only effective means for ablating significant residual tumor tissue derived from nonfunctioning tumors. PRL-, GH-, and ACTH-secreting tumor tissues are also amenable to medical therapy.

Side Effects In the short term, radiation may cause transient nausea and weakness. Alopecia and loss of taste and smell may be more long-lasting. Failure of pituitary hormone synthesis is common in patients who have undergone head and neck or pituitary-directed irradiation. More than 50% of patients develop failure of GH, ACTH, TSH, and/or gonadotropin secretion within 10 years, usually due to hypothalamic damage. Lifelong follow-up with testing of anterior pituitary hormone reserve is therefore necessary after radiation treatment. Optic nerve damage with impaired vision due to optic neuritis is reported in about 2% of patients who undergo pituitary irradiation. Cranial nerve damage is uncommon now that radiation doses are £2 Gy (200 rad) at any one treatment session and the maximum dose is <50 Gy (5000 rad). The advent of stereotactic radiotherapy may reduce damage to adjacent structures. The cumulative risk of developing a secondary tumor after conventional radiation is 1.3% after 10 years and 1.9% after 20 years.

Medical Medical therapy for pituitary tumors is highly specific and depends on tumor type. For prolactinomas, dopamine agonists are the treatment of choice. For acromegaly and TSH-secreting tumors, somatostatin analogues and, occasionally, dopamine agonists are indicated. ACTH-secreting tumors and nonfunctioning tumors are generally not responsive to medication and require surgery and/or irradiation.

PROLACTIN

SYNTHESIS

PRL consists of 198 amino acids and has a molecular mass of 21,500 kDa; it is weakly homologous to GH and human placental lactogen (hPL), reflecting the duplication and divergence of a common GH-PRL-hPL precursor gene on chromosome 6. PRL is synthesized in lactotropes, which comprise about 20% of anterior pituitary cells. Lactotropes and somatotropes are derived from a common precursor cell that may give rise to a tumor secreting both PRL and GH. Marked lactotrope cell hyperplasia develops during the last two trimesters of pregnancy and the first few months of lactation. These transient adaptive changes in the lactotrope population are induced by estrogen.

SECRETION

Fetal PRL synthesis begins at 12 weeks' gestation (about 4 weeks after GH). Normal adult serum PRL levels are about 10 to 25 ug/L in women and 10 to 20 ug/L in men. PRL secretion is pulsatile, with the highest secretory peaks occurring during rapid eye movement sleep. Peak serum PRL levels (up to 30 ug/L) occur between 4:00 and 6:00 A.M. The circulating half-life of PRL is about 50 min.

PRL is unique among the pituitary hormones in that the predominant central control mechanism is inhibitory, reflecting dopamine-mediated suppression of PRL release. This regulatory pathway is exemplified by the spontaneous PRL hypersecretion that occurs after pituitary stalk section, often a consequence of mass lesions at the skull

base.

Dopamine action is mediated by multiple receptor subtypes, each a member of the seven-transmembrane G protein-coupled receptor (GPCR) superfamily. In the pituitary, dopamine type 2 (D$_2$) receptors are predominant and mediate PRL inhibition. Targeted disruption (gene knockout) of the murine D$_2$ receptor results in hyperprolactinemia and lactotrope proliferation. Activation of D$_2$ receptors inhibits the cyclic AMP pathway, causing membrane hyperpolarization and closing of voltage-gated calcium channels; these events block secretory granule exocytosis by reducing intracellular free calcium. Because of the potent PRL inhibitory effects of dopamine, physiologic, pharmacologic, or pathologic alterations in dopamine action increase PRL levels. As discussed below, dopamine agonists play a central role in the management of hyperprolacinemic disorders.

TRH (pyro Glu-His-Pro-NH2) is a hypothalamic tripeptide that releases prolactin within 15 to 30 min after intravenous injection. The physiologic relevance of TRH for PRL regulation is unclear, as it appears to primarily regulate TSH (Chap. 330). *Vasoactive intestinal peptide* (VIP) also induces PRL release, whereas glucocorticoids and thyroid hormone suppress PRL secretion.

Serum PRL levels rise after exercise, meals, sexual intercourse, minor surgical procedures, general anesthesia, acute myocardial infarction, and other forms of acute stress. PRL levels also increase significantly (~tenfold) during pregnancy and decline rapidly within 2 weeks of parturition. If breastfeeding is initiated, basal PRL levels remain elevated; suckling stimulates reflex increases in PRL levels that last for about 30 to 45 min. Breast suckling activates neural afferent pathways in the hypothalamus that induce PRL release. With time, the suckling-induced responses diminish and interfeeding PRL levels return to normal.

ACTION

The PRL receptor is a member of the type I cytokine receptor family that also includes GH and interleukin (IL) 6 receptors. Ligand binding leads to receptor dimerization followed by intracellular signaling mediated by the Janus kinase (JAK) pathway, which phosphorylates components of the signal transduction and activators of transcription (STAT) family. The STAT proteins translocate to the nucleus, where they act as transcription factors on target genes. In the breast, the lobuloalveolar epithelium proliferates in response to PRL, placental lactogens, elevated progesterone, and local paracrine growth factors; lactogenesis occurs as a result of complex multihormonal interactions (Chap. 337).

PRL acts to induce and maintain lactation, decrease reproductive function, and suppress sexual drive. These functions are geared towards ensuring that maternal lactation is sustained and not interrupted by pregnancy. PRL inhibits reproductive function at multiple levels, including suppression of hypothalamic GnRH and pituitary gonadotropin secretion, as well as impairing gonadal steroidogenesis in both female and male subjects. In the hypothalamus, PRL-mediated suppression of GnRH leads to loss of pulsatile LH secretion and abrogation of the preovulatory LH surge. In the ovary, PRL blocks folliculogenesis and inhibits granulosa cell aromatase activity, leading to

hypoestrogenism and anovulation. PRL also has a luteolytic effect, generating a shortened, or inadequate, luteal phase of the menstrual cycle. In males, attenuated LH secretion leads to low testosterone levels and decreased spermatogenesis. These hormonal changes decrease libido and reduce fertility in patients with hyperprolactinemia (Chap. 54).

PRL exerts widespread metabolic effects to ensure maintenance of sustained lactation. Gastrointestinal calcium absorption is increased, bone calcium is mobilized, bile acids are elevated, and pancreatic bcell growth is induced by PRL. Centrally, PRL acts on brain centers involved in parenting behavior, appetite stimulation, and analgesia. Concomitant with these effects, maintenance of bone mineral density is abrogated; hyperprolactinemia is associated with enhanced risk for bone loss and the long-term development of osteoporosis. PRL receptors are abundant in the osteoblasts of developing bone, and the accompanying hypoestrogenemia contributes to accelerated bone loss in hyperprolactinemic women.

HYPERPROLACTINEMIA

Etiology Hyperprolactinemia is the most common pituitary hormone hypersecretion syndrome in both males and females. PRL-secreting pituitary adenomas (prolactinomas) are the most common cause of PRL levels>100 ug/L (see below). Less pronounced PRL elevation can also be caused by microprolactinomas but is more commonly caused by drugs, pituitary stalk compression, hypothyroidism, or renal failure (Table 328-8).

Pregnancy and lactation are the important physiologic causes of hyperprolactinemia. Sleep-associated hyperprolactinemia reverts to normal within an hour of awakening. Nipple stimulation and sexual orgasm may also cause acute PRL increases. Chest wall stimulation or trauma (including chest surgery and herpes zoster) invoke the reflex suckling arc with resultant hyperprolactinemia. Chronic renal failure elevates PRL by decreasing peripheral PRL clearance. Primary hypothyroidism is associated with mild hyperprolactinemia, probably because of enhanced TRH secretion.

Lesions of the hypothalamic-pituitary region that disrupt hypothalamic dopamine synthesis, portal vessel delivery, or lactotrope responses are associated with hyperprolactinemia. Thus, hypothalamic tumors, cysts, infiltrative disorders, and radiation-induced damage cause elevated PRL levels, usually in the range of 30 to 100 ug/L. Plurihormonal adenomas (including GH and ACTH tumors) may directly hypersecrete PRL. Clinically nonfunctioning pituitary tumors commonly cause stalk pressure and hyperprolactinemia.

Drug-induced inhibition or disruption of dopaminergic receptor function results in hyperprolactinemia (Table 328-8). Thus, many antipsychotics and antidepressants cause hyperprolactinemia. Methyldopa inhibits dopamine synthesis and verapamil blocks dopamine release, also leading to hyperprolactinemia. Hormonal agents that induce PRL include estrogens, antiandrogens, and TRH.

Presentation and Diagnosis Amenorrhea, galactorrhea, and infertility are the hallmarks of hyperprolactinemia in women. If hyperprolactinemia develops prior to the menarche, primary amenorrhea results. More commonly, hyperprolactinemia develops

later in life and leads to oligomenorrhea and, ultimately, to amenorrhea. Patients present with infertility, vaginal dryness, dyspareunia, and loss of libido. If hyperprolactinemia is sustained, vertebral bone mineral density can be reduced compared to age-matched controls, particularly when associated with pronounced hypoestrogenemia. Galactorrhea is present in up to 80% of hyperprolactinemic women. Though usually bilateral and spontaneous, it may be unilateral or only expressed manually. Patients may also complain of weight gain and mild hirsutism.

In men with hyperprolactinemia, diminished libido or visual loss (from optic nerve compression) are the usual presenting symptoms. Gonadotropin suppression leads to reduced testosterone, impotence, and oligospermia. True galactorrhea is uncommon in men with hyperprolactinemia. If the disorder is longstanding, secondary effects of hypogonadism are evident, including osteopenia, reduced muscle mass, and decreased beard growth.

The diagnosis of idiopathic hyperprolactinemia is made by exclusion of known causes of hyperprolactinemia in the setting of a normal pituitary MRI. Some of these patients may have small microadenomas below MRI sensitivity (~2 mm).

Laboratory Investigation Basal, fasting morning PRL levels (normally <20 ug/L) should be measured to assess hypersecretion. Because hormone secretion is pulsatile and levels vary widely in some individuals with hyperprolactinemia, it may be necessary to measure levels on several different occasions when clinical suspicion is high. Both false-positive and false-negative results may be encountered. In patients with markedly elevated PRL levels (>1000 ug/L), results may be falsely lowered because of assay artifacts; sample dilution is required to assess these high values accurately. Falsely elevated values may be caused by aggregated forms of circulating PRL, which are biologically inactive (macroprolactinemia). Hypothyroidism should be excluded by measuring TSH and T$_4$ levels.

TREATMENT

Treatment of hyperprolactinemia depends on the cause of elevated PRL levels. Regardless of the etiology, however, treatment should be aimed at normalizing PRL levels to alleviate suppressive effects on gonadal function, halt the galactorrhea, and preserve bone mineral density. Dopamine agonists are effective for many different causes of hyperprolactinemia (see "Treatment" for "Prolactinoma," below).

If the patient is taking a medication known to cause hyperprolactinemia, the drug should be withdrawn, if possible. For psychiatric patients who require neuroleptic agents, dose titration or the addition of a dopamine agonist can help restore normoprolactinemia and alleviate reproductive symptoms. However, dopamine agonists sometimes worsen the underlying psychiatric condition, especially at high doses. Hyperprolactinemia usually resolves after adequate thyroid hormone replacement in hypothyroid patients or after renal transplantation in patients or dialysis. Resection of hypothalamic or sellar mass lesions can reverse hyperprolactinemia caused by reduced dopamine tone. Granulomatous infiltrates rarely respond to glucocorticoid administration. In patients with irreversible hypothalamic damage, no treatment may be warranted. In up to 30% of patients with hyperprolactinemia -- with or without a visible pituitary microadenoma --

the condition resolves spontaneously.

PROLACTINOMA

Etiology and Prevalence Tumors arising from lactotrope cells account for about half of all functioning pituitary tumors, with an annual incidence of ~3/100,000 population. Mixed tumors secreting combinations of GH and PRL, ACTH and PRL, and rarely TSH and PRL, are also seen. These plurihormonal tumors are usually recognized by immunohistochemistry, without apparent clinical manifestations from the production of additional hormones. Microadenomas are classified as <1 cm in diameter and do not usually invade the parasellar region. Macroadenomas are >1 cm in diameter, are locally invasive, and may impinge on adjacent structures. The female:male ratio for microprolactinomas is 20:1, whereas the gender ratio is near 1:1 for macroadenomas. Tumor size generally correlates directly with PRL concentrations; values >100 ug/L are usually associated with macroadenomas. Males tend to present with larger tumors than females, possibly because the features of hypogonadism are less readily evident. PRL levels remain stable in most patients, reflecting the slow growth of these tumors. About 5% of microadenomas progress in the long term to macroadenomas. Hyperprolactinemia resolves spontaneously in about 30% of microadenomas.

Presentation and Diagnosis Women usually present with amenorrhea, infertility, and galactorrhea. If the tumor extends outside of the sella, visual field defects or other mass effects may be seen. Men often present with impotence, loss of libido, infertility, or signs of central CNS compression including headaches and visual defects. Assuming that known physiologic and medication-induced causes of hyperprolactinemia are excluded (Table 328-8), the diagnosis of prolactinoma is likely with a PRL level >100 ug/L. PRL levels <100ug/L may be caused by microadenomas, other sellar lesions that decrease dopamine inhibition, or nonneoplastic causes of hyperprolactinemia. For this reason, an MRI should be performed in all patients with hyperprolactinemia. It is important to remember that hyperprolactinemia caused by the mass effects of nonlactotrope lesions is also corrected by treatment with dopamine agonists. Consequently, PRL suppression by dopamine agonists does not necessarily indicate that the lesion is a prolactinoma.

TREATMENT

As microadenomas rarely progress to become macroadenomas, no treatment may be needed if fertility is not desired. Estrogen replacement is indicated to prevent bone loss and other consequences of hypoestrogenemia and does not appear to increase the risk of tumor enlargement. These patients should be monitored by regular serial PRL and MRI measurements.

For symptomatic microadenomas, therapeutic goals include control of hyperprolactinemia, reduction of tumor size, restoration of menses and fertility, and improvement of galactorrhea. Dopamine agonists should be titrated to achieve maximal PRL suppression and restoration of reproductive function (Fig. 328-7). A normalized PRL level does not assure reduced tumor size. However, tumor shrinkage is not usually seen in those who do not respond with lowered PRL levels. For macroadenomas, formal visual field testing should be performed before initiating dopamine agonists. MRI and visual fields should be assessed at 6- to 12-month intervals

until the mass shrinks and annually thereafter until maximum size reduction has occurred.

Medical Oral dopamine agonists (cabergoline or bromocriptine) are the mainstay of therapy for patients with micro- or macroprolactinomas. Dopamine agonists suppress PRL secretion and synthesis as well as lactotrope cell proliferation.

Bromocriptine The ergot alkaloid bromocriptine mesylate is a dopamine receptor agonist that suppresses prolactin secretion by binding directly to lactotrope D_2 dopamine receptors. Bromocriptine is used as initial therapy for both micro- and macroprolactinomas. In microadenomas the drug rapidly lowers serum prolactin levels to normal in up to 70% of patients, decreases tumor size, and restores gonadal function. In patients with macroadenomas, prolactin levels are also normalized in 70% of patients and tumor mass shrinkage (^{3}50%) is achieved in up to 40% of patients. Mass effect symptoms, including headaches and visual disorders, usually improve dramatically within days after bromocriptine initiation; improvement of sexual function requires several weeks of treatment but may occur before complete normalization of prolactin levels. Drug withdrawal usually results in recurrent hyperprolactinemia and tumor reexpansion, with the risk of visual compromise. After initial control of PRL levels has been achieved, bromocriptine should be reduced to the lowest effective maintenance dose. In ~5% of treated patients, hyperprolactinemia may resolve and not recur when bromocriptine is discontinued after long-term treatment.

Therapy is initiated by administering a low bromocriptine dose (0.625 to 1.25 mg) at bedtime with a snack, followed by gradually increasing the dose. Most patients are successfully controlled with a daily dose of £7.5 mg (2.5 mg tid). About 20% of patients are resistant to dopaminergic treatment; they may have decreased D_2 dopamine receptor numbers or a postreceptor defect. D_2 receptor gene mutations in the pituitary have not been reported.

Nausea, vomiting, and postural hypotension with faintness may occur in ~25% of patients after the initial dose. These symptoms may persist in some patients. Other side effects include constipation, nasal stuffiness, dry mouth, nightmares, insomnia, and vertigo; decreasing the dose usually alleviates these problems. For the approximately 15% of patients who cannot tolerate oral bromocriptine, intravaginal administration of tablets is often efficacious.

Auditory hallucinations, delusions, and mood swings have been reported in up to 5% of patients and may be due to the dopamine agonist properties or to the lysergic acid derivative of the compound. Rare reports of leukopenia, thrombocytopenia, pleural fibrosis, cardiac arrhythmias, and hepatitis have been described.

Cabergoline An ergoline derivative, cabergoline is a long-acting dopamine agonist with high D_2 receptor affinity. The drug effectively suppresses PRL for>14 days after a single oral dose and induces prolactinoma shrinkage in most patients. Cabergoline (0.5 to 1.0 mg twice weekly) achieves normoprolactinemia and resumption of normal gonadal function in ~80% of patients with microadenomas; galactorrhea improves or resolves in 90% of patients. Cabergoline normalizes PRL and shrinks ~70% of macroprolactinomas. It may also be effective in patients resistant to bromocriptine.

Adverse effects and drug intolerance are encountered less commonly than with bromocriptine.

Other dopamine agonists These include *pergolide mesylate*, an ergot derivative with dopaminergic properties; *lisuride*, an ergot derivative; and *quinagolide* (CV 205-502, Norprolac), a nonergot oral dopamine agonist with specific D_2 receptor activity.

Surgery Indications for surgical debulking include dopamine resistance or intolerance and the presence of an invasive macroadenoma with compromised vision that fails to improve rapidly after drug treatment. Initial PRL normalization is achieved in about 70% of microprolactinomas after surgical resection, but only 30% of macroadenomas can be successfully resected. However, follow-up studies have shown that recurrence of hyperprolactinemia occurs in up to 20% of patients within the first year after surgery; long-term recurrence rates exceed 50% for macroadenomas. Radiotherapy for prolactinomas is reserved for patients with aggressive tumors that do not respond to maximally tolerated dopamine agonists and/or surgery.

Pregnancy The pituitary increases in size during pregnancy, reflecting the stimulatory effects of estrogen and perhaps other growth factors. About 5% of microadenomas significantly increase in size, but 15 to 30% of macroadenomas may grow during pregnancy. Bromocriptine has been used for over 25 years to restore fertility in women with hyperprolactinemia, without evidence of untoward teratogenic effects. Nonetheless, most authorities recommend strategies to minimize fetal exposure to the drug. For women taking bromocriptine who desire pregnancy, mechanical contraception should be used through three regular menstrual cycles to allow for conception timing. When pregnancy is confirmed, bromocriptine should be discontinued and PRL levels followed serially, especially if headaches or visual symptoms occur. For women harboring macroadenomas, regular visual field testing is recommended, and the drug should be reinstituted if tumor growth is apparent. Although pituitary MRI may be safe during pregnancy, this procedure should be reserved for symptomatic patients with severe headache and/or visual field defects. Alternatively, surgical decompression may be indicated if vision is threatened. Though comprehensive data support the efficacy and relative safety of bromocriptine-facilitated fertility, patients should be advised of potential unknown deleterious effects and the risk of tumor growth during pregnancy. At present, the experience with cabergoline is too limited to recommend its routine use when fertility is desired.

GROWTH HORMONE

SYNTHESIS

GH is the most abundant anterior pituitary hormone and is expressed early in fetal life (at 8 weeks' gestation). GH-secreting somatotrope cells constitute up to 50% of the total anterior pituitary cell population. Mammosomatotrope cells, which coexpress PRL with GH, can be identified using double immunostaining techniques. Somatotrope development is determined by expression of the cell-specific Pit-1 nuclear transcription factor. In addition to controlling cell differentiation, it also enhances GH gene expression. Five distinct genes on chromosome 17q22 encode GH and related proteins. The pituitary GH gene (*hGH-N*) produces two alternatively spliced products that give

rise to 22-kDa GH (191 amino acids) and a less abundant, 20-kDa GH molecule, with similar biologic activity. Placental syncytiotrophoblast cells express a GH variant (*hGH-V*) gene; the related hormone human chorionic somatotropin (HCS) is expressed by distinct members of the gene cluster. HCS shares high homology with GH yet exhibits minimal growth-promoting properties.

SECRETION

GHsecretion is controlled by complex hypothalamic and peripheral factors.*GHRH*is a 44 amino acid hypothalamic peptide that stimulates GH synthesis and release. Synthetic agonists of the*GHRP*receptor stimulate GHRH and also directly stimulate GH release, but putative endogenous agonists remain incompletely characterized. *Somatostatin* (SRIF) is synthesized in the medial preoptic area of the hypothalamus and inhibits GH secretion. GHRH is secreted as discrete spikes that elicit GH pulses, whereas SRIF sets basal GH tone. SRIF is also expressed in many extrahypothalamic tissues, including theCNS, gastrointestinal system, and pancreas, where it also acts to inhibit the hormone secretion.*IGF*-I, the peripheral target hormone for GH, feeds back to inhibit GH; estrogens induce GH (Chap. 8), whereas glucocorticoid excess suppresses GH release.

Two distinct surface receptors on the somatotrope regulateGHsynthesis and secretion. TheGHRHreceptor is aGPCRthat signals through the intracellular cyclic AMP pathway. Activation of this receptor stimulates somatotrope cell proliferation as well as hormone production. Inactivating mutations of the GHRH receptor cause profound dwarfism (see below). A distinct surface receptor forGHRP has also been identified. This receptor is expressed in the hypothalamus and pituitary. A natural ligand, termed *ghrelin*, binds to the GHRP receptor; it is produced in large amounts in the stomach, though its physiologic role remains unknown. Hypothalamic somatostatin binds to five distinct receptor subtypes (SSTR1 to SSTR5) that are widely expressed in different tissues, including in the pituitary. SSTR2 and SSTR5 subtypes preferentially suppress GH (andTSH) secretion.

GHsecretion is pulsatile, with greater levels at night generally correlating with the onset of sleep. GH secretory rates decline markedly with age so that hormone production in middle age is about 15% of production during puberty. These changes are paralleled by an age-related decline in lean muscle mass. GH secretion is also reduced in obese individuals, thoughIGF-I levels are preserved, suggesting a change in the setpoint for feedback control. Elevated GH levels occur within an hour of deep sleep onset as well as after exercise, physical stress, trauma, and during sepsis. Integrated 24-h GH secretion is higher in women and is also enhanced by estrogen replacement. Using assays in common clinical use, random GH measurements are undetectable in ~50% of daytime samples obtained from healthy subjects and are undetectable in most obese and elderly subjects. Thus, single random GH measurements do not distinguish patients with adult GH deficiency from normal persons.

GHsecretion is profoundly influenced by nutritional factors. Using newer ultrasensitive chemiluminescence-based GH assays with a sensitivity of 0.002 ug/L, a glucose load can be shown to suppress GH to<0.7 ug/L in female and to<0.07 ug/L in male subjects. Increased GH pulse frequency and peak amplitudes occur with chronic malnutrition or

prolonged fasting. GH is stimulated by high-protein meals and by L-arginine. GH secretion is induced by dopamine and apomorphine (a dopamine receptor agonist), as well as by a-adrenergic pathways.b-Adrenergic blockage induces basal GH and enhancesGHRH- and insulin-evoked GH release.

ACTION

The pattern ofGHsecretion may affect tissue responses. The higher GH pulsatility observed in males, as compared to the relatively continuous GH secretion in females, may be an important biologic determinant of linear growth patterns and liver enzyme induction.

The 70-kD peripheralGHreceptor protein shares structural homology with the cytokine/hematopoietic superfamily. A fragment of the receptor extracellular domain generates a soluble GH binding protein (GHBP) that interacts with GH in the circulation. The liver contains the greatest number of GH receptors. GH binding induces receptor dimerization by making distinct contact through two separate binding domains of the hormone. The dimerized receptor interacts with members of theJAK/STATfamily. The activated STAT proteins translocate to the nucleus, where they modulate expression of GH-regulated target genes. GH analogues that bind to the receptor, but are incapable of mediating receptor dimerization, are potent antagonists of GH action and are being investigated for potential use in the treatment of acromegaly and diabetic microangiopathy.

GHinduces protein synthesis and nitrogen retention and impairs glucose tolerance by antagonizing insulin action. GH also stimulates lipolysis, leading to increased circulating fatty acid levels, reduced omental fat mass, and enhanced lean body mass. GH promotes sodium, potassium, and water retention and elevates serum levels of inorganic phosphate. Linear bone growth occurs as a result of complex hormonal and growth factor actions, including those ofIGF-I. GH stimulates epiphyseal prechondrocyte differentiation. These precursor cells produce IGF-I locally and are also responsive to the growth factor.

INSULIN-LIKE GROWTH FACTORS

ThoughGHexerts direct effects in target tissues, many of its physiologic effects are mediated indirectly throughIGF-I, a potent growth and differentiation factor. The major source of circulating IGF-I is hepatic in origin. Peripheral tissue IGF-I exerts local paracrine actions that appear to be both dependent and independent of GH. Thus, GH administration induces circulating IGF-I level as well as stimulating IGF-I expression in multiple tissues.

BothIGF-I and -II are bound to one of six high-affinity circulating IGF-binding proteins (IGFBPs) that regulate IGF bioactivity. Levels of IGFBP3 areGH-dependent, and it serves as the major carrier protein for circulating IGF-I. GH deficiency and malnutrition are associated with low IGFBP3 levels. IGFBP1 and -2 regulate local tissue IGF action but do not bind appreciable amounts of circulating IGF-I.

SerumIGF-I concentrations are profoundly affected by various physiologic factors.

Levels increase during puberty, peak at 16 years, and subsequently decline by>80% during the aging process. IGF-I concentrations are higher in females than in males. BecauseGH is the major determinant of hepatic IGF-I synthesis, abnormalities of GH synthesis or action (e.g., pituitary failure,GHRHreceptor defect, or GH receptor defect) reduce IGF-I levels. Hypocaloric states are associated with GH resistance; IGF-I levels are therefore low with cachexia, malnutrition, and sepsis. In acromegaly, IGF-I levels are invariably high and reflect a log-linear relationship with GH concentrations.

IGF-I Physiology ThoughIGF-I is not an approved drug, investigational studies provide insight into its physiologic effects. High doses of injected IGF-I (100 ug/kg) induce hypoglycemia, primarily because of actions through the insulin receptor. Low IGF-I doses improve insulin sensitivity in patients with severe insulin resistance and diabetes. In cachectic subjects, IGF-I infusion (12 ug/kg per hour) enhances nitrogen retention and lowers cholesterol levels. Longer-term subcutaneous IGF-I injections exert a marked anabolic effect with enhanced protein synthesis. The impact of long-term IGF-I administration on bone mineral content is as yet unclear. Although bone formation markers are induced, bone turnover may also be stimulated by IGF-I.

Side effects ofIGF-I are dose-dependent. An acute overdose may result in hypoglycemia and hypotension. Fluid retention, temporomandibular jaw pain, and increased intracranial pressure are reversible. Avascular necrosis of the femoral head has been reported. Chronic excess IGF-I would presumably result in features of acromegaly.

DISORDERS OF GROWTH AND DEVELOPMENT

Skeletal Maturation and Somatic Growth Linear growth is a function of endochrondral bone formation whereby cartilage is converted into bony skeleton in the long bones and vertebrae (Chap. 340). Ossification occurs within central diaphyseal and peripheral epiphyseal centers. A cartilaginous growth plate forms between the two centers, and chondrocytes proliferate within the growth plate. Linear bone growth ceases when this cartilage layer ossifies and fuses with epiphyseal and diaphyseal bone.

The growth plate is dependent on a variety of hormonal stimuli includingGH,IGF-I, sex steroids, thyroid hormones, paracrine growth factors, and cytokines. GH directly stimulates prechondrocyte differentiation and clonal expansion, resulting in chondrocytes that express both IGF-I receptors and IGF-I protein. The growth-promoting process also requires caloric energy, amino acids, vitamins, and trace metals and consumes about 10% of normal energy production. Malnutrition impairs chondrocyte activity and reduces circulating IGF-I andIGFBP3 levels.

Bone age is delayed in patients with all forms of trueGHdeficiency or GH receptor defects that result in attenuated GH action. Rarely, GH excess accelerates growth, particularly in the setting of delayed bone age from concomitant hypogonadism. Thyroid hormone is permissive for GH synthesis and secretion as well as for maintaining normal circulatingIGF-I and binding protein levels. Bone age is delayed by thyroid hormone deficiency. Consequently, congenital or acquired hypothyroidism is associated with stunted growth, which is partially reversed by thyroid hormone replacement (Chap. 330). Elevated pubertal sex steroid levels (especially estrogen) induce theGHRH-GH-IGF-I

axis and also directly stimulate epiphyseal growth. High doses of estrogen lead to epiphyseal closure. A mutation of the estrogen receptora prevented epiphyseal closure, confirming the important role of this pathway in bone maturation. Several pathologic conditions accompanied by increased levels of sex steroids, including precocious puberty, androgen exposure (exogenous or endogenous), congenital adrenal hyperplasia, and obesity, are associated with accelerated bone maturation. Thus, children with these conditions have accelerated early growth, but end up with reduced final height. In contrast to sex steroids, glucocorticoid excess inhibits linear growth. Glucocorticoids also stimulate SRIF and inhibit peripheral GH and IGF-I receptor signaling.

Linear bone growth rates are very high in infancy and are pituitary-dependent. Mean growth velocity is ~6 cm/year in later childhood and is usually maintained within a given range on a standardized percentile chart. Peak growth rates occur during midpuberty when bone age is 12 (girls) or 13 (boys) (Chap. 8). Secondary sexual development is associated with elevated sex steroids that cause progressive epiphyseal growth plate closure.

Short stature may occur as a result of constitutive intrinsic growth defects or because of acquired extrinsic factors that impair growth (Table 328-9). Genetic disorders, including pituitary transcription factor defects, mutations in growth-related genes, and pituitary hypoplasia syndromes, may all be associated with growth delay and short stature. In general, delayed bone age in a child with short stature is suggestive of a hormonal or systemic disorder, whereas normal bone age in a short child is more likely to be caused by a genetic growth plate disorder (Chap. 351). Other bone and cartilage dysplasia syndromes are associated with specific limb-body proportion phenotypes, and some involve associated calcium disorders (Chap. 343).

Intrauterine growth retardation results in short stature and may be caused by specific congenital anomalies (e.g.,IGF-I deficiency, *Russell-Silver syndrome*, chromosomal disomy) or maternal factors such as diabetes mellitus, infections, hypoxia, drug addiction, or placental dysfunction. Long-term responses of these children toGHtreatment are currently being evaluated.

Turner syndrome is caused by loss of all, or part, of an X chromosome in females (XO). It is characterized by short stature, in addition to gonadal dysgenesis and other characteristic features (Chap. 338). Short stature may be improved with a combination ofGH and an anabolic steroid (oxandrolone); estrogen is required to induce and sustain sexual development. *Noonan syndrome* resembles Turner syndrome phenotypically, but patients have apparently normal sex chromosomes. These patients have delayed pubertal development but not primary gonadal failure.

GH Deficiency in Children

GH Deficiency Isolated GH deficiency is characterized by short stature, micropenis, increased fat, high-pitched voice, and a propensity to hypoglycemia. The etiology of GH deficiency is not identifiable in most children with the disorder. Familial modes of inheritance are seen in one-third of these individuals and may be autosomal dominant, recessive, or X-linked, indicating that multiple genetic abnormalities can lead to GH

deficiency. About 10% of children with growth hormone deficiency have mutations in the GH-N gene. These include gene deletions and a wide range of point mutations, including some that function in a dominant negative manner (heterozygous mutations) to impair the synthesis or function of GH expressed from the normal allele. Mutations in transcription factors Pit-1 and Prop-1, which control somatotrope development, cause GH deficiency in combination with other pituitary hormone deficiencies. The diagnosis of *idiopathic GH deficiency* (IGHD) should be made only after known molecular defects have been excluded.

GHRHReceptor Mutations Recessive mutations of the GHRH receptor gene have been described in several unrelated families with severe proportionate dwarfism. The low basalGHlevels in these patients cannot be stimulated by exogenous GHRH,GHRP, or insulin-induced hypoglycemia, confirming the importance of the GHRH receptor for somatotrope cell proliferation and hormonal responsiveness.

Growth Hormone Insensitivity This is caused by defects ofGHreceptor structure or signaling. Homozygous or heterozygous exonic and intronic mutations of the GH receptor occur mainly in the extracellular ligand-binding domain and are associated with partial or complete GH insensitivity and growth failure (*Laron syndrome*). The diagnosis of this syndrome is based on normal or high GH levels, with decreased circulatingGHBP, and lowIGF-I levels. Very rarely, defective IGF-I, IGF-I receptor, or IGF-I signaling defects are also encountered.

Nutritional Short Stature Caloric deprivation and malnutrition, uncontrolled diabetes, and chronic renal failure represent secondary causes of abrogatedGHreceptor function. These conditions also stimulate the production of proinflammatory cytokines, including tumor necrosis factor (TNF)a andILs, which can block GH-mediated signal transduction. Children with these conditions typically exhibit features of acquired short stature with elevated GH and lowIGF-I levels. Circulating GH receptor antibodies may rarely cause peripheral GH insensitivity.

Psychosocial Short Stature Emotional and social deprivation lead to growth retardation accompanied by delayed speech, discordant hyperphagia, and attenuated response to administeredGH. A nurturing environment restores growth rates.

Presentation and Diagnosis Short stature is commonly encountered in clinical practice, but the criteria for biochemical diagnosis of trueGHdeficiency have been difficult to define. The decision to evaluate these children requires clinical judgement in association with auxologic data and family history. Short stature should be comprehensively evaluated if a patient's height is >3 SD below the mean for age or if the growth rate has decelerated. Skeletal maturation is best evaluated by measuring a radiologic bone age, which is based mainly on the degree of growth plate fusion. Final height can be predicted using standardized scales (Bayley-Pinneau or Tanner-Whitehouse) or estimated by adding 6.5 cm (boys) or subtracting 6.5 cm (girls) from the midparental height.

Laboratory Investigation BecauseGHsecretion is pulsatile, GH deficiency is best assessed by examining the response to provocative stimuli. Random GH measurements do not distinguish normal children from those with true GH deficiency.

Adequate adrenal and thyroid hormone replacement should be assured before testing. Provocative stimuli such as exercise, insulin-induced hypoglycemia, and other pharmacologic tests normally increase GH to >7 ug/L in children. Insulin-induced hypoglycemia testing requires the blood sugar nadir to be<50% of baseline levels. This test should be performed under close supervision and is contraindicated in children with seizure disorders. IGF-I levels are not sufficiently sensitive or specific to make the diagnosis but can be useful to confirm GH deficiency; they must be controlled for age and gender. Pituitary MRI may reveal pituitary mass lesions or structural defects.

TREATMENT

Replacement therapy with recombinant GH (0.02 to 0.05 mg/kg per day subcutaneously) restores growth velocity in GH-deficient children to ~10 cm/year. If pituitary insufficiency is documented, other associated hormone deficits should be corrected -- especially adrenal steroids. GH treatment is also moderately effective for accelerating growth rates in patients with Turner syndrome and chronic renal failure.

In patients with GH insensitivity and growth retardation due to mutations of the GH receptor, treatment with IGF-I bypasses the dysfunctional GH receptor. Growth rates have been maintained for several years, and this therapy now portends improved final adult stature in this group of patients.

ADULT GH DEFICIENCY (AGHD)

This disorder is usually caused by hypothalamic or pituitary somatotrope damage. Acquired pituitary hormone deficiency follows a typical sequential pattern whereby loss of adequate GH reserve foreshadows subsequent hormone deficits. The sequential order of hormone loss is usually GH ® FSH/LH ® TSH ® ACTH. The presence of documented central hypogonadism, hypothyroidism, and/or hypoadrenalism invariably assures the presence of GH deficiency. Conversely, ~40% of patients with incipient preclinical pituitary insufficiency already manifest GH deficiency, if rigorously tested.

Presentation and Diagnosis The clinical features of AGHD include changes in body composition, lipid metabolism, and quality of life and cardiovascular dysfunction (Table 328-10). Body composition changes are common and include reduced lean body mass, increased fat mass with selective deposition of intraabdominal visceral fat, and increased waist-to-hip ratio. Hyperlipidemia, left ventricular dysfunction, hypertension, and increase plasma fibrinogen levels may also be present. Bone mineral content is reduced, with resultant increased fracture rates. Patients may exhibit social isolation, depression, and difficulty in maintaining gainful employment. Adult hypopituitarism is associated with a three-fold increased cardiovascular mortality rate in comparison to age- and sex-matched controls, and this may be due to GH deficiency.

Laboratory Investigation AGHD is rare, and in light of the nonspecific nature of associated clinical symptoms, patients appropriate for testing should be carefully selected on the basis of well-defined criteria. With few exceptions, testing should be restricted to patients with the following predisposing factors: (1) pituitary surgery, (2) pituitary or hypothalamic tumor or granulomas, (3) cranial irradiation, (4) radiologic evidence of a pituitary lesion, (5) childhood requirement for GH replacement therapy, or,

rarely, (6) unexplained low age- and sex-matched IGF-I level. The transition of the GH-deficient adolescent to adulthood requires retesting to document adult GH deficiency. Up to 20% of patients treated for childhood-onset GH deficiency are found to be GH-sufficient on repeat testing as adults.

A significant proportion (~25%) of truly GH-deficient adults have low-normal IGF-I levels. Thus, as in the evaluation of GH deficiency in children, valid age- and gender-matched IGF-I measurements provide a useful index of therapeutic responses but are not sufficiently sensitive for diagnostic purposes. AGHD is diagnosed by demonstrating a subnormal GH response (<3 ug/L) to a standard provocative test. None of the available stimulation tests provides standardized GH responses that clearly discriminate normal subjects from truly GH-deficient adults. The age-related decline of GH blurs this distinction further in elderly individuals. The most validated test to distinguish pituitary-sufficient patients from those with AGHD is insulin-induced (0.05 to 0.1 U/kg) hypoglycemia. After glucose reduction to ~40 mg/dL, most individuals experience neuroglycopenic symptoms (Chap. 334), and peak GH release occurs at 60 min and remains elevated for up to 2 h. About 90% of healthy adults exhibit GH responses>5 ug/L; AGHD is defined by a peak GH response to hypoglycemia of <3 ug/L. An attenuated GH response is observed in patients with pituitary damage, obesity, untreated hypothyroidism, depression, or chronic renal failure. Although an *insulin tolerance test* (ITT) is safe when performed under appropriate supervision, it is contraindicated in patients with diabetes, ischemic heart disease, cerebrovascular disease, or epilepsy and in elderly patients. Alternative stimulatory tests include L-dopa (500 mg orally) and intravenous arginine (30 g), GHRH(1 ug/kg), and GHRP-6 (90 ug). Combinations of these tests may evoke GH secretion in subjects not responsive to a single test.

TREATMENT

Once the diagnosis of AGHD is unequivocally established, replacement of GH may be indicated. Contraindications to therapy include the presence of an active neoplasm, intracranial hypertension, or uncontrolled diabetes and retinopathy. The starting dose of 0.15 to 0.3 mg/d should be titrated (up to a maximum of 1.25 mg/d) to maintain IGF-I levels in the mid-normal range for age- and gender-matched controls (Fig. 328-8). Women require higher doses than men, and elderly patients require less GH. Long-term GH maintenance sustains normal IGF-I levels and is associated with persistent body composition changes (e.g., enhanced lean body mass and lower body fat). High-density lipoprotein cholesterol increases, but total cholesterol and insulin levels do not change significantly. Lumbar spine bone mineral density increases, but this response is gradual (>1 year). Many patients note significant improvement in quality of life when evaluated by standardized questionnaires. The effect of GH replacement on mortality rates in GH-deficient patients is currently the subject of long-term prospective investigation.

About 30% of patients exhibit reversible dose-related fluid retention, joint pain, and carpal tunnel syndrome, and up to 40% exhibit myalgias and paresthesias. Patients receiving insulin require careful monitoring for dosing adjustments, as GH is a potent counterregulatory hormone for insulin action. Patients with type 2 diabetes mellitus initially develop further insulin resistance. However, glycemic control improves with the sustained loss of abdominal fat associated with long-term GH replacement. Headache,

increased intracranial pressure, hypertension, atrial fibrillation, and tinnitus occur rarely. Prevalence of pituitary tumor regrowth and potential progression of skin lesions are currently being assessed in long-term surveillance programs. To date, development of these potential side effects does not appear significant. For some patients, the expense of long-term GH replacement is prohibitive.

ACROMEGALY

Etiology GH hypersecretion is usually the result of somatotrope adenomas but is also rarely caused by extrapituitary lesions (Table 328-11). In addition to typical GH-secreting somatotrope adenomas, mixed mammosomatotrope tumors and acidophilic stem-cell adenomas can secrete both GH and PRL. In patients with acidophilic stem-cell adenomas, features of hyperprolactinemia (hypogonadism and galactorrhea) predominate over the less clinically evident signs of acromegaly. Occasionally, mixed plurihormonal tumors are encountered that secrete ACTH, the glycoprotein hormone a subunit, or TSH, in addition to GH. Patients with partially empty sella may present with GH hypersecretion due to a small GH-secreting adenoma within the compressed rim of pituitary tissue; some of these may reflect the spontaneous necrosis of tumors that were previously larger. GH-secreting tumors rarely arise from ectopic pituitary tissue remnants in the nasopharynx or midline sinuses.

There are case reports of ectopic GH secretion by tumors of pancreatic, ovarian, or lung origin. Excess GHRH production may cause acromegaly because of chronic stimulation of somatotropes. These patients present with classic features of acromegaly, elevated GH levels, pituitary enlargement on MRI, and pathologic characteristics of pituitary hyperplasia. The most common cause of GHRH-mediated acromegaly is a chest or abdominal carcinoid tumor. Although these tumors usually express positive GHRH immunoreactivity, clinical features of acromegaly are evident in only a minority of patients with carcinoid disease. Excessive GHRH may also be elaborated by hypothalamic tumors, usually choristomas or neuromas.

Presentation and Diagnosis Protean manifestations of GH and IGF-I hypersecretion are indolent and often are not clinically diagnosed for 10 years or more. Acral bony overgrowth results in frontal bossing, increased hand and foot size, mandibular enlargement with prognathism, and widened space between the lower incisor teeth. In children and adolescents, initiation of GH hypersecretion prior to epiphyseal long bone closure is associated with the development of pituitary gigantism (Fig. 328-9). Soft tissue swelling results in increased heel pad thickness, increased shoe or glove size, ring tightening, characteristic coarse facial features, and a large fleshy nose. Other commonly encountered clinical features include hyperhidrosis, deep and hollow-sounding voice, oily skin, arthropathy, kyphosis, carpal tunnel syndrome, proximal muscle weakness and fatigue, acanthosis nigricans, and skin tags. Generalized visceromegaly occurs, including cardiomegaly, macroglossia, and thyroid gland enlargement.

The most significant clinical impact of GH excess occurs with respect to the cardiovascular system. Coronary heart disease, cardiomyopathy with arrhythmias, left ventricular hypertrophy, decreased diastolic function, and hypertension occur in about 30% of patients. Upper airway obstruction with sleep apnea occurs in about 60% of

patients and is associated with both soft tissue laryngeal airway obstruction and central sleep dysfunction. Diabetes mellitus develops in 25% of patients with acromegaly, and most patients are intolerant of a glucose load (as GH counteracts the action of insulin). Acromegaly is associated with an increased risk of colon polyps and colonic malignancy; polyps are diagnosed in up to one-third of acromegalic patients. Overall mortality is increased about three-fold and is due primarily to cardiovascular and cerebrovascular disorders, malignancy, and respiratory disease. Unless GH levels are controlled, survival is reduced by an average of 10 years compared with an age-matched control population.

Laboratory Investigation Age- and gender-matched serumIGF-I levels are invariably elevated in acromegaly. Consequently, an IGF-I level provides a useful laboratory screening measure when clinical features raise the possibility of acromegaly. Due to the pulsatility ofGHsecretion, measurement of a single random GH level is not useful for the diagnosis or exclusion of acromegaly and does not correlate with disease severity. The diagnosis of acromegaly is confirmed by demonstrating the failure of GH suppression to < 1 ug/L within 1 to 2 h of an oral glucose load (75 g). About 20% of patients exhibit a paradoxical GH rise after glucose. About 60% of patients with GH-secreting tumors may exhibit paradoxical GH responses toTRHadministration.PRLshould be measured as it is elevated in ~25% of patients with acromegaly. Thyroid function, gonadotropins, and sex steroids may be attenuated because of tumor mass effects. Because most patients will undergo surgery with glucocorticoid coverage, tests ofACTHreserve in asymptomatic patients are more efficiently deferred until after surgery.

TREATMENT

Surgical resection ofGH-secreting adenomas is the initial treatment for most patients (Fig. 328-10). Somatostatin analogues are used as adjuvant treatment for preoperative shrinkage of large invasive macroadenomas, immediate relief of debilitating symptoms, and reduction of GH hypersecretion, in elderly patients experiencing morbidity, in patients who decline surgery, or, when surgery fails, to achieve biochemical control. Irradiation or repeat surgery may be required for patients who cannot tolerate or do not respond to adjunctive medical therapy. The high rate of late hypopituitarism and the slow rate (5 to 15 years) of biochemical response are the main disadvantages of radiotherapy. Irradiation is relatively ineffective in normalizingIGF-I levels. Stereotactic ablation of GH-secreting adenomas by gamma-knife radiotherapy is promising, but long-term results are not available and the side effects have not been clearly delineated. Somatostatin analogues may be given while awaiting the full effect of radiotherapy. Systemic sequelae of acromegaly, including cardiovascular disease, diabetes, and arthritis, should also be managed aggressively. Maxillofacial surgery for mandibular repair may also be indicated.

Surgery Transsphenoidal surgical resection by an experienced surgeon is the preferred primary treatment for both microadenomas (cure rate ~70%) and macroadenomas (<50% cured). Soft tissue swelling improves immediately after tumor resection.GHlevels return to normal within an hour, andIGF-I levels are normalized within 3 to 4 days. In ~10% of patients, acromegaly may recur several years after apparently successful surgery; hypopituitarism develops in up to 15% of patients.

Somatostatin Analogues Somatostatin analogues exert their therapeutic effects through SSTR2 and -5 receptors, both of which are expressed by GH-secreting tumors. Octreotide acetate is an 8-amino-acid synthetic somatostatin analogue. In contrast to native somatostatin, the analogue is relatively resistant to plasma degradation. It has a 2-h serum half-life and possesses at least 40-fold greater potency than native somatostatin to suppress GH. These properties often allow effective pharmacologic control of GH hypersecretion without prior surgery or radiotherapy. Octreotide is administered by subcutaneous injection, beginning with 50 ug tid; the dose can be gradually increased up to 1500 ug/d. Fewer than 10% of patients do not respond to the analogue. Octreotide suppresses integrated GH levels to <5 ug/L in ~70% of patients and to <2 ug/L in up to 60% of patients. It normalizes IGF-I levels in ~75% of treated patients (Fig. 328-11). Prolonged use of the analogue is not associated with desensitization, even after ³10 years of treatment. Rapid relief of headache and soft tissue swelling occurs in ~75% of patients within days to weeks of treatment initiation. Subjective clinical benefits of octreotide therapy occur more frequently than biochemical remission, and most patients report symptomatic improvement, including amelioration of headache, perspiration, obstructive apnea, and cardiac failure. Modest pituitary tumor size reduction occurs in about 40% of patients, but this effect is reversed when treatment is stopped.

Two long-acting somatostatin depot formulations, octreotide and lanreotide, are becoming the preferred medical treatment for acromegalic patients. *Sandostatin-LAR* is a sustained-release, long-acting formulation of octreotide incorporated into microspheres that sustain drug levels for several weeks after intramuscular injection. GH suppression occurs for as long as 6 weeks after a 30-mg injection; long-term monthly treatment sustains GH and IGF-I suppression and reduction of pituitary tumor size. *Lanreotide*, a slow-release depot somatostatin preparation, is a cyclic somatostatin octapeptide analogue that suppresses GH and IGF-I hypersecretion for 10 to 14 days after a 30-mg intramuscular injection. Long-term administration controls GH hypersecretion in two-thirds of treated patients and improves patient compliance because of the long interval required between drug injections.

Side Effects Somatostatin analogues are well tolerated in most patients. Adverse effects are short-lived and mostly relate to drug-induced suppression of gastrointestinal motility and secretion. Nausea, abdominal discomfort, fat malabsorption, diarrhea, and flatulence occur in one-third of patients, though these symptoms usually remit within 2 weeks. Octreotide suppresses postprandial gallbladder contractility and delays gallbladder emptying; up to 30% of patients treated long-term develop echogenic sludge or asymptomatic cholesterol gallstones. Other side effects include mild glucose intolerance due to transient insulin suppression, asymptomatic bradycardia, hypothyroxinemia, and local pain at the injection site. The cost of chronic treatment may be prohibitive.

Dopamine Agonists Bromocriptine may suppress GH secretion in some acromegalic patients, particularly those with cosecretion of PRL. High doses (³20 mg/d), administered as three to four daily doses, are usually required to lower GH, and therapeutic efficacy is modest. GH levels are suppressed to <5 ug/L in ~20% of patients, and IGF-I levels are normalized in only 10% of patients. Cabergoline also suppresses GH and decreases adenoma size when given at a relatively high dose of 0.5 mg/d. Combined treatment

with octreotide and bromocriptine induces additive biochemical control compared to either drug alone.

GH Antagonists Investigational GH analogues antagonize endogenous GH action by blocking peripheral GH binding to its receptor. Consequently, serumIGF-I levels are suppressed, potentially reducing the deleterious effects of excess endogenous GH.

Radiation External radiation therapy or high-energy stereotactic techniques are used as adjuvant therapy for acromegaly. An advantage of radiation is that patient compliance with long-term treatment is not required. Tumor mass is reduced, andGHlevels are attenuated over time. However, 50% of patients require at least 8 years for GH levels to be suppressed to <5 ug/L; this suboptimal level of GH reduction is achieved in about 90% of patients after 18 years. Patients may require interim medical therapy for several years prior to attaining maximal radiation benefits. Most patients also experience hypothalamic-pituitary damage, leading to gonadotropin,ACTH, and/orTSHdeficiency within 10 years of therapy.

In summary, surgery is the preferred primary treatment forGH-secreting microadenomas (Fig. 328-10). The high frequency of GH hypersecretion after macroadenoma resection usually necessitates adjuvant or primary medical therapy for these larger tumors. Patients unable to receive or respond to medical treatment can be offered radiation.

ADRENOCORTICOTROPIN HORMONE (See also Chap. 331)

SYNTHESIS

ACTH-secreting corticotrope cells constitute about 20% of the pituitary cell population. ACTH (39 amino acids) is derived from thePOMCprecursor protein (266 amino acids) that also generates several other peptides, including b-lipotropin,b-endorphin, met-enkephalin,a melanocyte-stimulating hormone (MSH), and corticotropin-like intermediate lobe protein (CLIP). The POMC gene is located on chromosome 2 and possesses at least three different promoter regions that account for pituitary and peripheral tissue-specific POMC expression. A proximal promoter mediates POMC expression in corticotropes. The gonads, placenta, gastrointestinal tissues, liver, kidney, adrenal medulla, lung, and lymphocytic tissue express shorter POMC transcripts derived from a downstream promoter region. Tumors arising from peripheral neuroendocrine tissues that secrete ectopic ACTH express the longer form of POMC. The POMC gene is powerfully suppressed by glucocorticoids and induced by CRH, arginine vasopressin (AVP), and gp 130 proinflammatory cytokines, includingIL-6, and leukemia inhibitory factor.

CRH, a 41-amino-acid hypothalamic peptide synthesized in the paraventricular nucleus as well as in higher brain centers, is the predominant stimulator ofACTHsynthesis and release. The CRH receptor is aGPCRthat is expressed on the corticotrope. CRH signaling inducesPOMCtranscription and is mediated by cyclic AMP, as well as mitogen activated protein (MAP) kinase-activator protein-1 (AP-1) cascades.

SECRETION

ACTH secretion is pulsatile and exhibits a characteristic circadian rhythm, peaking at 6 A.M. and reaching a nadir about midnight. Adrenal glucocorticoid secretion, which is driven by ACTH, follows a parallel diurnal pattern. ACTH circadian rhymicity is determined by variations in secretory pulse amplitude rather than changes in pulse frequency. Superimposed on this endogenous rhythm, ACTH levels are increased by AVP, physical stress, exercise, acute illness, and insulin-induced hypoglycemia.

Loss of cortisol feedback inhibition, as occurs in primary adrenal failure, results in extremely high ACTH levels. Glucocorticoid-mediated negative regulation of the hypothalamo-pituitary-adrenal (HPA) axis occurs as a consequence of both hypothalamic CRH suppression and direct attenuation of pituitary POMC gene expression and ACTH release. Hypothalamic AVP stimulates the protein kinase C pathway and acts synergistically with CRH to enhance ACTH production.

Acute inflammatory or septic insults activate the HPA axis through the integrated actions of proinflammatory cytokines, bacterial toxins, and neural signals. The overlapping cascade of ACTH-inducing cytokines (TNF; IL-1, -2, and -6; and leukemia inhibitory factor) activates hypothalamic CRH and AVP secretion, pituitary POMC gene expression, and local paracrine pituitary cytokine networks. The resulting cortisol elevation restrains the inflammatory response and provides host protection. Concomitantly, cytokine-mediated central glucocorticoid receptor resistance impairs glucocorticoid suppression of the HPA. Thus, the neuroendocrine stress response reflects the net result of highly integrated hypothalamic, intrapituitary, and peripheral hormone and cytokine signals.

ACTION

The major function of the HPA axis is to maintain metabolic homeostasis and to mediate the neuroendocrine stress response. Peripheral and central afferent signals, which are integrated by the pituitary corticotrope cell, ultimately affect the pattern and quantity of adrenal cortisol secretion. ACTH induces cortical steroidogenesis by maintaining adrenal cell proliferation and function. The receptor for ACTH, designated *melanocortin-2 receptor*, is a GPCR that activates cyclic AMP and MAP kinase pathways; it induces steroidogenesis by stimulating a cascade of steroidogenic enzymes (Chap. 331).

ACTH DEFICIENCY

Presentation and Diagnosis Secondary adrenal insufficiency occurs as a result of pituitary ACTH deficiency. It is characterized by fatigue, weakness, anorexia, nausea, vomiting, and, occasionally, hypoglycemia (due to diminished insulin counterregulation). In contrast to primary adrenal failure, hypocortisolism associated with pituitary failure is not usually accompanied by pigmentation changes or mineralocorticoid deficiency. ACTH deficiency is commonly due to glucocorticoid withdrawal following treatment-associated suppression of the HPA axis. Isolated ACTH deficiency may occur after surgical resection of an ACTH-secreting pituitary adenoma that has suppressed the HPA axis; this phenomenon is suggestive of a surgical cure. The mass effects of other pituitary adenomas or sellar lesions may lead to ACTH deficiency, but usually in combination with other pituitary hormone deficiencies. Partial ACTH deficiency may be unmasked in the presence of an acute medical or surgical illness, when clinically

significant hypocortisolism reflects diminished ACTH reserve.

Laboratory Diagnosis Inappropriately low ACTH levels in the setting of low cortisol levels are characteristic of diminished ACTH reserve. Low basal serum cortisol levels are associated with blunted cortisol responses to ACTH provocative stimulation and impaired cortisol response to insulin-induced hypoglycemia, or testing with metyrapone or CRH. *For description of provocative ACTH tests, see "Tests of Pituitary-Adrenal Responsiveness" in Chap. 331.*

TREATMENT

Glucocorticoid replacement therapy improves most features of ACTH deficiency. The total daily dose of hydrocortisone replacement should not exceed 30 mg daily, divided into two or three doses. Prednisone (5 mg each morning; 2.5 mg each evening) is longer acting and has fewer mineralocorticoid effects than hydrocortisone. Some authorities advocate lower maintenance doses in an effort to avoid cushingoid side effects. Doses should be increased several-fold during periods of acute illness or stress.

CUSHING'S DISEASE (ACTH-PRODUCING ADENOMA) (See also Chap. 331)

Etiology and Prevalence Pituitary corticotrope adenomas account for 70% of patients with endogenous causes of Cushing's syndrome. However, it should be recalled that iatrogenic hypercortisolism is the most common cause of cushingoid features. Ectopic tumor ACTH production, cortisol-producing adrenal adenomas, carcinoma, and hyperplasia account for the other causes; rarely, ectopic tumor CRH production is encountered.

ACTH-producing adenomas account for about 10 to 15% of all pituitary tumors. Because the clinical features of Cushing's syndrome often lead to early diagnosis, most ACTH-producing pituitary tumors are relatively small microadenomas. However, macroadenomas are also seen, and some ACTH-secreting adenomas are clinically silent. Cushing's disease is 5 to 10 times more common in women than in men. These pituitary adenomas exhibit unrestrained ACTH secretion, with resultant hypercortisolemia. However, they retain partial suppressibility in the presence of high doses of administered glucocorticoids, providing the basis for dynamic testing to distinguish pituitary and nonpituitary causes of Cushing's syndrome.

Presentation and Diagnosis The diagnosis of Cushing's syndrome presents two great challenges: (1) to distinguish patients with pathologic cortisol excess from those with physiologic or other disturbances of cortisol production; and (2) to determine the etiology of cortisol excess, which can include iatrogenic administration of glucocorticoids, adrenal adenomas or carcinomas, pituitary adenomas, and ectopic sources of ACTH and CRH.

Typical features of chronic cortisol excess include thin, brittle skin, central obesity, hypertension, plethoric moon facies, purple striae and easy bruisability, glucose intolerance or diabetes mellitus, gonadal dysfunction, osteoporosis, proximal muscle weakness, signs of hyperandrogenism (acne, hirsutism), and psychologic disturbances (depression, mania, and psychoses) (Table 328-12). Hematopoietic features of

hypercortisolism include leukocytosis, lymphopenia, and eosinopenia. Immune suppression includes delayed hypersensitivity. The protean manifestations of hypercortisolism make it challenging to decide which patients mandate formal laboratory evaluation. Certain features make pathologic causes of hypercortisolism more likely -- these include characteristic central redistribution of fat, thin skin with striae and bruising, and proximal muscle weakness. In children and in young females, early osteoporosis may be particularly prominent. The primary cause of death is cardiovascular disease, but infections and risk of suicide are also increased.

Rapid development of features of hypercortisolism associated with skin hyperpigmentation and severe myopathy suggests the possibility of ectopic sources of ACTH. Hypertension, hypokalemic alkalosis, glucose intolerance, and edema are also more pronounced in these patients. Serum potassium levels<3.3 mmol/L are evident in ~70% of patients with ectopic ACTH secretion but are seen in <10% of patients with pituitary-dependent Cushing's disease.

Laboratory Investigation The diagnosis of Cushing's syndrome is based on laboratory documentation of endogenous hypercortisolism. Measurements of 24-h urine free cortisol (UFC) is a precise and cost-effective screening test. Alternatively, the failure to suppress plasma cortisol after an overnight 1-mg dexamethasone suppression test can be used to identify patients with hypercortisolism. As nadir levels of cortisol occur at night, elevated midnight samples of cortisol are suggestive of Cushing's syndrome. Basal plasma ACTH levels often distinguish patients with ACTH-independent (adrenal or exogenous glucocorticoid) from those with ACTH-dependent (pituitary, ectopic ACTH) Cushing's disease. Mean basal ACTH levels are about eight-fold higher in patients with ectopic ACTH secretion compared to those with pituitary ACTH-secreting adenomas. However, extensive overlap of ACTH levels in these two disorders precludes using ACTH to make the distinction. Instead, dynamic testing, based on differential sensitivity to glucocorticoid feedback, or ACTH stimulation in response to CRH or cortisol reduction is used to discriminate ectopic versus pituitary sources of excess ACTH (Table 328-13). Very rarely, circulating CRH levels are elevated, reflecting ectopic tumor-derived secretion of CRH and often ACTH. *For discussion of dynamic testing for Cushing's syndrome, see Chap. 331.*

Most ACTH-secreting pituitary tumors are <5 mm in diameter, and about half are undetectable by sensitive MRI. The high prevalence of incidental pituitary microadenomas diminishes the ability to distinguish ACTH-secreting pituitary tumors accurately by MRI.

Inferior Petrosal Venous Sampling Because pituitary MRI with gadolinium enhancement is insufficiently sensitive to distinguish small (<2 mm) pituitary ACTH-secreting adenomas from ectopic ACTH-secreting tumors that may have similar clinical and biochemical characteristics, bilateral inferior petrosal sinus ACTH sampling before and after CRH administration may be required. Simultaneous assessment of ACTH concentrations in each inferior petrosal vein and in the peripheral circulation provides a strategy for confirming and localizing pituitary ACTH production. Sampling is performed at baseline and 2, 5, and 10 min after intravenous ovine CRH (1 ug/kg) injection. An increased ratio (>2) of inferior petrosal:peripheral vein ACTH confirms pituitary Cushing's disease. After CRH injection, peak petrosal:peripheral ACTH ratios of³3

confirm the presence of a pituitary ACTH-secreting tumor. The sensitivity of this test is 99%, with very rare false-positive results. False-negative results may be encountered in patients with aberrant venous anatomic drainage. Petrosal sinus catheterizations are technically difficult, and about 0.05% of patients develop neurovascular complications. The procedure should not be performed in patients with hypertension or in the presence of a well-visualized pituitary adenoma on MRI.

TREATMENT

Selective transsphenoidal resection is the treatment of choice for Cushing's disease (Fig. 328-12). The remission rate for this procedure is ~80% for microadenomas but <50% for macroadenomas. After successful tumor resection, most patients experience a postoperative period of adrenal insufficiency that lasts for up to 12 months. This usually requires low-dose cortisol replacement, as patients experience steroid withdrawal symptoms as well as having a suppressed HPA axis. Biochemical recurrence occurs in approximately 5% of patients in whom surgery was initially successful.

When initial surgery is unsuccessful, repeat surgery is sometimes indicated, particularly when a pituitary source for ACTH is well documented. In older patients in whom growth and fertility are no longer important, hemi- or total hypophysectomy may be necessary if an adenoma is not recognized. Pituitary irradiation may be used after unsuccessful surgery, but it cures only about 15% of patients. Because radiation is slow and only partially effective in adults, steroidogenic inhibitors are used in combination with pituitary irradiation to block the adrenal effects of persistently high ACTH levels.

Mitotane (*o*, *p¢*-DDD) suppresses cortisol hypersecretion by inhibiting 11b-hydroxylase and cholesterol side-chain cleavage enzymes and by destroying adrenocortical cells. Side effects of mitotane include gastrointestinal symptoms, dizziness, gynecomastia, hyperlipidemia, skin rash, and hepatic enzyme elevation. It may also lead to hypoaldosteronism. *Ketoconazole*, an imidazole derivative antimycotic agent, inhibits several P450 enzymes and effectively lowers cortisol in most patients with Cushing's disease when administered twice daily (600 to 1200 mg/d). Elevated hepatic transaminases, gynecomastia, impotence, gastrointestinal upset, and edema are common side effects. *Metyrapone* (2 to 4 g/d) inhibits 11b-hydroxylase activity and normalizes plasma cortisol in up to 75% of patients. Side effects include nausea and vomiting, rash, and exacerbation of acne or hirsutism. Other agents include *aminoglutethimide* (250 mg tid), *trilostane* (200 to 1000 mg/d), *cyproheptadine* (24 mg/d), and IV *etomidate* (0.3 mg/kg per hour). Glucocorticoid insufficiency is a potential side effect of agents used to block steroidogenesis.

The use of steroidogenic inhibitors has decreased the need for bilateral adrenalectomy. Removal of both adrenal glands corrects hypercortisolism but may be associated with significant morbidity and necessitates permanent glucocorticoid and mineralocorticoid replacement. Adrenalectomy in the setting of residual corticotrope adenoma tissue predisposes to the development of *Nelson's syndrome*, a disorder characterized by rapid pituitary tumor enlargement and increased pigmentation secondary to high ACTH levels. Radiation therapy may be indicated to prevent the development of Nelson's syndrome after adrenalectomy.

GONADOTROPINS: FSH AND LH

SYNTHESIS AND SECRETION

Gonadotrope cells comprise about 10% of anterior pituitary cells and produce two gonadotropins -- LH and FSH. LikeTSH andhCG, LH and FSH are glycoprotein hormones consisting of a andb subunits. The a subunit is common to these glycoprotein hormones; specificity is conferred by theb subunits, which are expressed by separate genes.

Gonadotropin synthesis and release are dynamically regulated. This is particularly true in females, in whom the rapidly fluctuating gonadal steroid levels vary throughout the menstrual cycle. HypothalamicGnRH, a 10-amino-acid peptide synthesized in the preoptic region, regulates the synthesis and secretion of bothLH andFSH. GnRH is secreted in discrete pulses every 60 to 120 min, which in turn elicit LH and FSH pulses (Fig. 328-3). GnRH acts through aGPCR to stimulate phospholipase C, protein kinase C, and calcium signaling pathways. The pulsatile mode of GnRH input is essential to its action; pulses prime gonadotrope responsiveness, whereas continuous GnRH exposure induces desensitization. Based on this phenomenon, long-acting GnRH agonists are used to suppress gonadotropin levels in children with precocious puberty (Chap. 8) and in men with prostate cancer (Chap. 95) and are used in some ovulation-induction protocols to reduce endogenous gonadotropins (Chap. 336). Estrogens act at the hypothalamic and pituitary levels to control gonadotropin secretion. Chronic estrogen exposure is inhibitory, whereas rising estrogen levels, as occurs during the preovulatory surge, exert positive feedback to increase gonadotropin pulse frequency and amplitude. Progesterone slows GnRH pulse frequency but enhances gonadotropin responses to GnRH. Testosterone feedback in males also occurs at the hypothalamic and pituitary levels and partially reflects its conversion to estrogens (Chap. 335).

ThoughGnRH is the main regulator ofLH andFSHsecretion, FSH synthesis is also under separate control by the gonadal peptides inhibin and activin, which are members of the transforming growth factorb (TGF-b) family. Inhibin selectively suppresses FSH, whereas activin stimulates FSH synthesis (Chap. 336).

ACTION

The gonadotropin hormones interact with their respectiveGPCRsexpressed in the ovary and testis, evoking germ-cell development and maturation and steroid hormone biosynthesis. In women,FSHregulates ovarian follicle development and stimulates ovarian estrogen production.LHmediates ovulation and maintenance of the corpus luteum. In men, LH induces Leydig cell testosterone synthesis and secretion and FSH stimulates seminiferous tubule development and regulates spermatogenesis.

GONADOTROPIN DEFICIENCY

Hypogonadism is the most common presenting feature of adult hypopituitarism, even when other pituitary hormones are also deficient. It is often a harbinger of hypothalamic or pituitary diseases that impairGnRHproduction or delivery through the pituitary stalk. As noted above, hypogonadotropic hypogonadism is a common presenting feature of

hyperprolactinemia.

A variety of inherited and acquired disorders are associated with *isolated hypogonadotropic hypogonadism* (IHH) (Chap. 335). Hypothalamic defects associated with GnRH deficiency include two X-linked disorders, Kallmann syndrome (see above) and mutations in the *DAX1* gene. GnRH receptor mutations and inactivating mutations of the LHb and FSHb subunit genes are rare causes of selective gonadotropin deficiency. Acquired forms of GnRH deficiency leading to hypogonadotropism are seen in association with anorexia nervosa (Chap. 78), stress, starvation, and extreme exercise, but may also be idiopathic. Hypogonadotropic hypogonadism in these disorders is reversed by removal of the stressful stimulus.

Presentation and Diagnosis In premenopausal women, hypogonadotropic hypogonadism presents as diminished ovarian function leading to oligomenorrhea or amenorrhea, infertility, decreased vaginal secretions, decreased libido, and breast atrophy. In hypogonadal adult males, secondary testicular failure is associated with decreased libido and potency, infertility, decreased muscle mass with weakness, reduced beard and body hair growth, soft testes, and characteristic fine facial wrinkles. Osteoporosis occurs in both untreated hyogonadal females and males.

Laboratory Investigation Central hypogonadism is associated with low or inappropriately low serum gonadotropin levels and low sex hormone concentrations (testosterone in males, estradiol in females). Three pooled serum samples drawn 20 min apart are used for accurate measurement of serum LH and FSH levels, thus allowing for the effects of hormone secretory pulses. Male patients have abnormal semen analysis.

Intravenous GnRH (100 ug) stimulates gonadotropes to secrete LH (which peaks within 30 min) and FSH (which plateaus during the ensuing 60 min). Normal responses vary according to menstrual cycle stage, age, and sex of the patient. Generally, LH levels increase about threefold, whereas FSH responses are less pronounced. In the setting of gonadotropin deficiency, a normal gonadotropin response to GnRH indicates intact gonadotrope function and suggests a hypothalamic abnormality. An absent response, however, cannot reliably distinguish pituitary from hypothalamic causes of hypogonadism. For this reason, GnRH testing usually adds little to the information gained from baseline evaluation of the hypothalamic-pituitary-gonadotrope axis, except in cases of isolated GnRH deficiency (e.g., Kallmann syndrome).

MRI examination of the sellar region and assessment of other pituitary functions are usually indicated in patients with documented central hypogonadism.

TREATMENT

In males, testosterone replacement is necessary to achieve and maintain normal growth and development of the external genitalia, secondary sex characteristics, male sexual behavior, and androgenic anabolic effects including maintenance of muscle function and bone mass. Testosterone may be administered by intramuscular injections every 1 to 4 weeks or using patches that are replaced daily (Chap. 335). Gonadotropin injections [hCG or human menopausal gonadotropin (hMG)] over 12 to 18 months are used to

restore fertility. Pulsatile GnRH therapy (25 to 150 ng/kg every 2 h), administered by a subcutaneous infusion pump, is also effective for treatment of hypothalamic hypogonadism when fertility is desired.

In premenopausal women, cyclical replacement of estrogen and progesterone maintains secondary sexual characteristics and genitourinary tract integrity and prevents premature osteoporosis and possibly coronary artery disease (Chap. 336). Gonadotropin therapy is used for ovulation induction. Follicular growth and maturation are initiated using hMG or recombinant FSH; hCG is subsequently injected to induce ovulation. As in men, pulsatile GnRH therapy can be used to treat hypothalamic causes of gonadotropin deficiency.

NONFUNCTIONING AND GONADOTROPIN-PRODUCING PITUITARY ADENOMAS

Etiology and Prevalence Nonfunctioning pituitary adenomas include those that secrete little or no pituitary hormones, as well as tumors that produce too little hormone to result in recognizable clinical features. They are the most common type of pituitary adenoma and are usually macroadenomas at the time of diagnosis because clinical features are inapparent until tumor mass effects occur. Based on immunohistochemistry, most clinically nonfunctioning adenomas can be shown to originate from gonadotrope cells. These tumors typically produce small amounts of intact gonadotropins (usually FSH) as well as uncombined a and LH b and FSH b subunits. Tumor secretion may lead to elevated a and FSHb subunits and, rarely, to increased LHb subunit levels. Some adenomas express a subunits without FSH or LH. TRH administration often induces an atypical increase of tumor-derived gonadotropins or subunits.

Presentation and Diagnosis Clinically nonfunctioning tumors may present with optic chiasm pressure and other symptoms of local expansion or be incidentally discovered on an MRI performed for another indication. Menstrual disturbances or ovarian hyperstimulation rarely occur in women with large tumors that produce FSH and LH. More commonly, adenoma compression of the pituitary stalk or surrounding pituitary tissue leads to attenuated LH and features of hypogonadism. Prolactin levels are usually slightly increased, also because of stalk compression. It is important to distinguish this circumstance from true prolactinomas, as most nonfunctioning tumors respond poorly to treatment with dopamine agonists.

Laboratory Investigation The goal of laboratory testing in clinically nonfunctioning tumors is to classify the type of the tumor, to identify hormonal markers of tumor activity, and to detect possible hypopituitarism. Free a subunit levels may be elevated in 10 to 15% of patients with nonfunctioning tumors. In female patients, peri- or postmenopausal basal FSH concentrations are difficult to distinguish from tumor-derived FSH elevation. Premenopausal women have cycling FSH levels, also preventing clear-cut diagnostic distinction from tumor-derived FSH. In men, gonadotropin-secreting tumors may be diagnosed because of slightly increased gonadotropins (FSH>LH) in the setting of a pituitary mass. Testosterone levels are usually low, despite the normal or increased LH level, perhaps reflecting reduced LH bioactivity or the loss of normal LH pulsatility. Because this pattern of hormone tests is also seen in primary gonadal failure and, to some extent, with aging (Chap. 335), the increased gonadotropins alone are insufficient for the diagnosis of a gonadotropin-secreting tumor. In the majority of patients with

gonadotrope adenomas,TRHadministration stimulates LH b subunit secretion; this response is not seen in normal individuals.GnRHtesting is not helpful for making the diagnosis. For nonfunctioning and gonadotropin-secreting tumors, the diagnosis usually rests on immunohistochemical analyses of resected tumor tissue, as the mass effects of these tumors usually necessitate resection.

Although acromegaly or Cushing's syndrome usually presents with unique clinical features, clinically inapparent somatotrope or corticotrope adenomas can be excluded by a normalIGF-I value and normal 24-h urinary free cortisol levels. IfPRLlevels are <100 ug/L in a patient harboring a pituitary mass, a nonfunctioning adenoma causing pituitary stalk compression should be considered.

TREATMENT

Asymptomatic small nonfunctioning adenomas with no threat to vision may be followed with regularMRI and visual field testing without immediate intervention. However, for larger macroadenomas, transsphenoidal surgery is the only effective way to reduce tumor size and relieve mass effects (Fig. 328-13). Although it is not usually possible to remove all adenoma tissue surgically, vision improves in 70% of patients with preoperative visual field defects. Preexisting hypopituitarism that results from tumor mass effects commonly improves or may resolve completely. Early postoperative complications include diabetes insipidus and/or inappropriate antidiuretic hormone secretion. Beginning about 6 months postoperatively, MRI scans should be performed yearly to detect tumor regrowth. Within 5 to 6 years following successful surgical resection, ~15% of nonfunctioning tumors recur. When substantial tumor remains after transsphenoidal surgery, adjuvant radiotherapy may be indicated to prevent tumor growth. Radiotherapy may be deferred if no postoperative residual mass is evident.

Nonfunctioning pituitary tumors respond poorly to dopamine agonist treatment, with modest tumor shrinkage occurring in <10% of patients. Although SSTR subtypes 2 and 5 have been identified on nonfunctioning pituitary adenomas, octreotide does not shrink these tumors and only modestly suppresses gonadotropin and a subunit levels. Visual improvement sometimes occurs without evident reduction of tumor size byMRI, presumably reflecting relief of pressure on the optic tracts. The selectiveGnRHantagonist, Nal-Glu GnRH, suppressesFSHhypersecretion but has no effect on adenoma size.

THYROID-STIMULATING HORMONE

SYNTHESIS AND SECRETION

TSH-secreting thyrotrope cells comprise 5% of the anterior pituitary cell population. TSH is structurally related toLH andFSH. It shares a common a subunit with these hormones but contains a specific TSH b subunit.TRH is a hypothalamic tripeptide (pyroglutamyl histidylprolinamide) that acts through aGPCR to stimulate phospholipase C, protein kinase C, and calcium pathways. TRH stimulates TSH synthesis and secretion; it also stimulates the lactotrope cell to secretePRL. TSH secretion is stimulated by TRH, whereas thyroid hormones, dopamine, SRIF, and glucocorticoids suppress TSH by overriding TRH induction.

The thyrotrope is stimulated when TSH is released from the negative feedback inhibition of thyroid hormones. Thus, thyroid damage, including surgical thyroidectomy, radiation-induced hypothyroidism, chronic thyroiditis, or prolonged goitrogen exposure, are associated with increased TSH. Long-standing untreated hypothyroidism can lead to thyrotrope hyperplasia and pituitary enlargement, which may be evident on MRI.

ACTION

TSH is secreted in pulses, though the excursions are modest in comparison to other pituitary hormones because of the relatively low amplitude of the pulses and the relatively long half-life of TSH. Consequently, single determinations of TSH suffice to assess its circulating levels. TSH binds to a GPCR on thyroid follicular cells to stimulate thyroid hormone synthesis and release (Chap. 330).

TSH DEFICIENCY

Features of central hypothyroidism, due to TSH deficiency, mimic those seen whith primary hypothyroidism but are generally less severe. Pituitary hypothyroidism is characterized by low basal TSH levels in the setting of low free thyroid hormone. In contrast, patients with hypothyroidism of hypothalamic origin (presumably due to a lack of endogenous TRH) may exhibit normal or even slightly elevated TSH levels. There is evidence that the TSH produced in this circumstance has reduced biologic activity because of altered glycosylation.

TRH (200 ug) injected intravenously causes a two- to threefold increase in TSH (and PRL) levels within 30 min. Although TRH testing can be used to assess TSH reserve, abnormalities of the thyroid axis can usually be detected based on basal free T_4 and TSH levels, without the need for TRH testing.

Thyroid-replacement therapy should be initiated after establishing adequate adrenal function. Dose adjustment is based on thyroid hormone levels and clinical parameters rather than the TSH level.

TSH-SECRETING ADENOMAS

TSH-producing macroadenomas are rare but are often large and locally invasive when they occur. Patients usually present with thyroid goiter and hyperthyroidism, reflecting overproduction of TSH. Diagnosis is based on demonstrating elevated serum free T_4 levels, inappropriately normal or high TSH secretion, and MRI evidence of a pituitary adenoma. An elevated free asubunit level occurs in about half of patients and supports the diagnosis of a TSH-secreting adenoma.

It is important to exclude other causes of inappropriate TSH secretion, such as resistance to thyroid hormone, an autosomal dominant disorder caused by mutations in the thyroid hormone b receptor (Chap. 330). The presence of a pituitary mass and elevated a subunit levels are suggestive of a TSH-secreting tumor. Dysalbuminemic hyperthyroxinemia syndromes, caused by various mutations in serum thyroid hormone binding proteins, are also characterized by elevated thyroid hormone levels, but with

normal rather than suppressed TSH levels. However, free thyroid hormone levels are normal in these disorders, most of which are familial.

TREATMENT

The initial therapeutic approach is to remove or debulk the tumor mass surgically, using either a transsphenoidal or subfrontal approach. Total resection is not often achieved as most of these adenomas are large and locally invasive. Normal circulating thyroid hormone levels are achieved in about two-thirds of patients after surgery. Thyroid ablation or antithyroid drugs (methimazole or propylthiouracil) can be used to reduce thyroid hormone levels. Dopamine agonists are rarely effective for suppressingTSHsecretion from these tumors. However, somatostatin analogue treatment effectively normalizes TSH and a subunit hypersecretion, shrinks the tumor mass in 50% of patients, and improves visual fields in 75% of patients; euthyroidism is restored in most patients. In some patients, octreotide markedly suppresses TSH, causing biochemical hypothyroidism that requires concomitant thyroid hormone replacement. Lanreotide (30 mg intramuscularly), a long-acting somatostatin analogue (see above), effectively suppresses TSH and thyroid hormone in patients treated every 14 days.

DIABETES INSIPIDUS

See Chap. 329 for diagnosis and treatment of diabetes insipidus.

(Bibliography omitted in Palm version)

Back to Table of Contents

329. DISORDERS OF THE NEUROHYPOPHYSIS - *Gary L. Robertson*

The neurohypophysis, or posterior pituitary gland, is formed by axons that project from large cell bodies in the supraoptic and paraventricular nuclei of the hypothalamus to the posterior portion of the sella turcica. The neurohypophysis produces two hormones: (1) arginine vasopressin (AVP), also known as antidiuretic hormone (ADH); and (2) oxytocin. AVP acts on the renal tubules to induce water retention, leading to concentration of the urine. Oxytocin stimulates postpartum milk letdown in response to suckling. AVP deficiency causes diabetes insipidus (DI), characterized by the production of large amounts of dilute urine. Excessive or inappropriate production of AVP predisposes to hyponatremia, reflecting water retention. There are no known clinical disorders associated with oxytocin deficiency or excess.

VASOPRESSIN

ACTION

AVPis a nonapeptide composed of a six-membered disulfide ring and a tripeptide tail on which the C-terminal carboxy group is amidated (Fig. 329-1). The most important, if not the only, physiologic action of AVP is to influence the rate of water excretion by promoting concentration of urine. This antidiuretic effect is achieved by increasing the hydroosmotic permeability of cells that line the distal tubule and medullary collecting ducts of the nephron. In the absence of AVP, these cells are impermeable to water and reabsorb little, if any, of the relatively large volume of dilute filtrate that enters from the cortical nephron. Low AVP concentration results in the production of large amounts of urine (as much as 0.2 mL/kg per minute) that is maximally dilute (specific gravity and osmolality ~1.000 and 50 mmol/L, respectively), a condition known as a *water diuresis*. In the presence of AVP, the luminal surface of the cells lining the terminal collecting duct becomes selectively permeable to water, allowing it to diffuse back down the osmotic gradient created by the hypertonic renal medulla. As a result, the dilute fluid passing through the tubules is concentrated and the rate of urine flow decreases. The magnitude of this antidiuretic effect varies in direct proportion to the plasma AVP concentration. It is mediated via binding of AVP to G protein-coupled V_2receptors on the serosal surface of the cell, activation of adenyl cyclase, and insertion into the luminal surface of water channels composed of a protein known as *aquaporin 2* (Fig. 329-2). The genes encoding the V_2receptors and aquaporin 2 have been cloned and appear to be expressed exclusively in the distal and collecting tubules of the kidney. Nonpeptide, as well as peptide, AVP analogues with potent agonist or antagonist actions at human V_2receptors have been developed for treating disorders of water metabolism caused by deficient or excessive production of AVP (see below).

At high concentrations,AVP also has several other actions, including contraction of smooth muscle in blood vessels in the skin and gastrointestinal tract, glycogenolysis in the liver, and potentiation of adrenocorticotropic hormone (ACTH) release by corticotropin-releasing factor (CRF). These effects are mediated by V_{1a} or V_{1b}receptors that are coupled to phospholipase C. The genes that encode these receptors have also been cloned and sequenced and are expressed in many organs, including blood vessels, the anterior and posterior pituitary, and certain other areas of brain. Their role, if any, in human physiology/pathophysiology is still uncertain.

SYNTHESIS

AVPis synthesized via a polypeptide precursor that includes a binding protein known as *neurophysin II* and a glycosylated peptide called *copeptin*. The gene encoding the AVP precursor is located on chromosome 20 and has three exons. It is expressed in distinct subpopulations of magno- and parvocellular neurons in the supraoptic and paraventricular nuclei. Like other peptide hormones destined for secretion, newly synthesized AVP-neurophysin II precursor is translocated from the cytosol to the endoplasmic reticulum, where the signal peptide is removed and the prohormone folds and oligomerizes before moving through the Golgi apparatus to the neurosecretory vesicles; there it is transported down the axons and further cleaved to AVP, neurophysin II, and copeptin. Stimulation of the neurons results in an influx of calcium, fusion of the neurosecretory vesicle with the cell membrane, and extrusion of its contents into the systemic circulation.

SECRETION

The secretion ofAVP is regulated primarily by the "effective" osmotic pressure of body fluids. This control is mediated by specialized cells, known as *osmoreceptors*, which appear to be located in the anteromedial hypothalamus near the supraoptic nucleus. These osmoreceptors are extremely sensitive to small changes in the plasma concentration of sodium and certain other effective solutes such as mannitol but show little or no response to other solutes such as urea or glucose. They appear to have inhibitory as well as stimulatory components that function in concert to regulate AVP secretion around a specific set point. Thus, when plasma osmolality/sodium are depressed to a certain minimum or threshold level of ~280 mosmol/kg or 135 meq/L, respectively, plasma AVP is suppressed to low or undetectable levels and a water diuresis ensues. Conversely, when plasma osmolality/sodium rise above this "threshold," plasma AVP rises steeply in direct proportion, reaching a concentration sufficient to produce a maximum antidiuresis when plasma osmolality/sodium reach ~295 mosmol/kg and 143 meq/L. However, the exact "set" and "sensitivity" of this osmoregulatory system vary appreciably from person to person, apparently as a result of genetic influences, and change during pregnancy, the menstrual cycle, and normal aging; they can also be altered or disrupted by various pathologic conditions.

AVPsecretion can also be influenced by acute changes in blood volume or pressure. This baroregulation is mediated largely by neuronal afferents that originate in transmural pressure receptors of the cardiac atria, aorta, and carotid arteries; project via the vagus and glossopharyngeal nerves to the nucleus tractus solitarius of the brain stem; and then ascend to the paraventricular and supraoptic nucleii of the hypothalmus. These pathways regulate AVP release by maintaining a tonic inhibitory tone that decreases when blood volume or pressure falls by >10 to 20%. This baroregulatory system is probably of minor importance in the physiology of AVP secretion because the hemodynamic changes required to affect it are larger than those usually occurring in the course of normal activities. Moreover, moderate hemodynamic stimuli do not disrupt or override the osmoregulatory system but do lower its threshold or set point by an amount proportional to the magnitude of the hypovolemia or hypotension. However, the baroregulatory system undoubtedly plays an important role in AVP secretion in patients

with large, acute disturbances of hemodynamic function.

AVPsecretion can also be stimulated by a variety of other nonosmotic variables including nausea, acute hypoglycemia, glucocorticoid deficiency, smoking, and, possibly, hyperangiotensinemia. The emetic stimuli are extremely potent since they typically elicit immediate, 50- to 100-fold increases in plasma AVP, even when the nausea is transient and unassociated with vomiting or other symptoms. They appear to act via the emetic center in the medulla and can be completely blocked by treatment with antiemetics such as fluphenazine. There is no evidence that pain or other noxious stresses have any affect on AVP unless they elicit a vasovagal reaction with its associated nausea and hypotension.

METABOLISM

From the venous circulation,AVPdistributes rapidly into a space roughly equal in size to the extracellular fluid volume. It is cleared irreversibly from this space with a $t_1/2$ of 10 to 30 min. Most AVP clearance is due to degradation in the liver and kidneys. Urinary clearance of the hormone is normally much less than creatinine clearance but can vary as much as tenfold, depending on individual differences and changes in total solute clearance. Therefore, measurement of urinary AVP excretion rates are not always a reliable indicator of changes in secretion or plasma levels of the hormone. During pregnancy, the metabolic clearance of AVP is increased three- to fourfold due to placental production of an N-terminal peptidase.

THIRST

ThoughAVPregulates the effective osmotic pressure of body fluids by varying the rate of urinary free-water excretion, it cannot reduce insensible or urinary water output below a certain minimum obligatory level. Thus, an additional mechanism -- thirst -- is essential to prevent hypertonic dehydration. Like AVP, thirst is regulated primarily by an osmostat that is located in the anteromedial hypothalamus and is able to detect very small changes in the plasma concentration of sodium and certain other effective solutes. It functions like the AVP osmostat except that it appears to be "set" slightly higher. This arrangement ensures that thirst, polydipsia, and dilution of body fluids do not occur until dehydration and the resultant rise in plasma osmolality start to exceed the defensive capacity of the antidiuretic mechanism.

OXYTOCIN

Oxytocin is also a nonapeptide and differs fromAVP only at positions 3 and 8 (Fig. 329-1). However, it has relatively little antidiuretic effect and seems to act mainly on mammary ducts to facilitate milk letdown during nursing (Chap. 337). It may also help to initiate or facilitate labor by stimulating contraction of uterine smooth muscle, but it is not yet clear if this action is physiologic or necessary for normal delivery. Both the mammary and uterine effects are mediated by a G protein-coupled receptor that is linked to phospholipase C. Antagonists for this receptor have been developed and tested in humans for possible use in treating premature labor.

Oxytocin is also synthesized via a macromolecular precursor that is encoded by a gene

located on chromosome 20, very near the AVP gene. However, it differs in orientation and size and encodes only a signal peptide, the hormone, and its associated neurophysin. Oxytocin is also expressed in different magnocellular neurons than AVP and, in humans, is not subject to any of the same regulatory influences. Indeed, with the possible exception of nipple stimulation in the postpartum period, no other stimuli are known to consistently induce release of the hormone in humans. Plasma oxytocin is not increased during pregnancy or at the initiation of labor, although the latter condition may be facilitated by upregulation of oxytocin receptors. The distribution and clearance of oxytocin are similar to those of AVP. Oxytocin is also degraded by the liver and kidneys and an N-terminal peptidase produced by the placenta.

DEFICIENCIES OF VASOPRESSIN SECRETION AND ACTION

DIABETES INSIPIDUS

Clinical Characteristics Decreased secretion or action of AVP usually manifests as DI, a syndrome characterized by the production of abnormally large volumes of dilute urine. The 24-h urine volume is >50 mL/kg body weight and the osmolality is <300 mmol/kg. The polyuria produces symptoms of urinary frequency, enuresis, and/or nocturia, which may disturb sleep and cause mild daytime fatigue or somnolence. It is also associated with thirst and a commensurate increase in fluid intake (polydipsia). Clinical signs of dehydration are uncommon unless fluid intake is impaired.

Etiology Deficient secretion of AVP can be primary or secondary. The primary form usually results from agenesis or irreversible destruction of the neurohypophysis and is variously referred to as *neurohypophyseal DI*, *neurogenic DI*, *pituitary DI*, *cranial DI*, or *central DI*. It can be caused by a variety of congenital, acquired, or genetic disorders but almost half the time it is idiopathic (Table 329-1). The genetic form of neurohypophyseal DI is usually transmitted in an autosomal dominant mode and is caused by diverse mutations in the coding region of the AVP-neurophysin II gene. The mutant precursor cannot be processed properly or efficiently and eventually destroys the neuron, thereby accounting for the dominant mode of transmission and the delayed onset of the disorder. An X-linked recessive form also occurs. A primary deficiency of plasma AVP can also result from increased metabolism by an N-terminal aminopeptidase produced by the placenta. It is referred to as *gestational DI* since the signs and symptoms manifest during pregnancy and usually remit several weeks after delivery. However, a subclinical deficiency in AVP secretion can often be demonstrated in the nonpregnant state in these individuals, indicating that damage to the neurohypophysis may also contribute to the AVP deficiency. Finally, a primary deficiency of AVP can also result from malformation or destruction of the neurohypophysis by a variety of diseases or toxins (Table 329-1).

Secondary deficiencies of AVP result from inhibition of secretion by excessive intake of fluids. They are referred to as *primary polydipsia* and can be divided into three subcategories. One of them, called *dipsogenic DI*, seems to be caused by an inappropriate increase in thirst caused by a reduction in the "set" of the osmoregulatory mechanism. It sometimes occurs in association with multifocal diseases of the brain such as neurosarcoid, tuberculous meningitis, or multiple sclerosis but is often idiopathic. The second subtype, called *psychogenic polydipsia*, is not associated with

thirst, and the polydipsia seems to be a feature of psychosis. The third subtype, which may be referred to as *iatrogenic polydipsia*, results from recommendations of health professionals or the popular media to increase fluid intake for its presumed preventive or therapeutic benefits for other disorders.

Primary deficiencies in the antidiuretic action of AVP result in *nephrogenic DI* (Table 329-1). It can be genetic, acquired, or caused by exposure to various drugs. The genetic form is usually transmitted in an X-linked mode and is caused by mutations in the coding region of the V_2 receptor gene. An autosomal recessive form is caused by mutations in the gene encoding the aquaporin protein that forms the water channels in the distal nephron.

Secondary deficiencies in the antidiuretic response to AVP result from polyuria per se. They appear to be caused by washout of the medullary concentration gradient and/or suppression of aquaporin function. They usually resolve 24 to 48 h after the polyuria is corrected but often complicate interpretation of certain acute tests commonly used for differential diagnosis.

Pathophysiology When the net secretion or antidiuretic effect of AVP is decreased by >80 to 85%, the amount of hormone produced under basal conditions is insufficient to concentrate the urine and the rate of output increases exponentially to symptomatic levels. If the AVP defect is primary (e.g., the patient has pituitary, gestational, or nephrogenic DI), the polyuria results in a small (1 to 2%) decrease in body water and a commensurate increase in plasma osmolality and sodium concentration that stimulate thirst and a compensatory increase in water intake. As a result, *overt physical or laboratory signs of dehydration do not develop unless the patient also has a defect in thirst* (see below) *or fails to drink for some other reason*.

The severity of the defect in antidiuretic function varies markedly among patients with pituitary, gestational, or nephrogenic DI. In some, the deficiencies in AVP secretion or action are so severe that basal urine output approximates the maximum (10 to 15 mL/min); even an intense stimulus such as nausea or severe dehydration does not increase plasma AVP enough to concentrate the urine. In others, however, the deficiency in AVP secretion or action is less pronounced, and a modest stimulus such as a few hours of fluid deprivation, smoking, or a vasovagal reaction increases plasma AVP sufficiently to produce a profound antidiuresis. The maximum urine osmolality achieved in these patients is usually less than normal, largely because their maximal concentrating capacity is temporarily impaired by chronic polyuria per se. However, in a few patients with partial pituitary or nephrogenic DI, it can reach levels as high as 800 mosmol/kg (Fig. 329-3).

In primary polydipsia, the pathogenesis of the polydipsia and polyuria is the reverse of that in neurohypophyseal, nephrogenic, and gestational DI. Thus, the excessive intake of fluids slightly increases body water, thereby reducing plasma osmolality, AVP secretion, and urinary concentration. The latter results in a compensatory increase in urinary free-water excretion that varies in direct proportion to intake. Therefore, clinically appreciable overhydration is uncommon unless the compensatory water diuresis is impaired by a drug or disease that stimulates or mimics endogenous AVP.

In the dipsogenic form of primary polydipsia, fluid intake is excessive because the osmotic threshold for thirst appears to be reset to the left, often well below that for AVP release. As a result, thirst is abnormally increased and cannot be completely relieved because plasma AVP is suppressed and an offsetting water diuresis develops before plasma osmolality is reduced sufficiently to eliminate the dipsogenic stimulus. Typically, therefore, patients with dipsogenic DI present with complaints of chronic thirst, polydipsia, and polyuria indistinguishable from those in patients with pituitary, gestational, or nephrogenic DI. When deprived of fluids or subjected to some other acute osmotic or nonosmotic stimulus, they invariably increase plasma AVP normally, but the resultant increase in urine concentration is usually subnormal because their renal capacity to concentrate the urine is also blunted by chronic polyuria. Thus, their antidiuretic response to these stimuli may be indistinguishable from that in patients with partial pituitary, partial gestational, or partial nephrogenic DI (Fig. 329-3).

Differential Diagnosis When symptoms of urinary frequency, enuresis, nocturia, and/or persistent thirst are present, causes other than polyuria should be excluded. A 24-h urine output > 50 mL/kg per day (>3500 mL in a 70-kg man) is suspicious for DI. If the osmolality of the 24-h urine is >300 mosmol/kg, the patient has a solute diuresis and should be evaluated for uncontrolled diabetes mellitus or other less common causes of excessive solute excretion. However, if the 24-h urine osmolality is <300 mosmol/kg, the patient has a water diuresis and should be evaluated further to determine which type of DI is present.

In differentiating between the various types of DI, the history, physical examination, and routine laboratory tests may be helpful but are rarely sufficient because few, if any, of the findings are pathognomonic. Except in the rare patient who is clearly dehydrated under basal conditions of *ad libitum* fluid intake, this evaluation should begin with a *fluid deprivation test*. To minimize patient discomfort, avoid excessive dehydration, and maximize the information obtained, the test should be started in the morning and water balance should be monitored closely with hourly measurements of body weight, plasma osmolality and/or sodium concentration, and urine volume and osmolality.

If fluid deprivation does not result in urine concentration (osmolality>300 mosmol/kg, specific gravity>1.010) before body weight decreases by 5% or plasma osmolality/sodium exceed the upper limit of normal, primary polydipsia and a partial defect in AVP secretion or action are largely excluded (Fig. 329-3). In these patients, severe pituitary or nephrogenic DI can usually be distinguished by administering desmopressin (DDAVP, 0.03 ug/kg subcutaneously or intravenously) and repeating the measurement of urine osmolality 1 to 2 h later. An increase of >50% indicates severe pituitary DI, whereas a smaller or absent response is strongly suggestive of nephrogenic DI.

However, these indirect criteria are not useful for diagnosis in patients who concentrate their urine during fluid deprivation, because the changes in urine osmolality are remarkably similar in primary polydipsia and partial pituitary and partial nephrogenic DI (Fig. 329-3). In this situation, the safest and most reliable way to differentiate these conditions is to measure plasma or urine AVP collected before and during the fluid deprivation test and analyze the result in relation to the concurrent plasma or urine osmolality (Fig. 329-4). This approach invariably differentiates partial

nephrogenic DI from partial pituitary DI and primary polydipsia. It also differentiates pituitary DI from primary polydipsia if the hormone is measured when plasma osmolality or sodium is clearly above the normal range. However, the requisite level of hypertonic dehydration is difficult to produce by fluid deprivation alone when urine concentration occurs. Therefore, it is usually necessary to add an infusion of hypertonic (3%) saline and repeat the AVP measurements when plasma osmolality rises to >300 mmol/kg (Na_+> 145 mmol/L). This endpoint is usually reached within 30 to 120 min if the hypertonic saline is infused at a rate of 0.1 mL/kg per minute and the fluid deprivation is maintained.

The differential diagnosis of DI may also be facilitated by magnetic resonance imaging (MRI) of the pituitary and hypothalamus. In most healthy adults and children, the posterior pituitary emits a hyperintense signal in T1 weighted mid-saggital images. This "bright spot" is almost invariably absent or abnormally small in patients with pituitary DI but is present in 80 to 90% of those with primary polydipsia. Thus, the presence of a normal bright spot virtually excludes pituitary DI, whereas its absence supports but does not prove this diagnosis. Therefore, the MRI findings must be interpreted with caution and only in conjunction with other diagnostic studies based on assays of AVP or the differential responses to treatment.

TREATMENT

The signs and symptoms of uncomplicated pituitary DI can be eliminated completely by treatment with DDAVP (Fig. 329-5). It is a synthetic analogue of AVP (Fig. 329-1) that acts selectively at V_2 receptors to increase urine concentration and decrease urine flow in a dose-dependent manner. However, it is more resistant to degradation than AVP and has a three- to fourfold longer duration of action. This property makes it particularly useful in the treatment of gestational DI or pituitary DI during pregnancy. DDAVP can be given by intravenous or subcutaneous injection, nasal inhalation, or oral tablet. The doses required to control pituitary DI completely vary widely, depending on the patient and the route of administration. However, they usually range from 1 to 2 ug qd or bid by injection, 10 to 20 ug bid or tid by nasal spray, and 100 to 400 ug bid or tid orally. The onset of action is rapid, ranging from as little as 15 min after injection to 60 min after oral administration. When given in doses sufficient to completely normalize urinary osmolality and flow, DDAVP produces a slight (1 to 3%) increase in total-body water and a commensurate decrease in plasma osmolality and sodium concentration that rapidly eliminates thirst and polydipsia. Consequently, water balance is maintained and hyponatremia does not develop unless the patient has an associated abnormality in the osmoregulation of thirst or ingests/receives excessive amounts of fluid for some other reason. Fortunately, abnormal thirst occurs in <10% of patients with pituitary DI, and the other causes of excessive intake can usually be eliminated by educating the patient about the risks of drinking for reasons other than thirst. Therefore, most patients with pituitary DI can take desmopressin in doses sufficient to maintain a normal urine output continuously and do not need to endure the inconvenience and discomfort of allowing intermittent escape to prevent water intoxication.

Pituitary DI can also be treated with chlorpropamide (Diabinese). The mechanism of its antidiuretic action is uncertain but may involve potentiation of the effect of small amounts of AVP or direct activation of the V_2 receptor. In patients with severe as well as

partial pituitary DI, doses of chlorpropamide similar to those used in the treatment of diabetes mellitus (125 to 500 mg once daily) increase urine concentration and decrease urine flow, thirst, and polydipsia in a manner similar to DDAVP. The antidiuresis is usually less rapid and smaller than that produced by DDAVP but is almost always sufficient to reduce urine output by 30 to 70%. Moreover, its antidiuretic effect can be enhanced appreciably by cotreatment with a thiazide diuretic. The ability of water loading to reduce the antidiuretic effect of chlorpropamide makes it particularly useful in the treatment of patients who have pituitary DI and abnormal thirst since it is less likely than DDAVP to produce water intoxication. However, unlike DDAVP, chlorpropamide can have other side effects including hypoglycemia, which can be precipitated by severe reductions in caloric intake or heavy exercise, and it exhibits a disulfuram (Antabuse)-like reaction to ethanol. Chlorpropamide is contraindicated in the treatment of gestational DI because its teratogenicity is unknown.

Primary polydipsia cannot be treated with DDAVP in the usual way because a sustained inhibition of the compensatory water diuresis almost invariably results in the development of water intoxication within 24 to 48 h. This complication can also be caused by administration of a thiazide diuretic, smoking, or other nonosmotic stimuli to endogenousAVPsecretion. Iatrogenic polydipsia can often be corrected by patient counseling; however, there is no effective treatment for either psychogenic or dipsogenicDI. In the latter, nocturia or nocturnal enuresis can often be controlled safely by administering a single small dose of DDAVP at bedtime. If the dose is adjusted carefully to provide no more than 8 to 10 h of antidiuresis, it will not result in water intoxication, because patients with dipsogenic, as well as other forms of DI, tend to drink less fluid at night than during the day. Family or other caregivers of patients with psychogenic or dipsogenic DI should also be warned about the hazards of water intoxication caused by a variety of diseases or drugs that can stimulate or mimic the antidiuretic effects of endogenous AVP (see below).

The symptoms and signs of nephrogenicDI are not affected by treatment with DDAVP or chlorpropamide but may be reduced by treatment with a thiazide diuretic and/or amiloride in conjunction with a low-sodium diet. Inhibitors of prostaglandin synthesis (e.g., indomethacin) are also effective in many patients.

ADIPSIC HYPERNATREMIA

Clinical Characteristics Adipsic hypernatremia is characterized by chronic or recurrent hypertonic dehydration and a deficientAVPresponse to osmotic stimulation. Despite their dehydration, the patients have little or no thirst and may even resist efforts to increase their oral intake of fluids. The hypernatremia varies in severity and is associated with commensurate signs of hypovolemia such as tachycardia, postural hypotension, azotemia, hyperuricemia, and hypokalemia. Muscle weakness, pain, rhabdomyolysis, hyperglycemia, hyperlipidemia, and acute renal failure may also occur. Most patients remain conscious unless they have severe hyperglycemia and/or hypertonicity or go on to develop hyponatremia as a result of excessive rehydration.

Etiology Adipsic hypernatremia is caused by agenesis or destruction of the hypothalamic osmoreceptors that normally regulate thirst andAVPsecretion. The osmoreceptor deficiency can usually be traced to an identifiable congenital or acquired

disease in the hypothalamus but is sometimes idiopathic ([Table 329-2](#)). The neurohypophysis and its other regulatory afferents are usually spared; an[MRI](#)typically shows a normal posterior pituitary bright spot, and the AVP response to nonosmotic stimuli is also normal. Occasionally, the neurohypophysis is also affected, resulting in a combined defect in water balance that is particularly severe and difficult to manage.

Pathophysiology Lack of thirst and failure to drink enough water to replenish renal and extrarenal losses decrease total-body water and increase plasma osmolality/sodium. Plasma renin activity and aldosterone secretion also increase, and plasma potassium falls due to increased urinary excretion. The severity, frequency, and speed with which hypertonic dehydration develops vary markedly from patient to patient, or from time to time in the same patient, owing largely to differences in the rate of insensible and/or renal loss.

The osmoregulation of[AVP](#)secretion is also impaired in nearly all patients with adipsic hypernatremia ([Fig. 329-6](#)). This deficiency is obvious when the hormone is measured in the presence of hypertonic dehydration but is rarely severe enough to produce[DI](#). During rehydration, however, some patients exhibit a further decrease in their plasma AVP and develop DI before their hypernatremia is fully corrected. In other patients, the osmoregulatory deficiency appears to be complete because basal plasma AVP remains relatively fixed, irrespective of whether plasma osmolality and sodium are above, within, or below the normal range. Thus, if overhydrated, these patients do not mount a compensatory water diuresis and quickly develop a hyponatremic syndrome that is clinically and biochemically indistinguishable from acute syndrome of inappropriate antidiuretic hormone, which is commonly referred to as SIADH (see below). In all but a few patients, the abnormality of AVP secretion is limited to the osmoregulatory system since the hormone responds normally to all nonosmotic stimuli, such as nausea.

Differential Diagnosis Adipsic hypernatremia should be distinguished from the hypernatremia that results from various other causes. These distinctions can usually be made from the history, physical examination, and routine laboratory tests. If a conscious patient denies thirst and/or does not drink vigorously in the presence of significant hypernatremia, the diagnosis of hypodipsia or adipsia can be made with confidence. This diagnosis is supported by laboratory evidence of hypovolemia (azotemia, hypokalemia, hyperuricemia, hyperreninemia) and a relative deficiency of plasma[AVP](#). Close monitoring of these variables and urine osmolality during rehydration is useful for differentiating the patients who develop[DI](#) or SIADH in response to forced hydration. If the patient is obtunded or otherwise unable to answer questions or drink at the time of presentation, the possibility of adipsic hypernatremia can be evaluated after treatment by assessing the thirst and plasma AVP response to a controlled fluid deprivation-hypertonic saline infusion test similar to that described for evaluation of DI.

TREATMENT

Adipsic hypernatremia should be treated by administering water by mouth, if the patient is alert, or 0.45% saline by vein, if the patient is obtunded or uncooperative. The number of liters of free water that will be required to correct the deficit (DFW) can be estimated from body weight in kg (BW) and the serum sodium concentration in mmol/L (S_{Na}) by the formula $DFW = 0.5BW \times [(S_{Na} - 140)/140]$. If serum glucose ($S_{Glu}$) is elevated, the

measured S_{Na}should be corrected ($S_{Na}*$) by the formula $S_{Na}*=S_{Na}+[(S_{Glu}- 90)/36]$. This amount plus an allowance for continuing insensible and urinary losses should be given over a 24- to 48-h period. IfDI is present or develops during rehydration, DDAVP should also be given in standard doses to minimize urinary losses. If hyperglycemia and/or hypokalemia are present, insulin and/or potassium supplements should be given. These variables plus urine output and plasma urea/creatinine should be monitored closely during treatment for signs of emerging DI, SIADH, or acute renal failure.

Once the acute fluid and electrolyte imbalances are corrected, anMRI of the brain and tests of anterior pituitary function should be performed. A long-term management plan to prevent or minimize recurrence of the fluid and electrolyte imbalance should also be developed, including a practical method that the patient can use to regulate fluid intake in accordance with day-to-day variations in water balance. The most effective way to accomplish these objectives is to prescribe DDAVP or chlorpropamide to completely controlDI, if it is present, and teach the patient how to use day-to-day changes in body weight as a guide for adjusting fluid intake. Prescribing a constant fluid intake is less satisfactory because it does not take into account the large, uncontrolled variations in insensible loss that inevitably occur.

EXCESS VASOPRESSIN SECRETION AND ACTION

HYPONATREMIA (See also Chap. 49)

Clinical Characteristics Excessive secretion or action ofAVPresults in the production of decreased volumes of more highly concentrated urine. If not accompanied by a commensurate reduction in fluid intake, the reduced suppressibility of AVP results in water retention and a decrease in plasma osmolality/sodium. If the hyponatremia develops gradually or has been present for more than a few days, it may be asymptomatic. However, if it develops acutely, it is almost always accompanied by symptoms and signs of water intoxication that may include mild headache, confusion, anorexia, nausea, vomiting, coma, and convulsions. Severe hyponatremia may be lethal. Depending on the cause of the increased antidiuresis, osmotically inappropriate thirst and/or fluid intake and other disturbances of fluid and electrolyte balance may also be present.

Etiology Osmotically inappropriate antidiuresis can be caused by a primary defect inAVPsecretion or action or can be secondary to a recognized nonosmotic stimulus such as hypovolemia, hypotension, or glucocorticoid deficiency (Table 329-3). The primary forms are generally referred to as SIADH or euvolemic (type III) hyponatremia. They have many different causes, including ectopic production of AVP by lung cancer or other neoplasms, eutopic release by various diseases or drugs, and exogenous administration of AVP, DDAVP, or large doses of oxytocin. The ectopic forms result from abnormal and presumably unregulated expression of the AVP-NPII gene by primary or metastatic malignancies. They do not usually remit unless the ectopic source is eliminated. The eutopic forms manifest most often in patients with acute infections or strokes, but the mechanisms by which these diseases disrupt osmoregulation are not known. A form of acute or chronic hyponatremia very similar to SIADH can also result from stimulation of AVP secretion by protracted nausea or isolated glucocorticoid deficiency. In these patients the excess AVP secretion can be corrected quickly and

completely by specific treatments (antiemetics or glucocorticoids) that are not useful in other forms of SIADH.

The secondary forms of osmotically inappropriate antidiuresis are usually divided into two groups: type I (hypervolemic) and type II (hypovolemic) hyponatremia. Type I occurs in sodium-retaining, edema-forming states such as congestive heart failure, cirrhosis, or nephrosis and is thought to be due to a reduction in "effective" blood volume. Type II occurs in sodium-depleted states such as severe gastroenteritis, diuretic abuse, or mineralocorticoid deficiency and is probably due to a reduction in blood volume and/or pressure.

Pathophysiology In SIADH, interference with the osmotic suppression of AVP results in significant expansion and dilution of body fluids only if water intake exceeds the rate of insensible and urinary output. These abnormalities in water intake often result from an associated defect in the osmoregulation of thirst but can also be due to psychogenic or iatrogenic factors, including the administration of intravenous fluids.

In SIADH, the defect in the osmoregulation of antidiuretic function can take any of four distinct forms (Fig. 329-6). In one of them, AVP secretion remains fully responsive to changes in plasma osmolality/sodium, but the threshold or set point of the osmoregulatory system is abnormally low. Patients with this kind of downward resetting of the osmostat differ from those with the other types of osmoregulatory defect in that they are able to maximally suppress plasma AVP and dilute their urine if their fluid intake is high enough to reduce their plasma osmolality/sodium to the new set point. Another, smaller subgroup (about 10% of the total) do not have a demonstrable defect in the osmoregulation of AVP (Fig. 329-6). Thus, their inappropriate antidiuresis may be due to other abnormalities such as enhanced renal sensitivity to the antidiuretic effect of normally low levels of AVP or activation of aquaporin 2 water channels by a mechanism that is independent of AVP and V_2 receptors.

The extracellular volume expansion that results from excessive retention of water in SIADH also produces an increase in atrial natriuretic hormone, suppression of plasma renin activity, and a compensatory increase in urinary sodium excretion that serves to reduce the hypervolemia but aggravates the hyponatremia. Thus, hyponatremia is due to a decrease in total-body sodium as well as an increase in total-body water. The acute retention of water and fall in plasma sodium also cause a rise in intracellular volume. The resultant brain swelling increases intracranial pressure and probably causes the acute symptoms of water intoxication. After several days, this intracellular volume expansion may be reduced by inactivation or elimination of intracellular solutes, resulting in the remission of symptoms that often occur with hyponatremia of this duration.

In type I (edematous) or type II (hypovolemic) hyponatremia, the osmotic inhibition of AVP and urine concentration is counteracted by a hemodynamic stimulus that results from a substantial reduction in effective or absolute blood volume. In both cases, the inadequate suppression of AVP appears to be due to downward resetting of the osmostat. The resultant antidiuresis is usually enhanced by decreased distal delivery of filtrate that results from increased reabsorption of sodium in proximal nephrons secondary to the hypovolemia. If it is not associated with a commensurate reduction in

water intake, the marked reduction in urine output that ensues also leads to expansion and dilution of body fluids with symptoms of hyponatremia. This attenuates, but does not completely eliminate, the antidiuresis because the amount of water retained is usually insufficient to fully correct the effective or absolute hypovolemia. Unlike SIADH, therefore, plasma renin activity is elevated, causing secondary hyperaldosteronism and hypokalemia. The disturbance in salt and water balance that underlies the hyponatremia also differs from SIADH in that total-body sodium as well as water is increased in type I, whereas both are decreased in type II.

Differential Diagnosis When unexplained symptoms or signs consistent with water intoxication are present, serum sodium should be measured. If it is low and the reduction cannot be accounted for by an increase in plasma glucose or other solutes such as mannitol (e.g., plasma osmolality is also low), the type of hypotonic hyponatremia present can be determined by estimating extracellular fluid volume from the history, physical examination, and routine chemistries. If these findings are ambiguous or contradictory, measuring the rate of urinary sodium excretion or plasma renin activity may be helpful. These measurements can be misleading, however, if SIADH is stable or resolving or if the patient has type II hyponatremia due to a primary defect in renal conservation of sodium, surreptitious diuretic abuse, or hyporeninemic hypoaldosteronism. The latter may be suspected if serum potassium is elevated instead of low as is usually seen in types I and II hyponatremia. Measurements of plasma AVP are currently of no diagnostic value since they exhibit the same wide variation in abnormalities in all three types of hyponatremia. In patients who fulfill the clinical criteria for SIADH, plasma cortisol should also be measured to rule out unsuspected secondary adrenal insufficiency. If this is normal and there is no other obvious cause for SIADH, a careful search for occult lung cancer should also be undertaken.

TREATMENT

In acute SIADH, the keystone to treatment of hyponatremia is to restrict total fluid intake to less than the sum of insensible losses and urinary output. Total intake should include the water derived from food (300 to 500 mL/d). Because insensible losses in adults usually approximate 500 mL/d total discretionary intake (all water in liquid form) should be at least 500 mL less than urinary output. If achieved, this deficit usually reduces body water and increases serum sodium by about 1 to 2% per day. If more rapid correction of the hyponatremia is desired to eliminate severe symptoms or signs, the fluid restriction can be supplemented by intravenous infusion of hypertonic (3%) saline. This treatment has the advantage of correcting the sodium deficiency that is partly responsible for the hyponatremia as well as producing a solute diuresis that serves to remove some of the excess water. However, if the hyponatremia has been present for more than 24 to 48 h and is corrected too rapidly, the saline infusion also has the potential to produce central pontine myelinolysis, an acute, potentially fatal neurologic syndrome characterized by quadriparesis, ataxia, and abnormal extraocular movements (Chap. 376). The following guidelines appear to minimize, if not eliminate, the risk of this complication: the 3% saline should be infused a rate £0.05 mL/kg body weight per minute; the effect should be monitored continuously by STAT measurements of serum sodium at least once every hour; and the infusion should be stopped as soon as serum sodium increases by 12 mmol/L or to 130 mmol/L, whichever comes first. Urinary output should also be monitored continuously since spontaneous remission of the SIADH can result in an

acute water diuresis that greatly accelerates the rate of rise in serum sodium produced by fluid restriction and 3% saline infusion.

In chronic SIADH, the hyponatremia can be minimized or eliminated by treatment with demeclocycline, 150 to 300 mg orally three or four times a day, or fludrocortisone, 0.05 to 0.2 mg orally twice a day. The effect of the demeclocycline manifests in 7 to 14 days and is due to production of a reversible form of nephrogenicDI. Potential side effects include phototoxicity and azotemia. The effect of fludrocortisone also manifests in 1 to 2 weeks and is partly due to increased retention of sodium and possibly inhibition of thirst. It also increases urinary potassium excretion, which may require replacement through dietary adjustments or supplements. Fludrocortisone may induce hypertension, occasionally necessitating discontinuation of the treatment.

One or more nonpeptideAVPantagonists that block the antidiuretic effect of AVP may soon be approved for use in the United States. Preliminary studies with these antagonists in acute or chronic SIADH indicate that they produce a dose-dependent increase in urinary free-water excretion, which, if combined with a modest restriction of fluid intake, gradually reduces body water and corrects the hyponatremia without any recognized adverse effect. Thus, they may become the treatment of choice for those forms of SIADH in which there is inappropriate secretion of AVP that cannot be corrected by other, more specific therapy such as antiemetics or glucocorticoids.

When an SIADH-like syndrome is due to protracted nausea and vomiting or isolated glucocorticoid deficiency, all abnormalities can be corrected quickly and completely by giving an antiemetic or hydrocortisone. As with other treatments, care must be taken to ensure that serum sodium does not rise too quickly or too far.

In type I hyponatremia, the only treatment currently available is severe fluid restriction, administration of urea or mannitol to produce a solute diuresis, and/or administration of cardiotonics or serum albumin to correct the effective hypovolemia. None of these treatments is particularly effective, and some (e.g., administration of mannitol or albumin) carry significant risks. Infusion of hypertonic saline is contraindicated because it worsens the sodium retention and edema and may precipitate cardiovascular decompensation. However, preliminary studies indicate that theAVPantagonists may be almost as effective and safe in type I hyponatremia as they are in SIADH. Thus, they may become the treatment of choice for this form of hyponatremia also.

In type II hyponatremia, the defect inAVPsecretion and water balance can usually be corrected easily and quickly by stopping the loss of sodium and water and/or replacing the deficits by mouth or intravenous infusion of normal or hypertonic saline. As with the treatment of other forms of hyponatremia, care must be taken to ensure that plasma sodium does not increase too rapidly. Fluid restriction or administration of AVP antagonists is contraindicated as they would only aggravate the underlying volume depletion and could result in cardiovascular decompensation.

(Bibliography omitted in Palm version)

330. DISORDERS OF THE THYROID GLAND - *J. Larry Jameson*, *Anthony P. Weetman*

The thyroid gland produces two related hormones, thyroxine (T_4) and triiodothyronine (T_3) ([Fig. 330-1](#)). These hormones play a critical role in cell differentiation during development and help to maintain thermogenic and metabolic homeostasis in the adult. Thyroid hormones act through nuclear hormone receptors to modulate gene expression. Disorders of the thyroid gland result primarily from autoimmune processes that either stimulate the overproduction of thyroid hormones (*thyrotoxicosis*) or cause glandular destruction and underproduction of thyroid hormones (*hypothyroidism*). In addition, neoplastic processes in the thyroid gland can lead to benign nodules and various forms of thyroid cancer.

ANATOMY AND DEVELOPMENT

The thyroid gland is located in the neck, anterior to the trachea, between the cricoid cartilage and the suprasternal notch. The thyroid (Greek *thyreos*, shield, plus *eidos*, form) consists of two lobes that are connected by an isthmus. It is normally 12 to 20 g in size, highly vascular, and soft in consistency. Four parathyroid glands, which produce parathyroid hormone ([Chap. 341](#)), are located in the posterior region of each pole of the thyroid. The recurrent laryngeal nerves traverse the lateral borders of the thyroid gland and must be identified during thyroid surgery to avoid vocal cord paralysis.

The thyroid gland develops from the floor of the primitive pharynx during the third week of gestation. The gland migrates from the foramen cecum, at the base of the tongue, along the thyroglossal duct to reach its final location in the neck. This feature accounts for the rare ectopic location of thyroid tissue at the base of the tongue (lingual thyroid), as well as for the presence of thyroglossal duct cysts along this developmental tract. Thyroid hormone synthesis normally begins at about 11 weeks' gestation.

The parathyroid glands migrate from the third (inferior glands) and fourth (superior glands) pharyngeal pouches before becoming embedded in the thyroid gland. Neural crest derivatives from the ultimobranchial body give rise to thyroid medullary C cells that produce calcitonin, a calcium-lowering hormone. The C cells are interspersed throughout the thyroid gland, although their density is greatest in the juncture of the upper one-third and lower two-thirds of the gland.

Thyroid gland development is controlled by a series of developmental transcription factors. Thyroid transcription factor (TTF) 1 (also known as NKX2A), TTF-2 (also known as FKHL15), and paired homeobox-8 (PAX-8) are expressed selectively, but not exclusively, in the thyroid gland. In combination, they orchestrate thyroid cell development and the induction of thyroid-specific genes such as thyroglobulin (Tg), thyroid peroxidase (TPO), the sodium iodide symporter (NIS), and the thyroid-stimulating hormone receptor (TSH-R). Mutations in these developmental transcription factors or their downstream target genes are rare causes of thyroid agenesis or dyshormonogenesis and can cause congenital hypothyroidism ([Table 330-1](#)). Congenital hypothyroidism is common enough (approximately 1 in 3000 to 4000 newborns) that neonatal screening is now performed in most industrialized countries (see below). Though the underlying causes of most cases of congenital hypothyroidism

are unknown, early treatment with thyroid hormone replacement precludes potentially severe developmental abnormalities.

The mature thyroid gland contains numerous follicles composed of thyroid follicular cells that surround secreted colloid, a proteinaceous fluid that contains large amounts of thyroglobulin, the protein precursor of thyroid hormones (Fig. 330-2). The thyroid follicular cells are polarized -- the basolateral surface is apposed to the bloodstream and an apical surface faces the follicular lumen. Increased demand for thyroid hormone, usually signaled by thyroid-stimulating hormone (TSH) binding to its receptor on the basolateral surface of the follicular cells, leads to Tg reabsorption from the follicular lumen and proteolysis within the cell to yield thyroid hormones for secretion into the bloodstream.

REGULATION OF THE THYROID AXIS

TSH, secreted by the thyrotrope cells of the anterior pituitary, plays a pivotal role in control of the thyroid axis and serves as the most useful physiologic marker of thyroid hormone action. TSH is a 31-kDa hormone composed of a and b subunits; the a subunit is common to the other glycoprotein hormones [luteinizing hormone, follicle-stimulating hormone, human chorionic gonadotropin (hCG)], whereas the TSH b subunit is unique to TSH. The extent and nature of carbohydrate modification are modulated by thyrotropin-releasing hormone (TRH) stimulation and influence the biologic activity of the hormone. TSH has been produced recombinantly and is approved for use in the detection of residual thyroid cancer (see "Follow-up Whole-Body Scanning and Thyroglobulin Determinations," below).

The thyroid axis is a classic example of an endocrine feedback loop. Hypothalamic TRH stimulates pituitary production of TSH, which, in turn, stimulates thyroid hormone synthesis and secretion. Thyroid hormones feed back negatively to inhibit TRH and TSH production (Fig. 330-2). The "set-point" in this axis is established by TSH, the level of which is a sensitive and specific marker of thyroid function. TRH is the major positive regulator of TSH synthesis and secretion. TRH acts through a seven-transmembrane G protein-coupled receptor (GPCR) that activates phospholipase C to generate phosphatidylinositol turnover and the release of intracellular calcium. Peak TSH secretion occurs ~15 min after administration of exogenous TRH. Dopamine, glucocorticoids, and somatostatin suppress TSH but are not of major physiologic importance except when these agents are administered in pharmacologic doses. Reduced levels of thyroid hormone increase basal TSH production and enhance TRH-mediated stimulation of TSH. High thyroid hormone levels rapidly and directly suppress TSH and inhibit TRH-mediated stimulation of TSH, indicating that thyroid hormones are the dominant regulator of TSH production. Like other pituitary hormones, TSH is released in a pulsatile manner and exhibits a diurnal rhythm; its highest levels occur at night. However, these TSH excursions are modest in comparison to those of other pituitary hormones, in part because TSH has a relatively long plasma half-life (50 min). Consequently, single measurements of TSH are adequate for assessing its circulating level. TSH is measured using immunoradiometric assays that are highly sensitive and specific. These assays are capable of distinguishing between normal and suppressed TSH values, thus allowing TSH to be used for the diagnosis of hyperthyroidism (low TSH) as well as hypothyroidism (high TSH).

THYROID HORMONE SYNTHESIS, METABOLISM, AND ACTION

THYROID HORMONE SYNTHESIS

Thyroid hormones are derived from Tg, a large iodinated glycoprotein. After secretion into the thyroid follicle, Tg is iodinated on selected tyrosine residues that are subsequently coupled via an ether linkage. Reuptake of Tg into the thyroid follicular cell allows proteolysis and the release of T_4 and T_3.

Iodine Metabolism and Transport Iodide uptake is a critical first step in thyroid hormone synthesis. Ingested iodine is bound to serum proteins, particularly albumin. Unbound iodine is excreted in the urine. The thyroid gland extracts iodine from the circulation in a highly efficient manner. For example, 10 to 25% of radioactive tracer (e.g., $_{123}$I) is taken up by the normal thyroid gland over 24 h; this value can rise to 70 to 90% in Graves' disease.

Iodide uptake is mediated by the Na_+/I-symporter (NIS), which is expressed at the basolateral membrane of thyroid follicular cells. NIS is most highly expressed in the thyroid gland but is also expressed at low levels in the salivary glands, lactating breast, and placenta. The iodide transport mechanism is highly regulated, allowing adaptation to variations in dietary supply. Low iodine levels increase the amount of NIS and stimulate uptake, whereas high iodine levels suppress NIS expression and uptake. The selective expression of the NIS in the thyroid allows isotopic scanning, treatment of hyperthyroidism, and ablation of thyroid cancer with radioisotopes of iodine, without significant effects on other organs. Mutation of the *NIS* gene is a rare cause of congenital hypothyroidism, underscoring its importance in thyroid hormone synthesis.

Iodine deficiency is prevalent in many mountainous regions and in central Africa, central South America, and northern Asia. In areas of relative iodine deficiency, there is an increased prevalence of goiter and, when deficiency is severe, hypothyroidism and cretinism. *Cretinism* is characterized by mental and growth retardation and occurs when children who live in iodine-deficient regions are not treated with iodine or thyroid hormone to restore normal thyroid hormone levels during early childhood. These children are often born to mothers with iodine deficiency, suggesting that maternal thyroid hormone deficiency worsens the condition. Concomitant selenium deficiency may also contribute to the neurologic manifestations of cretinism. Iodine supplementation of salt, bread, and other food substances has markedly reduced the prevalence of cretinism. Unfortunately, however, iodine deficiency remains the most common cause of preventable mental deficiency, often because of resistance to the use of food additives or the cost of supplementation. In addition to overt cretinism, mild iodine deficiency can lead to subtle reduction of IQ. Iodine intake is assessed by determination of excretion in a 24-h urine collection. Oversupply of iodine, through supplements or foods enriched in iodine (e.g., shellfish, kelp), is associated with an increased incidence of autoimmune thyroid disease. The recommended average daily intake of iodine is 150 ug/d for adults, 90 to 120 ug/d for children, and 200 ug/d for pregnant women.

Organification, Coupling, Storage, Release After iodide enters the thyroid, it is

trapped and transported to the apical membrane of thyroid follicular cells where it is oxidized in an organification reaction that involves TPO and hydrogen peroxide. The reactive iodine atom is added to selected tyrosyl residues within Tg, a large (660 kDa) dimeric protein consisting of 2769 amino acids. The iodotyrosines in Tg are then coupled via an ether linkage in a reaction that is also catalyzed by TPO. Either T_4 or T_3 can be produced by this reaction, depending on the number of iodine atoms present in the iodotyrosines. After coupling, Tg is taken back into the thyroid cell where it is processed in lysosomes to release T_4 and T_3. Uncoupled mono- and diiodotyrosines (MIT, DIT) are deiodinated by the enzyme dehalogenase, thereby recycling any iodide that is not converted into thyroid hormones.

Disorders of thyroid hormone synthesis are rare causes of congenital hypothyroidism. The vast majority of these disorders are due to recessive mutations in TPO or Tg, but defects have also been identified in the TSH-R, NIS, the pendrin anion transporter, hydrogen peroxide generation, and in dehalogenase. Because of the biosynthetic defect, the gland is incapable of synthesizing adequate amounts of hormone, leading to increased TSH and a large goiter.

TSH Action TSH regulates thyroid gland function through the TSH-R, a seven-transmembrane GPCR. The TSH-R is coupled to the a subunit of stimulatory G protein (G_{sa}) and activates adenylyl cyclase, leading to increased production of cyclic AMP. TSH also stimulates phosphatidylinositol turnover by activating phospholipase C. The functional role of the TSH-R has been underscored by naturally occurring mutations. Recessive loss-of-function mutations are a rare cause of thyroid hypoplasia and congenital hypothyroidism. Dominant gain-of-function mutations cause sporadic or familial nonautoimmune hyperthyroidism that is characterized by goiter, thyroid cell hyperplasia, and autonomous function. Most of these activating mutations involve in amino acid substitutions in the transmembrane domain of the receptor. They are thought to mimic conformational changes in the receptor similar to those induced by TSH binding or the interactions of thyroid-stimulating immunoglobulins (TSI) in Graves' disease. Activating TSH-R mutations also occur as somatic events and lead to clonal selection and expansion of the affected thyroid follicular cell (see below).

Factors that Influence Hormone Synthesis and Release TSH is the dominant hormonal regulator of thyroid gland growth and function. However, a variety of growth factors, most produced locally in the thyroid gland, also influence thyroid hormone synthesis. These include insulin-like growth factor I (IGF-I), epidermal growth factor, transforming growth factorb (TGF-b), endothelins, and various cytokines. The quantitative roles of these factors are not well understood, but they are important in selected disease states. In acromegaly, for example, increased levels of growth hormone and IGF-I are associated with goiter and predisposition to multinodular goiter. Certain cytokines and interleukins (ILs) produced in association with autoimmune thyroid disease induce thyroid growth, whereas others lead to apoptosis. As noted above, iodine is an important regulator of thyroid function. For example, iodine deficiency increases thyroid blood flow and stimulates uptake by the NIS. Excess iodide transiently inhibits thyroid iodide organification, a phenomenon known as the *Wolff-Chaikoff effect*. In individuals with a normal thyroid, the gland escapes from this inhibitory effect and iodide organification resumes; the suppressive action of high iodide may persist, however, in patients with underlying autoimmune thyroid disease.

THYROID HORMONE TRANSPORT AND METABOLISM

Serum Binding Proteins T_4 is secreted from the thyroid gland in at least 20-fold excess over T_3(Table 330-2). Both hormones circulate bound to plasma proteins, including thyroxine-binding globulin (TBG), transthyretin (TTR, formerly known as thyroxine-binding prealbumin, or TBPA), and albumin. The functions of serum-binding proteins are to increase the pool of circulating hormone, delay hormone clearance, and perhaps to modulate hormone delivery to selected tissue sites. The concentration of TBG is relatively low (1 to 2 mg/dL), but because of its high affinity for thyroid hormones ($T_4>T_3$), it carries about 80% of the bound hormones. Albumin has relatively low affinity for thyroid hormones but has a high plasma concentration (~3.5 g/dL), and it binds up to 10% of T_4 and 30% of T_3. TTR also carries about 10% of T_4 but little T_3.

When the effects of the various binding proteins are combined, approximately 99.98% of $T_4$and 99.7% of T_3 are protein-bound. Because T_3 is less tightly bound than T_4, the amount of free T_3 is greater than free T_4, even though there is less total T_3 in the circulation. The unbound, or free, concentrations of the hormones are ~$2´10_{-11}M$ for T_4 and ~$6´10_{-12}M$ for T_3, which roughly correspond to the thyroid hormone receptor binding constants for these hormones (see below). Only the free hormone is biologically available to tissues. Therefore, homeostatic mechanisms that regulate the thyroid axis are directed towards maintenance of normal concentrations of free hormones.

Dysalbuminemic Hyperthyroxinemia A number of inherited and acquired abnormalities affect thyroid hormone binding proteins. X-linkedTBGdeficiency is associated with very low levels of total T_4 and T_3. However, because free hormone levels are normal, patients are euthyroid andTSHlevels are normal. The importance of recognizing this disorder is to avoid efforts to normalize total T_4levels, as this leads to thyrotoxicosis and is futile because of rapid hormone clearance in the absence of TBG. TBG levels are elevated by estrogen because of increased sialylation and delayed TBG clearance. Consequently, in women who are pregnant or taking estrogen-containing contraceptives, elevated TBG increases total $T_4$and T_3levels; however, free T_4 and T_3levels are normal. Mutations in TBG,TTR, and albumin that increase binding affinity for $T_4$and/or T_3 cause disorders known as *euthyroid hyperthyroxinemia* or *familial dysalbuminemic hyperthyroxinemia* (FDH) (Table 330-3). These disorders are usually dominantly transmitted and result in increased total $T_4$and/or T_3, but free hormone levels are normal. The familial nature of the disorders, and the fact that TSH levels are normal rather than suppressed, should suggest the diagnosis. Free hormone levels (ideally measured by dialysis) are normal in FDH. The diagnosis can be confirmed, if necessary, by using tests that measure the affinities of radiolabeled hormone binding to specific transport proteins or by performing DNA sequence analyses of the abnormal transport protein genes.

Certain medications, such as salicylates and salsalate, can displace thyroid hormones from circulating binding proteins. Though these drugs transiently perturb the thyroid axis by increasing free thyroid hormone levels,TSH is suppressed until a new steady state is reached, thereby restoring euthyroidism. Circulating factors associated with acute illness may also displace thyroid hormone from binding proteins (see "Sick Euthyroid Syndrome," below).

Deiodinases In many respects, T_4 may be thought of as a precursor for the more potent T_3. T_4 is converted to T_3 by the deiodinase enzymes ([Fig. 330-1](#)). Type I deiodinase, which is located primarily in thyroid, liver, and kidney, has a relatively low affinity for T_4. Type II deiodinase has a higher affinity for T_4 and is found primarily in the pituitary gland, brain, brown fat, and thyroid gland. The presence of type II deiodinase allows it to regulate T_3concentrations locally, a property that may be important in the context of levothyroxine (T_4) replacement. Type II deiodinase is also regulated by thyroid hormone -- hypothyroidism induces the enzyme, resulting in enhanced T_4®T_3conversion in tissues such as brain and pituitary. T_4®T_3conversion may be impaired by fasting, systemic illness or acute trauma, oral contrast agents, and a variety of medications (e.g., propylthiouracil, propranolol, amiodarone, glucocorticoids). Type III deiodinase inactivates T_4 and T_3 and is the most important source of reverse T_3(rT_3).

THYROID HORMONE ACTION

Nuclear Thyroid Hormone Receptors Thyroid hormones act by binding to nuclear receptors, termed *thyroid hormone receptors* (TRs) a andb. Both TRa and TRb are expressed in most tissues, but their relative levels of expression vary among organs; TRa is particularly abundant in brain, kidney, gonads, muscle, and heart, whereas TRb expression is relatively high in the pituitary and liver. Both receptors are variably spliced to form unique isoforms. The TRb2 isoform, which has a unique amino terminus, is selectively expressed in the hypothalamus and pituitary, where it appears to play a role in feedback control of the thyroid axis. The TRa2 isoform contains a unique carboxy terminus that prevents thyroid hormone binding; it may function to block the action of other TR isoforms.

The[TRs](#)contain a central DNA-binding domain and a C-terminal ligand-binding domain. They bind to specific DNA sequences, termed *thyroid response elements* (TREs), in the promoter regions of target genes ([Fig. 330-3](#)). The activated receptor can either stimulate gene transcription (e.g., myosin heavy chaina) or inhibit transcription (e.g.,[TSH](#)b-subunit gene), depending on the nature of the regulatory elements in the target gene. The receptors bind as homodimers or as heterodimers with retinoic acid X receptors (RXRs) ([Chap. 327](#)).

Thyroid hormones bind with similar affinities to[TR](#)a and TRb. However, T_3 is bound to its receptors with about 10 to 15 times greater affinity than T_4, which explains its increased hormonal potency. Though T_4 is produced in excess of T_3, receptors are occupied mainly by T_3, reflecting T_4®T_3conversion by peripheral tissues, greater T_3bioavailability in the plasma, and receptors' greater affinity for T_3. After binding to TRs, thyroid hormone induces conformational changes in the receptors that modify its interactions with accessory transcription factors. In the absence of thyroid hormone binding, the aporeceptors bind to corepressor proteins that inhibit gene transcription. Hormone binding dissociates the corepressors and allows the recruitment of coactivators that enhance transcription. The discovery of TR interactions with corepressors explains the fact that TR silences gene expression in the absence of hormone binding. Consequently, hormone deficiency has a profound effect on gene expression because it causes active gene repression as well as loss of hormone-induced stimulation. This concept has been corroborated by the finding that targeted deletion of the TR genes in

mice has a less pronounced phenotypic effect than hormone deficiency.

Thyroid Hormone Resistance Resistance to thyroid hormone (RTH) is an autosomal dominant disorder characterized by elevated free thyroid hormone levels and inappropriately normal or elevatedTSH. Individuals with RTH do not, in general, exhibit signs and symptoms that are typical of hypothyroidism, apparently because hormone resistance is compensated by increased levels of thyroid hormone. The clinical features of RTH can include goiter, attention deficit disorder, mild reduction in IQ, delayed skeletal maturation, tachycardia, and impaired metabolic responses to thyroid hormone.

The disorder is caused by mutations in the TRbreceptor gene. These mutations, located in restricted regions of the ligand-binding domain, cause loss of receptor function. However, because the mutant receptors retain the capacity to dimerize withRXRs, bind to DNA, and recruit corepressor proteins, they function as antagonists of the remaining, normal TRb and TRa receptors. This property, referred to as "dominant negative" activity, explains the autosomal dominant mode of transmission. The diagnosis is suspected when free thyroid hormone levels are increased without suppression ofTSH. Similar hormonal abnormalities are common in other affected family members, though the TRb mutation arises de novo in about 20% of patients. DNA sequence analysis of the TRb gene provides a definitive diagnosis.RTH must be distinguished from other causes of euthyroid hyperthyroxinemia (e.g., familial dysalbuminemic hyperthyroxinemia) and inappropriate secretion of TSH by TSH-secreting pituitary adenomas (Chap. 328). In most patients, no treatment is indicated; the importance of making the diagnosis is to avoid inappropriate treatment of mistaken hyperthyroidism and to provide genetic counseling.

PHYSICAL EXAMINATION

In addition to the examination of the thyroid itself, the physical examination should include a search for signs of abnormal thyroid function and the extrathyroidal features of ophthalmopathy and dermopathy (see below). Examination of the neck begins by inspecting the seated patient from the front and side, and noting any surgical scars, obvious masses, or distended veins. The thyroid can be palpated with both hands from behind or the examiner can face the patient, using the thumbs to palpate each lobe. Most often it is best to use a combination of these methods, especially in cases of doubt or when there are small nodules. The patient's neck should be slightly flexed to relax the neck muscles. After locating the cricoid cartilage, the isthmus can be identified and followed laterally to locate either lobe (the right lobe is normally slightly larger than the left). By asking the patient to swallow sips of water, thyroid consistency can be better appreciated as the gland moves beneath the examiner's fingers.

Features to be noted include thyroid size, consistency, nodularity, and any tenderness or fixation. An estimate of thyroid size (normally 12 to 20 g) should be made, and a drawing is often the best way to record findings. However, ultrasound is the method of choice when it is important to determine thyroid size accurately. The size, location, and consistency of any nodules should also be depicted. A bruit over the gland indicates increased vascularity, as occurs in hyperthyroidism. If the lower borders of the thyroid lobes are not clearly felt, a goiter may be retrosternal. Large retrosternal goiters can cause venous distention over the neck and difficulty breathing, especially when the

arms are raised (Pemberton's sign). With any central mass above the thyroid, the patient should be asked to stick out his or her tongue, as thyroglossal cysts then move upward. The thyroid examination is not complete without assessment for lymphadenopathy in the supraclavicular and cervical regions of the neck.

LABORATORY EVALUATION

MEASUREMENT OF THYROID HORMONES

The enhanced sensitivity and specificity of *TSH assays* have greatly improved laboratory assessment of thyroid function. Because TSH levels change dynamically in response to alterations of free T_4 and T_3, a logical approach to thyroid testing is to determine first whether TSH is suppressed, normal, or elevated. With rare exceptions (see below), a normal TSH level excludes a primary abnormality of thyroid function. This strategy depends on the use of immunoradiometric assays (IRMAs) for TSH that are sensitive enough to discriminate between the lower limit of the reference range and the suppressed values that occur with thyrotoxicosis. Extremely sensitive (fourth generation) assays can detect TSH levels £0.004 mU/L, but for practical purposes assays sensitive to £0.1 mU/L are sufficient. The widespread availability of the TSH IRMA has rendered the TRH stimulation test virtually obsolete, as the failure of TSH to rise after an intravenous bolus of 200 to 400 ug TRH has the same implications as a suppressed basal TSH measured by IRMA.

The finding of an abnormal TSH level must be followed by measurements of circulating thyroid hormone levels to confirm the diagnosis of hyperthyroidism (suppressed TSH) or hypothyroidism (elevated TSH). Radioimmunoassays are widely available for serum *total T_4* and *total T_3*. T_4 and T_3 are highly protein-bound, and numerous factors (illness, medications, genetic factors) can influence protein binding. It is useful, therefore, to measure the free or unbound hormone levels, which correspond to the biologically available hormone pool. Two direct methods are used to measure *free thyroid hormones*: (1) free thyroid hormone competition with radiolabeled T_4 (or an analogue) for binding to a solid-phase antibody, and (2) physical separation of the free hormone fraction by ultracentrifugation or equilibrium dialysis. Though early free hormone immunoassays suffered from artifacts, newer assays agree well with the results of the more technically demanding and expensive physical separation methods. An indirect method to estimate free thyroid hormone levels is to calculate the free T_3 or free T_4 index from the total T_4 or T_3 concentration and the *thyroid hormone binding ratio* (THBR). The latter is derived from the *T_3-resin uptake test*, which determines the distribution of radiolabeled T_3 between an absorbent resin and the unoccupied thyroid hormone binding proteins in the sample. The binding of the labeled T_3 to the resin is increased when there is reduced unoccupied protein binding sites (e.g., TBG deficiency) or increased total thyroid hormone in the sample; it is decreased under the opposite circumstances. The product of THBR and total T_3 or T_4 provides the *free T_3 or T_4 index*. In effect, the index corrects for anomalous total hormone values caused by abnormalities in hormone-protein binding.

Total thyroid hormone levels are elevated when TBG is increased due to estrogens (pregnancy, oral contraceptives, hormone replacement therapy, tamoxifen), and decreased when TBG binding is decreased (androgens, the nephrotic syndrome).

Genetic disorders and acute illness can also cause abnormalities in thyroid hormone binding proteins, and various drugs (phenytoin, carbamazepine, salicylates, and nonsteroidal anti-inflammatory drugs) can interfere with thyroid hormone binding. Because free thyroid hormone levels are normal and the patient is euthyroid in all of these circumstances, assays that measure free hormone are preferable to those for total thyroid hormones.

For most purposes, the free T_4 level is sufficient to confirm thyrotoxicosis, but 2 to 5% of patients have only an elevated T_3 level (T_3toxicosis). Thus, free T_3levels should be measured in patients with a suppressedTSH but normal free T_4levels. Free T_3levels are normal in about 25% of patients with hypothyroidism and provide little useful information in this setting.

There are several clinical conditions in which the use ofTSH as a screening test may be misleading, particularly without simultaneous free T_4determinations. Any severe nonthyroidal illness can cause abnormal TSH levels (see below). Although hypothyroidism is the most common cause of an elevated TSH level, rare causes include a TSH-secreting pituitary tumor (Chap. 328), thyroid hormone resistance, and assay artifact. Conversely, a suppressed TSH level, particularly<0.1 mU/L, usually indicates thyrotoxicosis but may also be seen during the first trimester of pregnancy (due to hCG secretion), after treatment of hyperthyroidism (because TSH remains suppressed for several weeks), and in response to certain medications (e.g., high doses of glucocorticoids or dopamine). Importantly, secondary hypothyroidism, caused by hypothalamic-pituitary disease, is associated with a variable (low to high-normal) TSH level, which is inappropriate for the low free T_4level. Thus, *TSH should not be used to assess thyroid function in patients with suspected or known pituitary disease*.

Tests for the end-organ effects of thyroid hormone excess or depletion, such as estimation of basal metabolic rate, tendon reflex speed, or serum cholesterol, are not useful as clinical determinants of thyroid function.

TESTS TO DETERMINE THE ETIOLOGY OF THYROID DYSFUNCTION

Autoimmune thyroid disease is detected most easily by measuring circulating antibodies againstTPO andTg. As antibodies to Tg alone are rare, it is reasonable to measure only TPO antibodies. About 5 to 15% of euthyroid women and up to 2% of euthyroid men have thyroid antibodies; such individuals are at increased risk of developing thyroid dysfunction. Almost all patients with autoimmune hypothyroidism, and up to 80% of those with Graves' disease, have TPO antibodies, usually at high levels.

TSIare antibodies that stimulate theTSH-R in Graves' disease. They can be measured in bioassays or indirectly in assays that detect antibody binding to the receptor. The main use of these assays is to predict neonatal thyrotoxicosis caused by high maternal levels of TSI in the last trimester of pregnancy.

Serum Tg levels are increased in all types of thyrotoxicosis except thyrotoxicosis factitia. The main role forTgmeasurement, however, is in the follow-up of thyroid cancer patients. After total thyroidectomy and radioablation, Tg levels should be undetectable; measurable levels (>1 to 2 ng/mL) suggest incomplete ablation or recurrent cancer.

RADIOIODINE UPTAKE AND THYROID SCANNING

The thyroid gland selectively transports radioisotopes of iodine (^{123}I, ^{125}I, ^{131}I) and ^{99m}Tc pertechnetate, allowing thyroid imaging and quantitation of radioactive tracer fractional uptake.

Graves' disease is characterized by an enlarged gland and increased tracer uptake that is distributed homogeneously. Toxic adenomas appear as focal areas of increased uptake, with suppressed tracer uptake in the remainder of the gland. In toxic multinodular goiter, the gland is enlarged -- often with distorted architecture -- and there are multiple areas of relatively increased or decreased tracer uptake. Subacute thyroiditis is associated with very low uptake because of follicular cell damage and TSH suppression. *Thyrotoxicosis factitia*, caused by self-administration of thyroid hormone, is also associated with low uptake.

Although the use of fine-needle aspiration (FNA) biopsy has diminished the use of thyroid scans in the evaluation of solitary thyroid nodules, the functional features of thyroid nodules have some prognostic significance. So-called cold nodules, which have diminished tracer uptake, are usually benign. However, these nodules are more likely to be malignant (~5 to 10%) than so-called hot nodules, which are almost never malignant.

Thyroid scanning is also used in the follow-up of thyroid cancer. After thyroidectomy and ablation using ^{131}I, there is diminished radioiodine uptake in the thyroid bed, allowing the detection of metastatic thyroid cancer deposits that retain the ability to transport iodine. Whole-body scans using 111 to 185 MBq (3 to 5 mCi) ^{131}I are typically performed after thyroid hormone withdrawal to raise the TSH level or after the administration of recombinant human TSH.

THYROID ULTRASOUND

Ultrasonography is used increasingly to assist in the diagnosis of nodular thyroid disease, a reflection of the limitations of the physical examination and improvements in ultrasound technology. Using 10-MHz instruments, spatial resolution and image quality are excellent, allowing the detection of nodules and cysts >3 mm. In addition to detecting thyroid nodules, ultrasound is useful for monitoring nodule size, for guiding FNA biopsies, and for the aspiration of cystic lesions. Ultrasound is also used in the evaluation of recurrent thyroid cancer, including possible spread to cervical lymph nodes.

AUTOIMMUNE BASIS OF THYROID DISEASE

PREVALENCE

Thyroid autoimmunity can cause several forms of thyroiditis and may lead to hypothyroidism as well as Graves' disease. Focal thyroiditis is present in 20 to 40% of autopsy cases and is associated with serologic evidence of autoimmunity, particularly the presence of TPO antibodies. These antibodies are 4 to 10 times more common in otherwise healthy women than men. About 5% of women experience self-limited *postpartum (silent) thyroiditis* in the months after pregnancy, often with transient clinical

symptoms. This condition is associated with the presence of TPO antibodies ante-partum. Up to 20% of women with an episode of postpartum thyroiditis develop permanent hypothyroidism 5 to 10 years after delivery. Autoimmune-mediated hypothyroidism affects about 5 to 10% of middle-aged and elderly women, depending on diagnostic criteria and geographic location. Graves' disease is about one-tenth as common as hypothyroidism and tends to occur in younger individuals. Although seemingly diverse, these disorders have many pathophysiologic features in common, and patients may progress from one state to the other as the autoimmune process changes.

SUSCEPTIBILITY FACTORS

As with most autoimmune disorders, susceptibility is determined by a combination of genetic and environmental factors. The concordance rate for Graves' disease in monozygotic twins is 20 to 30%. The risk of autoimmune thyroid disease is increased among siblings, who may exhibit features of either Graves' disease or autoimmune hypothyroidism. The autoimmune polyglandular syndrome type 2 (Chap. 339) involves the occurrence of autoimmune thyroid dysfunction with other autoimmune diseases (type 1 diabetes mellitus, Addison's disease, pernicious anemia, vitiligo). Shared genetic factors are likely in this group of autoimmune disorders.

HLA-DR3 is the best documented genetic risk factor for Graves' disease and autoimmune hypothyroidism in Caucasians, though different HLA associations exist for other racial groups, such as the Japanese and Chinese. A weak association with polymorphisms in the T cell regulatory gene CTLA-4 has been found in several racial groups. Other loci, including a region on chromosome 18q21, may be linked to Graves' disease as well as to several other autoimmune disorders such as type 1 diabetes mellitus, rheumatoid arthritis, and systemic lupus erythematosus (SLE). The female preponderance of thyroid autoimmunity is most likely due to the influence of sex steroids. Some studies suggest an association between antecedent major life events and Graves' disease, but a causal role for stress in the autoimmune process remains to be clearly established. Smoking is a minor risk factor for Graves' disease but a major risk factor for the development of ophthalmopathy. There is no convincing evidence for a role of infection in susceptibility, except for the congenital rubella syndrome, which is associated with a high frequency of autoimmune hypothyroidism. Viral thyroiditis does not induce subsequent autoimmune thyroid disease.

HUMORAL FACTORS

The thyrotoxicosis of Graves' disease is caused by TSH-R-stimulating immunoglobulins that bind to the receptor and mimic the action of TSH. These TSI can cross the placenta and cause *transient neonatal thyrotoxicosis*, a phenomenon that complicates 1 to 2% of pregnancies in women with active or previous Graves' disease. However, the autoimmune response against the TSH-R can also result in antibodies that block TSH function, causing hypothyroidism. Stimulating and blocking antibodies bind to separate epitopes on the receptor. TSH-R blocking antibodies are found in about 20% of Asian patients with autoimmune hypothyroidism and are associated with thyroid atrophy; blocking antibodies are less common in Caucasians. Patients may have a mixture of TSH-R antibodies, and thyroid function can oscillate between hyperthyroidism and

hypothyroidism as stimulating or blocking antibodies become dominant. Predicting the course of disease in such individuals is difficult, and close monitoring of thyroid function is required. Assays that measure the binding of antibodies to the receptor by competition with radiolabeled TSH [TSH-binding inhibiting immunoglobulins (TBII)] provide no information about functional effects and are used primarily to demonstrate the presence of TSH-R antibodies in atypical patients. Bioassays measure antibody-mediated stimulation of cyclic AMP production in cultured thyroid cells or cells transfected with the TSH-R. The use of these assays does not generally alter clinical management.

Antibodies to Tg and TPO, readily measured by immunofluorescence, hemagglutination, enzyme-linked immunosorbent assay, or radioimmunoassay, are clinically useful markers of thyroid autoimmunity, as discussed above. Any pathogenic effect is likely to be restricted to a secondary role in amplifying an ongoing autoimmune response. For instance, T cell- or cytokine-mediated injury to thyroid follicles could expose the enzyme on the apical border of follicles to TPO antibodies, which may then bind to the autoantigen and fix complement. There is evidence for intrathyroidal complement activation in both Graves' disease and autoimmune hypothyroidism. Tg antibodies do not fix complement, but could be involved in antibody-dependent, natural killer cell-mediated cytotoxicity. The NIS is also a target of autoantibody production in up to one-third of patients with autoimmune thyroid disease, but the functional consequences, if any, have not been established.

CELL-MEDIATED FACTORS

Activated circulating T cells are increased in autoimmune thyroid disease, and the gland is infiltrated with CD4+ and CD8+ T cells. The latter are believed to mediate perforin-dependent cytotoxicity, leading ultimately to thyroid cell destruction. In addition, thyroid cells undergo apoptosis through cytokine-mediated upregulation of Fas and possibly Fas ligand. Cytokines produced by the infiltrating immune cells also induce expression of thyroid cell-surface molecules that lead to: (1) engagement by immune cells (e.g., adhesion molecules, HLA class I and II molecules); (2) induction of cytokine secretion by the thyroid cells themselves; (3) production of nitric oxide; and (4) reduction of thyroid hormone production through inhibition of TSH-R, TPO, and Tg synthesis. Administration of high concentrations of cytokines for therapeutic purposes [especially interferon (IFN)a] is associated with increased autoimmune thyroid disease, presumably via mechanisms similar to those that occur in sporadic autoimmune disease.

Cytokines appear to play a major role in thyroid-associated ophthalmopathy. There is infiltration of the extraocular muscles by activated T cells; the release of cytokines results in fibroblast activation and increased synthesis of glycosaminoglycans that trap water, thereby leading to characteristic muscle swelling. Late in the disease, there is fibrosis and only then do the muscle cells show evidence of injury. Orbital fibroblasts may be uniquely sensitive to cytokines, perhaps explaining the anatomic localization of the immune response. Though the pathogenesis of thyroid-associated ophthalmopathy remains unclear, there is mounting evidence that expression of the TSH-R may provide an important orbital autoantigen. In support of this idea, injection of TSH-R into certain strains of mice induces autoimmune hyperthyroidism, as well as features of ophthalmopathy. A variety of autoantibodies against orbital muscle and fibroblast

antigens have been detected in patients with ophthalmopathy, but these antibodies most likely arise as a secondary phenomenon, dependent on T cell-mediated autoimmune responses.

HYPOTHYROIDISM

Iodine deficiency remains the most common cause of hypothyroidism worldwide. In areas of iodine sufficiency, autoimmune disease (Hashimoto's thyroiditis) and iatrogenic causes (treatment of hyperthyroidism) are most common (Table 330-4).

CONGENITAL HYPOTHYROIDISM

Prevalence Hypothyroidism occurs in about 1 in 3000 to 4000 newborns. It may be transient, especially if the mother has TSH-R blocking antibodies or has received antithyroid drugs, but permanent hypothyroidism occurs in the majority. Neonatal hypothyroidism is due to thyroid gland dysgenesis in 85%, inborn errors of thyroid hormone synthesis in 10 to 15%, and is TSH-R antibody-mediated in 5% of affected newborns. The developmental abnormalities are twice as common in girls. Mutations that cause congenital hypothyroidism are being increasingly recognized, but the vast majority remain idiopathic (Table 330-1).

Clinical Manifestations The majority of infants appear normal at birth, and <10% are diagnosed based on clinical features, which include prolonged jaundice, feeding problems, hypotonia, enlarged tongue, delayed bone maturation, and umbilical hernia. Importantly, permanent neurologic damage results if treatment is delayed. Typical features of adult hypothyroidism may also be present (Table 330-5).

Diagnosis and Treatment Because of the severe neurologic consequences of untreated congenital hypothyroidism, neonatal screening programs have been established in developed countries (Chap. 68). These are generally based on measurement of TSH or T_4 levels in heel-prick blood specimens. When the diagnosis is confirmed, T_4 is instituted at a dose of 10 to 15 ug/kg per day and the dosage is adjusted by close monitoring of TSH levels. T_4 requirements are relatively great during the first year of life, and a high circulating T_4 level is usually needed to normalize TSH. Early treatment with T_4 results in normal IQ levels, but subtle neurodevelopmental abnormalities may be detected in those with the most severe hypothyroidism at diagnosis or when treatment is suboptimal.

AUTOIMMUNE HYPOTHYROIDISM

Classification Autoimmune hypothyroidism may be associated with a goiter (Hashimoto's, or *goitrous thyroiditis*) or, at the later stages of the disease, minimal residual thyroid tissue (*atrophic thyroiditis*). Because the autoimmune process gradually reduces thyroid function, there is a phase of compensation during which normal thyroid hormone levels are maintained by a rise in TSH. Though some patients may have minor symptoms, this state is called *subclinical hypothyroidism*. Later, free T_4 levels fall and TSH levels rise further; symptoms become more readily apparent at this stage (usually TSH > 10 mU/L), which is referred to as *clinical hypothyroidism* (*overt hypothyroidism*).

Prevalence The mean annual incidence rate of autoimmune hypothyroidism is up to 4 per 1000 women and 1 per 1000 men. It is more common in certain populations, such as the Japanese, probably as a consequence of genetic factors and chronic exposure to a high-iodine diet. The mean age at diagnosis is about 60 years, and the prevalence of overt hypothyroidism increases with age. Subclinical hypothyroidism is found in 6 to 8% of women (10% over the age of 60) and 3% of men. The annual risk of developing clinical hypothyroidism is about 4% when subclinical hypothyroidism is associated with positiveTPOantibodies.

Pathogenesis In Hashimoto's thyroiditis, there is a marked lymphocytic infiltration of the thyroid with germinal center formation, atrophy of the thyroid follicles accompanied by oxyphil metaplasia, absence of colloid, and mild to moderate fibrosis. In atrophic thyroiditis, the fibrosis is much more extensive, lymphocyte infiltration is less pronounced, and thyroid follicles are almost completely absent. Atrophic thyroiditis likely represents the end stage of Hashimoto's thyroiditis rather than a distinct disorder. Autoimmune features are similar in both types of hypothyroidism, thoughTSH-Rblocking antibodies may be more frequent in Asian patients with atrophic thyroiditis. The mechanisms that result in thyroid follicular destruction are predominantly T cell mediated, but antibodies may also contribute to thyroid dysfunction by complement fixation or inhibition of thyroid cell function (see "Autoimmune Basis of Thyroid Disease," above).

Clinical Manifestations The main clinical features of hypothyroidism are summarized inTable 330-5. The onset is usually insidious, and the patient may become aware of symptoms only when euthyroidism is restored. Patients with Hashimoto's thyroiditis may present because of goiter rather than symptoms of hypothyroidism. The goiter may not be large but is usually irregular and firm in consistency. It is often possible to palpate a pyramidal lobe, normally a vestigial remnant of thyroglossal duct. Rarely, uncomplicated Hashimoto's thyroiditis is associated with pain.

Patients with atrophic thyroiditis, or the late stage of Hashimoto's thyroiditis, present with symptoms and signs of hypothyroidism. The skin is dry, and there is decreased sweating, thinning of the epidermis, and hyperkeratosis of the stratum corneum. Increased dermal glycosaminoglycan content traps water, giving rise to skin thickening without pitting (*myxedema*). Typical features include a puffy face with edematous eyelids and nonpitting pretibial edema (Figs. 330-4,330-CD1and330-CD2). There is pallor, often with a yellow tinge due to carotene accumulation. Nail growth is retarded, and hair is dry, brittle, difficult to manage, and falls out easily. In addition to diffuse alopecia, there is thinning of the outer third of the eyebrows.

Other common features include constipation and weight gain (despite a poor appetite). In contrast to popular perception, the weight gain is usually modest and due mainly to fluid retention in the myxedematous tissues. Libido is decreased in both sexes, and there may be oligomenorrhea or amenorrhea in long-standing disease, but menorrhagia is also common. Fertility is reduced and the incidence of miscarriage is increased. Prolactin levels are often modestly increased (Chap. 328) and may contribute to alterations in libido and fertility as well as causing galactorrhea.

Myocardial contractility and pulse rate are reduced, leading to a reduced stroke volume

and bradycardia. Increased peripheral resistance may be accompanied by hypertension, particularly diastolic. Blood flow is diverted from the skin, producing the cool extremities. Pericardial effusions occur in up to 30% of patients but rarely compromise cardiac function. Though alterations in myosin heavy chain isoform expression have been documented, cardiomyopathy is unusual. Fluid may also accumulate in other serous cavities and in the middle ear, giving rise to conductive deafness. Pulmonary function is generally normal, but dyspnea may be due to pleural effusion, impaired respiratory muscle function, diminished ventilatory drive, or sleep apnea.

Carpal tunnel and other entrapment syndromes are common, as is impairment of muscle function with stiffness, cramps, and pain. On examination, there may be slow relaxation of tendon reflexes (Video 330-1) and pseudomyotonia. Memory and concentration are impaired. Rare neurologic problems include reversible cerebellar ataxia, dementia, psychosis, and myxedema coma. *Hashimoto's encephalopathy* is a rare and distinctive syndrome associated with myoclonus and slow-wave activity on electroencephalography, which can progress to confusion, coma, and death. It is steroid-responsive and may occur in the presence of autoimmune thyroiditis, without hypothyroidism. The hoarse voice and occasionally clumsy speech of hypothyroidism are due to fluid accumulation in the vocal cords and tongue.

The features described above are due to a shortage of thyroid hormone. However, autoimmune hypothyroidism may be associated with signs or symptoms of other autoimmune diseases, particularly vitiligo, pernicious anemia, Addison's disease, alopecia areata, and type 1 diabetes mellitus. Less common associations include celiac disease, dermatitis herpetiformis, chronic active hepatitis, rheumatoid arthritis, SLE, and Sjogren's syndrome. Thyroid-associated ophthalmopathy, which usually occurs in Graves' disease (see below), occurs in about 5% of patients with autoimmune hypothyroidism.

Autoimmune hypothyroidism is uncommon in children and usually presents with slow growth and delayed facial maturation. The appearance of permanent teeth is also delayed. Myopathy, with muscle swelling, is more common than in adults. In most cases, puberty is delayed, but precocious puberty sometimes occurs. There may be intellectual impairment if the onset is before 3 years and the hormone deficiency is severe.

Laboratory Evaluation A summary of the investigations used to determine the existence and cause of hypothyroidism is provided in Fig. 330-5. A normal TSH level excludes primary (but not secondary) hypothyroidism. If the TSH is elevated, a free T_4 level is needed to confirm the presence of clinical hypothyroidism, but free T_4 is inferior to TSH when used as a screening test, as it will not detect subclinical or mild hypothyroidism. Circulating free T_3 levels are normal in about 25% of patients, reflecting adaptive responses to hypothyroidism. T_3 measurements are therefore not indicated.

Once clinical or subclinical hypothyroidism is confirmed, the etiology is usually easily established by demonstrating the presence of TPO antibodies, which are present in 90 to 95% of patients with autoimmune hypothyroidism. TBII can be found in 10 to 20% of patients, but these determinations are not needed routinely. If there is any doubt about

the cause of a goiter associated with hypothyroidism,FNAbiopsy can be used to confirm the presence of autoimmune thyroiditis. Other abnormal laboratory findings in hypothyroidism may include increased creatine phosphokinase, elevated cholesterol and triglycerides, and anemia (usually normocytic or macrocytic). Except when accompanied by iron deficiency, the anemia and other abnormalities gradually resolve with thyroxine replacement.

Differential Diagnosis An asymmetric goiter in Hashimoto's thyroiditis may be confused with a multinodular goiter or thyroid carcinoma, even when thyroid antibodies are present. Ultrasound can be used to show the presence of a solitary lesion or a multinodular goiter, rather than the heterogeneous thyroid enlargement typical of Hashimoto's thyroiditis.FNAbiopsy is useful in the investigation of focal nodules. Other causes of hypothyroidism are discussed below but rarely cause diagnostic confusion (Table 330-4).

OTHER CAUSES OF HYPOTHYROIDISM

Iatrogenic hypothyroidism is a common cause of hypothyroidism and can often be detected by screening before symptoms develop. In the first 3 to 4 months after radioiodine treatment, transient hypothyroidism may occur due to reversible radiation damage rather than to cellular destruction. Low-dose thyroxine treatment can be withdrawn if recovery occurs. BecauseTSHlevels are suppressed by hyperthyroidism, free T4levels are a better measure of thyroid function than TSH in the months following radioiodine treatment. Mild hypothyroidism after subtotal thyroidectomy may also resolve after several months, as the gland remnant is stimulated by increased TSH levels.

Iodine deficiency is responsible for endemic goiter and cretinism but is an uncommon cause of adult hypothyroidism unless the iodine intake is very low or there are complicating factors, such as the consumption of thiocyanates in cassava or selenium deficiency. Though hypothyroidism due to iodine deficiency can be treated with thyroxine, public health measures to improve iodine intake should be advocated to eliminate this problem. Iodized salt or bread or the use of a single bolus of oral or intramuscular iodized oil have all been used successfully.

Paradoxically, chronic iodine excess can also induce goiter and hypothyroidism. The intracellular events that account for this effect are unclear, but individuals with autoimmune thyroiditis are especially susceptible. Iodine excess is responsible for the hypothyroidism that occurs in up to 13% of patients treated with amiodarone (see below). Other drugs, particularly lithium, may also cause hypothyroidism.

Secondary hypothyroidism is usually diagnosed in the context of other anterior pituitary hormone deficiencies; isolatedTSHdeficiency is very rare (Chap. 328). TSH levels may be low, normal, or even slightly increased in secondary hypothyroidism; the latter is due to secretion of immunoactive but bioinactive forms of TSH. The diagnosis is confirmed by detecting a low free T4level. The goal of treatment is to maintain free T4levels in the upper half of the reference range, as TSH levels cannot be used to monitor therapy.

TREATMENT

Clinical Hypothyroidism If there is no residual thyroid function, the daily replacement dose of levothyroxine is usually 1.5 ug/kg body weight (typically 100 to 150 ug). In many patients, however, lower doses suffice until residual thyroid tissue is destroyed. In patients who develop hypothyroidism after the treatment of Graves' disease, there is often underlying autonomous function, necessitating lower replacement doses (typically 75 to 125 ug/d).

Adult patients under 60 without evidence of heart disease may be started on 50 to 100ug levothyroxine (T_4) daily. The dose is adjusted on the basis of TSH levels, with the goal of treatment being a normal TSH, ideally in the lower half of the reference range. TSH responses are gradual and should be measured about 2 months after instituting treatment or after any subsequent change in levothyroxine dosage. The clinical effects of levothyroxine replacement are often slow to appear. Patients may not experience full relief from symptoms until 3 to 6 months after normal TSH levels are restored. Adjustment of levothyroxine dosage is made in 12.5- or 25-ug increments if the TSH is high; decrements of the same magnitude should be made if the TSH is suppressed. Patients with a suppressed TSH of any cause, including T_4 overtreatment, have an increased risk of atrial fibrillation and reduced bone density.

Although dessicated animal thyroid preparations (thyroid extract USP) are available, they are not recommended as potency and composition vary between batches. Interest in using levothyroxine combined with liothyronine (triiodothyronine, T_3) has been revived, based on studies suggesting that patients feel better when taking the T_4/T_3 combination compared to T_4 alone. However, a long-term benefit from this combination is not established. There is no place for liothyronine alone as long-term replacement, because the short half-life necessitates three or four daily doses and is associated with fluctuating T_3 levels.

Once full replacement is achieved and TSH levels are stable, follow-up measurement of TSH is recommended at annual intervals and may be extended to every 2 to 3 years, if a normal TSH is maintained over several years. It is important to ensure ongoing compliance, however, as patients do not feel any difference after missing a few doses of levothyroxine, sometimes leading to self-discontinuation.

In patients of normal body weight who are taking ³200 ug of levothyroxine per day, an elevated TSH level is often a sign of poor compliance. This is also the likely explanation for fluctuating TSH levels, despite a constant levothyroxine dosage. Such patients often have normal or high free T_4 levels, despite an elevated TSH, because they remember to take medication for a few days before testing; this is sufficient to normalize T_4 but not TSH levels. It is important to consider variable compliance, as this pattern of thyroid function tests is otherwise suggestive of disorders associated with inappropriate TSH secretion (Table 330-3). Because T_4 has a long half-life (7 days), patients who miss doses can be advised to take up to three doses of the skipped tablets at once. Other causes of increased levothyroxine requirements must be excluded, particularly malabsorption (e.g., celiac disease, small-bowel surgery) and drugs that interfere with T_4 absorption or clearance such as cholestyramine, ferrous sulfate, calcium supplements, lovastatin, aluminum hydroxide, rifampicin, amiodarone, carbamazepine, and phenytoin.

Subclinical Hypothyroidism By definition, subclinical hypothyroidism refers to biochemical evidence of thyroid hormone deficiency in patients who have few or no apparent clinical features of hypothyroidism. There are no generally accepted guidelines for the treatment of subclinical hypothyroidism. As long as excessive treatment is avoided, there is little risk in correcting a slightly increased TSH, and some patients likely derive modest clinical benefit from treatment. Moreover, there is some risk that patients will progress to overt hypothyroidism, particularly when TPO antibodies are present. Treatment is administered by starting with a low dose of levothyroxine (25 to 50 ug/d) with the goal of normalizing TSH.

Special Treatment Considerations Rarely, levothyroxine replacement is associated with pseudotumor cerebri in *children*. Presentation appears to be idiosyncratic and occurs months after treatment is begun. Women with a history or high risk of hypothyroidism should ensure that they are euthyroid prior to conception and during early pregnancy as maternal hypothyroidism may adversely affect fetal neural development. TSH and free T$_4$ levels should be measured once pregnancy is confirmed and at the beginning of the second and third trimesters. The dose of levothyroxine may need to be increased by ³50% during pregnancy and returned to previous levels after delivery. In the *elderly*, especially patients with known coronary artery disease, the starting dose of levothyroxine is 12.5 to 25 ug/d with similar increments every 2 to 3 months until TSH is normalized. In some patients it may be impossible to achieve full replacement, despite optimal antianginal treatment. *Emergency surgery* is generally safe in patients with untreated hypothyroidism, although routine surgery in a hypothyroid patient should be deferred until euthyroidism is achieved.

Myxedema coma still has a high mortality rate, despite intensive treatment. Clinical manifestations include reduced level of consciousness, sometimes associated with seizures, as well as the other features of hypothyroidism (Table 330-5). Hypothermia can reach 23°C (74°F). There may be a history of treated hypothyroidism with poor compliance, or the patient may be previously undiagnosed. Myxedema coma almost always occurs in the elderly and is usually precipitated by factors that impair respiration, such as drugs (especially sedatives, anesthetics, antidepressants), pneumonia, congestive heart failure, myocardial infarction, gastrointestinal bleeding, or cerebrovascular accidents. Sepsis should also be suspected. Exposure to cold may also be a risk factor. Hypoventilation, leading to hypoxia and hypercapnia, plays a major role in pathogenesis; hypoglycemia and dilutional hyponatremia also contribute to the development of myxedema coma.

Levothyroxine can initially be administered as a single intravenous bolus of 500 ug, which serves as a loading dose. Although further levothyroxine is not strictly necessary for several days, it is usually continued at a dose of 50 to 100 ug/d. If a suitable intravenous preparation is not available, the same initial dose of levothyroxine can be given by nasogastric tube (though absorption may be impaired in myxedema). An alternative is to give liothyronine (T$_3$) intravenously or via nasogastric tube, in doses ranging from 10 to 25 ug every 8 to 12 h. This treatment has been advocated because T$_4$®T$_3$ conversion is impaired in myxedema coma. However, excess liothyroxine has the potential to provoke arrhythmias. Another commonly used option is to combine levothyroxine (200 ug) and liothyronine (25 ug) as a single, initial intravenous bolus

followed by daily treatment with levothyroxine (50 to 100 ug/d) and liothyronine (10 ug every 8 h).

Supportive therapy should be provided to correct any associated metabolic disturbances. External warming is indicated only if the temperature is<30°C, as it can result in cardiovascular collapse (Chap. 17). Space blankets should be used to prevent further heat loss. Parenteral hydrocortisone (50 mg every 6 h) should be administered, as there is impaired adrenal reserve in profound hypothyroidism. Any precipitating factors should be treated, including the early use of broad-spectrum antibiotics, pending the exclusion of infection. Ventilatory support with regular blood gas analysis is usually needed during the first 48 h. Hypertonic saline or intravenous glucose may be needed if there is hyponatremia or hypoglycemia; hypotonic intravenous fluids should be avoided because they may exacerbate water retention secondary to reduced renal perfusion and inappropriate vasopressin secretion. The metabolism of most medications is impaired, and sedatives should be avoided if possible or used in reduced doses. Blood levels should be monitored, when available, to guide medication dosage.

THYROTOXICOSIS

Thyrotoxicosis is defined as the state of thyroid hormone excess and is not synonymous with *hyperthyroidism*, which is the result of excessive thyroid function. However, the major etiologies of thyrotoxicosis are hyperthyroidism caused by Graves' disease, toxic multinodular goiter, and toxic adenomas. Other causes are listed in Table 330-6.

GRAVES' DISEASE

Epidemiology Graves' disease accounts for 60 to 80% of thyrotoxicosis, though the prevalence varies among populations, depending mainly on iodine intake (high iodine intake is associated with an increased prevalence of Graves' disease). Graves' disease occurs in up to 2% of women but is one-tenth as frequent in men. The disorder rarely begins before adolescence and typically occurs between 20 and 50 years of age, though it also occurs in the elderly.

Pathogenesis The hyperthyroidism of Graves' disease is caused byTSI that are directed to theTSH-R(see "Autoimmune Basis of Thyroid Disease," below). Other thyroid autoimmune responses coexist in these patients, and therefore there is no direct correlation between the levels of TSI and thyroid hormones. The extrathyroidal manifestations of Graves' disease -- i.e., ophthalmopathy and dermopathy -- are due to immunologically mediated activation of fibroblasts in the extraocular muscles and skin, with accumulation of glycosaminoglycans, leading to the trapping of water and edema. Later, fibrosis becomes prominent. The fibroblast activation is caused by cytokines (IFN-g, tumor necrosis factor,IL-1) derived from locally infiltrating T cells and macrophages.

Clinical Manifestations Signs and symptoms include features that are common to any cause of thyrotoxicosis (Table 330-7) as well as those specific for Graves' disease. The clinical presentation depends on the severity of thyrotoxicosis, the duration of the disease, individual susceptibility to excess thyroid hormone, and the age of the patient. In the elderly, features of thyrotoxicosis may be subtle or masked, and patients may

present mainly with fatigue and weight loss, leading to *apathetic hyperthyroidism*.

Thyrotoxicosis may cause unexplained weight loss, despite an enhanced appetite, and is due to the increased metabolic rate. Weight gain occurs in 5 to 10% of patients, however, as a result of increased food intake. Other prominent features include hyperactivity, nervousness, and irritability, ultimately leading to a sense of easy fatiguability in some patients. Insomnia and impaired concentration are common; apathetic thyrotoxicosis may be mistaken for depression in the elderly. Fine tremor is a very frequent finding, best elicited by asking patients to stretch out the fingers and feeling the fingertips with the palm. Common neurologic manifestations include hyperreflexia, muscle wasting, and proximal myopathy without fasciculation. Chorea is a rare feature. Thyrotoxicosis is sometimes associated with a form of hypokalemic periodic paralysis; this disorder is particularly common in Asian males with thyrotoxicosis.

The most common cardiovascular manifestation is sinus tachycardia, often associated with palpitations and sometimes due to supraventricular tachycardia. The high cardiac output produces a bounding pulse, widened pulse pressure, and an aortic systolic murmur, and can lead to worsening of angina or heart failure in the elderly or those with preexisting heart disease. Atrial fibrillation is more common in patients>50. Treatment of the thyrotoxic state alone reverts atrial fibrillation to normal sinus rhythm in fewer than half of patients, suggesting the existence of an underlying cardiac problem in the remainder.

The skin is usually warm and moist, and the patient complains of sweating and heat intolerance, particularly during warm weather. Palmar erythema; onycholysis; and, less commonly, pruritus, urticaria, and diffuse hyperpigmentation may be evident. Hair texture may become fine, and a diffuse alopecia occurs in up to 40% of patients, persisting for months after restoration of euthyroidism. Gastrointestinal transit time is decreased, leading to increased stool frequency, often with diarrhea and occasionally mild steatorrhea. Women frequently experience oligomenorrhea or amenorrhea; in men there may be impaired sexual function and, rarely, gynecomastia. The direct effect of thyroid hormones on bone resorption leads to osteopenia in long-standing thyrotoxicosis; mild hypercalcemia occurs in up to 20% of patients, but hypercalcuria is more common. There is a small increase in fracture rate in patients with a previous history of thyrotoxicosis.

In Graves' disease the thyroid is usually diffusely enlarged to two to three times its normal size. The consistency is firm, but less so than in multinodular goiter. There may be a thrill or bruit due to the increased vascularity of the gland and the hyperdynamic circulation.

Lid retraction, causing a staring appearance, can occur in any form of thyrotoxicosis and is the result of sympathetic overactivity. However, Graves' disease is associated with specific eye signs that comprise *Graves' ophthalmopathy* (Fig. 330-6*A* and 330-CD2). This condition is also called *thyroid-associated ophthalmopathy*, as it occurs in the absence of Graves' disease in 10% of patients. Most of these individuals have autoimmune hypothyroidism or thyroid antibodies. The onset of Graves' ophthalmopathy occurs within the year before or after the diagnosis of thyrotoxicosis in 75% of patients

but can sometimes precede or follow thyrotoxicosis by several years, accounting for some cases of euthyroid ophthalmopathy.

Many patients with Graves' disease have little clinical evidence of ophthalmopathy. However, the enlarged extraocular muscles typical of the disease, and other subtle features, can be detected in almost all patients when investigated by ultrasound or computed tomography (CT) imaging of the orbits. Unilateral signs are found in up to 10% of patients. The earliest manifestations of ophthalmopathy are usually a sensation of grittiness, eye discomfort, and excess tearing. About a third of patients have proptosis, best detected by visualization of the sclera between the lower border of the iris and the lower eyelid, with the eyes in the primary position. Proptosis can be measured using an exophthalmometer. In severe cases, proptosis may cause corneal exposure and damage, especially if the lids fail to close during sleep. Periorbital edema, scleral injection, and chemosis are also frequent. In 5 to 10% of patients, the muscle swelling is so severe that diplopia results, typically but not exclusively when the patient looks up and laterally. The most serious manifestation is compression of the optic nerve at the apex of the orbit, leading to papilledema, peripheral field defects, and, if left untreated, permanent loss of vision.

Many scoring systems have been used to gauge the extent and activity of the orbital changes in Graves' disease. The NO SPECS scheme is an acronym derived from the following classes of eye change:

0= No signs or symptoms

1 =Only signs (lid retraction or lag), no symptoms

2 =Soft tissue involvement (periorbital edema)

3 =Proptosis (>22 mm)

4 =Extraocular muscle involvement (diplopia)

5 =Corneal involvement

6= Sight loss

Although useful as a mnemonic, the NO SPECS scheme is inadequate to describe the eye disease fully, and patients do not necessarily progress from one class to another. When Graves' eye disease is active and severe, referral to an ophthalmologist is indicated and objective measurements are needed, such as lid fissure width; corneal staining with fluorescein; and evaluation of extraocular muscle function (e.g., Hess chart), intraocular pressure and visual fields, acuity, and color vision.

Thyroid dermopathy occurs in<5% of patients with Graves' disease (Fig. 330-6*B*), almost always in the presence of moderate or severe ophthalmopathy. Although most frequent over the anterior and lateral aspects of the lower leg (hence the term *pretibial myxedema*), skin changes can occur at other sites, particularly after trauma. The typical lesion is a noninflamed, indurated plaque with a deep pink or purple color and an

"orange-skin" appearance. Nodular involvement can occur, and the condition can rarely extend over the whole lower leg and foot, mimicking elephantiasis. *Thyroid acropachy* refers to a form of clubbing found in<1% of patients with Graves' disease ([Fig. 330-6*C*](#)). It is so strongly associated with thyroid dermopathy that an alternative cause of clubbing should be sought in a Graves' patient without coincident skin and orbital involvement.

Laboratory Evaluation Investigations used to determine the existence and cause of thyrotoxicosis are summarized in [Fig. 330-7](#). In Graves' disease, the[TSH](#)level is suppressed and free and total thyroid hormone levels are increased. In 2 to 5% of patients (and more in areas of borderline iodine intake), only T_3 is increased (T_3toxicosis). The converse state of T_4toxicosis, with elevated total and free T_4 and normal T_3levels, is occasionally seen when hyperthyroidism is induced by excess iodine, providing surplus substrate for thyroid hormone synthesis. Measurement of[TPO](#)antibodies is useful in differential diagnosis, but assays for[TSH-R](#)antibodies are not usually needed. Associated abnormalities that may cause diagnostic confusion in thyrotoxicosis include elevation of bilirubin, liver enzymes, and ferritin. Microcytic anemia and thrombocytopenia occur less often.

Differential Diagnosis Diagnosis of Graves' disease is straightforward in a patient with biochemically confirmed thyrotoxicosis, diffuse goiter on palpation, ophthalmopathy, positive[TPO](#)antibodies, and often a personal or family history of autoimmune disorders. For patients with thyrotoxicosis who lack these features, the most reliable diagnostic method is a radionuclide ($_{99m}$Tc, $_{123}$I, or$_{131}$I) scan of the thyroid, which will distinguish the diffuse, high uptake of Graves' disease from nodular thyroid disease, destructive thyroiditis, ectopic thyroid tissue, and factitious thyrotoxicosis. In secondary hyperthyroidism due to a[TSH](#)-secreting pituitary tumor, there is also a diffuse goiter. The presence of a nonsuppressed TSH level, and the finding of a pituitary tumor on[CT](#) or magnetic resonance imaging (MRI) scan readily identify such patients.

Clinical features of thyrotoxicosis can mimic certain aspects of other disorders including panic attacks, mania, pheochromocytoma, and the weight loss associated with malignancy. The diagnosis of thyrotoxicosis can be easily excluded if the[TSH](#)level is normal. A normal TSH also excludes Graves' disease as a cause of diffuse goiter.

Clinical Course Clinical features generally worsen without treatment; mortality was 10 to 30% before the introduction of satisfactory therapy. Some patients with mild Graves' disease experience spontaneous relapses and remissions. Rarely, there may be fluctuation between hypo- and hyperthyroidism due to changes in the functional activity of[TSH-R](#)antibodies. About 15% of patients who enter remission after treatment with antithyroid drugs develop hypothyroidism 10 to 15 years later as a result of the destructive autoimmune process. The clinical course of ophthalmopathy does not follow that of the thyroid disease. Ophthalmopathy typically worsens over the initial 3 to 6 months, followed by a plateau phase over the next 12 to 18 months, with spontaneous improvement, particularly in the soft tissue changes. However, the course is more fulminant in up to 5% of patients, requiring intervention in the acute phase if there is optic nerve compression or corneal ulceration. Diplopia may appear late in the disease due to fibrosis of the extraocular muscles. Some studies suggest that radioiodine treatment for hyperthyroidism worsens the eye disease in a small proportion of patients (especially smokers). Antithyroid drugs or surgery have no adverse effects on the

clinical course of ophthalmopathy. Thyroid dermopathy, when it occurs, usually appears 1 to 2 years after the development of Graves' hyperthyroidism; it may improve spontaneously.

TREATMENT

The *hyperthyroidism* of Graves' disease is treated by reducing thyroid hormone synthesis, using antithyroid drugs, or by reducing the amount of thyroid tissue with radioiodine ($_{131}$I) treatment or subtotal thyroidectomy. Antithyroid drugs are the predominant therapy in many centers in Europe and Japan, whereas radioiodine is more often the first line of treatment in North America. These differences reflect the fact that no single approach is optimal and that patients may require multiple treatments to achieve remission.

The main *antithyroid drugs* are the thionamides, such as propylthiouracil, carbimazole, and the active metabolite of the latter, methimazole. All inhibit the function of TPO, reducing oxidation and organification of iodide. These drugs also reduce thyroid antibody levels by mechanisms that remain unclear, and they appear to enhance rates of remission. Propylthiouracil inhibits deiodination of T_4®T_3. However, this effect is of minor benefit, except in the most severe thyrotoxicosis, and is offset by the much shorter half-life of this drug (90 min) compared to methimazole (6 h).

There are many variations of antithyroid drug regimens. The initial dose of carbimazole or methimazole is usually 10 to 20 mg every 8 or 12 h, but once-daily dosing is possible after euthyroidism is restored. Propylthiouracil is given at a dose of 100 to 200 mg every 6 to 8 h, and divided doses are usually given throughout the course. Lower doses of each drug may suffice in areas of low iodine intake. The starting dose of antithyroid drugs can be gradually reduced (titration regimen) as thyrotoxicosis improves. Alternatively, high doses may be given combined with levothyroxine supplementation (block-replace regimen) to avoid drug-induced hypothyroidism. Initial reports suggesting superior remission rates with the block-replace regimen have not been reproduced in several other trials. The titration regimen is often preferred to minimize the dose of antithyroid drug and provide an index of treatment response.

Thyroid function tests and clinical manifestations are reviewed 3 to 4 weeks after starting treatment, and the dose is titrated based on free T_4 levels. Most patients do not achieve euthyroidism until 6 to 8 weeks after treatment is initiated. TSH levels often remain suppressed for several months and therefore do not provide a sensitive index of treatment response. The usual daily maintenance doses of antithyroid drugs in the titration regimen are 2.5 to 10 mg of carbimazole or methimazole and 50 to 100 mg of propylthiouracil. In the block-replace regimen, the initial dose of antithyroid drug is held constant and the dose of levothyroxine is adjusted to maintain normal free T_4 levels.

Maximum remission rates (up to 30 to 50% in some populations) are achieved by 18 to 24 months. For unclear reasons, remission rates appear to vary in different geographic regions. Patients with severe hyperthyroidism and large goiters are most likely to relapse when treatment stops, but outcome is difficult to predict. All patients should be followed closely for relapse during the first year after treatment and at least annually thereafter.

The common side effects of antithyroid drugs are rash, urticaria, fever, and arthralgia (1 to 5% of patients). These may resolve spontaneously or after substituting an alternative antithyroid drug. Rare but major side effects include hepatitis, an SLE-like syndrome, and, most importantly, agranulocytosis (<1%). It is essential that antithyroid drugs are stopped and not restarted if a patient develops major side effects. Patients should be given written instructions regarding the symptoms of possible agranulocytosis (e.g., sore throat, fever, mouth ulcers) and the need to stop treatment pending a complete blood count to confirm that agranulocytosis is not present. Management of agranulocytosis is described in Chap. 109. Most physicians do not prospectively monitor blood counts, as the onset of agranulocytosis is idiosyncratic and abrupt.

Propranolol (20 to 40 mg every 6 h) or longer acting beta blockers, such as atenolol, may be useful to control adrenergic symptoms, especially in the early stages before antithyroid drugs take effect. Anticoagulation with warfarin should be considered in all patients with atrial fibrillation. If digoxin is used, increased doses are often needed in the thyrotoxic state.

Radioiodine causes progressive destruction of thyroid cells and can be used as initial treatment or for relapses after a trial of antithyroid drugs. There is a small risk of thyrotoxic crisis (see below) after radioiodine, which can be avoided by pretreatment with antithyroid drugs for at least a month before treatment. Antecedent treatment with antithyroid drugs should be considered in all elderly patients, or in those with cardiac problems, to deplete thyroid hormone stores before administration of radioiodine. Antithyroid drugs must be stopped 3 to 5 days before radioiodine administration to achieve optimum iodine uptake.

Efforts to calculate an optimal dose of radioiodine that achieves euthyroidism, without a high incidence of relapse or progression to hypothyroidism, have not been successful. Some patients inevitably relapse after a single dose because the biologic effects of radiation vary between individuals, and hypothyroidism cannot be uniformly avoided even using accurate dosimetry. A practical strategy is to give a fixed dose based on clinical features, such as the severity of thyrotoxicosis, the size of the goiter (increases the dose needed), and the level of radioiodine uptake (decreases the dose needed). ^{131}I dosage generally ranges between 185 MBq (5 mCi) to 555 MBq (15 mCi). Incomplete treatment or early relapse is more common in males and in patients <40 years of age. Many authorities favor an approach aimed at thyroid ablation (as opposed to euthyroidism), given that levothyroxine replacement is straightforward and most patients ultimately progress to hypothyroidism over 5 to 10 years anyway, frequently with some delay in the diagnosis of hypothyroidism.

Certain radiation safety precautions are necessary in the first few days after radioiodine treatment, but the exact guidelines vary depending on local protocols. In general, patients need to avoid close, prolonged contact with children and pregnant women for several days because of possible transmission of residual isotope and excessive exposure to radiation emanating from the gland. Rarely there may be mild pain due to radiation thyroiditis 1 to 2 weeks after treatment. Hyperthyroidism can persist for 2 to 3 months before radioiodine takes full effect. For this reason, b-adrenergic blockers or antithyroid drugs can be used to control symptoms during this interval. Persistent

hyperthyroidism can be treated with a second dose of radioiodine, usually 6 months after the first dose. The risk of hypothyroidism after radioiodine depends on the dosage but is at least 10 to 20% in the first year and 5% per year thereafter. Patients should be informed of this possibility before treatment and require close follow-up during the first year and annual thyroid function testing thereafter.

Pregnancy and breast feeding are absolute contraindications to radioiodine treatment, but patients can conceive safely 6 to 12 months after treatment. The presence of severe ophthalmopathy requires caution, and some authorities advocate the use of prednisone, 40 mg/d, at the time of radioiodine treatment, tapered over 2 to 3 months to prevent exacerbation of ophthalmopathy. The overall risk of cancer after radioiodine treatment in adults is not increased, but many physicians avoid radioiodine in children and adolescents because of the theoretical risks of malignancy.

Subtotal thyroidectomy is an option for patients who relapse after antithyroid drugs and prefer this treatment to radioiodine. Some experts recommend surgery in young individuals, particularly when the goiter is very large. Careful control of thyrotoxicosis with antithyroid drugs, followed by potassium iodide (3 drops SSKI orally tid) is needed prior to surgery to avoid thyrotoxic crisis and to reduce the vascularity of the gland. The major complications of surgery -- i.e., bleeding, laryngeal edema, hypoparathyroidism, and damage to the recurrent laryngeal nerves -- are unusual when the procedure is performed by highly experienced surgeons. Recurrence rates in the best series are <2%, but the rate of hypothyroidism is only slightly less than that following radioiodine treatment.

The titration regimen of antithyroid drugs should be used to manage Graves' disease in *pregnancy*, as blocking doses of these drugs produce fetal hypothyroidism. Propylthiouracil is usually used because of relatively low transplacental transfer and its ability to block $T_4 \circledR T_3$ conversion. Also, carbimazole and methimazole have been associated with rare cases of fetal *aplasia cutis*. The lowest effective dose of propylthiouracil should be given, and it is often possible to stop treatment in the last trimester since TSH-R antibodies tend to decline in pregnancy. Nonetheless, the transplacental transfer of these antibodies rarely causes *fetal thyrotoxicosis* or *neonatal thyrotoxicosis*. Poor intrauterine growth, a fetal heart rate of >160 beats/min, and high levels of maternal TSH-R antibodies should suggest this complication. Antithyroid drugs given to the mother can be used to treat the fetus and may be needed for 1 to 3 months after delivery, until the maternal antibodies disappear from the baby's circulation. The post-partum period is a time of major risk for relapse of Graves' disease. Breast feeding is safe with low doses of antithyroid drugs. Graves' disease in *children* is best managed with antithyroid drugs, often given as a prolonged course of the titration regimen. Surgery may be indicated for severe disease. Radioiodine can also be used in children, though most experts defer this treatment until adolescence or later.

Thyrotoxic crisis, or *thyroid storm*, is rare and presents as a life-threatening exacerbation of hyperthyroidism, accompanied by fever, delirium, seizures, coma, vomiting, diarrhea, and jaundice. The mortality rate due to cardiac failure, arrhythmia, or hyperthermia is ~30%, even with treatment. Thyrotoxic crisis is usually precipitated by acute illness (e.g., stroke, infection, trauma, diabetic ketoacidosis), surgery (especially on the thyroid), or radioiodine treatment of a patient with partially treated or untreated

hyperthyroidism. Management requires intensive monitoring and supportive care, identification and treatment of the precipitating cause, and measures that reduce thyroid hormone synthesis. Large doses of propylthiouracil (600-mg loading dose and 200 to 300 mg every 6 h) should be given orally or by nasogastric tube or per rectum; the drug's inhibitory action on $T_4 ® T_3$ conversion makes it the agent of choice. One hour after the first dose of propylthiouracil, stable iodide is given to block thyroid hormone synthesis via the Wolff-Chaikoff effect (the delay allows the antithyroid drug to prevent the excess iodine from being incorporated into new hormone). A saturated solution of potassium iodide (5 drops SSKI every 6 h), or ipodate or iopanoic acid (0.5 mg every 12 h), may be given orally. (Sodium iodide, 0.25 g intravenously every 6 h is an alternative but is not generally available.) Propranolol should also be given to reduce tachycardia and other adrenergic manifestations (40 to 60 mg orally every 4 h; or 2 mg intravenously every 4 h). Although other b-adrenergic blockers can be used, high doses of propranolol have been documented to decrease $T_4 ® T_3$ conversion, and the doses can be easily adjusted. Caution is needed to avoid acute negative inotropic effects, but controlling the heart rate is important, as some patients develop a form of high-output heart failure. Additional therapeutic measures include glucocorticoids (e.g., dexamethasone, 2 mg every 6 h), antibiotics if infection is present, cooling, and intravenous fluids.

Ophthalmopathy requires no active treatment when it is mild or moderate, as there is usually spontaneous improvement. General measures include meticulous control of thyroid hormone levels, advice about cessation of smoking, and an explanation of the natural history of ophthalmopathy. Discomfort can be relieved with artificial tears (e.g., 1% methylcellulose) and the use of dark glasses with side frames. Periorbital edema responds to a more upright sleeping position. Corneal exposure during sleep can be avoided by taping the eyelids shut. Minor degrees of diplopia improve with prisms fitted to spectacles. Severe ophthalmopathy, with optic nerve involvement or chemosis resulting in corneal damage, is an emergency requiring joint management with an ophthalmologist. Short-term benefit can be gained in about two-thirds of patients by the use of high-dose glucocorticoids (e.g., prednisone, 40 to 80 mg daily), sometimes combined with cyclosporine. Glucocorticoid doses are tapered by 5 mg every 1 to 2 weeks, but the taper often results in reemergence of congestive symptoms. Pulse therapy with intravenous methylprednisolone (1 g of methylprednisolone in 250 mL of saline infused over 2 h daily for 1 week) followed by an oral regimen is also used. Once the eye disease has stabilized, surgery may be indicated for relief of diplopia and correction of the appearance of the eyes. Orbital decompression can be achieved by removing bone from any wall of the orbit, thereby allowing displacement of fat and swollen extraocular muscles. The transantral route is used most often, as it requires no external incision. Proptosis recedes an average of 5 mm, but there may be residual or even worsened diplopia. Alternatively, retroorbital tissue can be decompressed without removal of bony tissue. External beam radiotherapy of the orbits has been used for many years, but the objective evidence that this therapy is beneficial remains equivocal.

Thyroid dermopathy does not usually require treatment but can cause cosmetic problems or interfere with the fit of shoes. Surgical removal is not indicated. Treatment consists of topical, high-potency glucocorticoid ointment under an occlusive dressing. Octreotide may be beneficial.

OTHER CAUSES OF THYROTOXICOSIS

Destructive thyroiditis (subacute or silent thyroiditis) typically presents with a short thyrotoxic phase due to the release of preformed thyroid hormones and catabolism of Tg (see "Subacute Thyroiditis," below). True hyperthyroidism is absent, as demonstrated by a low radionuclide uptake. Circulating Tg and IL-6 levels are usually increased. Other causes of thyrotoxicosis with low or absent thyroid radionuclide uptake include *thyrotoxicosis factitia*; iodine excess and, rarely, ectopic thyroid tissue, particularly teratomas of the ovary (*struma ovarii*); and functional metastatic follicular carcinoma. Whole-body radionuclide studies can demonstrate ectopic thyroid tissue, and thyrotoxicosis factitia can be distinguished from destructive thyroiditis by the clinical features and low levels of Tg. Amiodarone treatment is associated with thyrotoxicosis in up to 10% of patients, particularly in areas of low iodine intake.

TSH-secreting pituitary adenoma is a rare causes of thyrotoxicosis. It can be identified by the presence of an inappropriately normal or increased TSH level in a patient with hyperthyroidism, diffuse goiter, and elevated free T4 and T3 levels (Chap. 328). Elevated levels of the a subunit of TSH, released by the TSH-secreting adenoma, support this diagnosis, which can be confirmed by demonstrating the pituitary tumor on CT or MRI scan. A combination of transsphenoidal surgery, sella irradiation, and octreotide may be required to normalize TSH, as many of these tumors are large and locally invasive at the time of diagnosis. Radioiodine or antithyroid drugs can be used to control thyrotoxicosis.

Thyrotoxicosis caused by *toxic multinodular goiter* and *hyperfunctioning solitary nodules* is discussed below.

THYROIDITIS

A clinically useful classification of thyroiditis is based on the onset and duration of disease (Table 330-8).

ACUTE THYROIDITIS

Acute thyroiditis is rare and is due to suppurative infection of the thyroid. In children and young adults, the most common cause is the presence of a piriform sinus, a remnant of the fourth branchial pouch that connects the oropharynx with the thyroid. Such sinuses are predominantly left sided. A long-standing goiter and degeneration in a thyroid malignancy are risk factors in the elderly. The patient presents with thyroid pain, often referred to the throat or ears, and a small, tender goiter that may be asymmetric. Fever, dysphagia, and erythema over the thyroid are common, as are systemic symptoms of a febrile illness and lymphadenopathy.

The differential diagnosis of *thyroid pain* includes subacute or, rarely, chronic thyroiditis, hemorrhage into a cyst, malignancy including lymphoma, and, rarely, amiodarone-induced thyroiditis or amyloidosis. However, the abrupt presentation and clinical features of acute thyroiditis rarely cause confusion. The erythrocyte sedimentation rate (ESR) and white cell count are usually increased, but thyroid function is normal. FNA biopsy shows infiltration by polymorphonuclear leukocytes; culture of the sample can identify the organism. Caution is needed in immunocompromised patients

as fungal or *Pneumocystis* thyroiditis can occur in this setting. Antibiotic treatment is guided initially by Gram stain and subsequently by cultures of the FNA biopsy. Surgery may be needed to drain an abscess, which can be localized by CT scan or ultrasound. Tracheal obstruction, septicemia, retropharyngeal abscess, mediastinitis, and jugular venous thrombosis may complicate acute thyroiditis but are uncommon with prompt use of antibiotics.

SUBACUTE THYROIDITIS

This is also termed *de Quervain's thyroiditis*, *granulomatous thyroiditis*, or *viral thyroiditis*. Many viruses have been implicated, including mumps, coxsackie, influenza, adenoviruses, and echoviruses, but attempts to identify the virus in an individual patient are often unsuccessful and do not influence management. The diagnosis of subacute thyroiditis is often overlooked because the symptoms can mimic pharyngitis. The peak incidence occurs at 30 to 50 years, and women are affected three times more frequently than men.

Pathophysiology The thyroid shows a characteristic patchy inflammatory infiltrate with disruption of the thyroid follicles and multinucleated giant cells within some follicles. The follicular changes progress to granulomas accompanied by fibrosis. Finally, the thyroid returns to normal, usually several months after onset. During the initial phase of follicular destruction, there is release of Tg and thyroid hormones, leading to increased circulating free T_4 and T_3 and suppression of TSH (Fig. 330-8). During this destructive phase, radioactive iodine uptake is low or undetectable. After several weeks, the thyroid is depleted of stored thyroid hormone and a phase of hypothyroidism typically occurs, with low free T_4 (and sometimes T_3) and moderately increased TSH levels. Radioactive iodine uptake returns to normal or is even increased as a result of the rise in TSH. Finally, thyroid hormone and TSH levels return to normal as the disease subsides.

Clinical Manifestations The patient usually presents with a painful and enlarged thyroid, sometimes accompanied by fever. There may be features of thyrotoxicosis or hypothyroidism, depending on the phase of the illness. Malaise and symptoms of an upper respiratory tract infection may precede the thyroid-related features by several weeks. In other patients, the onset is acute, severe, and without obvious antecedent. Though the patient typically complains of a sore throat, examination reveals a small goiter that is exquisitely tender, and asymmetry is common. Pain is often referred to the jaw or ear. Complete resolution is the usual outcome, but permanent hypothyroidism can occur, particularly in those with coincidental thyroid autoimmunity. A prolonged course over many months, with one or more relapses, occurs in a small percentage of patients.

Laboratory Evaluation As depicted in Fig. 330-8, thyroid function tests characteristically evolve through three distinct phases over about 6 months: (1) thyrotoxic phase, (2) hypothyroid phase, and (3) recovery phase. In the thyrotoxic phase, T_4 and T_3 levels are increased, reflecting their discharge from the damaged thyroid cells, and TSH is suppressed. The T_4/T_3 ratio is greater than in Graves' disease or thyroid autonomy, in which T_3 is often disproportionately increased. The diagnosis is confirmed by a high ESR and low radioiodine uptake. Serum IL-6 levels increase during the thyrotoxic phase. The white blood cell count may be increased, and thyroid

antibodies are negative. If the diagnosis is in doubt, FNA biopsy may be useful, particularly to distinguish unilateral involvement from bleeding into a cyst or neoplasm.

TREATMENT

Relatively large doses of aspirin (e.g., 600 mg every 4 to 6 h) or nonsteroidal anti-inflammatory drugs are sufficient to control symptoms in most cases. If this treatment is inadequate, or if the patient has marked local or systemic symptoms, glucocorticoids should be given. The usual starting dose is 40 to 60 mg prednisone, depending on severity. The dose is gradually tapered over 6 to 8 weeks, in response to improvement in symptoms and the ESR. If a relapse occurs during glucocorticoid withdrawal, treatment should be started again and withdrawn more gradually. In these patients, it is useful to wait until the radioactive iodine uptake normalizes before stopping treatment. Thyroid function should be monitored every 2 to 4 weeks using TSH and free T_4 levels. Symptoms of thyrotoxicosis improve spontaneously but may be ameliorated by b-adrenergic blockers; antithyroid drugs play no role in treatment of the thyrotoxic phase. Levothyroxine replacement may be needed if the hypothyroid phase is prolonged, but doses should be low enough (50 to 100 ug daily) to allow TSH-mediated recovery.

SILENT THYROIDITIS

Painless thyroiditis, or *"silent" thyroiditis*, occurs in patients with underlying autoimmune thyroid disease. It has a clinical course similar to that of subacute thyroiditis, except that there is little or no thyroid tenderness. The condition occurs most frequently 3 to 6 months after pregnancy and is then termed *post-partum thyroiditis*. Typically, patients have a brief phase of thyrotoxicosis, lasting 2 to 4 weeks, followed by hypothyroidism for 4 to 12 weeks, and then resolution; often, however, only one phase is apparent. As in subacute thyroiditis, the radioactive iodine uptake is initially suppressed. In addition to the painless goiter, silent thyroiditis can be distinguished from subacute thyroiditis by the normal ESR and the presence of TPO antibodies. Glucocorticoid treatment is not indicated for silent thyroiditis. Severe thyrotoxic symptoms can be managed with a brief course of propranolol, 20 to 40 mg three or four times daily. Thyroxine replacement may be needed for the hypothyroid phase but should be withdrawn after 6 to 9 months, as recovery is the rule. Annual follow-up thereafter is recommended, as a proportion of these individuals develop permanent hypothyroidism.

DRUG-INDUCED THYROIDITIS

Patients receiving IFN-a, IL-2, or amiodarone may develop painless thyroiditis. INF-a, which is used to treat chronic hepatitis B or C, causes thyroid dysfunction in up to 5% of treated patients. It has been associated with painless thyroiditis, hypothyroidism, and Graves' disease. IL-2, which has been used to treat various malignancies, has also been associated with thyroiditis and hypothyroidism, though fewer patients have been studied. For discussion of amiodarone, see "Amiodarone Effects on Thyroid Function," below.

CHRONIC THYROIDITIS

The most common cause of chronic thyroiditis is *Hashimoto's thyroiditis*, an autoimmune disorder that often presents as a firm or hard goiter of variable size (see above). *Riedel's thyroiditis* is a rare disorder that typically occurs in middle-aged women. It presents with an insidious, painless goiter with local symptoms due to compression of the esophagus, trachea, neck veins, or recurrent laryngeal nerves. Dense fibrosis disrupts normal gland architecture and can extend outside the thyroid capsule. Despite these extensive histologic changes, thyroid dysfunction is uncommon. The goiter is hard, nontender, often asymmetric and fixed, leading to suspicion of a malignancy. Diagnosis requires open biopsy as FNAbiopsy is usually unhelpful. Treatment is surgical and directed to relief of compressive symptoms. There is an association between Riedel's thyroiditis and idiopathic fibrosis at other sites (retroperitoneum, mediastinum, biliary tree, lung, and orbit).

SICK EUTHYROID SYNDROME

Any acute, severe illness can cause abnormalities of circulatingTSH or thyroid hormone levels in the absence of underlying thyroid disease, making these measurements potentially misleading. The major cause of these hormonal changes is the release of cytokines. Unless a thyroid disorder is strongly suspected, the routine testing of thyroid function should be avoided in acutely ill patients.

The most common hormone pattern in sick euthyroid syndrome (SES) is a decrease in total and free T_3levels (low T_3syndrome) with normal levels of T_4 andTSH. The magnitude of the fall in T_3correlates with the severity of the illness. T_4conversion to T_3 via peripheral deiodination is impaired, leading to increased reverse T_3(rT_3). Despite this effect, decreased clearance rather than increased production is the major basis for increased rT_3. Also, $T_4$is alternately metabolized to the hormonally inactive T_3sulfate. It is generally assumed that this low T_3 state is adaptive, as it can be induced in normal individuals by fasting. Teleologically, the fall in T_3 may provide a mechanism for limiting catabolism in starved or ill patients.

Very sick patients may have a fall in total $T_4$and T_3levels (low T_4syndrome). This state has a poor prognosis. A key factor in the fall in T_4levels is altered binding toTBG. Free T_4assays usually demonstrate a normal free T_4 level in such patients, depending on the assay method used. Fluctuation inTSHlevels also creates challenges in the interpretation of thyroid function in sick patients. TSH levels may range from <0.1 to>20 mU/L; these alterations reverse after recovery, confirming the absence of underlying thyroid disease. A rise in cortisol or administration of glucocorticoids may provide a partial explanation for decreased TSH levels. However, the exact mechanisms underlying the subnormal TSH seen in 10% of sick patients and the increased TSH seen in 5% remain unclear.

Any severe illness can induce changes in thyroid hormone levels, but certain disorders exhibit a distinctive pattern of abnormalities. Acute liver disease is associated with an initial rise in total (but not free) T_3 and T_4levels, due toTBGrelease; these levels become subnormal with progression to liver failure. A transient increase in total and free T_4levels, usually with a normal T_3level, is seen in 5 to 30% of acutely ill psychiatric patients.TSHvalues may be transiently low, normal, or high in these patients. In the early stage of HIV infection, T_3 and T_4levels rise, even if there is weight loss. T_3levels fall

with progression to AIDS, but TSH levels usually remain normal. Renal disease is often accompanied by low T_3 concentrations, but with normal rather than increased rT_3 levels, due to an unknown factor that increases uptake of rT_3 into the liver.

The diagnosis of the SES is challenging. Historic information may be limited, and patients often have multiple metabolic derangements. Useful features to consider include previous history of thyroid disease and thyroid function tests, evaluation of the severity and time course of the patient's acute illness, documentation of medications that may affect thyroid function or thyroid hormone levels, and measurements of rT_3 together with free thyroid hormones and TSH. The diagnosis of SES is frequently presumptive, given the clinical context and pattern of laboratory values; only resolution of the test results with clinical recovery can clearly establish this disorder. Treatment of SES with thyroid hormone (T_4 and/or T_3) is controversial, but most authorities recommend monitoring the patient's thyroid function tests during recovery, without administering thyroid hormone, unless there is historic or clinical evidence suggestive of hypothyroidism. Sufficiently large randomized controlled trials using thyroid hormone are unlikely to resolve this therapeutic controversy in the near future, because clinical presentations and outcomes are highly variable.

AMIODARONE EFFECTS ON THYROID FUNCTION

Amiodarone is a commonly used type III antiarrhythmic agent (Chap. 230). It is structurally related to thyroid hormone and contains 39% iodine by weight. Thus, typical doses of amiodarone (200 mg/d) are associated with very high iodine intake, leading to >40-fold increases in plasma and urinary iodine levels. Moreover, because amiodarone is stored in adipose tissue, high iodine levels persist for >6 months after discontinuation of the drug. Amiodarone inhibits deiodinase activity, and its metabolites function as weak antagonists of thyroid hormone action. Amiodarone has the following multiple effects on thyroid function: (1) acute, transient changes in thyroid function; (2) hypothyroidism in patients susceptible to the inhibitory effects of a high iodine load; and (3) thyrotoxicosis that may be caused by at least three mechanisms -- a Jod-Basedow effect from the iodine load in the setting of multinodular goiter, a thyroiditis-like condition, and possibly induction of autoimmune Graves' disease.

The initiation of amiodarone treatment is associated with a transient decrease of T_4 levels, reflecting the inhibitory effect of iodine on T_4 release. Soon thereafter, most individuals escape from iodide-dependent suppression of the thyroid (Wolff-Chaikoff effect), and the inhibitory effects on deiodinase activity and thyroid hormone receptor action become predominant. These events lead to the following pattern of thyroid function tests: increased T_4, decreased T_3, increased rT_3, and a transient increase of TSH (up to 20 mU/L). TSH levels normalize or are slightly suppressed after about 1 to 3 months.

The incidence of hypothyroidism from amiodarone varies geographically, apparently correlating with iodine intake. Hypothyroidism occurs in up to 13% of amiodarone-treated patients in iodine-replete countries, such as the United States, but is less common (<6% incidence) in areas of lower iodine intake, such as Italy or Spain. The pathogenesis appears to involve an inability of the thyroid to escape from the high iodine load. Consequently, amiodarone-associated hypothyroidism is more common in

women and individuals with positive TPO antibodies. It is usually unnecessary to discontinue amiodarone for this side effect, as levothyroxine can be used to normalize thyroid function. TSH levels should be monitored, because T4 levels are often increased for the reasons described above.

The management of amiodarone-induced thyrotoxicosis (AIT) is complicated by the fact that there are several causes of thyrotoxicosis and because the increased thyroid hormone levels exacerbate underlying arrhythmias and coronary artery disease. Amiodarone treatment causes thyrotoxicosis in 10% of patients living in areas of low iodine intake and in 2% of patients in regions of high iodine intake. There are two major forms of AIT. Type 1 AIT is associated with an underlying thyroid abnormality (preclinical Graves' disease or nodular goiter). Thyroid hormone synthesis becomes excessive as a result of increased iodine exposure (Jod-Basedow phenomenon). Type 2 AIT occurs in individuals with no intrinsic thyroid abnormalities and is the result of drug-induced lysosomal activation leading to destructive thyroiditis with histiocyte accumulation in the thyroid. Mild forms of type 2 AIT can resolve spontaneously or can occasionally lead to hypothyroidism. Color-flow doppler thyroid scanning shows increased vascularity in type 1 but decreased vascularity in type 2 AIT; IL-6 levels are markedly raised in type 2 but only slightly increased in type 1 AIT. Thyroid scans are difficult to interpret in this setting, because the high endogenous iodine levels diminish tracer uptake. However, the presence of normal or increased uptake favors type 1 AIT.

In amiodarone-induced thyrotoxicosis the drug should be stopped, if possible, though this is often impractical because of the underlying cardiac disorder. Discontinuation of amiodarone will not have an acute effect because of its storage and prolonged half-life. High doses of antithyroid drugs can be used in type 1 AIT but are often ineffective. Potassium perchlorate, 200 mg every 6 h, has been used to reduce thyroidal iodide content. Perchlorate treatment has been associated with agranulocytosis, though the risk appears relatively low with short-term use. Glucocorticoids, administered as for subacute thyroiditis, are beneficial in type 2 AIT. Lithium blocks thyroid hormone release and can provide modest benefit. Near-total thyroidectomy rapidly decreases thyroid hormone levels and may be the most effective long-term solution, if the patient can undergo the procedure safely.

THYROID FUNCTION IN PREGNANCY

Three factors alter thyroid function in pregnancy: (1) the transient increase in hCG during the first trimester, which stimulates the TSH-R; (2) the estrogen-induced rise in TBG during the first trimester, which is sustained during pregnancy; and (3) increased urinary iodide excretion, which can cause impaired thyroid hormone production in areas of marginal iodine sufficiency. Women with a precarious iodine intake (<50 ug/d) are most at risk of developing a goiter during pregnancy, and iodine supplementation should be considered to prevent maternal and fetal hypothyroidism and the development of neonatal goiter.

The rise in circulating hCG levels during the first trimester is accompanied by a reciprocal fall in TSH that persists into the middle of pregnancy. This appears to reflect weak binding of hCG, which is present at very high levels, to the TSH-R. Rare individuals have been described with variant TSH-R sequences that enhance hCG binding and TSH-R

activation. Occasionally these hCG-induced changes in thyroid function result in transient gestational hyperthyroidism and/or *hyperemesis gravidarum*, a condition characterized by severe nausea and vomiting and risk of volume depletion. Antithyroid drugs are rarely needed, and parenteral fluid replacement usually suffices until the condition resolves.

Maternal hypothyroidism occurs in 2 to 3% of women of child-bearing age and is associated with increased risk of developmental delay in the offspring. Thyroid hormone requirements are increased by 25 to 50 ug/d during pregnancy.

GOITER AND NODULAR THYROID DISEASE

Goiter refers to an enlarged thyroid gland. Biosynthetic defects, iodine deficiency, autoimmune disease, and nodular diseases can each lead to goiter, though by different mechanisms. Biosynthetic defects and iodine deficiency are associated with reduced efficiency of thyroid hormone synthesis, leading to increased TSH, which stimulates thyroid growth as a compensatory mechanism to overcome the block in hormone synthesis. Graves' disease and Hashimoto's thyroiditis are also associated with goiter. In Graves' disease, the goiter results mainly from the TSH-R-mediated effects of TSI. The goitrous form of Hashimoto's thyroiditis occurs because of acquired defects in hormone synthesis, leading to elevated levels of TSH and its consequent growth effects. Lymphocytic infiltration and immune system-induced growth factors also contribute to thyroid enlargement in Hashimoto's thyroiditis. Nodular disease is characterized by the disordered growth of thyroid follicles, often combined with the gradual development of fibrosis. The management of goiter differs in patients depending on the etiology, and the detection of thyroid enlargement on physical examination should prompt further evaluation to identify its cause.

Nodular thyroid disease is common, occurring in about 3 to 7% of adults when assessed by physical examination. Using more sensitive techniques, such as ultrasound, it is present in >25% of adults. Thyroid nodules may be solitary or multiple, and they may be functional or nonfunctional.

DIFFUSE NONTOXIC (SIMPLE) GOITER

Etiology and Pathogenesis When diffuse enlargement of the thyroid occurs in the absence of nodules and hyperthyroidism, it is referred to as a *diffuse nontoxic goiter*. This is sometimes called *simple goiter*, because of the absence of nodules, or *colloid goiter*, because of the presence of uniform follicles that are filled with colloid. Worldwide, diffuse goiter is most commonly caused by iodine deficiency and is termed *endemic goiter* when it affects >5% of the population. In nonendemic regions, *sporadic goiter* occurs, and the cause is usually unknown. Thyroid enlargement in teenagers is sometimes referred to as *juvenile goiter*. In general, goiter is more common in women than men, probably because of the greater prevalence of underlying autoimmune disease and the increased iodine demands associated with pregnancy.

In *iodine-deficient areas*, thyroid enlargement reflects a compensatory effort to trap iodide and produce sufficient hormone under conditions in which hormone synthesis is relatively inefficient. Somewhat surprisingly, TSH levels are usually normal or only slightly

increased, suggesting increased sensitivity to TSH or activation of other pathways that lead to thyroid growth. Iodide appears to have direct actions on thyroid vasculature and may indirectly affect growth through vasoactive substances such as endothelins and nitric oxide. Endemic goiter is also caused by exposure to environmental *goitrogens* such as cassava root, which contains a thiocyanate, vegetables of the Cruciferae family (e.g., brussels sprouts, cabbage, and cauliflower), and milk from regions where goitrogens are present in grass. Though relatively rare, inherited defects in thyroid hormone synthesis also lead to a diffuse nontoxic goiter. These involve abnormalities at each step in hormone synthesis including iodide transport (NIS),Tgsynthesis, organification and coupling (TPO), and the regeneration of iodide (dehalogenase).

Clinical Manifestations and Diagnosis If thyroid function is preserved, most goiters are asymptomatic. Spontaneous hemorrhage into a cyst or nodule may cause the sudden onset of localized pain and swelling. Examination of a diffuse goiter reveals a symmetrically enlarged, nontender, generally soft gland without palpable nodules. Goiter is defined, somewhat arbitrarily, as a lateral lobe with a volume greater than the thumb of the individual being examined. If the thyroid is markedly enlarged, it can cause tracheal or esophageal compression. These features are unusual, however, in the absence of nodular disease and fibrosis. *Substernal goiter* may obstruct the thoracic inlet. *Pemberton's sign* refers to symptoms of faintness with evidence of facial congestion and external jugular venous obstruction when the arms are raised above the head, a maneuver that draws the thyroid into the thoracic inlet. Respiratory flow measurements andCT orMRIshould be used to evaluate substernal goiter in patients with obstructive signs or symptoms.

Thyroid function tests should be performed in all patients with goiter to exclude thyrotoxicosis or hypothyroidism. It is not unusual, particularly in iodine deficiency, to find a low total T_4, with normal T_3 andTSH, reflecting enhanced T_4®T_3conversion. A low TSH, particularly in older patients, suggests the possibility of thyroid autonomy or undiagnosed Graves' disease, causing subclinical thyrotoxicosis.TPOantibodies may be useful to identify patients at increased risk of autoimmune thyroid disease. Low urinary iodine levels (<100 ug/L) support a diagnosis of iodine deficiency. Thyroid scanning is not generally necessary but will reveal increased uptake in iodine deficiency and most cases of dyshormonogenesis. Ultrasound is not generally indicated in the evaluation of diffuse goiter, unless a nodule is palpable on physical examination.

TREATMENT

Iodine or thyroid hormone replacement induces variable regression of goiter in iodine deficiency, depending on how long it has been present and the degree of fibrosis that has developed. For other causes of nontoxic diffuse goiter, levothyroxine can be used in an attempt to reduce goiter size. Because of the possibility of underlying thyroid autonomy, caution should be exercised when instituting suppressive thyroxine therapy, particularly if the baselineTSH is in the low-normal range. In younger patients, the dose can be started at 100 ug/d and adjusted to suppress the TSH into the low-normal but detectable range. Treatment of elderly patients should be initiated at 50 ug/d. The efficacy of suppressive treatment is greater in younger patients and in those with soft goiters. Significant regression is usually seen within 3 to 6 months of treatment; after this time it is unlikely to occur. In older patients, and in those with some degree of

nodular disease or fibrosis, fewer than one-third demonstrate significant shrinkage of the goiter. Surgery is rarely indicated for diffuse goiter. Exceptions include documented evidence of tracheal compression or obstruction of the thoracic inlet, which are more likely to be associated with substernal multinodular goiters (see below). Subtotal or near-total thyroidectomy for these or cosmetic reasons should be performed by an experienced surgeon to minimize complication rates, which occur in up to 10% of cases. Surgery should be followed by mild suppressive treatment with levothyroxine to prevent regrowth of the goiter. Radioiodine reduces goiter size by about 50% in the majority of patients. It is rarely associated with transient acute swelling of the thyroid, which is usually inconsequential unless there is severe tracheal narrowing. If not treated with levothyroxine, patients should be followed after radioiodine treatment for the possible development of hypothyroidism.

NONTOXIC MULTINODULAR GOITER

Etiology and Pathogenesis Depending on the geographic region and the sensitivity of the methods used to detect the disorder, multinodular goiter (MNG) is common, occurring in between 1 and 12% of the population. MNG is more common in women than men and increases in prevalence with age. It is more common in iodine-deficient regions but also occurs in regions of iodine sufficiency, reflecting multiple genetic, autoimmune, and environmental influences on the pathogenesis.

Individual patients exhibit wide variation in nodule size. Histology reveals a spectrum of morphologies ranging from hypercellular regions to cystic areas filled with colloid. Fibrosis is often extensive, and areas of hemorrhage or lymphocytic infiltration may be seen. Using molecular techniques, most nodules within a MNG are polyclonal in origin, suggesting a hyperplastic response to locally produced growth factors and cytokines. TSH, which is usually not elevated, may play a permissive or contributory role. Monoclonal lesions also occur within a MNG, reflecting mutations in genes that confer a selective growth advantage to the progenitor cell.

Clinical Manifestations Most patients with nontoxic MNG are asymptomic and, by definition, euthyroid. MNG typically develops over many years and is detected on routine physical examination or because an individual notices an enlargement in the neck. If the goiter is large enough, it can ultimately lead to compressive symptoms including difficulty swallowing, respiratory distress (tracheal compression), or plethora (venous congestion), but these symptoms are uncommon. Symptomatic MNGs are usually extraordinarily large and/or develop fibrotic areas that cause compression. Sudden pain in a MNG is often caused by hemorrhage into a nodule but should raise the possibility of invasive malignancy. Hoarseness, reflecting laryngeal nerve involvement, also suggests malignancy.

Diagnosis On examination, thyroid architecture is distorted and multiple nodules of varying size can be appreciated. Substernal goiter is suggested by Pemberton's sign. Because many nodules are deeply embedded in thyroid tissue or reside in posterior or substernal locations, it is not possible to palpate all nodules. A TSH level should be measured to exclude subclinical hyper- or hypothyroidism, but thyroid function is usually normal. Tracheal deviation is common, but compression must usually exceed 70% of the tracheal diameter before there is significant airway compromise. Pulmonary function

testing can be used to assess the functional effects of compression and to detect tracheomalacia, which characteristically causes inspiratory stridor.CT orMRI can be used to evaluate the anatomy of the goiter and the extent of substernal extension, which is often much greater than is apparent on physical examination. A barium swallow may reveal the extent of esophageal obstruction.MNG does not appear to predispose to thyroid carcinoma or to more aggressive carcinoma. For this reason, and because it is not possible to biopsy all nodular lesions, thyroid biopsies should only be performed if malignancy is suspected because of a dominant or enlarging nodule.

TREATMENT

Most nontoxicMNGs can be managed conservatively. T_4suppression is rarely effective for reducing goiter size and introduces the risk of thyrotoxicosis, if there is underlying autonomy or if it develops during treatment. If levothyroxine is used, it should be started at low doses (50 ug) and advanced gradually while monitoring theTSHlevel to avoid excessive suppression. Contrast agents and other iodine-containing substances should be avoided because of the risk of inducing the *Jod-Basedow effect*, characterized by enhanced thyroid hormone production by autonomous nodules. Radioiodine is being used with increasing frequency because it often decreases goiter size and may selectively ablate regions of autonomy. Dosage of$_{131}$I depends on the size of the goiter and radioiodine uptake but is usually about 3.7 MBq (0.1 mCi) per gram of tissue, corrected for uptake [typical dose, 370 to 1070 Mbq (10 to 29 mCi)]. Repeat treatment may be needed. It is possible to achieve a 40 to 50% reduction in goiter size in most patients. Earlier concerns about radiation-induced thyroid swelling and tracheal compression have diminished as recent studies have shown this complication to be rare. When acute compression occurs, glucocorticoid treatment or surgery may be needed. Radiation-induced hypothyroidism is less common than occurs after treatment for Graves' disease. However, posttreatment autoimmune thyrotoxicosis may occur in up to 5% of patients treated for nontoxic MNG. Surgery remains highly effective but is not without risk, particularly in older patients with underlying cardiopulmonary disease.

TOXIC MULTINODULAR GOITER

The pathogenesis of toxicMNGappears to be similar to that of nontoxic MNG, the major difference being the presence of functional autonomy in toxic MNG. The molecular basis for autonomy in toxic MNG remains unknown. As in nontoxic goiters, many nodules are polyclonal, while others are monoclonal and vary in their clonal origins. Genetic abnormalities known to confer functional autonomy, such as activatingTSH-R or G_{sa}mutations (see below), are not usually found in the autonomous regions of toxic MNG goiter.

In addition to features of goiter, the clinical presentation of toxicMNGincludes subclinical hyperthyroidism or mild thyrotoxicosis. The patient is usually elderly and may present with atrial fibrillation or palpitations, tachycardia, nervousness, tremor, or weight loss. Recent exposure to iodine, from contrast dyes or other sources, may precipitate or exacerbate thyrotoxicosis. TheTSHlevel is low. The T_4 level may be normal or minimally increased; T_3 is often elevated to a greater degree than T_4. Thyroid scan shows heterogeneous uptake with multiple regions of increased and decreased uptake; 24-h uptake of radioiodine may not be increased.

TREATMENT

The management of toxicMNG is challenging. Antithyroid drugs, often in combination with beta blockers, can normalize thyroid function and address clinical features of thyrotoxicosis. This treatment, however, often stimulates the growth of the goiter, and, unlike in Graves' disease, spontaneous remission does not occur. Radioiodine can be used to treat areas of autonomy, as well as to decrease the mass of the goiter. Usually, however, some degree of autonomy remains, presumably because multiple autonomous regions emerge as soon as others are treated. Nonetheless, a trial of radioiodine should be considered before subjecting patients, many of whom are elderly, to surgery. Surgery provides definitive treatment of underlying thyrotoxicosis as well as goiter. Patients should be rendered euthyroid using antithyroid drugs before operation.

HYPERFUNCTIONING SOLITARY NODULE

A solitary, autonomously functioning thyroid nodule is referred to as *toxic adenoma*. The pathogenesis of this disorder has been unraveled by demonstrating the functional effects of mutations that stimulate theTSH-Rsignaling pathway. Most patients with solitary hyperfunctioning nodules have acquired somatic, activating mutations in the TSH-R (Fig. 330-9). These mutations, located primarily in the receptor transmembrane domain, induce constitutive receptor coupling to G_{sa}, increasing cyclic AMP levels and leading to enhanced thyroid follicular cell proliferation and function. Less commonly, somatic mutations are identified in G_{sa}. These mutations, which are similar to those seen in McCune-Albright syndrome (Chap. 336) or in a subset of somatotrope adenomas (Chap. 328), impair GTP hydrolysis, also causing constitutive activation of the cyclic AMP signaling pathway. In most series, activating mutations in either the TSH-R or the G_{sa}subunit genes are identified in >90% of patients with solitary hyperfunctioning nodules.

Thyrotoxicosis is usually mild. The disorder is suggested by the presence of the thyroid nodule, which is generally large enough to be palpable, and by the absence of clinical features suggestive of Graves' disease or other causes of thyrotoxicosis. A thyroid scan provides a definitive diagnostic test, demonstrating focal uptake in the hyperfunctioning nodule and diminished uptake in the remainder of the gland, as activity of the normal thyroid is suppressed.

TREATMENT

Radioiodine ablation is usually the treatment of choice. Because normal thyroid function is suppressed,$_{131}$I is concentrated in the hyperfunctioning nodule with minimal uptake and damage to normal thyroid tissue. Relatively large radioiodine doses [e.g., 370 to 1110 MBq (10 to 29.9 mCi)$_{131}$I] have been shown to correct thyrotoxicosis in about 75% of patients within 3 months. Hypothyroidism occurs in <10% of patients over the next 5 years. Surgical resection is also effective and is usually limited to enucleation of the adenoma or lobectomy, thereby preserving thyroid function and minimizing risk of hypoparathyroidism or damage to the recurrent laryngeal nerves. Medical therapy using antithyroid drugs and beta blockers can normalize thyroid function but is not an optimal long-term treatment. Ethanol injection under ultrasound guidance has been used

successfully in some centers to ablate hyperfunctioning nodules. Repeated injections (often more than 5 sessions) are required but reduce nodule size. Normal thyroid function can be achieved in most patients using this technique.

BENIGN NEOPLASMS

The various types of benign thyroid nodules are listed in Table 330-9. These lesions are common (5 to 10% adults) and often multiple, particularly when assessed by sensitive techniques such as ultrasound. The risk of malignancy is very low for *macrofollicular adenomas* and *normofollicular adenomas. Microfollicular, trabecular, and Hurthle cell variants* raise greater concern, partly because the histology is more difficult to interpret. About one-third of palpable nodules are *thyroid cysts*. These may be recognized by their ultrasound appearance or based on aspiration of large amounts of pink or straw-colored fluid (colloid). Many are mixed cystic/solid lesions, in which case it is desirable to aspirate cellular components under ultrasound or harvest cells after cytospin of cyst fluid. Cysts frequently recur, even after repeated aspiration, and may require surgical excision if they are large or if the cytology is suspicious. Sclerosis has been used with variable success but is often painful and may be complicated by infiltration of the sclerosing agent.

The treatment approach for benign nodules is similar to that forMNG.TSHsuppression with levothyroxine decreases the size of about 30% of nodules and may prevent further growth. The TSH level should be suppressed into the low-normal range, assuming there are no contraindications; alternatively, nodule size can be monitored without suppression. If a nodule has not decreased in size after 6 to 12 months of suppressive therapy, treatment should be discontinued as little benefit is likely to accrue from long-term treatment.

THYROID CANCER

Thyroid carcinoma is the most common malignancy of the endocrine system. Malignant tumors derived from the follicular epithelium are classified according to histologic features. Differentiated tumors, such as papillary thyroid cancer (PTC) or follicular thyroid cancer (FTC), are often curable, and the prognosis is good for patients identified with early-stage disease. In contrast, anaplastic thyroid cancer (ATC) is aggressive, responds poorly to treatment, and is associated with a bleak prognosis.

The incidence of thyroid cancer (~9/100,000 per year) increases with age, plateauing after about age 50 (Fig. 330-10). Age is also an important prognostic factor -- thyroid cancer at young age (<20) or in older persons (>65) is associated with a worse prognosis. Thyroid cancer is twice as common in women as men, but male sex is associated with a worse prognosis. Additional important risk factors include a history of childhood head or neck irradiation, large nodule size (^{3}4 cm), evidence for local tumor fixation or invasion into lymph nodes, and the presence of metastases (Table 330-10).

Several unique features of thyroid cancer facilitate its management: (1) thyroid nodules are readily palpable, allowing early detection and biopsy byFNA; (2) iodine radioisotopes can be used to diagnose ($_{123}$I) and treat ($_{131}$I) differentiated thyroid cancer, reflecting the unique uptake of this anion by the thyroid gland; and (3) serum markers allow the

detection of residual or recurrent disease, including the use of Tg levels for PTC and FTC and calcitonin for medullary thyroid cancer (MTC).

CLASSIFICATION

Thyroid neoplasms can arise in each of the cell types that populate the gland, including thyroid follicular cells, calcitonin-producing C cells, lymphocytes, and stromal and vascular elements, as well as metastases from other sites (Table 330-9). The American Joint Committee on Cancer (AJCC) has designated a staging system using the TNM classification (Table 330-11). Several other classification and staging systems are also widely used, some of which place greater emphasis on histologic features or risk factors such as age or gender.

PATHOGENESIS AND GENETIC BASIS

Radiation Early studies of the pathogenesis of thyroid cancer focused on the role of external radiation, which predisposes to chromosomal breaks, presumably leading to genetic rearrangements and loss of tumor-suppressor genes. External radiation of the mediastinum, face, head, and neck region was administered in the past to treat an array of conditions including acne and enlargement of the thymus, tonsils, and adenoids. Radiation exposure increases the risk of benign and malignant thyroid nodules, is associated with multicentric cancers, and shifts the incidence of thyroid cancer to an earlier age group. Radiation from nuclear fallout also predisposes to thyroid cancer. Children seem more predisposed to the effects of radiation than adults. Of note, radiation derived from ^{131}I therapy appears to contribute little, if any, increased risk of thyroid cancer.

TSH and Growth Factors Thyroid growth is regulated primarily by TSH but also by a variety of growth factors and cytokines. Many differentiated thyroid cancers express TSH receptors and, therefore, remain responsive to TSH. This observation provides the rationale for T_4 suppression of TSH in patients with thyroid cancer. Residual expression of TSH receptors also allows TSH-stimulated uptake of ^{131}I therapy (see below).

Oncogenes and Tumor-Suppressor Genes Thyroid cancers are monoclonal in origin, consistent with the idea that they originate as a consequence of mutations that confer a growth advantage to a single cell. In addition to increased rates of proliferation, some thyroid cancers exhibit impaired apoptosis and features that enhance invasion, angiogenesis, and metastasis (Chap. 83). By analogy with the model of multistep carcinogenesis proposed for colon cancer (Chap. 81), thyroid neoplasms have been analyzed for a variety of genetic alterations, but without clear evidence of an ordered acquisition of somatic mutations as they progress from the benign to the malignant state. On the other hand, certain mutations are relatively specific for thyroid neoplasia, some of which correlate with histologic classification (Table 330-12). For example, activating mutations of the TSH-R and the $G_{s\alpha}$ subunit are associated with autonomously functioning nodules. Though these mutations induce thyroid cell growth, this type of nodule is almost always benign. A variety of rearrangements involving the *RET* gene on chromosome 10 bring this receptor tyrosine kinase under the control of other promoters, leading to receptor overexpression. *RET* rearrangements occur in 20 to 40% of PTCs in different series and were observed with increased frequency in tumors developing after

the Chernobyl radiation disaster. Rearrangements in PTC have also been observed for another tyrosine kinase gene, *TRK1*, which is located on chromosome 1. To date, the identification of PTC with *RET* or *TRK1* rearrangements has not proven useful for predicting prognosis or treatment responses. *RAS* mutations are found in about 20 to 30% of thyroid neoplasms, including adenomas as well as PTC and FTC, suggesting that these mutations do not strongly affect tumor phenotype. Loss of heterozygosity (LOH), consistent with deletions of tumor-suppressor genes, is particularly common in FTC, often involving chromosomes 3p or 11q. Mutations of the tumor suppressor, p53, appear to play an important role in the development of ATC. Because p53 plays a role in cell cycle surveillance, DNA repair, and apoptosis, its loss may contribute to the rapid acquisition of genetic instability as well as poor treatment responses (Chap. 82). The role of other tumor-suppressor genes in thyroid cancer is under investigation (Table 330-12).

MTC, when associated with multiple endocrine neoplasia (MEN) type 2, harbors an inherited mutation of the *RET* gene. Unlike the rearrangements of *RET* seen in PTC, the mutations in MEN-2 are point mutations that induce constitutive activity of the tyrosine kinase (Chap. 339). MTC is preceded by hyperplasia of the C cells, raising the likelihood that as-yet-unidentified "second hits" lead to cellular transformation. A subset of sporadic MTC contain somatic mutations that activate *RET*.

WELL-DIFFERENTIATED THYROID CANCER

Papillary PTC is the most common type of thyroid cancer, accounting for 70 to 90% of well-differentiated thyroid malignancies. Microscopic PTC is present in as many 25% of thyroid glands at autopsy, but most of these lesions are very small (several millimeters) and are not clinically significant. Characteristic cytologic features of PTC help make the diagnosis by FNA or after surgical resection; these include psammoma bodies, cleaved nuclei with an "orphan-Annie" appearance caused by large nucleoli, and the formation of papillary structures.

PTC tends to be multifocal and to invade locally within the thyroid gland as well as through the thyroid capsule and into adjacent structures in the neck. It has a propensity to spread via the lymphatic system but can metastasize as well, particularly to bone and lung. Because of the relatively slow growth of the tumor, a significant burden of pulmonary metastases may accumulate, sometimes with remarkably few symptoms. The prognostic implication of lymph node spread is debated. Lymph node involvement by thyroid cancer can be remarkably well tolerated but probably increases the risk of recurrence and mortality, particularly in older patients. The staging of PTC by the TNM system is outlined in Table 330-11. Most papillary cancers are identified in the early stages (>80% stages I or II) and have an excellent prognosis, with survival curves similar to expected survival (Fig. 330-11*A*). Mortality is markedly increased in stage IV disease (distant metastases), but this group comprises only about 1% of patients. The treatment of PTC is described below.

Follicular The incidence of FTC varies widely in different parts of the world; it is more common in iodine-deficient regions. FTC is difficult to diagnose by FNA because the distinction between benign and malignant follicular neoplasms rests largely on evidence of invasion into vessels, nerves, or adjacent structures. FTC tends to spread by

hematogenous routes leading bone, lung, and central nervous system metastases. Mortality rates associated with FTC are less favorable than for PTC, in part because a larger proportion of patients present with stage IV disease (Fig. 330-11*B*). Poor prognostic features include distant metastases, age >50 years, primary tumor size >4 cm, Hurthle cell histology, and the presence of marked vascular invasion.

TREATMENT

Surgery All well-differentiated thyroid cancers should be surgically excised. In addition to removing the primary lesion, surgery allows accurate histologic diagnosis and staging, and multicentric disease is commonly found in the contralateral lobe. Lymph node spread can also be assessed at the time of surgery, and involved nodes can be removed. Recommendations about the extent of surgery vary for stage I disease, as survival rates are similar for lobectomy and near-total thyroidectomy. Lobectomy is associated with a lower incidence of hypoparathyroidism and injury to the recurrent laryngeal nerves. However, it is not possible to monitor Tg levels or to perform whole-body ^{131}I scans in the presence of the residual lobe. Moreover, if final staging or subsequent follow-up indicates the need for radioiodine scanning or treatment, repeat surgery is necessary to remove the remaining thyroid tissue. The authors favor near-total thyroidectomy in almost all patients; complication rates are acceptably low if the surgeon is highly experienced in the procedure. This approach, in combination with postsurgical radioablation of the remnant thyroid tissue, facilitates the use of radioiodine scanning and Tg determinations to assess disease recurrence.

TSH Suppression Therapy As most tumors are still TSH-responsive, levothyroxine suppression of TSH is a mainstay of thyroid cancer treatment. Though TSH suppression clearly provides therapeutic benefit, there are no prospective studies that identify the optimal level of TSH suppression. A reasonable goal is to suppress TSH to as low as possible without subjecting the patient to unnecessary side effects from excess thyroid hormone, such as atrial fibrillation, osteopenia, anxiety, and other manifestations of thyrotoxicosis. For patients at low risk of recurrence, TSH should be suppressed into the low but detectable range (0.1 to 0.5 IU/L). For patients at high risk of recurrence, or with known metastatic disease, complete TSH suppression is indicated, if there are no strong contraindications to mild thyrotoxicosis. In this instance, free T$_4$ or free T$_3$ levels must also be monitored to avoid excessive treatment.

Radioiodine Treatment Well-differentiated thyroid cancer still incorporates radioiodine, though less efficiently than normal thyroid follicular cells. Radioiodine uptake is determined primarily by expression of the NIS and is stimulated by TSH, requiring expression of the TSH-R. The retention time for radioactivity is influenced by the extent to which the tumor retains differentiated functions such as iodide trapping and organification. After near-total thyroidectomy, substantial thyroid tissue remains, particularly in the thyroid bed and surrounding the parathyroid glands. Consequently, ^{131}I ablation is necessary to eliminate remaining normal thyroid tissue and may treat residual tumor cells.

Indications The use of therapeutic doses of radioiodine remains an area of controversy in thyroid cancer management. Postoperative thyroid ablation and radiodine treatment of known residual PTC or FTC reduce recurrence rates. ^{131}I ablation of remaining normal

thyroid tissue also facilitates the detection of recurrent disease, using either whole-body iodine scanning or measurements of Tg. For tumors that take up iodine, 131I treatment can reduce or eliminate residual disease with relatively little associated toxicity. However, it is not clear that prophylactic radioiodine treatment reduces mortality for patients at relatively low risk. Most patients with stage 1 PTC with primary tumors <1.5 cm in size can usually be managed safely with thyroxine suppression, without radiation treatment, as the risk of recurrence and mortality is very low. For patients with larger papillary tumors, spread to the adjacent lymph nodes, FTC, or evidence of metastases, thyroid ablation and radioiodine treatment are generally indicated.

131I thyroid ablation and treatment As noted above, the decision to use 131I for thyroid ablation should be coordinated with the surgical approach, as radioablation is much more effective when there is minimal remaining normal thyroid tissue. A typical strategy is to treat the patient for several weeks postoperatively with liothyronine (25 ug bid or tid), followed by thyroid hormone withdrawal. Ideally, the TSH level should increase to >50 IU/L over about 3 to 4 weeks. The level to which TSH rises is dictated largely by the amount of normal thyroid tissue remaining postoperatively. A scanning dose of 131I [usually 148 to 185 MBq (4 to 5 mCi)] will reveal the amount of residual tissue and provides guidance about the dose needed to accomplish ablation. A maximum outpatient dose of 1110 MBq (29.9 mCi) 131I can be administered in the United States, though ablation is often more complete using greater doses [1850 to 2775 MBq (50 to 75 mCi)]. In patients with known residual cancer, the larger doses ensure thyroid ablation and may destroy remaining tumor cells. A whole-body scan following the high-dose radioiodine treatment is useful to identify possible metastatic disease.

Follow-up whole-body thyroid scanning and thyroglobulin determinations An initial whole-body scan should be performed about 6 months after surgery and thyroid ablation. The strategy for follow-up management of thyroid cancer has been altered by the availability of recombinant human TSH (rhTSH) to stimulate 131I uptake and by the improved sensitivity of Tg assays to detect residual or recurrent disease. A scheme for using either rhTSH or thyroid hormone withdrawal for thyroid scanning is summarized in Fig. 330-12. After thyroid ablation, rhTSH can be used to stimulate 131I uptake without subjecting patients to thyroid hormone withdrawal and its associated symptoms of hypothyroidism and the risk of prolonged TSH-stimulated tumor growth. This approach is recommended for patients predicted to be at low risk of disease recurrence, since rhTSH is not currently approved for use in conjunction with therapeutic doses of 131I. Alternatively, in patients who are likely to require 131I treatment, the traditional approach of thyroid hormone withdrawal can be used to increase TSH. This involves switching patients from levothyroxine (T4) to the more rapidly cleared hormone, liothyronine (T3), thereby allowing TSH to increase more quickly. If residual disease is detected on the initial whole-body scan [148 to 185 MBq (4 to 5 mCi)], a larger treatment dose, usually between 2775 and 5550 MBq (75 and 150 mCi), can be administered depending on the degree of residual uptake and assessment of cancer risk. Because TSH stimulates Tg levels, Tg measurements should be obtained after administration of rhTSH or when TSH levels have risen after thyroid hormone withdrawal. If the initial whole-body scan is negative and Tg levels are low, a repeat scan should be performed 1 year later. If still negative, the patient can be managed with suppressive therapy and measurements of Tg every 6 to 12 months. If a second follow-up scan is negative, no further scanning may be necessary if the patient is at low risk and there is no clinical or laboratory

evidence of recurrence. Many authorities advocate radioiodine treatment for scan-negative, Tg-positive (Tg >5 to 10 ng/mL) patients, as many derive therapeutic benefit from a large dose of ^{131}I.

In addition to radioiodine, external beam radiotherapy is also used to treat specific metastatic lesions, particularly when they cause bone pain or threaten neurologic injury (e.g., vertebral metastases).

ANAPLASTIC AND OTHER FORMS OF THYROID CANCER

Anaplastic Thyroid Cancer As noted above, ATC is a poorly differentiated and aggressive cancer. The prognosis is poor, and most patients die within 6 months of diagnosis. Because of the undifferentiated state of these tumors, radioiodine uptake is usually negligible but can be used therapeutically if there is residual uptake. Chemotherapy has been attempted with multiple agents, including anthracyclines and paclitaxel, but is usually futile. External radiation therapy can be attempted and continued if tumors are responsive.

Thyroid Lymphoma Lymphoma in the thyroid gland often arises in the background of Hashimoto's thyroiditis. A rapidly expanding thyroid mass should suggest the possibility of this diagnosis. Diffuse large cell lymphoma is the most common type in the thyroid. Biopsies reveal sheets of lymphoid cells that can be difficult to distinguish from small cell lung cancer or ATC. These tumors are often highly sensitive to external radiation. Surgical resection should be avoided as initial therapy because it may spread disease that is otherwise localized to the thyroid. If staging indicates disease outside of the thyroid, treatment should follow guidelines used for other forms of lymphoma (Chap. 112).

MEDULLARY THYROID CARCINOMA

MTC can be sporadic or familial and accounts for about 5 to 10% of thyroid cancers. There are three familial forms of MTC: MEN-2A, MEN-2B, and familial MTC without other features of MEN (Chap. 339). In general, MTC is more aggressive in MEN-2B than in MEN-2A, and familial MTC is more aggressive than sporadic MTC. Elevated serum calcitonin provides a marker of residual or recurrent disease. It is reasonable to test all patients with MTC for *RET* mutations, as genetic counseling and testing of family members can be offered to those individuals who test positive for mutations.

The management of MTC is primarily surgical. Unlike tumors derived from thyroid follicular cells, these tumors do not take up radioiodine. External radiation treatment and chemotherapy may provide palliation in patients with advanced disease (Chap. 339).

Approach to the Patient

Palpable thyroid nodules are found in about 5% of adults, though the prevalence varies considerably worldwide. Given this high prevalence rate, it is common for the practitioner to identify and evaluate thyroid nodules. The main goal of this evaluation is to identify, in a cost-effective manner, the small subgroup of individuals with malignant lesions.

As described above, nodules are more common in iodine-deficient areas, in women, and with aging. Most palpable nodules are >1 cm in diameter, but the ability to feel a nodule is influenced by its location within the gland (superficial versus deeply embedded), the anatomy of the patient's neck, and the experience of the examiner. More sensitive methods of detection, such as thyroid ultrasound and pathologic studies, reveal thyroid nodules in>20% of glands. These findings have led to much debate about how to detect nodules and which nodules to investigate further. Most authorities still rely on physical examination to detect thyroid nodules, reserving ultrasound for monitoring nodule size or as an aid in thyroid biopsy.

It is important to distinguish whether a patient presents with a solitary thyroid nodule or a prominent nodule in the context of aMNG, as the incidence of malignancy is greater in solitary nodules. An approach to the evaluation of a solitary nodule is outlined inFig. 330-13. Most patients with thyroid nodules have normal thyroid function tests. Nonetheless, thyroid function should be assessed by measuring aTSHlevel, which may be suppressed by one or more autonomously functioning nodules. If the TSH is suppressed, a radionuclide scan is indicated to determine if the identified nodule is "hot," as lesions with increased uptake are almost never malignant andFNA is unnecessary. Otherwise, FNA biopsy should be the first step in the evaluation of a thyroid nodule. FNA has good sensitivity and specificity when performed by physicians familiar with the procedure and when the results are interpreted by experienced cytopathologists. The technique is particularly accurate for detectingPTC. The distinction of benign and malignant follicular lesions is often not possible using cytology alone.

In several large studies,FNAbiopsies yield the following findings: 70% benign, 10% malignant or suspicious for malignancy, and 20% nondiagnostic or yielding insufficient material for diagnosis. Characteristic features of malignancy mandate surgery. A diagnosis of follicular neoplasm also warrants surgery, as benign and malignant lesions cannot be distinguished based on cytopathology or frozen section. The management of patients with benign lesions is more variable. Many authorities advocateTSHsuppression, whereas others monitor nodule size without suppression. With either approach, thyroid nodule size should be monitored, either by palpation or ultrasound. Repeat FNA is indicated if a nodule enlarges, and most authorities recommend a second biopsy within 2 to 5 years to confirm the benign status of the nodule.

Nondiagnostic biopsies occur for many reasons, including a fibrotic reaction with relatively few cells available for aspiration, a cystic lesion in which cellular components reside along the cyst margin, or a nodule that may be too small for accurate aspiration. For these reasons, ultrasound-guidedFNA is useful when the FNA is repeated. Ultrasound is also increasingly used for initial biopsies in an effort to enhance nodule localization and the accuracy of sampling.

The evaluation of a thyroid nodule is stressful for most patients. They are concerned about the possibility of thyroid cancer, whether verbalized or not. It is constructive, therefore, to review the diagnostic approach and to reassure patients when malignancy is not found. When a suspicious lesion or thyroid cancer is identified, an explanation of

the generally favorable prognosis and available treatment options should be provided.

(Bibliography omitted in Palm version)

331. DISORDERS OF THE ADRENAL CORTEX - *Gordon H. Williams*, *Robert G. Dluhy*

BIOCHEMISTRY AND PHYSIOLOGY

The adrenal cortex produces three major classes of steroids: (1) glucocorticoids, (2) mineralocorticoids, and (3) adrenal androgens. Consequently, normal adrenal function is important for modulating intermediary metabolism and immune responses through glucocorticoids; blood pressure, vascular volume, and electrolytes through mineralocorticoids; and secondary sexual characteristics (in females) through androgens. The adrenal axis plays an important role in the stress response by rapidly increasing cortisol levels. Adrenal disorders include hyperfunction (Cushing's syndrome) and hypofunction (adrenal insufficiency) as well as a variety of genetic abnormalities of steroidogenesis.

STEROID NOMENCLATURE

Steroids contain as their basic structure a cyclopentenoperhydrophenanthrane nucleus consisting of three 6-carbon hexane rings and a single 5-carbon pentane ring (Fig. 331-1). The carbon atoms are numbered in a sequence beginning with ring A. Adrenal steroids contain either 19 or 21 carbon atoms. The C_{19} steroids have methyl groups at C-18 and C-19. C_{19} steroids with a ketone group at C-17 are termed *17-ketosteroids*; C_{19} steroids have predominantly androgenic activity. The C_{21} steroids have a 2-carbon side chain (C-20 and C-21) attached at position 17 and methyl groups at C-18 and C-19; C_{21} steroids with a hydroxyl group at position 17 are termed *17-hydroxycorticosteroids*. The C_{21} steroids have either glucocorticoid or mineralcorticoid properties.

BIOSYNTHESIS OF ADRENAL STEROIDS

Cholesterol, derived from the diet and from endogenous synthesis, is the substrate for steroidogenesis. Uptake of cholesterol by the adrenal cortex is mediated by the low-density lipoprotein (LDL) receptor. With long-term stimulation of the adrenal cortex by adrenocorticotropic hormone (ACTH), the number of LDL receptors increases. The three major adrenal biosynthetic pathways lead to the production of glucocorticoids (cortisol), mineralocorticoids (aldosterone), and adrenal androgens (dehydroepiandrosterone). Separate zones of the adrenal cortex synthesize specific hormones (Fig. 331-2). This zonation is accompanied by the selective expression of the genes encoding the enzymes unique to the formation of each type of steroid: aldosterone synthase is normally expressed only in the outer (glomerulosa) cell layer, whereas 17-hydroxylase is expressed only in the (inner) faciculata-reticularis cell layers, which are the sites of cortisol and androgen biosynthesis, respectively.

STEROID TRANSPORT

Cortisol occurs in the plasma in three forms: free cortisol, protein-bound cortisol, and cortisol metabolites. *Free cortisol* is physiologically active hormone that is not protein-bound and, can act therefore, directly on tissue sites. Normally, <5% of circulating cortisol is free. Only the unbound cortisol and its metabolites are filterable at

the glomerulus. Increased quantities of free steroid are excreted in the urine in states characterized by hypersecretion of cortisol, because the unbound fraction of plasma cortisol rises. Plasma has two cortisol-binding systems. One is a high-affinity, low-capacity α_2-globulin termed *transcortin* or *cortisol-binding globulin* (CBG), and the other is a low-affinity, high-capacity protein, *albumin*. The binding affinity of CBG for cortisol is reduced in areas of inflammation, thus increasing the local concentration of free cortisol. When the concentration of cortisol exceeds 700 nmol/(25 ug/dL), part of the excess binds to albumin, and a greater proportion than usual circulates unbound. The CBG level is increased in high-estrogen states (e.g., pregnancy, oral contraceptive administration). The rise in CBG is accompanied by a parallel rise in *protein-bound cortisol*, with the result that the plasma cortisol concentration is elevated. However, the free cortisol level probably remains normal, and manifestations of glucocorticoid excess are absent. Most synthetic glucocorticoid analogues bind less efficiently to CBG (~70% binding). This may explain the propensity of some synthetic analogues to produce cushingoid effects at low doses. *Cortisol metabolites* are biologically inactive and bind only weakly to circulating plasma proteins.

Aldosterone is bound to proteins to a smaller extent than cortisol, and an ultrafiltrate of plasma contains as much as 50% of the circulating concentration of aldosterone.

STEROID METABOLISM AND EXCRETION

Glucocorticoids The daily secretion of cortisol ranges between 40 and 80 umol (15 and 30 mg, 8-10 mg/m2), with a pronounced circadian cycle. The plasma concentration of cortisol is determined by the rate of secretion, the rate of inactivation, and the rate of excretion of free cortisol. The liver is the major organ responsible for steroid inactivation. A major enzyme regulating cortisol metabolism is 11b-hydroxysteroid dehydrogenase (11b-HSD). There are two isoforms: 11b-HSD I is primarily expressed in the liver and acts as a reductase, converting the inactive cortisone to the active glucocorticoid, cortisol; the 11b-HSD II isoform is expressed in a number of tissues and converts cortisol to the inactive metabolite cortisone. The oxidative reaction of 11b-HSD I is increased in hyperthyroidism.

Mineralocorticoids In normal individuals with a normal salt intake, the average daily secretion of aldosterone ranges between 0.1 and 0.7 umol (50 and 250 ug). During a single passage through the liver, >75% of circulating aldosterone is normally inactivated by ring A reduction and conjugation with glucuronic acid because it is only weakly bound to proteins. However, under certain conditions, such as congestive failure, this rate of inactivation is reduced.

Adrenal Androgens The major androgen secreted by the adrenal is dehydroepiandrosterone (DHEA) and its sulfuric acid ester (DHEAS). From 15 to 30 mg of these compounds is secreted daily. Smaller amounts of androstenedione, 11b-hydroxyandrostenedione, and testosterone are secreted. DHEA is the major precursor of the urinary 17-ketosteroids. Two-thirds of the urine 17-ketosteroids in the male are derived from adrenal metabolites, and the remaining one-third comes from testicular androgens. In the female, almost all urine 17-ketosteroids are derived from the adrenal.

Steroids diffuse passively through the cell membrane and bind to intracellular receptors (Chap. 327). Glucocorticoids and mineralocorticoids bind with nearly equal affinity to the mineralocorticoid receptor (MR). However, only glucocorticoids bind to the glucocorticoid receptor (GR). After the steroid binds to the receptor, the steroid-receptor complex is transported to the nucleus, where it binds to specific sites on steroid-regulated genes, altering levels of transcription. Some actions of glucocorticoids (e.g., anti-inflammatory effects) are mediated by GR-mediated inhibition of other transcription factors, such as activating protein-1 (AP-1) or nuclear factor kappa-B (NFKB), which normally stimulate the activity of various cytokine genes. Because cortisol binds to the MR with the same affinity as aldosterone, mineralocorticoid specificity is achieved by local metabolism of cortisol to the inactive compound cortisone by 11b-HSDII. The glucocorticoid effects of other steroids, such as high-dose progesterone, correlate with their relative binding affinities for the GR. Inherited defects in the GR cause glucocorticoid resistance states. Individuals with GR defects have high levels of cortisol but do not have manifestations of hypercortisolism.

In addition to the classic genomic effects, which are mediated by steroids binding to cytosolic receptors, evidence is accumulating that mineralocorticoids also have acute, nongenomic effects, presumably by activating a cell-surface receptor yet to be identified. This effect uses a G-protein signaling pathway; among the actions is modification of the sodium-hydrogen exchanger. This effect has been demonstrated in both epithelial and nonepithelial cells, e.g., myocytes and leukocytes.

ACTHPHYSIOLOGY

ACTH and a number of other peptides (lipotropins, endorphins, and melanocyte-stimulating hormones) are processed from a larger precursor molecule of 31,000 mol wt -- pro-opiomelanocortin (POMC) (Chap. 328). POMC is made in a variety of tissues, including brain, anterior and posterior pituitary, and lymphocytes. The constellation of POMC-derived peptides secreted depends on the tissue. ACTH, a 39-amino acid peptide, is synthesized and stored in basophilic cells of the anterior pituitary. The *N*-terminal 18-amino acid fragment of ACTH has full biologic potency, and shorter *N*-terminal fragments have partial biologic activity. Release of ACTH and related peptides from the anterior pituitary gland is stimulated by corticotropin-releasing hormone (CRH), a 41-amino acid peptide produced in the median eminence of the hypothalamus (Fig. 331-3). Urocortin, a neuropeptide related to CRH, also binds to CRH receptors. Urocortin mimics many of the central effects of CRH (e.g., appetite suppression, anxiety), but its role in ACTH regulation is unclear. Some related peptides such as b-lipotropin (b-LPT) are released in equimolar concentrations with ACTH, suggesting that they are cleaved enzymatically from the parent POMC before or during the secretory process. However, b-endorphin levels may or may not correlate with circulating levels of ACTH, depending on the nature of the stimulus. The functions and regulation of secretion of the related peptides derived from POMC are poorly understood.

The major factors controllingACTHrelease includeCRH, the free cortisol concentration in plasma, stress, and the sleep-wake cycle (Fig. 331-3). The plasma level of ACTH varies during the day as a result of its pulsatile secretion, and it follows a circadian pattern with a peak just prior to waking and a nadir before sleeping. If a new sleep-wake cycle is

adopted, the pattern changes over several days to conform to it. ACTH and cortisol levels also increase in response to eating. Stress (e.g., pyrogens, surgery, hypoglycemia, exercise, and severe emotional trauma) causes the release of CRH and arginine vasopressin (AVP) and activation of the sympathetic nervous system. These changes in turn enhance ACTH release, acting individually or in concert. For example, AVP release acts synergistically with CRH to amplify ACTH secretion; CRH also stimulates the locus coeruleus/sympathetic system. Stress-related secretion of ACTH abolishes the circadian periodicity of ACTH levels but is, in turn, suppressed by prior high-dose glucocorticoid administration. The normal pulsatile, circadian pattern of ACTH release is regulated by CRH; this mechanism is the so-called open feedback loop. CRH secretion, in turn, is influenced by hypothalamic neurotransmitters. For example, serotoninergic and cholinergic systems stimulate the secretion of CRH and ACTH; there is contradictory evidence regarding the inhibitory effects of a-adrenergic agonists andg-aminobutyric acid (GABA) on CRH release. In addition, there may be direct pituitary effects of these neurotransmitters. There is also evidence for peptidergic regulation of ACTH release. For example,b-endorphin and enkephalin inhibit the secretion of ACTH, whereas vasopressin and angiotensin II augment it. The immune system also influences the hypothalamic-pituitary-adrenal axis (Fig. 331-4). For example, inflammatory cytokines [tumor necrosis factor (TNF)-a, interleukin (IL)-1a, IL-1b, and IL-6] produced by monocytes increase ACTH release by stimulating secretion of CRH and/or AVP. Finally, ACTH release is regulated by the level of free cortisol in plasma. Cortisol decreases the responsiveness of pituitary corticotropic cells to CRH; the response of thePOMCmRNA to CRH is also inhibited by glucocorticoids. In addition, glucocorticoids inhibit the locus coeruleus/sympathetic system and CRH release. The latter servomechanism establishes the primacy of cortisol in the control of ACTH secretion. The inhibition of ACTH occurs in two phases: (1) an early fast feedback, mediated via the MR, which lasts <10 min and depends on both the rate of increase of glucocorticoid levels and the specific glucocorticoid administered; and (2) a time-dependent, delayed feedback, likely mediated by the GR, which is probably due to inhibition of synthesis of the precursor protein. The suppression of ACTH secretion that results in adrenal atrophy following *prolonged* glucocorticoid therapy is caused primarily by suppression of hypothalamic CRH release, as exogenous CRH administration in this circumstance produces a rise in plasma ACTH. Cortisol also exerts feedback effects on higher brain centers (hippocampus, reticular system, and septum) and perhaps on the adrenal cortex (Fig. 331-4).

The biologic half-life ofACTH in the circulation is <10 min. The action of ACTH is also rapid; within minutes of its release, the concentration of steroids in the adrenal venous blood increases. ACTH stimulates steroidogenesis via activation of the membrane-bound adenyl cyclase. Adenosine-3¢,5¢-monophosphate (cyclic AMP), in turn, activates protein kinase enzymes, thereby resulting in the phosphorylation of proteins that activate steroid biosynthesis.

RENIN-ANGIOTENSIN PHYSIOLOGY (See also Chap. 246)

Renin is a proteolytic enzyme that is produced and stored in the granules of the juxtaglomerular cells surrounding the afferent arterioles of glomeruli in the kidney. Renin exists in both active and inactive forms. Whether the inactive form is a precursor ("prorenin") or is a product formed after release is uncertain. Renin acts on the basic

substrate angiotensinogen (a circulating a_2-globulin made in the liver) to form the decapeptide angiotensin I ([Fig. 331-5](#)). Angiotensin I is then enzymatically transformed by angiotensin-converting enzyme (ACE), which is present in many tissues (particularly the pulmonary vascular endothelium), to the octapeptide angiotensin II by the removal of the two C-terminal amino acids. Angiotensin II is a potent pressor agent and exerts its action by a direct effect on arteriolar smooth muscle. In addition, angiotensin II stimulates production of aldosterone by the zona glomerulosa of the adrenal cortex; the heptapeptide angiotensin III may also stimulate aldosterone production. The two major classes of angiotensin receptors are termed *AT1* and *AT2*; AT1 may exist as two subtypesa and b. Most of the effects of angiotensins II and III are mediated by the AT1 receptor. Angiotensinases rapidly destroy angiotensin II (half-life, approximately 1 min), while the half-life of renin is more prolonged (10 to 20 min). In addition to circulating renin-angiotensin, many tissues have a local renin-angiotensin system and the ability to produce angiotensin II. These tissues include the uterus, placenta, vascular tissue, heart, brain, and, particularly, the adrenal cortex and kidney. Although the role of locally generated angiotensin II is not established, it may be involved in the growth and modulation of function of the adrenal cortex and vascular smooth muscle.

The amount of renin released reflects the combined effects of four interdependent factors. The *juxtaglomerular cells*, which are specialized myoepithelial cells that cuff the afferent arterioles, act as miniature pressure transducers, sensing renal perfusion pressure and corresponding changes in afferent arteriolar perfusion pressures. For example, under conditions of a reduction in circulating blood volume, there is a corresponding reduction in renal perfusion pressure and, therefore, in afferent arteriolar pressure ([Fig. 331-5](#)). This change is perceived by the juxtaglomerular cells as a decreased stretch exerted on the afferent arteriolar walls. The juxtaglomerular cells then release more renin into the renal circulation. This results in the formation of angiotensin I, which is converted in the kidney and peripherally to angiotensin II by[ACE](#). Angiotensin II influences sodium homeostasis via two major mechanisms: it changes renal blood flow so as to maintain a constant glomerular filtration rate, thereby changing the filtration fraction of sodium, and it stimulates the adrenal cortex to release aldosterone. Increasing plasma levels of aldosterone enhance renal sodium retention and thus result in expansion of the extracellular fluid volume, which, in turn, dampens the stimulus for renin release. In this context, the renin-angiotensin-aldosterone system regulates volume by modifying renal hemodynamics and tubular sodium transport.

A second control mechanism for renin release is centered in the *macula densa* cells, a group of distal convoluted tubular epithelial cells directly apposed to the juxtaglomerular cells. They may function as chemoreceptors, monitoring the sodium (or chloride) load presented to the distal tubule, and such information may be conveyed to the juxtaglomerular cells, where appropriate modifications in renin release take place. Under conditions of increased delivery of filtered sodium to the macula densa, increasing release of renin decreases the glomerular filtration rate, thereby reducing the filtered load of sodium.

The *sympathetic nervous system* regulates the release of renin in response to assumption of the upright posture. The mechanism is either a direct effect on the juxtaglomerular cell to increase adenyl cyclase activity or an indirect effect on either the juxtaglomerular or the macula densa cells via vasoconstriction of the afferent arteriole.

Finally, circulating factors influence renin release. Increased dietary intake of *potassium* decreases, and decreased potassium intake increases, renin release. The significance of these effects is unclear. *Angiotensin II* exerts negative feedback control on renin release that is independent of alterations in renal blood flow, blood pressure, or aldosterone secretion. *Atrial natriuretic peptides* also inhibit renin release. Thus, the control of renin release involves both *intrarenal* (pressor receptor and macula densa) and *extrarenal* (sympathetic nervous system, potassium, angiotensin, etc.) mechanisms. Steady-state renin levels reflect all these factors, with the intrarenal mechanism predominating.

GLUCOCORTICOID PHYSIOLOGY

The division of adrenal steroids into glucocorticoids and mineralocorticoids is arbitrary in that most glucocorticoids have some mineralocorticoid-like properties. The descriptive term *glucocorticoid* is used for adrenal steroids whose predominant action is on intermediary metabolism. Their overall actions are directed at enhancing the production of the high-energy fuel, glucose, and reducing all other metabolic activity not directly involved in that process. Sustained activation, however, results in a pathophysiologic state, e.g., Cushing's syndrome. The principal glucocorticoid is cortisol (hydrocortisone). The effect of glucocorticoids on intermediary metabolism is mediated by the GR. Physiologic effects of glucocorticoids include the regulation of protein, carbohydrate, lipid, and nucleic acid metabolism. Glucocorticoids raise the blood glucose level by acting as an insulin antagonist and by suppressing the secretion of insulin, thereby inhibiting peripheral glucose uptake, which promotes hepatic glucose synthesis (gluconeogenesis) and increases hepatic glycogen content. The actions on protein metabolism are mainly catabolic in effect, resulting in an increase in protein breakdown and nitrogen excretion. In large part, these actions reflect a mobilization of glycogenic amino acid precursors from peripheral supporting structures, such as bone, skin, muscle, and connective tissue, due to protein breakdown and inhibition of protein synthesis and amino acid uptake. Hyperaminoacidemia also facilitates gluconeogenesis by stimulating glucagon secretion. Glucocorticoids act directly on the liver to stimulate the synthesis of certain enzymes, such as tyrosine aminotransferase and tryptophan pyrrolase. Glucocorticoids regulate fatty acid mobilization by enhancing the activation of cellular lipase by lipid-mobilizing hormones (e.g., catecholamines and pituitary peptides).

The actions of cortisol on protein and adipose tissue vary in different parts of the body. For example, pharmacologic doses of cortisol can deplete the protein matrix of the vertebral column (trabecular bone), whereas long bones (which are primarily compact bone) are affected only minimally; similarly, peripheral adipose tissue mass decreases, whereas abdominal and interscapular fat expand.

Glucocorticoids have anti-inflammatory properties, which are probably related to effects on the microvasculature and to suppression of inflammatory cytokines. In this sense, glucocorticoids modulate the immune response via the so-called immune-adrenal axis (Fig. 331-4). This "loop" is one mechanism by which a stress, such as sepsis, increases adrenal hormone secretion, and the elevated cortisol level in turn suppresses the immune response. For example, cortisol maintains vascular responsiveness to

circulating vasoconstrictors and opposes the increase in capillary permeability during acute inflammation. Glucocorticoids cause a leukocytosis due to release from the bone marrow of mature cells as well as to inhibition of their egress through the capillary wall. Glucocorticoids produce a depletion of circulating eosinophils and of lymphoid tissue, specifically T cells, by causing a redistribution from the circulation into other compartments. Thus, cortisol impairs cell-mediated immunity. Glucocorticoids also inhibit the production and action of the mediators of inflammation, such as the lymphokines and prostaglandins. These actions occur via the GR and are blocked by inhibitors of RNA and protein synthesis. Glucocorticoids inhibit the production and action of interferon by T lymphocytes and the production of IL-1 and IL-6 by macrophages. The antipyretic action of glucocorticoids may be explained by the effect on IL-1, which appears to be an endogenous pyrogen (Chap. 17). Glucocorticoids also inhibit the production of T cell growth factor (IL-2) by T lymphocytes. Glucocorticoids reverse macrophage activation and antagonize the action of migration-inhibiting factor (MIF), leading to reduced adherence of macrophages to vascular endothelium. Glucocorticoids inhibit prostaglandin and leukotriene production by inhibiting the activity of phospholipase A$_2$, thus blocking release of arachidonic acid from phospholipids. Finally, glucocorticoids inhibit the production and inflammatory effects of bradykinin, platelet-activating factor, and serotonin. It is probably only at pharmacologic dosages that antibody production is reduced and lysosomal membranes are stabilized, the latter effect suppressing the release of acid hydrolases.

Cortisol levels respond within minutes to stress, whether physical (trauma, surgery, exercise), psychological (anxiety, depression), or physiologic (hypoglycemia, fever). The reasons why elevated glucocorticoid levels protect the organism under stress are not understood, but in conditions of glucocorticoid deficiency, such stresses may cause hypotension, shock, and death. Consequently, in individuals with adrenal insufficiency, glucocorticoid administration should be increased during stress.

Cortisol has major effects on body water. It helps regulate the extracellular fluid volume by retarding the migration of water into cells and by promoting renal water excretion, the latter effect mediated by suppression of vasopressin secretion, by an increase in the rate of glomerular filtration, and by a direct action on the renal tubule. The consequence is to prevent water intoxication by increasing solute-free water clearance. Glucocorticoids also have weak mineralocorticoid-like properties, and high doses promote renal tubular sodium reabsorption and increased urine potassium excretion. Glucocorticoids also can influence behavior; emotional disorders may occur with either an excess or a deficit of cortisol. Last, cortisol suppresses the secretion of pituitary POMC and its derivative peptides (ACTH, b-endorphin, and b-LPT) and the secretion of hypothalamic CRH and vasopressin.

MINERALOCORTICOID PHYSIOLOGY

Mineralocorticoids are major regulators of extracellular fluid volume and the major determinant of potassium metabolism. These effects are mediated by the binding of aldosterone to the MR in target tissues, primarily the kidney. Volume is regulated through a direct effect on the collecting duct, where aldosterone causes an increase in sodium retention and an increase in potassium excretion. The reabsorption of sodium ions causes a fall in the transmembrane potential, thus enhancing the flow of positive

ions, such as potassium, out of the cell into the lumen. The reabsorbed sodium ions are transported out of the tubular epithelium into the renal interstitial fluid and from there into the renal capillary circulation. Water passively follows the transported sodium.

Because the concentration of hydrogen ion is greater in the lumen than in the cell, hydrogen ion is also actively secreted. Mineralocorticoids also act on the epithelium of the salivary ducts, sweat glands, and gastrointestinal tract to cause reabsorption of sodium in exchange for potassium.

When normal individuals are given aldosterone, an initial period of sodium retention is followed by natriuresis, and sodium balance is reestablished after 3 to 5 days. As a result, edema does not develop. This process is referred to as the *escape phenomenon*, signifying an "escape" by the renal tubules from the sodium-retaining action of aldosterone. While renal hemodynamic factors may play a role in the escape, the level of atrial natriuretic peptide also increases. However, it is important to realize that there is no escape from the potassium-losing effects of mineralocorticoids.

There are additional nonclassic effects of mineralocorticoids, primarily on nonepithelial cells. These effects are likely genomic and therefore mediated through activation of the cytosolic MR, but they do not include a modification of sodium-potassium homeostasis. They are probably mediated by mineralocorticoids modifying the expression of several collagen genes and/or genes controlling tissue growth factors, e.g., transforming growth factor b(TGF-b) and plasminogen activator inhibitor (PAI-1). The resultant effects lead to microangiopathy, necrosis (acutely), and fibrosis in a variety of tissues, e.g., heart, kidney, and vasculature. Increased levels of aldosterone are not necessary to produce this damage; rather, an imbalance between the level of aldosterone and the volume and/or sodium balance state appears to be the critical factor.

Three primary mechanisms control aldosterone release -- the renin-angiotensin system, potassium, and ACTH(Table 331-1). The renin-angiotensin system controls extracellular fluid volume via regulation of aldosterone secretion (Fig. 331-5). In effect, the renin-angiotensin system maintains the circulating blood volume constant by causing aldosterone-induced sodium retention during volume deficiency and by decreasing aldosterone-dependent sodium retention when volume is ample.

Potassium ion directly stimulates aldosterone secretion, independent of the circulating renin-angiotensin system, which it suppresses (Fig. 331-5). In addition to potassium's direct effect, it also modifies aldosterone secretion indirectly by activating the local renin-angiotensin system in the zona glomerulosa. This effect can be blocked by the administration ofACEinhibitors that reduce the local production of angiotensin II and thereby reduce the acute aldosterone response to potassium. An increase in serum potassium of as little as 0.1 mmol/L increases plasma aldosterone levels under certain circumstances. Oral potassium loading therefore increases aldosterone secretion, excretion, and plasma levels.

Physiologic amounts ofACTHstimulate aldosterone secretion acutely, but this action is not sustained unless ACTH is administered in a pulsatile fashion. Most studies relegate ACTH to a minor role in the control of aldosterone. For example, subjects receiving high-dose glucocorticoid therapy, and with presumed complete suppression of ACTH,

have normal aldosterone secretion in response to sodium restriction.

Prior dietary intake of both potassium and sodium can alter the magnitude of the aldosterone response to acute stimulation. This effect results from a change in the expression and activity of aldosterone synthase. Increasing potassium intake or decreasing sodium intake sensitizes the response of the glomerulosa cells to acute stimulation by ACTH, angiotensin II, and/or potassium. Thus, regulation of aldosterone secretion occurs both early and late in its synthetic pathway.

Neurotransmitters (dopamine and serotonin) and some peptides, such as atrial natriuretic peptide, g-melanocyte-stimulating hormone (g-MSH), and b-endorphin, also participate in the regulation of aldosterone secretion (Table 331-1). Thus, the control of aldosterone secretion involves both stimulatory and inhibitory factors.

ANDROGEN PHYSIOLOGY

Androgens regulate male secondary sexual characteristics and can cause virilizing symptoms in women (Chap. 53). Adrenal androgens have a minimal effect in males whose sexual characteristics are predominately determined by gonadal steroids (testosterone). In females, however, several androgen-like effects, e.g., sexual hair, are largely mediated by adrenal androgens. The principal adrenal androgens are DHEA, androstenedione, and 11-hydroxyandrostenedione. DHEA and androstenedione are weak androgens and exert their effects via conversion to the potent androgen testosterone in extraglandular tissues. DHEA also has poorly understood effects on the immune and cardiovascular systems. Adrenal androgen formation is regulated by ACTH, not by gonadotropins. It follows that adrenal androgens are suppressed by exogenous glucocorticoid administration.

LABORATORY EVALUATION OF ADRENOCORTICAL FUNCTION

A basic assumption is that measurements of the plasma or urinary level of a given steroid reflects the rate of adrenal *secretion* of that steroid. However, urine *excretion* values may not truly reflect the secretion rate because of improper collection or altered metabolism. Plasma levels reflect the level of secretion only at the time of measurement. The plasma level (*PL*) depends on two factors: the secretion rate (*SR*) of the hormone and the rate at which it is metabolized, i.e., its metabolic clearance rate (*MCR*). These three factors can be related as follows:

BLOOD LEVELS

Peptides The plasma levels of ACTH and angiotensin II can be measured by immunoassay techniques. Basal ACTH secretion shows a circadian rhythm, with lower levels in the early evening than in the morning. However, ACTH is secreted in a pulsatile manner, leading to rapid fluctuations superimposed on this circadian rhythm. Angiotensin II levels also vary diurnally and are influenced by dietary sodium and potassium intakes and posture. Both upright posture and sodium restriction elevate angiotensin II levels.

Most clinical determinations of the renin-angiotensin system, however, involve measurements of peripheral *plasma renin activity* (PRA) in which the renin activity is gauged by the generation of angiotensin I during a standardized incubation period. This method depends on the presence of sufficient angiotensinogen in the plasma as substrate. The generated angiotensin I is measured by radioimmunoassay. The PRA depends on the dietary sodium intake and on whether the patient is ambulatory. In normal humans, the PRA shows a diurnal rhythm characterized by peak values in the morning and decreases in activity in the afternoon. An alternative approach is to measure plasma active renin, which is easier and not dependent on endogenous substrate concentration. PRA and active renin correlate very well on low-sodium diets (r = 0.85 to 0.9) but less well on high-sodium diets.

Steroids Cortisol and aldosterone are both secreted episodically, and levels generally vary during the day, with peak values in the morning and low levels in the evening. In addition, the plasma level of aldosterone, but not of cortisol, is increased by dietary potassium loading, by sodium restriction, or by assumption of the upright posture. Measurement of the sulfate conjugate of DHEA may be a useful index of adrenal androgen secretion, as little DHEA sulfate is formed in the gonads and because the half-life of DHEA sulfate is 7 to 9 h. However, DHEA sulfate levels reflect both DHEA production and sulfatase activity.

URINE LEVELS

For the assessment of glucocorticoid secretion, the urine *17-hydroxycorticosteroid* assay has been replaced by measurement of urinary free cortisol. Elevated levels of urinary free cortisol correlate with states of hypercortisolism, reflecting changes in the levels of unbound, physiologically active circulating cortisol. Normally, the rate of excretion is higher in the daytime (7 A.M. to 7 P.M.) than at night (7 P.M. to 7 A.M.).

Urinary *17-ketosteroids* originate in either the adrenal gland or the gonad. In normal women, 90% of urinary 17-ketosteroids is derived from the adrenal, and in men 60 to 70% is of adrenal origin. Urine 17-ketosteroid values are highest in young adults and decline with age.

A carefully timed urine collection is a prerequisite for all excretory determinations. Urinary creatinine should be measured simultaneously to determine the accuracy and adequacy of the collection procedure.

STIMULATION TESTS

Stimulation tests are useful in the diagnosis of hormone deficiency states.

Tests of Glucocorticoid Reserve Within minutes after administration of ACTH, cortisol levels increase. This responsiveness can be used as an index of the functional reserve of the adrenal gland for production of cortisol. Under maximal ACTH stimulation, cortisol secretion increases tenfold, to 800 umol/d (300 mg/d), but maximal stimulation can be achieved only with prolonged ACTH infusions.

A screening test (the so-called rapid ACTH stimulation test) involves the administration of 25 units (0.25 mg) of cosyntropin intravenously or intramuscularly and measurement of plasma cortisol levels before and 30 and 60 min after administration; the test can be performed at any time of the day. The most clear-cut criterion for a normal response is a stimulated cortisol level of >500 nmol/L (>18 ug/dL), and the minimal stimulated normal increment of cortisol is >200 nmol/L (>7 ug/dL) above baseline. Severely ill patients with elevated basal cortisol levels may show no further increases following acute ACTH administration.

Tests of Mineralocorticoid Reserve and Stimulation of the Renin-Angiotensin System Stimulation tests use protocols designed to create a programmed volume depletion, such as sodium restriction, diuretic administration, or upright posture. A simple, potent test consists of severe sodium restriction and upright posture. After 3 to 5 days of a 10-mmol/d sodium intake, rates of aldosterone secretion or excretion should increase two- to threefold over the control values. Supine morning plasma aldosterone levels are usually increased three- to sixfold, and they increase a further two- to fourfold in response to 2 to 3 h of upright posture.

When the dietary sodium intake is normal, stimulation testing requires the administration of a potent diuretic, such as 40 to 80 mg furosemide, followed by 2 to 3 h of upright posture. The normal response is a two- to fourfold rise in plasma aldosterone levels.

SUPPRESSION TESTS

Suppression tests to document hypersecretion of adrenal hormones involve measurement of the target hormone response after standardized suppression of its tropic hormone.

Tests of Pituitary-Adrenal Suppressibility The ACTH release mechanism is sensitive to the circulating glucocorticoid level. When blood levels of glucocorticoid are increased in normal individuals, less ACTH is released from the anterior pituitary and less steroid is produced by the adrenal gland. The integrity of this feedback mechanism can be tested clinically by giving a glucocorticoid and judging the suppression of ACTH secretion by analysis of urine steroid levels and/or plasma cortisol and ACTH levels. A potent glucocorticoid such as dexamethasone is used, so that the agent can be given in an amount small enough not to contribute significantly to the pool of steroids to be analyzed.

The best *screening* procedure is the overnight dexamethasone suppression test. This involves the measurement of plasma cortisol levels at 8 A.M. following the oral administration of 1 mg dexamethasone the previous midnight. The 8 A.M. value for plasma cortisol in normal individuals should be<140 nmol/L (5 ug/dL).

The definitive test of adrenal suppressibility consists in administering 0.5 mg dexamethasone every 6 h for two successive days while collecting urine over a 24-h period for determination of creatinine and free cortisol and/or measuring plasma cortisol levels. In a patient with a normal hypothalamic-pituitary ACTH release mechanism, a fall in the urine free cortisol to<80 nmol/d (30 ug/d) or of plasma cortisol to<140 nmol/L (5 ug/dL) is seen on the second day of administration.

A normal response to either suppression test implies that the glucocorticoid regulation of ACTH and its control of the adrenal glands is physiologically normal. However, an isolated abnormal result, particularly to the overnight suppression test, does not in itself imply pituitary and/or adrenal disease.

Tests of Mineralocorticoid Suppressibility These tests rely on an expansion of extracellular fluid volume, which should decrease circulating plasma renin activity and decrease the secretion and/or excretion of aldosterone. Various tests differ in the rate at which extracellular fluid volume is expanded. One convenient suppression test involves the intravenous infusion of 500 mL/h of normal saline solution for 4 h, which normally suppresses plasma aldosterone levels to <220 pmol/L (<8 ng/dL) on a sodium-restricted diet or to <140 pmol/L (<5 ng/dL) on a normal sodium intake. Alternatively, a high-sodium diet can be administered for 3 days with 0.2 mg fludrocortisone twice daily. Aldosterone excretion is measured on the third day and should be <28 nmol/d (10 ug/d). These tests should not be performed in potassium-depleted individuals since they carry a risk of precipitating hypokalemia.

TESTS OF PITUITARY-ADRENAL RESPONSIVENESS

Stimuli such as insulin-induced hypoglycemia, AVP, and pyrogens cause the release of ACTH from the pituitary by an action on higher neural centers or on the pituitary itself. Insulin-induced hypoglycemia is particularly useful, because it stimulates the release of both growth hormone and ACTH. In this test, regular insulin (0.05 to 0.1 U/kg body weight) is given intravenously as a bolus to reduce the fasting glucose level to at least 50% below basal. The normal cortisol response is a rise to more than 500 nmol/L (18 ug/dL).

One of the best ways to test the integrity of the pituitary-adrenal axis is the metyrapone test. Metyrapone inhibits 11b-hydroxylase in the adrenal. As a result, the conversion of 11-deoxycortisol (compound S) to cortisol is impaired, causing 11-deoxycortisol to accumulate in the blood and the blood level of cortisol to decrease (Fig. 331-2). The hypothalamic-pituitary axis responds to the declining cortisol blood levels by releasing more ACTH. Note that assessment of the response depends on both an intact hypothalamic-pituitary axis and an intact adrenal gland.

Although modifications of the original metyrapone test have been described, we believe the best involves administering 750 mg of the drug by mouth every 4 h over a 24-h period and comparing the control and postmetyrapone plasma levels of 11-deoxycortisol, cortisol, and ACTH. In normal individuals, plasma 11-deoxycortisol levels should exceed 210 nmol/L (7 ug/dL) and ACTH levels should exceed 17 pmol/L (75 pg/mL) following metyrapone administration. The metyrapone test does not accurately reflect ACTH reserve if subjects are ingesting exogenous glucocorticoids or drugs that accelerate the metabolism of metyrapone (e.g., phenytoin).

A direct and selective test of the pituitary corticotrophs can be achieved with CRH. The bolus injection of ovine CRH (corticorelin, ovine triflutate; 1 ug/kg body weight) stimulates secretion of ACTH and b-LPT in normal human subjects within 15 to 60 min. In normal individuals, the mean increment in ACTH is 9 pmol/L (40 pg/mL). However, the

magnitude of the ACTH response is less than that produced by the insulin tolerance test, which implies that additional factors (such as vasopressin) augment stress-induced increases in ACTH secretion.

Although the rapidACTHstimulation test is useful for the diagnosis of primary adrenal insufficiency, normal cortisol responsiveness may be seen in some patients with a partial ACTH deficit and absence of adrenal atrophy. These patients have an inadequate pituitary ACTH reserve and fail to increase ACTH secretion in response to a stress such as surgery or hypoglycemia. Because the use of a bolus of exogenous ACTH does not invariably exclude a diagnosis of secondary adrenocortical insufficiency, direct tests of pituitary ACTH reserve (metyrapone test, insulin tolerance testing) may be required in the appropriate clinical setting. Alternatively, ACTH at a physiologic dose (1 ug), the so-called low-dose ACTH test, may be used. Abnormal response is similar to the rapid ACTH test. However, levels need to be measured at 30 min, and the ACTH needs to be injected directly intravenously because it can be absorbed to plastic tubing. On the other hand, the rapid ACTH test can distinguish between primary and secondary adrenal insufficiency, because aldosterone secretion is preserved in secondary adrenal failure by the renin-angiotensin system and potassium. Cosyntropin (25 units) is given intravenously or intramuscularly, and plasma cortisol and aldosterone levels are measured before and 30 and 60 min after administration. The cortisol response is abnormal in both groups, but patients with secondary insufficiency show an increase in aldosterone levels by at least 140 pmol/L (5 ng/dL). No aldosterone response is seen in patients in whom the adrenal cortex is destroyed.

HYPERFUNCTION OF THE ADRENAL CORTEX

Excess cortisol is associated with Cushing's syndrome; excess aldosterone causes aldosteronism; and excess adrenal androgens cause adrenal virilism. These syndromes do not always occur in the "pure" form but may have overlapping features.

CUSHING'S SYNDROME

Etiology Cushing described a syndrome characterized by truncal obesity, hypertension, fatigability and weakness, amenorrhea, hirsutism, purplish abdominal striae, edema, glucosuria, osteoporosis, and a basophilic tumor of the pituitary. As awareness of this syndrome has increased, the diagnosis of Cushing's syndrome has been broadened into the classification shown in Table 331-2. Regardless of etiology, all cases of endogenous Cushing's syndrome are due to increased production of cortisol by the adrenal. In most cases the cause is *bilateral adrenal hyperplasia* due to hypersecretion of pituitaryACTH or ectopic production of ACTH by a nonpituitary source. The incidence of pituitary-dependent adrenal hyperplasia is three times greater in women than in men, and the most frequent age of onset is the third or fourth decade. Most evidence indicates that the primary defect is the de novo development of a pituitary adenoma, as tumors are found in >90% of patients with pituitary-dependent adrenal hyperplasia. Alternatively, the defect may occasionally reside in the hypothalamus or in higher neural centers, leading to release ofCRHinappropriate to the level of circulating cortisol. The consequence would be that a higher level of cortisol is required to reduce ACTH secretion to normal. This primary defect leads to hyperstimulation of the pituitary, resulting in hyperplasia or tumor formation. In surgical series, most individuals with

hypersecretion of pituitary ACTH are found to have a microadenoma (<10 mm in diameter; 50% are£5 mm in diameter), but a pituitary macroadenoma (>10 mm) or diffuse hyperplasia of the corticotropic cells may be found. In some studies, the recurrence rate is>20%. Unfortunately, it may be difficult to distinguish between recurrence and inadequate primary therapy. Traditionally, only an individual who has an ACTH-producing pituitary tumor is defined as having *Cushing's disease*.

Nonpituitary tumors may secrete polypeptides that are biologically, chemically, and immunologically indistinguishable from eitherACTH orCRH and that cause bilateral adrenal hyperplasia (Chap. 100). The ectopic production of CRH results in clinical, biochemical, and radiologic features indistinguishable from those caused by hypersecretion of pituitary ACTH. The typical signs and symptoms of Cushing's syndrome may be absent or minimal with ectopic ACTH production, and hypokalemic alkalosis is a prominent manifestation. Most of these cases are associated with the primitive small cell (oat cell) type of bronchogenic carcinoma or with tumors of the thymus, pancreas, or ovary; medullary carcinoma of the thyroid; or bronchial adenomas. The onset of Cushing's syndrome may be sudden, particularly in patients with carcinoma of the lung, and this feature accounts in part for the failure of these patients to exhibit the classic manifestations. On the other hand, patients with carcinoid tumors or pheochromocytomas have longer clinical courses and usually exhibit the typical cushingoid features. The ectopic secretion of ACTH is also accompanied by the accumulation of ACTH fragments in plasma and by elevated plasma levels of ACTH precursor molecules. Because such tumors may produce large amounts of ACTH, baseline steroid values are usually markedly elevated, and increased skin pigmentation may be present. Indeed, hyperpigmentation in patients with Cushing's syndrome almost always points to an extraadrenal tumor, either in an extracranial location or within the cranium.

Approximately 20 to 25% of patients with Cushing's syndrome have an adrenal neoplasm. These tumors are usually unilateral, and about half are malignant. Occasionally, patients have biochemical features both of pituitaryACTHexcess and of an adrenal adenoma. These individuals usually have *nodular hyperplasia* of both adrenal glands often the result of prolonged ACTH stimulation in the absence of a pituitary adenoma. Two additional entities cause nodular hyperplasia: a familial disorder in children or young adults (so-called pigmented micronodular dysplasia; see below) and an abnormal cortisol response to gastric inhibitory polypeptide or luteinizing hormone, probably secondary to expression of receptors for these hormones in the adrenal cortex.

The most common cause of Cushing's syndrome is *iatrogenic* administration of steroids for a variety of reasons. Although the clinical features bear some resemblance to those seen with adrenal tumors, these patients are usually distinguishable on the basis of history and laboratory studies.

Clinical Signs, Symptoms, and Laboratory Findings Many of the signs and symptoms of Cushing's syndrome follow logically from the known action of glucocorticoids (Table 331-3). Mobilization of peripheral supportive tissue causes muscle weakness and fatigability, osteoporosis, broad violaceous cutaneous striae, and easy bruisability. The latter signs are secondary to weakening and rupture of collagen

fibers in the dermis. Osteoporosis may cause collapse of vertebral bodies and pathologic fractures of other bones. Decreased bone mineralization is particularly pronounced in children. Increased hepatic gluconeogenesis and insulin resistance can cause impaired glucose tolerance. Overt diabetes mellitus occurs in <20% of patients, who probably are individuals with a predisposition to this disorder. Hypercortisolism promotes the deposition of adipose tissue in characteristic sites, notably the upper face (producing the typical "moon" facies), the interscapular area (producing the "buffalo hump"), and the mesenteric bed (producing "truncal" obesity) (Fig. 331-6). Rarely, episternal fatty tumors and mediastinal widening secondary to fat accumulation occur. The reason for this peculiar distribution of adipose tissue is not known, but it is associated with insulin resistance and/or elevated insulin levels. The face appears plethoric, even in the absence of any increase in red blood cell concentration (Fig. 331-CD1). Hypertension is common, and emotional changes may be profound, ranging from irritability and emotional lability to severe depression, confusion, or even frank psychosis. In women, increased levels of adrenal androgens can cause acne, hirsutism, and oligomenorrhea or amenorrhea. Some signs and symptoms in patients with hypercortisolism -- i.e., obesity, hypertension, osteoporosis, and diabetes -- are nonspecific and therefore are less helpful in diagnosing the condition. On the other hand, easy bruising, typical striae, myopathy, and virilizing signs (although less frequent) are, if present, more suggestive of Cushing's syndrome (Table 331-3).

Except in iatrogenic Cushing's syndrome, plasma and urine cortisol levels are variably elevated. Occasionally, hypokalemia, hypochloremia, and metabolic alkalosis are present, particularly with ectopic production of ACTH.

Diagnosis The diagnosis of Cushing's syndrome depends on the demonstration of increased cortisol production and failure to suppress cortisol secretion normally when dexamethasone is administered (Chap. 328). Once the diagnosis is established, further testing is designed to determine the etiology (Fig. 331-7 and Table 331-4).

For initial screening, the overnight dexamethasone suppression test is recommended (see above). In difficult cases (e.g., in obese patients), measurement of a 24-h urine free cortisol also can be used as a screening test. A level>140 nmol/d (50 ug/d) is suggestive of Cushing's syndrome. The definitive diagnosis is then established by failure of urinary cortisol to fall to less than <25 nmol/d (10 ug/d) or of plasma cortisol to fall to <140 nmol/L (5 ug/dL) after a standard low-dose dexamethasone suppression test (0.5 mg every 6 h for 48 h). Owing to circadian variability, plasma cortisol and, to a certain extent, ACTH determinations are not meaningful when performed in isolation, but the absence of the normal fall of plasma cortisol at midnight is consistent with Cushing's syndrome.

The task of determining the etiology of Cushing's syndrome is complicated by the fact that all the available tests lack specificity and by the fact that the tumors producing this syndrome are prone to spontaneous and often dramatic changes in hormone secretion (periodic hormonogenesis). No test has a specificity >95%, and it may be necessary to use a combination of tests to arrive at the correct diagnosis. A useful step to distinguish patients with an ACTH-secreting pituitary microadenoma or hypothalamic-pituitary dysfunction from those with other forms of Cushing's syndrome is to determine the response of cortisol output to administration of high-dose dexamethasone (2 mg every 6

h for 2 days). An alternative 8-mg, overnight high-dose dexamethasone test has been developed; however, this test has a lower sensitivity and specificity than the standard test. When the diagnosis of Cushing's syndrome is clear-cut on the basis of baseline urinary and plasma assays, the high-dose dexamethasone suppression test may be used without performing the preliminary low-dose suppression test. The high-dose suppression test provides close to 100% specificity if the criterion used is suppression of urinary free cortisol by>90%. Occasionally, in individuals with bilateral nodular hyperplasia and/or ectopicCRHproduction, steroid output is also suppressed. Failure of low- and high-dose dexamethasone administration to suppress cortisol production (Table 331-4) is usual in patients with adrenal hyperplasia secondary to an ACTH-secreting pituitary macroadenoma or an ACTH-producing tumor of nonendocrine origin and in those with adrenal neoplasms.

PlasmaACTHlevels can be useful in distinguishing the various causes of Cushing's syndrome, particularly in separating ACTH-dependent from ACTH-independent causes. In general, measurement of plasma ACTH is useful in the diagnosis of ACTH-independent etiologies of the syndrome, since most adrenal tumors cause low or undetectable ACTH levels [<2 pmol/L (10 pg/mL)]. Furthermore, ACTH-secreting pituitary macroadenomas and ACTH-producing nonendocrine tumors usually result in elevated ACTH levels. In the ectopic ACTH syndrome, ACTH levels may be elevated to >110 pmol/L (500 pg/mL), and in most patients the level is>40 pmol/L (200 pg/mL). In Cushing's syndrome as the result of a microadenoma or pituitary-hypothalamic dysfunction, ACTH levels range from 6 to 30 pmol/L (30 to 150 pg/mL) [normal,<14 pmol/L (<60 pg/mL)], with half of values falling in the normal range. However, the main problem with the use of ACTH levels in the differential diagnosis of Cushing's syndrome is that ACTH levels may be similar in individuals with hypothalamic-pituitary dysfunction, pituitary microadenomas, ectopicCRHproduction, and ectopic ACTH production (especially carcinoid tumors) (Table 331-4).

Because of these difficulties, several additional tests have been advocated, such as the metyrapone andCRHinfusion tests. The rationale underlying these tests is that steroid hypersecretion by an adrenal tumor or the ectopic production ofACTHwill suppress the hypothalamic-pituitary axis so that inhibition of pituitary ACTH release can be demonstrated by either test. Thus, most patients with pituitary-hypothalamic dysfunction and/or a microadenoma have an increase in steroid or ACTH secretion in response to metyrapone or CRH administration, whereas most patients with ectopic ACTH-producing tumors do not. Most pituitary macroadenomas also respond to CRH, but their response to metyrapone is variable. However, false-positive and -negative CRH tests can occur in patients with ectopic ACTH and pituitary tumors.

The main diagnostic dilemma in Cushing's syndrome is to distinguish those instances due to microadenomas of the pituitary and/or pituitary-hypothalamic dysfunction from those due to tumors (e.g., carcinoids or pheochromocytoma) that produceCRHand/orACTHectopically. The clinical manifestations are similar unless the ectopic tumor produces other symptoms, such as diarrhea and flushing from a carcinoid tumor or episodic hypertension from a pheochromocytoma. Sometimes, one can distinguish between ectopic and pituitary ACTH production by using metyrapone or CRH tests, as noted above. In these situations, computed tomography (CT) of the pituitary gland is usually normal. Magnetic resonance imaging (MRI) with the enhancing

agent gadolinium may be better than CT for this purpose but demonstrates pituitary microadenomas in only half of patients with Cushing's disease. In subjects with negative imaging studies, selective petrosal sinus venous sampling for ACTH is employed in some centers. Demonstration of an ACTH gradient between the petrosal sinus and peripheral blood localizes the source of ACTH overproduction to the pituitary gland but does not distinguish pituitary-dependent adrenal hyperplasia from pituitary hyperplasia secondary to a tumor producing CRH. CRH levels should be measured in the peripheral blood prior to petrosal sinus sampling. In centers where petrosal sinus sampling is performed frequently, it has proved useful for distinguishing pituitary and nonpituitary sources of ACTH excess. However, the catheterization procedure is technically difficult, and complications have occurred.

The diagnosis of a *cortisol-producing adrenal adenoma* is suggested by disproportionate elevations in baseline urine free-cortisol levels with only modest changes in urinary 17-ketosteroids or plasmaDHEAsulfate. Adrenal androgen secretion is usually reduced in these patients owing to the cortisol-induced suppression ofACTH and subsequent involution of the androgen-producing zona reticularis.

The diagnosis of *adrenal carcinoma* is suggested by a palpable abdominal mass and by *markedly* elevated baseline values of *both* urine 17-ketosteroids and plasmaDHEAsulfate. Plasma and urine cortisol levels are variably elevated. Adrenal carcinoma is usually resistant to bothACTHstimulation and dexamethasone suppression. Elevated adrenal androgen secretion often leads to virilization in the female. Estrogen-producing adrenocortical carcinoma usually presents with gynecomastia in men and dysfunctional uterine bleeding in women. These adrenal tumors secrete increased amounts of androstenedione, which is converted peripherally to the estrogens estrone and estradiol (Chap. 337). Adrenal carcinomas that produce Cushing's syndrome are often associated with elevated levels of the intermediates of steroid biosynthesis (especially 11-deoxycortisol), suggesting inefficient conversion of the intermediates to the final product. This feature also accounts for the characteristic increase in 17-ketosteroids. Approximately 20% of adrenal carcinomas are not associated with endocrine syndromes and are presumed to be nonfunctioning or to produce biologically inactive steroid precursors. In addition, the excessive production of steroids is not always clinically evident (e.g., androgens in adult men).

Differential Diagnosis

Pseudo-Cushing's Syndrome Problems in diagnosis include patients with obesity, chronic alcoholism, depression, and acute illness of any type. Extreme *obesity* is uncommon in Cushing's syndrome; furthermore, with exogenous obesity, the adiposity is generalized, not truncal. On adrenocortical testing, abnormalities in patients with exogenous obesity are usually modest. Basal urine steroid excretion levels in obese patients are also either normal or slightly elevated. Some patients have elevated conversion of secreted cortisol into excreted metabolites. Urinary and blood cortisol levels are usually normal, and the diurnal pattern in blood and urine levels is normal. Patients with *chronic alcoholism* and those with *depression* share similar abnormalities in steroid output: modestly elevated urine cortisol, blunted circadian rhythm of cortisol levels, and resistance to suppression using the overnight dexamethasone test. In contrast to alcoholic subjects, depressed patients do not have signs and symptoms of

Cushing's syndrome. Following discontinuation of alcohol and/or improvement in the emotional status, results of steroid testing usually return to normal. One or more of three tests have been used to differentiate mild Cushing's syndrome and pseudo-Cushing's syndrome. The serum cortisol level following the standard 2-day low-dose dexamethasone test has very high sensitivity and specificity. While the CRH test alone is less useful, in combination with the low-dose dexamethasone test, there is nearly complete discrimination between these two conditions. Finally, a midnight cortisol level obtained in awake patients may have similar predictive value as the low-dose dexamethasone test if a cut-off of 210 nmol/L (7.5 ug/dL) is used. Patients with *acute illness* often have abnormal results on laboratory tests and fail to exhibit pituitary-adrenal suppression in response to dexamethasone, since major stress (such as pain or fever) interrupts the normal regulation of ACTH secretion. *Iatrogenic Cushing's syndrome*, induced by the administration of glucocorticoids or other steroids such as megestrol that bind to the glucocorticoid receptor, is indistinguishable by physical findings from endogenous adrenocortical hyperfunction. The distinction can be made, however, by measuring blood or urine cortisol levels in a basal state; in the iatrogenic syndrome these levels are low secondary to suppression of the pituitary-adrenal axis. The severity of iatrogenic Cushing's syndrome is related to the total steroid dose, the biologic half-life of the steroid, and the duration of therapy. Also, individuals taking afternoon and evening doses of glucocorticoids develop Cushing's syndrome more readily and with a smaller total daily dose than do patients taking morning doses only. The enzymatic disposition and binding of administered steroids differ among patients.

Radiologic Evaluation for Cushing's Syndrome The preferred radiologic study for visualizing the adrenals is a CT scan of the abdomen (Fig. 331-8). CT is of value both for localizing adrenal tumors and for diagnosing bilateral hyperplasia. All patients believed to have hypersecretion of pituitary ACTH should have a pituitary MRI scan with the contrast agent gadolinium. Even with this technique, small microadenomas may be undetectable; alternatively, false-positive masses due to cysts or nonsecretory lesions of the normal pituitary may be imaged. In patients with ectopic ACTH production, chest CT is a useful first step.

Evaluation of Asymptomatic Adrenal Masses With abdominal CT scanning, many incidental adrenal masses (so-called incidentalomas) are discovered. This is not surprising, since 10 to 20% of subjects at autopsy have adrenocortical adenomas. The first step in evaluating such patients is to determine whether the tumor is functioning by means of appropriate screening tests, e.g., measurement of 24-h urine catecholamines and metabolites and serum potassium and assessment of adrenal cortical function by dexamethasone-suppression testing. However, 90% of incidentalomas are nonfunctioning. If an extraadrenal malignancy is present, there is a 30 to 50% chance that the adrenal tumor is a metastasis. If the primary tumor is being treated and there are no other metastases, it is prudent to obtain a fine-needle aspirate of the adrenal mass to establish the diagnosis. In the absence of a known malignancy the next step is unclear. The probability of adrenal carcinoma is <0.01 percent, the vast majority of adrenal masses being benign adenomas. Features suggestive of malignancy include large size (a size >4 to 6 cm suggests carcinoma); irregular margins; and inhomogeneity, soft tissue calcifications visible on CT (Fig. 331-8), and findings characteristic of malignancy on a chemical-shift MRI image. If surgery is not performed, a repeat CT scan should be obtained in 3 to 6 months. Fine needle aspiration is not useful

to distinguish between benign and malignant primary adrenal tumors.

TREATMENT

Adrenal Neoplasm When an adenoma or carcinoma is diagnosed, adrenal exploration is performed with excision of the tumor. Adenomas may be resected using laparoscopic techniques. Because of the possibility of atrophy of the contralateral adrenal, the patient is treated pre- and postoperatively as if for total adrenalectomy, even when a unilateral lesion is suspected, the routine being similar to that for an Addisonian patient undergoing elective surgery (seeTable 331-8).

Despite operative intervention, most patients with adrenal carcinoma die within 3 years of diagnosis. Metastases occur most often to liver and lung. The principal drug for the treatment of adrenocortical carcinoma is mitotane (*o,p¢*-DDD), an isomer of the insecticide DDT. This drug suppresses cortisol production and decreases plasma and urine steroid levels. Although its cytotoxic action is relatively selective for the glucocorticoid-secreting zone of the adrenal cortex, the zona glomerulosa may also be inhibited. Because mitotane also alters the extraadrenal metabolism of cortisol, plasma and urinary cortisol levels must be assessed to titrate the effect. The drug is usually given in divided doses three to four times a day, with the dose increased gradually to tolerability (usually<6 g daily). At higher doses, almost all patients experience side effects, which may be gastrointestinal (anorexia, diarrhea, vomiting) or neuromuscular (lethargy, somnolence, dizziness). All patients treated with mitotane should receive long-term glucocorticoid maintenance therapy, and, in some, mineralocorticoid replacement is appropriate. In approximately one-third of patients, both tumor and metastases regress, but long-term survival is not altered. In many patients, mitotane only inhibits steroidogenesis and does not cause regression of tumor metastases. Osseous metastases are usually refractory to the drug and should be treated with radiation therapy. Mitotane can also be given as adjunctive therapy after surgical resection of an adrenal carcinoma, although there is no evidence that this improves survival. Because of the absence of a long-term benefit with mitotane, alternative chemotherapeutic approaches based on platinum therapy have been used. However, there are no data presently available indicating a prolongation of life.

Bilateral Hyperplasia Patients with hyperplasia have a relative or absolute increase inACTHlevels. Since therapy would logically be directed at reducing ACTH levels, the ideal primary treatment for ACTH- orCRH-producing tumors, whether pituitary or ectopic, is surgical removal. Occasionally (particularly with ectopic ACTH production) surgical excision is not possible because the disease is far advanced. In this situation, "medical" or surgical adrenalectomy may correct the hypercortisolism.

Controversy exists as to the proper treatment for bilateral adrenal hyperplasia when the source of theACTHoverproduction is not apparent. In some centers, these patients (especially those who suppress after the administration of a high-dose dexamethasone test) undergo surgical exploration of the pituitary via a transsphenoidal approach in the expectation that a microadenoma will be found. However, in most circumstances selective petrosal sinus venous sampling is recommended, and the patient is referred to an appropriate center if the procedure is not available locally. If a microadenoma is not found at the time of exploration, total hypophysectomy may be needed. Complications

of transsphenoidal surgery include cerebrospinal fluid rhinorrhea, diabetes insipidus, panhypopituitarism, and optic or cranial nerve injuries.

In other centers, total adrenalectomy is the treatment of choice. The cure rate with this procedure is close to 100%. The adverse effects include the certain need for lifelong mineralocorticoid and glucocorticoid replacement and a 10 to 20% probability of a pituitary tumor developing over the next 10 years (Nelson's syndrome; Chap. 328). Many of these tumors require surgical therapy. It is uncertain whether they arise de novo in these patients or were present prior to adrenalectomy but were too small to be detected. Periodic radiologic evaluation of the pituitary gland by MRI as well as serial ACTH measurements should be performed in all individuals after bilateral adrenalectomy for Cushing's disease. Such pituitary tumors may become locally invasive and impinge on the optic chiasm or extend into the cavernous or sphenoid sinuses.

Except in children, pituitary irradiation is rarely used as primary treatment, being reserved rather for postoperative tumor recurrences. In some centers, high levels of gamma radiation can be focused on the desired site with less scattering to surrounding tissues by using stereotactic techniques. Side effects of radiation include ocular motor palsy and hypopituitarism. There is a long lag time between treatment and remission, and the remission rate is usually <50%.

Finally, in occasional patients in whom a surgical approach is not feasible, "medical" adrenalectomy may be indicated (Table 331-5). Inhibition of steroidogenesis may also be indicated in severely cushingoid subjects prior to surgical intervention. Chemical adrenalectomy may be accomplished by the administration of the inhibitor of steroidogenesis ketoconazole (600 to 1200 mg/d). In addition, mitotane (2 or 3 g/d) and/or the blockers of steroid synthesis aminoglutethimide (1 g/d) and metyrapone (2 or 3 g/d) may be effective either alone or in combination. Mitotane is slow to take effect (weeks). Mifepristone, a competitive inhibitor of the binding of glucocorticoid to its receptor, may be a treatment option. Adrenal insufficiency is a risk with all these agents, and replacement steroids may be required.

ALDOSTERONISM

Aldosteronism is a syndrome associated with hypersecretion of the mineralocorticoid aldosterone. In *primary* aldosteronism the cause for the excessive aldosterone production resides within the adrenal gland; in *secondary* aldosteronism the stimulus is extraadrenal.

Primary Aldosteronism In the original case of excessive and inappropriate aldosterone production, the disease was the result of an *aldosterone-producing adrenal adenoma* (Conn's syndrome). Most cases involve a unilateral adenoma, which is usually small and may occur on either side. Rarely, primary aldosteronism is due to an adrenal carcinoma. Aldosteronism is twice as common in women as in men, usually occurs between the ages of 30 and 50, and is present in approximately 1% of unselected hypertensive patients. However, the prevalence may be as high as 10%, depending on the criteria and study population. Most of this difference is not secondary to the prevalence of patients with an aldosteronoma but rather because of the inclusion of

those with bilateral hyperplasia. In many patients with clinical and biochemical features of primary aldosteronism, a solitary adenoma is not found at surgery. Instead, these patients have *bilateral cortical nodular hyperplasia*. In the literature, this disease is also termed *idiopathic hyperaldosteronism*, and/or *nodular hyperplasia*. The cause is unknown.

Signs and Symptoms Hypersecretion of aldosterone increases the renal distal tubular exchange of intratubular sodium for secreted potassium and hydrogen ions, with progressive depletion of body potassium and development of hypokalemia. Most patients have diastolic hypertension, which may be very severe, and headaches. The hypertension is probably due to the increased sodium reabsorption and extracellular volume expansion. *Potassium depletion* is responsible for the muscle weakness and fatigue and is due to the effect of potassium depletion on the muscle cell membrane. The polyuria results from impairment of urinary concentrating ability and is often associated with polydipsia.

Electrocardiographic and roentgenographic signs of left ventricular enlargement are, in part, secondary to the hypertension. However, the left ventricular hypertrophy is disproportionate to the level of blood pressure when compared to individuals with essential hypertension, and regression of the hypertrophy occurs even if blood pressure is not reduced after removal of an aldosteronoma. Electrocardiographic signs of potassium depletion include prominent U waves, cardiac arrhythmias, and premature contractions. In the absence of associated congestive heart failure, renal disease, or preexisting abnormalities (such as thrombophlebitis), edema is characteristically absent. However, structural damage to the cerebral circulation, retinal vasculature, and kidney occurs more frequently than would be predicted based on the level and duration of the hypertension. Proteinuria may occur in as many as 50% of patients with primary aldosteronism, and renal failure occurs in up to 15%. Thus, it is probable that excess aldosterone production induces cardiovascular damage independent of its effect on blood pressure.

Laboratory Findings Laboratory findings depend on both the duration and the severity of potassium depletion. An overnight concentration test often reveals impaired ability to concentrate the urine, probably secondary to the hypokalemia. Urine pH is neutral to alkaline because of excessive secretion of ammonium and bicarbonate ions to compensate for the metabolic alkalosis.

Hypokalemia may be severe (<3 mmol/L) and reflects body potassium depletion, usually >300 mmol. In mild forms of primary aldosteronism, potassium levels may be normal. *Hypernatremia* is due to sodium retention, a concomitant water loss from polyuria, and a resetting of the osmostat. Metabolic alkalosis and elevation of serum bicarbonate are a result of hydrogen ion loss into the urine and migration into potassium-depleted cells. The alkalosis is perpetuated by potassium deficiency, which increases the capacity of the proximal convoluted tubule to reabsorb filtered bicarbonate. If hypokalemia is severe, serum magnesium levels are also reduced.

Diagnosis The diagnosis is suggested by persistent hypokalemia in a nonedematous patient with a normal sodium intake who is not receiving potassium-wasting diuretics (furosemide, ethacrynic acid, thiazides). If hypokalemia occurs in a hypertensive patient

taking a potassium-wasting diuretic, the diuretic should be discontinued and the patient should be given potassium supplements. After 1 to 2 weeks, the potassium level should be remeasured, and if hypokalemia persists, the patient should be evaluated for a mineralocorticoid excess syndrome (Fig. 331-9).

The criteria for the diagnosis of primary aldosteronism are (1) diastolic hypertension without edema, (2) hyposecretion of renin (as judged by low plasma renin activity levels) that fails to increase appropriately during volume depletion (upright posture, sodium depletion), and (3) hypersecretion of aldosterone that does not suppress appropriately in response to volume expansion.

Patients with primary aldosteronism characteristically *do not have edema*, since they exhibit an "escape" phenomenon from the sodium-retaining aspects of mineralocorticoids. Rarely, pretibial edema is present in patients with associated nephropathy and azotemia.

The estimation of plasma renin activity is of limited value in separating patients with primary aldosteronism from those with hypertension of other causes. Although failure of plasma renin activity to rise normally during volume-depletion maneuvers is a criterion for a diagnosis of primary aldosteronism, suppressed renin activity also occurs in about 25% of patients with essential hypertension.

Although a renin measurement alone lacks specificity, the ratio of serum aldosterone to plasma renin activity is a very useful screening test. A high ratio (>30), when aldosterone is expressed as ng/dL and plasma renin activity as ng/mL per hour, strongly suggests autonomy of aldosterone secretion. Aldosterone levels need to be>500 pmol/L (>15 ng/dL) and the salt intake not be restricted in making this assessment. Ultimately, it is necessary to demonstrate a lack of aldosterone suppression to diagnose primary aldosteronism (Fig. 331-9). The autonomy exhibited in these patients refers only to the resistance to suppression of secretion during volume expansion; aldosterone can and does respond in a normal or above-normal fashion to the stimulus of potassium loading orACTHinfusion.

Once hyposecretion of renin and failure of aldosterone secretion suppression are demonstrated, aldosterone-producing adenomas should be localized by abdominalCT scan, using a high-resolution scanner as many aldosteronomas are<1 cm in size. If the CT scan is negative, percutaneous transfemoral bilateral adrenal vein catheterization with adrenal vein sampling may demonstrate a two- to threefold increase in plasma aldosterone concentration on the involved side. In cases of hyperaldosteronism secondary to cortical nodular hyperplasia, no lateralization is found. It is important for samples to be obtained simultaneously if possible and for cortisol levels to be measured to ensure that false localization does not reflect dilution or anACTH- or stress-induced rise in aldosterone levels. In a patient with an adenoma, the aldosterone/cortisol ratio lateralizes to the side of the lesion.

Differential Diagnosis Patients with hypertension and hypokalemia may have either primary or secondary hyperaldosteronism (Fig. 331-10). A useful maneuver to distinguish between these conditions is the measurement of plasma renin activity. Secondary hyperaldosteronism in patients with accelerated hypertension is due to

elevated plasma renin levels; in contrast, patients with primary aldosteronism have suppressed plasma renin levels. Indeed, in patients with a serum potassium concentration of <2.5 mmol/L, a high ratio of plasma aldosterone to plasma renin activity in a random sample is usually sufficient to establish the diagnosis of primary aldosteronism without additional testing.

Primary aldosteronism must also be distinguished from other *hypermineralocorticoid states*. Nonaldosterone mineralocorticoid states will have suppressed plasma renin activity but low aldosterone levels. The most common problem is to distinguish between hyperaldosteronism due to an adenoma and that due to idiopathic bilateral nodular hyperplasia. This distinction is of importance because hypertension associated with idiopathic hyperplasia is usually not benefited by bilateral adrenalectomy, whereas hypertension associated with aldosterone-producing tumors is usually improved or cured by removal of the adenoma. Although patients with idiopathic bilateral nodular hyperplasia tend to have less severe hypokalemia, lower aldosterone secretion, and higher plasma renin activity than do patients with primary aldosteronism, differentiation is impossible solely on clinical and/or biochemical grounds. An anomalous postural decrease in plasma aldosterone and elevated plasma 18-hydroxycorticosterone levels are present in most patients with a unilateral lesion. However, these tests are also of limited diagnostic value in the individual patient, because some adenoma patients have an increase in plasma aldosterone with upright posture, so-called renin-responsive aldosteronoma. A definitive diagnosis is best made by radiographic studies, including bilateral adrenal vein catheterization, as noted above.

In a few instances, hypertensive patients with hypokalemic alkalosis have adenomas that secrete deoxycorticosterone (DOC). Such patients have reduced plasma renin activity levels, but aldosterone levels are either normal or reduced, suggesting the diagnosis of mineralocorticoid excess due to a hormone other than aldosterone. Several inherited disorders have clinical features similar to those of primary aldosteronism (see below).

TREATMENT

Primary aldosteronism due to an adenoma is usually treated by surgical excision of the adenoma. Where possible a laparoscopic approach is favored. However, dietary sodium restriction and the administration of an aldosterone antagonist, e.g., spironolactone, are effective in many cases. Hypertension and hypokalemia are usually controlled by doses of 25 to 100 mg spironolactone every 8 h. In some patients medical management has been successful for years, but chronic therapy in men is usually limited by side effects of spironolactone such as gynecomastia, decreased libido, and impotence.

When idiopathic bilateral hyperplasia is suspected, surgery is indicated only when significant, symptomatic hypokalemia cannot be controlled with medical therapy, e.g., by spironolactone, triamterene, or amiloride. Hypertension associated with idiopathic hyperplasia is usually not benefited by bilateral adrenalectomy.

Secondary Aldosteronism *Secondary aldosteronism* refers to an appropriately increased production of aldosterone in response to activation of the renin-angiotensin system (Fig. 331-10). The production rate of aldosterone is often higher in patients with

secondary aldosteronism than in those with primary aldosteronism. Secondary aldosteronism usually occurs in association with the accelerated phase of hypertension or on the basis of an underlying edema disorder. Secondary aldosteronism in pregnancy is a normal physiologic response to estrogen-induced increases in circulating levels of renin substrate and plasma renin activity and to the antialdosterone actions of progestogens.

Secondary aldosteronism in hypertensive states is due either to a primary overproduction of renin (primary reninism) or to an overproduction of renin secondary to a decrease in renal blood flow and/or perfusion pressure (Fig. 331-10). Secondary hypersecretion of renin can be due to a narrowing of one or both of the major renal arteries by atherosclerosis or by fibromuscular hyperplasia. Overproduction of renin from both kidneys also occurs in severe arteriolar nephrosclerosis (malignant hypertension) or with profound renal vasoconstriction (the accelerated phase of hypertension). The secondary aldosteronism is characterized by hypokalemic alkalosis, moderate to severe increases in plasma renin activity, and moderate to marked increases in aldosterone levels (Chap. 246).

Secondary aldosteronism with hypertension can also be caused by rare renin-producing tumors (primary reninism). These patients have the biochemical characteristics of renal vascular hypertension, but the primary defect is renin secretion by a juxtaglomerular cell tumor. The diagnosis can be made by demonstration of normal renal vasculature and/or demonstration of a space-occupying lesion in the kidney by radiographic techniques and documentation of a unilateral increase in renal vein renin activity. Rarely, these tumors arise in tissues such as the ovary.

Secondary aldosteronism is present in many forms of *edema*. The rate of aldosterone secretion is usually increased in patients with edema caused by either cirrhosis or the nephrotic syndrome. In congestive heart failure, elevated aldosterone secretion varies depending on the severity of cardiac failure. The stimulus for aldosterone release in these conditions appears to be *arterial hypovolemia* and/or hypotension. Thiazides and furosemide often exaggerate secondary aldosteronism via volume depletion; hypokalemia and, on occasion, alkalosis can then become prominent features. On occasion secondary hyperaldosteronism occurs without edema or hypertension (Bartter's and Gitelman's syndromes, see below).

SYNDROMES OF ADRENAL ANDROGEN EXCESS

Adrenal androgen excess results from excess production of DHEA and androstenedione, which are converted to testosterone in extraglandular tissues; elevated testosterone levels account for most of the virilization. Adrenal androgen excess may be associated with the secretion of greater or smaller amounts of other adrenal hormones and may, therefore, present as "pure" syndromes of virilization or as "mixed" syndromes associated with excessive glucocorticoids and Cushing's syndrome. *For further discussion of hirsutism and virilization, see Chap. 53.*

HYPOFUNCTION OF THE ADRENAL CORTEX

Cases of adrenal insufficiency can be divided into two general categories: (1) those

associated with primary inability of the adrenal to elaborate sufficient quantities of hormone, and (2) those associated with a secondary failure due to inadequate ACTH formation or release (Table 331-6).

PRIMARY ADRENOCORTICAL DEFICIENCY (ADDISON'S DISEASE)

The original description of Addison's disease -- "general languor and debility, feebleness of the heart's action, irritability of the stomach, and a peculiar change of the color of the skin" -- summarizes the dominant clinical features. Advanced cases are usually easy to diagnose, but recognition of the early phases can be a real challenge.

Incidence Primary insufficiency is relatively rare, may occur at any age, and affects both sexes equally. Because of the common therapeutic use of steroids, secondary adrenal insufficiency is relatively common.

Etiology and Pathogenesis Addison's disease results from progressive destruction of the adrenals, which must involve >90% of the glands before adrenal insufficiency appears. The adrenal is a frequent site for chronic granulomatous diseases, predominantly tuberculosis but also histoplasmosis, coccidioidomycosis, and cryptococcosis. In early series, tuberculosis was responsible for 70 to 90% of cases, but the most frequent cause now is *idiopathic* atrophy, and an autoimmune mechanism is probably responsible. Rarely, other lesions are encountered, such as adrenoleukodystrophy, bilateral hemorrhage, tumor metastases, HIV, cytomegalovirus (CMV), amyloidosis, adrenomyeloneuropathy, familial adrenal insufficiency, or sarcoidosis.

Although half of patients with idiopathic atrophy have circulating adrenal antibodies, autoimmune destruction is probably secondary to cytotoxic T lymphocytes. Specific adrenal antigens to which autoantibodies may be directed include 21-hydroxylase (CYP21A2) and side chain cleavage enzyme but the significance of these antibodies in the pathogenesis of adrenal insufficiency is unknown. Some antibodies cause adrenal insufficiency by blocking the binding of ACTH to its receptors. Some patients also have antibodies to thyroid, parathyroid, and/or gonadal tissue (Chap. 339). There is also an increased incidence of chronic lymphocytic thyroiditis, premature ovarian failure, type 1 diabetes mellitus, and hypo- or hyperthyroidism. The presence of two or more of these autoimmune endocrine disorders in the same person defines the polyglandular autoimmune syndrome type II. Additional features include pernicious anemia, vitiligo, alopecia, nontropical sprue, and myasthenia gravis. Within families, multiple generations are affected by one or more of the above diseases. Type II polyglandular syndrome is the result of a mutant gene on chromosome 6 and is associated with the HLA alleles B8 and DR3.

The combination of parathyroid and adrenal insufficiency and chronic mucocutaneous moniliasis constitutes type I polyglandular autoimmune syndrome. Other autoimmune diseases in this disorder include pernicious anemia, chronic active hepatitis, alopecia, primary hypothyroidism, and premature gonadal failure. There is no HLA association; this syndrome is inherited as an autosomal recessive trait. It is caused by mutations in *a*utoimmune *p*olyendocrinopathy *c*andidiasis *e*ctodermal *d*ystrophy (APECED) located on chromosome 21q22.3. The gene encodes a transcription factor thought to be

involved in lymphocyte function. The type I syndrome usually presents during childhood, whereas the type II syndrome is usually manifested in adulthood.

Clinical suspicion of adrenal insufficiency should be high in patients with AIDS (Chap. 309).CMVregularly involves the adrenal glands (so-called CMV necrotizing adrenalitis), and involvement with *Mycobacterium avium-intracellulare*, *Cryptococcus*, and Kaposi's sarcoma has been reported. Adrenal insufficiency in AIDS patients may not be manifest, but tests of adrenal reserve frequently give abnormal results. When interpreting tests of adrenocortical function, it is important to remember that medications such as rifampin, phenytoin, ketoconazole, megace, and opiates may cause or potentiate adrenal insufficiency. Adrenal hemorrhage and infarction occur in patients on anticoagulants and in those with circulating anticoagulants and hypercoagulable states, such as the antiphospholipid syndrome.

There are several rare genetic causes of adrenal insufficiency that present primarily in infancy and childhood (see below).

Clinical Signs and Symptoms Adrenocortical insufficiency caused by gradual adrenal destruction is characterized by an insidious onset of fatigability, weakness, anorexia, nausea and vomiting, weight loss, cutaneous and mucosal pigmentation, hypotension, and occasionally hypoglycemia (Table 331-7). Depending on the duration and degree of adrenal hypofunction, the manifestations vary from mild chronic fatigue to fulminating shock associated with acute destruction of the glands, as described by Waterhouse and Friderichsen.

Asthenia is the cardinal symptom. Early it may be sporadic, usually most evident at times of stress; as adrenal function becomes more impaired, the patient is continuously fatigued, and bed rest is necessary.

Hyperpigmentation may be striking or absent. It commonly appears as a diffuse brown, tan, or bronze darkening of parts such as the elbows or creases of the hand and of areas that normally are pigmented such as the areolae about the nipples. Bluish-black patches may appear on the mucous membranes. Some patients develop dark freckles, and irregular areas of vitiligo may paradoxically be present. As an early sign, tanning following sun exposure may be persistent.

Arterial hypotension with postural accentuation is frequent, and blood pressure may be in the range of 80/50 or less.

Abnormalities of gastrointestinal function are often the presenting complaint. Symptoms vary from mild anorexia with weight loss to fulminating nausea, vomiting, diarrhea, and ill-defined abdominal pain, which may be so severe as to be confused with an acute abdomen. Patients may have personality changes, usually consisting of excessive irritability and restlessness. Enhancement of the sensory modalities of taste, olfaction, and hearing is reversible with therapy. Axillary and pubic hair may be decreased in women due to loss of adrenal androgens.

Laboratory Findings In the early phase of gradual adrenal destruction, there may be no demonstrable abnormalities in the routine laboratory parameters, but adrenal reserve

is decreased -- that is, while basal steroid output may be normal, a subnormal increase occurs after stress. Adrenal stimulation with ACTH uncovers abnormalities in this stage of the disease, eliciting a subnormal increase of cortisol levels or no increase at all. In more advanced stages of adrenal destruction, serum sodium, chloride, and bicarbonate levels are reduced, and the serum potassium level is elevated. The hyponatremia is due both to loss of sodium into the urine (due to aldosterone deficiency) and to movement into the intracellular compartment. This extravascular sodium loss depletes extracellular fluid volume and accentuates hypotension. Elevated plasma vasopressin and angiotensin II levels may contribute to the hyponatremia by impairing free water clearance. Hyperkalemia is due to a combination of aldosterone deficiency, impaired glomerular filtration, and acidosis. Basal levels of cortisol and aldosterone are subnormal and fail to increase following ACTH administration. Mild to moderate hypercalcemia occurs in 10 to 20% of patients for unclear reasons. The electrocardiogram may show nonspecific changes, and the electroencephalogram exhibits a generalized reduction and slowing. There may be a normocytic anemia, a relative lymphocytosis, and a moderate eosinophilia.

Diagnosis The diagnosis of adrenal insufficiency should be made only with ACTH stimulation testing to assess adrenal reserve capacity for steroid production (see above for ACTH test protocols). In brief, the best screening test is the cortisol response 60 min after 250 ug of cosyntropin given intramuscularly or intravenously. Cortisol levels should exceed 495 nmol/L (18 ug/dL). If the response is abnormal, then primary and secondary adrenal insufficiency can be distinguished by measuring aldosterone levels from the same blood samples. In secondary, but not primary, adrenal insufficiency the aldosterone increment will be normal [³150 pmol/l (5 ng/dL)]. Furthermore, in primary adrenal insufficiency, plasma ACTH and associated peptides (b-LPT) are elevated because of loss of the usual cortisol-hypothalamic-pituitary feedback relationship, whereas in secondary adrenal insufficiency, plasma ACTH values are low or "inappropriately" normal (Fig. 331-11).

Differential Diagnosis Since weakness and fatigue are common, diagnosis of early adrenocortical insufficiency may be difficult. However, the combination of mild gastrointestinal distress, weight loss, anorexia, and a suggestion of increased pigmentation makes it mandatory to perform ACTH stimulation testing to rule out adrenal insufficiency, particularly before steroid treatment is begun. Weight loss is useful in evaluating the significance of weakness and malaise. Racial pigmentation may be a problem, but a *recent* and progressive *increase* in pigmentation is usually reported by the patient with gradual adrenal destruction. Hyperpigmentation is usually absent when adrenal destruction is rapid, as in bilateral adrenal hemorrhage. The fact that hyperpigmentation occurs with other diseases may also present a problem, but the appearance and distribution of pigment in adrenal insufficiency are usually characteristic. When doubt exists, measurement of ACTH levels and testing of adrenal reserve with the infusion of ACTH provide clear-cut differentiation.

TREATMENT

All patients with adrenal insufficiency should receive specific hormone replacement. Like diabetics, these patients require careful education about the disease. Replacement therapy should correct both glucocorticoid and mineralocorticoid deficiencies.

Hydrocortisone (cortisol) is the mainstay of treatment. The dose for most adults (depending on size) is 20 to 30 mg/d. Patients are advised to take glucocorticoids with meals or, if that is impractical, with milk or an antacid, because the drugs may increase gastric acidity and exert direct toxic effects on the gastric mucosa. To simulate the normal diurnal adrenal rhythm, two-thirds of the dose is taken in the morning, and the remaining one-third is taken in the late afternoon. Some patients exhibit insomnia, irritability, and mental excitement after initiation of therapy; in these, the dosage should be reduced. Other situations that may necessitate smaller doses are hypertension and diabetes mellitus. Obese individuals and those on anticonvulsive medications may require increased dosages. Measurements of plasmaACTH or cortisol or of urine cortisol levels do not appear to be useful in determining optimal glucocorticoid dosages.

Since the replacement dosage of hydrocortisone does not replace the mineralocorticoid component of the adrenal hormones, mineralocorticoid supplementation is usually needed. This is accomplished by the administration of 0.05 to 0.1 mg fludrocortisone per day by mouth. Patients should also be instructed to maintain an ample intake of sodium (3 to 4 g/d).

The adequacy of mineralocorticoid therapy can be assessed by measurement of blood pressure and serum electrolytes. Blood pressure should be normal and without postural changes; serum sodium, potassium, creatinine, and urea nitrogen levels should also be normal. Measurement of plasma renin levels may also be useful in titrating the dose.

In female patients with adrenal insufficiency, androgen levels are also low. Thus, some physicians believe that daily replacement with 25 to 50 mg ofDHEAorally may improve quality of life and skeletal density.

Complications of glucocorticoid therapy, with the exception of gastritis, are *rare* at the dosages recommended for treatment of adrenal insufficiency. Complications of mineralocorticoid therapy include hypokalemia, hypertension, cardiac enlargement, and even congestive heart failure due to sodium retention. Periodic measurements of body weight, serum potassium level, and blood pressure are useful. All patients with adrenal insufficiency should carry medical identification, should be instructed in the parenteral self-administration of steroids, and should be registered with a medical alerting system.

Special Therapeutic Problems During periods of intercurrent illness, especially in the setting of fever, the dose of hydrocortisone should be doubled. With severe illness it should be increased to 75 to 150 mg/d. When oral administration is not possible, parenteral routes should be employed. Likewise, before surgery or dental extractions, supplemental glucocorticoids should be administered. Patients should also be advised to increase the dose of fludrocortisone and to add salt to their otherwise normal diet during periods of strenuous exercise with sweating, during extremely hot weather, and with gastrointestinal upsets such as diarrhea. A simple strategy is to supplement the diet one to three times daily with salty broth (1 cup of beef or chicken bouillon contains 35 mmol of sodium). For a representative program of steroid therapy for the patient with adrenal insufficiency who is undergoing major surgery, seeTable 331-8. This schedule is designed so that on the day of surgery it will mimic the output of cortisol in normal individuals undergoing prolonged major stress (10 mg/h, 250 to 300 mg/d). Thereafter, if the patient is improving and is afebrile, the dose of hydrocortisone is tapered by 20 to

30% daily. Mineralocorticoid administration is unnecessary at hydrocortisone doses >100 mg/d because of the mineralocorticoid effects of hydrocortisone at such dosages.

SECONDARY ADRENOCORTICAL INSUFFICIENCY

ACTHdeficiency causes *secondary* adrenocortical insufficiency; it may be a selective deficiency, as is seen following prolonged administration of excess glucocorticoids, or it may occur in association with deficiencies of multiple pituitary hormones (panhypopituitarism) (Chap. 328). Patients with secondary adrenocortical hypofunction have many symptoms and signs in common with those having primary disease but are *characteristically not hyperpigmented*, since ACTH and related peptide levels are low. In fact, plasma ACTH levels distinguish between primary and secondary adrenal insufficiency, since they are elevated in the former and decreased to absent in the latter. Patients with total pituitary insufficiency have manifestations of multiple hormone deficiencies. An additional feature distinguishing primary adrenocortical insufficiency is the *near-normal level of aldosterone secretion* seen in pituitary and/or isolated ACTH deficiencies (Fig. 331-11). Patients with pituitary insufficiency may have hyponatremia, which can be dilutional or secondary to a subnormal increase in aldosterone secretion in response to severe sodium restriction. However, severe *dehydration*, *hyponatremia*, and *hyperkalemia* are characteristic of severe mineralocorticoid insufficiency and favor a diagnosis of primary adrenocortical insufficiency.

Patients receiving long-term steroid therapy, despite physical findings of Cushing's syndrome, develop adrenal insufficiency because of prolonged pituitary-hypothalamic suppression and adrenal atrophy secondary to the loss of endogenousACTH. These patients have two deficits, a loss of adrenal responsiveness to ACTH and a failure of pituitary ACTH release. They are characterized by low blood cortisol and ACTH levels, a low baseline rate of steroid excretion, and abnormal ACTH and metyrapone responses. Most patients with steroid-induced adrenal insufficiency eventually recover normal hypothalamic-pituitary-adrenal responsiveness, but recovery time varies from days to months. The rapid ACTH test provides a convenient assessment of recovery of hypothalamic-pituitary-adrenal function. Because the plasma cortisol concentrations after injection of cosyntropin and during insulin-induced hypoglycemia are usually similar, the rapid ACTH test assesses the integrated hypothalamic-pituitary-adrenal function (see "Tests of Pituitary-Adrenal Responsiveness," above). Some investigators suggest using the low-dose (1 ug) ACTH test for suspected secondary ACTH deficiency. Additional tests to assess pituitary ACTH reserve include the standard metyrapone and insulin-induced hypoglycemia tests.

Glucocorticoid therapy in patients with secondary adrenocortical insufficiency does not differ from that for the primary disorder. Mineralocorticoid therapy is usually not necessary, as aldosterone secretion is preserved.

ACUTE ADRENOCORTICAL INSUFFICIENCY

Acute adrenocortical insufficiency may result from several processes. On the one hand, *adrenal crisis* may be a rapid and overwhelming intensification of chronic adrenal insufficiency, usually precipitated by sepsis or surgical stress. Alternatively, acute hemorrhagic destruction of both adrenal glands can occur in previously well subjects. In

children, this event is usually associated with septicemia with *Pseudomonas* or meningococcemia (Waterhouse-Friderichsen syndrome). In adults, anticoagulant therapy or a coagulation disorder may result in bilateral adrenal hemorrhage. Occasionally, bilateral adrenal hemorrhage in the newborn results from birth trauma. Hemorrhage has been observed during pregnancy, following idiopathic adrenal vein thrombosis, and as a complication of venography (e.g., infarction of an adenoma). The third and most frequent cause of acute insufficiency is the rapid withdrawal of steroids from patients with adrenal atrophy owing to chronic steroid administration. Acute adrenocortical insufficiency may also occur in patients with congenital adrenal hyperplasia or those with decreased adrenocortical reserve when they are given drugs capable of inhibiting steroid synthesis (mitotane, ketoconazole) or of increasing steroid metabolism (phenytoin, rifampin).

Adrenal Crisis The long-term survival of patients with adrenocortical insufficiency depends largely on the prevention and treatment of adrenal crisis. Consequently, the occurrence of infection, trauma (including surgery), gastrointestinal upsets, or other stresses necessitates an immediate increase in hormone. In untreated patients, preexisting symptoms are intensified. Nausea, vomiting, and abdominal pain may become intractable. Fever may be severe or absent. Lethargy deepens into somnolence, and hypovolemic vascular collapse ensues. In contrast, patients previously maintained on chronic glucocorticoid therapy may not exhibit dehydration or hypotension until they are in a preterminal state, since mineralocorticoid secretion is usually preserved. In all patients in crisis, a precipitating cause should be sought.

TREATMENT

Treatment is directed primarily toward repletion of circulating glucocorticoids and replacement of the sodium and water deficits. Hence an intravenous infusion of 5% glucose in normal saline solution should be started with a bolus intravenous infusion of 100 mg hydrocortisone followed by a continuous infusion of hydrocortisone at a rate of 10 mg/h. An alternative approach is to administer a 100-mg bolus of hydrocortisone intravenously every 6 h. However, only continuous infusion maintains the plasma cortisol constantly at stress levels [>830 nmol/L (30 ug/dL)]. Effective treatment of hypotension requires glucocorticoid replacement and repletion of sodium and water deficits. If the crisis was preceded by prolonged nausea, vomiting, and dehydration, several liters of saline solution may be required in the first few hours. Vasoconstrictive agents (such as dopamine) may be indicated in extreme conditions as adjuncts to volume replacement. With large doses of steroid, e.g., 100 to 200 mg hydrocortisone, the patient receives a maximal mineralocorticoid effect, and supplementary mineralocorticoid is superfluous. Following improvement, the steroid dosage is tapered over the next few days to maintenance levels, and mineralocorticoid therapy is reinstituted if needed (Table 331-8).

HYPOALDOSTERONISM

Isolated aldosterone deficiency accompanied by normal cortisol production occurs in association with hyporeninism, as an inherited biosynthetic defect, postoperatively following removal of aldosterone-secreting adenomas, during protracted heparin or heparinoid administration, in pretectal disease of the nervous system, and in severe

postural hypotension.

The feature common to all forms hypoaldosteronism is the inability to increase aldosterone secretion appropriately in response to salt restriction. Most patients have unexplained hyperkalemia, which often is exacerbated by restriction of dietary sodium intake. In severe cases, urine sodium wastage occurs at a normal salt intake, whereas in milder forms, excessive loss of urine sodium occurs only with salt restriction.

Most cases of isolated hypoaldosteronism occur in patients with a deficiency in renin production (so-called hyporeninemic hypoaldosteronism), most commonly in adults with diabetes mellitus and mild renal failure and in whom hyperkalemia and metabolic acidosis are out of proportion to the degree of renal impairment. Plasma renin levels fail to rise normally following sodium restriction and postural changes. The pathogenesis is uncertain. Possibilities include renal disease (the most likely), autonomic neuropathy, extracellular fluid volume expansion, and defective conversion of renin precursors to active renin. Aldosterone levels also fail to rise normally after salt restriction and volume contraction; this effect is probably related to the hyporeninism, since biosynthetic defects in aldosterone secretion usually cannot be demonstrated. In these patients, aldosterone secretion increases promptly after ACTH stimulation, but it is uncertain whether the magnitude of the response is normal. On the other hand, the level of aldosterone appears to be subnormal in relationship to the hyperkalemia.

Hypoaldosteronism can also be associated with high renin levels and low or elevated levels of aldosterone (see below). Severely ill patients may also have hyperreninemic hypoaldosteronism; such patients have a high mortality rate (80%). Hyperkalemia is not present. Possible explanations for the hypoaldosteronism include adrenal necrosis (uncommon) or a shift in steroidogenesis from mineralocorticoids to glucocorticoids, possibly related to prolonged ACTH stimulation.

Before the diagnosis of isolated hypoaldosteronism is considered for a patient with hyperkalemia, "pseudohyperkalemia" (e.g., hemolysis, thrombocytosis) should be excluded by measuring the *plasma* potassium level. The next step is to demonstrate a normal cortisol response to ACTH stimulation. Then, the response of renin and aldosterone levels to stimulation (upright posture, sodium restriction) should be measured. Low renin and aldosterone levels establish the diagnosis of hyporeninemic hypoaldosteronism. A combination of high renin levels and low aldosterone levels is consistent with an aldosterone biosynthetic defect or a selective unresponsiveness to angiotensin II. Finally, there is a condition that clinically and biochemically mimics hypoaldosteronism with elevated renin levels. However, the aldosterone levels are not low but high -- so-called pseudohypoaldosteronism. This inherited condition is caused by a mutation in the epithelial sodium channel (see below).

TREATMENT

The treatment is to replace the mineralocorticoid deficiency. For practical purposes, the oral administration of 0.05 to 0.15 mg fludrocortisone daily should restore electrolyte balance if salt intake is adequate (e.g., 150 to 200 mmol/d). However, patients with hyporeninemic hypoaldosteronism may require higher doses of mineralocorticoid to correct hyperkalemia. This need poses a potential risk in patients with hypertension,

mild renal insufficiency, or congestive heart failure. An alternative approach is to reduce salt intake and to administer furosemide, which can ameliorate acidosis and hyperkalemia. Occasionally, a combination of these two approaches is efficacious.

GENETIC CONSIDERATIONS

Glucocorticoid Diseases

Congenital Adrenal Hyperplasia Congenital adrenal hyperplasia (CAH) is the consequence of recessive mutations that cause one of several distinct enzymatic defects (see below). Because cortisol is the principal adrenal steroid regulating ACTH elaboration and because ACTH stimulates adrenal growth and function, a block in cortisol synthesis may result in the enhanced secretion of adrenal androgens and/or mineralocorticids depending on the site of the enzyme block. In severe congenital virilizing hyperplasia, the adrenal output of cortisol may be so compromised as to cause adrenal deficiency despite adrenal hyperplasia.

CAH is the most common adrenal disorder of infancy and childhood (Chap. 338). Partial enzyme deficiencies can be expressed after adolescence, predominantly in women with hirsutism and oligomenorrhea but minimal virilization. Late-onset adrenal hyperplasia may account for 5 to 25% of cases of hirsutism and oligomenorrhea in women, depending on the population.

ETIOLOGY Enzymatic defects have been described in 21-hydroxylase (CYP21A2), 17a-hydroxylase/17,20-Lyase (CYP17), 11b-hydroxylase (CYP11B1), and in (3b-HSD2) (Fig. 331-2). Although the cDNAs for these enzymes have been cloned, the diagnosis of specific enzyme deficiencies with genetic techniques is not practical for routine use. CYP21A2 deficiency is closely linked to the HLA-B locus of chromosome 6 so that HLA typing and/or DNA polymorphism can be used to detect the heterozygous carriers and to diagnose affected individuals in some families (Chap. 306). The clinical expression in the different disorders is variable, ranging from virilization of the female (CYP2/A2) to feminization of the male (3b-HSD2) (Chap. 338).

Adrenal virilization in the female at birth is associated with ambiguous external genitalia (*female pseudohermaphroditism*). Virilization probably begins after the fifth month of intrauterine development. At birth there may be enlarged genitalia in the male infant and enlargement of the clitoris, partial or complete fusion of the labia, and sometimes a urogenital sinus in the female. If the labial fusion is nearly complete, the female infant has external genitalia resembling a penis with hypospadias. In the *postnatal* period, CAH is associated with virilization in the female and isosexual precocity in the male. The excessive androgen levels result in accelerated growth, so that bone age exceeds chronologic age. Because epiphyseal closure is hastened by excessive androgens, growth stops, but truncal development continues, the characteristic appearance being a short child with a well-developed trunk.

The most common form of CAH (95% of cases) is a result of impairment of CYP21A2. In addition to cortisol deficiency, aldosterone secretion is decreased in approximately one-third of the patients. Thus, with CYP21A2 deficiency, adrenal virilization occurs with or without a salt-losing tendency due to aldosterone deficiency (Fig. 331-2).

CYP11B1 deficiency causes a "hypertensive" variant of CAH. Hypertension and hypokalemia occur because of the impaired conversion of 11-deoxycorticosterone to corticosterone, resulting in the accumulation of 11-deoxycorticosterone, a potent mineralocorticoid. The degree of hypertension is variable. Increased shunting again occurs into the androgen pathway.

CYP17 deficiency is characterized by hypogonadism, hypokalemia, and hypertension. This rare disorder causes decreased production of cortisol and shunting of precursors into the mineralocorticoid pathway with hypokalemic alkalosis, hypertension, and suppressed plasma renin activity. Usually, 11-deoxycorticosterone production is elevated. Because CYP17 hydroxylation is required for biosynthesis of both adrenal androgens and gonadal testosterone and estrogen, this defect is associated with sexual immaturity, high urinary gonadotropin levels, and low urinary 17-ketosteroid excretion. Female patients have primary amenorrhea and lack of development of secondary sexual characteristics. Because of deficient androgen production, male patients have either ambiguous external genitalia or a female phenotype (*male pseudohermaphroditism*). Exogenous glucocorticoids can correct the hypertensive syndrome, and treatment with appropriate gonadal steroids results in sexual maturation.

With 3b-HSD2 deficiency, conversion of pregnenolone to progesterone is impaired, so that the synthesis of both cortisol and aldosterone is blocked, with shunting into the adrenal androgen pathway via 17a-hydroxypregnenolone and DHEA. Because DHEA is a weak androgen, and because this enzyme deficiency is also present in the gonad, the genitalia of the male fetus may be incompletely virilized or feminized. Conversely, in the female, overproduction of DHEA may produce partial virilization.

DIAGNOSIS The diagnosis of CAH should be considered in infants having episodes of acute adrenal insufficiency or salt-wasting or with hypertension. The diagnosis is further suggested by the finding of hypertrophy of the clitoris, fused labia, or a urogenital sinus in the female or of isosexual precocity in the male. In infants and children with a CYP21A2 defect, increased urine 17-ketosteroid excretion and increased plasma DHEA sulfate levels are typically associated with an increase in the blood levels of 17-hydroxyprogesterone and the excretion of its urinary metabolite pregnanetriol. Demonstration of elevated levels of 17-hydroxyprogesterone in amniotic fluid at 14 to 16 weeks of gestation allows prenatal detection of affected female infants.

The diagnosis of a *salt-losing form* of CAH due to defects in CYP21A2 is suggested by episodes of acute adrenal insufficiency with hyponatremia, hyperkalemia, dehydration, and vomiting. These infants and children often crave salt and have laboratory findings indicating deficits in both cortisol and aldosterone secretion.

With the *hypertensive form* of CAH due to CYP11B1 deficiency, 11-deoxycorticosterone and 11-deoxycortisol accumulate. The diagnosis is confirmed by demonstrating increased levels of 11-deoxycortisol in the blood or increased amounts of tetrahydro-11-deoxycortisol in the urine. Elevation of 17-hydroxyprogesterone levels does not imply a coexisting CYP21A2 deficiency.

Very high levels of urine DHEA with low levels of pregnanetriol and of cortisol metabolites

in urine are characteristic of children with 3b-HSD2 deficiency. Marked salt-wasting may also occur.

Adults with *late-onset adrenal hyperplasia* (partial deficiency of CYP21A2, CYP11B1, or 3b-HSD2) are characterized by normal or moderately elevated levels of urinary 17-ketosteroids and plasmaDHEAsulfate. A high basal level of a precursor of cortisol biosynthesis (such as 17-hydroxyprogesterone, 17-hydroxypregnenolone, or 11-deoxycortisol), or elevation of such a precursor afterACTHstimulation, confirms the diagnosis of a partial deficiency. Measurement of steroid precursors 60 min after bolus administration of ACTH is usually sufficient. Adrenal androgen output is easily suppressed by the standard low-dose (2 mg) dexamethasone test.

TREATMENT

Patients with CAHhave a fundamental defect of cortisol deficiency with resultant excessiveACTHsecretion, producing hyperplasia of the adrenal glands and causing additional shunting into the precursor steroid pathways. Therapy in these patients consists of daily administration of glucocorticoids to suppress pituitary ACTH secretion. Because of its cost and intermediate half-life, prednisone is the drug of choice except in infants, in whom hydrocortisone is usually used. In adults with late-onset adrenal hyperplasia, the smallest single bedtime dose of a long- or intermediate-acting glucocorticoid that suppresses pituitary ACTH secretion should be administered. The amount of steroid required by children with CAH is approximately 1 to 1.5 times the normal cortisol production rate of 27 to 35 umol (10 to 13 mg) of cortisol per square meter of body surface per day and is given in divided doses two or three times per day. The dosage schedule is governed by repetitive analysis of the urinary 17-ketosteroids, plasmaDHEAsulfate, and/or precursors of cortisol biosynthesis. Skeletal growth and maturation must also be monitored closely, as overtreatment with glucocorticoid replacement therapy retards linear growth.

Receptor Mutations Much less common thanCAH are three syndromes secondary to mutation(s) in a key receptor involved in adrenal function. *Isolated glucocorticoid deficiency* is a rare autosomal recessive disease secondary to a mutation in theACTHreceptor. Usually mineralocorticoid function is normal. Adrenal insufficiency is manifest within the first 2 years of life usually as hyperpigmentation, convulsions, and/or frequent episodes of hypoglycemia. In some patients the adrenal insufficiency is associated with achalasia and alacrima -- Allgrove's, or triple A, syndrome. However, in some triple A syndrome patients, no mutation in the ACTH receptor has been identified, suggesting that a distinct genetic abnormality causes this syndrome. *Adrenal hypoplasia congenita* is a rare X-linked disorder caused by a mutation in the *DAX1* gene located on the X chromosome. This gene encodes an orphan nuclear receptor that plays an important role in the development of the adrenal cortex and also the hypothalamic-pituitary-gonadal axis. Thus, patients present with signs and symptoms secondary to deficiencies of all three major adrenal steroids -- cortisol, aldosterone, and adrenal androgens -- as well as gonadotropin deficiency. Finally a rare cause of hypercortisolism without cushingoid stigmata is *primary cortisol resistance* due to mutations in the glucocorticoid receptor. The resistance is incomplete because patients do not exhibit signs of adrenal insufficiency. Thus, these three rare inherited disorders have in common an elevated ACTH. However, the clinical manifestations range from no

evidence of adrenal insufficiency to only cortisol deficiency (similar to secondary adrenal insufficiency) to a clinical picture indistinguishable from classic Addison's disease.

Miscellaneous Conditions Adrenoleukodystrophy causes severe demyelination and early death in children, and adrenomyeloneuropathy is associated with a mixed motor and sensory neuropathy with spastic paraplegia in adults; both disorders are associated with elevated circulating levels of very long chain fatty acids and cause adrenal insufficiency. Autosomal recessive mutations in the *steroidogenic acute regulatory* (STAR) protein gene cause congenital lipoid adrenal hyperplasia (Chap. 338), which is characterized by adrenal insufficiency and defective gonadal steroidogenesis. Because STAR mediates cholesterol transport into the mitochondrion, mutations in the protein cause massive lipid accumulation in steroidogenic cells, ultimately leading to cell toxicity. Thus, these three rare inherited disorders have in common an elevated ACTH. However, the clinical manifestations range from no evidence of adrenal insufficiency to only cortisol deficiency (similar to secondary adrenal insufficiency) to a clinical picture indistinguishable from classic Addison's disease.

MINERALOCORTICOID DISEASES

Some forms of CAH have a mineralocorticoid component (see above). Others are caused by a mutation in other enzymes or ion channels important in mediating or mimicking aldosterone's action.

Hypermineralocorticoidism

Low Plasma Renin Activity Rarely, hypermineralocorticoidism is due to a defect in cortisol biosynthesis, specifically 11- or 17-hydroxylation. ACTH levels are increased, with a resultant increase in the production of the mineralocorticoid 11-deoxycorticosterone. Hypertension and hypokalemia can be corrected by glucocorticoid administration. The definitive diagnosis is made by demonstrating an elevation of precursors of cortisol biosynthesis in the blood or urine or by direct demonstration of the genetic defect.

Glucocorticoid administration can also ameliorate hypertension or produce normotension even though a hydroxylase deficiency cannot be identified (Fig. 331-9). These patients have normal to slightly elevated aldosterone levels that do not suppress in response to saline but do suppress in response to 2 days of dexamethasone (2 mg/d). The condition is inherited as an autosomal dominant trait and is termed *glucocorticoid-remediable aldosteronism* (GRA). This entity is secondary to a chimeric gene duplication whereby the 11-bhydroxylase gene promoter (which is under the control of ACTH) is fused to the aldosterone synthase coding sequence. Thus, aldosterone synthase activity is ectopically expressed in the zona fasciculata and is regulated by ACTH, in a fashion similar to the regulation of cortisol secretion. Screening for this defect is best performed by assessing the presence or absence of the chimeric gene. Because the abnormal gene may be present in the absence of hypokalemia, its frequency as a cause of hypertension is unknown. Individuals with suppressed plasma renin levels and juvenile-onset hypertension or a history of early-onset hypertension in first-degree relatives should be screened for this disorder. Early hemorrhagic stroke also occurs in GRA-affected individuals.

Glucocorticoid-remediable hyperaldosteronism documented by genetic analysis may be treated with glucocorticoid administration or antimineralocorticoids, e.g., spironolactone, triamterene, or amiloride. Glucocorticoids should be used only in small doses to avoid inducing iatrogenic Cushing's syndrome. A combination approach is often necessary.

High Plasma Renin Activity Bartter's syndrome is characterized by severe hyperaldosteronism (hypokalemic alkalosis) with moderate to marked increases in renin activity and hypercalciuria, but normal blood pressure and no edema; this disorder usually begins in childhood. Renal biopsy shows juxtaglomerular hyperplasia. The pathogenesis involves a defect in the renal conservation of sodium or chloride. The renal loss of sodium is thought to stimulate renin secretion and aldosterone production. Hyperaldosteronism produces potassium depletion, and hypokalemia further elevates prostaglandin production and plasma renin activity. In some cases, the hypokalemia may be potentiated by a defect in renal conservation of potassium. Increased production of prostaglandins is probably not a primary abnormality, since administration of inhibitors of prostaglandin synthesis reverses the features only temporarily (Chap. 276). Bartter's syndrome is caused by a mutation in the renal Na-K-2Cl co-transporter gene.

Gitelman's syndrome is an autosomal recessive trait characterized by renal salt wasting and as a result, as in Bartter's syndrome, activation of the renin-angiotensin-aldosterone system. As a consequence affected individuals have low blood pressure, low serum potassium, low serum magnesium, and high serum bicarbonate. In contrast to Bartter's syndrome, urinary calcium excretion is reduced. Gitelman's syndrome results from loss-of-function mutations of the renal thiazide-sensitive Na-Cl co-transporter.

Increased Mineralocorticoid Action Liddle's syndrome is a rare autosomal dominant disorder that mimicks hyperaldosteronism. The defect is in the genes encoding theb or g subunits of the epithelial sodium channel. Both renin and aldosterone levels are low, owing to the constitutively activated sodium channel and the resulting excess sodium reabsorption in the renal tubule.

A rare autosomal recessive cause of hypokalemia and hypertension is 11b-HSDII deficiency, in which cortisol cannot be converted to cortisone and hence binds to theMR and acts as a mineralocorticoid. This condition, also termed *apparent mineralocorticoid excess syndrome*, is caused by a defect in the gene encoding the renal isoform of this enzyme, 11b-HSD II. Patients can be identified either by documenting an increased ratio of cortisol to cortisone in the urine or by genetic analysis. Patients with the 11b-HSD deficiency syndrome can be treated with small doses of dexamethasone. Although dexamethasone is a potent glucocorticoid that suppressesACTH and endogenous cortisol production, it binds less well to the mineralocorticoid receptor than does cortisol.

The ingestion of candies or chewing tobacco containing certain forms of licorice produces a syndrome that mimics primary aldosteronism. The component of such agents that causes sodium retention is glycyrrhizinic acid, which inhibits the 11b-HSDII and hence allows cortisol to act as a mineralocorticoid and cause sodium retention, expansion of the extracellular fluid volume, hypertension, depressed plasma renin levels, and suppressed aldosterone levels. The diagnosis is established or excluded by

a careful history.

Decreased Mineralocorticoid Production or Action In patients with these conditions, disorders of aldosterone biosynthesis or action are associated with high renin levels, salt wasting, and hyperkalemia. The aldosterone levels may be low or elevated. In patients with a deficiency in aldosterone biosynthesis, the transformation of corticosterone into aldosterone is impaired, owing to a mutation in the aldosterone synthase (CYP11B2) gene. These patients have low to absent aldosterone secretion, elevated plasma renin levels, and elevated levels of the intermediates of aldosterone biosynthesis (corticosterone and 18-hydroxycorticosterone).

Pseudohypoaldosteronism type I (PHA-I) is an autosomal recessive disorder that is seen in the neonatal period and is characterized by salt wasting, hypotension, hyperkalemia, and high renin and aldosterone levels. In contrast to the gain-of-function mutations in the epithelial sodium channel (ENaC) in Liddle's syndrome, mutations in PHA-I result in loss of ENaC function.

NONSPECIFIC CLINICAL USE OF ADRENAL STEROIDS

The widespread use of glucocorticoids emphasizes the need for a thorough understanding of the metabolic effects of these agents. Before adrenal hormone therapy is instituted, the expected gains should be weighed against undesirable effects.

HOW SERIOUS IS THE DISORDER?

In a patient who has unexplained shock or in whom other measures have failed, the physician need not hesitate to employ high-dose steroid therapy. In contrast, one should exercise restraint in administering steroids to a patient with early rheumatoid arthritis for whom physiotherapy, anti-inflammatory agents, disease-modifying agents, and general medical care have not been tried (Chap. 312).

HOW LONG WILL GLUCOCORTICOID THERAPY BE REQUIRED?

The use of intravenous steroids for 24 to 48 h for a life-threatening situation such as status asthmaticus or pseudotumor cerebri has few or no contraindications, in contrast to the initiation of chronic steroid therapy for asthma, arthritis, or psoriasis. In the latter instances, the almost certain development of some degree of Cushing's syndrome must be weighed against the potential benefit. These side effects should be minimized by a careful choice of steroid preparations, alternate-day or interrupted therapy; the use of topical steroids, e.g., inhaled, intranasal, or dermal, whenever possible; and the judicious use of supplementary adjuvants.

WHICH PREPARATION IS BEST?

Several considerations should be taken into account in deciding which steroid preparation to use.

1. *The biologic half-life.* The rationale behind alternate-day therapy is to decrease the metabolic effects of the steroids for a significant part of each 48 h period while still

producing a pharmacologic effect durable enough to be effective. Too long a half-life would defeat the first purpose, and too short a half-life would defeat the second. In general, the more potent the steroid, the longer its biologic half-life.

2. *The importance of the mineralocorticoid effects of the steroid.* Most synthetic steroids have less mineralocorticoid effect than hydrocortisone (Table 331-9).

3. *The biologically active form of the steroid.* Cortisone and prednisone have to be converted to biologically active metabolites before anti-inflammatory effects can occur. Because of this, in a condition for which steroids are known to be effective and when an adequate dose has been given without response, one should consider substituting hydrocortisone or prednisolone for cortisone or prednisone.

4. *The cost of the medication.* This is a serious consideration if chronic administration is planned. Prednisone is the least expensive of available steroid preparations.

5. *The type of formulation.* Topical steroids have the distinct advantage over oral steroids in reducing the likelihood of systemic side effects. In addition, some inhaled steroids have been designed to minimize side effects by increasing their hepatic inactivation if they are swallowed (Chap. 252). However, all topical steroids can be absorbed into the systemic circulation.

EVALUATION OF PATIENTS PRIOR TO INITIATING STEROID THERAPY (SeeTable 331-10)

Chronic Infection Three issues demand attention: (1) Any active infection, particularly tuberculosis, should be identified. (2) If tuberculosis is present, steroid therapy should be employed only in conjunction with antituberculous chemotherapy. The chest film and tuberculin test provide baseline information for future comparison. Since high-dose steroids may impair the tuberculin reaction, serial chest roentgenograms may be necessary. (3) Infection due to opportunistic pathogens should be constantly considered in patients on steroid therapy, especially when combined with other immunosuppressive agents.

Diabetes Mellitus Prolonged glucocorticoid therapy may unmask or aggravate diabetes mellitus. The presence of diabetes mellitus or the demonstration of impaired glucose tolerance may affect the decision to institute glucocorticoid therapy.

Osteoporosis Patients receiving long-term steroid therapy are at risk for osteoporosis. Indeed, osteoporosis with vertebral fractures or compression is a dreaded complication for patients at high risk (postmenopausal women, elderly men, patients with restricted physical activity, and especially organ transplant patients who are maintained on high doses to prevent rejection). Alternate-day or interrupted steroid therapy does not prevent the risk of this complication (Table 331-11). Adjunctive therapies with antiresorptive agents, such as parenteral or oral bisphosphonates, have been shown to be effective in treating the osteoporosis (Chap. 342). Bone mineral density should be assessed periodically with dual-energy X-ray absorptiometry (DEXA) scans.

Peptic Ulcer, Gastric Hypersecretion, or Esophagitis In conventional doses

(equivalent to£15 mg prednisone per day), glucocorticoids probably do not cause peptic ulceration; whether higher doses cause ulcer disease is not established and probably depends on duration, dose of treatment, and predisposing factors such as hypoalbuminemia or cirrhosis and also whether subjects are concomitantly ingesting nonsteroidal anti-inflammatory agents (NSAIDs). However, even at conventional doses, patients with a history of ulcer may experience aggravation of symptoms while receiving glucocorticoids. Consequently, all individuals with a positive history or with known risk factors should be managed with a vigorous antiulcer program (antacids, H$_2$receptor antagonists, or ATPase inhibitors) along with glucocorticoids. *The development of anemia in a patient receiving glucocorticoids should suggest gastrointestinal bleeding as a cause.*

Hypertension or Cardiovascular Disease The capacity of many adrenal steroid preparations to promote sodium retention makes it necessary to exercise caution when using them in patients with preexisting hypertension or cardiovascular or renal disease. The use of preparations with minimal sodium-retaining activity, restriction of dietary sodium intake, and the use of diuretic agents and supplementary potassium salts minimize the mineralocorticoid effects of steroids. However, hypertension may be exacerbated by steroid-induced increases in renin substrate and angiotensin II levels and by reduction in vasodilator prostaglandin production. Steroids also accelerate atherogenesis by induction of hypertension, glucose intolerance, and unfavorable lipid profiles. Glucocorticoid-associated lipid abnormalities include hypertriglyceridemia, hypercholesterolemia, and increasedLDLcholesterol levels.

Psychological Difficulties Steroid therapy may cause psychological disturbances. In general, these disturbances correlate better with the patient's personality than with the dose of hormone, although larger doses cause more serious reactions. There is no reliable method of predicting the psychological reaction to steroid therapy; moreover, previous tolerance of steroids does not necessarily ensure the safety of subsequent courses. Likewise, untoward psychological reactions on one occasion do not invariably mean that the patient will respond unfavorably to a second course. However, patients with depressive symptoms during a first course of steroids may benefit from prophylactic treatment prior to a second course.

Sleeplessness is common and can be minimized by using the shorter-acting steroids and by prescribing the total daily amount as a single early-morning dose.

ALTERNATE-DAY STEROID THERAPY

The most effective way to minimize the cushingoid effects of glucocorticoids is to administer the total 48-h dose as a *single* dose of *intermediate-acting steroid* in the morning, *every other day*. If symptoms of the underlying disorder can be controlled by this technique, it offers distinct advantages. Three considerations deserve mention: (1) The alternate-day schedule may be approached through transition schedules that allow the patient to adjust gradually; (2) supplementary nonsteroid medications may be needed on the "off" day to minimize symptoms of the underlying disorder; and (3) many symptoms that occur during the off day (e.g., fatigue, joint pain, muscle stiffness or tenderness, and fever) may represent relative adrenal insufficiency rather than exacerbation of the underlying disease.

The alternate-day approach capitalizes on the fact that cortisol secretion and plasma levels normally are highest in the early morning and lowest in the evening. The normal pattern is mimicked by administering an intermediate-acting steroid in the morning (7 to 8 A.M.) (Table 331-9).

Initially, the steroid program often requires daily or more frequent doses of steroid to achieve the desired anti-inflammatory or immunity-suppressing action. *Only after this desired effect is achieved is an attempt made to switch to an alternate-day program.* A number of schedules can be used for transferring from a daily to an alternate-day program. The key points to be considered are flexibility in arranging a program and the use of supportive measures on the off day. One may attempt a gradual transition to the alternate-day schedule rather than an abrupt changeover. One approach is to keep the steroid dose constant on one day and gradually reduce it on the alternate day. Alternatively, the steroid dose can be increased on one day and reduced on the alternate day. In any case, it is important to anticipate that some increase in pain or discomfort may occur in the 36 to 48 h following the last dose.

WITHDRAWAL OF GLUCOCORTICOIDS FOLLOWING LONG-TERM USE

It is possible to reduce gradually and eventually to discontinue a daily steroid dose, but under most circumstances withdrawal of steroids should be initiated by first implementing an alternate-day schedule. Patients who have been on an alternate-day program for a month or more experience less difficulty during termination regimens. The dosage is gradually reduced and finally discontinued after a replacement dosage has been reached (e.g., 5 to 7.5 mg prednisone). Complications rarely ensue unless undue stress is experienced, and patients should understand that for 1 year or longer after withdrawal from long-term high-dose steroid therapy, supplementary hormone should be given in the event of a serious infection, operation, or injury. A useful strategy in patients with symptoms of adrenal insufficiency on a tapering regimen is to measure plasma cortisol levels prior to the steroid dose. A level<140 nmol/L (5 ug/dL) indicates continuing suppression of the pituitary-adrenal axis and implies that a more cautious tapering of steroids is indicated.

In patients on high-dose daily steroid therapy, it is advised to reduce dosage to approximately 20 mg prednisone daily as a single morning dose before beginning the transition to alternate-day therapy. If a patient cannot tolerate an alternate-day program, consideration should be given to the possibility that the patient has developed primary adrenal insufficiency.

(Bibliography omitted in Palm version)

332. PHEOCHROMOCYTOMA - *Lewis Landsberg, James B. Young*

Pheochromocytomas produce, store, and secrete catecholamines. They are usually derived from the adrenal medulla but may develop from chromaffin cells in or about sympathetic ganglia (extraadrenal pheochromocytomas or paragangliomas). Related tumors that secrete catecholamines and produce similar clinical syndromes include chemodectomas derived from the carotid body and ganglioneuromas derived from the postganglionic sympathetic neurons.

The clinical features are due predominantly to the release of catecholamines and, to a lesser extent, to the secretion of other substances. Hypertension is the most common sign, and hypertensive paroxysms or crises, often spectacular and alarming, occur in over half the cases.

Pheochromocytoma occurs in approximately 0.1% of the hypertensive population but is, nevertheless, an important correctable cause of high blood pressure. Indeed, it is usually curable if properly diagnosed and treated, but it may be fatal if undiagnosed or mistreated. Postmortem series indicate that most pheochromocytomas are unsuspected clinically, even when the tumor is related to the fatal outcome.

PATHOLOGY

Location and Morphology In adults, approximately 80% of pheochromocytomas are unilateral and solitary, 10% are bilateral, and 10% are extraadrenal. In children, a fourth of tumors are bilateral, and an additional fourth are extraadrenal. Solitary lesions inexplicably favor the right side. Although pheochromocytomas may grow to large size (over 3 kg), most weigh less than 100 g and are less than 10 cm in diameter. The tumors are highly vascular.

The tumors are made up of large, polyhedral, pleomorphic chromaffin cells. Less than 10% of these tumors are malignant. As with other endocrine tumors, malignancy cannot be determined from the histologic appearance; tumors that contain large numbers of aneuploid or tetraploid cells, as determined by flow cytometry, are more likely to recur. Local invasion of surrounding tissues or distant metastases indicate malignancy.

Extraadrenal Pheochromocytomas Extraadrenal pheochromocytomas usually weigh 20 to 40 g and are <5 cm in diameter. Most are located within the abdomen in association with the celiac, superior mesenteric, and inferior mesenteric ganglia. Approximately 10% are in the thorax, 1% are within the urinary bladder, and <3% are in the neck, usually in association with the sympathetic ganglia or the extracranial branches of the ninth or tenth cranial nerves.

Catecholamine Synthesis, Storage, and Release Pheochromocytomas synthesize and store catecholamines by processes resembling those of the normal adrenal medulla (Chap. 72). Little is known about the mechanisms of catecholamine release from pheochromocytomas, but changes in blood flow and necrosis within the tumor may be the cause in some instances. These tumors are not innervated, and catecholamine release does not result from neural stimulation. Pheochromocytomas also store and secrete a variety of peptides, including endogenous opioids, adrenomedullin,

endothelin, erythropoietin, parathyroid hormone-related protein, neuropeptide Y, and chromagranin A (Chap. 72). These peptides contribute to the clinical manifestations in selected cases, as noted below.

Epinephrine, Norepinephrine, and Dopamine Most pheochromocytomas contain and secrete both norepinephrine and epinephrine, and the percentage of norepinephrine is usually greater than in the normal adrenal. Most extraadrenal pheochromocytomas secrete norepinephrine exclusively. Rarely, pheochromocytomas produce epinephrine alone, particularly in association with multiple endocrine neoplasia (MEN). Although epinephrine-producing tumors may cause a preponderance of metabolic and beta-receptor effects, in general the major catecholamine secreted cannot be predicted from the clinical presentation. Increased production of dopamine and homovanillic acid (HVA) is uncommon with benign lesions but may occur with malignant pheochromocytoma.

FAMILIAL PHEOCHROMOCYTOMA

In approximately 5% of cases, pheochromocytoma is inherited as an autosomal dominant trait either alone or in combination with other abnormalities such as MEN type 2a (Sipple's syndrome) or type 2b (mucosal neuroma syndrome) (Chap. 339), von Hippel-Lindau's retinal cerebellar hemangioblastomosis, or von Recklinghausen's neurofibromatosis. Bilateral adrenal pheochromocytomas are common in the familial syndromes; within MEN kindreds, over half of pheochromocytomas are bilateral. A familial syndrome should be suspected in any patient with bilateral pheochromocytomas.

GENETIC CONSIDERATIONS

Several molecular genetic abnormalities have been identified in the familial pheochromocytoma syndromes. The MEN2A and B syndromes are associated with abnormalities in the RET protooncogene located in pericentromeric region of chromosome 10 (Chap. 339). These mutations result in the constitutive activation of the receptor tyrosine kinase exposing adrenal medullary chromaffin cells and parafollicular C cells to hyperplasia and rendering them susceptible to malignant transformation. The RET mutations are located in the extracellular domain in MEN 2A and in the intracellular portion of the receptor in families with the MEN 2B syndrome. Interestingly, mutations at specific sites in the RET protooncogene are highly predictive of pheochromocytoma. The different phenotypic manifestations of the syndrome in different families, therefore, reflect differences in the specific mutations.

In the von Hippel-Landau (VHL) syndrome, mutation of one copy of the VHL tumor suppressor gene is associated with the development of tumors characteristic of the syndrome including pheochromocytomas. Loss of function of the VHL tumor suppressor gene promotes tumor formation by mechanisms that are incompletely understood but may involve mRNA transcript elongation. In the VHL syndrome, the frequency of pheochromocytoma varies considerably in different kindreds. As in the MEN 2 syndromes, certain VHL mutations are highly associated with the development of pheochromocytoma. Of further interest is the recent finding that the VHL mutation has been identified in some kindreds with familial pheochromocytoma as the sole

manifestation without other clinical evidence of the VHL syndrome. Missense mutations, as opposed to deletions, insertions, or non-sense mutations, appear to be more commonly associated with pheochromocytoma. There is also a high incidence of germ-line VHL mutations in patients with thoracic extraadrenal pheochromocytomas.

Interestingly, neither the RET protooncogene nor the VHL mutation occurs commonly as a somatic mutation in sporadic pheochromocytomas. Screening apparently sporadic cases, however, may uncover a germ-line mutation and lead to the identification of an involved family that was unsuspected on clinical grounds.

CLINICAL FEATURES

Pheochromocytoma occurs at all ages but is most common in young to midadult life. Some series show a slight female preponderance. Most patients come to medical attention as a result of hypertensive crisis, paroxysmal symptoms suggestive of seizure disorder or anxiety attacks, or hypertension that responds poorly to conventional treatment. Less commonly, unexplained hypotension or shock in association with surgery or trauma will suggest the diagnosis. Most patients have hypertension in association with headaches, excessive sweating, and/or palpitations.

Hypertension Hypertension is the most common manifestation. In approximately 60% of cases the hypertension is sustained, although significant blood pressure lability is usually present, and half of patients with sustained hypertension have distinct crises or paroxysms. The other 40% have blood pressure elevations only during an attack. The hypertension is often severe, occasionally malignant, and may be resistant to treatment with standard antihypertensive drugs.

Paroxysms or Crises The paroxysm or crisis occurs in over half of patients. In an individual patient, the symptoms are often similar with each attack. The paroxysms may be frequent or sporadic, occurring at intervals as long as weeks or months. With time, the paroxysms usually increase in frequency, duration, and severity.

The attack usually has a sudden onset. It may last from a few minutes to several hours or longer. Headache, profuse sweating, palpitations, and apprehension, often with a sense of impending doom, are common. Pain in the chest or abdomen may be associated with nausea and vomiting. Either pallor or flushing may occur during the attack. The blood pressure is elevated, often to alarming levels, and the elevation is usually accompanied by tachycardia.

The paroxysm may be precipitated by any activity that displaces the abdominal contents. In some cases a particular stimulus may induce an attack in a characteristic fashion, but in others no clearly defined precipitating event can be found. Although anxiety may accompany the attacks, mental or psychological stress does not usually provoke a crisis.

Other Distinctive Clinical Features Symptoms and signs of an increased metabolic rate, such as profuse sweating and mild to moderate weight loss, are common. Orthostatic hypotension is a consequence of diminished plasma volume and blunted sympathetic reflexes. Both these factors predispose the patient with unsuspected

pheochromocytoma to hypotension or shock during surgery or trauma. Secretion of the hypotensive peptide adrenomedullin may contribute to the hypotension in some patients.

Cardiac Manifestations Sinus tachycardia, sinus bradycardia, supraventricular arrhythmias, and ventricular premature contractions all have been noted. Angina and acute myocardial infarction may occur even in the absence of coronary artery disease. A catecholamine-induced increase in myocardial oxygen consumption and, perhaps, coronary spasm may play a role in these ischemic events. Electrocardiographic changes, including nonspecific ST-T wave changes, prominent U waves, left ventricular strain patterns, and right and left bundle branch blocks may be present in the absence of demonstrable ischemia or infarction. Cardiomyopathy, either congestive with myocarditis and myocardial fibrosis or hypertrophic with concentric or asymmetric hypertrophy, may be associated with heart failure and cardiac arrhythmias. Multiorgan system failure with noncardiogenic pulmonary edema may be the presenting manifestation. Elevated levels of amylase originating from damaged pulmonary endothelium and abdominal pain may suggest acute pancreatitis, although serum lipase levels are normal.

Carbohydrate Intolerance Over half of patients have impaired carbohydrate tolerance due to suppression of insulin and stimulation of hepatic glucose output. The impaired glucose tolerance rarely requires treatment with insulin and disappears after removal of the tumor.

Hematocrit The elevated hematocrit is secondary to diminished plasma volume. Rarely, production of erythropoietin by the tumor may cause a true erythrocytosis.

Other Manifestations Hypercalcemia has been attributed to the ectopic secretion of parathyroid hormone-related protein. Fever and an elevated erythrocyte sedimentation rate have been reported in association with the production of interleukin 6. Elevated temperature more commonly reflects catecholamine-mediated increases in metabolic rate and diminished heat dissipation secondary to vasoconstriction. Polyuria is an occasional finding, and rhabdomyolysis with myoglobinuric renal failure may result from extreme vasoconstriction with muscle ischemia.

Pheochromocytoma of the Urinary Bladder Pheochromocytoma in the wall of the urinary bladder may result in typical paroxysms in relation to micturition. The location in the bladder wall is responsible for the occurrence of symptoms while the tumors are quite small, and, consequently, catecholamine excretion may be normal or minimally elevated. Hematuria is present in over half of patients, and the tumor can often be visualized at cystoscopy.

Adverse Drug Interactions Severe and occasionally fatal paroxysms have been induced by opiates, histamine, adrenocorticotropin, saralasin, and glucagon. These agents appear to release catecholamines directly from the tumor. Indirect-acting sympathomimetic amines, including methyldopa (when administered intravenously), may cause an increase in blood pressure by releasing catecholamines from the augmented stores within nerve endings. Drugs that block neuronal uptake of catecholamines, such as tricyclic antidepressants or guanethidine, may enhance the

physiologic effects of circulating catecholamines. Indeed, all medications should be considered carefully and administered cautiously in patients with known or suspected pheochromocytoma.

Associated Diseases Pheochromocytoma is associated with medullary carcinoma of the thyroid in the MEN syndrome types 2a and 2b and with hyperparathyroidism in MEN 2a (Chap. 339). Hypercalcemia, resolving after tumor resection, also has been described in the absence of parathyroid disease, as described above. Individuals at risk for MEN 2a and 2b should be screened periodically for pheochromocytoma by assay of a 24-h urine sample for catecholamines, including measurement of epinephrine. Pheochromocytoma should be excluded or removed before thyroid or parathyroid surgery.

The association of pheochromocytoma and neurofibromatosis is not common. Nevertheless, since incomplete forms of neurofibromatosis may be associated with pheochromocytoma, minor manifestations such as cafe au lait spots, vertebral abnormalities, or kyphoscoliosis should increase the suspicion of pheochromocytoma in a patient with hypertension. The incidence of pheochromocytoma in some kindreds with von Hippel-Lindau disease may be as high as 10 to 25%. Many of these are unsuspected clinically and diagnosed on a computed tomography (CT) scan or at postmortem.

The incidence of cholelithiasis is 15 to 20%. Cushing's syndrome is a rare association, usually a consequence of ectopic secretion of adrenocorticotropic hormone by the pheochromocytoma or, less commonly, by a coexistent medullary carcinoma of the thyroid.

Diagnosis The diagnosis is established by the demonstration of increased excretion of catecholamines or catecholamine metabolites. The diagnosis can usually be made by the analysis of a single 24-h urine sample, provided the patient is hypertensive or symptomatic at the time of collection.

Biochemical Tests The assays employed include those for vanillylmandelic acid (VMA), the metanephrines, and unconjugated or "free" catecholamines (Chap. 72). The VMA assay is both less sensitive and less specific than assays of metanephrines or catecholamines. Accuracy of diagnosis is improved when two of three determinations are employed. The following considerations apply to all the urinary tests: (1) Despite claims for the adequacy of determinations made on random urine samples, analysis of a full 24-h urine sample is preferable. Creatinine should also be determined to assess the adequacy of collection. (2) Where possible, the collection should be made when the patient is at rest, on no medication, and without recent exposure to radiographic contrast media. When it is not practical to discontinue all medications, drugs known specifically to interfere with these assays (as noted below) should be avoided. (3) The urine should be acidified and refrigerated during and after collection. (4) With high-quality assays, dietary restrictions are minimal and should be specified by the laboratory performing the analyses. (5) Although most patients with pheochromocytoma excrete increased amounts of catecholamines and catecholamine metabolites at all times, the yield is increased in patients with paroxysmal hypertension if a 24-h urine collection is initiated during a crisis.

Free Catecholamines The upper limit of normal for total urinary catecholamines is between 590 and 885 nmol (100 and 150 ug) per 24 h. In most patients with pheochromocytoma, values in excess of 1480 nmol (250 ug) per day are obtained. Measurement of epinephrine is often of value, since increased epinephrine excretion [over 275 nmol (50 ug) per 24 h] is usually due to an adrenal lesion and may be the only abnormality in cases associated with MEN. False-positive increases in catecholamine excretion result from exogenous catecholamines and related drugs such as methyldopa, levodopa, labetalol, and sympathomimetic amines, which may elevate catecholamine excretion for up to 2 weeks. Endogenous catecholamines from stimulation of the sympathoadrenal system also may increase urinary catecholamine excretion. Relevant clinical situations that cause such increases include hypoglycemia, strenuous exertion, central nervous system disease with increased intracranial pressure, severe hypoxia, and clonidine withdrawal.

Metanephrines and VMA In most laboratories, the upper limit of normal is 7 umol (1.3 mg) of total metanephrines and 35 umol (7.0 mg) of VMA excretion per 24 h. In most patients with pheochromocytoma, the increase in these urinary metabolites is considerable, often to more than three times the normal range. Metanephrine excretion is increased by exogenous and endogenous catecholamines and by treatment with monoamine oxidase inhibitors; propranolol may cause a spurious increase in metanephrine excretion, since a propranolol metabolite interferes in the commonly used spectrophotometric assay. VMA is less affected by endogenous and exogenous catecholamines but is spuriously increased by a variety of drugs, including carbidopa. VMA excretion is decreased by monoamine oxidase inhibitors.

Plasma Catecholamines Measurement of plasma catecholamines has a limited application. The care required in obtaining basal levels (Chap. 72) and the satisfactory results with urinary determinations make measurement of plasma catecholamines unnecessary in most cases. Plasma catecholamine levels are affected by the same drugs and physiologic perturbations that increase urinary catecholamine excretion. In addition, a- and b-adrenergic receptor blocking agents may elevate plasma catecholamines by impairing clearance.

When the clinical features suggest pheochromocytoma and the urinary assay results are borderline, measurement of plasma catecholamines may be worthwhile. Markedly elevated basal levels of total catecholamines support the diagnosis, although approximately one-third of patients with pheochromocytoma have normal or slightly elevated basal values. The usefulness of plasma catecholamine determinations may be increased by agents that suppress sympathetic nervous system activity. Clonidine and ganglionic blocking agents (Chap. 72) reduce plasma catecholamine levels in normal subjects and in patients with essential hypertension. These drugs have little effect on catecholamine levels in patients with pheochromocytoma. In patients with elevated or borderline basal catecholamine values, failure to suppress plasma or urinary levels with clonidine supports the diagnosis of pheochromocytoma.

Pharmacologic Tests Reliable methods for the measurement of catecholamines and catecholamine metabolites in urine have rendered obsolete both the provocative and adrenolytic tests, which are nonspecific and entail considerable risk. A modified version

of the adrenolytic test may be of some use, however, as a therapeutic trial in a patient in hypertensive crisis with features suggestive of pheochromocytoma. A positive response to phentolamine (5-mg bolus following a test dose of 0.5 mg) is a reduction in blood pressure of at least 35/25 mmHg that peaks after 2 min and persists for 10 to 15 min. The pharmacologic response is never diagnostic, and biochemical confirmation is essential. Provocative tests in normotensive patients are potentially dangerous and rarely indicated. However, a glucagon provocative test may be of use in patients with paroxysmal hypertension and nondiagnostic basal catecholamine levels. Glucagon has a negligible effect on blood pressure or plasma catecholamine levels in normal or hypertensive subjects. In patients with pheochromocytoma, on the other hand, glucagon may increase both blood pressure and circulating catecholamine levels. The elevation in plasma catecholamine concentration, moreover, may occur without a blood pressure response. It must be emphasized, however, that life-threatening pressor crises have occurred after administration of glucagon to patients with pheochromocytoma, so the test should never be performed casually. Careful continuous monitoring of the blood pressure is required, intravenous access must be adequate, and phentolamine must be at hand to terminate the test if a significant pressor reaction ensues.

Differential Diagnosis Since the manifestations of pheochromocytoma can be protean, the diagnosis must be considered and excluded in many patients with suggestive clinical features. In patients with essential hypertension and "hyperadrenergic" features such as tachycardia, sweating, and increased cardiac output, and in patients with anxiety attacks associated with blood pressure elevations, analysis of a 24-h urine collection is usually decisive in excluding the diagnosis. Repeated determinations on urine collected during attacks may be necessary, however, before the diagnosis can be excluded with certainty. The clonidine suppression and glucagon stimulation tests may be helpful in excluding the diagnosis in difficult cases. Pressor crises associated with clonidine withdrawal and the use of cocaine or monoamine oxidase inhibitors (Chap. 72) may mimic the paroxysms of pheochromocytoma. Factitious crises may be produced by self-administration of sympathomimetic amines in psychiatrically disturbed patients.

Intracranial lesions, particularly posterior fossa tumors or subarachnoid hemorrhage, may cause hypertension and increased excretion of catecholamines or catecholamine metabolites. While this is most common in patients with an obvious neurologic catastrophe, the possibility of subarachnoid or intracranial hemorrhage secondary to pheochromocytoma should be considered. Diencephalic or autonomic epilepsy may be associated with paroxysmal spells, hypertension, and increased plasma catecholamine levels. This rare entity may be difficult to distinguish from pheochromocytoma, but an aura, an abnormal electroencephalogram, and a beneficial response to anticonvulsant medications will often suggest the proper diagnosis.

TREATMENT

Preoperative Management The induction of stable a-adrenergic blockade is the basis of preoperative management and provides the foundation for successful surgical treatment. Once the diagnosis is established, the patient should be placed on phenoxybenzamine to induce a long-lived, noncompetitivea-receptor blockade. The usual initial dose is 10 mg every 12 h with increments of 10 to 20 mg added every few days until the blood pressure is controlled and the paroxysms disappear. Because of the

long duration of action, the therapeutic effects are cumulative, and the optimal dose must be achieved gradually with careful monitoring of supine and upright blood pressures. Most patients require between 40 and 80 mg phenoxybenzamine per day, although 200 mg or more may be necessary. Phenoxybenzamine should be administered for at least 10 to 14 days prior to surgery. Over this time, the combination of a-receptor blockade and a liberal salt intake will restore the contracted plasma volume to normal. Before adequatea-adrenergic blockade with phenoxybenzamine is achieved, paroxysms may be treated with oral prazosin or noncompetitive intravenous phentolamine. Selectivea$_1$antagonists have been employed for preoperative preparation, but their role in preparative management should be limited to the treatment of individual paroxysms. They may be useful as antihypertensive agents in patients with suspected pheochromocytoma while workup is in progress, since they are usually better tolerated than phenoxybenzamine and will prevent serious pressor crises if pheochromocytoma is present. Nitroprusside, calcium channel blocking agents, and possibly angiotensin-converting enzyme inhibitors reduce blood pressure in patients with pheochromocytoma. Nitroprusside may also be useful in the treatment of pressor crises.

b-Adrenergic receptor blocking agents should be given only after alpha blockade has been induced, since administration of such agents by themselves may cause a paradoxic increase in blood pressure by antagonizing beta-mediated vasodilation in skeletal muscle. Beta blockade is usually initiated when tachycardia develops during the induction of a-adrenergic blockade. Low doses often suffice, and a reasonable starting dose is 10 mg propranolol three to four times per day, increased as needed to control the pulse rate. Beta blockade is effective for catecholamine-induced arrhythmias, particularly those potentiated by anesthetic agents.

Preoperative Localization of the Tumor Surgical removal of pheochromocytoma is facilitated if the location of the tumor or tumors can be established preoperatively. Once pheochromocytoma is diagnosed, localization should be undertaken while the patient is being prepared for surgery.CT or magnetic resonance imaging (MRI) of the adrenals is usually successful in identifying intraadrenal lesions. Extraadrenal tumors within the chest can frequently be identified by conventional chest films or CT. MRI is useful in identifying extraadrenal tumors in the abdomen. If these studies are negative, abdominal aortography (once a-adrenergic blockade is complete) may identify extraadrenal pheochromocytomas in the abdomen, since these lesions are often supplied by a large aberrant artery. If aortography, CT, and MRI fail to localize the lesion, venous sampling at different levels of the inferior and superior vena cava may reveal catecholamine gradients in the region drained by the tumor; this area may then be restudied by selective angiography or scanning by CT or MRI. An additional localization technique involves a radionuclide scintiscan after administration of the radiopharmaceutical [131I]metaiodobenzylguanidine (MIBG). This agent is concentrated by the amine uptake process and produces an external scintigraphic image at the site of the tumor. This type of scanning may be useful in characterizing lesions discovered by CT when biochemical confirmation is indeterminate, as well as in localizing extraadrenal pheochromocytomas. Percutaneous fine-needle aspiration of chromaffin tumors is contraindicated; indeed, pheochromocytoma should be considered before any adrenal lesions are aspirated.

Surgery Surgical treatment of pheochromocytoma is best performed in centers with experience in the preoperative, anesthetic, and intraoperative management of pheochromocytoma. In experienced hands, surgical mortality is<2 or 3%.

Monitoring during the surgical procedure should include continuous recording of arterial pressure and central venous pressure as well as electrocardiography; in the presence of cardiac disease or if congestive failure has been present, pulmonary capillary wedge pressure should be monitored. Adequate fluid replacement is crucial. Intraoperative hypotension responds better to volume replacement than to vasoconstrictors. Hypertension and cardiac arrhythmias are most likely during induction of anesthesia, intubation, and manipulation of the tumor. Intravenous phentolamine is usually sufficient to control the blood pressure, but nitroprusside may be required. Propranolol may be given in the treatment of tachycardia or ventricular ectopy.

Pheochromocytoma in Pregnancy Spontaneous labor and vaginal delivery in unprepared patients are usually disastrous for mother and fetus. In early pregnancy, the patient should be prepared with phenoxybenzamine, and the tumor should be removed as soon as the diagnosis is confirmed. The pregnancy need not be terminated, but the operative procedure itself may result in spontaneous abortion. In the third trimester, treatment with adrenergic blocking agents should be undertaken; when the fetus is of sufficient size, cesarean section may be followed by extirpation of the tumor. Although the safety of adrenergic blocking drugs in pregnancy is not established, these agents have been administered in several cases without obvious adverse effect. Antepartum diagnosis and treatment lowers the maternal death rate to that approaching nonpregnant pheochromocytoma patients; fetal death rate, however, remains elevated.

Unresectable and Malignant Tumors In cases of metastatic or locally invasive tumor in patients with intercurrent illness that precludes surgery, long-term medical management is required. When the manifestations cannot be adequately controlled by adrenergic blocking agents, the concomitant administration of metyrosine may be required. This agent inhibits tyrosine hydroxylase, diminishes catecholamine production by the tumor, and often simplifies chronic management. Malignant pheochromocytoma frequently recurs in the retroperitoneum, and it metastasizes most commonly to bone and lung. Although these malignant tumors are resistant to radiotherapy, combination chemotherapy has had limited success in controlling them. Use of$_{131}$I-MIBG has had limited success in the treatment of malignant pheochromocytoma, due to poor uptake of the radioligand.

PROGNOSIS AND FOLLOW-UP

The 5-year survival rate after surgery is usually over 95%, the recurrence rate is<10%. After successful surgery, catecholamine excretion returns to normal in about 2 weeks and should be measured to ensure complete tumor removal. Catecholamine excretion should be assessed at the reappearance of suggestive symptoms or yearly if the patient remains asymptomatic. For malignant pheochromocytoma, the 5-year survival rate is<50%.

Complete removal cures the hypertension in approximately three-fourths of patients. In the remainder, hypertension recurs but is usually well controlled by standard

antihypertensive agents. In this group, either underlying essential hypertension or irreversible vascular damage induced by catecholamines may cause the persistence of the hypertension.

(Bibliography omitted in Palm version)

[Back to Table of Contents](#)

333. DIABETES MELLITUS - *Alvin C. Powers*

Diabetes mellitus (DM) comprises a group of common metabolic disorders that share the phenotype of hyperglycemia. Several distinct types of DM exist and are caused by a complex interaction of genetics, environmental factors, and life-style choices. Depending on the etiology of the DM, factors contributing to hyperglycemia may include reduced insulin secretion, decreased glucose usage, and increased glucose production. The metabolic dysregulation associated with DM causes secondary pathophysiologic changes in multiple organ systems that impose a tremendous burden on the individual with diabetes and on the health care system. In the United States, DM is the leading cause of end-stage renal disease, nontraumatic lower extremity amputations, and adult blindness. With an increasing incidence worldwide, DM will likely continue to be a leading cause of morbidity and mortality for the foreseeable future.

CLASSIFICATION

Recent advances in the understanding of the etiology and pathogenesis of diabetes have led to a revised classification (Table 333-1). Although all forms of DM are characterized by hyperglycemia, the pathogenic mechanisms by which hyperglycemia arises differ widely. Some forms of DM are characterized by an absolute insulin deficiency or a genetic defect leading to defective insulin secretion, whereas other forms share insulin resistance as their underlying etiology. Recent changes in classification reflect an effort to classify DM on the basis of the pathogenic process that leads to hyperglycemia, as opposed to criteria such as age of onset or type of therapy (Fig. 333-1).

The two broad categories of DM are designated type 1 and type 2. Type 1A DM results from autoimmune beta cell destruction, which usually leads to insulin deficiency. Type 1B DM is also characterized by insulin deficiency as well as a tendency to develop ketosis. However, individuals with type 1B DM lack immunologic markers indicative of an autoimmune destructive process of the beta cells. The mechanisms leading to beta cell destruction in these patients are unknown. Relatively few patients with type 1 DM fall into the type 1B idiopathic category; many of these individuals are either African-American or Asian in heritage.

Type 2 DM is a heterogeneous group of disorders usually characterized by variable degrees of insulin resistance, impaired insulin secretion, and increased glucose production. Distinct genetic and metabolic defects in insulin action and/or secretion give rise to the common phenotype of hyperglycemia in type 2 DM (see below). The identification of distinct pathogenic processes in type 2 DM has important potential therapeutic implications, as pharmacologic agents that target specific metabolic derangements become available.

Two features of the current classification of DM diverge from previous classifications. First, the terms *insulin-dependent diabetes mellitus* (IDDM) and *noninsulin-dependent diabetes mellitus* (NIDDM) are obsolete. These previous designations reflected the observation that most individuals with type 1 DM (previously IDDM) have an absolute requirement for insulin treatment, whereas many individuals with type 2 DM (previously NIDDM) do not require insulin therapy to prevent ketoacidosis. However, because many

individuals with type 2 DM eventually require insulin treatment for control of glycemia, the use of the latter term generated considerable confusion.

A second difference is that age is no longer used as a criterion in the new classification system. Although type 1 DM most commonly develops before the age of 30, an autoimmune beta cell destructive process can develop at any age. In fact, it is estimated that between 5 and 10% of individuals who develop DM after age 30 have type 1A DM. Likewise, although type 2 DM more typically develops with increasing age, it also occurs in children, particularly in obese adolescents.

OTHER TYPES OF DM

Other etiologies for DM include specific genetic defects in insulin secretion or action, metabolic abnormalities that impair insulin secretion, and a host of conditions that impair glucose tolerance (Table 333-1). *Maturity onset diabetes of the young* (MODY) is a subtype of DM characterized by autosomal dominant inheritance, early onset of hyperglycemia, and impairment in insulin secretion (discussed below). Mutations in the insulin receptor cause a group of rare disorders characterized by severe insulin resistance.

DM can result from pancreatic exocrine disease when the majority of pancreatic islets (>80%) are destroyed. Several endocrinopathies can lead to DM as a result of excessive secretion of hormones that antagonize the action of insulin. Notable within this group are acromegaly and Cushing's disease, both of which may present with DM. Viral infections have been implicated in pancreatic islet destruction, but are an extremely rare cause of DM. Congenital rubella greatly increases the risk for DM; however, most of these individuals also have immunologic markers indicative of autoimmune beta cell destruction.

GESTATIONAL DIABETES MELLITUS (GDM)

Glucose intolerance may develop and first become recognized during pregnancy. Insulin resistance related to the metabolic changes of late pregnancy increases insulin requirements and may lead to hyperglycemia or impaired glucose tolerance. GDM is seen in approximately 4% of pregnancies in the United States; most women revert to normal glucose tolerance post-partum but have a substantial risk (30 to 60%) of developing DM later in life.

EPIDEMIOLOGY

The worldwide prevalence of DM has risen dramatically over the past two decades. It is projected that the number of individuals with DM will continue to increase in the near future. Between 1976 and 1994, for example, the prevalence of DM among adults in the United States increased from 8.9% to 12.3%. These findings, based on national epidemiologic data, include individuals with a diagnosis of DM and those with undiagnosed DM (based on identical diagnostic criteria). Likewise, prevalence rates of impaired fasting glucose (IFG) increased from 6.5% to 9.7% over the same period. Although the prevalence of both type 1 and type 2 DM is increasing worldwide, the prevalence of type 2 DM is expected to rise more rapidly in the future because of

increasing obesity and reduced activity levels.

There is considerable geographic variation in the incidence of both type 1 and type 2 DM. For example, Scandinavia has the highest rate of type 1 DM (in Finland, incidence is 35/100,000 per year). The Pacific Rim has a much lower rate (in Japan and China, incidence is 1 to 3/100,000 per year) of type 1 DM; Northern Europe and the United States share an intermediate rate (8 to 17/100,000 per year). Much of the increased risk of type 1 DM is believed to reflect the frequency of high-risk HLA alleles among ethnic groups in different geographic locations.

The prevalence of type 2 DM and its harbinger, impaired glucose tolerance (IGT), is highest in certain Pacific islands, intermediate in countries such as India and the United States, and relatively low in Russia and China. This variability is likely due to both genetic and environmental factors. There is also considerable variation in DM prevalence among different ethnic populations within a given country.

In 1998, approximately 16 million individuals in the United States met the diagnostic criteria for DM. This represents ~6% of the population. About 800,000 individuals in the United States develop DM each year. The vast majority of these (>90%) have type 2 DM. The number of people with DM increases with the age of the population, ranging from an incidence of ~1.5% in individuals from 20 to 39 years to ~20% of individuals>75 years. The incidence of DM is similar in men and women throughout most age ranges but is slightly greater in men>60 years. The prevalence of DM is approximately twofold greater in African Americans, Hispanic Americans, and Native Americans than in non-Hispanic whites, and the onset of type 2 DM occurs, on average, at an earlier age in the former groups than in the non-Hispanic white population. The incidence of type 2 DM in these ethnic groups is rapidly increasing. The reasons for these differences are not yet clear.

DIAGNOSIS

Revised criteria for diagnosing DM have been issued by consensus panels of experts from the National Diabetes Data Group and the World Health Organization (Table 333-2). The revised criteria reflect new epidemiologic and metabolic evidence and are based on the following premises: (1) the spectrum of fasting plasma glucose (FPG) and the response to an oral glucose load varies in normal individuals, and (2) DM defined as the level of glycemia at which diabetes-specific complications are noted and not on the level of glucose tolerance from a population-based viewpoint. For example, the prevalence of retinopathy in Native Americans (Pima Indian population) begins to increase at a FPG > 6.4 mmol/L (116 mg/dL) (Fig. 333-2).

Glucose tolerance is classified into three categories based on the FPG: (1) FPG < 6.1 mmol/L (110 mg/dL) is considered normal; (2) FPG ³ 6.1 mmol/L (110 mg/dL) but< 7.0 mmol/L (126 mg/dL) is defined as IFG; and (3) FPG ³ 7.0 mmol/L (126 mg/dL) warrants the diagnosis of DM. IFG is a new diagnostic category defined by the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. It is analogous to IGT, which is defined as plasma glucose levels between 7.8 and 11.1 mmol/L (140 and 200 mg/dL) 2 h after a 75-g oral glucose load (Table 333-2). Individuals with IFG or IGT are at substantial risk for developing type 2 DM and cardiovascular disease in the future,

though they may not meet the criteria for DM.

The revised criteria for the diagnosis of DM emphasize the FPG as the most reliable and convenient test for diagnosing DM in asymptomatic individuals. A random plasma glucose concentration $\geq$11.1 mmol/L (200 mg/dL) accompanied by classic symptoms of DM (polyuria, polydipsia, weight loss) is sufficient for the diagnosis of DM (Table 333-2). Oral glucose tolerance testing, although still a valid mechanism for diagnosing DM, is not recommended as part of routine screening.

Some investigators have advocated the hemoglobin A1c (HbA1c) as a diagnostic test for DM. Though there is a strong correlation between elevations in the plasma glucose and the HbA1c (discussed below), the relationship between the FPG and the HbA1c in individuals with normal glucose tolerance or mild glucose intolerance is less clear, and the test is not universally standardized or available.

The diagnosis of DM has profound implications for an individual from both a medical and financial standpoint. Thus, the health care provider must be certain that these criteria are completely satisfied before assigning the diagnosis of DM to an individual. The revised criteria also allow for the diagnosis of DM to be withdrawn in situations where the FPG no longer exceeds these criteria. Abnormalities on screening tests for diabetes should be repeated before making a definitive diagnosis of DM, unless acute metabolic derangements or a markedly elevated plasma glucose are present (Table 333-2).

SCREENING

Widespread use of the FPG as a screening test for type 2 DM is strongly encouraged because: (1) a large number of individuals who meet the current criteria for DM are unaware that they have the disorder, (2) epidemiologic studies suggest that type 2 DM may be present for up to a decade before diagnosis, and (3) as many as 50% of individuals with type 2 DM have one or more diabetes-specific complications at the time of their diagnosis. The Expert Committee suggests screening all individuals>45 years every 3 years and screening asymptomatic individuals with additional risk factors (Table 333-3) at an earlier age. In contrast to type 2 DM, it is rare for an individual to have a long asymptomatic period of hyperglycemia prior to the diagnosis of type 1 DM. A number of immunologic markers for type 1 DM are becoming available (discussed below), but their use is currently discouraged pending the identification of clinically beneficial interventions for individuals at high risk for developing type 1 DM.

INSULIN BIOSYNTHESIS, SECRETION, AND ACTION

BIOSYNTHESIS

Insulin is produced in the beta cells of the pancreatic islets. It is initially synthesized as a single-chain 86-amino-acid precursor polypeptide, preproinsulin. Subsequent proteolytic processing removes the aminoterminal signal peptide, giving rise to proinsulin. Proinsulin is structurally related to insulin-like growth factors I and II, which bind weakly to the insulin receptor (Chap. 327). Cleavage of an internal 31-residue fragment from proinsulin generates the C peptide and the A (21 amino acids) and B (30 amino acids) chains of insulin, which are connected by disulfide bonds. The mature insulin molecule

and C peptide are stored together and cosecreted from secretory granules in the beta cells. Because the C peptide is less susceptible than insulin to hepatic degradation, it is a useful a marker of insulin secretion and allows discrimination of endogenous and exogenous sources of insulin in the evaluation of hypoglycemia (Chap. 334). Human insulin is now produced by recombinant DNA technology; structural alterations at one or more residues are useful for modifying its physical and pharmacologic characteristics (see below).

SECRETION

Glucose is the key regulator of insulin secretion by the pancreatic beta cell, although amino acids, ketones, various nutrients, gastrointestinal peptides, and neurotransmitters also influence insulin secretion. Glucose levels>3.9 mmol/L (70 mg/dL) stimulate insulin synthesis, primarily by enhancing protein translation and processing, as well as inducing insulin secretion. Glucose stimulates insulin secretion through a series of regulatory steps that begin with transport into the beta cell by the GLUT2 glucose transporter (Fig. 333-3). Glucose phosphorylation by glucokinase is the rate-limiting step that controls glucose-regulated insulin secretion.

Further metabolism of glucose-6-phosphate via glycolysis generates ATP, which inhibits the activity of an ATP-sensitive K+channel. This channel is a complex of two separate proteins, one of which is the receptor for certain oral hypoglycemics (e.g., sulfonylureas, meglitinides); the other subunit is an inwardly rectifying K+ channel protein. Inhibition of this K+channel induces beta cell membrane depolarization, opening of voltage-dependent calcium channels (leading to an influx of calcium), and stimulation of insulin secretion. Careful studies of insulin secretory profiles reveal pulsatile pattern of hormone release, with small secretory bursts occurring about every 10 min, superimposed upon greater amplitude oscillations of about 80 to 150 min. Meals or other major stimuli of insulin secretion induce large (four- to fivefold increase versus baseline) bursts of insulin secretion that usually last for 2 to 3 h before returning to baseline. Derangements in these normal secretory patterns are one of the earliest signs of beta cell dysfunction inDM (see below).

ACTION

Once insulin is secreted into the portal vein, ~50% is removed and degraded by the liver. Unextracted insulin enters the systemic circulation and binds to its receptor in target sites. The insulin receptor belongs to the tyrosine kinase class of membrane-bound receptors (Chap. 327). Insulin binding to the receptor stimulates intrinsic tyrosine kinase activity, leading to receptor autophosphorylation and the recruitment of intracellular signaling molecules, such as insulin receptor substrates (IRS) 1 and 2 (Fig. 333-4). These and other adaptor proteins initiate a complex cascade of phosphorylation and dephosphorylation reactions, ultimately resulting in the widespread metabolic and mitogenic effects of insulin. As an example, activation of the phosphatidylinositol-3¢-kinase (PI-3 kinase) pathway stimulates translocation of glucose transporters (e.g., GLUT4) to the cell surface, an event that is crucial for glucose uptake by skeletal muscle and fat. Activation of other insulin receptor signaling pathways induces glycogen synthesis, protein synthesis, lipogenesis, and regulation of various genes in insulin-responsive cells.

Glucose homeostasis reflects a precise balance between hepatic glucose production and peripheral glucose uptake and utilization. Insulin is the most important regulator of this metabolic equilibrium, but the effects of other pathways including neural input, metabolic signals, and hormones (e.g., glucagon) result in integrated control of glucose supply and utilization (Chap. 334;Fig. 334-1). In the fasting state, low insulin levels promote hepatic gluconeogenesis and glycogenolysis to prevent hypoglycemia. Low insulin levels decrease glycogen synthesis, reduce glucose uptake in insulin-sensitive tissues, and promote mobilization of stored precursors. Reduced insulin levels are also permissive in allowing glucagon to stimulate glycogenolysis and gluconeogenesis by the liver and renal medulla. These processes are of critical importance to ensure an adequate glucose supply for the brain. Postprandially, a large glucose load elicits a rise in insulin and fall in glucagon, leading to a reversal of these processes. The major portion of postprandial glucose is utilized by skeletal muscle. Other tissues, most notably the brain, utilize glucose in an insulin-independent fashion.

PATHOGENESIS

TYPE 1 DM

Type 1ADMdevelops as a result of the synergistic effects of genetic, environmental, and immunologic factors that ultimately destroy the pancreatic beta cells. The temporal development of type 1A DM is shown schematically as a function of beta cell mass in Fig. 333-5. Individuals with a genetic susceptibility have normal beta cell mass at birth but begin to lose beta cells secondary to autoimmune destruction that occurs over months to years. This autoimmune process is thought to be triggered by an infectious or environmental stimulus and to be sustained by a beta cell-specific molecule. In the majority of individuals, immunologic markers appear after the triggering event but before diabetes becomes clinically overt. Beta cell mass then begins to decline, and insulin secretion becomes progressively impaired, although normal glucose tolerance is maintained. The rate of decline in beta cell mass varies widely among individuals, with some patients progressing rapidly to clinical diabetes and others evolving more slowly. Features of diabetes do not become evident until a majority of beta cells are destroyed (~80%). At this point, residual functional beta cells still exist but are insufficient in number to maintain glucose tolerance. The events that trigger the transition from glucose intolerance to frank diabetes are often associated with increased insulin requirements, as might occur during infections or puberty. Following the initial clinical presentation of type 1A DM, a "honeymoon" phase may ensue during which time glycemic control is achieved with modest doses of insulin or, rarely, insulin is not needed. However, this fleeting phase of endogenous insulin production from residual beta cells disappears as the autoimmune process destroys the remaining beta cells, and the individual becomes completely insulin deficient.

GENETIC CONSIDERATIONS

The genetic contributions to type 1ADMinvolve multiple genes. The development of the disease appears to require inheritance of a sufficient complement of genes to confer susceptibility to the disorder. The concordance of type 1A DM in identical twins ranges between 30 and 70%, indicating that additional modifying factors must be involved in

determining whether diabetes develops. The major susceptibility gene for type 1A DM is located in the HLA region on chromosome 6. Polymorphisms in the HLA complex appear to account for 40 to 50% of the genetic risk of developing type 1A DM. This region contains genes that encode the class II MHC molecules, which present antigen to helper T cells and thus are involved in initiating the immune response (Chaps. 305,306,307). The ability of class II MHC molecules to present antigen is dependent on the amino acid composition of their antigen-binding sites. Amino acid substitutions may influence the specificity of the immune response by altering the binding affinity of different antigens for the class II molecules.

Most individuals with type 1A DM have the HLA DR3 and/or DR4 haplotype. Refinements in genotyping of HLA loci have shown that the haplotypes DQA1*0301, DQB1*0302 and DQA1*501, DQB1*0201 have the strongest association with type 1A DM. These haplotypes are present in 40% of children with type 1A DM as compared to 2% of the normal U.S. population.

In addition to MHC class II associations, at least 17 different genetic loci may contribute susceptibility to type 1A DM. For example, polymorphisms in the promoter region of the insulin gene appear to account for ~10% of the predisposition to type 1A DM. Genes that confer protection against the development of the disease also exist. For example, the haplotype DQA1*0102, DQB1*0602 is present in 20% of the U.S. population but is extremely rare in individuals with type 1A DM (<1%).

Although type 1A DM is clearly associated with certain predisposing genotypes, most individuals with these haplotypes do not develop diabetes. In addition, most individuals with type 1A DM do not have a first-degree relative with this disorder. Nevertheless, the risk of developing type 1A DM for relatives of individuals with the disease is considerably higher compared to the risk for the general population.

Autoimmune Factors Although other islet cell types [alpha cells (glucagon-producing), delta cells (somatostatin-producing) or PP cells (pancreatic polypeptide-producing)] are functionally and embryologically similar to beta cells and express most of the same proteins as beta cells, they are inexplicably spared from the autoimmune process. Pathologically, the pancreatic islets are infiltrated with lymphocytes (in a process termed *insulitis*). After all beta cells are destroyed, the inflammatory process abates, the islets become atrophic, and immunologic markers disappear. Studies of the insulitis and autoimmune process in humans and animal models of type 1A DM (NOD mouse and BB rat) have identified the following abnormalities in both the humoral and cellular arms of the immune system: (1) islet cell autoantibodies; (2) activated lymphocytes in the islets, peripancreatic lymph nodes, and systemic circulation; (3) T lymphocytes that proliferate when stimulated with islet proteins; and (4) release of cytokines within the insulitis. Beta cells seem to be particularly susceptible to the toxic effect of some cytokines (tumor necrosis factor a, interferong, and interleukin 1). The precise mechanisms of beta cell death are not known but may involve formation of nitric oxide metabolites, apoptosis, and direct CD8+ T cell cytotoxicity. Islet autoantibodies are not thought to be involved in the destructive process, as these antibodies do not generally react with the cell surface of islet cells and are not capable of transferring diabetes mellitus to animals.

Pancreatic islet molecules targeted by the autoimmune process include insulin, glutamic

acid decarboxylase (GAD; the biosynthetic enzyme for the neurotransmitter GABA), ICA-512/IA-2 (homology with tyrosine phosphatases), and phogrin (insulin secretory granule protein). Other less clearly defined autoantigens include an islet ganglioside and carboxypeptidase H. With the exception of insulin, none of the autoantigens are beta cell specific, which raises the question of how the beta cells are selectively destroyed. Current theories favor initiation of an autoimmune process directed at one beta cell molecule, which then spreads to other islet molecules as the immune process destroys beta cells and creates a series of secondary autoantigens. The beta cells of individuals who develop type 1ADM do not differ from beta cells of normal individuals, since transplanted islets are destroyed by a recurrence of the autoimmune process of type 1A DM.

Immunologic Markers Islet cell autoantibodies (ICAs) are a composite of several different antibodies directed at pancreatic islet molecules such asGAD, insulin, IA-2/ICA512, and an islet ganglioside and serve as a marker of the autoimmune process of type 1ADM. Testing for ICAs can be useful in classifying the type of DM as type IA and in identifying nondiabetic individuals at risk for developing type 1A DM. ICAs are present in the majority of individuals (>75%) diagnosed with new-onset type 1A DM, in a significant minority of individuals with newly diagnosed type 2 DM, and occasionally in individuals withGDM(<5%). ICAs are present in 3 to 4% of first-degree relatives of individuals with type 1A DM. In conjunction with impaired insulin secretion on intravenous glucose tolerance testing, they predict a>50% risk of developing type 1A DM within 5 years. Without this impairment in insulin secretion, the presence of ICAs predicts a 5-year risk of <25%. Based on these data, the risk of a first-degree relative developing type 1A DM is relatively low, and even ICA-positive individuals are not destined to develop diabetes. At present, the ICAs are used predominantly as a research tool and not in clinical practice, in part because of the technically demanding nature of the assay but also because no treatments have been proven to prevent the occurrence or progression of type 1A DM.

Environmental Factors Numerous environmental events have been proposed to trigger the autoimmune process in genetically susceptible individuals; however, none have been conclusively linked to diabetes. Identification of an environmental trigger has been difficult because the event may precede the onset ofDM by several years (Fig. 333-5). Putative environmental triggers include viruses (coxsackie and rubella most prominently), early exposure to bovine milk proteins, and nitrosourea compounds. Epidemiologic studies have noted an association between bovine milk intake and type 1A DM; studies are ongoing to investigate a possible relationship between exposure to bovine milk and the autoimmune process of type 1A DM.

Prevention of Type 1A DM A number of interventions have successfully delayed or prevented diabetes in animal models. Some interventions have targeted the immune system directly (immunosuppression, selective T cell subset deletion, induction of immunologic tolerance to islet proteins), whereas others have prevented islet cell death by blocking cytotoxic cytokines or increasing islet resistance to the destructive process. Though results in animal models are promising, most of these interventions have not been successful in preventing type 1ADM in humans. Clinical trials of several interventions are underway in the United States and Europe. The Diabetes Prevention Trial -- type 1 is being conducted to determine whether administering insulin to

individuals at high risk for developing type 1A DM can induce immune tolerance and alter the autoimmune process of type 1A DM.

TYPE 2 DM

Type 2DM is a heterogeneous disorder with a complex etiology that develops in response to genetic and environmental influences. Central to the development of type 2 DM are insulin resistance and abnormal insulin secretion. Although controversy remains regarding the primary defect, most studies support the view that insulin resistance precedes insulin secretory defects.

GENETIC CONSIDERATIONS

Type 2DM has a strong genetic component. Although the major genes that predispose to this disorder have yet to be identified, it is clear that the disease is polygenic and multifactorial. Various genetic loci contribute to susceptibility, and environmental factors (such as nutrition and physical activity) further modulate phenotypic expression of the disease. The concordance of type 2 DM in identical twins is between 70 and 90%. Individuals with a parent with type 2 DM have an increased risk of diabetes; if both parents have type 2 DM, the risk in offspring may reach 40%. Insulin resistance, as demonstrated by reduced glucose utilization in skeletal muscle, is present in many nondiabetic, first-degree relatives of individuals with type 2 DM. However, definition of the genetic abnormalities of type 2 DM remains a challenge because the genetic defect in insulin secretion or action may not manifest itself unless an environmental event or another genetic defect, such as obesity, is superimposed.

The identification of individuals with mutations in various molecules involved in insulin action (e.g., the insulin receptor and enzymes involved in glucose homeostasis) has been useful for characterizing key steps in insulin action. However, mutations in these molecules account for a very small fraction of type 2DM.Likewise, genetic defects in proteins involved in insulin secretion have not been found in most individuals with type 2 DM. Genome-wide scanning for mutations or polymorphisms associated with type 2 DM is being used in an effort to identify genes associated with type 2 DM.

Pathophysiology Type 2DM is characterized by three pathophysiologic abnormalities: impaired insulin secretion, peripheral insulin resistance, and excessive hepatic glucose production. Obesity, particularly visceral or central, is very common in type 2 DM. Insulin resistance associated with obesity augments the genetically determined insulin resistance of type 2 DM. Adipocytes secrete a number of biologic products (leptin, tumor necrosis factor a, free fatty acids) that modulate processes such as insulin secretion, insulin action, and body weight and may contribute to the insulin resistance. In the early stages of the disorder, glucose tolerance remains normal, despite insulin resistance, because the pancreatic beta cells compensate by increasing insulin output. As insulin resistance and compensatory hyperinsulinemia progress, the pancreatic islets become unable to sustain the hyperinsulinemic state.IGT, marked by elevations in postprandial glucose, then develops. A further decline in insulin secretion and an increase in hepatic glucose production lead to overt diabetes with fasting hyperglycemia. Ultimately, beta cell failure may ensue.

Metabolic Abnormalities

Insulin Resistance This is caused by the decreased ability of insulin to act effectively on peripheral target tissues (especially muscle and liver) and is a prominent feature of type 2DM. This resistance is relative, since supernormal levels of circulating insulin will normalize the plasma glucose. Insulin dose-response curves exhibit a rightward shift, indicating reduced sensitivity, and a reduced maximal response, indicating an overall decrease in maximum glucose utilization (30 to 60% lower than normal individuals). Resistance to the action of insulin impairs glucose utilization by insulin-sensitive tissues and increases hepatic glucose output -- both effects contributing to the hyperglycemia of diabetes. Increased hepatic glucose output predominantly accounts for increasedFPGlevels, whereas decreased peripheral glucose usage results in postprandial hyperglycemia. In skeletal muscle, there is a greater impairment in nonoxidative glucose usage (glycogen formation) than in oxidative glucose metabolism through glycolysis. Glucose usage in insulin-independent tissues is not decreased in type 2 DM.

The precise molecular mechanism of insulin resistance in type 2DM has yet to be elucidated. Insulin receptor levels and tyrosine kinase activity in skeletal muscle are reduced, but these alterations are most likely secondary to hyperinsulinemia and are not a primary defect. Therefore, postreceptor defects are believed to play the predominant role in insulin resistance (Fig. 333-4). Polymorphisms inIRS-1 may be associated with glucose intolerance, raising the possibility that polymorphisms in various postreceptor molecules may combine to create an insulin-resistant state.

A current focus for the pathogenesis of insulin resistance focuses on a PI-3 kinase signaling defect, which causes reduced translocation of GLUT4 to the plasma membrane, among other abnormalities. Of note, not all insulin signal transduction pathways are resistant to the effects of insulin (e.g., those controlling cell growth and differentiation). Consequently, hyperinsulinemia may actually increase the insulin action through these pathways.

Another emerging theory proposes that elevated levels of free fatty acids, a common feature of obesity, may contribute to the pathogenesis of type 2DM in several different ways. Free fatty acids can impair glucose utilization in skeletal muscle, promote glucose production by the liver, and impair beta cell function.

Impaired Insulin Secretion Insulin secretion and sensitivity are interrelated (Fig. 333-6). In type 2DM, insulin secretion initially increases in response to insulin resistance in order to maintain normal glucose tolerance. Initially, the insulin secretory defect is mild and selectively involves glucose-stimulated insulin secretion. The response to other nonglucose secretagogues, such as arginine, is preserved. Eventually, the insulin secretory defect progresses to a state of grossly inadequate insulin secretion. Some endogenous insulin production continues, but the amount secreted is less than the amount secreted by normal individuals at the same plasma glucose concentration.

The reason(s) for the decline in insulin secretory capacity in type 2DM is unclear. Despite the assumption that a second genetic defect -- superimposed upon insulin resistance -- leads to beta cell failure, intense genetic investigation has so far excluded

mutations in islet candidate genes. Islet amyloid polypeptide or amylin is cosecreted by the beta cell and likely forms the amyloid fibrillar deposit found in the islets of individuals with longstanding type 2 DM. Whether such islet amyloid deposits are a primary or secondary event is not known. The metabolic environment may also impact islet function negatively. For example, chronic hyperglycemia paradoxically impairs islet function ("glucose toxicity") and leads to a worsening of hyperglycemia. Improvement in glycemic control is often associated with improved islet function. In addition, elevation of free fatty acid levels ("lipotoxicity") also worsens islet function.

Increased Hepatic Glucose Production The liver maintains plasma glucose during periods of fasting through glycogenolysis and gluconeogenesis using substrates derived from skeletal muscle and fat (alanine, lactate, glycerol, and fatty acids). Insulin promotes the storage of glucose as hepatic glycogen and suppresses gluconeogenesis. In type 2 DM, insulin resistance in the liver arises from the failure of hyperinsulinemia to suppress gluconeogenesis, which results in fasting hyperglycemia and decreased glucose storage by the liver in the postprandial state. Increased hepatic glucose production occurs early in the course of diabetes, though likely after the onset of insulin secretory abnormalities and insulin resistance in skeletal muscle.

Insulin Resistance Syndromes It is likely that the insulin resistance condition comprises a spectrum of disorders, with hyperglycemia representing one of the most readily diagnosed features. *Syndrome X* is a term used to describe a constellation of metabolic derangements that includes insulin resistance, hypertension, dyslipidemia, central or visceral obesity, endothelial dysfunction, and accelerated cardiovascular disease. Epidemiologic evidence supports hyperinsulinemia as a marker for coronary artery disease risk, though an etiologic role has not been demonstrated.

A number of forms of severe insulin resistance may be associated with a phenotype similar to that in type 2 DM or IGT (Table 333-1). *Acanthosis nigricans* and signs of hyperandrogenism (hirsutism, acne, and oligomenorrhea) are common physical features. In addition to rare genetic syndromes seen in early childhood, two distinct syndromes of severe insulin resistance have been described in adults: (1) type A, which affects young women and is characterized by severe hyperinsulinemia, obesity, and features of hyperandrogenism; and (2) type B, which affects middle-aged women and is characterized by severe hyperinsulinemia, features of hyperandrogenism, and autoimmune disorders. Individuals with the type A insulin resistance syndrome have an undefined defect in the insulin signaling pathway; individuals with the type B insulin resistance syndrome have autoantibodies directed at the insulin receptor. These receptor autoantibodies may block insulin binding or may stimulate the insulin receptor, leading to intermittent hypoglycemia.

Polycystic ovary syndrome (PCOS) is a common disorder that affects premenopausal women and is characterized by chronic anovulation and hyperandrogenism. Insulin resistance is seen in a significant subset of women with PCOS, and the disorder substantially increases the risk for type 2 DM, independent of the effects of obesity. Both metformin and thiazolidinediones may attenuate hyperinsulinemia, ameliorate hyperandrogenism, and induce ovulation, but are not approved for this indication.

Prevention Because type 2 DM is preceded by a period of IGT, a number of life-style

modifications and pharmacologic agents have been suggested to prevent or delay its onset. Individuals with a strong family history or those at high risk for developing DM should be strongly encouraged to maintain a normal body mass index and to engage in regular physical activity. Beyond this general advice, however, there are no specific interventions proven to prevent type 2 DM. Clinical trials of various interventions in individuals with IGT or early DM are underway in the United States and worldwide.

MODY: GENETICALLY DEFINED, MONOGENIC FORMS OF DIABETES MELLITUS

Several monogenic forms of DM have recently been identified. MODY comprises a phenotypically and genetically heterogeneous subtype of DM. Onset of the disease typically occurs between the ages of 10 and 25. Five different variants of MODY, due to mutations in genes encoding islet cell transcription factors or glucokinase (Fig. 333-3), have been identified so far, and all are transmitted as autosomal dominant disorders (Table 333-1). MODY 2, the most common variant, is caused by mutations in the glucokinase gene. Glucokinase catalyzes the formation of glucose-6-phosphate from glucose, a reaction that is important for glucose sensing by the beta cells and for glucose utilization by the liver. As a result of glucokinase mutations, higher glucose levels are required to elicit insulin secretory responses, thus altering the set point for insulin secretion. MODY 1, MODY 3, and MODY 5 are caused by mutations in the hepatocyte nuclear transcription factors HNF-4a, HNF-1a, and HNF-1b, respectively. As their names imply, these transcription factors are expressed in the liver but also in other tissues, including the pancreatic islets. The mechanisms by which such mutations lead to DM is not well understood, but it is likely that these factors affect islet development or the transcription of genes that are important in stimulating insulin secretion. MODY 4 is a rare variant caused by mutations in the insulin promoter factor (IPF-1), which is a transcription factor that regulates both pancreatic development and insulin gene transcription. Homozygous inactivating mutations lead to pancreatic agenesis, whereas heterozygous mutations result in early-onset DM. Studies of populations with type 2 DM suggest that mutations in the glucokinase gene and various islet cell transcription factors do not account for ordinary type 2 DM. Nevertheless, elucidation of the molecular genetics underlying these rare forms of DM has been important in identifying critical steps in the control of pancreatic beta cell function.

COMPLICATIONS OF DM

ACUTE COMPLICATIONS

Diabetic ketoacidosis (DKA) and nonketotic hyperosmolar state (NKHS) are acute complications of diabetes. DKA is seen primarily in individuals with type 1 DM, and NKHS is seen in individuals with type 2 DM. Both disorders are associated with absolute or relative insulin deficiency, volume depletion, and altered mental status. DKA and NKHS exist along a continuum of hyperglycemia, with or without ketosis. The metabolic similarities and differences in DKA and NKHS are highlighted in Table 333-4. Both disorders are associated with potentially serious complications if not promptly diagnosed and treated.

DIABETIC KETOACIDOSIS

Clinical Features The symptoms and physical signs of DKA are listed in Table 333-5. DKA may be the initial symptom complex that leads to a diagnosis of type 1 DM, but more frequently it occurs in individuals with established diabetes. Nausea and vomiting are often prominent, and their presence in an individual with diabetes warrants laboratory evaluation for DKA. Abdominal pain may be severe and sometimes suggests acute pancreatitis or ruptured viscous. Hyperglycemia leads to glucosuria, volume depletion, tachycardia, and possibly hypotension. Kussmaul respirations and an acetone odor on the patient's breath (both secondary to metabolic acidosis) are classic signs of the disorder. Lethargy and central nervous system depression may evolve into coma with severe DKA. Cerebral edema, an extremely serious complication of DKA, is seen most frequently in children. Signs of infection, which may precipitate DKA, should be sought on physical examination, even in the absence of fever.

Pathophysiology DKA results from insulin deficiency combined with counterregulatory hormone excess (glucagon, catecholamines, cortisol, and growth hormone). Both insulin deficiency and glucagon excess, in particular, are necessary for DKA to develop. The hyperglycemia of DKA results from increased hepatic glucose production (gluconeogenesis and glycogenolysis) and impaired peripheral glucose utilization. The decreased ratio of insulin to glucagon promotes gluconeogenesis, glycogenolysis, and ketone body formation in the liver, as well as increasing substrate delivery from fat and muscle (free fatty acids, amino acids) to the liver.

The combination of insulin deficiency and hyperglycemia reduces the hepatic level of fructose-2,6-phosphate, which alters the activity of phosphofructokinase and fructose-1,6-bisphosphatase. Glucagon excess decreases the activity of pyruvate kinase, whereas insulin deficiency increases the activity of phosphoenolpyruvate carboxykinase. These hepatic changes shift the handling of pyruvate toward glucose synthesis and away from glycolysis. Glycogenolysis is promoted by the increased levels of glucagon and catecholamines in the face of low insulin levels. Insulin deficiency also reduces levels of the GLUT4 glucose transporter, which impairs glucose uptake into skeletal muscle and fat and reduces intracellular glucose metabolism (Fig. 333-4).

Ketosis results from a marked increase in free fatty acid release from adipocytes, with a resulting shift toward ketone body synthesis in the liver. Reduced insulin levels, in combination with elevations in catecholamines and growth hormone, lead to an increase in lipolysis and release of free fatty acids. Normally, these free fatty acids are converted to triglycerides or very low density lipoproteins (VLDL) in the liver, but in DKA, hyperglucagonemia alters hepatic metabolism to favor ketone body formation, through activation of the enzyme carnitine palmitoyltransferase I. This enzyme is crucial for regulating fatty acid transport into the mitochondria, where beta oxidation and conversion to ketone bodies occurs. At physiologic pH, ketone bodies exist as ketoacids, which are neutralized by bicarbonate. As bicarbonate stores are depleted, metabolic acidosis ensues. Increased lactic acid production also contributes to the acidosis. The increased free fatty acids result in increased triglyceride production and increased hepatic production of VLDL. VLDL clearance is also reduced because the activity of insulin-sensitive lipoprotein lipase is decreased. Hypertriglyceridemia may be severe enough to cause pancreatitis.

DKA can be precipitated by inadequate levels of plasma insulin for a variety of reasons

([Table 333-5](#)). Most commonly, DKA is precipitated when relatively insufficient insulin is available when insulin requirements increase, as might occur during a concurrent illness. Failure to augment insulin therapy appropriately by the patient or health care team compounds the problem. Occasionally, complete omission of insulin by the patient or health care team (in a hospitalized patient with type 1[DM](#)) precipitates DKA. Patients using insulin infusion devices with short-acting insulin have a greater potential for DKA, since even a brief interruption in insulin delivery (e.g., mechanical malfunction) quickly leads to insulin deficiency.

Laboratory Abnormalities and Diagnosis The timely diagnosis of[DKA](#) is crucial and allows for prompt initiation of therapy. DKA is characterized by hyperglycemia, ketosis, and metabolic acidosis (increased anion gap) along with a number of secondary metabolic derangements ([Table 333-4](#))). Serum bicarbonate is frequently<10 mmol/L, and arterial pH ranges between 6.8 and 7.3, depending on the severity of the acidosis. Despite a total-body potassium deficit, the serum potassium at presentation is typically at the high end of the normal range or mildly elevated, secondary to the acidosis. Total-body stores of sodium, chloride, phosphorous, and magnesium are also reduced in DKA, but are not accurately reflected by their levels in the serum. Elevated blood urea nitrogen (BUN) and serum creatinine levels reflect intravascular volume depletion. Interference from acetoacetate may falsely elevate the serum creatinine measurement. Leukocytosis, hypertriglyceridemia, and hyperlipoproteinemia are commonly found as well. Hyperamylasemia may suggest a diagnosis of pancreatitis, especially when accompanied by abdominal pain. However, in DKA the amylase is usually of salivary origin and thus is not diagnostic of pancreatitis.

The measured serum sodium is reduced as a consequence of the hyperglycemia [1.6 meq (1.6 mmol/L) reduction in serum sodium for each 100 mg/dL (5.6 mmol/L) rise in the serum glucose]. A normal serum sodium in the setting of[DKA](#)indicates a more profound water deficit. In "conventional" units, the calculated serum osmolality [2 ´ (serum sodium+ serum potassium) + plasma glucose (mg/dL)/18 +[BUN](#)/2.8] is mildly to moderately elevated, though to a lesser degree than that found in[NKHS](#)hyperosmolar state (see below).

In[DKA](#), the ketone body, b-hydroxybutyrate, is synthesized at a threefold greater rate than acetoacetate; however, the latter ketone body is preferentially detected by a commonly used ketosis detection reagent (nitroprusside). Serum ketones are present at significant levels (usually positive at serum dilution of 1:8 or greater). The nitroprusside tablet, or stick, is often used to detect urine ketones; certain medications such as captopril or penicillamine may cause false-positive reactions. Serum or plasma assays for b-hydroxybutyrate more accurately reflect the true ketone body level.

The metabolic derangements of[DKA](#)exist along a spectrum, beginning with mild acidosis with moderate hyperglycemia evolving into more severe findings. The degree of acidosis and hyperglycemia do not necessarily correlate closely, as a variety of factors determine the level of hyperglycemia (oral intake, urinary glucose loss). Ketonemia is a consistent finding in DKA and distinguishes it from simple hyperglycemia.

TREATMENT

The management of DKA is outlined in Table 333-6. After initiating intravenous fluid replacement and insulin therapy, the agent or event that precipitated the episode of DKA should be sought and aggressively treated. If the patient is vomiting or has altered mental status, a nasogastric tube should be inserted to prevent aspiration of gastric contents. Central to successful treatment of DKA is careful patient monitoring and frequent reassessment to ensure that the patient and the metabolic derangements are improving. A comprehensive flow sheet should record chronologic changes in vital signs, fluid intake and output, and laboratory values as a function of insulin administered.

After the initial bolus of normal saline, replacement of the sodium and free water deficit is carried out over the next 24 h (fluid deficit is often 3 to 5 L). When hemodynamic stability and adequate urine output are achieved, intravenous fluids should be switched to 0.45% saline at a rate of 200 to 300 mL/h, depending on the calculated volume deficit. The change to 0.45% saline helps reduce the trend toward hyperchloremia later in the course of DKA. Alternatively, initial use of lactated Ringer's intravenous solution may reduce the hyperchloremia that commonly occurs with normal saline.

A bolus of intravenous or intramuscular insulin (10 to 20 units) should be administered immediately (Table 333-6)), and subsequent treatment should provide continuous and adequate levels of circulating insulin. Intravenous administration is preferred, because it assures rapid distribution and allows adjustment of the infusion rate as the patient responds to therapy. Intravenous insulin should be continued until the acidosis resolves and the patient is metabolically stable. As the acidosis and insulin resistance associated with DKA resolve, the insulin infusion rate can be decreased (to 1 to 4 units/h). Intermediate or long-acting insulin, in combination with subcutaneous regular insulin, should be administered as soon as the patient resumes eating, as this facilitates transition to an outpatient insulin regimen and reduces length of hospital stay. It is crucial to continue the insulin infusion until adequate insulin levels are achieved by the subcutaneous route. Even relatively brief periods of inadequate insulin administration in this transition phase may allow for DKA relapse.

Hyperglycemia usually improves at a rate of 4.2 to 5.6 mmol/L (75 to 100 mg/dL per hour) as a result of insulin-mediated glucose disposal, reduced hepatic glucose release, and rehydration. The latter reduces catecholamines, increases urinary glucose loss, and expands the intravascular volume. The decline in the plasma glucose within the first 1 to 2 h may be more rapid and is mostly related to volume expansion. When the plasma glucose reaches 13.9 mmol/L (250 mg/dL), glucose should be added to the 0.45% saline infusion to maintain the plasma glucose in the 11.1 to 13.9 mmol/L (200 to 250 mg/dL) range, and the insulin infusion should be continued. Ketoacidosis begins to resolve as insulin reduces lipolysis, increases peripheral ketone body use, suppresses hepatic ketone body formation, and promotes bicarbonate regeneration. However, the acidosis and ketosis resolve at a slower rate than does the hyperglycemia. As ketoacidosis improves, b-hydroxybutyrate is converted to acetoacetate. Ketone body levels may appear to increase if measured by laboratory assays that use the nitroprusside reaction, which only detects acetoacetate and acetone levels. The improvement in acidosis and anion gap, a result of bicarbonate regeneration and decline in ketone bodies, is reflected by a rise in the serum bicarbonate level and the

arterial pH. Depending on the rise of serum chloride, the anion gap (but not bicarbonate) will normalize. A hyperchloremic acidosis [serum bicarbonate of 15 to 18 mmol/L (15 to 18 meq/L)] often follows successful treatment and is minimized by the use of hypotonic intravenous solutions. This gradually resolves as the kidney regenerates bicarbonate and excretes chloride.

Potassium stores are depleted in DKA [estimated deficit 3 to 5 mmol/kg (3 to 5 meq/kg)], but the serum potassium may be normal or even elevated at the time of presentation. During treatment with insulin and fluids, various factors contribute to the development of hypokalemia. These include insulin-mediated potassium transport into cells, resolution of the acidosis (which also promotes potassium entry into cells), and urinary loss of potassium salts of organic acids. Thus, potassium repletion should commence as soon as adequate urine output and a normal serum potassium are documented. If the initial serum potassium level is elevated, then potassium repletion should be delayed until the potassium falls into the normal range. Inclusion of 20 to 40 meq of potassium in each liter of intravenous fluid is reasonable, but additional potassium supplements may also be required. To reduce the amount of chloride administered, potassium phosphate or acetate can be substituted for the chloride salt. The goal is to maintain the serum potassium>3.5 mmol/L (3.5 meq/L).

Despite a bicarbonate deficit, bicarbonate replacement is not usually necessary or advisable. In fact, theoretical arguments suggest that bicarbonate administration and rapid reversal of acidosis may impair cardiac function, impair tissue oxygenation, and promote hypokalemia. The results of most clinical trials do not support the routine use of bicarbonate replacement. In the presence of severe acidosis (arterial pH< 7.0 or hypotension unresponsive to fluid resuscitation), some physicians administer bicarbonate [50 to 150 mmol/L (meq/L) of sodium bicarbonate in 250 mL of 0.45% saline over 1 to 2 h until the serum bicarbonate rises to approximately 10 mmol/L (meq/L)]. Hypophosphatemia may result from increased glucose usage, but randomized clinical trials have not demonstrated that phosphate replacement is beneficial in DKA. If the serum phosphate is < 0.32 mmol/L (1.0 mg/dl), then phosphate supplement should be considered and the serum calcium monitored. Hypomagnesemia may develop during DKA therapy and may also require supplementation.

With appropriate therapy, the mortality of DKA is low (<5%) and is related more to the underlying or precipitating event, such as infection or myocardial infarction. The major nonmetabolic complication of DKA therapy is cerebral edema, which most often develops in children as DKA is resolving. The etiology and optimal therapy for cerebral edema are not well established, but overreplacement of free water should be avoided. Venous thrombosis and adult respiratory distress syndrome occasionally complicate DKA.

Following successful treatment of DKA, the physician and patient should review the sequence of events that led to DKA to prevent future recurrences. Foremost is patient education about the symptoms of DKA, its precipitating factors, and the management of diabetes during a concurrent illness. During illness or when oral intake is compromised, patients should: (1) frequently measure the capillary blood glucose; (2) measure urinary ketones when the serum glucose >16.5 mmol/L (300 mg/dL); (3) drink fluids to maintain hydration; (4) continue or increase insulin; and (5) seek medical attention if dehydration,

persistent vomiting, or uncontrolled hyperglycemia develop. In this way, early DKA can be detected and treated appropriately on an outpatient basis.

NONKETOTIC HYPEROSMOLAR STATE

Clinical Features NKHS is most commonly seen in elderly individuals with type 2DM. Its most prominent features include polyuria; orthostatic hypotension; and a variety of neurologic symptoms that include altered mental status, lethargy, obtundation, seizure, and possibly coma. The prototypical patient is a mildly diabetic, elderly individual with a several week history of polyuria, weight loss, and diminished oral intake that culminates in mental confusion, lethargy, or coma. The physical examination reflects profound dehydration and hyperosmolality and reveals hypotension, tachycardia, and altered mental status. Notably absent are symptoms of nausea, vomiting, and abdominal pain and the Kussmaul respirations characteristic ofDKA. NKHS is often precipitated by a serious, concurrent illness such as myocardial infarction or stroke. Sepsis, pneumonia, and other serious infections are frequent precipitants and should be sought thoroughly. In addition, a debilitating condition (prior stroke or dementia) or social situation that compromises water intake may contribute to the development of the disorder. Finally, the development of NKHS can be associated with the use of certain medications (thiazide diuretics, glucocorticoids, phenytoin).

Pathophysiology Insulin deficiency and inadequate fluid intake are the underlying causes ofNKHS. Insulin deficiency increases hepatic glucose production (through glycogenolysis and gluconeogenesis) and impairs glucose utilization in skeletal muscle (see above discussion underDKA). Hyperglycemia induces an osmotic diuresis that leads to profound intravascular volume depletion, which is exacerbated by inadequate fluid replacement. The absence of ketosis in NKHS is not completely understood. Presumably, the insulin deficiency is only relative and less severe than in DKA. Lower levels of counterregulatory hormones and free fatty acids have been found in NKHS than in DKA in some studies. It is also possible that the liver is less capable of ketone body synthesis or that the insulin/glucagon ratio does not favor ketogenesis.

Laboratory Abnormalities and Diagnosis The laboratory features inNKHS are summarized in Table 333-4. Most notable are the marked hyperglycemia [plasma glucose may be >55.5 mmol/L (1000 mg/dL)], hyperosmolality (>350 mosmol/L), and prerenal azotemia. The measured serum sodium may be normal or slightly low despite the marked hyperglycemia. The corrected serum sodium is usually increased [add 1.6 meq to measured sodium for each 5.6 mmol/L (100 mg/dL) rise in the serum glucose]. In contrast toDKA, acidosis and ketonemia are absent or mild. A small anion gap metabolic acidosis may be present secondary to increased lactic acid. Moderate ketonuria, if present, is secondary to starvation.

TREATMENT

Volume depletion and hyperglycemia are prominent features of bothNKHS andDKA. Consequently, therapy of these disorders involves several shared elements (Table 333-6). In both disorders, careful monitoring of the patient's fluid status, laboratory values, and insulin infusion rate is crucial. Underlying or precipitating problems should be aggressively sought and treated. In NKHS, the volume depletion, free water deficit,

and hyperosmolality are greater than in DKA. The patient with NKHS is usually older, more likely to have mental status changes, and thus more likely to have a life-threatening precipitating event with accompanying comorbidities. Even with proper treatment, NKHS has a substantially higher mortality than DKA (up to 50% in some clinical series).

Fluid replacement should initially stabilize the hemodynamic status of the patient (1 to 3 L of 0.9% normal saline over the first 2 to 3 h). Because the fluid deficit inNKHS is accumulated over a period of days to weeks, the rapidity of reversal of the hyperosmolar state must balance the need for free water repletion and the observation that too rapid a reversal may worsen neurologic function. If the serum sodium is >150mmol/L (150 meq/L), 0.45% saline should be used. After hemodynamic stability is achieved, the intravenous fluid administration is directed at reversing the free water deficit using hypotonic fluids (0.45% saline initially then 5% dextrose in water, D_5W). The calculated free water deficit (which averages 9 to 10 L) should be reversed over the next 1 to 2 days (infusion rates of 200 to 300 mL/h of hypotonic solution). Potassium repletion is usually necessary and should be dictated by repeated measurements of the serum potassium. In patients taking diuretics, the potassium deficit can be quite large and may be accompanied by magnesium deficiency. Hypophosphatemia may occur during therapy and can be improved by using KPO_4 and beginning nutrition.

As inDKA, rehydration and volume expansion lower the plasma glucose initially, but insulin is eventually required. InNKHS, patients tend to be more sensitive to insulin than in DKA and dose requirements are not usually as large. A reasonable regimen for NKHS begins with an intravenous insulin bolus of 5 to 10 units followed by intravenous insulin at a constant infusion rate (3 to 7 units/h). As in DKA, glucose should be added to intravenous fluid when the plasma glucose falls to 13.9 mmol/L (250 mg/dL), and the insulin infusion rate should be decreased to 1 to 2 units/h. The insulin infusion should be continued until the patient has resumed eating and can be transferred to a subcutaneous insulin regimen. The patient should be discharged from the hospital on insulin, though some patients can later undergo a trial of oral glucose-lowering agents.

CHRONIC COMPLICATIONS

The chronic complications ofDMaffect many organ systems and are responsible for the majority of morbidity and mortality associated with the disease. Chronic complications can be divided into vascular and nonvascular complications (Table 333-7). The vascular complications of DM are further subdivided into microvascular (retinopathy, neuropathy, nephropathy) and macrovascular complications (coronary artery disease, peripheral vascular disease, cerebrovascular disease). Nonvascular complications include problems such as gastroparesis, sexual dysfunction, and skin changes. This division is rather arbitrary since it is likely that multiple pathogenic processes are involved in all forms of complications.

The risk of chronic complications increases as a function of the duration of hyperglycemia; they usually become apparent in the second decade of hyperglycemia. Since type 2DM may have a long asymptomatic period of hyperglycemia, many individuals with type 2 DM have complications at the time of diagnosis.

The microvascular complications of both type 1 and type 2 DM result from chronic hyperglycemia. Randomized, prospective clinical trials involving large numbers of individuals with type 1 or type 2 DM have conclusively demonstrated that a reduction in chronic hyperglycemia prevents or reduces retinopathy, neuropathy, and nephropathy. Other incompletely defined factors also modulate the development of complications. For example, despite longstanding DM, some individuals never develop nephropathy or retinopathy. Many of these patients have glycemic control that is indistinguishable from those who develop microvascular complications. Because of these observations, it is suspected that a genetic susceptibility for developing particular complications exists. However, the genetic loci responsible for these susceptibilities have not yet been identified.

Evidence implicating a causative role for chronic hyperglycemia in the development of macrovascular complications is less conclusive, but some results suggest a role for chronic hyperglycemia in the development of macrovascular disease. For example, coronary heart disease events and mortality are two to four times greater in patients with type 2 DM. These events correlate with fasting and postprandial plasma glucose levels as well as with the HbA1c. Other factors (dyslipidemia and hypertension) also play important roles in macrovascular complications.

MECHANISMS OF COMPLICATIONS

Although chronic hyperglycemia is an important etiologic factor leading to complications of DM, the mechanism(s) by which it leads to such diverse cellular and organ dysfunction is unknown. Three major theories, which are not mutually exclusive, have been proposed to explain how hyperglycemia might lead to the chronic complications of DM (Fig. 333-7).

One hypothesis is that increased intracellular glucose leads to the formation of advanced glycosylation end products (AGEs) via the nonenzymatic glycosylaton of cellular proteins. Nonenzymatic glycosylation results from the interaction of glucose with amino groups on proteins. AGEs have been shown to cross-link proteins (e.g., collagen, extracellular matrix proteins), accelerate atherosclerosis, promote glomerular dysfunction, reduce nitric oxide synthesis, induce endothelial dysfunction, and alter extracellular matrix composition and structure. The serum level of AGEs correlates with the level of glycemia, and these products accumulate as glomerular filtration rate declines.

A second hypothesis proposed to explain how chronic hyperglycemia leads to complications of DM is based on the observation that hyperglycemia increases glucose metabolism via the sorbitol pathway. Intracellular glucose is predominantly metabolized by phosphorylation and subsequent glycolysis, but when intracellular glucose is increased, some glucose is converted to sorbitol by the enzyme aldose reductase. Increased sorbitol concentrations affect several aspects of cellular physiology (decreased myoinositol, altered redox potential) and may lead to cellular dysfunction. However, testing of this theory in humans, using aldose reductase inhibitors, has not demonstrated beneficial effects on clinical endpoints of retinopathy, neuropathy, or nephropathy.

A third hypothesis proposes that hyperglycemia increases the formation of diacylglycerol leading to activation of certain isoforms of protein kinase C (PKC), which, in turn, affect a variety of cellular events that lead to DM-related complications. For example, PKC activation by glucose alters the transcription of genes for fibronectin, type IV collagen, contractile proteins, and extracellular matrix proteins in endothelial cells and neurons in vitro. Growth factors appear to play an important role in DM-related complications. Vascular endothelial growth factor (VEGF) is increased locally in diabetic proliferative retinopathy and decreases after laser photocoagulation. Transforming growth factor b (TGF-b) is increased in diabetic nephropathy and appears to stimulate basement membrane production of collagen and fibronectin by mesangial cells. Other growth factors, such as platelet-derived growth factor, epidermal growth factor, insulin-like growth factor I, growth hormone, basic fibroblast growth factor, and even insulin, have been suggested to play a role in DM-related complications.

Although hyperglycemia serves as the initial trigger for complications of diabetes, it is still unknown whether the same pathophysiologic processes are operative in all complications or whether certain processes predominate in certain organs. Finally, oxidative stress and free radical generation, as a consequence of the hyperglycemia, may also promote the development of complications.

GLYCEMIC CONTROL AND COMPLICATIONS

The Diabetes Control and Complications Trial (DCCT) provided definitive proof that reduction in chronic hyperglycemia can prevent many of the early complications of type 1 DM. This large multicenter clinical trial randomized over 1400 individuals with type 1 DM to either intensive or conventional diabetes management, and then evaluated the development of retinopathy, nephropathy, and neuropathy. Individuals in the intensive diabetes management group received multiple administrations of insulin each day along with intense educational, psychological, and medical support. Individuals in the conventional diabetes management group received twice daily insulin injections and quarterly nutritional, educational, and clinical evaluation. The goal in the former group was normoglycemia; the goal in the latter group was prevention of symptoms of diabetes. Individuals in the intensive diabetes management group achieved a substantially lower HbA1c (7.2%) than individuals in the conventional diabetes management group (HbA1c of 9.0%).

Results from the DCCT demonstrated that improvement of glycemic control reduced nonproliferative and proliferative retinopathy (47% reduction), microalbuminuria (39% reduction), clinical nephropathy (54% reduction), and neuropathy (60% reduction). Improved glycemic control also slowed the progression of early diabetic complications. There was a nonsignificant trend in reduction of macrovascular events. The results of the DCCT predicted that individuals in the intensive diabetes management group would gain 7.7 additional years of sight, 5.8 additional years free from end-stage renal disease (ESRD), and 5.6 years free from lower extremity amputations. If all complications of DM were combined, individuals in the intensive diabetes management group would experience 15.3 more years of life without significant microvascular or neurologic complications of DM as compared to individuals who received standard therapy. This translates into an additional 5.1 years of life expectancy for individuals in the intensive diabetes management group. The benefit of the improved glycemic control during the

DCCT persisted even after the study concluded and glycemic control worsened.

The benefits of an improvement in glycemic control occurred over the entire range of HbA1c values ([Fig. 333-8]), suggesting that at any HbA1c level, an improvement in glycemic control is beneficial. Therefore, there is no threshold beneath which the HbA1c can be reduced and the complications of DM prevented. The clinical implication of this finding is that the goal of therapy is to achieve an HbA1c level as close to normal as possible, without subjecting the patient to excessive risk of hypoglycemia.

Considerable debate has emerged as to whether the DCCT findings are applicable to individuals with type 2 DM, in whom insulin resistance, hyperinsulinemia, and obesity predominate. Concerns have been raised that therapies associated with weight gain and additional insulin therapy may worsen underlying insulin resistance and hyperinsulinemia. Despite these concerns, most available data support extrapolation of the results of the DCCT to individuals with type 2 DM.

The United Kingdom Prospective Diabetes Study (UKPDS) studied the course of >5000 individuals with type 2 DM for >10 years. This complex and important study utilized multiple treatment regimens and monitored the effect of intensive glycemic control and risk factor treatment on the development of diabetic complications. Newly diagnosed individuals with type 2 DM were randomized to (1) intensive management using various combinations of insulin, a sulfonylurea, or metformin; or (2) conventional therapy using dietary modification and pharmacotherapy with the goal of symptom prevention. In addition, individuals were randomly assigned to different antihypertensive regimens. Individuals in the intensive treatment arm achieved an HbA1c of 7.0%, compared to a 7.9% HbA1c in the standard treatment group. The UKPDS demonstrated that each percentage point reduction in HbA1c was associated with a 35% reduction in microvascular complications, a 25% reduction in DM-related deaths, and a 7% reduction in all-cause mortality. As in the DCCT, there was a continuous relationship between glycemic control and development of complications. Although there was no statistically significant effect of glycemic control on cardiovascular complications, there was a 16% reduction in fatal and nonfatal myocardial infarctions.

One of the major findings of the UKPDS was the observation that strict blood pressure control significantly reduced both macro- and microvascular complications. In fact, the beneficial effects of blood pressure control were greater than the beneficial effects of glycemic control. Lowering blood pressure to moderate goals (144/82 mmHg) reduced the risk of DM-related death, stroke, microvascular end points, retinopathy, and heart failure (risk reductions between 32 and 56%). Improved glycemic control did not conclusively reduce (nor worsen) cardiovascular mortality but was associated with improvement with lipoprotein risk profiles, such as reduced triglycerides and increased high-density lipoprotein (HDL).

Similar reductions in the risks of retinopathy and nephropathy were also seen in a small trial of lean Japanese individuals with type 2 DM randomized to either intensive glycemic control or standard therapy with insulin (Kumamoto study). These results demonstrate the effectiveness of improved glycemic control in individuals of different ethnicity with a presumably different etiology of DM (i.e., phenotypically different from those in the DCCT and UKPDS).

The findings of the DCCT, UKPDS, and Kumamoto study support the idea that chronic hyperglycemia plays a causative role in the pathogenesis of diabetic microvascular complications. These landmark studies prove the value of metabolic control and emphasize the importance of (1) intensive glycemic control in all forms of DM, and (2) early diagnosis and strict blood pressure control in type 2 DM.

OPHTHALMOLOGIC COMPLICATIONS OF DIABETES MELLITUS

DM is the leading cause of blindness between the ages of 20 and 74 in the United States. The gravity of this problem is highlighted by the finding that individuals with DM are 25 times more likely to become legally blind than individuals without DM. Blindness is primarily the result of progressive diabetic retinopathy and clinically significant macular edema. Diabetic retinopathy is classified into two stages: nonproliferative and proliferative. *Nonproliferative diabetic retinopathy* usually appears late in the first decade or early in the second decade of the disease and is marked by retinal vascular microaneurysms, blot hemorrhages, and cotton wool spots (see Plate IV-15). Mild nonproliferative retinopathy progresses to more extensive disease, characterized by changes in venous vessel caliber, intraretinal microvascular abnormalities, and more numerous microaneurysms and hemorrhages. The pathophysiologic mechanisms invoked in nonproliferative retinopathy include loss of retinal pericytes, increased retinal vascular permeability, alterations in retinal blood flow, and abnormal retinal microvasculature, all of which lead to retinal ischemia.

The appearance of neovascularization in response to retinal hypoxia is the hallmark of *proliferative diabetic retinopathy*. These newly formed vessels may appear at the optic nerve and/or macula and rupture easily, leading to vitreous hemorrhage, fibrosis, and ultimately retinal detachment. Not all individuals with nonproliferative retinopathy develop proliferative retinopathy, but the more severe the nonproliferative disease, the greater the chance of evolution to proliferative retinopathy within 5 years. This creates a clear opportunity for early detection and treatment of diabetic retinopathy (discussed below). In contrast, *clinically significant macular edema* may appear when only nonproliferative retinopathy is present. Fluorescein angiography is often useful to detect macular edema, which is associated with a 25% chance of moderate visual loss over the next 3 years.

Duration of DM and degree of glycemic control are the best predictors of the development of retinopathy. Nonproliferative retinopathy is found in almost all individuals who have had DM for >20 years (25% incidence with 5 years, and 80% incidence with 15 years of type 1 DM). Although there is genetic susceptibility for retinopathy, it confers less influence on the development of retinopathy than either the duration of DM or the degree of glycemic control.

TREATMENT

The most effective therapy for diabetic retinopathy is prevention. Intensive glycemic control will greatly delay the development or slow the progression of retinopathy in individuals with either type 1 or type 2 DM. Paradoxically, during the first 6 to 12 months of improved glycemic control, established diabetic retinopathy may transiently worsen.

Fortunately, this progression is temporary, and in the long term, improved glycemic control is associated with less diabetic retinopathy. Individuals with known retinopathy should be considered candidates for prophylactic photocoagulation when initiating intensive therapy. Once advanced retinopathy is present, improved glycemic control imparts less benefit, though adequate ophthalmologic care can prevent most blindness.

Equally as important as glycemic control are regular, comprehensive eye examinations for all individuals with DM. Most diabetic eye disease can be successfully treated if detected early. Routine, nondilated eye examinations by the primary care provider or diabetes specialist are *inadequate* to detect diabetic eye disease properly. The treatment of diabetic eye disease requires an ophthalmologist experienced in these disorders. Laser photocoagulation is very successful in preserving vision. Proliferative retinopathy is usually treated with panretinal laser photocoagulation, whereas macular edema is treated with focal laser photocoagulation. Although exercise has not been conclusively shown to worsen proliferative diabetic retinopathy, most ophthalmologists advise individuals with advanced diabetic eye disease to limit physical activities associated with repeated Valsalva maneuvers. Aspirin therapy (650 mg/d) does not appear to influence the natural history of diabetic retinopathy, but studies of other antiplatelet agents are under way.

RENAL COMPLICATIONS OF DIABETES MELLITUS

Diabetic nephropathy is the leading cause of ESRD in the United States and a leading cause of DM-related morbidity and mortality. Proteinuria in individuals with DM is associated with markedly reduced survival and increased risk of cardiovascular disease. Individuals with diabetic nephropathy almost always have diabetic retinopathy also.

Like other microvascular complications, the pathogenesis of diabetic nephropathy is related to chronic hyperglycemia (Fig. 334-7). The mechanisms by which chronic hyperglycemia leads to ESRD, though incompletely defined, involve the following: interaction of soluble factors (growth factors, angiotensin II, endothelin, AGEs), hemodynamic alterations in the renal microcirculation (glomerular hyperfiltration, increased glomerular capillary pressure), and structural changes in the glomerulus (increased extracellular matrix, basement membrane thickening, mesangial expansion, fibrosis). Some of these effects may be mediated through angiotensin receptors. Smoking accelerates the decline in renal function.

The natural history of diabetic nephropathy is shown schematically in Fig. 333-9 and is characterized by a fairly predictable pattern of events. Although this sequence of events was defined for individuals with type 1 DM, a similar pattern is also likely in type 2 DM. Glomerular hyperfusion and renal hypertrophy occur in the first years after the onset of DM and are reflected by an increased glomerular filtration rate (GFR). During the first 5 years of DM, thickening of the glomerular basement membrane, glomerular hypertrophy, and mesangial volume expansion occur as the GFR returns to normal. After 5 to 10 years of type 1 DM, ~40% of individuals begin to excrete small amounts of albumin in the urine (microalbuminuria). *Microalbuminuria* is defined as 30 to 300 mg/d in a 24-h collection or 30 to 300 ug/mg creatinine in a spot collection. The appearance of microalbuminuria (incipient nephropathy) in type 1 DM is a very important predictor of progression to overt proteinuria (>300 mg/d). Blood pressure may rise slightly at this

point but usually remains in the normal range. Once overt proteinuria is present, there is a steady decline in GFR, and ~50% of individuals reach ESRD in 7 to 10 years. The early pathologic changes and albumin excretion abnormalities are reversible with normalization of plasma glucose. However, once nephropathy becomes overt, the pathologic changes are likely irreversible.

The nephropathy that develops in type 2 DM differs from that of type 1 DM in the following respects: (1) microalbuminuria or overt nephropathy may be present when type 2 DM is diagnosed, reflecting its long asymptomatic period; (2) hypertension more commonly accompanies microalbuminuria or overt nephropathy in type 2 DM; and (3) microalbuminuria may be less predictive of progression to overt nephropathy in type 2 DM. Finally, it should be noted that albuminuria in type 2 DM may be secondary to factors unrelated to DM, such as hypertension, congestive heart failure, prostate disease, or infection.

Other renal problems may also occur in individuals with DM. Type IV renal tubular acidosis (hyporeninemic hypoaldosteronism) occurs in many individuals with DM. These individuals develop a propensity to hyperkalemia, which may be exacerbated by medications [especially angiotensin-converting enzyme (ACE) inhibitors]. Patients with DM are predisposed to radiocontrast-induced nephrotoxicity. Individuals with DM undergoing radiographic procedures with contrast dye should be well hydrated before and after dye exposure, and the serum creatinine should be monitored for several days following the procedure.

TREATMENT

The optimal therapy for diabetic nephropathy is prevention. As part of comprehensive diabetes care, microalbuminuria should be detected at an early stage when effective therapies can be instituted. The recommended strategy for detecting microalbuminuria is outlined in Fig. 333-10. Interventions effective in slowing progression from microalbuminuria to overt nephropathy include: (1) near normalization of glycemia, (2) strict blood pressure control, and (3) administration of ACE inhibitors.

Improved glycemic control reduces the rate at which microalbuminuria appears and progresses in both type 1 and type 2 DM. However, once overt nephropathy exists, it is unclear whether improved glycemic control will slow progression of renal disease. During the phase of declining renal function, insulin requirements may fall as the kidney is a site of insulin degradation. Furthermore, glucose-lowering medications (sulfonylureas and metformin) may accumulate and are contraindicated in renal insufficiency.

Many individuals with type 1 or type 2 DM develop hypertension. Numerous studies in both type 1 and type 2 DM demonstrate the effectiveness of strict blood pressure control in reducing albumin excretion and slowing the decline in renal function. Blood pressure should be maintained at <130/85 mmHg in diabetic individuals without proteinuria. A slightly lower blood pressure (120/80) should be targeted for individuals with microalbuminuria or overt nephropathy. Treatment of hypertension is discussed below.

ACE inhibitors reduce the progression of overt nephropathy in individuals with type 1 or

type 2DM and should be prescribed in individuals with type 1 or type 2 DM and microalbuminuria. After 2 to 3 months of therapy, measurement of proteinuria should be repeated and the drug dose increased until either the albuminuria disappears or the maximum dose is reached. If an ACE inhibitor has an unacceptable side-effect profile (hyperkalemia, cough, and renal insufficiency), angiotensin II receptor blockers and calcium channel blockers (phenylalkylamine class) are alternatives. However, their efficacy in slowing the fall in glomerular filtration rate is not proven. Blood pressure control with any agent is extremely important, but a drug-specific benefit in diabetic nephropathy, independent of blood pressure control, has been shown only for ACE inhibitors.

A consensus panel of the American Diabetes Association (ADA) suggests modest restriction of protein intake in diabetic individuals with microalbuminuria (0.8 g/kg per day, which is the adult Recommended Daily Allowance, and about 10% of the daily caloric intake). Protein intake should be restricted further in individuals with overt diabetic nephropathy (0.6 g/kg per day), though conclusive proof of the efficacy of protein restriction is lacking.

Nephrology consultation should be considered after the diagnosis of early nephropathy. Once overt nephropathy ensues, the likelihood ofESRD is very high. As compared to nondiabetic individuals, hemodialysis in patients withDM is associated with more frequent complications, such as hypotension (autonomic neuropathy, loss of reflex tachycardia), more difficult vascular access, and accelerated progression of retinopathy. Survival after the onset of ESRD is shorter in the diabetic population compared to nondiabetics with similar clinical features. Atherosclerosis is the leading cause of death in diabetic individuals on dialysis, and hyperlipidemia should be aggressively treated. Renal transplantation from a living-related donor is the preferred therapy but requires chronic immunosuppression. Combined pancreas-kidney transplant offers the promise of normoglycemia but requires substantial expertise.

NEUROPATHY AND DIABETES MELLITUS

Diabetic neuropathy occurs in approximately 50% of individuals with long-standing type 1 and type 2DM. It may manifest as polyneuropathy, mononeuropathy, and/or autonomic neuropathy. As with other complications of DM, the development of neuropathy correlates with the duration of diabetes and glycemic control. Because the clinical features of diabetic neuropathy are similar to those of other neuropathies, the diagnosis of *diabetic* neuropathy should be made only after other possible etiologies are excluded (Chap. 378).

Polyneuropathy/Mononeuropathy The most common form of diabetic neuropathy is distal symmetric *polyneuropathy*. It most frequently presents with distal sensory loss. Hyperesthesia, parathesia, and pain also occur. Any combination of these symptoms may develop as neuropathy progresses. Physical examination reveals sensory loss, loss of ankle reflexes, and abnormal position sense. Parethesia is characteristically perceived as a sensation of numbness, tingling, sharpness, or burning that begins in the feet and spreads proximally. Neuropathic pain develops in some of these individuals, occasionally preceded by improvement in their glycemic control. Pain typically involves the lower extremities, is usually present at rest, and worsens at night. Both an acute

(lasting<12 months) and a chronic form of painful diabetic neuropathy have been described. As diabetic neuropathy progresses, the pain subsides and eventually disappears, and a sensory deficit in the lower extremities persists.

Diabetic polyradiculopathy is a syndrome characterized by severe disabling pain in the distribution of one or more nerve roots. It may be accompanied by motor weakness. Intercostal or truncal radiculopathy causes pain over the thorax or abdomen. Involvement of the lumbar plexus or femoral nerve may cause pain in the thigh or hip and may be associated with muscle weakness in the hip flexors or extensors (diabetic amyotrophy). Fortunately, diabetic polyradiculopathies are usually self-limited and resolve over 6 to 12 months.

Mononeuropathy (dysfunction of isolated cranial or peripheral nerves) is less common than polyneuropathy in DM and presents with pain and motor weakness in the distribution of a single nerve. A vascular etiology is favored, but the pathogenesis is unknown. Involvement of the third cranial nerve is most common and is heralded by diplopia. Physical examination reveals ptosis and opthalmoplegia with normal papillary constriction to light. Sometimes cranial nerves IV, VI, or VII (Bell's palsy) are affected. Peripheral mononeuropathies or simultaneous involvement of more than one nerve (mononeuropathy multiplex) may also occur.

Autonomic Neuropathy Individuals with long-standing type 1 or 2 DM may develop signs of autonomic dysfunction involving the cholinergic, noradrenergic, and peptidergic (peptides such as pancreatic polypeptide, substance P, etc.) systems. DM-related autonomic neuropathy can involve multiple systems, including: the cardiovascular, gastrointestinal, genitourinary, sudomotor, and metabolic systems. Autonomic neuropathies affecting the cardiovascular system cause a resting tachycardia and orthostatic hypotension. Reports of sudden death have also been attributed to autonomic neuropathy. Gastroparesis and bladder-emptying abnormalities are also likely related to the autonomic neuropathy seen in DM (discussed below). Hyperhidrosis of the upper extremities and anhidrosis of the lower extremities result from sympathetic nervous system dysfunction. Anhidrosis of the feet can promote dry skin with cracking, which increases the risk of skin ulceration. Autonomic neuropathy may reduce counterregulatory hormone release, leading to an inability to sense hypoglycemia appropriately (*hypoglycemia unawareness*;Chap. 334), thereby subjecting the patient to the risk of severe hypoglycemia and complicating efforts to improve glycemic control.

TREATMENT

Treatment of diabetic neuropathy is less than satisfactory. Improved glycemic control should be pursued and will improve nerve conduction velocity, but the symptoms of diabetic neuropathy may not necessarily improve. Efforts to improve glycemic control may be confounded by autonomic neuropathy and hypoglycemia unawareness. Avoidance of neurotoxins (alcohol), supplementation with vitamins for possible deficiencies (B_{12}, B_6, folate; Chap. 75), and symptomatic treatment are the mainstays of therapy. Aldose reductase inhibitors do not currently offer significant symptomatic relief. Loss of sensation in the foot places the patient at risk for ulceration and its sequelae; consequently, prevention of such problems is of paramount importance. Since the pain of acute diabetic neuropathy may resolve over the first year, analgesics may be

discontinued as progressive neuronal damage from DM occurs. Chronic, painful diabetic neuropathy is difficult to treat but may respond to tricyclic antidepressants (amitriptyline, desipramine, nortriptyline), gabapentin, nonsteroidal anti-inflammatory agents (avoid in renal dysfunction), and other agents (mexilitine, phenytoin, carbamazepine, capsaicin cream). Referral to a pain management center may be necessary.

Therapy of orthostatic hypotension secondary to autonomic neuropathy is difficult. A variety of agents have limited success (fludrocortisone, midodrine, clonidine, octreotide, and yohimbine) but have significant side effects. Nonpharmacologic maneuvers (adequate salt intake, avoidance of dehydration and diuretics, and lower extremity support hose) may offer some benefit.

GASTROINTESTINAL/GENITOURINARY DYSFUNCTION

Long-standing type 1 and 2 DM may affect the motility and function of gastrointestinal (GI) and genitourinary systems. The most prominent GI symptoms are delayed gastric emptying (gastroparesis) and altered small- and large-bowel motility (constipation or diarrhea). *Gastroparesis* may present with symptoms of anorexia, nausea, vomiting, early satiety, and abdominal bloating. Nuclear medicine scintigraphy after ingestion of a radiolabeled meal is the best study to document delayed gastric emptying, but noninvasive "breath tests" following ingestion of a radiolabeled meal are under development. Though parasympathetic dysfunction secondary to chronic hyperglycemia is important in the development of gastroparesis, hyperglycemia itself also impairs gastric emptying. Nocturnal diarrhea, alternating with constipation, is a common feature of DM-related GI autonomic neuropathy. In type 1 DM, these symptoms should also prompt evaluation for celiac sprue because of its increased frequency. Esophageal dysfunction in long-standing DM is common but usually asymptomatic.

Diabetic autonomic neuropathy may lead to genitourinary dysfunction including cystopathy, erectile dysfunction, and female sexual dysfunction (reduced sexual desire, dyspareunia, reduced vaginal lubrication). Symptoms of diabetic cystopathy begin with an inability to sense a full bladder and a failure to void completely (Chap. 48). As bladder contractility worsens, bladder capacity and the postvoid residual increase, leading to symptoms of urinary hesitancy, decreased voiding frequency, incontinence, and recurrent urinary tract infections. Diagnostic evaluation includes cystometry and urodynamic studies.

Erectile dysfunction and retrograde ejaculation are very common in DM and may be one of the earliest signs of diabetic neuropathy. Erectile dysfunction, which increases in frequency with the age of the patient and the duration of diabetes, may occur in the absence of other signs of diabetic autonomic neuropathy.

TREATMENT

Current treatments for these complications of DM are inadequate. Improved glycemic control should be a primary goal, as some aspects (neuropathy, gastric function) may improve as near-normoglycemia is achieved. Smaller, more frequent meals that are easier to digest (liquid) and low in fat and fiber may minimize symptoms of gastroparesis. Cisapride (10 to 20 mg before each meal) is probably the most effective

medication but has been removed from use in the U.S. market except under special circumstances. Other agents with some efficacy include dopamine agonists (metoclopramide, 5 to 10 mg, and domperidone, 10 to 20 mg, before each meal) and bethanechol (10 to 20 mg before each meal). Erythromycin interacts with the motilin receptor and may promote gastric emptying. Diabetic diarrhea in the absence of bacterial overgrowth is treated symptomatically with loperamide but may respond to clonidine at higher doses (0.6 mg tid) or octreotide (50 to 75 ug tid subcutaneously). Treatment of bacterial overgrowth with antibiotics is sometimes useful (Chap. 286).

Diabetic cystopathy should be treated with timed voiding or self-catherization. Medications (bethanechol) are inconsistently effective. The drug of choice for erectile dysfunction is sildenafil, but the efficacy in individuals with DM is slightly lower than in the nondiabetic population (Chap. 51). Sexual dysfunction in women may be improved with use of vaginal lubricants, treatment of vaginal infections, and systemic or local estrogen replacement.

CARDIOVASCULAR MORBIDITY AND MORTALITY

Cardiovascular disease is increased in individuals with type 1 or type 2 DM. The Framingham Heart Study revealed a marked increase in several cardiovascular diseases in DM including peripheral vascular disease, congestive heart failure, coronary artery disease, myocardial infarction, and sudden death (risk increase from one- to fivefold). The American Heart Association recently designated DM as a major risk factor for cardiovascular disease (same category as smoking, hypertension, and hyperlipidemia). Because of the extremely high frequency of underlying cardiovascular disease in individuals with diabetes (especially in type 2 DM), evidence of atherosclerotic vascular disease should be sought in the individual with diabetes who has symptoms suggestive of cardiac ischemia, peripheral or carotid arterial disease, a resting electrocardiogram indicative of prior infarction, plans to initiate an exercise program, proteinuria, or two other cardiac risk factors (ADA recommendations). The absence of chest pain ("silent ischemia") is common in individuals with diabetes, and a thorough cardiac evaluation is indicated in individuals undergoing major surgical procedures.

The increase in morbidity and mortality appears to relate to the synergism of hyperglycemia with other cardiovascular risk factors. For example, after controlling for all known cardiovascular risk factors, type 2 DM increases the cardiovascular death rate by twofold in men and fourfold in women. Risk factors for macrovascular disease in diabetic individuals include dyslipidemia, hypertension, obesity, reduced physical activity, and cigarette smoking. Additional risk factors specific to the diabetic population include microalbuminuria, gross proteinuria, an elevation in serum creatinine, and altered platelet function. Insulin resistance, as reflected by elevated serum insulin levels, is associated with an increased risk of cardiovascular complications in individuals with and without DM. Individuals with insulin resistance and type 2 DM have elevated levels of plasminogen activator inhibitors (especially PAI-1) and fibrinogen, which enhances the coagulation process and impairs fibrinolysis, thus favoring the development of thrombosis.

Despite proof that improved glycemic control reduces microvascular complications

in DM, it is possible that macrovascular complications may be unaffected or even worsened by such therapy. Concerns about the anabolic and atherogenic potential of insulin remain, since in nondiabetic individuals, higher serum insulin levels (indicative of insulin resistance) are associated with a greater risk of cardiovascular morbidity and mortality. In the DCCT, the number of cardiovascular events did not differ between the standard and intensively treated groups. However, the duration of DM in these individuals was relatively short, and the total number of events was very low. An improvement in the lipid profile of individuals in the intensive group [lower total and low-density lipoprotein (LDL) cholesterol, lower triglycerides] suggested that intensive therapy may reduce the risk of cardiovascular morbidity and mortality associated with DM. In the UKPDS, improved glycemic control did not conclusively reduce cardiovascular mortality. Importantly, treatment with insulin and the sulfonylureas did not appear to increase the risk of cardiovascular disease in individuals with type 2 DM, refuting prior claims about the atherogenic potential of these agents.

In addition to coronary artery disease, cerebrovascular disease is increased in individuals with DM (threefold increase in stroke). Individuals with DM have an increased incidence of congestive heart failure (diabetic cardiomyopathy). The etiology of this abnormality is probably multifactorial and includes factors such as myocardial ischemia from atherosclerosis, hypertension, and myocardial cell dysfunction secondary to chronic hyperglycemia.

TREATMENT

In general, the treatment of coronary disease is no different in the diabetic individual (Chap. 244), though overall prognosis after myocardial infarction is worse in the diabetic population. Revascularization procedures for coronary artery disease, including percutaneous transluminal coronary angioplasty (PTCA) and coronary artery bypass grafting (CABG), are less efficacious in the diabetic individual. Initial success rates of PTCA in diabetic individuals are similar to those in the nondiabetic population, but diabetic patients have higher rates of restenosis and lower long-term patency and survival rates. Perioperative mortality from CABG is not altered in DM, but both short- and long-term survival are reduced. Recent trials indicate that diabetic individuals with multivessel coronary artery disease or who recently suffered a Q-wave myocardial infarction have better long-term survival with CABG than PTCA.

Results of studies investigating the effect of intensive diabetes management on survival rates and cardiovascular events after myocardial infarction have been conflicting. In the face of conflicting data, the ADA has emphasized the importance of glycemic control and aggressive cardiovascular risk modification in all individuals with DM. Despite past trepidation about using beta blockers in individuals who have diabetes, these agents clearly benefit diabetic patients after myocardial infarction, analogous to the benefit in nondiabetic individuals. ACE inhibitors may also be particularly beneficial in reducing mortality after myocardial infarction in patients with DM.

Antiplatelet therapy reduces cardiovascular events in individuals with DM who have coronary artery disease. Current recommendations by the ADA suggest the use of aspirin as secondary prevention of additional coronary events. Although data demonstrating efficacy in primary prevention of coronary events are lacking, antiplatelet

therapy should be considered, especially in diabetic individuals with other coronary risk factors such as hypertension, smoking, or hyperlipidemia. The aspirin dose (81 to 325 mg) is the same as that in nondiabetic individuals. Aspirin therapy does not have detrimental effects on renal function or hypertension, nor does it influence the course of diabetic retinopathy or maculopathy.

Cardiovascular Risk Factors

Dyslipidemia Individuals with DM may have several forms of dyslipidemia (Chap. 344). Because of the additive cardiovascular risk of hyperglycemia and hyperlipidemia, lipid abnormalities should be aggressively detected and treated as part of comprehensive diabetes care (Fig. 333-11). The most common pattern of dyslipidemia is hypertriglyceridemia and reduced HDL cholesterol levels. DM itself does not increase levels of LDL, but the small dense LDL particles found in type 2 DM are more atherogenic because they are more easily glycated and susceptible to oxidation.

According to guidelines of the ADA and the American Heart Association, the lipid profile in diabetic individuals without cardiovascular disease (primary prevention) should be: LDL < 3.4 mmol/L (130 mg/dL); HDL > 0.9 mmol/L (35 mg/dL) in men and >1.2 mmol/L (45 mg/dL) in women; and triglycerides < 2.3 mmol/L (200 mg/dL). In diabetic individuals with cardiovascular disease, the LDL goal is < 2.6 mmol/L (100 mg/dL). Because of the risk of cardiovascular disease in diabetic individuals, many authorities recommend that optimal lipid levels for all individuals with DM (with or without cardiovascular disease) should be: LDL <2.6 mmol/L (100 mg/dL), HDL > 1.15 mmol/L (45 mg/dL) in men and > 1.41 mmol/L (55 mg/dL) in women; and triglycerides < 2.3 mmol/L (200 mg/dL). The ADA recommends dietary modification in diabetic individuals without cardiovascular disease and a LDL cholesterol of 2.6 to 3.3 mmol/L (100 to 129 mg/dL). If multiple cardiovascular risk factors are present, the goal should be a LDL < 2.6 mmol/L (100 mg/dL) even without known cardiovascular disease.

Almost all studies of diabetic dyslipidemia have been performed in individuals with type 2 DM because of the greater frequency of dyslipidemia in this form of diabetes. Interventional studies have shown that the beneficial effects of LDL reduction are similar in the diabetic and nondiabetic populations. Large prospective trials of primary and secondary intervention for coronary heart disease have included a small number of individuals with type 2 DM, and subset analyses have consistently found that reductions in LDL reduce cardiovascular events and morbidity in individuals with DM (Fig. 333-CD1). Most clinical trials used HMG CoA reductase inhibitors, although a fibric acid derivative was also beneficial in one trial. No prospective studies have addressed similar questions in individuals with type 1 DM.

Based on the guidelines provided by the ADA and the American Heart Association, the order of priorities in the treatment of hyperlipidemia is: (1) lower the LDL cholesterol, (2) raise the HDL cholesterol, and (3) decrease the triglycerides. A treatment strategy depends on the pattern of lipoprotein abnormalities (Fig. 333-11). Initial therapy for all forms of dyslipidemia should include dietary changes, as well as the same life-style modifications recommended in the nondiabetic population (smoking cessation, control of blood pressure, weight loss, increased physical activity). The dietary recommendations for individuals with DM are similar to those advocated by the National Cholesterol

Education Program ([Chap. 344](#)) and include an increase in monounsaturated fat and carbohydrates and a reduction in saturated fats and cholesterol. Though viewed as important, the response to dietary alterations is often modest [<0.6-mmol/L (<25-mg/dL) reduction in the LDL]. Improvement in glycemic control will lower triglycerides and have a modest beneficial effect on raising HDL. Most medications that improve glycemic control are useful in lowering triglycerides and may raise the HDL slightly. Though fibric acid derivatives have some efficacy and are well tolerated, nicotinic acid may worsen glycemic control and increase insulin resistance; thus, niacin is relatively contraindicated in diabetic patients on oral glucose-lowering agents. As noted above, HMG CoA reductase inhibitors have proven benefit in patients with DM, even with modest elevations in LDL. Combination therapy with an HMG CoA reductase inhibitor and fibric acid derivative may be useful but increases the possibility of myositis. Bile acid-binding resins should not be used if hypertriglyceridemia is present.

Hypertension Hypertension can accelerate other complications of[DM](#), particularly cardiovascular disease and nephropathy. Hypertension therapy should first emphasize life-style modifications such as weight loss, exercise, stress management, and sodium restriction ([Chap. 35](#)). Antihypertensive agents should be selected based on the advantages and disadvantages of the therapeutic agent in the context of an individual patient's risk factor profile.[ACE](#)inhibitors are glucose- and lipid-neutral and thus positively impact the cardiovascular risk profile. For example, captopril actually improves insulin resistance, reduces[LDL](#)slightly, and increases[HDL](#)slightly. In one study of nondiabetic individuals, the ACE inhibitor ramipril reduced the risk of developing type 2 DM. Other effective agents include a-adrenergic blockers (prazocin, terazosin, doxazosin), calcium channel blockers, beta blockers (both b_1 selective and nonselective), thiazide diuretics (hydrochlorothiazide and its derivatives), central adrenergic antagonists (clonidine, methyldopa), and vasodilators (minoxidil, hydralazine). DM-related considerations include the following:

1. a-Adrenergic blockers slightly improve insulin resistance and positively impact the lipid profile, whereas beta blockers and thiazide diuretics can increase insulin resistance, negatively impact the lipid profile, and slightly increase the risk of developing type 2 diabetes.

2. Beta blockers, often questioned because of the potential masking of hypoglycemic symptoms, are effective agents and hypoglycemic events are rare when cardioselective (b_1) agents are used.

3. Central adrenergic antagonists and vasodilators are lipid- and glucose-neutral.

4. Sympathetic inhibitors and a-adrenergic blockers may be associated with orthostatic hypotension in the diabetic individual with autonomic neuropathy.

5. Calcium channel blockers are glucose- and lipid-neutral, and some evidence suggests that they reduce cardiovascular morbidity and mortality in type 2 DM, particularly in elderly patients with systolic hypertension.

If microalbuminuria or overt albuminuria is present, the optimal antihypertensive agent is an[ACE](#)inhibitor. If albumin excretion is normal, then an ACE inhibitor or other

antihypertensive agent may be used. Low-dose diuretics and beta blockers are sometimes preferred as initial agents because of their clear efficacy in the nondiabetic population. Since hypertension is often difficult to control with a single agent (especially in type 2 DM), multiple antihypertensive agents are usually required when blood pressure goals (<130/85 mmHg) are not achieved. In this setting, long-acting calcium channel antagonists should be considered as additional, or second-line, agents, as these drugs appear to provide protection against cardiovascular events. ACE inhibitors are contraindicated in pregnant diabetic patients and those anticipating pregnancy. Because of the high prevalence of atherosclerotic disease in individuals with DM, the possibility of renovascular hypertension should be considered when the blood pressure is not readily controlled.

LOWER EXTREMITY COMPLICATIONS

DM is the leading cause of nontraumatic lower extremity amputation in the United States. Foot ulcers and infections are also a major source of morbidity in individuals with DM. The reasons for the increased incidence of these disorders in DM are complex and involve the interaction of several pathogenic factors: neuropathy, abnormal foot biomechanics, peripheral vascular disease, and poor wound healing. The peripheral sensory neuropathy interferes with normal protective mechanisms and allows the patient to sustain major or repeated minor trauma to the foot, often without knowledge of the injury. Disordered proprioception causes abnormal weight bearing while walking and subsequent formation of callus or ulceration. Motor and sensory neuropathy leads to abnormal foot muscle mechanics and to structural changes in the foot (hammer toe, claw toe deformity, prominent metatarsal heads). Autonomic neuropathy results in anhidrosis and altered superficial blood flow in the foot, which promote drying of the skin and fissure formation. Peripheral vascular disease and poor wound healing impede resolution of minor breaks in the skin, allowing them to enlarge and to become infected.

Approximately 15% of individuals with DM develop a foot ulcer, and a significant subset of those individuals will at some time undergo amputation (14 to 24% risk with that ulcer or subsequent ulceration). Risk factors for foot ulcers or amputation include: male sex, diabetes >10 years' duration, peripheral neuropathy, abnormal structure of foot (bony abnormalities, callus, thickened nails), peripheral vascular disease, smoking, and history of previous ulcer or amputation. Glycemic control is also a risk factor -- each 2% increase in the HbA1c increases the risk of a lower extremity ulcer by 1.6 times and the risk of lower extremity amputation by 1.5 times.

TREATMENT

The optimal therapy for foot ulcers and amputations is prevention through identification of high-risk patients, education of the patient, and institution of measures to prevent ulceration. High-risk patients should be identified during the routine foot examination performed on all patients with DM (see "Ongoing Aspects of Comprehensive Diabetes Care," below). Patient education should emphasize: (1) careful selection of footwear, (2) daily inspection of the feet to detect early signs of poor-fitting footwear or minor trauma, (3) daily foot hygiene to keep the skin clean and moist, (4) avoidance of self-treatment of foot abnormalities and high-risk behavior (e.g., walking barefoot), and (5) prompt consultation with a health care provider if an abnormality arises. Patients at high risk for

ulceration or amputation may benefit from evaluation by a foot care specialist. Interventions directed at risk factor modification include orthotic shoes and devices, callus management, nail care, and prophylactic measures to reduce increased skin pressure from abnormal bony architecture. Attention to other risk factors for vascular disease (smoking, dyslipidemia, hypertension) and improved glycemic control are also important.

Despite preventive measures, foot ulceration and infection are common and represent a potentially serious problem. Due to the multifactorial pathogenesis of lower extremity ulcers, management of these lesions must be multidisciplinary and often demands expertise in orthopedics, vascular surgery, endocrinology, podiatry, and infectious diseases. The plantar surface of the foot is the most common site of ulceration. Cellulitis without ulceration is also frequent and should be treated with antibiotics that provide broad-spectrum coverage, including anaerobes (see below).

An infected ulcer is a clinical diagnosis, since superficial culture of any ulceration will likely find multiple possible bacterial pathogens. The infection surrounding the foot ulcer is often the result of multiple organisms (gram-positive and -negative cocci and anaerobes), and gas gangrene may develop in the absence of clostridial infection. Cultures taken from the debrided ulcer base or from purulent drainage are most helpful. Wound depth should be determined by inspection and probing with a blunt-tipped sterile instrument. Plain radiographs of the foot should be performed to assess the possibility of osteomyelitis in chronic ulcers that have not responded to therapy. Nuclear medicine bone scans may be helpful, but overlying subcutaneous infection is often difficult to distinguish from osteomyelitis. Indium-labeled white cell studies are more useful in determining if the infection involves bony structures or only soft tissue, but they are technically demanding. Magnetic resonance imaging of the foot may be the most specific modality, although distinguishing bony destruction due to osteomyelitis from destruction secondary to Charcot arthropathy is difficult. If surgical debridement is necessary, bone biopsy and culture usually provide the answer.

Osteomyelitis is best treated by a combination of prolonged antibiotics and debridement of infected bone. The possible contribution of vascular insufficiency should be considered in all patients. Noninvasive blood-flow studies are often unreliable in DM, and angiography may be required, recognizing the risk of contrast-induced nephrotoxicity. Peripheral vascular bypass procedures are often effective in promoting wound resolution and in decreasing the need for amputation of the ischemic limb.

A growing number of possible treatments for diabetic foot ulcers exist, but they have yet to demonstrate clear efficacy in prospective, controlled trials. A recent consensus statement from the ADA identified six interventions with demonstrated efficacy in diabetic foot wounds: (1) off-loading, (2) debridement, (3) wound dressings, (4) appropriate use of antibiotics, (5) revascularization, and (6) limited amputation. Off-loading is the complete avoidance of weight bearing on the ulcer, which removes the mechanical trauma that retards wound healing. Bed rest and a variety of orthotic devices limit weight bearing on wounds or pressure points. Surgical debridement of neuropathic wounds is important and effective, but clear efficacy of other modalities for wound cleaning (enzymes, soaking, whirlpools) is lacking. Dressings promote wound healing by creating a moist environment and protecting the wound. Antiseptic agents and topical antibiotics

should be avoided. Referral for physical therapy, orthotic evaluation, and rehabilitation may be useful once the infection is controlled.

Mild or non-limb-threatening infections can be treated with oral antibiotics (cephalosporin, clindamycin, amoxicillin/clavulanate, and fluoroquinolones), surgical debridement of necrotic tissue, local wound care (avoidance of weight bearing over the ulcer), and close surveillance for progression of infection. More severe ulcers may require intravenous antibiotics as well as bed rest and local wound care. Urgent surgical debridement may be required. Intravenous antibiotics should provide broad-spectrum coverage directed toward *Staphylococcus aureus*, streptococci, gram-negative aerobes, and anaerobic bacteria. Initial antimicrobial regimens include cefotetan, ampicillin/sulbactam, or the combination of clindamycin and a fluoroquinolone. Severe infections, or infections that do not improve after 48 h of antibiotic therapy, require expansion of antimicrobial therapy to treat methicillin-resistant *S. aureus* (e.g., vancomycin) and *Pseudomonas aeruginosa*. If the infection surrounding the ulcer is not improving with intravenous antibiotics, reassessment of antibiotic coverage and reconsideration of the need for surgical debridement or revascularization are indicated. With clinical improvement, oral antibiotics and local wound care can be continued on an outpatient basis with close follow-up. As infection improves, a comprehensive assessment of modifiable risk factors for foot ulceration should be performed and should involve health professionals with expertise in podiatry, orthotics, vascular surgery, and orthopedics.

New information about wound biology has led to a number of new technologies (e.g., living skin equivalents and growth factors such as basic fibroblast growth factor) that may prove useful. Recombinant platelet-derived growth factor has some benefit and complements the basic therapies of off-loading, debridement, and antibiotics. Hyperbaric oxygen has been used, but rigorous proof of efficacy is lacking.

INFECTIONS

Individuals with DM exhibit a greater frequency and severity of infection. The reasons for this increase include incompletely defined abnormalities in cell-mediated immunity and phagocyte function associated with hyperglycemia, as well as diminished vascularization secondary to long-standing diabetes. Hyperglycemia likely aids the colonization and growth of a variety of organisms (*Candida* and other fungal species). Many common infections are more frequent and severe in the diabetic population, whereas several rare infections are seen almost exclusively in the diabetic population. Examples of this latter category includes rhinocerebral mucormycosis and malignant otitis externa, which is usually secondary to *P. aeruginosa* infection in the soft tissue surrounding the external auditory canal. Malignant otitis externa begins with pain and discharge and may progress rapidly to osteomyelitis and meningitis. These infections should be sought, in particular, in patients presenting with NKHS.

Pneumonia, urinary tract infections, and skin and soft tissue infections are all more common in the diabetic population. In general, the organisms that cause pulmonary infections are similar to those found in the nondiabetic population; however, gram-negative organisms, *S. aureus*, and *Mycobacterium tuberculosis* are more frequent pathogens. Urinary tract infections (either lower tract or pyelonephritis) are the

result of common bacterial agents such as *Escherichia coli*, though several yeast species (*Candida* and *Torulopsis glabrata*) are commonly observed. Complications of urinary tract infections include emphysematous pyelonephritis and emphysematous cystitis. Bacteriuria occurs frequently in individuals with diabetic cystopathy. Susceptibility to furunculosis, superficial candidal infections, and vulvovaginitis is increased. Poor glycemic control is a common denominator in individuals with these infections. Diabetic individuals have an increased rate of colonization of *S. aureus* in the skin folds and nares. Diabetic patients also have a greater risk of postoperative wound infections.

DERMATOLOGIC MANIFESTATIONS

The most common skin manifestations of DM are protracted wound healing and skin ulcerations. Diabetic dermopathy, sometimes termed *pigmented pretibial papules*, or "diabetic skin spots," begins as an erythematous area and evolves into an area of circular hyperpigmentation (Fig. 333-CD2). These lesions result from minor mechanical trauma in the pretibial region and are more common in elderly men with DM. Bullous diseases (shallow ulcerations or erosions in the pretibial region) are also seen. *Necrobiosis lipoidica diabeticorum* is a rare disorder of DM that predominantly affects young women with type 1 DM, neuropathy, and retinopathy. It usually begins in the pretibial region as an erythematous plaque or papules that gradually enlarge, darken, and develop irregular margins, with atrophic centers and central ulceration. They may be painful. *Acanthosis nigricans* (hyperpigmented velvety plaques seen on the neck or extensor surfaces) is sometimes a feature of severe insulin resistance and accompanying diabetes (Fig. 333-CD3). Generalized or localized *granuloma annulare* (erythematous plaques on the extremities or trunk) and *scleredema* (areas of skin thickening on the back or neck at the site of previous superficial infections) are more common in the diabetic population. *Lipoatrophy* and *lipohypertrophy* can occur at insulin injection sites but are unusual with the use of human insulin. Xerosis and pruritus are common and are relieved by skin moisturizers.

Approach to the Patient

DM and its complications produce a wide range of symptoms and signs; those secondary to acute hyperglycemia may occur at any stage of the disease, whereas those related to chronic complications begin to appear during the second decade of hyperglycemia. Individuals with previously undetected type 2 DM may present with chronic complications of DM at the time of diagnosis. The history and physical examination should assess for symptoms or signs of acute hyperglycemia and should screen for the chronic complications and conditions associated with DM.

History A complete medical history should be obtained with special emphasis on DM-relevant aspects such as weight, family history of DM and its complications, risk factors for cardiovascular disease, prior medical conditions, exercise, smoking, and ethanol use. Symptoms of hyperglycemia include polyuria, polydipsia, weight loss, fatigue, weakness, blurry vision, frequent superficial infections (vaginitis, fungal skin infections), and slow healing of skin lesions after minor trauma. Metabolic derangements relate mostly to hyperglycemia (osmotic diuresis, reduced glucose entry into muscle) and to the catabolic state of the patient (urinary loss of glucose and

calories, muscle breakdown due to protein degradation and decreased protein synthesis). Blurred vision results from changes in the water content of the lens and resolves as the hyperglycemia is controlled.

In a patient with established DM, the initial assessment should also include special emphasis on prior diabetes care, including the type of therapy, prior HbA1c levels, self-monitoring blood glucose results, frequency of hypoglycemia, presence of DM-specific complications, and assessment of the patient's knowledge about diabetes. The chronic complications may afflict several organ systems, and an individual patient may exhibit some, all, or none of the symptoms related to the complications of DM (see above). In addition, the presence of DM-related comorbidities should be sought (cardiovascular disease, hypertension, dyslipidemia).

PHYSICAL EXAMINATION

In addition to a complete physical examination, special attention should be given to DM-relevant aspects such as weight or body mass index, retinal examination, orthostatic blood pressure, foot examination, peripheral pulses, and insulin injection sites. Careful examination of the lower extremities should seek evidence of peripheral neuropathy, calluses, superficial fungal infections, nail disease, and foot deformities (such as hammer or claw toes and Charcot foot) in order to identify sites of potential skin ulceration. Vibratory sensation (128-MHz tuning fork at the base of the great toe) and the ability to sense touch with a monofilament (5.07, 10-g monofilament) are useful to detect moderately advanced diabetic neuropathy. Since dental disease is more frequent in DM, the teeth and gums should also be examined.

Classification of DM in an Individual Patient The etiology of diabetes in an individual with new-onset disease can usually be assigned on the basis of clinical criteria. Individuals with type 1 DM tend to have the following characteristics: (1) onset of disease prior to age 30; (2) lean body habitus; (3) requirement of insulin as the initial therapy; (4) propensity to develop ketoacidosis; and (5) an increased risk of other autoimmune disorders such as autoimmune thyroid disease, adrenal insufficiency, pernicious anemia, and vitiligo. In contrast, individuals with type 2 DM often exhibit the following features: (1) develop diabetes after the age of 30; (2) are usually obese (80% are obese, but elderly individuals may be lean); (3) may not require insulin therapy initially; and (4) may have associated conditions such as insulin resistance, hypertension, cardiovascular disease, dyslipidemia, or polycystic ovary syndrome. In type 2 DM, insulin resistance is often associated with abdominal obesity (as opposed to hip and thigh obesity) and hypertriglyceridemia. Although most individuals diagnosed with type 2 DM are older, the age of diagnosis appears to be declining in some ethnic groups, and there is a marked increase among overweight teenagers. On the other hand, some individuals (<10%) with the phenotypic appearance of type 2 DM do not have absolute insulin deficiency but have autoimmune markers suggestive of type 1 DM. Thus, despite the revised classification of DM, it is remains difficult to categorize some patients unequivocally. Individuals who deviate from the clinical profile of type 1 and type 2 DM, or who have other associated defects such as deafness, pancreatic exocrine disease, and other endocrine disorders, should be classified accordingly (Table 333-1).

Laboratory Assessment The laboratory assessment should first determine whether

the patient meets the diagnostic criteria for DM (Table 333-2) and should then assess the degree of glycemic control (HbA1c, discussed below). In addition to the standard laboratory evaluation, the patient should be screened for DM-associated conditions (e.g., microalbuminuria, dyslipidemia, thyroid dysfunction). Individuals at high risk for cardiovascular disease should be screened for asymptomatic coronary artery disease by appropriate cardiac stress testing, when indicated.

The classification of the type of DM does not usually require laboratory assessments. Serum insulin or C-peptide measurements do not clearly distinguish type 1 from type 2 DM at the time of diabetes onset; a low C-peptide level merely confirms a patient's need for insulin. Conversely, many individuals with new-onset type 1 DM retain some C-peptide production. Measurement of islet cell antibodies at the time of diabetes onset may be useful if the type of DM is not clear based on the characteristics discussed above, but this knowledge does not usually alter therapy, which is based primarily on empirical metabolic features.

LONG-TERM TREATMENT

OVERALL PRINCIPLES

The goals of therapy for type 1 or type 2 DM are to: (1) eliminate symptoms related to hyperglycemia, (2) reduce or eliminate the long-term microvascular and macrovascular complications of DM, and (3) allow the patient to achieve as normal a life-style as possible. To reach these goals, the physician should identify a target level of glycemic control for each patient, provide the patient with the educational and pharmacologic resources necessary to reach this level, and monitor/treat DM-related complications. Symptoms of diabetes usually resolve when the plasma glucose is <11.1 mmol/L (200 mg/dL), and thus most DM treatment focuses on achieving the second and third goals.

The care of an individual with either type 1 or type 2 DM requires a multidisciplinary team. Central to the success of this team are the patient's participation, input, and enthusiasm, all of which are essential for optimal diabetes management. Members of the health care team include the primary care provider and/or the endocrinologist or diabetologist, a certified diabetes educator, and a nutritionist. In addition, when the complications of DM arise, subspecialists (including neurologists, nephrologists, vascular surgeons, cardiologists, ophthalmologists, and podiatrists) with experience in DM-related complications are essential.

A number of names are sometimes applied to different approaches to diabetes care, such as intensive insulin therapy, intensive glycemic control, and "tight control." The current chapter, however, will use the term *comprehensive diabetes care* to emphasize the fact that optimal diabetes therapy involves more than plasma glucose management. Though glycemic control is central to optimal diabetes therapy, comprehensive diabetes care of both type 1 and type 2 DM should also detect and manage DM-specific complications and modify risk factors for DM-associated diseases.

In addition to assessing the physical aspects of the patient with DM, the physician and members of the diabetes management team should consider social, family, financial, cultural, and employment-related issues that may have an impact on diabetes care.

With this information, the physician can work with the patient and his or her family to establish therapeutic goals and design a comprehensive and feasible plan for optimal diabetes care.

EDUCATION OF THE PATIENT ABOUT DM, NUTRITION, AND EXERCISE

Patient participation is an essential component of comprehensive diabetes care. The patient with type 1 or type 2 DM should receive education about nutrition, exercise, care of diabetes during illness, and medications to lower the plasma glucose. Along with improved compliance, patient education allows individuals with DM to assume greater responsibility for their care. Patient education should be viewed as a continuing process with regular visits for reinforcement; it should *not* be a process that is completed after one or two visits to a nurse educator or nutritionist.

Diabetes Education The diabetes educator is a health care professional (nurse, dietician, or pharmacist) with specialized patient education skills who is certified in diabetes education (indicating demonstrated skills in diabetes knowledge and education and certification by the American Association of Diabetes Educators). The educator is a vital member of the comprehensive diabetes care program and educates the patient about a number of issues important for optimal diabetes care, including self-monitoring of blood glucose; urine ketone monitoring (type 1 DM); insulin administration; guidelines for diabetes management during illnesses; management of hypoglycemia; foot and skin care; diabetes management before, during, and after exercise; and risk factor-modifying activities.

Nutrition *Medical nutrition therapy* (MNT) is a term used by the ADA to describe the optimal coordination of caloric intake with other aspects of diabetes therapy (insulin, exercise, weight loss). Historically, nutrition has imposed restrictive, complicated regimens on the patient. Current practices have greatly changed, though many patients and health care providers still view the diabetic diet as monolithic and static. For example, modern MNT now includes foods with sucrose and seeks to modify other risk factors such as hyperlipidemia and hypertension rather than focusing exclusively on weight loss in individuals with type 2 DM. Like other aspects of DM therapy, MNT must be adjusted to meet the goals of the individual patient. Furthermore, MNT education is an important component of comprehensive diabetes care and should be reinforced by regular patient education. In general, the components of optimal MNT are similar for individuals with type 1 or type 2 DM (Table 333-8).

The goal of MNT in the individual with type 1 DM is to coordinate and match the caloric intake, both temporally and quantitatively, with the appropriate amount of insulin. MNT in type 1 DM and self-monitoring of blood glucose must be integrated to define the optimal insulin regimen. MNT must be flexible enough to allow for exercise, and the insulin regimen must allow for deviations in caloric intake. An important component of MNT in type 1 DM is to minimize the weight gain often associated with intensive diabetes management.

The goals of MNT in type 2 DM are slightly different and address the greatly increased prevalence of cardiovascular risk factors (hypertension, dyslipidemia, obesity) and disease in this population. The majority of these individuals are obese, and weight loss

is still strongly encouraged and should remain an important goal. Medical treatment of obesity is a rapidly evolving area and is discussed in Chap. 77. Hypocaloric diets and modest weight loss often result in rapid and dramatic glucose lowering in individuals with new-onset type 2 DM. Nevertheless, numerous studies document that long-term weight loss is uncommon. Therefore, current MNT for type 2 DM should emphasize modest caloric reduction, increased physical activity, and reduction of hyperlipidemia and hypertension. Increased consumption of soluble, dietary fiber may improve glycemic control in individuals with type 2 DM.

Exercise Exercise is an integral component of comprehensive diabetes care that can have multiple positive benefits (cardiovascular benefits, reduced blood pressure, maintenance of muscle mass, reduction in body fat, weight loss, etc.). For individuals with type 1 or type 2 DM, exercise is also useful for lowering plasma glucose (during and following exercise) and increasing insulin sensitivity.

Despite its benefits, exercise presents several challenges for individuals with DM because they lack the normal glucoregulatory mechanisms. Skeletal muscle is a major site for metabolic fuel consumption in the resting state, and the increased muscle activity during vigorous, aerobic exercise greatly increases fuel requirements. Individuals with type 1 DM are prone to either hyperglycemia or hypoglycemia during exercise, depending on the preexercise plasma glucose, the circulating insulin level, and the level of exercise-induced catecholamines. If the insulin level is too low, the rise in catecholamines may increase the plasma glucose excessively, promote ketone body formation, and possibly lead to ketoacidosis. Conversely, if the circulating insulin level is excessive, this relative hyperinsulinemia may reduce hepatic glucose production (decreased glycogenolysis, decreased gluconeogenesis) and increase glucose entry into muscle, leading to hypoglycemia.

To avoid exercise-related hyper- or hypoglycemia, individuals with type 1 DM should: (1) monitor blood glucose before, during, and after exercise; (2) delay exercise if blood glucose is >14 mmol/L (250 mg/dL), <5.5 mmol/L (100 mg/dL), or if ketones are present; (3) eat a meal 1 to 3 h before exercise and take supplemental carbohydrate feedings at least every 30 min during vigorous or prolonged exercise; (4) decrease insulin doses (based on previous experience) before exercise and inject insulin into a nonexercising area; and (5) learn individual glucose responses to different types of exercise and increase food intake for up to 24 h after exercise, depending on intensity and duration of exercise. In individuals with type 2 DM, exercise-related hypoglycemia is less common but can occur in individuals taking either insulin or sulfonylureas.

Because asymptomatic cardiovascular disease appears at a younger age in both type 1 and type 2 DM, formal exercise tolerance testing may be warranted in diabetic individuals with any of the following: age ³35 years, long-standing type 1 DM (>20 to 25 years' duration), microvascular complications of DM (retinopathy, microalbuminuria, or nephropathy), peripheral vascular disease, other risk factors of coronary artery disease, or autonomic neuropathy. Untreated proliferative retinopathy is a relative contraindication to vigorous exercise, since this may lead to vitreous hemorrhage or retinal detachment.

MONITORING THE LEVEL OF GLYCEMIC CONTROL

Optimal monitoring of glycemic control involves plasma glucose measurements by the patient and an assessment of long-term control by the physician (measurement of HbA1c and review of the patient's self-measurements of plasma glucose). These measurements are complementary: the patient's measurements provide a picture of short-term glycemic control, whereas the HbA1c reflects average glycemic control over the previous 2 to 3 months. Integration of both measurements provides an accurate assessment of the glycemic control achieved.

Self-Monitoring of Blood Glucose Self-monitoring of blood glucose (SMBG) is the standard of care in diabetes management and allows the patient to monitor his or her blood glucose at any time. In SMBG, a small drop of blood and an easily detectable enzymatic reaction allow measurement of the capillary plasma glucose. By combining glucose measurements with diet history, medication changes, and exercise history, the physician and patient can improve the treatment program.

The frequency of SMBG measurements must be individualized and adapted to address the goals of diabetes care as defined by the patient and the health care provider. Individuals with type 1 DM should routinely measure their plasma glucose four to eight times per day to estimate and select mealtime boluses of short-acting insulin and to modify long-acting insulin doses. Most individuals with type 2 DM require less frequent monitoring, though the optimal frequency of SMBG has not been clearly defined. Individuals with type 2 DM who are on oral medications should utilize SMBG as a means of assessing the efficacy of their medication and diet. Since plasma glucose levels fluctuate less in these individuals, one to two SMBG measurements per day (or fewer) may be sufficient. Individuals with type 2 DM who are on insulin should utilize SMBG more frequently than those on oral agents.

Two devices for continuous blood glucose monitoring have been recently approved by the U.S. Food and Drug Administration (FDA). The Glucowatch uses iontophoresis to assess glucose in interstitial fluid, whereas the Minimed device uses an indwelling subcutaneous catheter to monitor interstitial fluid glucose. Both devices utilize immobilized glucose oxidase to generate electrons in response to changing glucose levels. Though clinical experience with these devices is limited, they perform well in clinical trials and appear to provide useful short-term information about the patterns of glucose changes as well as an enhanced ability to detect hypoglycemic episodes.

Although urine glucose testing does not provide an accurate assessment of glycemic control, urine ketones are a sensitive indicator of early diabetic ketoacidosis and should be measured in individuals with type 1 DM when the plasma glucose is consistently >16.7 mmol/L (300 mg/dL); during a concurrent illness; or with symptoms such as nausea, vomiting, or abdominal pain.

Assessment of Long-Term Glycemic Control Measurement of glycated hemoglobin is the standard method for assessing long-term glycemic control. When plasma glucose is consistently elevated, there is an increase in nonenzymatic glycation of hemoglobin; this alteration reflects the glycemic history over the previous 2 to 3 months, since erythrocytes have an average life span of 120 days. There are numerous laboratory methods for measuring the various forms of glycated hemoglobin, and these have

significant interassay variations. Because of its superior specificity and reliability, the HbA1c assay performed by the high-performance liquid chromatography (HPLC) method has become the standard reference method for most glycated hemoglobin measurements. Since glycated hemoglobin measurements are usually compared to prior measurements, it is essential for the assay results to be comparable. Depending on the assay methodology for HbA1c, hemoglobinopathies, hemolytic anemia, and uremia may interfere with the HbA1c result.

Glycated hemoglobin or HbA1c should be measured in all individuals with DM during their initial evaluation and as part of their comprehensive diabetes care. As the primary predictor of long-term complications of DM, the HbA1c should mirror, to a certain extent, the short-term measurements of SMBG. These two measurements are complementary in that recent intercurrent illnesses may impact the SMBG measurements but not the HbA1c. Likewise, postprandial and nocturnal hyperglycemia may not be detected by the SMBG of fasting and preprandial capillary plasma glucose but will be reflected in the HbA1c. When measured by HPLC, the HbA1c approximates the following mean plasma glucose values: an HbA1c of 6% is 6.6 mmol/L (120 mg/dL), 7% is 8.3 mmol/L (150 mg/dL), 8% is 10.0 mmol/L (180 mg/dL), etc. [A 1% rise in the HbA1c translates into a 1.7-mmol/L (30 mg/dL) increase in the mean glucose.] The degree of glycation of other proteins, such as albumin, has been used as an alternative indicator of glycemic control when the HbA1c is inaccurate (hemolytic anemia, hemoglobinopathies). The fructosamine assay (using albumin) is an example of an alternative measurement of glycemic control and reflects the glycemic status over the 2 to 4 prior weeks. Current consensus statements do not favor the use of alternative assays of glycemic control, as there are no studies to indicate whether such assays accurately predict the complications of DM.

TREATMENT

Establishment of a Target Level of Glycemic Control Because the complications of DM are related to glycemic control, normoglycemia or near normoglycemia is the desired, but often elusive, goal for most patients. However, normalization of the plasma glucose for long periods of time is extremely difficult, as demonstrated by the DCCT. Regardless of the level of hyperglycemia, improvement in glycemic control will lower the risk of diabetes complications (Fig. 333-8).

The target for glycemic control (as reflected by the HbA1c) must be individualized, and the health care provider should establish the goals of therapy in consultation with the patient after considering a number of medical, social, and life-style issues. Some important factors to consider include the patient's age, ability to understand and implement a complex treatment regimen, presence and severity of complications of diabetes, ability to recognize hypoglycemic symptoms, presence of other medical conditions or treatments that might alter the response to therapy, life-style and occupation (e.g., possible consequences of experiencing hypoglycemia on the job), and level of support available from family and friends.

The ADA has established suggested glycemic goals based on the premise that glycemic control predicts development of DM-related complications. In general, the target HbA1c should be<7.0% (Table 333-9). Other consensus groups (such as the Veterans

Administration) have suggested HbA1c goals that take into account the patient's life expectancy at the time of diagnosis and the presence of microvascular complications. Such recommendations strive to balance the financial and personal costs of glycemic therapy with anticipated benefits (reduced health care costs, reduced morbidity). One limitation to this approach is that the onset of hyperglycemia in type 2 DM is difficult to ascertain and likely predates the diagnosis. Furthermore, though the life expectancy can be predicted for a patient population, the physician must treat an individual patient; consequently, the target HbA1c must be individualized to accommodate these other considerations.

Type 1 Diabetes Mellitus

General Aspects Comprehensive diabetes care should be instituted in all individuals with type 1DM and should involve attention to nutrition, exercise, and risk factor management in addition to insulin administration. TheADArecommendations for fasting and bedtime glycemic goals and HbA1c targets are summarized in Table 333-9. The goal is to design and implement insulin regimens that mimic physiologic insulin secretion. Because individuals with type 1 DM lack endogenous insulin production, administration of basal, exogenous insulin is essential for regulating glycogen breakdown, gluconeogenesis, lipolysis, and ketogenesis. Likewise, postprandial insulin replacement should be appropriate for the carbohydrate intake and promote normal glucose utilization and storage.

Intensive Management Intensive diabetes management is defined by theADA as "...a mode of treatment for the person withDM that has the goal of achieving euglycemia or near-normal glycemia using all available resources to accomplish this goal." These resources include thorough and continuing patient education, comprehensive recording of plasma glucose measurements and nutrition intake by the patient, and a variable insulin regimen that matches glucose intake and insulin dose. Insulin regimens usually include multiple-component insulin regimens, multiple daily injections (MDI), or insulin infusion devices (all discussed below).

The benefits of intensive diabetes management and improved glycemic control include a reduction in the microvascular complications ofDM and a possible delay or reduction in the macrovascular complications of DM. From a psychological standpoint, the patient experiences greater control over his or her diabetes and often notes an improved sense of well-being, greater flexibility in the timing and content of meals, and the capability to alter insulin dosing with exercise. In addition, intensive diabetes management in pregnancy reduces fetal malformation and morbidity. Intensive diabetes management is also strongly encouraged in newly diagnosed patients with type 1 DM because it may prolong the period of C-peptide production, which may result in better glycemic control and a reduced risk of serious hypoglycemia.

Although intensive management confers impressive benefits, it is also accompanied by significant personal and financial costs and is therefore not appropriate for all individuals. It requires a combination of dedication, persistence, and motivation on the part of the patient, as well as medical, educational, nursing, nutritional, and psychological expertise on the part of the diabetes management team. Circumstances in which intensive diabetes management should be strongly considered are listed in

.

Insulin Preparations Current insulin preparations are generated by recombinant DNA technology and consist of the amino acid sequence of human insulin. Animal insulin (beef or pork) is no longer used. Human insulin has been formulated with distinctive pharmacokinetics to mimic physiologic insulin secretion (Table 333-11). In the United States, all insulin is formulated as U-100 (100 units/mL), whereas in some other countries it is available in other units (e.g., U-40 = 40 units/mL). One short-acting insulin formulation, lispro, is an insulin analogue in which the 28th and 29th amino acids (lysine and proline) on the insulin B chain have been reversed by recombinant DNA technology. This insulin analogue has full biologic activity but less tendency toward subcutaneous aggregation, resulting in more rapid absorption and onset of action and a shorter duration of action. These characteristics are particularly advantageous for allowing entrainment of insulin injection and action to rising plasma glucose levels following meals, although improvement in HbA1c values have not been found consistently. The shorter duration of action also appears to be associated with a decreased number of hypoglycemic episodes, primarily because the decay of lispro action corresponds better to the decline in plasma glucose after a meal. Insulin glargine is a long-acting biosynthetic human insulin that differs from normal insulin in that asparagine is replaced by glycine at amino acid 21, and two arginine residues are added to the C-terminus of the B chain. Compared to NPH insulin, the onset of insulin glargine action is later, the duration of action is longer (~24 h), and there is no pronounced peak. A lower incidence of hypoglycemia, especially at night, was reported in one trial with insulin glargine when compared to NPH insulin. Since glargine has only recently approved, clinical experience is limited. Additional insulin analogues are currently under development.

Basal insulin requirements are provided by intermediate (NPH or lente) or long-acting (ultralente or glargine) insulin formulations. These are usually combined with short-acting insulin in an attempt to mimic physiologic insulin release with meals. Although mixing of intermediate and short-acting insulin formulations is common practice, this mixing may alter the insulin absorption profile (especially those of short-acting insulins). For example, the absorption of regular insulin is delayed when mixed for even short periods of time (<5 min) with lente or ultralente insulin, but not when mixed with NPH insulin. Lispro absorption is delayed by mixing with NPH but not ultralente. Insulin glargine should not be mixed with other insulins. The miscibility of human regular and NPH insulin allows for the production of combination insulins that contain 75% NPH and 25% regular (75/25), 70% NPH and 30% regular (70/30), or equal mixtures of NPH and regular. These combinations of insulin are more convenient for the patient but prevent adjustment of only one component of the insulin formulation. The alteration in insulin absorption when the patient mixes different insulin formulation should not discourage the patient from mixing insulin. However, the following guidelines should be followed: (1) mix the different insulin formulations in the syringe immediately before injection (inject within 2 min after mixing); (2) if possible, do not store insulin as a mixture; and (3) follow the same routine in terms of insulin mixing and administration to standardize the physiologic response to injected insulin.

Insulin Regimens Representations of the various insulin regimens that may be utilized in type 1DM are illustrated in Fig. 333-12. Although the insulin profiles are depicted as

"smooth," symmetric curves, there is considerable patient-to-patient variation in the peak and duration. In all regimens, long-acting insulins (NPH, lente, ultralente, or glargine insulin) supply basal insulin, whereas prandial insulin is provided by either regular or lispro insulin. Lispro should be injected just before a meal; regular insulin is given 30 to 45 min prior to a meal.

A shortcoming of current insulin regimens is that injected insulin immediately enters the systemic circulation, whereas endogenous insulin is secreted into the portal vein. Thus, exogenous insulin administration exposes the liver to subphysiologic insulin levels. No insulin regimen reproduces the precise insulin secretory pattern of the pancreatic islet. However, the most physiologic regimens entail more frequent insulin injections, greater reliance on short-acting insulin, and more frequent capillary plasma glucose measurements. In general, individuals with type 1DMrequire 0.5 to 1.0 U/kg per day of insulin divided into multiple doses. Initial insulin-dosing regimens should be conservative; approximately 40 to 50% of the insulin should be given as basal insulin. A single daily injection of insulin is not appropriate therapy in type 1 DM.

One commonly used regimen consists of twice-daily injections of an intermediate insulin (NPH or lente) mixed with a short-acting insulin before the morning and evening meal (Fig. 333-12A). Such regimens usually prescribe two-thirds of the total daily insulin dose in the morning (with about two-thirds given as intermediate-acting insulin and one-third as short-acting) and one-third before the evening meal (with approximately one-half given as intermediate-acting insulin and one-half as short-acting). The drawback to such a regimen is that it enforces a rigid schedule on the patient, in terms of daily activity and the content and timing of meals. Although it is simple and effective at avoiding severe hyperglycemia, it does not generate near-normal glycemic control in most individuals with type 1DM. Moreover, if the patient's meal pattern or content varies or if physical activity is increased, hyperglycemia or hypoglycemia may result. Moving the intermediate insulin from before the evening meal to bedtime may avoid nocturnal hypoglycemia and provide more insulin as glucose levels rise in the early morning (so-called dawn phenomenon). The insulin dose in such regimens should be adjusted based onSMBGresults with the following general assumptions: (1) the fasting glucose is primarily determined by the prior evening intermediate-acting insulin; (2) the pre-lunch glucose is a function of the morning short-acting insulin; (3) the pre-supper glucose is a function of the morning intermediate-acting insulin; and (4) the bedtime glucose is a function of the pre-supper, short-acting insulin.

Multiple-component insulin regimens refer to the combination of basal insulin; preprandial short-acting insulin; and changes in short-acting insulin doses to accommodate the results of frequentSMBG, anticipated food intake, and physical activity. Sometimes also referred to as *multiple daily injections*, such regimens offer the patient maximal flexibility in terms of life-style and the best chance for achieving near normoglycemia. One such regimen, shown inFig. 333-12B, consists of a basal insulin with ultralente twice a day and preprandial lispro. The lispro dose is based on individualized algorithms that integrate the preprandial glucose and the anticipated carbohydrate intake. An alternative multiple-component insulin regimen consists of bedtime intermediate insulin, a small dose of intermediate insulin at breakfast (20 to 30% of bedtime dose), and preprandial short-acting insulin. There are numerous variations of these regimens that can be optimized for individual patients. Frequent

SMBG (four to 8 times per day) is absolutely essential for these types of insulin regimens.

Continuous subcutaneous insulin infusion (CSII) is another multiple-component insulin regimen (Fig. 333-12*C*). Sophisticated insulin infusion devices are now available that can accurately deliver small doses of insulin (microliters per hour). For example, multiple basal infusion rates can be programmed to: (1) accommodate nocturnal versus daytime basal insulin requirement, (2) alter infusion rate during periods of exercise, or (3) select different waveforms of insulin infusion. A preprandial insulin ("bolus") is delivered by the insulin infusion device based on instructions from the patient, which follow individualized algorithms that account for preprandial plasma glucose and anticipated carbohydrate intake. These devices require a health professional with considerable experience with insulin infusion devices and very frequent patient interactions with the diabetes management team. Insulin infusion devices present unique challenges, such as infection at the infusion site, unexplained hyperglycemia because the infusion set becomes obstructed, or diabetic ketoacidosis if the pump becomes disconnected. Since most physicians use lispro insulin in CSII, the extremely short half-life of this insulin quickly leads to insulin deficiency if the delivery system is interrupted. Essential to the safe use of infusion devices is thorough patient education about pump function and frequentSMBG.

Type 2 Diabetes Mellitus

General Aspects The goals of therapy for type 2DM are similar to those in type 1: improved glycemic control with near normalization of the HbA1c. While glycemic control tends to dominate the management of type 1 DM, the care of individuals with type 2 DM must also include attention to the treatment of conditions associated with type 2 DM (obesity, hypertension, dyslipidemia, cardiovascular disease) and detection/management of DM-related complications (Fig. 333-13). DM-specific complications may be present in up to 20 to 50% of individuals with newly diagnosed type 2 DM. Reduction in cardiovascular risk is of paramount importance as this is the leading cause of mortality in these individuals.

Diabetes management should begin withMNT(discussed above). An exercise regimen to increase insulin sensitivity and promote weight loss should also be instituted. After MNT and increased physical activity have been instituted, glycemic control should be reassessed; if the patient's glycemic target is not achieved after 3 to 4 weeks of MNT, pharmacologic therapy is indicated. Pharmacologic approaches to the management of type 2DMinclude both oral glucose-lowering agents and insulin; most physicians and patients prefer oral glucose-lowering agents as the initial choice. Any therapy that improves glycemic control reduces "glucose toxicity" to the islet cells and improves endogenous insulin secretion.

Glucose-Lowering Agents Recent advances in the therapy of type 2DM have generated considerable enthusiasm for oral glucose-lowering agents that target different pathophysiologic processes in type 2 DM. Based on their mechanisms of action, oral glucose-lowering agents are subdivided into agents that increase insulin secretion, reduce glucose production, or increase insulin sensitivity (Table 333-12). Oral glucose-lowering agents (with the exception of a-glucosidase inhibitors) are ineffective

in type 1 DM and should not be used for glucose management of severely ill individuals with type 2 DM. Insulin is sometimes the initial glucose-lowering agent.

INSULIN SECRETAGOGUES Insulin secretagogues stimulate insulin secretion by interacting with the ATP-sensitive potassium channel on the beta cell (Fig. 333-1). These drugs are most effective in individuals with type 2DM of relatively recent onset (<5 years), who have endogenous insulin production and tend to be obese. At maximum doses, first-generation sulfonylureas are similar in potency to second-generation agents but have a longer half-life, a greater incidence of hypoglycemia, and more frequent drug interactions (Table 333-13). Thus, second-generation sulfonylureas are generally preferred. An advantage to a more rapid onset of action is better coverage of the postprandial glucose rise, but the shorter half-life of such agents requires more than once-a-day dosing. Sulfonylureas reduce both fasting and postprandial glucose and should be initiated at low doses and increased at 1- to 2-week intervals based onSMBG. In general, sulfonylureas increase insulin acutely and thus should be taken shortly before a meal; with chronic therapy, though, the insulin release is more sustained. Replaglinide is not a sulfonylurea but also interacts with the ATP-sensitive potassium channel. Because of its short half-life, it is usually given with or immediately before each meal to reduce meal-related glucose excursions.

Insulin secretagogues are well tolerated in general. All of these agents, however, have the potential to cause profound and persistent hypoglycemia, especially in elderly individuals. Hypoglycemia is usually related to delayed meals, increased physical activity, alcohol intake, or renal insufficiency. Individuals who ingest an overdose of these agents develop prolonged and serious hypoglycemia and should be monitored closely in the hospital (Chap. 334). Most sulfonylureas are metabolized in the liver to compounds that are cleared by the kidney. Thus, their use in individuals with significant hepatic or renal dysfunction is not advisable. Weight gain, a common side effect of sulfonylurea therapy, results from the increased insulin levels and improvement in glycemic control. Some sulfonylureas have significant drug interactions with other medications such as alcohol, warfarin, aspirin, ketoconazole, a-glucosidase inhibitors, and fluconazole. Despite prior concerns that use of sulfonylureas might increase cardiovascular risk, recent trials have refuted this claim.

BIGUANIDES Metformin is representative of this class of agents. It reduces hepatic glucose production through an undefined mechanism and may improve peripheral glucose utilization slightly (Table 333-12). Metformin reduces fasting plasma glucose and insulin levels, improves the lipid profile, and promotes modest weight loss. The initial starting dose of 500 mg once or twice a day can be increased to 850 mg tid or 1000 mg bid. Because of its relatively slow onset of action and gastrointestinal symptoms with higher doses, the dose should be escalated every 2 to 3 weeks based onSMBGmeasurements. The major toxicity of metformin, lactic acidosis, can be prevented by careful patient selection. Metformin should not be used in patients with renal insufficiency [serum creatinine >133 umol/L (1.5 mg/dL) in men or >124 umol/L (1.4 mg/dL) in women, with adjustments for age], any form of acidosis, congestive heart failure, liver disease, or severe hypoxia. Metformin should be discontinued in patients who are seriously ill, in patients who can take nothing orally, and in those receiving radiographic contrast material. Insulin should be used until metformin can be restarted. Though well tolerated in general, some individuals develop gastrointestinal side effects

(diarrhea, anorexia, nausea, and metallic taste) that can be minimized by gradual dose escalation. Because the drug is metabolized in the liver, it should not be used in patients with liver disease or heavy ethanol intake.

a-GLUCOSIDASE INHIBITORS a-Glucosidase inhibitors (acarbose and miglitol) reduce postprandial hyperglycemia by delaying glucose absorption; they do not affect glucose utilization or insulin secretion (Table 333-12). Postprandial hyperglycemia, secondary to impaired hepatic and peripheral glucose disposal, contributes significantly to the hyperglycemic state in type 2DM. These drugs, taken just before each meal, reduce glucose absorption by inhibiting the enzyme that cleaves oligosaccharides into simple sugars in the intestinal lumen. Therapy should be initiated at a low dose (25 mg of acarbose or miglitol) with the evening meal and may be increased to a maximal dose over weeks to months (50 to 100 mg for acarbose or 50 mg for miglitol with each meal). The major side effects (diarrhea, flatulence, abdominal distention) are related to increased delivery of oligosaccharides to the large bowel and can be reduced somewhat by gradual upward dose titration.a-Glucosidase inhibitors may increase levels of sulfonylureas and increase the incidence of hypoglycemia. Simultaneous treatment with bile acid resins and antacids should be avoided. These agents should not be used in individuals with inflammatory bowel disease, gastroparesis, or a serum creatinine >177 umol/L (2.0 mg/dL). This class of agents is not as potent as other oral agents in lowering the HbA1c but is unique in that it reduces the postprandial glucose rise even in individuals with type 1 DM.

THIAZOLIDINEDIONES Thiazolidinediones represent a new class of agents that reduce insulin resistance. These drugs bind to a nuclear receptor (peroxisome proliferator-activated receptor, PPAR-g) that regulates gene transcription. The PPAR-g receptor is found at highest levels in adipocytes but is expressed at lower levels in many other insulin-sensitive tissues. Agonists of this receptor promote adipocyte differentiation and may reduce insulin resistance in skeletal muscle indirectly. Thiazolidinediones reduce the fasting plasma glucose by improving peripheral glucose utilization and insulin sensitivity (Table 333-12). Circulating insulin levels decrease with use of the thiazolidinediones, indicating a reduction in insulin resistance. Although direct comparisons are not available, the two currently available thiazolidinediones appear to have similar efficacy; the therapeutic range for pioglitazone is 15 to 45 mg/d in a single daily dose and for rosiglitazone is 2 to 8 mg/d -- once a day at lower doses and bid at higher doses. The ability of thiazolidinediones to influence other features of the insulin resistance syndrome is under investigation.

The prototype of this class of drugs, troglitazone, was withdrawn from the U.S. market after reports of hepatotoxicity and an association with an idiosyncratic liver reaction that sometimes led to hepatic failure. The two other thiazolidinediones, rosiglitazone and pioglitazone, thus far do not appear to induce the liver abnormalities seen with troglitazone. However, long-term experience with the newer agents is limited. Consequently, theFDArecommends measurement of liver function tests prior to initiating therapy with a thiazolidinedione and at regular intervals (every two months for the first year and then periodically). The thiazolidinediones raiseLDL andHDLslightly and lower triglycerides by 10 to 15%, but the clinical significance of these changes is not known. Thiazolidinediones are associated with minor weight gain (1 to 2 kg), a small reduction in the hematocrit, and a mild increase in plasma volume. Cardiac function is not

affected, but the incidence of peripheral edema is increased. They are contraindicated in patients with liver disease or congestive heart failure (class III or IV). Thiazolidinediones have been shown to induce ovulation in premenopausal women with polycystic ovary syndrome (see "Insulin Resistance Syndromes," above). Women should be warned about the risk of pregnancy, since the safety of thiazolidinediones in pregnancy is not established.

INSULIN THERAPY IN TYPE 2 DM Modest doses of insulin are quite efficacious in controlling hyperglycemia in newly diagnosed type 2 DM. Insulin should be considered as the initial therapy in type 2 DM, particularly in lean individuals or those with severe weight loss, in individuals with underlying renal or hepatic disease that precludes oral glucose-lowering agents, or in individuals who are hospitalized or acutely ill. Insulin therapy is ultimately required by a substantial number of individuals with type 2 DM because of the progressive nature of the disorder and the relative insulin deficiency that develops in patients with long-standing diabetes.

Because endogenous insulin secretion continues and is capable of providing some coverage of mealtime caloric intake, insulin is usually initiated in a single dose of intermediate-acting insulin (0.3 to 0.4 U/kg per day), given either before breakfast or just before bedtime (or ultralente at bedtime). Since fasting hyperglycemia and increased hepatic glucose production are prominent features of type 2 DM, bedtime insulin is more effective in clinical trials than a single dose of morning insulin. Some physicians prefer a relatively low, fixed starting dose of intermediate-acting insulin (~15 to 20 units in the morning and 5 to 10 units at bedtime) to avoid hypoglycemia. The insulin dose may then be adjusted in 10% increments as dictated by SMBG results. Both morning and bedtime intermediate insulin may be used in combination with oral glucose-lowering agents (biguanides, a-glucosidase inhibitors, or thiazolidinediones).

CHOICE OF INITIAL GLUCOSE-LOWERING AGENT Though insulin is an effective primary therapy for type 2 DM, most patients and physicians currently prefer oral glucose-lowering drugs as the initial pharmacologic approach. The level of hyperglycemia should influence the initial choice of therapy. Assuming maximal benefit of MNT and increased physical activity has been realized, patients with mild to moderate hyperglycemia [fasting plasma glucose <11.1 to 13.9 mmol/L (200 to 250 mg/dL)] often respond well to a single oral glucose-lowering agent. Patients with more severe hyperglycemia [fasting plasma glucose >13.9 mmol/L (250 mg/dL)] may respond partially but are unlikely to achieve normoglycemia with oral monotherapy. Nevertheless, many physicians prefer a stepwise approach that starts with a single agent and adds a second agent to achieve the glycemic target (see "Combination Therapy," below). Some physicians begin insulin in individuals with severe hyperglycemia [fasting plasma glucose >13.9 to 16.7 mmol/L (250 to 300 mg/dL)]. This approach is based on the rationale that more rapid glycemic control will reduce "glucose toxicity" to the islet cells, improve endogenous insulin secretion, and possibly allow oral glucose-lowering agents to be more effective. If this occurs, the insulin may be discontinued.

Insulin secretagogues, biguanides, a-glucosidase inhibitors, thiazolidinediones, and insulin are approved for monotherapy of type 2 DM. Although each class of oral glucose-lowering agents has unique advantages and disadvantages, certain generalizations apply: (1) insulin secretagogues, biguanides, and thiazolidinediones

improve glycemic control to a similar degree (1 to 2% reduction in HbA1c) and are more effective thana-glucosidase inhibitors; (2) assuming a similar degree of glycemic improvement, no clinical advantage to one class of drugs has been demonstrated, and any therapy that improves glycemic control is beneficial; (3) insulin secretagogues and a-glucosidase inhibitors begin to lower the plasma glucose immediately, whereas the glucose-lowering effects of the biguanides and thiazolidinediones are delayed by several weeks to months; (4) not all agents are effective in all individuals with type 2 DM (primary failure); (5) biguanides, a-glucosidase inhibitors, and thiazolidinediones do not directly cause hypoglycemia; and (6) most individuals will eventually require treatment with more than one class of oral glucose-lowering agents, reflecting the progressive nature of type 2 DM.

Considerable clinical experience exists with sulfonylureas and metformin because they have been available for several decades. It is assumed that the a-glucosidase inhibitors and thiazolidinediones, which are newer classes of oral glucose-lowering drugs, will reduceDM-related complications by improving glycemic control, although long-term data are not yet available. The thiazolidinediones are theoretically attractive because they target a fundamental abnormality in type 2 DM, namely insulin resistance. However, these agents are currently more costly than others and require liver function monitoring.

A reasonable treatment algorithm for initial therapy proposes either a sulfonylurea or metformin as initial therapy because of their efficacy, known side-effect profile, and relatively low cost (Fig. 333-14). Metformin has the advantage that it promotes mild weight loss, lowers insulin levels, improves the lipid profile slightly, and may have a lower secondary failure rate. However, there is no difference in response rate or degree of glycemic control when metformin and sulfonylureas are compared in randomized, prospective clinical trials. Based onSMBGresults and the HbA1c, the dose of either the sulfonylurea or metformin should be increased until the glycemic target is achieved.a-Glucosidase inhibitors and thiazolidinediones are alternative, initial agents (Fig. 333-14).

When used as monotherapy, approximately one-third of individuals will reach their target glycemic goal with either a sulfonylurea or metformin. Approximately 25% of individuals will not respond to sulfonylureas or metformin; under these circumstances, the drug usually should be discontinued. Some individuals respond to one agent but not the other. The remaining individuals treated with either sulfonylureas or metformin alone will exhibit some improvement in glycemic control but will not achieve their glycemic target and should be considered for combination therapy.

COMBINATION THERAPY WITH GLUCOSE-LOWERING AGENTS A number of combinations of therapeutic agents are successful in type 2DM, and the dosing of agents in combination is the same as when the agents are used alone. Because mechanisms of action of the first and second agents are different, the effect on glycemic control is usually additive. Commonly used regimens include: (1) insulin secretagogue with metformin or thiazolidinedione, (2) sulfonylurea with a-glucosidase inhibitor, and (3) insulin with metformin or thiazolidinedione. The combination of metformin and a thiazolidinedione is also effective and complementary. If adequate control is not achieved with two oral agents, bedtime insulin or a third oral agent may be added stepwise. However, long-term experience with any triple combination is lacking, and

experience with two-drug combinations is relatively limited.

Insulin becomes required as type 2 DM enters the phase of relative insulin deficiency (as seen in long-standing DM) and is signaled by inadequate glycemic control on one or two oral glucose-lowering agents. Insulin can be used in combination with any of the oral agents in patients who fail to reach the glycemic target. For example, a single dose of intermediate-acting insulin at bedtime is effective in combination with metformin. As endogenous insulin production falls further, multiple injections of intermediate-acting and short-acting insulin regimens are necessary to control postprandial glucose excursions. These combination regimens are identical to the intermediate- and short-acting combination regimens discussed above for type 1 DM. Since the hyperglycemia of type 2 DM tends to be more "stable," these regimens can be increased in 10% increments every 2 to 3 days using SMBG results. The daily insulin dose required can become quite large (1 to 2 units/kg per day) as endogenous insulin production falls and insulin resistance persists. Individuals who require >1 unit/kg per day of intermediate-acting insulin should be considered for combination therapy with metformin or a thiazolidinedione. The addition of a thiazolidinedione can reduce insulin requirements in some individuals with type 2 DM, while maintaining or even improving glycemic control.

Intensive diabetes management (Table 333-10) is a treatment option in type 2 patients who cannot achieve optimal glycemic control and are capable of implementing such regimens. A recent study from the Veterans Administration found that intensive diabetes management is not associated with a greater degree of side effects (hypoglycemia, weight gain) than standard insulin therapy. The effect of higher insulin levels associated with intensive diabetes management on the prognosis of diseases commonly associated with type 2 DM (cardiovascular disease, hypertension) is still debated. In selected patients with type 2 DM, insulin pumps improve glycemic control and are well tolerated.

Emerging Therapies Whole pancreas transplantation (conventionally performed concomitantly with a renal transplant) may normalize glucose tolerance and is an important therapeutic option in type 1 diabetes, though it requires substantial expertise and is associated with the side effects of immunosuppression. Pancreatic islet transplantation has been plagued by limitations in pancreatic islet isolation and graft survival, but recent advances in specific immunomodulation have greatly improved the results. Islet transplantation is an area of active clinical investigation.

Advances in molecular biology and new insights into normal mechanisms of glucose homeostasis have led to a number of emerging therapies for diabetes and its complications. For example, glucagon-like peptide 1, a potent insulin secretagogue, may be efficacious in type 2 DM. Inhaled insulin and additional insulin analogues are in advanced stages of clinical trials. Aminoguanidine, an inhibitor of the formation of advanced glycosylation end products, and inhibitors of protein kinase C may reduce the complications of DM. Closed-loop pumps that infuse the appropriate amount of insulin in response to changing glucose levels are potentially feasible now that continuous glucose-monitoring technology has been developed.

COMPLICATIONS OF THERAPY FOR DIABETES MELLITUS

As with any therapy, the benefits of efforts directed towards glycemic control must be weighed against the risks of treatment. Side effects of intensive treatment include an increased frequency of serious hypoglycemia, weight gain, increased economic costs, and greater demands on the patient. In the DCCT, quality of life was very similar in the intensive therapy and standard therapy groups. The most serious complication of therapy for DM is hypoglycemia (Chap. 334). Weight gain occurs with most (insulin, insulin secretagogues, thiazolidinediones) but not all (metformin and a-glucosidase inhibitors) therapies that improve glycemic control due to the anabolic effects of insulin and the reduction in glucosuria. In the DCCT, individuals with the greatest weight gain exhibited increases in LDL cholesterol and triglycerides as well as increases in blood pressure (both systolic and diastolic) similar to those seen in individuals with type 2 DM and insulin resistance. These effects could increase the risk of cardiovascular disease in intensively managed patients. As discussed previously, improved glycemic control is sometimes accompanied by a transient worsening of diabetic retinopathy or neuropathy.

ONGOING ASPECTS OF COMPREHENSIVE DIABETES CARE

The morbidity and mortality of DM-related complications can be greatly reduced by timely and consistent surveillance procedures (Table 333-14). These screening procedures are indicated for all individuals with DM, but numerous studies have documented that most individuals with diabetes do not receive comprehensive diabetes care. Screening for dyslipidemia and hypertension should be performed annually. In addition to routine health maintenance, individuals with diabetes should also receive the pneumococcal and tetanus vaccines (at recommended intervals) and the influenza vaccine (annually).

An annual comprehensive eye examination should be performed by a qualified optometrist or ophthalmologist. If abnormalities are detected, further evaluation and treatment require an ophthalmologist skilled in diabetes-related eye disease. Because many individuals with type 2 DM have had asymptomatic diabetes for several years before diagnosis, a consensus panel from the ADA recommends the following ophthalmologic examination schedule: (1) individuals with onset of DM at <29 years should have an initial eye examination within 3 to 5 years of diagnosis, (2) individuals with onset of DM at >30 years should have an initial eye examination at the time of diabetes diagnosis, and (3) women with DM who are contemplating pregnancy should have an eye examination prior to conception and during the first trimester.

An annual foot examination should: (1) assess blood flow, sensation, and nail care; (2) look for the presence of foot deformities such as hammer or claw toes and Charcot foot; and (3) identify sites of potential ulceration. Calluses and nail deformities should be treated by a podiatrist; the patient should be discouraged from self-care of even minor foot problems.

An annual microalbuminuria measurement is advised in individuals with type 1 or type 2 DM and no protein on a routine urinalysis (Fig. 333-10). If the urinalysis detects proteinuria, the amount of protein should be quantified by standard urine protein measurements. If the urinalysis was negative for protein in the past, microalbuminuria should be the annual screening examination. Routine urine protein measurements do

not detect low levels of albumin excretion. Screening should commence 5 years after the onset of type 1 DM and at the time of onset of type 2 DM.

SPECIAL CONSIDERATIONS IN DIABETES MELLITUS

PSYCHOSOCIAL ASPECTS

As with any chronic, debilitating disease, the individual withDM faces a series of challenges that affect all aspects of daily life. The individual with DM must accept that he or she may develop complications related to DM. Even with considerable effort, normoglycemia can be an elusive goal, and solutions to worsening glycemic control may not be easily identifiable. The patient should view him- or herself as an essential member of the diabetes care team and not as someone who is cared for by the diabetes team. Emotional stress may provoke a change in behavior so that individuals no longer adhere to a dietary, exercise, or therapeutic regimen. This can lead to the appearance of either hyper- or hypoglycemia. Depression and eating disorders (in women) are more common in individuals with type 1 or type 2 DM (Chap. 78).

MANAGEMENT IN THE HOSPITALIZED PATIENT

Virtually all medical and surgical subspecialties may be involved in the care of hospitalized patients with diabetes. General anesthesia, surgery, and concurrent illness raise the levels of counterregulatory hormones (cortisol, growth hormone, catecholamines, and glucagon), and infection may lead to transient insulin resistance. These factors increase insulin requirements by increasing glucose production and impairing glucose utilization and thus may worsen glycemic control. On the other hand, the concurrent illness or surgical procedure may prevent the patient withDM from eating normally and may promote hypoglycemia. Glycemic control should be assessed (with HbA1c) and, if feasible, should be optimized prior to surgery. Electrolytes, renal function, and intravascular volume status should be assessed as well. The extremely high prevalence of asymptomatic cardiovascular disease in individuals with DM (especially in type 2 DM) may require preoperative cardiovascular evaluation.

The goals of diabetes management during hospitalization are avoidance of hypoglycemia, optimization of glycemic control, and transition back to the outpatient diabetes treatment regimen. Attention to each stage in this process requires integrating information regarding the plasma glucose, diabetes treatment regimen, and clinical status of the patient. For example, some surgical procedures utilizing local anesthesia or epidural anesthesia may have minimal effects on glycemic control. If the patient is eating soon after the procedure and there is no disruption of the patient's regular meal plans, then glycemic control is usually maintained.

The physician caring for an individual with diabetes in the perioperative period, during times of infection or serious physical illness, or simply when fasting for a diagnostic procedure must monitor the plasma glucose vigilantly, adjust the diabetes treatment regimen, and provide glucose infusion as needed. Several different treatment regimens (intravenous or subcutaneous insulin regimens) can be employed successfully. Individuals with type 1DMrequire continued insulin administration to maintain the levels of circulating insulin necessary to preventDKA. Prolongation of a surgical procedure or

delay in the recovery room is not uncommon and may result in periods of insulin deficiency. Even relatively brief periods without insulin may lead to mild DKA. Individuals with type 1 DM who are undergoing general anesthesia and surgery, or who are seriously ill, should receive continuous insulin, either through an intravenous insulin infusion or by subcutaneous administration of a reduced dose of long-acting insulin. Short-acting insulin alone is insufficient.

Individuals with type 2DM can be managed with either insulin infusion or a reduced dose of subcutaneous insulin. Oral glucose-lowering agents are discontinued at the time a combined insulin/glucose infusion is started. Oral agents such as sulfonylureas, metformin, acarbose, and thiazolidinediones are not useful in regulating the plasma glucose in clinical situations where the insulin requirements and glucose intake are changing rapidly. Moreover, these oral agents may be dangerous if the patient is fasting (e.g., hypoglycemia with sulfonylureas). Metformin should be withheld when radiographic contrast media will be given or if severe congestive heart failure, acidosis, or declining renal function is present.

Insulin infusions can effectively control plasma glucose in the perioperative period and when the patient is unable to take anything by mouth. The absorption of subcutaneous insulin may be variable in such situations because of changes in blood flow. The physician must consider carefully the clinical setting in which an insulin infusion will be utilized, including whether adequate ancillary personnel are available to monitor the plasma glucose frequently and whether they can adjust the insulin infusion rate, either based on an algorithm or in consultation with the physician. The initial rate for an insulin infusion may range from 0.5 to 5 units/h, depending on the degree of insulin resistance and the clinical situation. Based on hourly capillary glucose measurements, the insulin infusion rate is adjusted to maintain the plasma glucose within the desired range [5.6 to 11.1 mmol/L (100 to 200 mg/dL)]. Glucose infusion, initiated at the time the patient begins fasting, should be adjusted to deliver the equivalent of 50 to 150 mL of D$_5$W/h until the patient is reliably taking nutrition orally. The insulin infusion can be temporarily discontinued if hypoglycemia occurs and may be resumed at a lower infusion rate once the plasma glucose exceeds 5.6 mmol/L (100 mg/dL).

Insulin infusion is the preferred method for managing patients with type 1DM in the perioperative period or when serious concurrent illness is present. Individuals with type 2 DM can be managed with an insulin infusion, but subcutaneous insulin in reduced doses can be used effectively as well. If the diagnostic or surgical procedure is brief and performed under local or regional anesthesia, a reduced dose of subcutaneous, long-acting insulin may suffice. This approach facilitates the transition back to the long-acting insulin after the procedure. The dose of long-acting insulin should be reduced by 30 to 40%, and short-acting insulin is either held or, likewise, reduced by 30 to 40%. Glucose should be infused to prevent hypoglycemia.

Total Parenteral Nutrition (See Chap. 76) Total parenteral nutrition (TPN) greatly increases insulin requirements. In addition, individuals not previously known to haveDM may become hyperglycemic during TPN and require insulin treatment. Intravenous insulin infusion is the preferred treatment for hyperglycemia, and rapid titration to the required insulin dose is done most efficiently using a separate insulin infusion. After the total insulin dose has been determined, insulin may be added directly to the TPN

solution. Often, individuals receiving either TPN or enteral nutrition receive their caloric loads continuously and not at "meal times"; consequently, subcutaneous insulin regimens must be adjusted.

GLUCOCORTICOIDS

Glucocorticoids increase insulin resistance, decrease glucose utilization, increase hepatic glucose production, and impair insulin secretion. These changes lead to a worsening of glycemic control in individuals with DM and may precipitate diabetes in other individuals ("steroid-induced diabetes"). The effects of glucocorticoids on glucose homeostasis are dose-related, usually reversible, and most pronounced in the postprandial period. If the fasting plasma glucose is near the normal range, oral diabetes agents (sulfonylureas and acarbose) may be sufficient to reduce hyperglycemia. If the fasting plasma glucose >11.1 mmol/L (200 mg/dL), oral agents are usually not efficacious and insulin therapy is required. Short-acting insulin may be required to supplement long-acting insulin in order to control postprandial glucose excursions.

REPRODUCTIVE ISSUES

Reproductive capacity in either men or women with DM appears to be normal. Menstrual cycles may be associated with alterations in glycemic control in women with DM. Pregnancy is associated with marked insulin resistance; the increased insulin requirements often precipitate DM and lead to the diagnosis of GDM. Glucose, which at high levels is a teratogen to the developing fetus, readily crosses the placenta, but insulin does not. Thus, hyperglycemia or hypoglycemia from the maternal circulation may stimulate insulin secretion in the fetus. The anabolic and growth effects of insulin may result in macrosomia. GDM complicates approximately 4% of pregnancies in the United States. The incidence of GDM is greatly increased in certain ethnic groups, including African Americans and Hispanic Americans, consistent with a similar increased risk of type 2 DM. Current recommendations advise screening for glucose intolerance between weeks 24 and 28 of pregnancy in women with high risk for GDM (³25 years; obesity; family history of DM; member of an ethnic group such as Hispanic American, Native American, Asian American, African American, or Pacific Islander). Therapy for GDM is similar to that for individuals with pregnancy-associated diabetes and involves MNT and insulin, if hyperglycemia persists. Oral glucose-lowering agents have not been approved for use during pregnancy. With current practices, the morbidity and mortality of the mother with GDM and the fetus are no different from those in the nondiabetic population. Individuals who develop GDM are at marked increased risk for developing type 2 DM in the future and should be screened periodically for DM. After delivery, glucose homeostasis should be reassessed in the mother. Most individuals with GDM revert to normal glucose tolerance, but some will continue to have overt diabetes or impairment of glucose tolerance. In addition, children of women with GDM appear to be at risk for obesity and glucose intolerance and have an increased risk of diabetes beginning in the later stages of adolescence.

Pregnancy in individuals with known DM requires meticulous planning and adherence to strict treatment regimens. Intensive diabetes management and normalization of the HbA1c are the standard of care for individuals with existing DM who are planning

pregnancy. The crucial period of glycemic control is extremely early following fertilization. The risk of fetal malformations is increased 4 to 10 times in individuals with uncontrolled DM at the time of conception. The goals are normal plasma glucose during the preconception period and throughout the periods of organ development in the fetus.

LIPODYSTROPHIC DM (See also Chap. 354)

Lipodystrophy, or the loss of subcutaneous fat tissue, may be generalized in certain genetic conditions such as leprechaunism. Generalized lipodystrophy is associated with severe insulin resistance and is often accompanied by acanthosis nigricans and dyslipidemia. Localized lipodystrophy associated with insulin injections has been reduced considerably by the use of human insulin.

Protease Inhibitors and Lipodystrophy Protease inhibitors used in the treatment of HIV disease (Chap. 309) have been associated with a centripetal accumulation of fat (visceral and abdominal area), accumulation of fat in the dorsocervical region, loss of extremity fat, decreased insulin sensitivity (elevations of the fasting insulin level and reduced glucose tolerance on intravenous glucose tolerance testing), and dyslipidemia. Although many aspects of the physical appearance of these individuals resemble Cushing's syndrome, derangements in cortisol secretion have not been found consistently and do not appear to account for this appearance. Although some individuals have IGT, diabetes is not a common feature. The possibility remains that this is related to HIV infection by some undefined mechanism, since some features of the syndrome were observed before the introduction of protease inhibitors. Therapy for HIV-related lipodystrophy is not well established.

(Bibliography omitted in Palm version)

Back to Table of Contents

334. HYPOGLYCEMIA - *Philip E. Cryer*

Hypoglycemia occurs most commonly as a result of treating patients with diabetes mellitus. However, a number of other disorders, including insulinoma, large mesenchymal tumors, end-stage organ failure, alcoholism, endocrine deficiencies, postprandial reactive hypoglycemic conditions, and inherited metabolic disorders, are also associated with hypoglycemia (Table 334-1). Hypoglycemia is sometimes defined as a plasma glucose level <2.5 to 2.8 mmol/L (<45 to 50 mg/dL). However, as discussed below, the glucose thresholds for hypoglycemia-induced symptoms and physiologic responses vary widely, depending on the clinical setting. Therefore, *Whipple's triad* provides an important framework for making the diagnosis of hypoglycemia: (1) symptoms consistent with hypoglycemia, (2) a low plasma glucose concentration, and (3) relief of symptoms after the plasma glucose level is raised. Hypoglycemia can cause significant morbidity and can be lethal, if severe or prolonged; it should be considered in any patient who presents with confusion, altered level of consciousness, or seizures.

SYSTEMIC GLUCOSE BALANCE AND COUNTERREGULATION

Glucose is an obligate metabolic fuel for the brain under physiologic conditions. By contrast, other organs can use fatty acids, in addition to glucose, to generate energy. The brain cannot synthesize glucose and stores only a few minutes' supply as glycogen. It therefore requires a continuous supply of glucose, which is delivered by facilitated diffusion from arterial blood. As the plasma glucose concentration falls below the physiologic range, blood-to-brain glucose transport becomes insufficient for adequate brain energy metabolism and functioning. It is therefore not surprising that redundant physiologic mechanisms prevent or rapidly correct hypoglycemia.

Plasma glucose levels are maintained within a narrow range, usually between 3.3 and 8.3 mmol/L (60 and 150 mg/dL), despite wide variation in food intake and activity level. This delicate balance requires dynamic regulation of glucose influx into the circulation as glucose utilization in various tissues can change rapidly. The diet is normally a major source of glucose. However, between meals or during fasting, serum glucose levels are maintained primarily by the breakdown of glycogen in the liver and by gluconeogenesis (Fig. 334-1). In most people, hepatic glycogen stores are sufficient to maintain plasma glucose levels for 8 to 12 h, but this time period can be shorter if glucose demand is increased by exercise or if glycogen stores are depleted by illness or starvation.

As glycogen stores are depleted, glucose is generated by gluconeogenesis, which occurs primarily in the liver but also in the kidney. Gluconeogenesis requires a coordinated supply of precursors from liver, muscle, and adipose tissue. Muscle provides lactate, pyruvate, alanine, and other amino acids; triglycerides in adipose tissue are broken down into glycerol, which is a precursor for gluconeogenesis, and free fatty acids, which generate acetyl CoA for gluconeogenesis and provide an alternative fuel source to tissues other than brain.

The balance of glucose production and its uptake and utilization in peripheral tissues is exquisitely regulated by a network of hormones, neural pathways, and metabolic signals (Chap. 333). Among the factors that control glucose production and utilization, insulin

plays a dominant and pivotal role. In the fasting state, insulin is suppressed, allowing increased gluconeogenesis in the liver and the kidney and enhancing glucose generation by the breakdown of liver glycogen. Low insulin levels also reduce glucose uptake and utilization in peripheral tissues and allow lipolysis and proteolysis to occur, leading to the release of precursors for gluconeogenesis and providing alternative energy sources. In the fed state, insulin release from the pancreaticb cells reverses these processes. Glycogenolysis and gluconeogenesis are inhibited, thereby reducing hepatic and renal glucose output; peripheral glucose uptake and utilization are enhanced; lipolysis and proteolysis are restrained; and energy storage is promoted by the conversion of substrates into glycogen, triglycerides, and proteins. Other hormones including glucagon, epinephrine, growth hormone, and cortisol play less important roles in the control of glucose flux during normal physiologic circumstances. However, as described below, these hormones are critically important in the response to hypoglycemia.

As glucose levels approach, and ultimately enter, the hypoglycemic range, a characteristic sequence of *counterregulatory hormone responses* occurs. Glucagon is the first and most important of these responses. It promotes glycogenolysis and gluconeogenesis. Epinephrine can also play an important role in the acute response to hypoglycemia, particularly when glucagon is insufficient. It, too, stimulates glycogenolysis and gluconeogenesis as well as limiting glucose utilization by insulin-sensitive tissues. When hypoglycemia is prolonged, growth hormone and cortisol also reduce glucose utilization and support its production.

The glucose thresholds at which various counterregulatory hormone responses occur are quite similar in healthy subjects (Table 334-2). Nonetheless, these thresholds are dynamic and can be influenced by recent metabolic events. A person with poorly controlled diabetes can have symptoms of hypoglycemia at higher-than-normal glucose levels. Recurrent hypoglycemia, as may occur in individuals with diabetes or in the setting of an insulinoma, shifts thresholds for symptoms and counterregulatory responses to lower glucose levels.

CLINICAL MANIFESTATIONS

Symptoms of hypoglycemia can be divided into two categories, neuroglycopenic and neurogenic (or autonomic) responses. Neuroglycopenic symptoms are the direct result of central nervous system neuronal glucose deprivation. They include behavioral changes, confusion, fatigue, seizure, loss of consciousness, and, if hypoglycemia is severe and prolonged, death. Hypoglycemia-induced autonomic responses include adrenergic symptoms such as palpitations, tremor, and anxiety as well as cholinergic symptoms such as sweating, hunger, and paresthesia. Adrenergic symptoms are mediated by norepinephrine released from sympathetic postganglionic neurons and the release of epinephrine from the adrenal medullae. Increased sweating is mediated by cholinergic sympathetic nerve fibers. Patients with diabetes mellitus learn to recognize the characteristic symptoms of hypoglycemia, but these are less familiar to individuals with other causes of hypoglycemia. Symptoms may be less pronounced with repeated hypoglycemic episodes (see below).

Common signs of hypoglycemia include pallor and diaphoresis. Heart rate and the

systolic blood pressure are typically raised, but these findings may not be prominent. The neuroglycopenic manifestations are valuable, albeit nonspecific, signs. Transient focal neurologic deficits occur occasionally.

CAUSES

Hypoglycemia is traditionally classified as *postprandial* or *fasting*. However, in the clinical setting, hypoglycemia most commonly results from the treatment of diabetes. This topic is therefore addressed before considering the other causes of hypoglycemia.

HYPOGLYCEMIA IN DIABETES

Frequency and Impact Were it not for hypoglycemia, diabetes would be rather easy to treat by administering enough insulin (or any effective drug) to lower plasma glucose concentrations to, or below, the normal range. Because of imperfections in all current insulin-replacement regimens, individuals with type 1 diabetes are at ongoing risk for periods of relative hyperinsulinemia with resultant hypoglycemia. Those attempting to achieve near-normal glycemic control may experience several episodes of asymptomatic or symptomatic hypoglycemia each week. Plasma glucose levels may be <2.8 mmol/L (<50 mg/dL) as much as 10% of the time. At least 25% of such patients suffer an episode of severe, temporarily disabling hypoglycemia, often with seizure or coma, in a given year. Although seemingly complete recovery from the latter is the rule, the possibility of persistent cognitive deficits has been raised, but permanent neurologic defects are rare. About 2 to 4% of deaths associated with type 1 diabetes are estimated to result from hypoglycemia. Fear of hypoglycemia also can lead to disabling psychosocial morbidity.

Hypoglycemia is a less frequent problem in type 2 diabetes but occurs nevertheless in those treated with insulin or sulfonylureas. Transient, mild hypoglycemia may be seen with the shorter-acting sulfonylureas and repaglinide, which also acts by enhancing insulin secretion. Patients taking the long-acting sulfonylureas, chlorpropamide and glyburide, occasionally experience episodes of severe hypoglycemia that may last up to 24 to 36 h.

Conventional Risk Factors Insulin excess is the primary determinant of risk from iatrogenic hypoglycemia. Relative or absolute insulin excess occurs when: (1) insulin (or oral agent) doses are excessive, ill timed, or of the wrong type; (2) the influx of exogenous glucose is reduced, as during an overnight fast or following missed meals or snacks; (3) insulin-independent glucose utilization is increased, as during exercise; (4) insulin sensitivity is increased, as occurs with effective intensive therapy, in the middle of the night, late after exercise, or with increased fitness or weight loss; (5) endogenous glucose production is reduced, as following alcohol ingestion; and (6) insulin clearance is reduced, as in renal failure. However, analyses of the Diabetes Control and Complications Trial (DCCT) indicate that these conventional risk factors explain only a minority of episodes of severe iatrogenic hypoglycemia and that other causes are involved in the majority of episodes.

Hypoglycemia-Associated Autonomic Failure It is now clear that inadequate physiologic counterregulatory and behavioral responses greatly compound the problem

of hypoglycemia caused by insulin excess. Hypoglycemia-associated autonomic failure has two main components: (1) reduced counterregulatory hormone responses, which result in impaired glucose generation; and (2) hypoglycemia unawareness, which precludes appropriate behavioral responses, such as eating.

Defective Glucose Counterregulation The counterregulatory hormone response is fundamentally altered in all people with established (e.g., absent C peptide) type 1 diabetes. As the patient becomes totally insulin-deficient over the first few months or years of the disease, circulating insulin levels are no longer tightly coordinated with glucose levels and are a passive function of administered insulin. Thus, insulin levels do not always decline as glucose levels fall; the first defense against hypoglycemia is lost. Over the same time frame, the glucagon response to falling glucose levels diminishes, and the second defense against hypoglycemia is lost. The cause of defective glucagon production by the pancreatic islet a cells is unknown, but it is tightly linked to the loss of insulin production by the bcells. It is a functional abnormality rather than an absolute deficiency of glucagon, as responses to stimuli other than hypoglycemia are intact. The third defense against hypoglycemia is compromised when the epinephrine response to hypoglycemia is reduced. In contrast to the absent glucagon response, epinephrine deficiency is a threshold abnormality; an epinephrine response can still be elicited, but a lower plasma glucose concentration is required. This threshold shift is largely the result of recent antecedent hypoglycemia, although an additional anatomic component may be present as well in patients affected by classic diabetic autonomic neuropathy. The development of a reduced epinephrine response is a critical pathophysiologic event. Prospective studies have shown that patients with combined deficiencies of glucagon and epinephrine suffer severe hypoglycemia at rates 25-fold or greater than individuals with absent glucagon but intact epinephrine responses.

Hypoglycemia Unawareness Hypoglycemia unawareness refers to loss of the warning symptoms of hypoglycemia that normally alert individuals to the presence of hypoglycemia and prompt them to eat to abort the episode. Under these circumstances, the first manifestation of hypoglycemia is neuroglycopenia, and it is often too late for patients to treat themselves. Like defective counterregulation, the presence of hypoglycemia unawareness has been shown in prospective studies to be associated with a high frequency of severe hypoglycemia.

The interplay of factors involved in hypoglycemia-associated autonomic failure in type 1 diabetes, and consequent hypoglycemia unawareness, is summarized inFig. 334-2. Periods of relative or absolute therapeutic insulin excess, in the setting of absent glucagon responses, lead to episodes of iatrogenic hypoglycemia. These episodes, in turn, cause reduced autonomic (including adrenomedullary) responses to falling glucose concentrations. These impaired autonomic responses result in reduced symptoms of impending hypoglycemia (e.g., hypoglycemia unawareness) and, because epinephrine responses are reduced in the setting of absent glucagon responses, impaired physiologic defense against developing hypoglycemia. Thus, a vicious cycle of recurrent hypoglycemia is created and perpetuated. The syndrome of hypoglycemia unawareness and the reduced epinephrine component of defective glucose counterregulation are reversible after as little as 2 weeks of scrupulous avoidance of hypoglycemia. This involves a shift of glycemic thresholds back to higher plasma glucose concentrations.

Hypoglycemia Risk Factor Reduction It is possible to minimize the risk of hypoglycemia by applying the principles of modern therapy -- patient education and empowerment, frequent self-monitoring of blood glucose, flexible insulin (and other drug) regimens, rational glycemic goals, and ongoing professional guidance and support. With respect to the latter, the issue of hypoglycemia needs to be addressed in every patient contact. If hypoglycemia is a recognized problem, first consider each of the conventional risk factors summarized earlier and recommend the appropriate adjustments of medications, diet, and life-style. Nonselective beta blockers may attenuate the recognition of hypoglycemia and they impair glycogenolysis; a relatively selective b_1-antagonist (e.g., metoprolol or atenolol) is preferable when a beta blocker is indicated. One should consider the issue of compromised glucose counterregulation. Although it is possible to test for this abnormality using a low-dose insulin infusion test, this is not practical. A diagnosis of hypoglycemia unawareness can usually be made from the history. It should be remembered that hypoglycemia unawareness implies that previous episodes of hypoglycemia have occurred, whether these are documented or not. If low glucose levels are not apparent from the patient's self-monitoring log, one should suspect hypoglycemia during the night. The presence of clinical hypoglycemia unawareness makes defective glucose counterregulation quite likely. A 2 to 3 week period of conscientious avoidance of hypoglycemia is advisable.

REACTIVE HYPOGLYCEMIA

The postprandial (reactive) hypoglycemias occur only after meals, and hypoglycemia is self-limited. Postprandial hypoglycemia occurs in children with certain rare enzymatic defects in carbohydrate metabolism such as hereditary fructose intolerance and galactosemia (Chap. 350). Reactive hypoglycemia also occurs in some individuals who have undergone gastric surgery that results in the rapid passage of food from the stomach to the small intestine. This type of *alimentary hypoglycemia* causes a rapid postprandial rise in plasma glucose levels and the release of gut incretins, which induce an exuberant insulin response and subsequent hypoglycemia. Administration of ana-glucosidase inhibitor, which delays carbohydrate digestion and thus glucose absorption from the intestine, can be considered for treatment of reactive hypoglycemia, although its efficacy remains to be established in controlled trials.

If postprandial symptoms occur as an idiopathic disorder, caution should be exercised before labeling a person with the diagnosis of hypoglycemia. Indeed, a self-diagnosis of hypoglycemia has often been reinforced by the finding of a "low" venous glucose concentration late after glucose ingestion. An oral glucose tolerance test should not be used in this setting. Plasma glucose falls as low as 2.4 mmol/L (43 mg/dL) after a 100-g glucose load in 5% of normal asymptomatic individuals, making it difficult to identify hypoglycemia based on the results of this test. The diagnosis of postprandial hypoglycemia requires documentation of Whipple's triad after a typical mixed meal. The cause of repetitive postprandial symptoms in certain individuals is unknown, but they may be particularly sensitive to the normal autonomic responses that follow ingestion of a meal.

FASTING HYPOGLYCEMIA

There are many causes of fasting hypoglycemia (Table 334-1). In addition to insulin and

sulfonylureas used in the treatment of diabetes, ethanol use is a relatively common cause of hypoglycemia. Sepsis and renal failure are often complicated by hypoglycemia, but it is less common in other critical illnesses. Endocrine deficiencies, non-b-cell tumors, and endogenous hyperinsulinemia (including that caused by an insulinoma) are rare causes of hypoglycemia. Enzymatic metabolic errors that cause hypoglycemia are also rare but are being recognized more frequently in infants and children (Chaps. 350 and352).

Drugs In contrast to the sulfonylureas and benzoic acid derivatives (e.g., repaglinide), other oral hypoglycemic agents -- biguanides (e.g., metformin),a-glucosidase inhibitors (e.g., acarbose, miglitol), and thiazolidinediones (e.g., troglitazone, rosiglitazone, pioglitazone) -- do not act by stimulating insulin secretion. Therefore, with these agents, insulin levels usually decrease appropriately as plasma glucose levels fall. Nonetheless, these drugs can contribute to hypoglycemia in other ways. Treatment with ana-glucosidase inhibitor alters the management of hypoglycemia; pure glucose should be used rather than ingestion of complex carbohydrates. Thiazolidinediones, as well as metformin, can predispose to hypoglycemia in patients receiving combined treatment with insulin or an insulin secretagogue.

Ethanol blocks gluconeogenesis but not glycogenolysis. Thus, alcohol-induced hypoglycemia typically occurs after a several-day ethanol binge during which the person eats little food, thereby causing glycogen depletion. Hypoglycemia in this setting can be profound, with mortality rates as high as 10%. Blood ethanol levels correlate poorly with plasma glucose concentrations at the time of diagnosis, as hypoglycemia occurs late in the sequence and often precludes further alcohol consumption.

Pentamidine, which is used to treat *Pneumocystis* pneumonia and other parasitic infections, is toxic to the pancreatic b cell. It causes insulin release initially, with hypoglycemia in about 10% of treated patients, and predisposes to the development of diabetes mellitus later. Quinine also stimulates insulin secretion. However, the relative contribution of hyperinsulinemia to the pathogenesis of hypoglycemia in quinine-treated patients who are critically ill with malaria is debated. Salicylates and sulfonamides can cause hypoglycemia but do so rarely. There are reports of hypoglycemia attributed to nonselectiveb-adrenergic antagonists (e.g., propranolol) and a variety of other drugs.

Critical Illness Rapid and extensive hepatic destruction (e.g., severe toxic hepatitis) causes fasting hypoglycemia because the liver is the major site of endogenous glucose production. The mechanism of hypoglycemia reported in patients with cardiac failure is unknown but likely involves hepatic congestion. Although the kidneys are a source of glucose production, it is perhaps too simplistic to attribute hypoglycemia in people with renal failure to this mechanism alone. The clearance of insulin is reduced substantially in renal failure, and reduced mobilization of gluconeogenic precursors has been reported.

Sepsis is sometimes complicated by hypoglycemia, which is multifactorial in origin. There is impaired endogenous glucose production, perhaps the result of hepatic hypoperfusion, and increased glucose utilization, which is induced by cytokines in macrophage-rich tissues such as the liver, spleen, and ileum and in muscle. Nutrition is also often inadequate in the setting of sepsis. Hypoglycemia can be seen with

prolonged starvation, perhaps as a result of the loss of whole-body fat stores and the subsequent depletion of gluconeogenic precursors (e.g., amino acids), which necessitate increased glucose utilization.

Endocrine Deficiencies Neither cortisol nor growth hormone is critical to the prevention of acute hypoglycemia, at least in adults. Nonetheless, hypoglycemia can occur with prolonged fasting in patients with untreated primary adrenocortical failure (Addison's disease) or hypopituitarism. Anorexia and weight loss are typical features of chronic cortisol deficiency and likely result in glycogen depletion with increased reliance on gluconeogenesis. Cortisol deficiency is associated with low levels of gluconeogenic precursors, suggesting that substrate-limited gluconeogenesis, in the setting of glycogen depletion, is the cause of the impaired ability to tolerate fasting in cortisol-deficient individuals. Growth hormone deficiency can cause hypoglycemia in young children. In addition to extended fasting, high rates of glucose utilization (e.g., during exercise, pregnancy) or low rates of glucose production (e.g., following alcohol ingestion) can precipitate hypoglycemia in adults with hypopituitarism. Cortisol and growth hormone secretion should be evaluated in patients with fasting hypoglycemia when the history suggests pituitary or adrenal disease and when other causes of hypoglycemia are not apparent.

As discussed earlier, the combined loss of counterregulatory glucagon and epinephrine responses plays a central role in the pathogenesis of hypoglycemia in diabetes mellitus. However, hypoglycemia is not a feature of the epinephrine-deficient state that results from bilateral adrenalectomy when glucocorticoid replacement is adequate, nor does it occur during pharmacologic adrenergic blockage when other glucoregulatory systems are intact. There are case reports of fasting hypoglycemia attributed to isolated glucagon or epinephrine deficiency, although hyperinsulinemia was not excluded convincingly in the neonatal cases and other counterregulatory defects may have contributed in the adults. Thus, the regular assessment of glucagon and epinephrine secretion is not warranted.

Non-b-Cell Tumors Fasting hypoglycemia, often termed *non-islet cell tumor hypoglycemia*, occurs in some patients with large mesenchymal or other tumors (e.g., hepatoma, adrenocortical tumors, carcinoids). The glucose kinetic patterns resemble those of hyperinsulinism, but insulin secretion is suppressed appropriately during hypoglycemia. In most instances, hypoglycemia is due to overproduction of an incompletely processed form of insulin-like growth factor (IGF)II. Although total IGF-II levels are not consistently elevated, circulating free IGF-II levels are high. Hypoglycemia may result from IGF-II actions through the insulin or IGF-I receptors. Because of negative-feedback suppression of growth hormone secretion, IGF-I levels tend to be low, causing an increased IGF-II to IGF-I ratio.

Endogenous Hyperinsulinism Hypoglycemia due to excessive endogenous insulin secretion can be caused by: (1) a primary pancreatic islet bcell disorder, typically a b cell tumor (insulinoma), sometimes multiple insulinomas, or, especially in infants or young children, a functional b cell disorder without an anatomic correlate; (2) a b cell secretagogue, often a sulfonylurea, and, theoretically, a b cell-stimulating autoantibody; (3) an autoantibody to insulin; or (4) perhaps ectopic insulin secretion. None of these disorders is common. Endogenous hyperinsulinism is more likely in an overtly well

individual without other apparent causes of hypoglycemia such as a relevant drug history, critical illness, endocrine deficiencies, or a non-b-cell tumor. Accidental, surreptitious, or even malicious administration of a sulfonylurea or insulin should also be considered in such individuals.

The fundamental pathophysiologic feature of endogenous hyperinsulinism is the failure of insulin secretion to fall to very low rates during hypoglycemia. This is assessed by measuring insulin, proinsulin, and C peptide, which is derived from the processing of proinsulin. The critical diagnostic findings are a plasma insulin concentration³36 pmol/L (³6 uU/mL) and a plasma C-peptide concentration ³0.2 mmol/L (³0.6 ng/mL) when the plasma glucose concentration is £2.5 mmol/L (£45 mg/dL) in the fasting state with symptoms of hypoglycemia. Insulin and C-peptide levels do not need to be absolutely high (e.g., relative to euglycemic normal values) but only inappropriately high in the setting of fasting hypoglycemia. Plasma proinsulin concentrations are also inappropriately high, particularly in patients with an insulinoma. Sulfonylureas, because they stimulate insulin secretion, result in a pattern of glucose, insulin, and C-peptide levels that is indistinguishable from that produced by a primary b cell disorder. The measurement of sulfonylureas in plasma or urine distinguishes these conditions. Antibodies to insulin produce *autoimmune hypoglycemia* following the transition from the postprandial to the postabsorptive state, as insulin slowly dissociates from the antibodies. Total and free plasma insulin concentrations are inappropriately high. The distinguishing finding is the presence of circulating antibodies to insulin, but the need to measure these routinely is debated, since autoimmune hypoglycemia appears to be rare. Autoantibodies to the insulin receptor are another rare cause of hypoglycemia and usually occur in the context of other autoimmune diseases. A few cases of apparent ectopic secretion of insulin (from a non-b-cell tumor) have been reported.

Insulinoma and Other Primaryb Cell Disorders Insulinomas are rare, but because approximately 90% are benign, they are a treatable cause of potentially fatal hypoglycemia. The yearly incidence is estimated to be 1 in 250,000. About 60% of cases occur in women. The median age at presentation is 50 years in sporadic cases, but it usually presents in the third decade when associated with multiple endocrine neoplasia type 1 ([Chap. 339](#)). Insulinomas arise within the substance of the pancreas in >99% of cases and are usually small (1 to 2 cm). About 5 to 10% of insulinomas are malignant, as evidenced by the presence of metastases.

Insulinomas almost always come to clinical attention because of hypoglycemia rather than mass effects. As noted earlier, unusually low plasma glucose concentrations may be required to produce symptoms and signs of hypoglycemia because recurrent hypoglycemia shifts the glycemic thresholds. Although symptomatic hypoglycemia can be seen after an overnight fast, it often follows exercise. Rarely, symptomatic hypoglycemia occurs following meals, but most such patients have evidence of fasting hypoglycemia as well.

Octreotide scans localize about half of insulinomas. Arteriography has been used extensively in the past, but false-negative and false-positive results occur, and it is generally preferable to use less invasive computed tomography (CT) or magnetic resonance imaging (MRI) scans, which detect 45 to 75% of tumors. Preoperative ultrasound is of value in some patients. Intraoperative ultrasonography has high

sensitivity and may localize tumors not identified by palpation. Surgical resection of a solitary insulinoma is generally curative. Diazoxide, which inhibits insulin secretion, and the somatostatin analogue, octreotide, can be used to treat hypoglycemia in patients with unresectable insulinomas.

Factitious Hypoglycemia Factitious hypoglycemia, caused by malicious or self-administration of insulin or ingestion of a sulfonylurea, shares many clinical and laboratory features with insulinoma. It is most common among health care workers, patients with diabetes or their relatives, and people with a history of other factitious illnesses. When this diagnosis is suspected, it is useful to seek previous medical records, which may reveal admissions for similar episodes as well as relevant laboratory data. In individuals taking exogenous insulin, factitious hypoglycemia can be distinguished from insulinoma by the presence of high insulin levels without a concomitant increase in the C-peptide level, which is suppressed by the exogenous insulin. As noted above, sulfonylureas stimulate endogenous insulin and can therefore be detected only by measuring drug levels in plasma or urine. Factitious or surreptitious hypoglycemia should be considered in every patient requiring a fasting test for hypoglycemia. In addition to laboratory tests, observing the behavior of the patient may help make this diagnosis.

Approach to the Patient

In addition to recognition and documentation of hypoglycemia, and often urgent treatment, diagnosis of the hypoglycemic mechanism is critical for choosing a treatment that prevents, or at least minimizes, recurrent hypoglycemia. A diagnostic algorithm is shown in [Fig. 334-3](#).

Recognition and Documentation Urgent treatment is often necessary in patients with suspected hypoglycemia. Nevertheless, blood should be drawn, whenever possible, before the administration of glucose to allow documentation of the plasma glucose level. Convincing documentation of hypoglycemia requires the fulfillment of Whipple's triad. Thus, *the ideal time to test the plasma glucose is during an episode associated with hypoglycemic symptoms.* A normal plasma glucose concentration measured when the patient is free of symptoms does not exclude hypoglycemia at the time of earlier symptoms. When the cause of hypoglycemia is obscure, additional assays should include glucose, insulin, C peptide, sulfonylurea levels, cortisol, and ethanol.

Hypoglycemia is sometimes detected serendipitously. A distinctly low plasma glucose measurement in a person without a history of corresponding symptoms raises the possibility of a laboratory error caused by ongoing metabolism of glucose by the formed elements of the blood after the sample is drawn. This type of artifactually low glucose level is particularly likely when leukocyte, erythrocyte, or platelet counts are abnormally high, but it can also occur if separation of the plasma or serum from the formed elements is delayed.

Diagnosis of the Hypoglycemic Mechanism In an adult patient with documented hypoglycemia, a plausible hypoglycemic mechanism and further diagnostic evaluation can be guided by the history, physical examination, and available laboratory data ([Fig. 334-3](#)). Relevant historic elements include: drug history, particularly hypoglycemic

agents or alcohol use; relevant critical illness (hepatic, renal, or cardiac failure, sepsis, or inanition); previous gastric surgery associated with postprandial hypoglycemia; features suggestive of cortisol or growth hormone deficiency; inherited enzyme deficiencies associated with hypoglycemia; or features of a non-b-cell tumor. Absent these, one must consider medication error, endogenous hyperinsulinism, or surreptitious sulfonylurea or insulin administration. In the absence of documented spontaneous hypoglycemia, overnight fasting, or food deprivation during observation in the outpatient setting, will sometimes elicit hypoglycemia and allow diagnostic evaluation. If these maneuvers do not reveal hypoglycemia, and there is a high degree of clinical suspicion, an extended fast lasting up to 72 h is often required to make these diagnoses. This procedure should be performed in the hospital with careful supervision and should be terminated if the plasma glucose drops to<2.5 mmol/L (<45 mg/dL) and the patient has symptoms. It is essential to draw blood samples for appropriate tests before administering glucose or allowing the patient to eat.

Urgent Treatment Oral treatment with glucose tablets or glucose-containing fluids, candy, or food is appropriate if the patient is able and willing to take these. A reasonable initial dose is 20 g of glucose. If neuroglycopenia precludes oral feedings, parenteral therapy is necessary. Intravenous glucose (25 g) should be given using a 50% solution followed by a constant infusion of 5 or 10% dextrose. If intravenous therapy is not practical, subcutaneous or intramuscular glucagon can be used, particularly in people with type 1 diabetes mellitus. Because it acts primarily by stimulating glycogenolysis, glucagon is ineffective in glycogen-depleted individuals (e.g., those with alcohol-induced hypoglycemia). These treatments raise plasma glucose concentrations only transiently, and patients should be encouraged to eat as soon as practical in order to replete glycogen stores.

Prevention of Recurrent Hypoglycemia Prevention of recurrent hypoglycemia requires an understanding of the hypoglycemic mechanism. Offending drugs can be discontinued or their doses reduced. It should be remembered that hypoglycemia caused by sulfonylureas may recur after a period of many hours or days. Underlying critical illnesses can often be treated. Cortisol and growth hormone can be replaced, if deficient. Surgical, radiotherapeutic, or chemotherapeutic reduction of a non-b-cell tumor can alleviate hypoglycemia, even if the tumor cannot be cured; glucocorticoid or growth hormone administration may also reduce hypoglycemic episodes in such patients. Surgical resection of an insulinoma is often curative; medical therapy with diazoxide or octreotide can be used if resection is not possible and in patients with a nontumor primary b cell disorder. The treatment of autoimmune hypoglycemia (e.g., with a glucocorticoid) is more problematic, but this disorder is often self-limited. Failing these treatments, frequent feedings and avoidance of fasting may be required. Uncooked cornstarch at bedtime or an overnight infusion of intragastric glucose may be necessary in some patients.

(Bibliography omitted in Palm version)

335. DISORDERS OF THE TESTES - *James E. Griffin, Jean D. Wilson*

The testes produce sperm and the steroid hormones that regulate male sexual function. Both processes are under complex feedback control by the hypothalamic-pituitary system so that the testes have biosynthetic and regulatory features similar to those of the ovary and the adrenal. Testicular hormones are also responsible for the formation of the basic male phenotype during embryogenesis (Chap. 338). Disorders that affect testicular function are common. Infertility occurs in about 5% of men; Klinefelter syndrome (XXY) occurs in 1 in 500 men and often escapes diagnosis; and various disorders cause hypogonadism, a condition that can be treated by hormone replacement. The testes are also a site of malignancies, most of which are highly responsive to radiation and/or chemotherapy (Chap. 96).

PHYSIOLOGY AND REGULATION OF TESTICULAR FUNCTION

The testis consists of two components -- clusters of interstitial or Leydig cells, where androgenic steroids are synthesized, and a system of spermatogenic tubules for the production and transport of sperm. The components are regulated by the pituitary gonadotropins -- luteinizing hormone (LH), which stimulates Leydig cell function, and follicle-stimulating hormone (FSH), which controls Sertoli cell function and spermatogenesis (Fig. 335-1).

GONADOTROPIN REGULATION AND TESTICULAR FUNCTION

Gonadotropin-releasing hormone (GnRH), also called luteinizing hormone-releasing hormone (LHRH), regulates the production of the gonadotropins,LH andFSH(Chap. 328). Because hypothalamic GnRH is secreted in discrete pulses, the plasma concentrations of LH, FSH, and testosterone are not constant, but fluctuate in a pulsatile pattern that mirrors the secretion of GnRH (Fig. 335-2). Pulsatile secretion is most apparent for LH because it has a relatively short plasma half-life by comparison to FSH; pulsatile secretion of testosterone in response to LH is also apparent, although the pulses are dampened because of the need to stimulate steroid synthesis and secretion by the Leydig cell. In contrast to women, in whom the frequency and amplitude of LH pulses vary during the menstrual cycle, the frequency of LH pulses in adult men is relatively constant at about one pulse every 1 to 2 h.

Testosterone secretion is regulated primarily by pituitaryLH.FSH may augment testosterone secretion by stimulating a Sertoli-cell derived factor that enhances testosterone production. Testosterone feeds back to regulate the hypothalamic-pituitary production of LH. It decreases hypothalamic GnRH pulse frequency and diminishes pituitary sensitivity to GnRH, leading to lower LH levels. Although the pituitary can convert testosterone to dihydrotestosterone and to estrogens, testosterone itself is the primary regulator of gonadotropin secretion by the pituitary. Under ordinary circumstances, LH secretion is exquisitely sensitive to the feedback effects of testosterone, with almost complete suppression after the administration of amounts of exogenous androgen that approximate the normal daily secretory rate of testosterone (~20 umol or 6 mg). However, prolonged elevation of plasma LH (as in testicular deficiency) renders the pituitary less sensitive to negative feedback control by androgen.

FSHis regulated byGnRH and also by the gonadal peptides -- inhibins and activins. Inhibins A and B are heterodimeric proteins (composed of a-b$_A$ or a-b$_B$subunits) that selectively suppress FSH without affectingLH; activins are homodimeric proteins (composed of b$_A$ orb$_B$subunits) that selectively stimulate FSH production. Inhibin B, which is the major form of inhibin in the male, is produced by the Sertoli cell and provides feedback control of FSH production. Activins, which are produced in the pituitary as well as the gonad, stimulate FSH production through an autocrine-paracrine mechanism.

The interlocking system in which two pituitary hormones (LH andFSH) regulate testicular function provides a precise dual-control mechanism in which hormonal signals from Leydig cells and the spermatogenic tubules feed back on the hypothalamic-pituitary system to regulate their own function (Fig. 335-1).

THE LEYDIG CELL

Testosterone Synthesis The biochemical pathway by which the 27-carbon sterol cholesterol is converted to androgens and estrogens is depicted inFig. 335-3. Cholesterol, which can be either synthesized de novo in the Leydig cell or derived from plasma lipoproteins, is converted to testosterone as the result of five enzymatic reactions: (1) cholesterol side chain cleavage (CYP11A1); (2) 3b-hydroxysteroid dehydrogenase/isomerase 2 (3b-HSD2); (3) 17a-hydroxylase (CYP17); (4) 17,20-lyase (CYP17); and (5) 17b-hydroxysteroid dehydrogenase 3 (17b-HSD3). Both the 17a-hydroxylase and the 17,20-lyase reactions are catalyzed by a single cytochrome CYP17; post-translational modification (phosphorylation) of the enzyme and the presence of enzyme cofactors confers 17,20-lyase activity, thereby allowing androgen synthesis in the testis and zona reticularis of the adrenal gland. The first four reactions take place in the adrenal as well as the testis. The rate-limiting process in testosterone synthesis is the delivery of cholesterol by the steroid acute regulatory (StAR) protein to the inner mitochondrial membrane where it can undergo side chain cleavage by CYP11A1 to form pregnenolone.LH from the pituitary stimulates the activity of StAR protein and the enzymes in the steroid pathway. Additional steroids including estradiol are synthesized in small amounts in the Leydig cell.

TESTOSTERONE SECRETION AND TRANSPORT

Only about 70 nmol (20 ug) of testosterone is stored in the normal testes, so the total hormone content turns over about 200 times each day to provide the average of 17 to 20 umol (5 to 6 mg) that is secreted into plasma in normal young men (Fig. 335-4). Testosterone is transported in plasma bound to protein, largely to albumin and to a specific transport protein, sex hormone-binding globulin (SHBG), also called testosterone-binding globulin (TeBG). The bound and unbound fractions in plasma are in dynamic equilibrium, only ~1 to 3% being unbound. Because of rapid dissociation from albumin, the fraction of circulating testosterone available for entry into tissues (bioavailable testosterone) approximates the sum of the free and albumin-bound fractions or about half the total plasma level.

Peripheral Metabolism of Androgens Testosterone serves as a circulating precursor (or prohormone) for the formation of two other hormones that mediate many of the

physiologic processes involved in androgen action (Fig. 335-3). Testosterone can be 5a-reduced to dihydrotestosterone, which is responsible for many of the differentiative, growth-promoting, and functional aspects of male sexual differentiation and virilization. Circulating testosterone (and androstenedione) also can be converted to estrogens by aromatase (CYP19) in extraglandular tissues (Fig. 335-4). All estrone production [averaging about 240 nmol (66 ug) per day] can be accounted for by formation from circulating precursors. The mean estradiol production is approximately 170 nmol (45 ug) per day; ~35% is derived from circulating testosterone, 50% is derived from the estrone, and 15% is secreted directly by the testes. When gonadotropin levels are elevated, estradiol secretion by the testes increases. Thus, the physiologic effects of testosterone are the result of the combined actions of testosterone itself plus those of the active androgen and estrogen metabolites of the parent molecule.

The 5a-reduced and estrogenic metabolites can exert local (paracrine) actions in the tissues in which they are formed or enter the circulation and act as hormones at other sites. Circulating dihydrotestosterone is formed principally in the androgen target tissues; estrogen formation takes place in many tissues, the most significant being adipose tissue. The overall rate of extraglandular estrogen formation increases with age and with increased mass of adipose tissue.

Testosterone and its active metabolites are catabolized in the liver and excreted predominantly in the urine, approximately half in the form of urinary 17-ketosteroids (primarily androsterone and etiocholanolone) and half as polar metabolites (diols, triols, and conjugates).

Androgen Action The major functions of androgen are formation of the male phenotype during sexual differentiation, regulation of LH secretion, and induction of sexual maturation at puberty. The cellular process by which androgens perform these functions is schematized in Fig. 335-5. Testosterone enters the cell by passive diffusion and can be converted to dihydrotestosterone by either steroid 5a-reductase 1 or 2; 5a-reductase 2 is responsible for dihydrotestosterone formation in most androgen target tissues. Testosterone or dihydrotestosterone is then bound to the androgen-receptor protein in the nucleus. The hormone-receptor complex binds to specific DNA sequences to regulate the transcription of messenger RNA and, ultimately, the synthesis of cellular proteins. The androgen receptor, which is encoded by a gene on the long arm of the X chromosome, contains 917 amino acids and has a molecular mass of about 110 kDa. A polymorphic region in the amino terminus of the receptor, which contains a variable number of glutamine repeats, appears to modify the activity of the receptor. The androgen receptor is similar in structure to other steroid hormone receptors and has distinct hormone-binding, DNA-binding, and transcriptional regulatory domains (Chap. 327). Estradiol acts by a similar mechanism but has its own distinct estrogen receptors a and b (Chap. 336).

Although testosterone and dihydrotestosterone bind to the same receptor, their physiologic roles differ. The testosterone-receptor complex regulates gonadotropin secretion, spermatogenesis, and the virilization of the wolffian ducts during sexual differentiation (Chap. 338), whereas the dihydrotestosterone-receptor complex is responsible for external virilization during embryogenesis and for most androgen actions during sexual maturation and adult sexual life. The mechanism by which two hormones

can interact with the same receptor but have different physiologic effects is not well understood. However, dihydrotestosterone binds to the receptor much more tightly than does testosterone, and hence its formation serves to amplify the hormonal signal.

THE SEMINIFEROUS TUBULES AND SPERMATOGENESIS

Spermatogenesis is dependent on both pituitaryFSH and androgen production by the adjacent Leydig cells (Fig. 335-1). The function of FSH in gametogenesis has been clarified by rare, naturally occurring mutations in the *FSHb* gene and in the FSH receptor. Females with these mutations are hypogonadal and infertile because ovarian follicles do not mature, whereas males with mutations in the FSH pathway exhibit variable degrees of impaired spermatogenesis. Thus, while FSH is not absolutely required for spermatogenesis, it increases the number and maturation of sperm. The major site of FSH action is the Sertoli cell, which regulates germ cell proliferation and maturation in the seminiferous tubules. Androgen, which reaches very high concentrations locally in the testis, appears to be essential for spermatogenesis, acting through receptors located in the seminiferous tubules. Several cytokines and growth factors are also involved in the regulation of spermatogenesis by paracrine and autocrine mechanisms. The normal adult testes produce >100 million sperm per day.

The Sertoli cell cannot synthesize steroid hormones de novo but can convert testosterone that diffuses from adjacent Leydig cells to estradiol and to dihydrotestosterone. The Sertoli cell also produces inhibin B. Damage to the seminiferous tubules (e.g., by radiation) reduces inhibin B production, causing a selective increase inFSH.

ASSESSMENT OF TESTICULAR FUNCTION

LEYDIG CELL FUNCTION

History and Physical Examination The assessment of Leydig cell function and androgen status should include inquiry about the presence of developmental abnormalities of the urogenital tract; the timing and extent of sexual maturation at puberty; the rate of beard growth; and the current libido, sexual function, strength, and energy. Inadequate Leydig cell function or androgen action during embryogenesis may cause hypospadias, cryptorchidism, or microphallus. If Leydig cell failure occurs before puberty, sexual maturation will not occur, and the individual will develop the features termed *eunuchoidism*, including an infantile amount and distribution of body hair, poor development of skeletal muscles, and delayed closure of the epiphyses, so that the arm span is more than 5 cm greater than the height and the lower body segment (heel to pubic bone) is more than 5 cm longer than the upper body segment (pubic bone to crown). Detection of postpubertal Leydig cell failure requires a high index of suspicion and appropriate laboratory assessment because some functions that require androgens for initiation continue unabated when Leydig cell failure occurs, and functions that eventually regress may do so very slowly. For example, the frequency of shaving may not decrease for months or years because of slow decline in the rate of beard growth once established. Furthermore, decreased sexual function in adult men may be caused by nonendocrine as well as endocrine factors (Chap. 51).

Plasma Testosterone and Dihydrotestosterone Levels Plasma testosterone is measured by immunoassay. Testosterone is secreted into plasma in a pulsatile fashion every 60 to 90 min (Fig. 335-2). A single random testosterone sample provides a result within ±20% of the true mean value only two-thirds of the time; a pool of three samples spaced 15 to 20 min apart provides a more accurate assessment. The plasma testosterone level in normal adult men ranges from 10 to 35 nmol/L (3 to 10 ng/mL). However, in some normal men with long interpulse intervals of LH, testosterone levels can transiently fall below this normal range, emphasizing the importance of using pooled or repeated samples before making a diagnosis of testosterone deficiency. In adult men, plasma testosterone levels also vary somewhat throughout the day and at different times of the year. In young adult men the plasma testosterone level is ~30% higher in the morning than in the evening. Estimation of SHBG concentration by radioimmunoassay is sometimes useful in the interpretation of total plasma testosterone levels. Bioavailable testosterone in plasma can be estimated by measuring the non-SHBG-bound fraction of testosterone.

The plasma testosterone level is slightly higher in prepubertal boys than in girls, ranging in both from 0.2 to 0.7 nmol/L (0.05 to 0.2 ng/mL). The rise in plasma testosterone level at the start of male puberty begins as a result of sleep-related nocturnal gonadotropin surges, so that levels of plasma testosterone and LH are initially higher at night than during the day. Random daytime levels of plasma testosterone increase gradually as puberty progresses and reach adult levels at about age 17.

Dihydrotestosterone is also measured by immunoassay. In young men the plasma dihydrotestosterone level is about one-tenth the value for testosterone, averaging ~2 nmol/L (0.6 ng/mL). In older men with benign prostatic hyperplasia, plasma dihydrotestosterone levels average ~3 nmol/L (0.9 ng/mL).

Urinary 17-Ketosteroids The measurement of urinary 17-ketosteroids is not a valid way to assess testicular function because testosterone contributes only ~40% of urinary 17-ketosteroids in men, the bulk being derived from adrenal androgens.

Plasma LH Plasma LH is also measured by immunoassay. Dual-site immunometric assays have largely replaced radioimmunoassays. Because LH is secreted in a pulsatile fashion, assay of a pool of three samples drawn 15 to 20 min apart, as described above, provides a value approaching the true mean. In early puberty, plasma LH secretion increases only during sleep, but in the adult the pulsatile secretion is of similar magnitude during sleep and waking periods. The normal plasma LH values should be established for a given laboratory with an appropriate reference standard. A low plasma testosterone level can be interpreted correctly only if plasma LH is measured simultaneously, and, likewise, the "appropriateness" of a given plasma LH value must be interpreted in relation to the plasma testosterone level. For example, a low plasma testosterone level coupled with a low LH level implies hypothalamic or pituitary disease, whereas the finding of a low plasma testosterone level and a high LH level suggests primary testicular insufficiency.

Response to Gonadotropin Stimulation Leydig cell function is difficult to assess before puberty, when both LH and testosterone levels are low, but it is possible to measure response of plasma testosterone to gonadotropin stimulation as an index of

Leydig cell capacity. Normal prepubertal boys respond to 3 to 5 days of injection of 1000 to 2000 IU of human chorionic gonadotropin (hCG) with an increase in the plasma testosterone level to ~7 nmol/L (2 ng/mL); the response increases with the initiation of puberty and peaks in early puberty.

Response to GnRH Before puberty, there is minimal response of plasma LH and FSH to the administration of GnRH because the pituitary has not been "primed" by previous exposure to GnRH or gonadal steroids. After pubertal development, the LH response to acute administration of GnRH increases, while the FSH response is less robust. The amount of LH released after acute administration of GnRH probably reflects the amount of stored hormone in the pituitary. When 100 ug GnRH is given subcutaneously or intravenously to normal men, LH levels usually increase four- to fivefold, with the peak level at 30 min. However, the range of response is broad, and some normal men exhibit less than a doubling of LH levels. In primary testicular failure, measurement of basal LH is usually sufficient, and assessment of GnRH response is of little aid in diagnosis. Since men with either pituitary or hypothalamic disease can have a normal or an abnormal LH response to acute administration of GnRH, a normal response does not clearly distinguish these causes of gonadal deficiency. A subnormal response is, however, of value in establishing that an abnormality exists, even though the site of the defect is not clearly determined. If pulsatile GnRH or daily infusions of GnRH for a week lead to the development of a normal acute LH response, a hypothalamic etiology is likely.

SEMINIFEROUS TUBULE FUNCTION

Examination of the Testes Evaluation of the testes is an essential portion of the physical examination. The prepubertal testis measures about 2 cm in length and 2 mL in volume and grows during puberty to reach the adult proportions by age 16. When damage to the seminiferous tubules occurs before puberty, the testes are small and firm, whereas the testes are usually soft after postpubertal damage (the capsule, once enlarged, does not contract to its previous size). Testes in adult men average 4.6 cm in length (range, 3.5 to 5.5 cm), corresponding to a volume of 12 to 25 mL, and the seminiferous tubules account for ~60% of testicular mass. Advanced age does not influence testicular size, so the significance of small testes in the adult is the same at all ages. Asian men have smaller testes than western Europeans, independent of differences in body size. Because of its possible causal role in infertility, the presence of varicocele should be sought by palpation with the patient standing.

Semen Analysis Seminal fluid is analyzed on samples obtained by masturbation into a glass container after 24 to 36 h of abstinence. Analysis should be performed within an hour of collection. The normal ejaculate volume is 2 to 6 mL. Immediately after ejaculation, the seminal fluid coagulates, followed in 15 to 30 min by liquefaction. Motility should be assessed in undiluted seminal fluid; >60% of the sperm should be motile and of normal morphology. The normal range for sperm density is generally considered to be >20 million per milliliter, with a total count of >60 million per ejaculate, but the definition of a minimally adequate ejaculate is not clear. Some men with low sperm counts are nevertheless fertile. This uncertainty as to the lower level of sperm density, percent motility, and percent normal forms in fertile semen stems from two issues. First, the seminal fluid is routinely evaluated by tests that do not assess the

functional capacity of sperm. Second, many factors produce temporary aberrations in sperm count; in men with semen of equivocal quality, it is necessary to examine three or more ejaculates to determine whether the abnormal findings are permanent.

Plasma FSH Plasma FSH, as measured by immunoassay, usually correlates inversely with spermatogenesis. When damage to the germinal epithelium is severe, plasma levels of inhibin B fall and plasma levels of FSH increase.

Testicular Biopsy Testicular biopsy is useful in some patients with oligospermia and azoospermia both as an aid in diagnosis and as an indication of the feasibility of treatment. For example, normal findings on testicular biopsy and a normal FSH level in an azoospermic man suggest obstruction of the vas deferens, which may be correctable surgically. In some men with severe oligospermia, testicular biopsy allows retrieval of sperm for intracytoplasmic sperm injections (ICSI) into oocytes (Chap. 54).

ESTROGENIC FUNCTION

Examination of the Breasts Breast enlargement (gynecomastia) is the most consistent feature of feminizing states in men (Chap. 337). Gynecomastia is due to an increase in both glandular and adipose tissue. The presence of gynecomastia should be sought by examining the sitting patient, using the fingers to grasp glandular tissue. Early or minimal breast enlargement may be missed if the breast is palpated with the flat of the hand while the patient is supine. In obese men it is important to try to define the rim of the glandular tissue where it meets the adipose tissue of the chest wall.

Plasma Estrogen As discussed above, most of the estradiol and all of the estrone produced in normal men is formed by extraglandular aromatization of circulating androgens. The plasma level of estradiol usually is<180 pmol/L (50 pg/mL) in normal men; the plasma estrone level is somewhat higher but usually is<300 pmol/L (80 pg/mL). Elevations in estrogen production and estrogen plasma levels can be due to increases in plasma precursors (liver or adrenal disease), to increased extraglandular aromatization (obesity), or to increased production by the testes (testicular tumors, androgen resistance, gonadotropin stimulation).

PHASES OF NORMAL TESTICULAR FUNCTION

The phases of male sexual life can be defined in terms of the plasma testosterone value (Fig. 335-6). In the male embryo the production of testosterone by the testes commences at about 7 weeks of gestation and is stimulated in part by placental hCG. Shortly thereafter, plasma testosterone attains a high level that is maintained until late in gestation. The level then falls so that at the time of birth the plasma testosterone level is only slightly higher in males than in females. Shortly after birth, a transient increase of pituitary gonadotropins raises the plasma testosterone level in the male infant for ~3 months, before hormone levels again decrease to low levels by age 6 months to 1 year. The significance of the rise in plasma testosterone level during the first year of life is not certain. However, in other primates neonatal activation of the hypothalamic-pituitary-testicular axis is important for subsequent normal pubertal development. The testosterone concentration then remains low (but slightly higher in boys than in girls) until the onset of puberty, when it begins to rise in boys, reaching

adult levels by about age 17. The level of bioavailable testosterone remains constant until the 40s when it begins to decline at a rate of ~1.2% per year; the level of SHBG increases by ~1.2% per year so that there is little decline in total testosterone until the later decades of life. During the third, or adult, phase of male sexual life, sperm production becomes sufficient to allow reproduction to take place. The physiologic events during these various phases differ, as do the pathologic consequences of derangements in testicular function. Male sexual differentiation during embryogenesis is considered in Chap. 338.

ABNORMALITIES OF TESTICULAR FUNCTION

PUBERTY

The control of puberty is poorly understood and may reside in the hypothalamic-pituitary system, the testes, or the adrenals (Chap. 8). Before the onset of puberty, gonadotropin secretion by the pituitary is low, but prepubertal castration causes a rise in plasma gonadotropin levels. This finding suggests that before puberty the negative feedback control of gonadotropin secretion is exquisitely sensitive to the small amount of circulating testosterone. The onset of puberty is heralded by sleep-associated surges in gonadotropin secretion. Later in puberty the rises in LH and FSH levels persist throughout the day. Thus, with maturation, the hypothalamic-pituitary system becomes less sensitive to negative feedback control, and the consequences are higher mean plasma levels of testosterone and gonadotropins, maturation of the testes, and the onset of spermatogenesis. The rise in gonadotropin secretion is the consequence of an increase both in GnRH secretion and in the sensitivity of the pituitary to GnRH.

The somatic changes at the time of puberty are secondary to the rise in plasma testosterone, which induces the growth of the accessory organs of male reproduction (the penis, the prostate, the seminal vesicles, and the epididymides). Accelerated linear growth is accompanied by growth of muscle and connective tissue, which account for the major portion of nitrogen retention at puberty. The principal androgen-sensitive muscles are those of the pectoral region and the shoulder. The characteristic hair growth of male puberty involves development of mustache and beard; regression of the scalp line; appearance of body, extremity, and perianal hair; and extension of the pubic hair upward into a diamond-shaped pattern. Growth of axillary and pubic hair is initiated under the control of adrenal androgens and is promoted by testicular androgens. The larynx enlarges, and the vocal cords thicken, resulting in a lowering of the pitch of the voice. Hemoglobin levels increase by ~1 g/dL. These various androgen-mediated growth and maturation processes reach some limiting value, so that once puberty is completed, the administration of pharmacologic doses of androgen has no further effect. The entire process is heralded by testicular enlargement beginning at age 11 to 12 and is usually completed within 5 years, although some aspects of virilization, such as growth of the chest hair, may continue over a decade or more.

The events of normal male puberty are variable in onset, duration, and sequence. The central issue in dealing with disorders of puberty is separating true absence or precocity from the extremes of normal variation. The use of staging criteria that correlate developmental and anatomic landmarks with chronologic age is useful in making this distinction (Chap. 8).

Sexual Precocity Premature development of sexual characteristics that are phenotypically appropriate -- i.e., virilization in boys -- is termed *isosexual precocity*. *Heterosexual precocity* refers to feminizing syndromes in boys.

Isosexual precocity Sexual development before age 9 in boys is generally considered abnormal. *True precocious puberty* or *complete isosexual precocity* occurs when both virilization and spermatogenesis are premature. *Precocious pseudopuberty* or *incomplete isosexual precocity* refers to virilization unaccompanied by spermatogenesis. This distinction is blurred in practice, because pure virilizing syndromes may cause activation of gonadotropin secretion secondarily and thus be followed by development of spermatogenesis. Furthermore, local androgen production in the testis, as in Leydig cell tumors, can cause local areas of spermatogenesis and limited sperm production around the tumor. We therefore prefer a two-part classification: virilizing syndromes (in which hypothalamic-pituitary activity is appropriate for age) and premature activation of the hypothalamic-pituitary system.

Virilizing syndromes can result from Leydig cell tumors, hCG-secreting tumors, adrenal tumors, congenital adrenal hyperplasia (most commonly 21-hydroxylase deficiency), androgen administration, or Leydig cell hyperplasia. In these disorders plasma testosterone levels are inappropriately elevated for the age. Leydig cell tumors are rare in children but should be suspected when the testes are asymmetric in size (Chap. 96). Virilizing adrenal tumors mainly secrete androstenedione and dehydroepiandrosterone, some of which is converted to testosterone; consequently, they cause increased 17-ketosteroid excretion. Glucocorticoid administration does not reduce 17-ketosteroid excretion to normal in patients with testicular or adrenal tumors, in contrast to the prompt decrease that occurs after such treatment in patients with congenital adrenal hyperplasia. Congenital adrenal hyperplasia leads to elevated 17-hydroxyprogesterone levels and, as a consequence, elevated androgen levels (Chaps. 331 and338). When this disorder is treated with glucocorticoids, true precocious puberty can then result if the increased androgen levels have caused sufficient hypothalamic maturation.

Gonadotropin-independent sexual precocity in boys may occur as a result of autonomous Leydig cell hyperplasia in the absence of a Leydig cell tumor. The disorder can occur sporadically or can be inherited as a male-limited autosomal disorder either from affected fathers or from mothers who are unaffected carriers. It is due to point mutations in the LH receptor that cause constitutive activation of the receptor in the absence of LH. Virilization usually begins by age 2. Testosterone levels are elevated, often to the adult male range; however, immunoreactive and bioactive LH levels and the LH response to GnRH are prepubertal. In the past many of these boys were mistakenly thought to have true precocious puberty because spermatogenesis may be present.

Premature activation of the hypothalamic-pituitary system Central precocious puberty may be "idiopathic" or due to central nervous system (CNS) tumors, infections, or injuries. Early hypothalamic-pituitary activation typically is associated with features of normal puberty, i.e., sleep-related gonadotropin secretion, elevated plasma bioactive LH, and enhanced gonadotropin response to GnRH. Since the diagnosis of idiopathic true precocious puberty is one of exclusion, patients may later prove to have been misclassified and to have a CNS abnormality. With improved means of diagnosis, such

as magnetic resonance imaging, delays in diagnosis will probably be less frequent.

Management of sexual precocity due to steroid- or gonadotropin-producing tumors, congenital adrenal hyperplasia, or CNS abnormality is directed toward the primary disease. In boys with Leydig cell hyperplasia, attempts have been made to lower plasma testosterone with medroxyprogesterone acetate or ketoconazole, or to blunt hormone action with spironolactone, but treatment remains suboptimal. Idiopathic true precocious puberty and true precocious puberty due to inoperable CNS lesions are treated with long-acting GnRH analogue therapy, which inhibits gonadotropin production and testosterone synthesis, reversing pubertal maturation and decreasing the rate of skeletal development.

Heterosexual precocity Feminization in prepubertal boys can result from absolute or relative increases in estrogen due to a variety of causes (Chap. 337).

Delayed or Incomplete Puberty Separating failure of puberty from variants of normal development is one of the most difficult problems in endocrinology. Some boys fail to show the normal spurt of growth and sexual development at the usual time but eventually commence puberty by age 16 or older. Adolescence may then either progress rapidly, or slow pubertal development and growth may continue until age 20 to 22. Many men with delayed onset of puberty attain heights within the normal adult range. The history may reveal that a parent or sibling had a similar pattern of development. Panhypopituitarism and hypothyroidism can cause pubertal failure (Chaps. 328 and 330). Absent puberty also can result from primary testicular disease; this diagnosis is suspected on the basis of low plasma testosterone levels and elevated FSH and LH levels. Hereditary androgen resistance (in which plasma testosterone and LH levels are both high) usually causes male pseudohermaphroditism but in mild form may be manifested by absent or incomplete puberty (Chap. 338).

Most boys with absent puberty have low plasma levels of both testosterone and gonadotropins; in these boys it is necessary to distinguish delayed puberty from isolated gonadotropin deficiency or idiopathic *hypogonadotropic hypogonadism (Kallman syndrome)*. The manifestations of isolated gonadotropin deficiency vary from boys with eunuchoidal features and testes of prepubertal size to those with partial LH and/or FSH deficiency and some degree of testicular enlargement and pubertal development. Ansomia or hyposmia is caused by abnormal development of the olfactory tracts (which share progenitor cells with GnRH neurons) and is characteristically seen in Kallmann syndrome. X-linked *adrenal hypoplasia congenita (AHC)* is characterized by primary adrenal insufficiency, which usually presents in infancy, and hypogonadotropic hypogonadism, caused by deficient GnRH production and abnormal gonadotrope function. Congenital hypogonadotropic hypogonadism is frequently associated with cryptorchidism and a prepubertal manifestation can be microphallus, in which the size of the penis is below the fifth percentile for the age.

The pathogenesis of hypogonadotropic hypogonadism can involve several distinct abnormalities of GnRH formation or action. Some cases are inherited as an X-linked recessive trait associated with defects in a neural cell adhesion molecule (KAL) involved in the migration of GnRH neurons into the olfactory bulb. Other causes involve autosomal dominant disorders with variable expressivity; rare autosomal recessive

cases are due to mutations that impair the GnRH receptor. Serum FSH and LH levels are usually below the normal male range, and plasma testosterone levels are low for age. The secretion of other pituitary hormones is normal. The administration of pulsatile GnRH corrects the endocrine abnormalities and initiates spermatogenesis in patients with GnRH deficiency; all patients respond to gonadotropin replacement. If untreated, these patients usually remain in the prepubertal state indefinitely.

It is particularly difficult to distinguish hypogonatropic hypogonadism from delayed puberty in boys of early or midpubertal age; the presence of microphallus, anosmia, or a family history may suggest the diagnosis. In the absence of such evidence, observation through the teenage years may be required before it becomes clear whether a patient has delayed puberty or a permanent form of hypogonadotropic hypogonadism. Since delayed puberty is associated with a decreased bone mass, therapy should not be delayed too long. In some cases the response of plasma LH to GnRH stimulation may be helpful in suggesting that puberty is imminent.

ADULTHOOD

At the completion of puberty, plasma testosterone levels reach the adult level of 10 to 35 nmol/L (3 to 10 ng/mL) throughout the day, plasma gonadotropins are in the normal adult range, and sperm production is sufficient to allow reproduction. The adult pattern of hypothalamic-pituitary-gonadal regulation is sustained in the normal man for more than 40 years. However, the system is subject to a variety of influences, at the level of both the testes and the hypothalamic-pituitary system. Spermatogenesis is exquisitely sensitive to alterations in scrotal temperature, and brief increases in either systemic or local temperature (as in a hot bath) can be followed by temporary decreases in sperm production. The system also is influenced by diet, drugs, alcohol, environmental agents, and psychological stress, any of which may cause temporary decreases in sperm count.

Persistent abnormalities of testicular function in adult men can be due to hypothalamic-pituitary disorders (Chap. 328), testicular defects, or abnormalities of sperm transport. Certain of these conditions tend to affect Leydig cell function or spermatogenesis selectively, but most impair both androgenization and fertility (Table 335-1). The interlocking of Leydig cell function and fertility is due to the dependence of spermatogenesis on androgen. Even a partial decrease in testosterone production can cause infertility. Certain conditions (hyperprolactinemia, radiation therapy, cyclophosphamide therapy, autoimmunity, paraplegia, androgen resistance) can cause either isolated infertility or a combined defect in testicular function.

Hypothalamic-Pituitary Disorders Disorders of the hypothalamus and pituitary can impair the secretion of gonadotropins either as one manifestation of a generalized disease of the anterior pituitary (Chap. 328) or as an isolated defect. In the latter case the cause is usually hypogonadotropic hypogonadism, in which secretion of both LH and FSH are impaired. This disorder usually is congenital but may be acquired. Alternatively, gonadotropin secretion can be altered by factors other than hypothalamic-pituitary pathology. For example, elevation of plasma cortisol, as in the *Cushing syndrome,* can depress LH secretion independent of a space-occupying lesion of the pituitary. Critical illness also suppresses plasma gonadotropin levels. Some patients with uncontrolled *congenital adrenal hyperplasia* have elevated levels of

adrenal androgens, suppressed gonadotropin secretion, and consequent infertility. Likewise, the use of *androgens* for purposes other than replacement therapy can inhibit gonadotropin secretion and impair sperm production (see below). *Hyperprolactinemia* (as the consequence either of pituitary adenomas or of drugs such as phenothiazines) can cause combined Leydig cell and seminiferous tubule dysfunction, presumably due to inhibition of LH and FSH secretion by prolactin. Occasionally, impaired fertility in hyperprolactinemia is associated with normal gonadotropin and androgen levels and is presumed to result from direct inhibition of spermatogenesis by prolactin. *Hemochromatosis* usually impairs testicular function as the result of effects on the pituitary; less often it affects the testis directly (Chap. 345). In some conditions, testosterone levels may be decreased in association with normal LH levels, and the mechanism is less clear. Men with massive obesity have decreased levels of SHBG and of total and bioavailable testosterone, which return toward normal with weight loss. Obesity may also contribute to the decreased testosterone levels in the subset of such men with Pickwickian syndrome (Chap. 263). Some men with temporal lobe seizures also have hypogonadotropic hypogonadism.

Testicular Defects Abnormalities of testicular function in the adult man can be grouped into several categories: developmental and structural defects of the testes, acquired testicular defects, and disorders secondary to systemic disease.

Developmental abnormalities The *Klinefelter syndrome* (*XXY*, both the classic and the mosaic forms) and the *XX male syndrome* are usually not recognized until after the time of expected puberty (Chap. 338). Some developmental defects cause infertility in the presence of normal androgen production. These include varicocele, germinal cell aplasia, deletions or mutations of the azoospermia factor (*AZF*) genes on the Y chromosome, and cryptorchidism. *Varicocele* may be of etiologic importance in as much as one-third of all cases of male infertility. It is caused by retrograde flow of blood into the internal spermatic vein that eventuates in progressive, often palpable dilation of the peritesticular pampiniform plexus of veins. Varicocele occurs in ~10 to 15% of men in the general population and in 20 to 40% of men with infertility. It is thought to result from incompetence of the valve between the internal spermatic vein and the renal vein and is more common on the left side (85%). Unilateral varicocele increases the blood flow and the temperature of both testes as a result of the extensive anastomoses of the venous systems. The increased scrotal (and testicular) temperature is believed to be the cause of the poor-quality semen and infertility (the testes no longer are 2°C cooler than the abdominal cavity). The findings on semen analysis are usually nonspecific, with all parameters showing some abnormality. Surgical repair of varicocele results in fertility in about half of men, with the best results (70% pregnancy rate) in those whose preoperative sperm counts are >10 million per milliliter.

Some patients with *germinal cell aplasia* (the Sertoli cell-only syndrome) have a positive family history and may constitute a specific group in whom the germinal epithelium is missing with resulting azoospermia; plasma testosterone and LH values are normal, and plasma FSH levels are elevated. Other patients with identical histologic and clinical findings have androgen resistance or a history of viral orchitis or cryptorchidism; microdeletions of one or more genes (e.g., *Deleted in Azoospermia, DAZ*) on the Y chromosome have been documented in 10 to 20% of men with azoospermia or oligospermia (many of whom have germinal cell aplasia), depending on the criteria used

for selection.

Unilateral *cryptorchidism*, even when corrected before puberty, is associated with abnormal semen in many individuals, indicating that the testes can be bilaterally abnormal even in unilateral cryptorchidism.

The *immotile cilia syndrome* is an autosomal recessive defect characterized by immotility or poor motility of the cilia of the airways and of the sperm. Kartagener's syndrome is a subgroup of the immotile cilia syndrome associated with situs inversus, chronic sinusitis, and bronchiectasis (Chap. 256). The structural abnormality leading to impaired motility of cilia can usually be defined by the electron-microscopic appearance showing defects in the dynein arms, spokes, or microtubule doublets. Cilia from epithelia and sperm tails exhibit the same defects, but the pulmonary manifestations may be minor. Other less well understood structural defects can cause immotility of sperm without involvement of cilia in the lung.

Acquired testicular defects Acquired testicular failure in the adult man can be due to *viral orchitis*. The responsible viruses include mumps virus, echovirus, lymphocytic choriomeningitis virus, and group B arboviruses. The orchitis is due to actual infection of the tissue by virus rather than to indirect effects of infection. Orchitis occurs in as many as one-fourth of adult men with mumps; in about two-thirds the orchitis is unilateral, and in the remainder it is bilateral. Orchitis usually develops a few days after the onset of parotitis but may precede it. The testis may return to normal size and function or undergo atrophy. Atrophy is believed to be due both to direct effects of the virus on the seminiferous tubules and to ischemia secondary to pressure and edema within the taut tunica albuginea. Semen analysis returns to normal in three-fourths of men with unilateral involvement and in only one-third of men with bilateral orchitis. Atrophy is usually perceptible within 1 to 6 months after the acute illness, and the degree of atrophy is not necessarily proportional to the severity of the acute orchitis. Unilateral atrophy occurs in about one-third of patients, and bilateral atrophy occurs in about one-tenth.

Trauma, including torsion, can also cause secondary atrophy of the testes. The exposed position of the testes in the scrotum renders them susceptible to both thermal and physical trauma -- particularly in men with hazardous occupations.

The testes are sensitive to *radiation damage*; decreased secretion of testosterone appears to be a consequence of diminished testicular blood flow. Doses>200 mGy (20 rad) cause increases in plasmaFSH andLHlevels and damage to the spermatogonia. After about 800 mGy (80 rad), oligospermia or azoospermia develops, and higher doses may obliterate the germinal epithelium, except for occasional stem and Sertoli cells. Fractionated radiation may have a more profound effect than single-dose radiation. Recovery of sperm density occurs in a dose-related fashion, and complete recovery of sperm density may require as long as 5 years. Permanent infertility can occur after radiation therapy for malignant lymphoma despite shielding of the testes. Permanent androgen deficiency in adult men is uncommon after therapeutic radiation; however, most boys given direct testicular radiation therapy for acute lymphoblastic leukemia have permanently low plasma testosterone levels. Sperm banking should be considered in patients before they undergo radiation treatment or chemotherapy.

In general, *drugs* interfere with testicular function in one of four ways -- inhibition of testosterone synthesis, blockade of androgen action, enhancement of estrogen levels, or direct inhibition of spermatogenesis. Spironolactone and ketoconazole block the synthesis of androgen by interfering with the late steps in androgen biosynthesis. Spironolactone and cimetidine compete with androgen for binding to the androgen receptor and thus block androgen action in target cells. Testosterone levels may be low, and estradiol levels may be elevated in persons using marijuana, heroin, or methadone, although the exact reasons are unclear. Alcohol, when consumed in excess for prolonged periods, causes decreased plasma testosterone levels, independent of liver disease or malnutrition. Elevated plasma estradiol and decreased plasma testosterone levels may occur in men taking digitalis.

Antineoplastic and chemotherapeutic agents commonly interfere with spermatogenesis. Cyclophosphamide causes azoospermia or extreme oligospermia within a few weeks after the initiation of therapy. Cessation of therapy is followed by a return of spermatogenesis within 3 years in about half of patients. Combination chemotherapy for acute leukemia, Hodgkin's disease, and other malignancies also may impair Leydig cell function. In pubertal boys this impairment is manifested by decreased serum testosterone and elevated LH levels; in adult men testosterone levels do not decline, and the impaired Leydig cell function may be detected only as an enhanced LH response to GnRH. The alkylating agents in the chemotherapeutic regimens seem to be responsible for Leydig cell toxicity.

Because of the toxic effects of many physical and chemical agents on spermatogenesis, the occupational and recreational history should be carefully evaluated in all men with infertility. Known environmental hazards include microwaves, ultrasound, and chemicals such as the nematocide dibromochloropropane, cadmium, and lead. In some populations, sperm density is said to have declined by as much as 40% in the past 50 years, and it has been postulated that environmental estrogens or antiandrogens may be responsible.

Testicular failure also occurs as a part of *polyglandular autoimmune insufficiency* (Chap. 339). Sperm antibodies can cause isolated male infertility. In some instances these antibodies are secondary phenomena resulting from duct obstruction or vasectomy. *Granulomatous diseases* can destroy the testes, and testicular atrophy occurs in 10 to 20% of men with lepromatous leprosy owing to direct invasion of the tissue by the mycobacteria. The tubules are involved initially, followed by endarteritis and destruction of Leydig cells.

Testicular abnormalities associated with systemic disease In *cirrhosis of the liver*, a combined testicular and pituitary abnormality leads to decreased testosterone production independent of the direct toxic effects of ethanol. Although the plasma LH level is elevated, the level may be below the expected range given the degree of androgen deficiency. This situation most likely results from the inhibition of LH secretion by estrogen in patients with chronic liver disease. Increased estrogen production results from impaired hepatic extraction of adrenal androstenedione and subsequent increased extraglandular conversion to estrone and estradiol. In effect, estrogen precursors are shunted to sites of extraglandular aromatization. Testicular atrophy and gynecomastia

are present in about half of men with cirrhosis, and many such men are impotent. Successful liver transplantation reverses the effects of cirrhosis on the pituitary-testicular axis.

In chronic *renal failure*, androgen synthesis and sperm production decrease despite elevated plasma gonadotropins. The elevated LH level is due to increased production and reduced clearance but does not restore normal testosterone production. In addition, about one-fourth of men with renal failure have hyperprolactinemia; the role of hyperprolactinemia in decreasing testosterone production is unclear. Low testosterone coupled with normal or increased plasma estrogen levels cause gynecomastia in about half of men on chronic hemodialysis, and about half of men on dialysis have decreased libido and/or impotence. Improvement in testosterone production with hemodialysis is incomplete, but successful transplantation may return testicular function to normal.

Men with *sickle cell anemia* usually have impaired secondary sexual development, and testicular atrophy is present in one-third of them. The defect may be at either the testicular or the hypothalamic-pituitary level. Abnormalities in Leydig cell function, frequently accompanied by decreased sperm density, have been noted in a variety of chronic systemic diseases, including protein-energy *malnutrition*, advanced *Hodgkin's disease* and *cancer* before chemotherapy, and *amyloidosis*. Most of these disorders cause a lowered plasma testosterone level coupled with a normal to increased plasma LH level, suggesting combined hypothalamic-pituitary and testicular defects. Similar hormone changes occur after *surgery*, *myocardial infarction*, and severe *burns* and thus may be a nonspecific effect of illness.

In HIV-infected men, elevation of gonadotropins (a compensated state of hypogonadism) may precede the development of overt hypogonadism, but 35 to 50% of men with AIDS eventually develop low testosterone levels. Elevation of SHBG levels may partially mask the fall in testosterone. Some of the hormonal changes in this disorder are likely nonspecific and related to severe illness. Whether testosterone deficiency contributes to the muscle wasting and weight loss characteristic of this disorder is unclear, but androgen replacement therapy may increase muscle and lean body mass.

Sperm density can decrease temporarily after *acute febrile illness* in the absence of a change in testosterone production. Infertility in men with *celiac disease* is associated with a hormonal pattern typical of androgen resistance, namely, elevated testosterone and LH levels. *Neurologic* diseases associated with altered testicular function include myotonic dystrophy, spinobulbar muscular atrophy, and paraplegia. In myotonic dystrophy, small testes may be associated with impairment of both spermatogenesis and Leydig cell function. Spinobulbar muscular atrophy is caused by an expansion of the glutamine repeat sequences in the amino-terminal region of the androgen receptor; this expansion impairs function of the androgen receptor, but it is unclear how the alteration is related to the neurologic manifestations. Men with spinobulbar muscular atrophy often have as a late manifestation underandrogenization and infertility and the hormonal features of androgen resistance (Chap. 338). *Spinal cord lesions* that cause paraplegia lead to a temporary decrease in testosterone levels and may cause persistent defects in spermatogenesis; some patients retain the capacity for penile erection and ejaculation.

Androgen resistance Defects of the androgen receptor cause resistance to the action of androgen, usually associated with defective male phenotypic development, infertility, and underandrogenization ([Chap. 338](#)). Mutations of the androgen receptor that cause mild androgen resistance can cause infertility due to oligo- or azoospermia in otherwise phenotypically normal men.

Impairment of Sperm Transport Disorders of sperm transport may cause infertility in as many as 6% of infertile men with normal virilization. Obstruction of the ejaculatory system may be unilateral or bilateral, congenital or acquired. In men with unilateral obstruction, infertility may result from antisperm antibodies. Congenital defects of the vas deferens can occur as an isolated abnormality associated with absence of the seminal vesicles (and consequently absence of fructose in the ejaculate), in men whose mothers received *diethylstilbestrol* during pregnancy, and in men with *cystic fibrosis*. Furthermore, congenital bilateral absence of the vas deferens can be due to mutations in the cystic fibrosis conductance regulator (*CFTR*) gene; some of these mutations are distinct from those associated with the more typical pulmonary and gastrointestinal manifestations of cystic fibrosis. Acquired obstructive azoospermia can occur at the level of the epididymis in association with chronic infections of the paranasal sinuses and lungs and with tuberculosis, leprosy, and gonorrhea.

Empirical Therapy of Male Infertility Disorders for which there are logical or effective treatments (genital tract obstruction, sperm autoimmunity, gonadotropin deficiency) account for only 10% of infertile men, and pregnancies are infrequent when the male partner has genital tract obstruction or sperm autoimmunity. Severe oligospermia/azoospermia from other causes accounts for about one-fourth of cases of male infertility and has largely been considered untreatable. The other two-thirds of infertile men have a partial reduction in semen parameters and subfertility of a variable degree, and in this group spontaneous fertility may occur in untreated men (as high as 25% in one year). In the past various empirical therapies (e.g., testosterone rebound, gonadotropins, antiestrogens) have been tried without success. The only successful empirical therapy for men with mild to moderate defects in semen quality is in vitro fertilization. However, standard in vitro fertilization does not provide a good outcome in the presence of severe semen abnormalities, such as a sperm density of<5 million per milliliter, poor motility, and many abnormal forms. For such men the technique of intracytoplasmic sperm injection (ICSI) has been a major advance; indeed fertilization and pregnancy rates with this technique are similar to those for standard in vitro techniques in couples with fallopian tube pathology, e.g., a 50 to 70% fertilization rate and a 30% pregnancy rate per cycle. This technique is sometimes successful with spermatozoa recovered from testicular biopsies in men with azoospermia. *The management of male infertility is discussed in [Chap. 54](#).*

Fertility Control in Men (See also [Chap. 54](#)) A variety of approaches to fertility control in men have been tried, including use of the condom as an effective barrier that also prevents sexually transmitted diseases. Vasectomy, which involves transection or ligation of the vas deferens, has a high success rate and can be performed on an outpatient basis. The time required for azoospermia to occur after the operation depends on the number of sperm in the terminal vas deferens and ejaculatory ducts at the time of surgery, but it is usually less than 40 days. Azoospermia should be documented in each case to prove effectiveness. No deleterious effects on either

testosterone production or the hypothalamic-pituitary axis have been documented. Despite reports of immune-complex-associated accelerated atherosclerosis in vasectomized nonhuman primates, there does not appear to be any association between vasectomy and atherosclerosis in men. Vasectomy should be recommended only for men requesting permanent sterilization. Only about 30 to 40% of men subjected to vasovasostomy for reanastomosis of the vas subsequently achieve fertility. Suppression of gonadotropins with long-acting GnRH analogues or GnRH antagonists causes marked reduction in sperm counts but requires concomitant androgen replacement. The efficacy and acceptance of this approach to male contraception remain to be established.

GONADAL FUNCTION DURING AGING

Beginning at about age 40, mean plasma bioavailable testosterone concentrations decline gradually; about 40% of elderly men have low bioavailable testosterone levels. Although statistically lower than the levels in young men, the concentrations of total testosterone usually remain within the normal range, even in elderly men. The cause of the reduced testosterone level is likely a decreased number of Leydig cells. In older men seminiferous tubule function and sperm production also usually decline. Plasma LH and FSH levels are often slightly elevated, consistent with a decline in gonadal function. An increase in the conversion of androgen to estrogen in peripheral tissues results in a decrease in the effective ratio of androgen to estrogen. These latter hormonal changes may play a role in the development of prostatic hyperplasia and in the development of gynecomastia in aging men (Chaps. 95 and 337). Male sexual function gradually declines after early adulthood, but there is no convincing evidence that hormonal changes have any direct bearing on changes in sexual function with age in healthy men.

Prostatic Hyperplasia See Chap. 95.

Cancer of the Prostate See Chap. 95.

DISORDERS OF ALL AGES

Testicular Tumors (See also Chap. 96) Low levels of hCG are present in normal testes may be elevated in persons with testicular tumors. Indeed, an elevated plasma level of the b subunit of hCG (hCG-b) is a sensitive and specific marker of tumor activity in some men with germ cell tumors. Plasma levels of hCG-b are elevated in all men with choriocarcinoma, in one-third of those with embryonal carcinomas and teratocarcinomas, and rarely in those with seminomas. Changes in hCG-b levels correlate with response to therapy.

Testicular tumors can cause elevated estradiol and testosterone levels by at least two mechanisms: (1) Trophoblastic, Leydig, and Sertoli cell tumors produce both hormones autonomously; pituitary gonadotropin secretion and hormone production by the uninvolved portions of the testes are depressed, and azoospermia is common; (2) hCG secretion by the tumors can increase estradiol and testosterone production in the unaffected areas of the testes; azoospermia is uncommon with such tumors. When estrogens and androgens are formed (directly or indirectly) by the tumors, feminization, virilization, or no obvious change may result, depending on the hormones produced and

the age of the patient.a-Fetroprotein can provde another cellular marker of testicular tumor activity.

Gynecomastia See [Chap. 337](#).

TREATMENT

Androgens

Pharmacologic Preparations When testosterone is taken by mouth, it is absorbed into the portal blood and degraded promptly by the liver, so that only insignificant amounts reach the systemic circulation; when administered parenterally, testosterone is rapidly absorbed from the injection vehicle and rapidly degraded. As a consequence, effective androgen therapy requires the administration of either a slowly absorbed form of testosterone (dermal patches or micronized oral testosterone) or modified analogues. Chemical modifications either retard absorption or catabolism, or enhance the androgenic potency, so that full effects can be achieved at a lower blood level of drug. Three types of modification have had widespread clinical application ([Fig. 335-7](#)): (1) esterification of the 17b-hydroxyl group, (2) alkylation at the 17a position, and (3) alteration of the ring structure, particularly by substitutions at the 2, 9, and 11 positions. Most pharmacologic agents actually have combinations of ring structure alterations and either 17a-alkylation or esterification of the 17b-hydroxyl group. Esterification decreases the polarity of the molecule so that the steroid is more soluble in the fat vehicles used for injection, leading to slower release into the circulation. Most esters must be injected parenterally. The larger the acid esterified, the slower the release and the more prolonged the action. Esters such as testosterone cypionate and testosterone enanthate can be injected every 1 to 3 weeks, the usual regimen being 200 mg of either ester intramuscularly every 2 weeks. Because the esters are hydrolyzed before the hormones act, therapy can be monitored by assaying the plasma testosterone level at various times after administration.

The oral effectiveness of 17a-alkylated androgens (such as methyltestosterone and methandrostenolone) is due to slower hepatic catabolism, which allows the alkylated derivatives to reach the systemic circulation. For this reason, 17a-methyl or -ethyl substitution is a feature of most orally active androgens. Unfortunately, all 17a-alkylated steroids can cause abnormal liver function, and for this reason they have a limited role in therapy.

Other alterations of the ring structure have been adopted empirically; some slow the rate of inactivation, others enhance the potency of a given molecule, and some alter the conversion to other active metabolites. For example, the potency of fluoxymesterone may be due to the fact that, unlike most androgens, it is a poor precursor for conversion to estrogens in peripheral tissues.

Three transdermal preparations are available in which a testosterone-loaded patch is applied to the skin each day. One is a scrotal patch (Testoderm) that contains no permeation enhancers, which may irritate the skin, but it has a low rate of acceptance. The other two systems (Androderm, Testosterone TTS) are applied to the trunk, arms, or thighs. Each patch provides physiologic testosterone levels that mimic the normal

diurnal variation with higher levels in the morning hours. High rates of dermatologic problems have been reported with the Androderm transdermal system.

Side Effects of Androgens All androgens carry the risk of inducing virilization in women. Early manifestations include acne, coarsening of the voice, hirsutism, and menstrual irregularities. If treatment is discontinued as soon as these effects develop, the manifestations may slowly subside. Long-term side effects such as male-pattern baldness, marked hirsutism, voice changes, and hypertrophy of the clitoris are largely irreversible. At physiologic replacement doses, testosterone esters have few toxic effects in mature men. At supraphysiologic doses, however, gonadotropin secretion is inhibited, the testes shrink, and the sperm count falls (indeed, androgen abuse can be associated with low sperm counts that may persist for 9 months or longer after cessation of the steroid). In some older men, testosterone therapy may cause polycythemia (hematocrit>52%); in men predisposed to obstructive sleep apnea, androgen therapy may initiate or worsen symptoms. In older men, the presence of benign prostatic hyperplasia is not a contraindication for androgen therapy, but such men should be screened for prostate cancer before initiating androgen replacement (Chap. 95).

The so-called toxic side effects vary among the different agents and with the clinical setting in which they are used. Retention of a limited amount of sodium is an inevitable consequence of androgen therapy and may lead to edema in patients with underlying heart disease or renal failure, or when androgens are administered in enormous amounts. Although androgens do not cause malignancy, they may promote the growth of and intensify pain from carcinomas of the prostate and breast in men.

The feminizing side effects of androgen therapy in men are poorly understood. Testosterone (but not 5a-reduced androgens) can be converted (aromatized) in extraglandular tissues to estradiol. The most common manifestation of feminization is the development of gynecomastia. Such breast enlargement is common in children given androgens, possibly because of a greater capacity to convert androgens to estrogens in childhood. The administration of testosterone esters to men results in an increase in plasma estrogen levels, but in men with normal liver function gynecomastia usually develops only after use of high doses.

All 17a-alkylated androgens can produce liver function abnormalities such as elevated plasma levels of alkaline phosphatase and conjugated bilirubin. The incidence of clinical liver disease probably depends on the previous integrity of the liver, but jaundice may occur in the absence of preexisting liver disease. 17a-Alkylated drugs also increase the levels of a variety of plasma proteins that are synthesized in the liver. The most serious complications of 17a-alkylated androgens are peliosis hepatis (blood-filled cysts in the liver) and hepatoma. These disorders were initially described in patients with aplastic anemia, many of whom had Fanconi anemia, itself a predisposing factor for the development of malignancy. However, both lesions can occur after administration of substituted androgens for other indications, including use by athletes. These tumors may either follow a benign course after discontinuation of the drugs or be rapidly fatal.

One indication for 17a-alkylated androgens is in the treatment of hereditary angioedema in which the desired therapeutic benefit (increased level of the inhibitor of the first

component of complement) may actually be an effect of the 17-alkylated side chain rather than of the parent androgen. As a consequence, weak androgens such as danazol are effective in this disorder (Fig. 335-7). Danazol is also used in the management of endometriosis (Chap. 52).

Replacement Therapy The aim of androgen therapy in hypogonadal men is to restore or bring to normal male secondary sexual characteristics (beard, body hair, external genitalia) and male sexual behavior and to mimic the hormonal effects on somatic development (hemoglobin, muscle mass, nitrogen balance, and epiphyseal closure). Since an assay for plasma testosterone is available for monitoring therapy, the treatment of androgen deficiency is almost universally successful. The parenteral administration of a long-acting testosterone ester such as 100 to 200 mg testosterone enanthate at 1- to 2-week intervals results in a sustained increase in plasma testosterone to the normal male range. Alternatively, testosterone may be administered transdermally. Testosterone patches, which are available in different doses, are replaced daily. If hypogonadism is primary and of long duration (as in the Klinefelter syndrome), suppression of plasmaLH to the normal range may not occur for many weeks, if at all. There is considerable variability in the relation between plasma testosterone and male sexual behavior, but in cases of postpubertal testicular failure (even of many years duration), normal sexual activity usually is resumed after adequate replacement. Androgen administration does not restore spermatogenesis in hypogonadal states, but the volume of the ejaculate (derived largely from the prostate and seminal vesicles) and other male secondary sex characteristics return to normal. The effects of endogenous androgen on hemoglobin, nitrogen retention, and skeletal development are also reproduced.

In men of all ages in whom hypogonadism developed before expected puberty (such as men with isolated gonadotropin deficiency), it is appropriate to bring plasma testosterone into the adult range slowly. When therapy is commenced at the time of expected puberty in such men, the normal events of puberty proceed in the usual fashion. If therapy is delayed until after the time of usual puberty, the degree to which normal virilization will occur is variable, but many patients undergo a relatively complete anatomic and functional maturation. Intermittent low-dose androgen therapy is indicated in prepubertal hypogonadal boys with microphallus to bring the external genitalia into the normal range. If such patients are monitored closely and given androgens for only short periods, therapy usually has no adverse effects on somatic growth.

In boys of pubertal age with either isolated gonadotropin deficiency or primary testicular disease, the usual practice is to institute androgen therapy between the ages of 12 and 14 years, depending on the subjective need for sexual development. The initial administration of small doses of testosterone esters followed by a gradual increase to 100 to 150 mg/m2 of body surface area every 1 to 3 weeks should result in a normal pubertal growth spurt. The time from the start of treatment to the appearance of secondary sex characteristics is variable. Penile development, deepening of the voice, and the appearance of other secondary sexual characteristics usually commence during the first year of treatment. In normal boys, puberty extends over several years, and treatment designed to replicate normal development does not shorten the process greatly.

Testosterone exerts its full action only in the presence of a balanced hormonal environment and, particularly, in the presence of adequate levels of growth hormone. Consequently, prepubertal boys with coexisting growth hormone and androgen deficiency respond poorly to androgens unless growth hormone is given simultaneously.

Pharmacologic Uses Androgens have been used for a variety of disorders unassociated with hypogonadism in the hope that potential benefits from the nonvirilizing actions of the agents (such as increases in nitrogen retention, muscle mass, and hemoglobin) would outweigh any deleterious actions of the drugs. The most common nonreplacement uses of androgen have been attempts to improve nitrogen balance in catabolic states (e.g., AIDS), self-administration by athletes to increase muscle mass and/or athletic performance, attempts to enhance erythropoiesis in refractory anemias (including the anemia of renal failure), treatment of hereditary angioedema and endometriosis, and management of growth retardation of various etiologies. Most of the expected benefits in these disorders have not been realized for two reasons. First, modest pharmacologic doses of androgens have little physiologic effect in men when superimposed on normal testicular androgen, and in women the virilizing side effects of androgens are formidable. Second, no androgen has been devised that exhibits only the nonvirilizing effects of the hormone. This conclusion is not surprising in view of the fact that the physiologic actions of androgens are mediated by a single, high-affinity receptor (Fig. 335-5).

The most pervasive form of androgen abuse is by male athletes in the expectation that muscle development and athletic performance will be improved. In controlled studies using modest pharmacologic doses (two to four times the usual replacement doses), these agents do not improve performance consistently. However, at the doses frequently taken by athletes (which sometimes exceed 10 times the replacement dose), androgens do enhance nitrogen balance and muscle mass; since the drugs have multiple side effects at high doses, these benefits do not outweigh the risks associated with androgen abuse in man, and the use of androgens by female athletes is associated with disfiguring virilization. Thus this practice cannot be condemned too harshly. The only established indications for androgen therapy outside of male hypogonadism are in selected patients with anemia due to bone marrow failure or hereditary angioedema and as an adjunct to growth hormone therapy.

Gonadotropins Gonadotropin therapy is used to establish or restore fertility in patients with gonadotropin deficiency of any cause. Several gonadotropin preparations are available. Human menopausal gonadotropin (hMG) (purified from the urine of postmenopausal women) contains 75 IUFSH and 75 IU LHper vial.hCG(purified from the urine of pregnant women) has little FSH activity and resembles LH in its ability to stimulate testosterone production by Leydig cells. Because of the expense of hMG, treatment is usually begun with hCG alone, and hMG is added later to promote the FSH-dependent stages of spermatid development. A high ratio of LH to FSH activity and treatment for 6 to 18 months may be necessary to bring about maturation of the prepubertal testes. Recombinant human FSH is also available and has been used mainly for ovulation induction in women. Trials are underway to examine its effects on spermatogenesis in men with hypogonadotropic hypogonadism. Once spermatogenesis is restored with combined FSH and LH therapy, hCG alone is often sufficient to maintain spermatogenesis.

The dose of hCG required to maintain a normal testosterone level varies from 1000 to 5000 IU weekly. A number of regimens have been used to induce maturation of spermatogenesis. Most involve starting with 2000 IU hCG three or more times a week until most of the clinical parameters, including plasma testosterone levels, are normal. hMG (usually one ampule) is then added three times a week to complete the development of spermatogenesis. The length of therapy required to restore spermatogenesis may be as long as 12 months.

GnRH and GnRH Analogues GnRH (gonadorelin) is available for endocrine testing and is used by some physicians for chronic therapy of the infertility of hypogonadotropic hypogonadism. In the latter instance, it is necessary to administer GnRH in frequent boluses (25 to 200 ng/kg of body weight every 2 h) with the use of portable infusion pumps, analogous to those used for insulin administration. In general, pulsatile GnRH does not appear to be more efficacious than gonadotropin in returning sperm counts to normal. GnRH analogues (leuprolide, nafarelin, histrelin) are available for the suppression of gonadotropin secretion, leading to hypogonadism. In prostatic cancer, testicular androgen production can be blocked by monthly injection of 7.5 mg leuprolide in depot form.

(Bibliography omitted in Palm version)

336. DISORDERS OF THE OVARY AND FEMALE REPRODUCTIVE TRACT - *Bruce R. Carr, Karen D. Bradshaw*

The ovary is the source of ova for reproduction and of the hormones that regulate female sexual life. The anatomic structure, response to hormonal stimuli, and secretory capacity of the ovary vary at different periods of life. This chapter will review normal ovarian physiology as a background for understanding ovarian abnormalities and will consider other disorders of the female reproductive tract.

DEVELOPMENT, STRUCTURE, AND FUNCTION OF THE OVARY

EMBRYOLOGY

During the third week of gestation, the primordial germ cells differentiate from the endoderm lining the yolk sac at the caudal end of the embryo. The germ cells migrate to the genital ridge adjacent to the mesonephric kidney by the fifth week of gestation and undergo mitotic division. The gonads exist in an undifferentiated state until the seventh week of fetal life, at which time the primitive ovary can be distinguished from the testis (Chap. 338). Estrogen formation in the ovary commences between weeks 8 and 10, and by 10 to 11 weeks of gestation, oogonia in the ovarian cortex begin developing into primary oocytes. The ovary contains a finite number of germ cells, the number peaking at about 7 million oogonia by the fifth to sixth month of gestation. Subsequently, the germ cells decrease in number through a process of atresia so that only 1 million remain at birth, 400,000 are present at menarche, and only a few remain at menopause. Two normal X chromosomes are required for development of the ovary; in individuals with a 45,X karyotype, ovarian development occurs, but the rate of atresia is accelerated so that only a fibrous streak remains at birth (Chap. 338).

After the oogonia cease to proliferate, meiosis commences, continues until the diplotene stage of the first meiotic division is completed, and then is arrested until the onset of ovulation at puberty. From the fifth month of fetal life, the primordial follicle consists of the primary oocyte arrested in meiosis, a single surrounding layer of granulosa cells, and a basement membrane that separates the primordial follicle from surrounding stromal (interstitial) tissues.

PUBERTAL MATURATION

The final maturation of ovarian follicles commences during puberty. The two major hormones that regulate follicular development are the pituitary gonadotropins -- follicle-stimulating hormone (FSH) and luteinizing hormone (LH) (Fig. 336-1). During the second trimester of fetal development, the plasma gonadotropins rise to levels similar to those at menopause. This peak in gonadotropin levels may be responsible for the simultaneous peak in oocyte replication. After the second trimester, the hypothalamic-pituitary axis (the so-called gonadostat) becomes functional and is sensitive to negative feedback by steroid hormones, particularly estrogen and progesterone produced in the placenta. The levels of circulating gonadotropins consequently decrease, and gonadotropins are almost undetectable at the time of birth. In the neonate, concomitant with the decrease in estrogen and progesterone levels caused by separation from the placenta, there is a rebound increase in gonadotropin

secretion for the first few months of life. With continued maturation of the hypothalamic-pituitary system, the gonadostat becomes exquisitely sensitive to negative feedback by low levels of circulating steroid hormones, and plasma gonadotropins again decrease.

As the time of puberty nears, a decrease in the sensitivity of the gonadostat allows for increased secretion of FSH and LH, possibly secondary to increased episodic or pulsatile secretion of gonadotropin-releasing hormone (GnRH) by the hypothalamus (Chap. 328). A sleep-induced, pulsatile pattern of LH secretion then ensues, the first step in the development of a cyclic pattern of gonadotropin secretion (Fig. 336-1). The increase in estrogen secretion exerts a positive feedback, which leads to an exaggeration of the pulsatile release of LH and eventually to menarche and ovulation, after which plasma gonadotropin concentrations reach adult values, which are similar during day and night. After the menopause, plasma gonadotropin levels rise, then plateau 5 to 10 years after menopause and remain fairly constant until the eighth to ninth decade of life, when the levels may fall. Although ovarian function is regulated primarily by LH and FSH, the ovary is a source of peptide and protein hormones and growth factors such as inhibin and activin that may play a role in ovarian function and regulation. The production of inhibin by the mature ovary accounts, in part, for the relative reduction in FSH that is seen during the reproductive years (Fig. 336-1).

With puberty the sensitivity of the hypothalamic-pituitary centers to circulating steroid hormones is decreased, GnRH release by the hypothalamus increases, gonadotropin secretion by the pituitary is enhanced, ovarian estrogen secretion increases, and the anatomic changes of puberty ensue. At age 10 to 11, the first secondary sexual characteristics begin to appear in girls, namely, development of the breast buds (thelarche), followed by the development of pubic hair (pubarche), and later by the development of axillary hair (adrenarche). The growth of pubic and axillary hair is believed to be initiated by adrenal androgens, the levels of which begin to rise at approximately 6 to 8 years of age. A growth spurt ensues, and peak growth rate is attained by age 12.

The culmination of puberty is the onset of predictable, cyclic menses. The average time between the beginning of breast development and the onset of menses (menarche) is 2 years. During the first few years after menarche, menstrual cycles are often irregular and unpredictable due to anovulation. The age of menarche is variable and is influenced by socioeconomic and genetic factors and by general health. In the United States, the mean age of menarche is believed to have decreased at a rate of 3 to 4 months per decade over the past 100 years and is now around 12 years, a change believed to be due to improved nutrition. A body weight of around 48 kg or some critical combination of weight, body water, and body fat is associated with development of hypothalamic insensitivity to feedback by steroids that leads to increased secretion of gonadotropins and finally to menarche. Obese girls have earlier menarche than girls with normal weights. In contrast, active participation in sports or ballet, malnutrition, and chronic debilitating disease can delay menarche.

MATURE OVARY

Morphology The anatomic components and function of the adult ovary are illustrated

schematically in Fig. 336-2. Under the influence of gonadotropins, a group of primary follicles are recruited, and by day 6 to 8 of the menstrual cycle, one follicle becomes mature or "dominant," a process characterized by accelerated growth of granulosa cells and enlargement of the fluid-filled antrum. The recruited follicles not destined to ovulate undergo degeneration, similar to the atresia that occurs in other follicles during embryogenesis. Just prior to ovulation, meiosis resumes in the ovum of the dominant follicle, and the first meiotic division results in formation of the first polar body. The antrum rapidly enlarges (up to 10 to 25 mm in size), follicular fluid increases in amount, and the follicular surface thins and forms a conical stigma. Ovulation from the dominant follicle occurs some 16 to 23 h after the LH peak and some 24 to 38 h after the onset of the LH surge as the result of rupture of the follicular wall at the area of the stigma. The ovum is then expelled together with a mass of surrounding granulosa cells called *cumulus cells*. The rupture is believed to result from the action of hydrolytic enzymes on the surface of the follicle, possibly under the control of prostaglandins. The second meiotic division occurs after the egg is fertilized by a sperm, and the second polar body is then extruded. The formation of the *corpus luteum* begins in the retained remnant of the ovulated follicle; the remaining granulosa and theca cells increase in size and accumulate lipids and a yellow pigment, lutein, to become "luteinized." After a period of 14± 2 days (the functional life of the corpus luteum), the corpus luteum begins to atrophy, to be replaced in time by a fibrous scar, the *corpus albicans*. The factors that limit the life span of the human corpus luteum are not known, but if pregnancy occurs, the corpus luteum persists under the influence of placental or chorionic gonadotropins, and progesterone is produced by the corpus luteum for the support of pregnancy.

Hormone Formation

Steroid Hormones Like other steroid hormones, ovarian steroids are derived from cholesterol (Fig. 336-3). The ovary can synthesize cholesterol de novo and can also utilize cholesterol obtained from circulating lipoproteins as substrate for steroid hormone formation (Fig. 336-4). Virtually all ovarian cells are believed to possess the complete complement of enzymes required for the synthesis of estradiol from cholesterol (Fig. 336-3); however, different cell types in the ovary contain different amounts of these enzymes so that the main steroids produced differ in different compartments. For example, the corpus luteum forms mainly progesterone and 17-hydroxyprogesterone, whereas theca and stromal cells convert cholesterol to androstenedione and testosterone. Granulosa cells are particularly rich in the aromatase enzyme responsible for estrogen synthesis and utilize as substrates for this process androgens synthesized in the granulosa cells and the adjacent theca cells.

The principal sites of action of LH and FSH are also illustrated in Figs. 336-3 and 336-4. LH acts primarily to regulate the early steps in steroid hormone biosynthesis, namely, the transport of cholesterol into the mitochondria by steroidogenic acute regulatory (StAR) protein and its conversion to pregnenolone. FSH acts mainly to regulate the final process by which androgens are aromatized to estrogens. As a consequence, LH enhances substrate flow and the formation of androgens and/or progesterone in the absence of FSH, whereas FSH action is impeded in the absence of LH because of diminished substrate for aromatization.

ESTROGENS Naturally occurring estrogens are 18-carbon steroids characterized by an

aromatic A ring, a phenolic hydroxyl group at C-3, and either a hydroxyl group (estradiol) or a ketone (estrone) at C-17 (Fig. 336-3). (For the numbering of the steroid ring, see Fig. 335-1.) The principal estrogen secreted by the ovary and the most potent estrogen is estradiol. Estrone is also produced by the ovary, but most estrone is formed by extraglandular conversion of androstenedione in peripheral tissues. Estriol (16-hydroxyestradiol), the main estrogen in urine, arises from the 16-hydroxylation of estrone and estradiol. Catechol estrogens are formed by hydroxylation of estrogens at the C-2 or C-4 position and may act as the intracellular mediators of some estrogen action. Estrogens promote development of the secondary sexual characteristics in women and cause uterine growth, thickening of the vaginal mucosa, thinning of the cervical mucus, and development of the ductule system of the breasts. Estrogens also alter lipid profiles and exert vascular effects that help prevent cardiovascular disease. The mechanism of estrogen action in target tissues is similar to that for other steroid hormones and involves binding to a nuclear steroid receptor -- either estrogen receptor (ER)a or ERb -- and enhancement of the transcription of messenger RNA, which in turn causes increased protein synthesis in the cell cytoplasm (Chap. 327). These receptors have specific tissue site expression and bind various estrogens with different affinities, thereby conferring selective actions.

PROGESTERONE Progesterone, a 21-carbon steroid (Fig. 336-3), is the principal hormone secreted by the corpus luteum and is responsible for progestational effects, i.e., induction of secretory activity in the endometrium of the estrogen-primed uterus in preparation for implantation of the fertilized egg. Progesterone also induces a decidual reaction in endometrium. Other effects include inhibition of uterine contractions, an increase in the viscosity of cervical mucus, glandular development of the breasts, and an increase in basal body temperature (thermogenic effect).

ANDROGENS The ovary synthesizes a variety of 19-carbon steroids, including dehydroepiandrosterone, androstenedione, testosterone, and dihydrotestosterone, principally in stromal and thecal cells. The major ovarian 19-carbon steroid is androstenedione (Fig. 336-3), part of which is secreted into the plasma and part of which is converted to estrogen in granulosa cells or to testosterone in the interstitium. Androstenedione can also be converted to testosterone and estrogens in peripheral tissues. Only testosterone and dihydrotestosterone are true androgens that interact with the androgen receptor and induce virilizing signs in women (Chaps. 53 and335).

Other Hormones Inhibin is secreted in two forms (A and B) by the follicle and inhibits the release ofFSH by the hypothalamic-pituitary unit. *Activin* is also secreted by the follicle and may enhance FSH secretion as well as having local effects on ovarian steroidogenesis. *Follistatin* is an activin-binding protein that attenuates the actions of activin and other members of the transforming growth factor (TGF)bfamily.

Some ovarian hormones play an uncertain role in human physiology. *Relaxin*, a polypeptide hormone produced by the human corpus luteum and by the decidua, causes softening of the cervix and loosening of the symphysis pubis in preparation for parturition in animals. *Oxytocin*, *vasopressin*, and other hypothalamic and pituitary hormones are also present in granulosa and/or luteal cells, but their function in these cells is unknown. *Follicle regulatory protein* (FRP), found in human follicular fluid, inhibits granulosa secretion and growth. *Gonadocrinins*, peptides purified from rat

follicular fluid, stimulate the release of both FSH and LH from the pituitary in vitro and in vivo. Granulosa cells secrete *oocyte maturation inhibitor* (OMI), a factor that prevents premature ovulation. In addition, in the gonads of both sexes a *meiosis-inducing substance* (MIS) triggers the onset of meiosis, an event that occurs earlier in ovarian than in testicular development. Local growth factors [including insulin-like growth factors (IGFs) 1 and 2 and TGFa and -b] may also influence steroid secretion by the ovary.

The Normal Menstrual Cycle The menstrual cycle is divided into a follicular or proliferative phase and a luteal or secretory phase (Fig. 336-5). The secretion of FSH and LH is fundamentally under negative feedback control by ovarian steroids (particularly estradiol) and by inhibin (which selectively suppresses FSH), but the response of gonadotropins to different levels of estradiol varies. FSH secretion is inhibited progressively as estrogen levels increase -- typical negative feedback. In contrast, LH secretion is suppressed maximally by sustained low levels of estrogen and is enhanced by a rising level of estradiol -- positive feedback. Feedback of estrogen involves both the hypothalamus and pituitary. Negative feedback suppresses GnRH and inhibits gonadotropin production. Positive feedback is associated with an increased frequency of GnRH secretion and enhanced pituitary sensitivity to GnRH.

The length of the menstrual cycle is defined as the time from the onset of one menstrual bleeding episode to onset of the next. In women of reproductive age, the cycle averages 28± 3 days and the mean duration of flow is 4± 2 days. Longer menstrual cycles (usually characterized by anovulation) occur at menarche and near the onset of menopause. At the end of a cycle plasma levels of estrogen and progesterone fall, and circulating levels of FSH increase. Under the influence of FSH, follicular recruitment results in development of the follicle that will be dominant during the next cycle.

After the onset of menses, follicular development continues, but FSH levels decrease. Approximately 8 to 10 days prior to the midcycle LH surge, plasma estradiol levels begin to rise as the result of estradiol formation by the granulosa cells of the dominant follicle. During the second half of the follicular phase, LH levels also begin to rise (owing to positive feedback). Just before ovulation, estradiol secretion reaches a peak and then falls. Immediately thereafter, a further rise in the plasma level of LH mediates the final maturation of the follicle, followed by follicular rupture and ovulation 16 to 23 h after the LH peak. The rise in LH is accompanied by a smaller increase in the level of plasma FSH, the physiologic significance of which is unclear. The plasma progesterone level also begins to rise just prior to midcycle and facilitates the positive feedback action of estradiol on LH secretion.

At the onset of the luteal phase, plasma gonadotropins decrease and plasma progesterone increases. A secondary rise in estrogens causes further gonadotropin suppression. Near the end of the luteal phase, progesterone and estrogen levels fall, and FSH levels begin to rise to initiate the development of the next follicle (usually in the contralateral ovary) and the next menstrual cycle. Inhibin A levels are low in the follicular phase but reach a peak in the luteal phase. Inhibin B levels, in contrast, are increased in the follicular phase and low in the luteal phase.

The endometrium lining the uterine cavity undergoes marked alterations in response to the changing plasma levels of ovarian hormones (Fig. 336-5). Concurrent with the

decrease in plasma estrogen and progesterone and the decline of corpus luteum function in the late luteal phase, intense vasospasm occurs in the spiral arterioles supplying blood to the endometrium, causing ischemic necrosis, endometrial desquamation, and bleeding. This vasospasm is caused by locally synthesized prostaglandins. The onset of bleeding marks the first day of the menstrual cycle. By the fourth to fifth day of the cycle, the endometrium is thin. During the proliferative phase, glandular growth of the endometrium is mediated by estrogen. After ovulation, increased progesterone levels lead to further thickening of the endometrium, but the rapid growth slows. The endometrium then enters the secretory phase, characterized by tortuosity of the glands, curling of the spiral arterioles, and glandular secretion. As corpus luteum function begins to wane in the absence of conception, the sequence of events leading to menstruation is again set into action.

Biphasic changes in basal body temperature are characteristic of the ovulatory cycle and are mediated by alterations in progesterone levels (Fig. 336-5). An increase in basal body temperature by 0.3 to 0.5°C begins after ovulation, persists during the luteal phase, and returns to the normal baseline (36.2 to 36.4°C) after the onset of the subsequent menses.

Cellular Interactions in the Ovary during the Normal Cycle LHstimulates thecal cells surrounding the follicle to form androgens, and androstenedione diffuses across the basement membrane of the follicle into granulosa cells, where it is aromatized to estrogen (Figs. 336-3 and336-4).

The increase ofFSH late in the preceding menstrual cycle stimulates growth and recruitment of the primary follicles by enhancing granulosa cell proliferation, resulting ultimately in the formation of the dominant follicle. This function of FSH is underscored by the fact that only primary follicles are seen in patients with mutations in FSH or the FSH receptor. In the granulosa cells, FSH also stimulates estrogen synthesis. Enhanced secretion of estradiol causes an increase in the number of estradiol receptors and further proliferation of granulosa cells. In the late follicular phase, FSH, in concert with estradiol, causes induction ofLHreceptors on the granulosa cells. LH acts via these receptors to increase progesterone secretion at midcycle. The amount of progesterone formed by the follicle is believed to be limited by the availability of cholesterol to serve as substrate for steroidogenesis and by the fact that most of the progesterone is converted to androstenedione by thecal cells. Prior to ovulation, the granulosa cells of the follicle are bathed in follicular fluid but have limited access to circulating blood and consequently to plasma low-density lipoprotein (LDL). As depicted in Fig. 336-4, the granulosa cells become vascularized after ovulation, and plasma cholesterol is made available to serve as the major substrate for progesterone synthesis by the corpus luteum. Thus, increased progesterone synthesis by the corpus luteum is the consequence of increased substrate availability. The peak in progesterone secretion by the corpus luteum occurs 8 days after ovulation at the time of maximal vascularization of the granulosa cells.

MENOPAUSE

The *menopause* is defined as the final episode of menstrual bleeding in women. However, the term is used commonly to refer to the time interval that encompasses the

transitional period between the reproductive years up to and after the last episode of menstrual bleeding. During this period, there is a progressive loss of ovarian function and a variety of endocrine, somatic, and psychological changes.

The median age of women at the time of cessation of menstrual bleeding is 50 to 51 years. Since the life expectancy of women is close to 80 years, approximately one-third of life occurs after cessation of reproductive function. Preceding the menopause, the pattern of menstrual cycles is variable, but the interval between menses usually becomes shorter, as follicular recruitment is hastened by increases in FSH. Day 3 FSH and E$_2$ levels are often elevated. Ovulatory cycles continue for some period of time, then anovulation becomes common.

The menopause is the consequence of the exhaustion of ovarian follicles. The decrease in the number of ova begins in intrauterine life; by the time of the menopause, few ova remain, and these appear to be nonfunctional. Only a small number of ova are lost as the result of ovulation during reproductive life; the majority are lost by atresia. The cessation of follicular development results in decreased production of estradiol, inhibin, and other hormones, which causes a loss of negative feedback on the hypothalamic-pituitary centers. In turn, the levels of plasma gonadotropins increase, with FSH levels rising earlier and higher than LH levels (Figs. 336-1 and 336-6). The higher concentration of FSH than LH in postmenopausal women may result from the decrease in inhibin secretion by the ovary, from the fact that FSH is cleared from plasma less rapidly than LH, and possibly from the loss of positive feedback on LH production by estradiol.

The ovaries of postmenopausal women are small and wrinkled, and the residual cells are predominantly stromal. Estrogen and androgen levels in plasma are reduced but not absent (Fig. 336-6). Before the menopause, plasma androstenedione is derived almost equally from the adrenals and the ovaries; after menopause the ovarian contribution ceases so that the plasma levels of androstenedione fall by 50% (Fig. 336-6). However, the menopausal ovary continues to secrete testosterone, presumably formed in stromal cells.

Circulating estrogens in the ovulating woman are derived from two sources. Some 60% of mean estrogen formation during the menstrual cycle is in the form of estradiol, formed primarily by ovaries, and the remainder is estrone, formed mainly in extraglandular tissues from androstenedione. After menopause, extraglandular estrogen formation is the major pathway for estrogen synthesis. Because adipose tissue is a major site of extraglandular estrogen production, peripheral estrogen formation may actually be enhanced in obese postmenopausal women, so that total estrogen production rates may be as great or greater than in premenopausal women. The predominant estrogen formed is estrone rather than estradiol.

The most common menopausal symptoms are vasomotor instability (hot flashes), atrophy of the urogenital epithelium and skin, decreased size of the breasts, and osteoporosis. Approximately 40% of menopausal women develop symptoms serious enough to seek medical assistance.

The pathogenesis of the hot flash is uncertain. There is a close temporal relationship

between the onset of the hot flash and pulses of LH secretion, which reflect hypothalamic secretion of GnRH. Alterations in catecholamine, prostaglandin, endorphin, or neurotensin metabolism may play a role in conjunction with low estrogen production. Symptoms associated with the hot flash, including nervousness, anxiety, irritability, and depression, may or may not be caused entirely by estrogen deficiency.

The decrease in size of the tissues of the female reproductive tract and breasts in the menopause is due to estrogen deficiency. The vaginal mucosa and the endometrium usually become thin and atrophic (although endometrial hyperplasia occurs in one-fifth of postmenopausal women).

Osteoporosis is one of the dread afflictions of aging, and there is a close relationship between estrogen deprivation and its development. Approximately one-fourth of aging women and one-tenth of elderly men sustain a vertebral or hip fracture between the ages of 60 and 90, and the incidence is highest in elderly white women. Such fractures are a major cause of loss of independence, death and morbidity, and the fracture-related mortality increases from <10% in the 60- to 64-year age group to >30% in patients over 80 (Chap. 342). Many factors affect the development of osteoporosis, including ethnic origin, diet, activity, smoking, and general health, and estrogen deprivation is of particular importance. White and Asian postmenopausal women are more predisposed to osteoporosis and its consequences because bone mass in this group is lower prior to menopause, so loss in bone density has more severe consequences. Further evidence that osteoporosis is a disease of estrogen deprivation is suggested by the early development of osteoporosis in women with premature menopause due to either natural causes or surgical castration. After the menopause women experience an increase in the incidence of cardiovascular disease as the result of a decrease in the level of high-density lipoprotein (HDL) cholesterol as well as the effects of hypoestrogenism on vascular endothelium and reactivity.

LABORATORY AND CLINICAL ASSESSMENT OF HORMONAL STATUS

The hormonal status of women can usually be assessed by history and physical examination. In general, the presence of secondary sexual characteristics such as normal female breast development indicates adequate estrogen secretion in the past, and the presence of regular, predictable, cyclic menses implies that ovulation and the production of gonadotropins, estrogen, progesterone, and androgens are adequate and that the outflow tract is intact. Such a history may be more valuable than laboratory tests in evaluating ovarian hormone status. However, laboratory tests provide valuable ancillary information in the evaluation of women with endocrine dysfunction or infertility (Chap. 54).

PITUITARY GONADOTROPINS

Plasma gonadotropins are assessed by radioimmunoassay (RIA), fluoroimmunoassay (FIA), or immunoradiometric assay (IRMA). Because both FSH and LH are secreted in a pulsatile manner, the results obtained from a single serum sample may be difficult to interpret. Consequently, multiple samples taken at 20-min intervals over 2 h may be pooled to obtain a mean value. Serum gonadotropin measurements are of the most use in evaluating women with suspected ovarian failure and in supporting the diagnosis of

polycystic ovarian syndrome (PCOS) and hypogonadotropic hypogonadism. The normal ranges for serum LH and FSH in ovulating women are 0.8 to 57 and 1.4 to 21 IU/L, respectively. FSH levels that are persistently >40 IU/L are diagnostic of ovarian failure, and an LH value <0.8 IU/L suggests hypogonadotropic hypogonadism. In practice, however, gonadotropin values may be equivocal and must be interpreted in light of the remainder of the findings.

OVARIAN HORMONES

The mean plasma levels and production rates of the principal ovarian hormones are presented in Table 336-1.

Estrogen The presence of normal secondary sexual characteristics implies that estrogen production was adequate in the past. The current estrogen status can be estimated by pelvic examination. The presence of a moist, rugated vagina with copious, clear, thin cervical mucus that can be stretched and that exhibits arborization or ferning when spread on a slide is strong evidence of adequate estrogen production. Cytologic demonstration of mature vaginal epithelial cells and abundant cornified squamous epithelial cells with pyknotic nuclei confirms the presence of adequate estrogen levels.

The progesterone-withdrawal test provides a functional assessment of the endometrium, outflow tract, and estrogen status. If menses appear within a week to 10 days after the end of a trial of medroxyprogesterone acetate (10 mg by mouth once or twice a day for 5 days) or after a single intramuscular injection of progesterone (100 mg), then prior estrogen priming was adequate to allow withdrawal bleeding.

Owing to its variable level in plasma during the normal cycle and the difficulty of estimating the day of the cycle in women with abnormal cycles, the measurement of estrogen levels in plasma or urine is of little use in the routine assessment of estrogen status. Measurement of plasma estradiol is useful during attempts to induce ovulation with gonadotropins to prevent the development of the ovarian hyperstimulation syndrome and is used along with ultrasound assessment to monitor follicular growth in women who are to undergo in vitro fertilization.

Progesterone Cyclic, predictable menses also imply that adequate progesterone is secreted during the luteal phase of the menstrual cycle. Assessment of progesterone is useful to detect ovulation and to evaluate the adequacy of the luteal phase in infertile women. Several functional assays of progesterone can be used. The least expensive and most useful is the daily measurement of basal body temperature throughout a cycle. Owing to the thermogenic properties of progesterone, a normal biphasic monthly curve showing a temperature elevation lasting for approximately 2 weeks after ovulation is a valid indication of progesterone secretion during the luteal phase (Fig. 336-5). The presence of viscous cervical mucus that does not stretch or fern and of predominantly intermediate cells on vaginal cytology or demonstration of a secretory epithelium in an endometrial biopsy during the luteal phase on days 20 to 22 of the cycle provides additional assessment of progesterone secretion. In addition, serum progesterone can be measured to assess the function of the corpus luteum.

Androgen Under normal conditions, the ovary secretes androstenedione, testosterone,

and dehydroepiandrosterone. In conditions of androgen excess, hirsutism and/or virilization are common. The evaluation of androgen excess is discussed in Chap. 53.

DIAGNOSIS OF PREGNANCY

Pregnancy is usually recognized on the basis of the history and physical examination. That is, a woman with previously cyclic, predictable menses develops amenorrhea accompanied by breast tenderness, malaise, lassitude, and nausea, and on physical examination the uterus is soft and enlarged.

Assays of placental products facilitate the diagnosis of pregnancy. Human chorionic gonadotropin (hCG) is secreted by the trophoblastic cells of the placenta into the maternal plasma and excreted in the urine. Assays of the hCG content of serum or urine use antibodies against hCG and make it possible to detect pregnancies 8 to 10 days after ovulation, before the first missed menstrual period and long before pregnancy can be diagnosed by clinical assessments. Assay of the b subunit of hCG in serum or urine makes it possible to differentiate between excess LH and hCG, an important distinction in evaluating women with trophoblastic disease such as hydatidiform mole or choriocarcinoma. Sensitive and specific hCG-based pregnancy tests are now available for testing by patients at home.

DISORDERS OF OVARIAN FUNCTION

PREPUBERTAL YEARS

Puberty is said to be *precocious* if breast budding begins before age 8 or if menarche occurs before age 9. Those disorders in which the developing sexual characteristics are appropriate for the genetic and gonadal sex -- i.e., feminization in girls or virilization in boys -- are termed *isosexual precocity*, whereas *heterosexual precocity* occurs when sexual characteristics are not in accord with the genetic sex, namely, virilization in girls or feminization in boys. Pubertal disorders of boys are described in Chap. 335.

Isosexual Precocious Puberty Isosexual precocious puberty in girls can be divided into three major categories (Table 336-2).

True Precocious Puberty True precocious puberty is characterized by an early but otherwise normal sequence of pubertal development, including increased secretion of gonadotropins and ovulatory menstrual cycles. Constitutional or idiopathic precocious puberty accounts for 90% of cases. In these individuals, no cause for the premature maturation of the central nervous system-hypothalamic-pituitary axis can be identified, and the diagnosis is confirmed by finding an adult pattern of LH and FSH release on a GnRH stimulation test. As many as half these individuals have abnormal findings on electroencephalograms. Premature appearance of secondary sexual characteristics and of ovulatory cycles with the accompanying risk of fertility may cause significant emotional disturbance. Therefore, prompt initiation of therapy is imperative. GnRH analogues suppress gonadotropins and inhibit estrogen synthesis, thereby blocking precocious puberty; they may also prevent premature closure of the epiphyses and the resulting short stature.

About 10% of cases are due to organic brain diseases, including brain tumors (hypothalamic gliomas, astrocytomas, ependymomas, germinomas, and hamartomas), encephalitis, meningitis, hydrocephalus, head injury, tuberous sclerosis, and neurofibromatosis. It is essential to distinguish this group of patients from those with the idiopathic disorder, and patients whose disorder is designated as idiopathic occasionally prove to have such tumors. Fortunately, most patients with organic lesions serious enough to cause precocious puberty have obvious neurologic signs and symptoms. Evaluation of all patients with precocious puberty should include, at a minimum, skull films and computed tomography (CT) or magnetic resonance imaging (MRI) of the brain. The success of treatment depends on the nature of the lesion, but surgical and radiation treatment of well-localized tumors is occasionally successful.

A rare cause of isosexual precocity is congenital adrenal hyperplasia due to 21-hydroxylase deficiency in girls in whom treatment is delayed until 4 to 8 years of age. After initiation of glucocorticoid replacement, such individuals may undergo isosexual precocious puberty (Chap. 331).

Precocious Pseudopuberty Precocious pseudopuberty occurs when girls undergo feminization as a consequence of enhanced estrogen formation but do not ovulate or develop cyclic menses. Ovarian cysts or tumors that secrete estrogen (granulosa-theca cell tumors) are the most frequent cause of precocious pseudopuberty. Granulosa-theca cell tumors associated with intestinal polyps and pigmentation of the mucous membranes occur in the Peutz-Jeghers syndrome. Other ovarian tumors that secrete estrogens (or androgens that can be converted to estrogens at extraglandular sites) include dysgerminomas, teratomas, cystadenomas, and ovarian carcinomas (Chap. 97). Ovarian tumors can usually be detected by rectoabdominal examination or by sonography, CT, MRI, and/or laparoscopy. Ovarian teratomas and choriocarcinomas and other carcinomas that secrete hCG do not cause precocious puberty in girls unless they also secrete estrogen (hCG or LH in the absence of FSH does not induce ovarian estrogen production). Rarely, feminizing tumors of the adrenal cause isosexual precocious puberty by direct formation of estrogens or by secretion of weak androgens, which are converted to estrogens in extraglandular tissues.

Other causes of precocious pseudopuberty include the following:

1. The McCune-Albright syndrome (polyostotic fibrous dysplasia) is due to an activating mutation in the G-protein, Gsa, that occurs during embryogenesis, leading to a mosaic pattern of expression in various tissues. It is characterized by cafe au lait spots, cystic fibrous dysplasia of bones, and sexual precocity. In the ovary, the Gsa mutation mimics the action of FSH, leading to autonomous follicle development and estrogen formation. Occasionally, this disorder leads to true precocious puberty (Chap. 343).

2. Primary hypothyroidism is occasionally associated with enhanced secretion of FSH, inducing ovarian estrogen secretion. High levels of thyroid-stimulating hormone (TSH) caused by hypothyroidism may also stimulate the FSH receptor.

3. The Russell-Silver syndrome, or congenital asymmetry, is associated with short stature and precocious feminization.

4. Estrogen-containing medications, including use of estrogen-containing creams for diaper rash or the ingestion of meat from estrogen-treated animals or poultry or of any estrogen by mouth, can cause this disorder.

Incomplete Isosexual Precocity This term is used to describe the premature development of a single pubertal event and encompasses several entities. Breast budding prior to age 7 (*premature thelarche*) without other evidence of estrogen secretion and without premature bone maturation is believed to be due to a transient increase in estrogen secretion or to a temporary increase in sensitivity to the small amounts of circulating estrogens formed prior to puberty. Usually, the disorder is self-limited and resolves spontaneously. Occasionally, axillary hair and/or pubic hair (*premature adrenarche* and *premature pubarche*) appear without any other secondary sexual development. The phenomenon is associated with adrenal androgen secretion in the range of normal puberty and can be distinguished from syndromes of virilization by the absence of clitoromegaly. It requires no treatment, and patients enter puberty at about the average time.

Heterosexual Precocity Virilization in a prepubertal female is usually due to congenital adrenal hyperplasia or to androgen secretion by an ovarian or adrenal tumor. The manifestations of virilization are described inChaps. 53 and331. Virilization in girls with congenital adrenal hyperplasia usually takes place in a background of variable sexual ambiguity (seeChap. 338).

Evaluation of Sexual Precocity The evaluation of sexual precocity involves a careful history and physical examination, including rectoabdominal examination, abdominal sonography, determination of bone age, andGnRHstimulation test, and measurement of thyroid hormones,TSH, and gonadotropins (and androgen or estrogen levels when appropriate).MRIand/orCT scans should be obtained if a neurologic disorder is suspected and no evidence of ovarian or adrenal tumor is found.

REPRODUCTIVE YEARS

Disorders of the Menstrual Cycle

Abnormal Uterine Bleeding Between menarche and the menopause, almost every woman experiences one or more episodes of abnormal uterine bleeding, here defined as any bleeding pattern that differs in frequency, duration, or amount from the pattern observed during a normal menstrual cycle. A variety of descriptive terms (such as *menorrhagia*, *metrorrhagia*, and *menometrorrhagia*) have been used to characterize patterns of abnormal uterine bleeding. A more logical approach is to divide abnormal uterine bleeding into those patterns associated with ovulatory cycles and those associated with anovulatory cycles.

Ovulatory Cycles Normal menstrual bleeding with ovulatory cycles is spontaneous, regular, cyclic, and predictable and is frequently associated with discomfort (*dysmenorrhea*). Deviations from this pattern associated with cycles that are still regular and predictable are most often due to organic disease of the outflow tract. For example, regular but prolonged and excessive bleeding episodes unassociated with bleeding dyscrasias (hypermenorrhea or menorrhagia) can result from abnormalities of the uterus

such as submucous leiomyomas, adenomyosis, or endometrial polyps. Regular, cyclic, predictable menstruation characterized by spotting or light bleeding is termed *hypomenorrhea* and is due to obstruction of the outflow tract as from intrauterine synechiae or scarring of the cervix. Intermenstrual bleeding between episodes of regular, ovulatory menstruation is also often due to cervical or endometrial lesions. An exception to the association between organic disease and abnormal uterine bleeding is the occurrence of regular menstruation more frequently than 21 days apart (*polymenorrhea*). Such cycles may be a normal variant.

Anovulatory Cycles Uterine bleeding that is unpredictable with respect to amount, onset, and duration and usually painless is described as *dysfunctional uterine bleeding*. This disorder is not due to abnormalities of the uterus but rather to chronic anovulation and occurs when there is interruption of the normal sequence of follicular and luteal phases under the influence of a dominant follicle and its resulting corpus luteum. As discussed above, uterine bleeding in ovulatory cycles is due to progesterone withdrawal and requires that the endometrium first be primed with estrogen. (When castrates or postmenopausal women are given progesterone, withdrawal bleeding usually does not occur.)

Dysfunctional uterine bleeding can occur in women who have a transient disruption of the synchronous hypothalamic-pituitary-ovarian patterns necessary for ovulatory cycles, most often at the extremes of the reproductive life -- in the early menarche and in the perimenopausal period -- but also after temporary stress or intercurrent illness.

Primary dysfunctional uterine bleeding can result from three disorders.

1. *Estrogen withdrawal bleeding* occurs when estrogen is given to a castrated or postmenopausal woman and then withdrawn. As in other types of dysfunctional uterine bleeding, this form of menstrual bleeding is usually painless.

2. *Estrogen breakthrough bleeding* occurs when there is continuous estrogen stimulation of the endometrium not interrupted by cyclic progesterone secretion and withdrawal. This is the most common type of dysfunctional uterine bleeding and is usually due to anovulation associated with chronic acyclic estrogen production, as in women with PCOS. Such women may have histories of irregular, unpredictable menses, oligomenorrhea, or amenorrhea (see below). Alternatively, estrogen breakthrough bleeding can occur in hypogonadal women given estrogens chronically rather than intermittently and in women with estrogen-secreting tumors of the ovary. Estrogen breakthrough bleeding may be profuse and is unpredictable with respect to duration, amount of flow, and time of occurrence. The endometrium is typically thin because its repair between episodes of bleeding is incomplete.

3. *Progesterone breakthrough bleeding* occurs in the presence of abnormally high ratios of progesterone to estrogen, e.g., in women using continuous low-dose oral contraceptives.

The approach to a patient with dysfunctional uterine bleeding begins with a careful history of menstrual patterns and prior hormonal therapy. Since not all urogenital tract bleeding is from the uterus, rectal, bladder, and vaginal or cervical sources must be

excluded by physical examination. If the bleeding is from the uterus, a pregnancy-related disorder such as abortion or ectopic pregnancy must be ruled out.

TREATMENT

Once the diagnosis of dysfunctional uterine bleeding is established, a rational approach to management is as follows: During a first episode of dysfunctional bleeding the patient can simply be observed, provided the bleeding is not copious and no evidence of bleeding dyscrasia is present. If bleeding is moderately severe, control can be achieved with relatively high dose estrogen oral contraceptives for 3 weeks. Alternatively, a regimen of three or four low-dose oral contraceptive pills per day for 1 week followed by tapering to the usual dosage for up to 3 weeks is also effective. If uterine bleeding is more severe, hospitalization, bed rest, and intramuscular injections of estradiol valerate (10 mg) and hydroxyprogesterone caproate (500 mg) or intravenous or intramuscular conjugated estrogens (25 mg) usually control the bleeding. After initial treatment, iron replacement should be instituted, and recurrence can be prevented by cyclic oral contraceptives for 2 to 3 months (or more if pregnancy is not desired). Alternatively, menses can be induced every 2 to 3 months with medroxyprogesterone acetate, 10 mg by mouth once or twice a day for 10 days. If hormone therapy fails to control uterine bleeding, an endometrial biopsy, hysteroscopy, or dilatation and curettage may be required for diagnosis and therapy. Indeed, uterine sampling should be performed prior to hormone therapy in women at risk for endometrial cancer (i.e., in women who are approaching the age of menopause or are massively obese); endometrial cancer is rare in ovulatory women of reproductive age.

Amenorrhea An acceptable definition of amenorrhea is failure of menarche by age 15, irrespective of the presence or absence of secondary sexual characteristics, or the absence of menstruation for 6 months in a woman with previous periodic menses. However, women who do not fulfill these criteria should be evaluated if (1) the patient and/or her family are greatly concerned, (2) no breast development has occurred by age 13, or (3) any sexual ambiguity or virilization is present (Chap. 338). Amenorrhea is commonly categorized as either primary (the woman has never menstruated) or secondary (when menstruation has been present for a variable period of time in the past and has ceased). However, some disorders can cause either primary or secondary amenorrhea. For example, most women with gonadal dysgenesis have primary amenorrhea, but some have a few follicles and ovulate for short periods so that pregnancy occurs rarely. Furthermore, patients with chronic anovulation (PCOS) usually have secondary amenorrhea but on occasion have primary amenorrhea. For these reasons, categorization of amenorrhea into primary and secondary types is less helpful than a classification based on the underlying physiologic derangements: (1) anatomic defects, (2) ovarian failure, and (3) chronic anovulation with or without estrogen present.

ANATOMIC DEFECTS Anatomic or structural defects of the genital tract can preclude menstrual bleeding. Starting from the caudal end of the female genital tract, labial fusion is often associated with disorders of sexual development, particularly female pseudohermaphroditism (congenital adrenal hyperplasia or exposure to maternal androgens in utero; Chap. 338). Congenital defects of the vagina, imperforate hymen, and transverse vaginal septae can also cause amenorrhea. These women frequently have accumulation of menstrual blood behind the obstruction and may have cyclic,

predictable episodes of abdominal pain.

More severe mullerian anomalies include mullerian agenesis (the Mayer-Rokitansky-Kuster-Hauser syndrome;Chap. 338), second in frequency only to gonadal dysgenesis as a cause of primary amenorrhea. It can be caused by mutations in the genes encoding anti-mullerian hormone (AMH) or its receptor (AMHR). Women with this syndrome have a 46,XX karyotype, female secondary sex characteristics, and normal ovarian function, including cyclic ovulation, but have absence or hypoplasia of the vagina. The uterus usually consists of only rudimentary bicornuate cords, but if the uterus contains endometrium, cyclic abdominal pain and accumulation of blood may occur, as in other forms of outlet obstruction. One-third of women with this syndrome have abnormalities of the urogenital tract, and one-tenth have skeletal anomalies, usually involving the spine. The major diagnostic problem is distinguishing mullerian agenesis from complete testicular feminization, in which 46,XY genetic males with testes differentiate as phenotypic women but with a blind vaginal pouch and no uterus. Women with testicular feminization have feminized breasts but a paucity of pubic and axillary hair. The disorder is X-linked and is caused by mutations in the androgen receptor that result in profound resistance to the action of testosterone (Chap. 338). Testicular feminization can be diagnosed by demonstrating a male level of serum testosterone and a 46,XY karyotype, whereas demonstration of a 46,XX karyotype, the biphasic basal body temperature curve characteristic of ovulation, and elevated levels of progesterone during the luteal phase establish the diagnosis of mullerian agenesis.

A rare cause of absence of the uterus in 46,XY phenotypic women who are sexually infantile is the so-called testicular regression syndrome or testicular agenesis (Chap. 338).

Other abnormalities of the uterus that cause amenorrhea include obstruction due to scarring or stenosis of the cervix, often resulting from surgery, electrocautery, laser therapy, or cryosurgery. Such destruction of the endometrium (Asherman's syndrome) usually follows vigorous curettage for postpartum hemorrhage or after therapeutic abortion complicated by infection. This diagnosis is confirmed by hysterosalpingography or by direct visual examination of the endometrial scarring or synechiae using a hysteroscope.

Treatment of disorders of the outflow tract is surgical.

OVARIAN FAILURE Primary ovarian failure is associated with elevated plasma gonadotropin levels and can result from several causes. The most frequent cause is *gonadal dysgenesis*, in which the germ cells are absent and the ovary is replaced by a fibrous streak (Chaps. 65 and338). Women with gonadal dysgenesis can be divided into two broad groups on the basis of chromosomal karyotype. The most common type is due to deletion of genetic material in the X chromosomes and accounts for about two-thirds of cases of gonadal dysgenesis. A 45,X karyotype is found in about half of women with this disorder, and most have somatic defects, including short stature, webbed neck, shield chest, and cardiovascular defects, collectively termed the *Turner phenotype*. The remainder of women with X chromosome abnormalities have chromosomal mosaicism with or without associated structural abnormalities of the X. The most common form of mosaicism is 45,X/46,XX. Gonadal tumors are rare in 45,X

patients, but gonadal malignancies may occur in women with chromosomal mosaicism involving the Y chromosome. Therefore, chromosomal analysis should be performed in all cases of amenorrhea associated with ovarian failure, and the streak gonad should be removed if a Y chromosome is present. One means of identifying the presence of a Y chromosome is to amplify the sex-determining regions of the Y chromosome (SRY) by means of the polymerase chain reaction ([Chap. 338](#)). Approximately 90% of women with gonadal dysgenesis due to partial or complete deletion of the X never have menstrual bleeding, and the remaining 10% have sufficient follicles to experience menses and, rarely, fertility; the menstrual and reproductive lives of such individuals are invariably brief.

One-tenth of individuals identified as having bilateral streak gonads have a normal 46,XX or 46,XY karyotype and are said to have *pure gonadal dysgenesis*. These individuals have either normal or above-average stature, owing to failure of estrogen-mediated epiphyseal closure in the presence of a normal chromosomal constitution. Pure gonadal dysgenesis does not constitute a phenotypic or chromosomally homogeneous disorder ([Chap. 338](#)). Occasional women with a 46,XY karyotype develop signs of virilization, including clitoromegaly, and have an increased incidence of tumors in the gonadal streaks; as a consequence, gonadal streaks should be removed prophylactically, as discussed above, when a Y chromosome is present. Approximately two-thirds of women with 46,XX gonadal dysgenesis experience no menses, while the remainder have one or more menstrual episodes and are occasionally fertile.

Other causes of ovarian failure and amenorrhea include deficiency of the *CYP17* gene that encodes 17a-hydroxylase and 17,20-lyase activities, premature ovarian failure, the resistant-ovary syndrome, and ovarian failure secondary to chemotherapy or radiation therapy for malignancy. *17a-Hydroxylase deficiency* is characterized by primary amenorrhea, sexual infantilism, and hypertension, the latter due to increased production of desoxycorticosterone (DOC); whereas women with *17,20-lyase deficiency* have primary amenorrhea and sexual infantilism with normal blood pressure ([Chaps. 331](#) and [338](#)). The diagnosis of *premature ovarian failure* or *premature menopause* is applied to women who cease menstruating before age 40. The ovaries in such women are similar to the ovaries of postmenopausal women, containing few or no follicles as the result of accelerated follicular atresia. Premature ovarian failure due to ovarian antibodies may be one component of polyglandular failure, together with adrenal insufficiency, hypothyroidism, and other autoimmune disorders ([Chap. 339](#)). A rare form of ovarian failure is the *resistant-ovary syndrome*, in which the ovaries contain many follicles that are arrested in development prior to the antral stage, possibly because of resistance to the action of [FSH](#) in the ovary. A subset of these individuals have mutations in FSH or its receptor. To differentiate this disorder from the 46,XX variety of pure gonadal dysgenesis, which is also associated with sexual immaturity, it is necessary to perform ovarian biopsy. However, it is not clinically useful to make this distinction, since the conventional treatment of infertility in both conditions is usually unsuccessful. Women with ovarian failure who desire pregnancy have been treated with hormone replacement and transfer of donor embryos to the uterine cavity or fallopian tubes.

Chronic Anovulation At least 80% or more of gynecologic endocrine disorders result

from chronic anovulation. Women with chronic anovulation fail to ovulate spontaneously but may ovulate with appropriate therapy. The ovaries of such women do not secrete estrogen in a normal cyclic pattern; it is clinically useful to differentiate those women who produce enough estrogen to have withdrawal bleeding after progestogen therapy from those who do not; the latter often have hypothalamic-pituitary dysfunction.

CHRONIC ANOVULATION WITH ESTROGEN PRESENT Women with chronic anovulation who experience withdrawal bleeding after progestogen administration are said to be in a state of "estrus" due to the acyclic production of estrogen, largely estrone, by extraglandular aromatization of circulating androstenedione. This disorder is commonly termed *polycystic ovarian syndrome* (PCOS) and is characterized by infertility, hirsutism, obesity, and amenorrhea or oligomenorrhea. When spontaneous uterine bleeding occurs in women with PCOS, it is unpredictable as to time of onset, duration, and amount; on occasion the bleeding can be severe. The dysfunctional uterine bleeding is usually due to estrogen breakthrough (see above).

The disorder, as originally described by Stein and Leventhal, was characterized by enlarged, polycystic ovaries, but it is now known to be associated with a variety of pathologic findings in the ovaries, only some of which result in enlargement and none of which are pathognomonic. The most common finding is a white, smooth, sclerotic ovary with a thickened capsule, multiple follicular cysts in various stages of atresia, a hyperplastic theca and stroma, and rare or absent corpora albicans. Other ovaries have hyperthecosis in which the ovarian stroma is hyperplastic and may contain lipid-laden luteal cells. Thus, the diagnosis of PCOS is a clinical one, based on the coexistence of chronic anovulation and varying degrees of androgen excess. The fundamental defect that causes PCOS is unknown, and it is likely to have several distinct causes.

In most women with PCOS, menarche occurs at the expected time, but uterine bleeding is unpredictable in onset, duration, and amount. Amenorrhea ensues after a variable period, although primary amenorrhea occurs in some women. Signs of androgen excess (hirsutism) usually become evident around the time of menarche. One scenario suggests that this disorder originates as an exaggerated adrenarche in obese girls (Fig. 336-7). The combination of elevated levels of adrenal androgens and obesity leads to increased formation of extraglandular estrogen. This estrogen exerts a positive feedback on LH secretion and negative feedback on FSH secretion, resulting in a ratio of LH to FSH levels in plasma that is characteristically greater than 2. The increased LH levels can then lead to hyperplasia of the ovarian stroma and theca cells and increased androgen production, which in turn provides more substrate for peripheral aromatization and perpetuates the chronic anovulation. In the advanced stage of the disorder, the ovary is the major site of androgen production, but the adrenal may continue to secrete excess androgen as well. The greater the obesity, the more strongly this sequence would be perpetuated because more androgen is converted to estrogen by adipose tissue stromal cells, which in turn exaggerates inappropriate LH release by positive feedback. Ovarian follicles from women with PCOS have low aromatase activity, but normal aromatase can be induced by treatment with FSH. An association exists between PCOS/hyperthecosis, virilization, acanthosis nigricans, and insulin resistance; in the ovary, insulin may interact via the insulin-like growth factor receptors to enhance androgen synthesis in insulin-resistant states.

TREATMENT

Treatment of PCOS is directed toward interrupting the self-perpetuating cycle and can be accomplished in several ways, such as by decreasing ovarian androgen secretion (by wedge resection or the use of oral contraceptive agents), decreasing peripheral estrogen formation (by weight reduction), or enhancing FSH secretion [by administration of clomiphene, human menopausal gonadotropin (hMG), GnRH (gonadorelin) by portable infusion pump, or purified FSH (urofollitropin)]. The choice of therapy depends on the clinical findings and the needs of the patient. An attempt at weight reduction is appropriate in all who are obese. If the woman is not hirsute and does not desire pregnancy, periodic withdrawal menses can be induced with medroxyprogesterone acetate 10 days per month; such treatment prevents the development of endometrial hyperplasia. If the woman is hirsute and does not desire pregnancy, the ovarian (and possibly the adrenal) component of androgen production can be suppressed with combined estrogen-progestogen oral contraceptive agents. Combined oral contraceptives are also indicated if prolonged or excessive menstrual bleeding is present. Once androgen excess is controlled, treatment of previously existing hair growth by shaving, depilatories, or electrolysis may be indicated (Chap. 53). If pregnancy is desired, ovulation must be induced. The insulin-sensitizing drugs metformin and troglitazone improve fertility in women with PCOS. Clomiphene promotes ovulation in three-fourths of cases, or ovulation can be induced with hMG, urofollitropin, or gonadorelin (Chap. 54). Pretreatment with GnRH analogues prior to use of hMG, urofollitropin, or gonadorelin may improve the rates of ovulation and pregnancy. Women with PCOS are at increased risk of ovarian hyperstimulation after treatment with gonadotropins. They also experience increased rates of spontaneous abortion. An alternative therapy is ovarian drilling by laser or cautery performed at laparoscopy in women in whom hormonal therapy is not effective; however, the procedure is associated with a high incidence of ovarian adhesions.

Chronic anovulation with estrogen present also may occur with tumors of the ovary. These include granulosa-theca cell tumors, Brenner tumors, cystic teratomas, mucous cystadenomas, and Krukenberg tumors (Chap. 97). Such tumors can either secrete excess estrogen themselves or produce androgens that are aromatized in extraglandular sites. Chronic anovulation and the clinical features of PCOS result. Occasionally, areas of the ovary not involved with tumors show the characteristic histologic changes of PCOS. Other causes of chronic anovulation with estrogen present include adrenal production of excess androgen (usually adult-onset adrenal hyperplasia due to partial 21-hydroxylase deficiency) and various thyroid disorders.

CHRONIC ANOVULATION WITH ESTROGEN ABSENT Women with chronic anovulation who have low or absent estrogen production and do not experience withdrawal bleeding after progestogen treatment usually have hypogonadotropic hypogonadism due to disease of either the pituitary or the central nervous system.

Isolated hypogonadotropic hypogonadism associated with defects of smell (olfactory bulb defects) is known as the *Kallmann syndrome* (Chaps. 328 and 335). Affected women are sexually infantile and have a defect in the synthesis and/or release of GnRH. Hypothalamic lesions that impair GnRH production and cause hypogonadotropic hypogonadism include craniopharyngioma, germinoma (pinealoma), glioma,

Hand-Schuller-Christian disease, teratomas, endodermal-sinus tumors, tuberculosis, sarcoidosis, and metastatic tumors that cause suppression or destruction of the hypothalamus. Central nervous system trauma and irradiation can also cause hypothalamic amenorrhea and deficiencies in secretion of growth hormone, adrenocorticotropic hormone (ACTH), vasopressin, and thyroid hormone.

More commonly, gonadotropin deficiency leading to chronic anovulation is believed to arise from functional disorders of the hypothalamus or higher centers. A history of a stressful event in a young woman is frequent. For example, chronic anovulation can begin suddenly in a woman who leaves home for the first time or experiences the death of a loved one. Gonadotropin and estrogen levels are in the low to low-normal range as compared with normal women in the early follicular phase of the cycle. In addition, rigorous exercise, such as jogging or ballet, and diets that result in excessive weight loss may lead to chronic anovulation, particularly in girls with a history of prior menstrual irregularity. The amenorrhea in these women does not appear to be due to weight loss alone but to a combination of a decrease in body fat and chronic stress. An extreme form of weight loss with chronic anovulation occurs in anorexia nervosa. Anorexia nervosa is characterized by the development in a young woman of amenorrhea with associated severe weight loss, distorted attitudes toward eating and weight gain, and distorted body image. In anorexia nervosa amenorrhea can precede, follow, or coincide with weight loss (Chap. 78). During successful therapy, gonadotropin changes recapitulate those observed during normal puberty (Fig. 336-1).

In addition, chronic debilitating diseases such as end-stage kidney disease, malignancy, and malabsorption are believed to lead to development of hypogonadotropic hypogonadism via a hypothalamic mechanism.

Treatment of chronic anovulation due to hypothalamic disorders includes ameliorating the stressful situation, decreasing exercise, and correcting weight loss if appropriate. These women appear to be susceptible to the development of osteoporosis; estrogen replacement therapy is recommended to induce and maintain normal secondary sexual characteristics and prevent bone loss in those who do not desire pregnancy, and gonadotropin or gonadorelin therapy is indicated when pregnancy is desired (see "Treatment," below). When appropriate, therapy is directed at the primary disease of the hypothalamus.

Disorders of the pituitary can lead to the estrogen-deficient form of chronic anovulation by at least two mechanisms -- direct interference with gonadotropin secretion by lesions that either obliterate or interfere with the gonadotropic cells (chromophobe adenomas, Sheehan's syndrome) or inhibition of gonadotropin secretion in association with excess prolactin (prolactinoma). *Pituitary tumors* make up approximately 10% of all intracranial tumors and may secrete no hormone, one hormone, or more than one hormone (Chap. 328). Prolactin levels are elevated in 50 to 70% of patients with pituitary tumors, either because of prolactin secretion by the tumor itself (in the case of prolactinomas) or because the tumor mass interferes with the normal hypothalamic inhibition of prolactin secretion.

Prolactinomas can be divided into microadenomas (<10 mm in diameter) and macroadenomas (>10 mm). Prolactin excess associated with low levels of LH

and FSH constitutes a specific subtype of hypogonadotropic hypogonadism. One-tenth or more of amenorrheic women have increased levels of serum prolactin, and more than half of women with both galactorrhea and amenorrhea have elevated prolactin levels. The amenorrhea is most often associated with decreased or absent estrogen production, but prolactin-secreting tumors on occasion are associated with normal ovulatory menses or chronic anovulation with estrogen present. The increased frequency of diagnosis of prolactin-secreting adenomas is probably due to several factors, including increased awareness, improved radiographic detection methods, and availability of radioimmunoassays for prolactin. Since in older autopsy series a 9 to 23% prevalence of pituitary adenomas was observed in asymptomatic women, the clinical and prognostic significance of small microadenomas in asymptomatic individuals is unclear. However, when tumors of any size are associated with symptoms of amenorrhea or galactorrhea, particularly when visual field defects or severe headaches are present, bromocriptine therapy or neurosurgical evaluation is indicated. In the latter half of pregnancy, prolactin-secreting pituitary tumors may expand, leading to headaches, compression of the optic chiasm, bitemporal hemianopsia, and blindness. Therefore, before inducing ovulation for the purposes of achieving pregnancy, it is mandatory to exclude the presence of a pituitary tumor. *The evaluation, differential diagnosis, and management of hyperprolactinemia are described in Chap. 328.*

Large pituitary tumors such as null cell adenomas -- whether or not hyperprolactinemia is present -- are likely to be associated with deficiency of hormones in addition to gonadotropins (Chap. 328).

Craniopharyngiomas, which are thought to arise from remnants of Rathke's pouch, account for 3% of intracranial neoplasms, occur most frequently in the second decade of life, and may extend into the suprasellar region. Many of these tumors calcify and can be diagnosed by conventional skull films. Patients often present with sexual infantilism, delayed puberty, and amenorrhea due to gonadotropin deficiency; secretion of TSH, ACTH, growth hormone, and vasopressin may also be impaired.

Panhypopituitarism can occur spontaneously, be caused by mutations in transcription factors (Pit1; Prop1) involved in pituitary gland development, result from surgical or radiation treatment of pituitary adenomas, or develop after postpartum hemorrhage (Sheehan's syndrome). Patients with the latter disorder characteristically have failure to lactate or ovulate, loss of genital and axillary hair, hypothyroidism, and adrenal insufficiency (Chap. 328).

Evaluation of Amenorrhea A general scheme for the evaluation of women with amenorrhea is given in Fig. 336-8. On physical examination, attention should be given to three features: (1) the degree of maturation of the breasts, pubic and axillary hair, and external genitalia; (2) the current estrogen status; and (3) the presence or absence of a uterus. Pregnancy should be excluded in all women with amenorrhea; it is prudent to perform a suitable pregnancy screening test even when the history and physical examination are not suggestive. Once that is done, the cause of amenorrhea can frequently be diagnosed clinically. For example, Asherman's syndrome is suggested by a history of curettage in a woman who previously menstruated; in women with primary amenorrhea and sexual infantilism, the essential differential diagnosis is between gonadal dysgenesis and hypopituitarism, and the diagnosis of gonadal dysgenesis

(Turner's syndrome) or of anatomic defects of the outflow tract (mullerian agenesis, testicular feminization, and cervical stenosis) is frequently suggested on the basis of physical findings. When a specific cause is suspected, it is appropriate to proceed directly to confirm the diagnosis (obtaining a chromosomal karyotype or measurement of plasma gonadotropins). It is also useful to measure serum prolactin and FSH levels during the initial evaluation.

Estrogen status is evaluated by determining if the vaginal mucosa is moist and rugated and if the cervical mucus can be stretched and shown to fern upon drying. If these criteria are indeterminate, a progestational challenge is indicated, most often the administration of 10 mg medroxyprogesterone acetate by mouth once or twice daily for 5 days or 100 mg progesterone in oil intramuscularly. (It should be emphasized that progestogen should never be administered until pregnancy is excluded.) If estrogen levels are adequate (and the outflow tract is intact), menstrual bleeding should occur within 1 week of ending the progestogen treatment. If withdrawal bleeding occurs, the diagnosis is chronic anovulation with estrogen present, usually caused by PCOS.

If no withdrawal bleeding or only minimal vaginal spotting occurs, the nature of the subsequent workup depends on the results of the initial prolactin assay. If plasma prolactin is elevated, or if galactorrhea is present, radiography of the pituitary should be undertaken. When the plasma prolactin level is normal in an anovulatory woman with estrogen absent and with elevated FSH levels, the diagnosis is ovarian failure. If the gonadotropins are in the low or normal range, the diagnosis is either hypothalamic-pituitary disorder or an anatomic defect of the outflow tract. As indicated previously, the diagnosis of outflow tract disorder is usually suspected or established on the basis of the history and physical findings. When the physical findings are not clear-cut, it is useful to administer cyclic estrogen plus progestogen (1.25 mg oral conjugated estrogens per day for 3 weeks, with 10 mg medroxyprogesterone acetate added for the last 7 to 10 days of estrogen treatment), followed by 10 days of observation. If no bleeding occurs, the diagnosis of Asherman's syndrome or another anatomic defect of the outflow tract is confirmed by hysterosalpingography or hysteroscopy. If withdrawal bleeding occurs following the estrogen-progestogen combination, the diagnosis of chronic anovulation with estrogen absent (functional hypothalamic amenorrhea) is suggested. Radiologic evaluations of the pituitary-hypothalamic areas may be indicated in the latter cases -- irrespective of the prolactin level -- because of the danger of overlooking a pituitary-hypothalamic tumor and because the diagnosis of functional hypothalamic amenorrhea is one of exclusion (Chap. 328).

Infertility Infertility, the failure to become pregnant after 1 year of unprotected intercourse, affects approximately 10 to 15% of couples and is a common reason for seeking gynecologic assistance (Chap. 54). Male factors account for 40% of infertility problems (Chap. 335). In women, failure of ovulation accounts for 30% of cases; pelvic factors, such as tubal disease or endometriosis, account for half; and a cervical factor is implicated in about one-tenth. In 10 to 20% of infertile women no etiology is found.

Medical Aspects of Pregnancy (See also Chap. 7) The possibility of pregnancy should be considered in all women of reproductive age who are evaluated for medical illness or considered for surgery. Procedures such as x-ray exposure, drugs, and

anesthetics may be harmful to the developing fetus, and a variety of medical problems may worsen during pregnancy, including hypertension; diseases of the heart, lungs, kidney, and liver; and metabolic and endocrine disorders. Abnormal vaginal bleeding or amenorrhea during the reproductive years should prompt consideration of a complication of pregnancy, such as incomplete abortion, ectopic pregnancy, or trophoblastic disease (hydatidiform mole or choriocarcinoma). Women who present with these complications of pregnancy often have histories of abdominal pain and vaginal bleeding and may have evidence of intraabdominal hemorrhage.

Choriocarcinoma is a particular problem because of its protean manifestations. Half these malignancies follow pregnancies complicated by hydatidiform mole, and the remainder occur after spontaneous abortion, ectopic pregnancy, or normal deliveries. Patients may present with intraabdominal bleeding due to rupture of the uterus, liver, or ovary, with pulmonary manifestations (cough, hemoptysis, pleuritic pain, dyspnea, and respiratory failure) or gastrointestinal symptoms, usually chronic blood loss or melena. In addition, patients can present with cerebral metastases or renal involvement. The diagnosis can be established by demonstrating an elevated level of the b subunit of hCG in plasma. Treatment and cure are possible with chemotherapeutic agents (dactinomycin and/or methotrexate). *The manifestations of choriocarcinoma in men are discussed in Chap. 96.*

Ovarian Tumors See Chap. 97

TREATMENT

Progestogens The major use of progestogens is in conjunction with estrogen to ensure the full maturation of the endometrium, both in combination birth control pills and in the therapy of hypogonadal states. In certain circumstances, however, progestogen therapy is appropriate by itself -- to induce a progestational effect on the estrogen-primed endometrium (in diagnostic tests for the evaluation of amenorrhea), to inhibit pituitary gonadotropins for contraception (the progestogen-only birth control pill or progestogen-containing implants), for prophylaxis to prevent hyperplasia in PCOS, for palliation in cases of endometrial and breast carcinoma, and for treatment of endometriosis. Even when a direct progestational effect is desired, the available oral drugs substitute a synthetic derivative for the naturally occurring hormone. Oral progestogens include medroxyprogesterone acetate, megestrol acetate, norethindrone, norgestrel, and micronized progesterone. Parenteral agents include progesterone in oil, medroxyprogesterone acetate suspension, and 17-hydroxyprogesterone caproate. Vaginal progesterone suppositories are used for treatment of luteal-phase defects, and progestogen implants are available for long-term contraception.

The most common undesirable side effect is breakthrough bleeding, which occurs when progestogens are used continuously. Other complications include nausea, vomiting, and hirsutism. Abnormal liver function is a side effect of those derivatives with alkyl substitution in the 17a position. Synthetic progestogens are contraindicated if pregnancy is known or suspected, because of the risk of birth defects.

Estrogens Estrogens are used for the treatment of gonadal failure, the control of fertility, and the management of dysfunctional uterine bleeding. However, none of the

available oral or parenteral hormones replaces the pattern of circulating estradiol levels characteristic of the normally cycling, premenopausal woman (Fig. 336-5). Estrogens that can be given by mouth are either nonsteroidal agents (such as diethylstilbestrol) that mimic the action of estradiol, estrogen conjugates that must be hydrolyzed before they become active (usually estrone sulfate from pregnant mare's urine), or estrogen analogues that cannot be metabolized to estradiol (mestranol, quinestrol; Fig. 336-9). Even when micronized estradiol is given orally, it is rapidly converted in the body to estrone. Because oral therapy neither replaces nor mimics the daily secretory pattern of the deficient hormone, such therapy must be viewed as a pharmacologic substitution rather than a physiologic replacement. Likewise, the use of parenteral estrogens rarely mimics the physiologic situation. Parenteral preparations of conjugated estrogens, like the oral derivatives, are poor precursors of estradiol, and estradiol esters (estradiol benzoate and valerate) rarely result in plasma estradiol levels that mimic the normal monthly secretory cycle of the hormone. Transdermal estradiol results in constant levels of blood estradiol and is effective in the treatment of menopausal symptoms. Estrogen vaginal rings and creams can be used for local treatment of vaginal atrophy, but systemic absorption is variable. The side effects of estrogen substitution differ at various times of life.

Hypoestrogenism In women with decreased estrogen production, whether due to disease of the ovaries (gonadal dysgenesis) or to hypogonadotropic hypogonadism, treatment with cyclic estrogens should be instituted at the time of expected puberty to induce the development and maintenance of female secondary sexual characteristics and to prevent osteoporosis. The most commonly used medications are conjugated estrogens (0.625 to 1.25 mg/d by mouth) together with medroxyprogesterone acetate (2.5 mg/d or 5 to 10 mg during the last several days of monthly estrogen treatment to prevent development of endometrial hyperplasia). Alternatively, oral contraceptives may be given (Chap. 54). Abnormal bleeding in women receiving estrogen replacement mandates histologic evaluation of the endometrium. Such substitution therapy or the use of oral contraceptives may also be used for the purpose of suppressing pituitary gonadotropins, as in women with PCOS, in whom the major therapeutic aim is suppression of ovarian androgen production.

Temporary administration of estrogens in larger quantities (up to two times the usual adult maintenance dose) may be necessary to induce the full development of secondary sexual characteristics in girls and for the control of menopausal symptoms. Even larger doses of parenteral estrogens (10 mg estradiol valerate or 25 mg conjugated estrogen) in conjunction with progestogen may be required in some instances of dysfunctional uterine bleeding. In addition to the potential long-term side effects of all estrogens (see below), high doses may cause nausea, vomiting, and edema.

Contraceptives See Chap. 54

Estrogen Treatment of the Menopause The use of estrogens in postmenopausal women is based on evidence that they may relieve some of the complications of the postmenopausal state, including osteoporosis, and some manifestations of aging itself. In some parts of the United States, as many as half of women in the menopausal age group used one or more forms of estrogen replacement for a median period of 5 years.

The menopause is not, however, a state of simple estrogen deprivation, as some estrogens continue to be produced. It is instead a state of altered estrogen metabolism; the predominant estrogen becomes estrone, which is formed by extraglandular conversion of prehormone, rather than estradiol secreted by the ovary. As is true for all estrogen therapy, the estrogen treatment of the menopause is actually a pharmacologic substitution of one or another estrogen analogue for estradiol rather than a physiologic replacement of the missing steroid. The estrogens available for replacement therapy include conjugated estrogens, estrogen substitutes (diethylstilbestrol), synthetic estrogen (ethinyl estradiol or derivatives), micronized estradiol, estrogen-containing vaginal creams, and estrogen-containing dermal patches. Selective ER modulators (e.g., raloxifene) have estrogenic activity in the bone and the cardiovascular system but are antiestrogenic in breast and uterus. Raloxifene binds to the ER, with a conformational change in the domain of the receptor involved in transcription. Regimens associated with a low risk of complications include the following: (1) cyclic estrogen therapy in the lowest effective dose for 25 days per month or continuous estrogens given each day of the month, (2) estrogens plus the addition of progestogen during the last 10 to 14 days of estrogen therapy, (3) low-dose continuous progestogen plus estrogen given daily, and (4) daily selective ER modulators (SERM).

The most clear-cut early benefit of estrogen therapy in the menopause is the relief of vasomotor instability (hot flashes) and of atrophy of the urogenital epithelium and skin. Estrogen therapy ameliorates these symptoms in most cases. When estrogen therapy is intended to treat hot flashes alone, it should be continued for only a few years, since hot flashes tend to diminish after 3 to 4 years in untreated women. Selective ER modulators have no effect on hot flashes.

Several lines of evidence indicate that routine estrogen therapy is beneficial in preventing the complications of menopausal osteoporosis, especially in high-risk women (i.e., thin white women; Chap. 342). First, in women undergoing premature menopause, the incidence and complication rates of osteoporosis are increased, and long-term estrogen replacement appears to be beneficial. Second, estrogen therapy has short-term positive effects on calcium balance and long-term beneficial effects on bone density. Third, in women given estrogen therapy, the incidence of fractures is decreased.

Of the potential side effects, the possibility of an increased risk of endometrial carcinoma is perhaps most worrisome. The relative risk of developing endometrial adenocarcinoma in estrogen users is between six and eight times the risk in nonusers. This risk increases with increasing duration and dosage of estrogen but is largely negated in women given combination estrogen-progestogen therapy. Despite the large body of evidence linking endometrial carcinoma and estrogen use, the increased incidence primarily involves low-grade malignancies that may be difficult to distinguish histologically from hyperplasia. These forms of malignancy have little effect on life expectancy.

Apprehension concerning worsening of hypertension and thromboembolic disease appears to be due to reports of the effects of estrogen-progestogen oral contraceptives during the reproductive years and not to estrogen use in postmenopausal women. There is no conclusive evidence that low-dose estrogen therapy after menopause

increases the incidence or the severity of breast cancer or hypertension, but the risk for venous thromboembolism is slightly increased. Low-dose estrogen treatment after menopause does not appear to influence the development of atherosclerosis, myocardial infarction, or stroke. Strong evidence suggests that, in fact, estrogens may decrease the incidence of death from myocardial infarction. There is a slightly increased risk for the development of gallbladder disease with postmenopausal estrogen use.

A reasonable approach to the postmenopausal use of estrogens is as follows: (1) For long-term use, estrogens should be given in the minimal effective doses (0.625 mg conjugated estrogen orally or 1.0 mg micronized estradiol or transdermal estradiol 0.05 to 1.0 mg in a formulation that is changed every 3.5 days or every week). For women with an intact uterus, it is the practice in some clinics to give estrogens alone for 15 days and estrogen plus a daily progestogen dose for the remainder of the month. The most common regimens involve continuous estrogen plus low-dose continuous progestogen. (2) Such replacement therapy is indicated routinely in women undergoing premature menopause (surgically induced or spontaneous). (3) Estrogen therapy is also indicated routinely in women of all ages who have severe hot flashes or symptomatic atrophy of the urogenital epithelium. Hot flashes rarely persist for longer than 7 years, so, if therapy is given for this purpose, its duration can be limited. (4) In women who have had a hysterectomy, the potential benefits of treatment appear to outweigh the dangers, and in such women cyclic or continuous estrogen without progestogen is recommended. Whether estrogens should be given routinely to all women with intact uteri is an unsettled question, but the authors prescribe it in the absence of contraindications in hopes of ameliorating osteoporosis and reducing the risk of cardiovascular disease. (5) Raloxifene is given as a 60-mg tablet daily when the goal is to provide protection against bone loss without incurring additional risk of estrogen-dependent breast cancer. Each woman receiving estrogens must be monitored indefinitely at yearly intervals.

Induction of Ovulation See Chap. 54

OTHER DISORDERS OF THE FEMALE REPRODUCTIVE TRACT

VULVA

Most disorders of the vulva are due to venereal disease, most commonly syphilis (painless chancre), condylomata acuminata (venereal warts), and herpes vulvitis (painful ulcers; Chap. 132). All other lesions of the vulva, particularly in older women, must be biopsied. Early biopsy of cancer of the vulva is mandatory, because when it becomes symptomatic (pruritus and bleeding), it has often progressed to an advanced stage.

VAGINA

Infections of the vagina usually present as vaginal discharge and pruritus. The most frequent organisms are *Trichomonas*, *Candida albicans*, and *bacterial vaginalis* (Chap. 132). The diagnosis is made by microscopic examination of the discharge, and appropriate therapy can be instituted using vaginal or oral antibiotics.

Abnormalities of the vagina and cervix in female offspring of women given

diethylstilbestrol during pregnancy include adenosis of the vagina and structural abnormalities of the vagina, cervix, and uterus; the risk of developing a rare vaginal cancer (adenocarcinoma, clear cell type) is increased (2 per 10,000 exposed women). Periodic examination of women at risk should begin at age 12 to 14, and reexamination should be done after any episode of abnormal bleeding.

CERVIX

Preinvasive lesions of the cervix (also known as *cervical intraepithelial neoplasia*) and invasive carcinoma of the cervix can be detected reliably by obtaining a Papanicolaou (Pap) smear.

Evaluation of the Pap Smear The incidence of invasive cervical cancer has declined as a result of Pap smear screening. In the United States, approximately 2 to 3 million abnormal Pap smears are found each year. Most represent low-grade lesions but require appropriate follow-up. The follow-up of abnormal Pap smears requires an understanding of the Bethesda system for evaluating such smears (see below) and of the limitations of cytologic screening systems. Further evaluation may require repeat cytologic examination, colposcopy, or both.

Current Screening Recommendations Risk factors for cervical neoplasia include a history of multiple sexual partners, coitus beginning at an early age, a history of infection with human papilloma virus (HPV), infection with HIV or another immunosuppressed state, and a history of cancer of the lower genital tract. Cervical cancer screening is recommended annually beginning at 18 years of age or when the woman becomes sexually active, if earlier than age 18. "Less frequent" screening is performed when three consecutive, negative, satisfactory annual Pap smears have been obtained or if the woman is in a low-risk category. There is no upper age limit for screening, because the prevalence of invasive cancer shows a linear increase with age, most of these cancers being diagnosed after age 50. Even after hysterectomy, annual screening should be performed if there is a history of abnormal Pap smears or other lower genital tract neoplasia.

The Bethesda System of Cytologic Examination Pap smears are evaluated in regard to the adequacy of the specimen (satisfactory for evaluation, satisfactory but limited, or unsatisfactory for evaluation because of a stated reason), the general diagnosis (normal or abnormal), and a descriptive diagnosis if the smear is abnormal. The descriptive diagnoses include benign cellular changes, reactive cellular changes, and epithelial cell abnormalities, the latter including (1) atypical squamous cells of undetermined significance (ASCUS); (2) low-grade squamous intraepithelial lesion (LSIL), which is further categorized to include HPV infection, cervical intraepithelial neoplasia (CIN 1), and high-grade squamous intraepithelial lesion (HSIL, which is itself subdivided into CIN 2 and CIN 3); and (3) squamous cell carcinoma.

Guidelines for the Management of Women with Abnormal Pap Smears For ASCUS smears that are unqualified or suggest a reactive process, a repeat smear should be obtained every 4 to 6 months for 2 years until three consecutive negative smears have been obtained. For ASCUS smears that are unqualified but have severe inflammation, any specific cause should be treated, and the smear should be repeated in 2 to 3

months; because invasive carcinoma can be obscured by severe inflammation, clinical evaluation is mandatory. For postmenopausal women not using hormone replacement, a course of topical estrogen should be given before the test is repeated. For LSIL smears, the Pap test is repeated every 4 to 6 months for 2 years until three consecutive negative smears have been obtained; treatment of HPV is of no established benefit, and there is a high rate of regression of LSIL, so that in compliant, low-risk individuals, the outcome is usually favorable. If LSIL is persistent, colposcopy with directed biopsy is performed, and endocervical curettage is undertaken if a specific diagnosis is made by biopsy. Cervical cone biopsy or loop electrosurgical excision procedures are performed for higher-grade lesions such as HSIL. If cervical cancer is diagnosed by biopsy, clinical staging is performed, and the patient is treated with radiation therapy or surgery.

UTERUS

Only 40% of cases of endometrial adenocarcinoma are detected by Pap smear. In women at high risk for endometrial carcinoma (because of obesity, a history of chronic anovulatory cycles, diabetes mellitus, hypertension, or estrogen treatment), yearly endometrial sampling should be performed. Measurement of endometrial thickness by sonography can indicate which patients are at risk for endometrial pathology. Endometrial thickness <5 mm is rarely associated with either hyperplasia or cancer. Low-dose oral estrogen therapy rarely causes breakthrough or withdrawal bleeding in postmenopausal women. Therefore, irrespective of whether the patient is using estrogen therapy, the occurrence of postmenopausal bleeding makes it mandatory to obtain a tissue diagnosis by either endometrial sampling or curettage to exclude endometrial cancer.

One of the most common disorders of the uterus and the most frequent tumor of women (one of four women affected) is the uterine leiomyoma, or fibroid tumor. Three-fourths of women with leiomyoma are asymptomatic, and the diagnosis is made on routine pelvic examination. When the tumor is associated with excessive menstrual blood loss, is large or fast-growing, or causes significant pelvic pain (Chap. 52), the preferred treatment is hysterectomy if there is no desire for further childbearing. In young women, myomectomy is sometimes indicated when infertility or repeated fetal wastage is a manifestation or where future childbearing is desired.

FALLOPIAN TUBES AND OVARIES

Infectious pelvic inflammatory disease is a common disorder of the fallopian tubes and usually becomes symptomatic after a menstrual period; the symptoms include fever, chills, abdominal pain, and vaginal discharge, and pelvic tenderness on physical examination is common. The initiating organism most often is *Chlamydia trachomatis* or *Neisseria gonorrhoeae*, but tuboovarian abscess and sterility are probably caused by mixed aerobic and anaerobic superinfections and require wide-spectrum antibiotic treatment (Chap. 133).

Endometriosis is a benign disorder characterized by the presence and proliferation of endometrial tissue (stroma and glands) outside the endometrial cavity. The clinical manifestations are variable. Endometriosis occurs most commonly between the ages of 30 and 40 and is found incidentally at the time of surgery in approximately one-fifth of all

gynecologic operations. The fertility rate is reduced in affected women. The disorder usually involves the posterior cul-de-sac or the ovaries and can give rise to ovarian enlargement (endometriomas), although it may involve distant sites (lung, umbilicus). The major symptom is pelvic pain, characteristically dysmenorrhea (Chap. 52). However, the frequency and severity of symptoms correlate poorly with the extent of disease. Other manifestations include dyspareunia, pain with defecation, and infertility. The characteristic physical findings are multiple tender nodules palpable along the uterosacral ligament at the time of rectal-vaginal examination, a posteriorly fixed uterus, or enlarged, cystic ovaries. The diagnosis can only be confirmed by direct visualization, usually at diagnostic laparoscopy. Treatment depends on the degree of involvement and the desires of the patient and includes observation for mild disease with no associated infertility or pain, hormonal suppressive therapy, conservative surgery by laparoscopy or laparotomy if fertility is desired, or removal of the uterus, tubes, and ovaries in severe disease. Endometriosis is rare after the menopause.

Any adnexal mass that persists for more than 6 weeks or is larger than 6 cm must be evaluated. Although ovarian cysts and neoplasms are the most common pelvic adnexal masses (see above), tumors of the fallopian tubes, uterus, gastrointestinal tract, or urinary tract should also be considered. Sonography or radiographic evaluation is often helpful in identifying the nature of the adnexal mass prior to surgical exploration.

(Bibliography omitted in Palm version)

337. ENDOCRINE DISORDERS OF THE BREAST - *Jean D. Wilson*

The breasts are the site of fatal and preventable cancer in women and provide clues to underlying systemic illness in both men and women. Consequently, examination of the breasts is an important part of the physical examination. It is the duty of every physician to distinguish the abnormal from the normal at the earliest possible stage and to seek referral if there is any doubt. *For discussion of cancer of the breast, see Chap. 89.*

ENDOCRINE CONTROL OF THE BREAST

There is no histologic or functional difference in the breasts of prepubertal boys and girls, but a profound sexual dimorphism in breast development ensues at the time of puberty. The endocrine control of female breast development is illustrated inFig. 337-1. Growth of the female breast at puberty is mediated primarily by estradiol, which induces the enlargement, division, and elongation of the tubular duct system and maturation of the nipples. Administration of estrogen to men is equally effective in this regard. To produce true alveolar development at the ends of the ducts, however, the synergistic action of progesterone is required. Within the gland a variety of mediators influence epithelial cell division and differentiation, including stimulatory factors such as the insulin-like growth factors, transforming growth factor a, and epidermal growth factor and inhibitory factors such as transforming growth factor b. Once the anatomic development of the ducts and alveoli is complete, the continued action of estrogen and progesterone is not required for lactation itself.

The endocrine control of milk formation is complex, requiring, in addition to appropriate priming by estrogen and progesterone, lactogenic hormones and the permissive action of glucocorticoid, insulin, thyroxine, and, in some species, growth hormone. There are two lactogenic hormones: (1) human placental lactogen (hPL, or chorionic somatomammotropin) and (2) prolactin. hPL is secreted in large amounts by the placenta during the latter part of gestation and prepares the breast for milk production. It disappears from the maternal (and fetal) circulation shortly after termination of pregnancy. The secretion of pituitary prolactin (Chap. 328) rises during pregnancy and plays the critical role in the initiation and maintenance of lactation in the puerperium. During late pregnancy and lactation, 60 to 80% of the anterior pituitary may consist of prolactin-secreting lactotrope cells, reflecting the stimulatory effects of estrogen on these cells.

Unlike most pituitary hormones, prolactin secretion is controlled predominantly by tonic inhibition. Under basal conditions inhibitory hypothalamic hormones, the most important being dopamine, are delivered from the central nervous system to the pituitary via the hypothalamic portal system and inhibit the release of prolactin into the blood (Chap. 328). Most factors that influence prolactin secretion do so by affecting the synthesis or release of dopamine. Basal prolactin levels in the mother fall after delivery, but prolactin secretion is enhanced by stimulation of the breasts, such as the act of nursing (the so-called sucking reflex), a phenomenon that is probably mediated by the reflex release of oxytocin, which acts as a prolactin-releasing factor. Prolactin binds to specific receptors on the cell surface of the breast acinar cells and activates the JAK-STAT signal transduction cascade to stimulate the synthesis ofb-casein, whey acidic protein, and other milk constituents. In the postgestational state, the normal lactating woman

forms about a liter of milk per day containing 38 g fat, 70 g lactose, and 12 g protein. Lactation can be suppressed by the administration of estrogens or diethylstilbestrol, which inhibit milk production by direct effects on the breast, or dopamine agonists such as bromocriptine, which inhibit prolactin secretion by the pituitary. Alternatively, if a woman does not nurse or use breast pumps postpartum, lactation usually ceases in 1 to 2 weeks.

GALACTORRHEA

Galactorrhea refers to the nonpuerperal discharge of milk-containing fluid from the breast. The definition of exactly what constitutes galactorrhea is not always clearly defined in the literature. According to the studies of Friedman and Goldfein, breast secretions are absent in normal, regularly menstruating nulligravid women. However, breast secretions can be demonstrated in a fourth of normal women who have been pregnant in the past; thus breast secretions in small amounts may be of no clinical significance in these instances. Spontaneous leakage of milk from the breasts is usually of more significance than milk that must be expressed. When the secretion is milky or white, it is safe to assume that it contains fat, casein, and lactose and is in fact milk; the concentration of milk constituents may increase after repeated sampling. When the secretion is brown or greenish, it rarely contains normal milk constituents and consequently may not result from an underlying endocrinopathy. Bloody discharges may be due to neoplasms of the breast. With these issues in mind, galactorrhea can be defined as inappropriate production of milk that is persistent or worrisome to the patient, recognizing that in some instances no underlying pathology may be demonstrated.

Since the action of a lactogenic hormone is necessary for the initiation of milk production, it is logical to consider galactorrhea as a consequence of deranged prolactin physiology. However, as indicated above, a complex hormonal milieu is necessary for lactation. Milk production does not take place in many instances in which prolactin is elevated, both in men and in women who have not been exposed to the necessary hormonal environment. As a consequence, hyperprolactinemia is more common than galactorrhea. Furthermore, while enhanced prolactin secretion is necessary for the initiation of lactation, continued production can be maintained in the presence of minimally or intermittently elevated prolactin levels so that basal plasma prolactin levels are not always elevated in patients with galactorrhea. In some such women prolactin levels may be elevated during sleep or with stimulation of the nipple; in others, hyperprolactinemia may have been present transiently. Perhaps the strongest evidence for a critical role for prolactin in galactorrhea is the fact that administration of dopaminergic agents that suppress plasma prolactin levels corrects galactorrhea even when the basal plasma prolactin levels are normal.

Differential Diagnosis Galactorrhea can be due to failure of the normal hypothalamic inhibition of prolactin release, to increased prolactin-releasing factor(s), or to autonomous prolactin secretion by tumors (Table 337-1). Pituitary stalk section, whether traumatic or secondary to the mass effects of sellar tumors, results in increases in prolactin secretion due to interruption in the delivery of dopamine to the pituitary. Likewise, many drugs that influence the central nervous system (CNS) (including virtually all psychotropic agents, methyldopa, reserpine, and antiemetics) enhance prolactin release, presumably by inhibiting synthesis, release, or action of dopamine.

Estrogens increase prolactin secretion, but estrogen withdrawal (as in the discontinuation of oral contraceptives) may also trigger the onset of galactorrhea. CNS diseases outside the pituitary can cause galactorrhea presumably by interfering with the production or delivery of dopamine to the pituitary (CNS sarcoidosis, craniopharyngioma, pinealoma, encephalitis, meningitis, hydrocephalus, hypothalamic tumors).

In primary hypothyroidism, galactorrhea results from the enhanced production of thyrotropin-releasing hormone (TRH), which also stimulates prolactin release; thyroid hormone replacement corrects the galactorrhea. A similar mechanism, involving enhanced secretion of oxytocin, may cause the galactorrhea that follows breast surgery or breast trauma.

Enhanced prolactin release can also occur from pituitary or nonpituitary tumors. Three types of pituitary tumors ([Chap. 328](#)) can cause galactorrhea: (1) pure prolactin-secreting micro- or macroadenomas, (2) mixed tumors that secrete both growth hormone and prolactin and cause acromegaly with galactorrhea, and (3) large null cell adenomas. The latter may interfere with the delivery of dopamine to the pituitary, either by mass effects on the hypothalamus or by compressing the pituitary stalk. Excess growth hormone secretion, in the absence of hyperprolactinemia, on occasion causes galactorrhea. Rarely, prolactin is secreted by bronchogenic carcinomas, and hydatidiform moles and choriocarcinomas may secrete placental lactogen.

In series involving several hundred patients with galactorrhea, a pituitary tumor was identified in about one-fourth, other known causes were identified in another fourth or fifth, and the remaining half fell into the idiopathic category. Many of the latter group ultimately developed prolactin-secreting pituitary tumors, some probably had subtle disorders of hypothalamic function, and in others a drug-related cause may have been missed. The fact remains that no satisfactory diagnosis is reached in many patients. When menses are normal, the likelihood of establishing a cause for galactorrhea is poor.

Galactorrhea is unusual in men, even in the presence of profound elevations of plasma prolactin; when it does occur, it is usually upon the background of a feminizing state (see below).

Diagnostic Evaluation If hyperprolactinemia is present, the evaluation is fundamentally that of a pituitary tumor once drug causes and hypothyroidism are excluded ([Chap. 328](#)). Even when a specific cause cannot be identified and the diagnosis of idiopathic galactorrhea is made by exclusion, it is necessary to remember that pituitary tumors may subsequently become manifest. The higher the prolactin values and the more persistent the galactorrhea, the greater is the likelihood of such a development.

TREATMENT

Breast binders can be effective in patients with mild galactorrhea of unknown etiology, presumably by preventing stimulation of the nipple and the consequent perpetuation of lactation. The aim of treatment in other instances is to correct the elevated prolactin

level, and treatment of a pituitary tumor, cessation of causative drugs, or correction of hypothyroidism is often followed by the disappearance of galactorrhea. Dopamine agonists that suppress plasma prolactin have been used to treat idiopathic hyperprolactinemia, prolactin-secreting tumors of the pituitary (Chap. 328), and even normoprolactinemic galactorrhea. These drugs suppress lactation and may cause resumption of menstrual cycles (and even fertility) in women with amenorrhea and galactorrhea.

GYNECOMASTIA

Enlargement of the male breast, or *gynecomastia*, can be a normal physiologic phenomenon at certain times of life or the result of several pathologic states (Table 337-2). A central issue in the evaluation of breast tissue in adult men is the separation of the normal from the abnormal, as gynecomastia can be an important indicator of underlying disease. It is sometimes difficult to distinguish true breast tissue enlargement from adipose tissue (*lipomastia*). True glandular tissue is often palpable, especially around the areolae, as it is firmer and contains cordlike features that are distinct from the texture of adipose tissue. In difficult cases, true gynecomastia can be identified by mammography or ultrasonography. In this discussion, we shall assume that any palpable breast tissue in men (except for the three physiologic states see below) can be due to an underlying endocrinopathy and deserves, at a minimum, a limited evaluation.

Early gynecomastia is characterized by proliferation in the breast of both the fibroblastic stroma and the duct system, which elongates, buds, and duplicates. As gynecomastia persists, progressive fibrosis and hyalinization are associated with regression of epithelial proliferation and, eventually, a decrease in the number of ducts. When the cause of the gynecomastia is corrected early in the course, resolution occurs by reduction in size and epithelial content with gradual disappearance of the ducts, leaving hyaline bands that eventually disappear.

Growth of the breast in men, as in women, is mediated by estrogen and results from an decrease in the ratio of active androgen to estrogen. As described inChap. 335, estradiol formation in normal men occurs principally by the conversion of circulating androgen to estrogen in extraglandular tissues; the normal ratio of the two hormones in plasma is about 300:1. Growth of the breast ensues in men when the normal ratio decreases as the result of diminished testosterone production or action, enhanced estrogen formation, or both processes occurring simultaneously.

Physiologic Gynecomastia In the *newborn* transient enlargement of the breast is due to the action of maternal and/or placental estrogens. The enlargement usually disappears in a few weeks but may persist longer. *Adolescent* gynecomastia is common at some time during puberty. The median age of onset is 14; it is often asymmetric or transiently unilateral, frequently tender, and it regresses so that by age 20 only a small number of men have palpable vestiges of glandular tissue in one or both breasts. Although the origin of the excess estrogen has not been identified, the onset of gynecomastia correlates with the increase in adrenal androgens at adrenarche. In addition, the luteinizing hormone (LH) stimulation of androgen synthesis by the Leydig cell in early puberty may be associated with transient elevations of plasma estradiol so that the ratio of potent androgen to estrogen is low prior to the completion of puberty.

Gynecomastia of aging also occurs in 40% or more of otherwise healthy elderly men. A likely explanation is the increase with age in the conversion of androgens to estrogens in extraglandular tissues. Abnormal liver function or drug therapy may be contributing causes in such men.

Pathologic Gynecomastia Pathologic gynecomastia can result from one of three basic mechanisms: deficiency in testosterone production or action (with or without a secondary increase in estrogen production), increase in estrogen production, or drugs (Table 337-2). Most of the disorders that cause primary and secondary testicular failure are discussed in Chap. 335. The fact that deficient testosterone production can cause gynecomastia is illustrated by the syndrome of congenital anorchia in which normal (or slightly low) estradiol levels, in the presence of profoundly decreased testosterone production, results in florid gynecomastia. Decreased testosterone production is also responsible for gynecomastia in some men with Klinefelter syndrome or testicular failure of other causes. In the disorders of androgen resistance, such as testicular feminization, deficient androgen action and increased testicular estrogen production are both present.

A primary increase in estrogen production can result from a variety of causes. Increased secretion of testicular estrogen may result from elevations in plasma gonadotropins, as in cases of aberrant production of chorionic gonadotropin by testicular tumors or by bronchogenic carcinomas, from the ovarian elements in the gonads of men with true hermaphroditism, or as the result of formation by testicular tumors (particularly Leydig and Sertoli cell tumors). Increased conversion of androgen to estrogens in extraglandular tissues can be due either to increased availability of substrate (androstenedione) for extraglandular estrogen formation (congenital adrenal hyperplasia, hyperthyroidism, and most feminizing adrenal tumors), or to diminished catabolism of androstenedione (liver disease) so that estrogen precursors are shunted to aromatase in peripheral sites. Extraglandular aromatase can be increased in tumors of the liver or adrenal gland and rarely as a result of an inherited disorder manifested by gynecomastia in affected males and macromastia in females.

Drugs can cause gynecomastia by several mechanisms. Many drugs either act directly as estrogens or cause an increase in plasma estrogen activity (e.g., in men receiving diethylstilbestrol for prostatic carcinoma and in transsexuals in preparation for sex-change operations). Boys and young men are particularly sensitive to estrogen and can develop gynecomastia after the use of dermal ointments containing estrogen or after the ingestion of milk or meat from estrogen-treated animals. The gynecomastia of digitalis ingestion is usually attributed to an estrogen-like side effect of the drug, but it occurs most commonly in men with abnormal liver function. A second mechanism of drug-induced gynecomastia is illustrated by clomiphene and human chorionic gonadotropin (hCG), which cause enhanced testicular secretion of estrogen. Other drugs cause gynecomastia by interfering with testosterone synthesis (ketoconazole and alkylating agents) and/or testosterone action, for instance, by blocking the binding of androgen to its receptor protein in target tissues (spironolactone and cimetidine). Finally, drugs that cause gynecomastia by mechanisms that have not been defined include busulfan, ethionamide, isoniazid, methyldopa, tricyclic antidepressants, penicillamine, omeprazole, calcium channel blocking agents, angiotensin-converting enzyme inhibitors, metoclopramide, antiretroviral agents, diazepam, marijuana, and heroin. In some instances, the feminization is due to effects of drugs on liver function.

Treatment with growth hormone can cause gynecomastia even in prepubertal boys, suggesting that growth hormone itself or one of the insulin-like growth factors has a direct effect on the breast.

Diagnostic Evaluation The evaluation of patients with gynecomastia should include: (1) a careful drug history; (2) measurement and examination of the testes (if both are small, a chromosomal karyotype should be obtained; if the testes are asymmetric, a workup for testicular tumor should be instituted); (3) evaluation of liver function; and (4) endocrine workup to include measurement of serum androstenedione or 24-h urinary 17-ketosteroids (usually elevated in feminizing adrenal states), measurement of plasma estradiol and hCG (helpful if elevated but usually normal), and measurement of plasma LH and testosterone. If LH is high and testosterone is low, the diagnosis is usually testicular failure; if LH and testosterone are both low, the diagnosis is most likely increased primary estrogen production (e.g., a Sertoli cell tumor of the testis), provided hypogonadotropic hypogonadism has been excluded; and if both LH and testosterone are elevated, the diagnosis is either an androgen-resistance state or a gonadotropin-secreting tumor.

A satisfactory diagnosis can be made in only half or fewer of patients referred for gynecomastia. This implies either that the diagnostic techniques are not sufficiently refined to recognize mild disturbances, that many causes of gynecomastia are as yet undefined, that the causes may be transient and difficult to diagnose, or that gynecomastia may in some instances be normal rather than due to a pathologic state. Because of the problem of separating the normal from the abnormal, gynecomastia should probably be worked up routinely only if the drug history is negative, the breast is tender (indicating rapid growth), or the breast mass is >4 cm in diameter. However, the decision to perform an endocrine evaluation depends on the clinical context. For example, gynecomastia associated with signs of underandrogenization should always be evaluated. A firm or hard breast mass should raise suspicion of male breast cancer (Chap. 89).

TREATMENT

When the primary cause can be identified and corrected, the breast enlargement usually subsides promptly and eventually disappears. For example, androgen replacement therapy may produce dramatic improvement in men with testicular insufficiency. However, if the gynecomastia is of long duration (and fibrosis has replaced the original ductal hyperplasia), correction of the primary defect may not be followed by resolution. In such instances and when the primary cause cannot be corrected, surgery is the only effective therapy. Indications for surgery include several psychological and/or cosmetic problems, continued growth or tenderness, or suspected malignancy. Although the relative risk of carcinoma of the breast is increased in men with gynecomastia, it is rare nevertheless. Prophylactic radiation of the breasts prior to the institution of diethylstilbestrol or estrogen therapy is effective in preventing gynecomastia and has a low complication rate. In patients who have painful gynecomastia and who are not candidates for other therapy, treatment with antiestrogens such as tamoxifen may be indicated.

(Bibliography omitted in Palm version)

338. DISORDERS OF SEXUAL DIFFERENTIATION - *Jean D. Wilson, James E. Griffin*

Sexual differentiation is a sequential and ordered process. *Chromosomal sex*, established at the moment of fertilization, determines *gonadal sex*, and gonadal sex, in turn, causes the development of *phenotypic sex*, in which the male or female urogenital tract is formed (Fig. 338-1). A disturbance of any step in this process during embryogenesis may impair sexual differentiation (Table 338-1). Known causes of such disorders include environmental insults as in the ingestion of a virilizing drug during pregnancy, nonfamilial aberrations of the sex chromosomes as in 45,X gonadal dysgenesis, birth defects of multifactorial etiology as in most cases of hypospadias, and disorders due to single gene mutations as in the testicular feminization syndrome.

Limitations of knowledge make it necessary to make empirical assignments as to the type of derangement in some disorders, but specific diagnoses can usually be made as the result of genetic, endocrine, phenotypic, and chromosomal assessment. As a consequence, gender assignment in the newborn is usually appropriate, even in extreme instances of ambiguous genitalia.

NORMAL SEXUAL DIFFERENTIATION

The first event in sexual differentiation is the establishment of chromosomal sex, the heterogametic sex (XY) being male and the homogametic sex (XX) female. The embryos of both sexes then develop in an identical fashion until approximately 40 days of gestation.

The second phase of sexual differentiation is the conversion of the indifferent gonad into a testis or an ovary. No matter how many X chromosomes are present (as in 47,XXY, 48,XXXY, etc.), a testis will develop as long as a normal Y chromosome is present. Differentiation of the indifferent gonad into a testis is initiated by the actions of a single gene on the short arm of the Y chromosome (*SRY*); expression of an *SRY* transgene into female mice causes them to develop as males. The gene encodes a DNA-binding protein, but the mechanism by which *SRY* promotes testicular development remains poorly defined. At least four additional genes are also necessary for normal testicular development: (1) the Wilms' tumor-related gene (*WT1*), (2) steroidogenic factor 1 (*SF1*), (3) *SRY*-related HMG-box 9 (*SOX9*), and (4) *d*osage-sensitive sex reversal-*a*drenal hypoplasia congenita critical region on the *X* chromosome gene *1* (*DAX1*). These genes each encode putative transcription factors that regulate the expression of genes necessary for gonadal survival; mutations in *SRY* or any of the four downstream genes impair testicular development. It remains unclear if there are analogous "ovarian-determining genes," or if ovarian development is a default pathway in the absence of testicular determination genes. Mutations in some of the genes noted above (e.g., *SF1*) also prevent normal ovarian development.

The final process in sexual differentiation, the translation of gonadal sex into phenotypic sex (formation of the male or female urogenital tracts), is the consequence of the type of gonad formed and the endocrine secretions of the fetal gonads. The internal urogenital tract is derived from the wolffian and mullerian ducts that exist side by side in early embryos of both sexes (Fig. 338-1*A*). In the male the wolffian ducts give rise to the

epididymides, vasa deferentia, and seminal vesicles, and the mullerian ducts disappear. In the female the mullerian ducts are converted into the fallopian tubes, uterus, and upper vagina, and the wolffian ducts regress. The external genitalia and urethra in the two sexes develop from common anlage -- the urogenital sinus and the genital tubercle, folds, and swellings (Fig. 338-1*B*). The urogenital sinus gives rise to the prostate and prostatic urethra in the male and to the urethra and lower portion of the vagina in the female. The genital tubercle gives rise to the glans penis in the male and the clitoris in the female. The urogenital swellings become the scrotum or the labia majora, and the urethral folds form the labia minora or fuse to form the shaft of the penis and the male urethra.

In the absence of the testes, as in the normal female or in the male embryo castrated prior to the onset of gonadal differentiation, phenotypic sex develops along female lines. Thus, masculinization of the fetus is induced by hormones from the fetal testes, whereas female development does not require the presence of the ovary. Phenotypic sex normally conforms to chromosomal sex. That is, chromosomal sex determines gonadal sex, and gonadal sex, in turn, controls phenotypic sex.

Formation of the male phenotype is vested in the action of three hormones. Two -- antimullerian hormone (AMH) and testosterone -- are secreted by the fetal testis. AMH [also termed *mullerian-inhibiting substance* (MIS)] is a protein that suppresses the mullerian ducts and prevents development of the uterus and fallopian tubes in the male. Testosterone acts directly to virilize the wolffian duct and is the precursor for the third embryonic male hormone, dihydrotestosterone (Chap. 335), which promotes development of the male urethra and prostate and formation of the penis and scrotum. Testosterone and dihydrotestosterone induce formation of the male urogenital tract during fetal life by acting through the same nuclear androgen receptor by which they act in postembryonic life (Chap. 335).

The secretion of testosterone by the fetal testes approaches a maximum by the tenth week of gestation, and formation of the sexual phenotypes is largely completed by the end of the first trimester. During the latter phases of gestation, the ovarian follicles develop and the vagina matures in the female, and testicular descent and phallic growth take place in the male.

DISORDERS OF CHROMOSOMAL SEX

Disorders of chromosomal sex (Table 338-2) occur when the number or structure of the X or Y chromosomes is abnormal (Chap. 66).

KLINEFELTER SYNDROME

Clinical Features Klinefelter syndrome is characterized by small, firm testes, azoospermia, gynecomastia, and elevated levels of plasma gonadotropins in men with two or more X chromosomes. The common karyotype is either a 47,XXY chromosomal pattern (the classic form) or 46,XY/47,XXY mosaicism. It is the most frequent major abnormality of sexual differentiation, the incidence being around 1 in 500 men.

Prepubertally, the testes are small but otherwise appear normal. After puberty, the

disorder is manifest as infertility, gynecomastia, or occasionally underandrogenization (Table 338-3). Hyalinization of the seminiferous tubules and azoospermia are consistent features of the 47,XXY variety. The small, firm testes are usually <2 cm and always<3.5 cm in length (corresponding to 2- and 12-mL volume, respectively). Longer legs cause increased mean height. Gynecomastia is common and ordinarily develops during adolescence, is generally bilateral and painless, and may become disfiguring (Chap. 337). Obesity and varicose veins occur in one-third to one-half, and leg ulcers are associated with deficiency of plasminogen activator inhibitor-1. The diagnosis is made most frequently in boys with developmental delay and/or learning disabilities and social maladjustment; abnormalities of thyroid function, diabetes mellitus, and pulmonary disease are also common. The risk of breast cancer is 20 times that of normal men (but only about a fifth that in women). Most have male psychosexual orientation and function sexually as normal men.

About 10% of the patients have the mosaic form, as estimated by chromosomal karyotypes on peripheral blood leukocytes. The frequency of this variant may be underestimated, since chromosomal mosaicism can be present in the testes when the peripheral leukocyte karyotype is normal. The mosaic form is usually not as severe as the 47,XXY variety, and the testes may be normal (Table 338-3). The endocrine abnormalities are less severe, and gynecomastia and azoospermia are less common, and occasional mosaic individuals are fertile. In some the diagnosis may not be suspected because the manifestations are so mild.

Approximately 30 additional variants of Klinefelter syndrome have been described, including those with uniform cell lines (such as 48,XXYY, 48,XXXY, and 49,XXXXY) and a variety of mosaicisms of the X chromosome with or without associated structural abnormalities of the X. In general, the greater the chromosomal abnormality the more severe the manifestations.

Pathophysiology The classic form is due to meiotic nondisjunction of the chromosomes during gametogenesis (Fig. 338-2). About 40% of the responsible meiotic nondisjunctions occur during spermatogenesis, and 60% occur during oogenesis. Advanced maternal age is a predisposing factor. The mosaic form is thought to result from chromosomal mitotic nondisjunction after fertilization of the zygote and can take place in either a 46,XY zygote (Fig. 338-2) or a 47,XXY zygote. The latter situation, double nondisjunction (meiotic and mitotic), may be the usual cause and thus explain why the mosaic form is less frequent than the classic disorder.

Levels of plasma follicle stimulating hormone (FSH) and luteinizing hormone (LH) are usually high; FSH shows the best discrimination because of the consistent damage to the seminiferous tubules. The plasma testosterone level averages half normal but may overlap the normal range. Mean plasma estradiol levels are elevated; early, estradiol secretion by the testes may increase in response to the elevated plasma LH level, but the testicular secretion of estradiol and testosterone eventually declines. Elevated plasma estradiol late in the course is probably due both to decreased metabolic clearance and increased conversion of testosterone to estradiol in extragonadal tissues. The net result both early and late is a variable degree of feminization and virilization. Feminization, including gynecomastia, depends on the ratio of circulating estrogen to androgen (relative or absolute), and individuals with low plasma testosterone and high

plasma estradiol levels are more likely to develop gynecomastia (Chap. 337). Men with untreated Klinefelter syndrome may have enlarged pituitary glands, presumably due to hyperplasia of the gonadotrophs due to inadequate testosterone feedback.

TREATMENT

In men with some sperm production, or in whom spermatids can be recovered from testicular biopsy, fertility is possible with in vitro fertilization (Chap. 335). Gynecomastia should be treated surgically. Some underandrogenized men benefit from supplemental androgen, particularly in those with decreased bone density, but such treatment may worsen the gynecomastia, presumably by providing increased substrate for estrogen formation in peripheral tissues. Testosterone should be injected in the form of testosterone cypionate or testosterone enanthate or administered via the transdermal route (Chap. 335). Following the administration of testosterone, the plasmaLH level returns to normal only after several months, if at all.

XX MALE SYNDROME

A 46,XX karyotype is found in approximately 1 in 20,000 phenotypic males. The findings resemble those in Klinefelter syndrome: the testes are small and firm (generally<2 cm), gynecomastia is frequent, the penis is normal to small in size, azoospermia and hyalinization of the seminiferous tubules are usual, no female urogenital structures are present, psychosexual identification is male, mean plasma testosterone level is low, plasma estradiol level is elevated, and plasma gonadotropin levels are high. Affected individuals differ from typical Klinefelter patients only in that average height is less than in normal men, the incidence of cognitive problems is not increased, and the incidence of hypospadias is increased.

The majority of XX males have Y-related DNA (e.g., detected by polymerase chain reaction of the *SRY* gene); thus an X-Y or Y-autosome translocation appears to be the common cause. Some 46,XX males are negative for all Y-specific DNA, suggesting that their disorder is due to mutation in a downstream, autosomal or X-linked gene involved in development of the testes. The management is similar to that of Klinefelter syndrome.

GONADAL DYSGENESIS (TURNER SYNDROME)

Clinical Features Gonadal dysgenesis is characterized by primary amenorrhea, sexual infantilism, short stature, multiple congenital anomalies, and bilateral streak gonads in phenotypic women with any of several defects of the X chromosome. This condition should be distinguished from (1) mixed gonadal dysgenesis in which a unilateral testis and a contralateral streak gonad may be present; (2) pure gonadal dysgenesis in which bilateral streak gonads are associated with a normal 46,XX or 46,XY karyotype, normal stature, and primary amenorrhea; and (3) the Noonan syndrome, an autosomal dominant disorder in both sexes characterized by webbed neck, short stature, congenital heart disease, cubitus valgus, and other congenital defects despite normal karyotypes and normal gonads.

The incidence is estimated at 1 in 3000 newborn females; the prenatal incidence may be as high as 2% of all human conceptuses, only a small fraction of whom survive to

term. The diagnosis is made either at birth because of the associated anomalies or at puberty when amenorrhea and failure of sexual development are noted in conjunction with the associated anomalies. Gonadal dysgenesis is the most common cause of primary amenorrhea, accounting for a third of such patients. The external genitalia are unambiguously female but remain immature, and there is no breast development unless exogenous estrogen is given. The fallopian tubes and uterus are immature, and bilateral streak gonads are present in the broad ligaments. Primordial germ cells are present transiently during embryogenesis but disappear because of an accelerated rate of atresia (Chap. 336). After the age of expected puberty, these streaks lack identifiable follicles and ova and consist of fibrous tissue that is indistinguishable from normal ovarian stroma.

The somatic abnormalities primarily involve the skeleton and connective tissue. Lymphedema of the hands and feet, webbing of the neck, low hairline, redundant skin folds on the back of the neck, a shieldlike chest with widely spaced nipples, and growth retardation are features that suggest the diagnosis in infancy. Micrognathia, epicanthal folds, prominent low-set or deformed ears, a fishlike mouth, and ptosis may be present. Short fourth metacarpals are present in half, and 10 to 20% have coarctation of the aorta. In adults, the average height rarely exceeds 150 cm. Associated conditions include renal malformations, pigmented nevi, hypoplastic nails, tendency to keloid formation, perceptive hearing loss, unexplained hypertension, glucose intolerance, and autoimmune thyroid disease.

Pathophysiology About half have a 45,X karyotype, approximately one-fourth have mosaicism with no structural abnormality (46,XX/45,X), and the remainder have a structurally abnormal X chromosome with or without mosaicism (Chap. 66). The mechanism of chromosome loss is unknown and may occur during gametogenesis in either parent or as a mitotic error during one of the early cleavage divisions of the fertilized zygote (Fig. 338-2). Short stature and other somatic features result from haploinsufficiency of one or more genes encoded on the short arm of the X chromosome. Streak gonads result when genetic material is missing from either the long or short arm of the X. In individuals with mosaicism or structural abnormalities of the X, the phenotype on average is less severe than in the 45,X variety. In some patients with hypertrophy of the clitoris, an unidentified fragment of a chromosome is present and is assumed to be an abnormal Y; gonadoblastoma may develop in the streak gonads in this subset of patients. The Y-linked genes that predispose to gonadoblastoma are distinct from *SRY* because XY women with *SRY* deletions or mutations are also at risk for such tumors. Rarely, familial transmission of gonadal dysgenesis can be the result of a balanced X-autosome translocation (Chap. 66). Analysis of chromosomal karyotype is necessary to establish the diagnosis and to identify the group with Y chromosomal elements and hence a chance of developing malignancy in the streak gonads.

After the time of expected puberty, pubic and axillary hair remain sparse, the breasts are infantile, and no menses occur. Serum FSH is elevated in infancy, falls during midchildhood to the normal range, and increases to castrate levels at the age of 9 or 10. At this time, the serum LH level is also elevated, and plasma estradiol levels are low [<40 pmol/L (<10 pg/mL)]. Approximately 2% of 45,X and 12% of mosaic women have sufficient residual follicles to allow some menstruation, and occasionally minimally

affected women become pregnant; the reproductive life in such individuals is brief.

TREATMENT

At the anticipated time of puberty, replacement therapy with estrogen should be instituted to induce maturation of the breasts, labia, vagina, uterus, and fallopian tubes (Chap. 336). Linear growth and bone maturation rates usually double during the first year of treatment with estradiol, but the eventual height rarely approaches the predicted height. Combination therapy with oxandrolone and/or growth hormone accelerates growth and increases final height. Streak gonads should be removed in all women who are virilized or have Y-chromosome sequences.

MIXED GONADAL DYSGENESIS

Clinical Features Mosaicism for a Y-bearing cell line, usually the 45,X/46,XY karyotype, is responsible for most instances of mixed gonadal dysgenesis. Affected individuals usually have a testis on one side and a streak gonad on the other, but bilateral dysgenetic testes or bilateral streak gonads may be present. The incidence is unknown, but in most hospitals the disorder is the second most common cause of ambiguous genitalia in the neonate after congenital adrenal hyperplasia.

The phenotype varies depending on the proportion of XY cells and their distribution. About two-thirds of such children are reared as females. Many have ambiguous genitalia, including phallic enlargement, a urogenital sinus, and some labioscrotal fusion. In most the testis is intraabdominal; individuals with a testis in the inguinal or scrotal position are usually reared as males. A uterus, vagina, and at least one fallopian tube are almost invariably present.

The prepubertal testis appears relatively normal. The postpubertal testis contains abundant Leydig cells, but the seminiferous tubules lack germinal elements and contain only Sertoli cells. The streak gonad, a thin, pale, elongated structure located either in the broad ligament or along the pelvic wall, is composed of ovarian stroma. At puberty the testis secretes androgen, causing virilization and phallic enlargement. Feminization is rare; when it occurs, estrogen secretion from a gonadal tumor should be suspected.

Approximately a third of these individuals exhibit somatic features of 45,X gonadal dysgenesis. Approximately two-thirds have the 45,X/46,XY karyotype, and the remainder have a 46,XY karyotype or a variant mosaicism. The origin of 45,X/46,XY mosaicism is best explained by the loss of a Y chromosome during an early mitotic division of an XY zygote similar to the postulated loss of the X chromosome in the 46,XY/47,XXY mosaicism shown in Fig. 338-2.

Pathophysiology It is assumed (but has been difficult to prove) that the 46,XY cell line stimulates testicular differentiation, whereas the 45,X stem leads to the development of the contralateral streak gonad. Both masculinization and mullerian duct regression in utero are incomplete. Since Leydig cell function may be that of a normal male at puberty, inadequate virilization in utero may be due to delayed development of a testis that is ultimately capable of normal Leydig cell function.

TREATMENT

For the older child or adult in whom gender is established prior to diagnosis, the central consideration is the possibility of tumor development in the gonads, which can occur prior to puberty. The overall incidence of seminomas and gonadoblastomas may be as high as 25%. Such tumors occur most frequently in subjects with a female phenotype who lack the somatic features of 45,X gonadal dysgenesis and are more common in testes than in the streak gonad. When the diagnosis is established in phenotypic females, prophylactic gonadectomy should be performed because gonadal tumors may occur in childhood and because the testes secrete androgen at puberty and thus cause virilization. Such women, like those with gonadal dysgenesis, are then given estrogen to induce and maintain feminization.

When the diagnosis is established in phenotypic males during late childhood or in adults, the management is more complicated. Men with mixed gonadal dysgenesis are infertile (no germinal elements are present in the testes) and have a high risk of developing gonadal tumors. In deciding which testes can be safely conserved the following observations apply: (1) tumors develop in scrotal streak gonads but not in scrotal testes, (2) tumors that develop in intraabdominal testes are always associated with ipsilateral mullerian duct structures, and (3) tumors in streak gonads are always associated with tumors in the contralateral abdominal testis. Based on these observations, it is recommended that (1) all streak gonads should be removed, (2) scrotal testes should be preserved, and (3) intraabdominal testes should be excised unless they can be relocated in the scrotum and are not associated with ipsilateral mullerian duct structures. Reconstructive surgery of the phallus should be performed when appropriate.

When the diagnosis is established in early infancy and the genitalia are ambiguous, gender assignment is usually female, and resection of the phallus and gonadectomy can be performed in infancy, sometimes in one procedure. If the decision is for male gender assignment, the same criteria apply as to which testes should be removed as in older males.

TRUE HERMAPHRODITISM

Clinical Features True hermaphroditism is a condition in which both an ovary and a testis or one or more gonads with features of both (ovotestis) are present. To justify the diagnosis, both types of gonadal epithelium must be documented histologically, the presence of ovarian stroma without oocytes not being sufficient. The incidence is unknown, but more than 400 cases have been reported. Three categories are recognized: (1) one-fifth are bilateral -- testicular and ovarian tissue (ovotestes) on each side; (2) two-fifths are unilateral -- an ovotestis on one side and an ovary or a testis on the other; and (3) the remainder are lateral -- a testis on one side and an ovary on the other.

The external genitalia exhibit all gradations of the male-to-female spectrum. Two-thirds of affected individuals are sufficiently masculinized to be reared as males, but fewer than one-tenth have normal male external genitalia; most have hypospadias and incomplete labioscrotal fusion. Two-thirds of phenotypic females have an enlarged

clitoris, and most have a urogenital sinus. Differentiation of the internal ducts usually corresponds to the adjacent gonad. Although an epididymis usually develops adjacent to a testis, the vas deferens is usually incomplete. Of the patients with an ovotestis, three-fourths have an epididymis, two-thirds have a fallopian tube, one-tenth have a vas deferens, and one-tenth have both a vas deferens and a fallopian tube. The uterus may be hypoplastic or unicornuate. The ovary is usually in the normal position, but the testis or ovotestis may be found at any level along the route of testicular descent, frequently associated with an inguinal hernia. Testicular tissue is present in the scrotum or the labioscrotal fold in one-third, in the inguinal canal in one-third, and in the abdomen in one-third.

Variable feminization and virilization ensue at puberty; three-fourths develop gynecomastia, and about half menstruate. In phenotypic men, menstruation may cause cyclic hematuria. Ovulation occurs in approximately one-fourth and is more common than spermatogenesis. In men, ovulation may cause testicular pain. Fertility has been reported in women and more rarely in men. Congenital malformations of other systems are unusual.

Pathophysiology About two-thirds of individuals have a 46,XX karyotype, a tenth have a 46,XY karyotype, and the remainder are chimeras or mosaics in whom a Y cell line is present. The mechanism responsible for the abnormal gonadal development is unknown. Only 10% of 46,XX true hermaphrodites are *SRY* positive, presumably the consequence of mosaicism or translocation of a portion of the Y chromosome; the remainder are believed to result from gain-of-function mutations in downstream genes involved in *SRY* action. On occasion, multiple sibs with a 46,XX karyotype are affected, possibly the result of an autosomal or X-linked mutation.

Because corpora lutea are frequently present in the ovaries, it is presumed that the female neuroendocrine axis functions normally in such individuals. Feminization (gynecomastia and menstruation) is the result of secretion of estradiol by ovarian tissue. In masculinized patients, secretion of androgen predominates, and some produce sperm.

TREATMENT

When the diagnosis is made in early infancy, gender assignment depends on the anatomic features. In older children and adults, gonads and internal duct structures that are contradictory to the predominant phenotype (and the gender of rearing) should be removed, and the external genitalia should be modified when appropriate. Gonadal tumors are rare but have been reported in true hermaphrodites who carry Y chromosome sequences. Consequently, the possibility of future tumor development must be taken into account when the decision is made to preserve gonadal tissue.

DISORDERS OF GONADAL SEX

Disorders of gonadal sex result when chromosomal sex is normal but differentiation of the gonads is abnormal. Thus, gonadal sex does not correspond to chromosomal sex.

PURE GONADAL DYSGENESIS

Clinical Features Pure gonadal dysgenesis is a disorder in which phenotypic females with gonads and genitalia characteristic of gonadal dysgenesis (bilateral streaks, infantile uterus and fallopian tubes, and sexual infantilism) have normal height, few if any somatic anomalies, and either a normal 46,XX or 46,XY karyotype. This disorder is about one-tenth as common as gonadal dysgenesis. It is genetically distinct but cannot be distinguished clinically from those instances of gonadal dysgenesis with minimal somatic abnormalities. The height is normal or greater than normal, some individuals being >170 cm. Estrogen levels vary from profound deficiency typical of 45,X gonadal dysgenesis to some breast development and menses that terminate in an early menopause. About 40% have some feminization. Axillary and pubic hair is scanty, and the internal genitalia consist of mullerian derivatives only. In both the 46,XX and the 46,XY forms the disorder prevents differentiation of ovary or testis, respectively; the development of the female phenotype and the elevation of gonadotropin secretion are due to failure of gonadal development.

Tumors may develop in the streak gonads, particularly dysgerminoma or gonadoblastoma in the 46,XY disorder. Such tumors may be heralded by the development of virilizing signs or a pelvic mass.

Pathophysiology Although chromosomal mosaicisms have been described under this nosology, the designation here is restricted to women with uniform 46,XX or 46,XY karyotypes. (Those with mosaicism are variants of gonadal dysgenesis or mixed gonadal dysgenesis, as described above.) The rationale for this restricted definition is based on the fact that both the XX and XY varieties can result from single gene mutations that are presumed to involve gene(s) essential for gonadal development. Several sibships have been reported in which more than one individual is affected with the 46,XX disorder, frequently the result of consanguineous matings, suggesting an autosomal recessive inheritance.

The 46,XY variety may occur in families; in some the disorder appears to be inherited as an X-linked trait, and in others the pattern suggests a male-limited autosomal recessive inheritance. About 15% of 46,XY women have either a deletion or a mutation in the *SRY* coding sequence. Other instances could be due to mutations in *SRY* outside the coding sequence, in other genes that influence *SRY* expression, or in the downstream genes that are controlled by *SRY*. Indeed, mutations in several genes that are downstream of *SRY* are now known to be a cause of the dysgenetic testes syndrome (also termed *dysgenetic male pseudohermaphroditism*; see below).

TREATMENT

The management of the estrogen deficiency is identical to that in gonadal dysgenesis; namely, appropriate estrogen replacement therapy is initiated at the time of expected puberty and maintained in adult life (Chap. 336). Because of the high frequency of gonadal tumors in the 46,XY variety, the streak gonads should be removed once the diagnosis is made; development of virilizing signs is indication for immediate surgery. The natural history of the gonadal tumors is uncertain, but the prognosis after surgical removal is usually good.

DYSGENETIC TESTES

Individuals with these disorders are genetic males with disorders of testicular development that vary from streak gonads similar to those in gonadal dysgenesis to less severe defects. The disorders are frequently associated with systemic abnormalities, because many of the genes involved are also involved in the development of other tissues. The first of these genes to be identified was the Wilms' tumor gene *WT1*; mutations of this gene cause two disorders -- the Denys-Drash and Frasier syndromes. *Denys-Drash syndrome* is an autosomal dominant disorder characterized by development of Wilms' tumors in males with a spectrum of gonadal defects ranging from streak gonads to less severely affected testes; urogenital defects include diffuse mesangial sclerosis of the kidneys. The underlying mutations in the zinc finger region of WT-1 are believed to inhibit the function of the wild-type protein and hence act as dominant negative mutations. Patients with *Frasier syndrome* have gonadal dysgenesis, impaired virilization, and focal sclerosis of the kidney but do not develop Wilms' tumors. Mutations in intron 9 of the *WT1* gene that cause Frasier syndrome interfere with the synthesis of specific splice variants of *WT1*.

A second downstream gene that is essential for differentiation of the testes is *SF1* (also known as *FTZF1*), a member of the nuclear hormone receptor superfamily. SF-1 regulates the expression of many genes involved in adrenal and gonadal development and steroidogenesis, as well as the *AMH* gene. Heterozygous mutation of this autosomal gene has been associated with 46,XY gonadal dysgenesis with adrenal insufficiency.

Another downstream gene is *SOX9*, a close relative of *SRY* that maps to chromosome 17q. This gene is expressed in high levels in the testes, where it is believed to be a key regulator of male differentiation. Heterozygous mutations of this gene cause 46,XY gonadal dysgenesis and skeletal abnormalities (*campomelic dysplasia*).

46,XY gonadal dysgenesis is also associated with duplication of the short arm of the X chromosome (Xp21), a phenomenon termed *dosage-sensitive sex reversal*. Loss-of-function mutations of the *DAX1* gene in this region of the X chromosome are associated with adrenal hypoplasia congenita and hypogonadotrophic hypogonadism. The DAX-1 protein inhibits the expression of SF-1-regulated genes, providing a potential mechanism by which an excess of DAX-1 could cause gonadal dysgenesis in genetic males with Xp21 duplications.

THE ABSENT TESTES SYNDROME (ANORCHIA, TESTICULAR REGRESSION, GONADAL AGENESIS, AGONADISM)

Clinical Features A spectrum of phenotypes occurs in 46,XY males with absent or rudimentary testes but in whom unequivocal evidence exists that endocrine function of the testis (e.g., consistent mullerian duct regression and variable testosterone synthesis) was present at some time during embryonic life. In pure gonadal dysgenesis, in contrast, no evidence can be inferred for gonadal function during embryonic development. The manifestations vary from complete failure of virilization to incomplete virilization of the external genitalia to otherwise normal men with bilateral anorchia.

The purest form is represented by 46,XY females with absent testes, sexual infantilism, and absence of both mullerian duct derivatives and accessory organs of male reproduction. Such individuals differ from those with 46,XY pure gonadal dysgenesis in that no gonadal remnant can be identified, including no streak gonad and no mullerian derivatives. Testicular failure must have occurred, therefore, between the onset of AMH synthesis and the onset of testosterone secretion (e.g., after development of Sertoli cells but before the onset of Leydig cell function).

In others, testicular failure occurred later in gestation, and these individuals may constitute problems in gender assignment. In some, failure of mullerian regression is more pronounced than failure of testosterone secretion, but none exhibit normal mullerian development. In those with more extensive virilization, the external genitalia are phenotypically male, but rudimentary oviducts and vasa deferentia may coexist internally.

At the final extreme is the syndrome of bilateral anorchia in which phenotypic men have absence of mullerian structures and gonads but male wolffian duct derivatives and external genitalia. Microphallus implies that failure of testosterone secretion occurred late in embryogenesis after anatomic development of the male urethra was complete. Gynecomastia may or may not be present.

Pathophysiology The pathogenesis is not understood. Testicular regression could be the result of mutant genes, teratogen, or trauma, and the disorder may well be heterogeneous in origin. Several instances of agonadism have occurred in the same family, some unilateral and others bilateral. Some individuals in whom no testes can be identified at laparotomy have blood testosterone values above the castrate range, presumably derived from remnant testes.

TREATMENT

The management of the two extremes is clear-cut. Sexually infantile, phenotypic females should be given adequate estrogen to ensure appropriate feminization, and any coexisting vaginal agenesis should be treated by surgical or medical means. Likewise, phenotypic males with anorchia should be given androgen replacement to allow normal male secondary sexual development. Individuals with incomplete virilization or ambiguous external genitalia demonstrate a more complex problem and require careful assessment to determine appropriate gender assignment, hormonal therapy at the time of expected puberty, and surgical correction of the external genitalia when appropriate.

DISORDERS OF PHENOTYPIC SEX

Disorders of phenotypic sex occur in 46,XX or 46,XY individuals with appropriate gonadal sex but in whom development of the urogenital tract is inappropriate for the chromosomal/gonadal sex.

FEMALE PSEUDOHERMAPHRODITISM

Female pseudohermaphroditism occurs in 46,XX women with bilateral ovaries but with variable virilization of the urogenital tract because of androgen excess during fetal life.

Congenital Adrenal Hyperplasia

Clinical Features The pathways by which glucocorticoids are synthesized in the adrenal gland and androgens are formed in the testis and adrenal are summarized in Fig. 338-3. Three reactions are common to both pathways (cholesterol side chain cleavage, 3b-hydroxysteroid dehydrogenase/isomerase, and 17a-hydroxylase); impairment of any of these reactions results in deficiency of glucocorticoid and androgen synthesis and consequently causes both congenital adrenal hyperplasia (due to enhanced ACTH levels) and defective virilization of the male embryo (male pseudohermaphroditism). Two reactions are involved exclusively in androgen synthesis (17,20-lyase and 17b-hydroxysteroid dehydrogenase); deficiency in either results in pure male pseudohermaphroditism with normal glucocorticoid synthesis. Deficiency of either of the terminal two enzymes of glucocorticoid synthesis (21-hydroxylase and 11b-hydroxylase) impairs formation of hydrocortisone; the compensatory increase in ACTH secretion causes adrenal hyperplasia and a secondary increase in androgen formation that virilizes the female and induces precocious masculinization in the male.

The major features of congenital adrenal hyperplasia are listed in Table 338-4. The *adrenal insufficiency* can be equally severe and life-threatening in both sexes and is described in Chap. 331. Some defects in steroidogenesis cause female pseudohermaphroditism, and some cause male pseudohermaphroditism. (3b-hydroxysteroid dehydrogenase/isomerase deficiency can cause either male or female pseudohermaphroditism, but since incomplete virilization of the male is more common, the disorder will be discussed under male pseudohermaphroditism.)

Congenital adrenal hyperplasia due to classic 21-hydroxylase deficiency is the most common cause of ambiguous genitalia in the newborn, with an incidence of between 1 in 5000 and 1 in 15,000 live births in Europe and the United States; it may or may not be associated with mineralocorticoid deficiency (salt loss) (Table 338-4). Virilization is usually apparent at birth in the female and within the first 2 to 3 years of life in the male. Manifestations in females include hypertrophy of the clitoris with ventral chordee, partial fusion of the labioscrotal folds, and variable virilization of the urethra. The uterus, fallopian tubes, and ovaries are normal, and the wolffian ducts do not virilize, probably because adrenal function begins relatively late in embryogenesis. The external appearance of an affected female newborn is similar to that of a male with bilateral cryptorchidism and hypospadias. The labioscrotal folds are bulbous and rugated and resemble a scrotum. Rarely, the virilization is so severe that development of a complete penile urethra and prostate results in errors in sex assignment at birth. Radiography following the injection of radiopaque dye into the external genital orifice is helpful in demonstrating the presence of vagina, uterus, and (sometimes) fallopian tubes. Occasionally, virilization of the female is slight or absent at birth and becomes evident in later infancy, adolescence, or adulthood (the so-called nonclassic or late-onset form of the disorder). In both sexes, rapid somatic maturation results in premature epiphyseal closure and a short adult height. The untreated female with the classic disorder grows rapidly during the first year of life and has progressive virilization. At the time of expected puberty there is a failure of normal female sexual development and absence of menstruation.

Since male phenotypic differentiation is normal, the condition is usually not recognized at birth in boys in the absence of adrenal insufficiency. However, early maturation of the external genitalia, development of secondary sex characteristics, coarsening of the voice, frequent erections, and excessive muscular development are noticeable during the first few years of life. Virilization in the male can follow two patterns. Excessive adrenal androgens can inhibit gonadotropin production so that the testes remain infantile in size despite the acceleration of masculinization. Such untreated adult men are capable of erection and ejaculation but have no spermatogenesis. Alternatively, adrenal androgen secretion can induce premature maturation of the hypothalamic-pituitary axis and initiate a true precocious puberty including early maturation of spermatogenesis (Chap. 335). The untreated male is also subject to the development of ACTH-dependent "tumors" of the adrenal rest cells of the testes.

In classic 21-hydroxylase deficiency, which accounts for about 95% of congenital adrenal hyperplasia, decreased production of hydrocortisone leads to increased release of ACTH, enlargement of the adrenal glands, and partial or complete compensation of the defect in the secretion of hydrocortisone. In about half, the enzyme defect appears to be partial, and cortisol secretion is normal. This form is termed *simple virilizing*. When deficiency of the enzyme is more complete, the so-called salt-losing form of 21-hydroxylase deficiency, production of cortisol and aldosterone is inadequate, leading to severe salt wastage with anorexia, vomiting, volume depletion, and vascular collapse within the first few weeks of life. In untreated patients, there is overproduction of the cortisol precursors prior to the 21-hydroxylase step, causing an increase in plasma progesterone and 17-hydroxyprogesterone. These steroids are weak aldosterone antagonists at the receptor level; in the compensated state aldosterone production increases to attempt to maintain normal sodium balance. Increased substrate availability is also responsible for the enhanced androgen synthesis and hence for the virilization.

Female pseudohermaphroditism also occurs in 11b-hydroxylase deficiency. In this disorder, a block in hydroxylation at the 11-carbon results in the accumulation of 11-deoxycortisol and deoxycorticosterone (DOC), a potent salt-retaining hormone that causes hypertension rather than salt loss. The clinical features that stem from glucocorticoid deficiency and androgen excess are similar to those in 21-hydroxylase deficiency.

Pathophysiology Both disorders are due to autosomal recessive mutations. The carrier frequency for CYP21A2 deficiency is about 1 in 50. Because the gene is located on the sixth chromosome close to the HLA-B locus, heterozygous carriers and homozygotes within a given family can be identified on the basis of the HLA haplotype. At the molecular level the mutations that give rise to 21-hydroxylase deficiency are highly polymorphic; indeed, partial gene deletions (10 to 30%), conversion of the gene from a functional state to a form that is not transcribed normally (10%), and point mutations (60 to 75%) have been characterized in the disorder. The classic disorder is due to mutations that severely impair enzyme activity, and less severe mutations cause the nonclassic, or late-onset, variety. 11b-hydroxylase activity is encoded by two genes on chromosome 8; mutations of the *CYP11B1* gene are responsible for this disorder. The *CYP11B2* gene encodes aldosterone synthase. A late-onset variant of 11b-hydroxylase deficiency exists but has not been characterized at the molecular level.

Urinary excretion of 17-ketosteroids and of the metabolites that accumulate proximal to the enzymatic blocks is increased. Plasma ACTH is elevated. In CYP21A2 deficiency, 17-hydroxyprogesterone accumulates in blood and is excreted predominantly as pregnanetriol. In CYP11B1 deficiency, 11-deoxycortisol accumulates in blood and is excreted predominantly as tetrahydrocortexolone.*For additional discussion of the endocrine pathology, see Chap. 331.*

TREATMENT

Gender assignment should correspond to the chromosomal and gonadal sex, and appropriate surgical correction of the external genitalia should be undertaken as early as possible. This is of importance because appropriately treated men and women are capable of fertility. However, if the correct diagnosis is made late (after 3 years of age), gender assignment should be changed only after careful consideration of the psychosexual background.

Treatment with appropriate glucocorticoids prevents the consequences of hydrocortisone deficiency, arrests the rapid virilization, and prevents premature somatic and epiphyseal maturation. The suppression of the abnormal steroid secretion corrects the hypertension in CYP11B1 deficiency and in both disorders allows normal onset of menses and development of female secondary sex characteristics. In males, glucocorticoid therapy suppresses adrenal androgens and results in normal gonadotropin secretion, testicular development, and spermatogenesis. The usual maintenance dose of hydrocortisone is 10 to 20 mg/m$_2$per day, given in three divided doses. However, the dose must be adjusted on an individual basis to optimize ACTH suppression while avoiding glucocorticoid side effects, which include growth retardation as well as Cushingoid features. Measurements of plasma 17-hydroxyprogesterone, androstenedione, ACTH, and renin levels have all been used to assess adequacy of replacement therapy. In severe CYP21A2 deficiency associated with salt loss or elevated plasma renin activity, treatment with mineralocorticoids is also indicated, and plasma renin levels should be monitored to assess the adequacy of mineralocorticoid replacement. Trials are underway to assess the potential use of antiandrogens (e.g., spironolactone, cyproterone acetate, flutamide) or aromatase inhibitors (e.g., latrazole, testolactone) as adjunctive therapy that may allow reductions in glucocorticoid dose. Treatment of affected fetuses in utero (beginning at 4 to 6 weeks) has been accomplished by administering dexamethasone (which crosses the placenta) to the mother. Though this treatment can reduce the extent of virilization in some girls, it is associated with maternal side effects, and the long-term consequences have not been established.

Other Causes of Female Pseudohermaphroditism Placental aromatase deficiency due to mutations in the gene encoding aromatase (*CYP19*) causes virilization of female embryos because of defective conversion of androgens to estrogens in the placenta and the secondary increase in testosterone levels in the fetus; in postnatal life CYP19 deficiency in women causes hirsutism and development of polycystic ovaries. Female pseudohermaphroditism can also occur in babies born to mothers with virilizing tumors of the ovary (e.g., arrhenoblastomas or luteomas of pregnancy) and, rarely, to mothers with virilizing adrenal tumors. In the past, the administration to pregnant women of progestogens with androgenic side effects (such as 17a-ethinyl-19-nor-testosterone) to

prevent abortion resulted in masculinization of female fetuses.

Developmental Disorders of Mullerian Ducts (Congenital Absence of the Vagina, Mullerian Agenesis)

Clinical Features Congenital hypoplasia or absence of the vagina in combination with abnormal or absent uterus (the Mayer-Rokitansky-Kuster-Hauser syndrome) is second to gonadal dysgenesis as a cause of primary amenorrhea. Most patients are ascertained after the time of expected puberty because of failure to menstruate. The height is normal, and the breasts, axillary and pubic hair, and habitus are feminine in character. The uterus can vary from almost normal, lacking only a conduit to the introitus, to the characteristic rudimentary bicornuate cords with or without a lumen. In some patients cyclic abdominal pain indicates that sufficient functional endometrium is present to result in retrograde menstruation and/or hematometra.

About one-third have abnormal kidneys, most commonly agenesis, ectopy, fused kidneys of the horseshoe type, or solitary ectopic kidneys in the pelvis. Skeletal abnormalities are present in one-tenth; most involve the spine, and limb and rib defects account for the rest. Specific abnormalities include wedge vertebrae, fused rudimentary or asymmetric vertebral bodies, supernumerary vertebrae, and the Klippel-Feil syndrome (congenital fusion of the cervical spine, short neck, low posterior hairline, and painless limitation of cervical movement).

Pathophysiology The karyotype is 46,XX. Familial occurrence has been described, and the pattern of inheritance in most is consistent with a sex-limited autosomal dominant mutation. Sporadic cases may represent new mutations of the type responsible for the familial disorder or be multifactorial in etiology. In the familial cases, expressivity is variable; some have skeletal or renal abnormalities only, and some have other abnormalities of mullerian derivatives such as a double uterus. Bilateral renal aplasia in stillborn infants is commonly associated with absence of the uterus and vagina. Thus, the family history should be probed for isolated skeletal and renal abnormalities and for stillbirths that might result from congenital absence of both kidneys. Ovarian function is normal, and successful pregnancies have occurred after corrective vaginal surgery in patients with a normal uterus.

TREATMENT

Vaginal agenesis can be treated by surgical or nonsurgical means. Surgical repair generally utilizes a split-thickness skin graft around a solid rubber mold to create an artificial vagina. Medical therapy involves the repeated application of pressure against the vaginal dimple with a simple dilator to force development of adequate vaginal depth. In view of complication rates of 5 to 10% in surgical series, surgery should be reserved for patients in whom a well-formed uterus is present and the possibility of fertility exists. Frequent coitus or instrumental dilatation is essential for maintaining patency of neovaginas formed by either technique.

MALE PSEUDOHERMAPHRODITISM

Defective virilization of the 46,XY embryo (male pseudohermaphroditism) can result

from defects in androgen synthesis, defects in androgen action, defects in mullerian duct regression, and uncertain causes.

Abnormalities in Androgen Synthesis

Clinical Features Enzymatic defects that result in defective testosterone synthesis (Fig. 338-3) account for only about a fifth of cases of male pseudohermaphroditism (Tables 338-4 and338-5). Each of the defects blocks a step in the conversion of cholesterol to testosterone. Three enzymatic reactions are common to the synthesis of other adrenal hormones as well: cholesterol side chain cleavage, 3b-hydroxysteroid dehydrogenase/isomerase, and 17a-hydroxylase (CYP17). Consequently, their deficiency results in congenital adrenal hyperplasia (Table 338-4) as well as male pseudohermaphroditism. Two others (17,20-lyase and 17b-hydroxysteroid dehydrogenase 3) are unique to the pathway of androgen synthesis, and their deficiency results only in male pseudohermaphroditism. Since androgens are obligatory precursors of estrogens, synthesis of estrogen is also low in all but the terminal defect (17b-hydroxysteroid dehydrogenase 3 deficiency).

Adrenal dysfunction is described in Chap. 331, and the present discussion concerns the abnormal sexual development. In genetic males with defective testosterone synthesis, absence of the uterus and fallopian tubes indicates that mullerian duct inhibition was normal during embryogenesis. Masculinization of the urogenital tract and external genitalia and virilization at puberty vary from almost normal to absent, and, therefore, the manifestations vary from men with mild hypospadias to phenotypic women who prior to puberty resemble women with complete testicular feminization. This heterogeneity is the consequence of varying severity of the enzymatic defects, varying effects of the steroids that accumulate proximal to the various metabolic blocks, and the presence of alternative enzymatic pathways in some disorders. In patients with partial defects in whom the plasma testosterone level is normal, the diagnosis can only be made by measuring the steroids that accumulate proximal to the metabolic block.

Congenital lipoid adrenal hyperplasia is an autosomal recessive disorder in which virtually no urinary steroids (either 17-ketosteroids or 17-hydroxycorticoids) can be detected. Since the defect blocks the conversion of cholesterol to pregnenolone, a step catalyzed by the cholesterol side chain cleavage enzyme (CYP11A1), it was originally assumed that the defect must involve this enzyme. However, the disorder is instead caused by mutations in the *st*eroidogenic *a*cute *r*egulatory (*StAR*) gene, which encodes the protein that transports cholesterol from the cytosol to the inner mitochondrial membrane in the adrenal and gonads. Manifestations of the disorder include salt wasting and profound adrenal insufficiency, and most affected individuals die during infancy. At autopsy, the adrenals and testes are enlarged and infiltrated with lipid. Affected males are incompletely masculinized, whereas affected female infants have normal genital development because lipid accumulation in the ovary does not occur until there is follicular development and active steroidogenesis.

3b-Hydroxysteroid dehydrogenase/isomerase 2 deficiency causes varying failure of masculinization and the development of a vagina in male infants. Female infants may be modestly virilized at birth due to the weak androgenic potency of dehydroepiandrosterone, the major steroid secreted. If the enzyme is absent in both the

adrenal and testis, no urinary steroids contain a D_4-3-keto configuration, whereas in patients in whom the defect is partial or affects only the testis, the urine may contain normal or elevated levels of D_4-3-ketosteroids. Most patients have marked salt wasting and profound adrenal insufficiency, and long-term survival in untreated cases occurs only in states of partial deficiency. Minimally affected males may experience an otherwise normal male puberty except for profound gynecomastia. In these boys, a low-normal blood testosterone level is accompanied by elevated D_5 precursor steroids. 3b-Hydroxysteroid dehydrogenase activity is catalyzed by more than one isoenzyme. The type 2 isoenzyme is expressed in adrenals and gonads; the disorder described above is due to any of several mutations in this enzyme. The coding sequence of this gene is said to be normal in several individuals with the late-onset variant of the disease, the pathophysiology of which is unclear.

17a-Hydroxylase-17,20-lyase (CYP17) deficiency impairs the introduction of the 17-hydroxyl and the scission of the C-17,20 carbon bond that convert pregnenolone and progesterone to dehydroepiandrosterone and androstenedione, respectively. These reactions are mediated by a single enzyme CYP17, which is encoded on chromosome 10; it remains unclear why both reactions occur in the ovary and testis, whereas in the adrenal 17-hydroxyprogesterone is largely converted to glucocorticoids and mineralocorticoids rather than the 19-carbon steroids. Of note, some patients appear to have selective impairment of either 17a-hydroxylase or 17,20-lyase activity. These mutations have identified enzyme domains proposed to undergo posttranslational modification and selective interactions with cofactors that switch enzymatic activity. Whatever the mechanism, the consequences of 17a-hydroxylase and 17,20-lyase deficiencies are different.

17a-Hydroxylase deficiency is characterized by hypogonadism, absence of secondary sex characteristics, hypokalemic alkalosis, hypertension, and virtually undetectable hydrocortisone secretion in phenotypic women. Formation of both corticosterone and DOC by the adrenal is elevated, and urinary 17-ketosteroids are low. Aldosterone secretion is low due to high plasma DOC and depressed angiotensin levels and returns to normal after suppressive doses of glucocorticoids. In 46,XX individuals, amenorrhea, absent sexual hair, and hypertension are common, but the phenotype is that of a normal prepubertal woman. In males, the deficiency results in defective virilization that varies from complete male pseudohermaphroditism to ambiguous genitalia with perineoscrotal hypospadias and, in some, gynecomastia. Adrenal insufficiency does not develop, since the secretion of both corticosterone (a weak glucocorticoid) and DOC (a mineralocorticoid) is elevated. Hypertension and hypokalemia are prominent (even in the neonatal period) and remit after suppression of the DOC secretion by glucocorticoid replacement. A variety of point mutations, deletions, and insertions in the *CYP17* gene have been characterized in affected individuals.

17,20-Lyase deficiency in males is associated with normal function of the adrenal cortex and a variable pattern of male pseudohermaphroditism. In the majority there is genital ambiguity at birth, with some virilization at the time of expected puberty. Rare 46,XY patients have had a female phenotype and no virilization at the time of expected puberty. The disorder has been recognized in one 46,XX woman with sexual infantilism. Mutations of *CYP17* that cause this disorder involve an area of the gene that is known to encode a binding site for the redox-partner of the enzyme.

17b-Hydroxysteroid dehydrogenase 3 (17b-HSD-3) deficiency involves the final step in testosterone biosynthesis, reduction of the 17-keto group of androstenedione. This is the most common enzymatic defect in testosterone synthesis. Affected 46,XY males usually have a female phenotype with a blind-ending vagina and absence of mullerian derivatives, but inguinal or abdominal testes and virilized wolffian duct structures are present. At the time of expected puberty, both virilization (with phallic enlargement and development of facial and body hair) and a variable degree of feminization take place. In some untreated patients, reversal of gender behavior from female to male occurs at puberty. Plasma testosterone level may be in the low-normal range, making it essential to document elevation in plasma androstenedione to make the diagnosis. Isoenzymes encoded by several different genes possess 17b-hydroxysteroid dehydrogenase activity, but the isoenzyme 3 is expressed in the testes. A variety of mutations have been characterized in the 17b-HSD-3 gene in affected individuals.

Pathophysiology These various disorders are inherited as autosomal recessive traits. The pattern of steroid secretion and excretion depends on the site of the various metabolic blocks (Fig. 338-3). In general, gonadotropin secretion is high, and many individuals with incomplete defects are able to compensate so that the steady-state levels of testosterone may be normal or almost normal.

In rare cases of male pseudohermaphroditism, testosterone formation is deficient for reasons other than a single enzyme defect in androgen synthesis. These include disorders in which Leydig cell agenesis is due to autosomal recessive loss-of-function mutations of theLHreceptor or to the secretion of a biologically inactive LH molecule. In addition, as described above, several disorders, including familial 46,XY pure gonadal dysgenesis, sporadic dysgenetic testes, and the absent testis syndrome, are characterized by deficient testosterone production due to abnormal gonadal development.

TREATMENT

Therapy with glucocorticoids and in some instances mineralocorticoids is indicated in those disorders causing adrenal hyperplasia. The management of the genital abnormalities depends on the individual case. Fertility has not been reported, and its consideration does not enter into sex assignment. In genetic females there is no problem (except in diagnosis), in that affected individuals are raised appropriately as females and estrogen therapy is begun at the time of expected puberty to promote development of female secondary sex characteristics. Whether newborn males with ambiguous genitalia should be raised as males or females depends on the anatomic defect; in general, the more severely affected should be raised as females, and corrective surgery of the genitalia and removal of the testes should be undertaken as early as possible. In such women estrogen therapy is begun at the appropriate age to allow development of normal female secondary sex characteristics. In individuals raised as males, corrective surgery is indicated for any coexisting hypospadias, and plasma androgens should be monitored at the time of expected puberty to determine whether testosterone therapy is appropriate.

Abnormalities in Androgen Action Several disorders of male phenotypic development

result from abnormalities of androgen action. The spectrum of phenotypes is described in Tables 338-4 and 338-5. In these disorders, testosterone formation and mullerian regression are normal, but male development is impaired because of resistance to androgen action in target tissues.

Steroid 5a-Reductase 2 Deficiency This autosomal recessive disorder is characterized by (1) severe perineoscrotal hypospadias; (2) a blind vaginal pouch of variable size opening either into the urogenital sinus or into the urethra; (3) testes with normal epididymides, vasa deferentia, and seminal vesicles, and termination of the ejaculatory ducts in the blind-ending vagina; (4) a female habitus with normal axillary and pubic hair but without female breast development; (5) the absence of uterus and fallopian tubes; (6) normal male plasma testosterone; and (7) masculinization to a variable degree at the time of puberty.

The realization that virilization during embryogenesis is defective only in the urogenital sinus and the external genitalia provided insight into the fundamental abnormality. Testosterone, the androgen secreted by the fetal testis, is responsible for conversion of the wolffian duct into the epididymis, vas deferens, and seminal vesicle, whereas dihydrotestosterone mediates virilization of the urogenital sinus and the external genitalia. Consequently, impairment of dihydrotestosterone formation in a male embryo would be expected to cause the phenotype in this disorder -- normal male wolffian duct derivatives with defective masculinization of the external genitalia and urogenital sinus. Since testosterone itself regulates LH secretion (Chap. 335), plasma LH level is normal or minimally elevated. As a result, testosterone and estrogen production rates are those of normal men, and gynecomastia does not develop.

The fact that 5a-reductase 2 enzyme is deficient in this disorder was established by assay of biopsied tissues and cultured fibroblasts from affected individuals. Deletions or point mutations in the gene encoding steroid 5a-reductase 2 have been identified in most families studied. Approximately 40% are compound heterozygotes.

Receptor Disorders The androgen receptor is a member of the steroid/thyroid family of receptors with steroid-binding, DNA-binding, and functional domains and is encoded by a gene on the long arm of the X chromosome. Mutations of this gene impair receptor function and hence impair male phenotypic differentiation and/or virilization.

CLINICAL FEATURES Complete testicular feminization (also called *complete androgen insensitivity*) is a common form of male pseudohermaphroditism; estimates of frequency vary from 1 in 20,000 to 1 in 64,000 male births. It is the third most common cause of primary amenorrhea after gonadal dysgenesis and congenital absence of the vagina. The features are characteristic. Namely, a woman is ascertained either because of inguinal hernia (prepubertal) or primary amenorrhea (postpubertal). The development of the breasts, the habitus, and the distribution of body fat are female in character so that most have a truly feminine appearance. Axillary and pubic hair is absent or scanty, but some vulval hair is usually present. Scalp hair is that of a normal woman, and facial hair is absent. The external genitalia are unambiguously female, and the clitoris is normal. The vagina is short and blind-ending and may be absent or rudimentary. All internal genitalia are absent except for testes that contain normal Leydig cells and seminiferous tubules without spermatogenesis. The testes may be located in the abdomen, along the

course of the inguinal canal, or in the labia majora. Occasionally, remnants of mullerian or wolffian duct origin are present in the paratesticular fascia or in fibrous bands extending from the testis. Patients tend to be rather tall, and bone age is normal. Psychosexual development is unmistakably female with regard to behavior, outlook, and maternal instincts.

The major complication of undescended testes in this disorder, as in all forms of cryptorchidism, is the development of tumors (Chap. 96). Since affected individuals undergo normal pubertal growth and feminize at the time of expected puberty and since testicular tumors rarely develop until after puberty, it is usual to delay gonadectomy until after the time of expected puberty. Prepubertal gonadectomy is indicated if the testes are present in the inguinal region or labia majora and result in discomfort or hernia formation. (If hernia repair is indicated prepubertally, most physicians prefer to remove the testes at the same time to limit the number of operative procedures.) If the testes are removed prepubertally, estrogen therapy is required at the appropriate age to ensure normal growth and breast development. Postpubertal gonadectomy causes menopausal symptoms and other evidence of estrogen withdrawal, and suitable estrogen replacement is indicated (Chap. 336).

Incomplete testicular feminization is about one-tenth as frequent as the complete form. In this disorder there is minor virilization of the external genitalia (partial fusion of the labioscrotal folds and/or some degree of clitoromegaly), normal pubic hair, and mixed virilization and feminization at the time of expected puberty. The vagina is short and blind-ending, but in contrast to the complete form, the wolffian duct derivatives are often partially developed. Since women with the incomplete disorder virilize at the time of expected puberty, gonadectomy should be performed before the expected time of puberty in all prepubertal patients with clitoromegaly or posterior labial fusion.

Reifenstein syndrome (also called *partial androgen insensitivity*) is the term applied to forms of incomplete male pseudohermaphroditism initially described by a number of eponyms (Reifenstein syndrome, Gilbert-Dreyfus syndrome, Lubs syndrome). These syndromes are mutations that partially impair the function of the androgen receptor. The most common phenotype is a man with perineoscrotal hypospadias and gynecomastia, but the spectrum of defective virilization in affected families ranges from men with azoospermia to phenotypic women with pseudovaginas. Axillary and pubic hair is normal, but chest and facial hair is scanty. Cryptorchidism is common, the testes are small, and azoospermia is present. Some have defects in wolffian duct derivatives such as absence or hypoplasia of the vas deferens. Since the psychological development in most is unequivocally male, the hypospadias and cryptorchidism should be corrected surgically. The treatment of the gynecomastia is surgical removal.

A disorder of the androgen receptor that is not actually a form of male pseudohermaphroditism is manifested as *infertility and/or undervirilization in phenotypic men*. Some such individuals are minimally affected members of families with Reifenstein syndrome in whom azoospermia is the only manifestation of the receptor defect. The *undervirilized fertile male* is an even more subtle manifestation of androgen receptor defect. In these families, affected men have gynecomastia and undervirilization, and some are fertile. More commonly, however, individuals with negative family histories present with male infertility with or without undervirilization.

PATHOPHYSIOLOGY The karyotype is 46,XY, and the mutation is X-linked. The frequency of a positive family history varies from about two-thirds of patients with testicular feminization and Reifenstein syndrome to only occasional patients with the infertile male syndrome. The disorder in subjects with a negative family history is believed to be the result of new mutations.

Hormone dynamics are similar in all disorders of the androgen receptor. Plasma testosterone levels and rates of testosterone production by the testes are normal or high. Elevated testosterone production is caused by the high mean plasma level of LH, which, in turn, is due to defective feedback regulation caused by resistance to the action of androgen at the hypothalamic-pituitary level. Elevated LH concentration is responsible also for the increased estrogen production by the testes (Chap. 337). (In normal men, most estrogen is derived from peripheral formation from circulating androgens, but when the plasma LH level is elevated, the testes secrete increased amounts of estrogen into the circulation.) Thus resistance to the feedback regulation of LH secretion by circulating androgen results in elevated plasma LH levels, and this, in turn, results in the enhanced secretion of both testosterone and estradiol by the testes. Gonadotropin levels rise even higher (and menopausal symptoms may develop) when the testes are removed, indicating that gonadotropin secretion is under partial feedback control. Presumably, in the steady state and in the absence of an androgen effect, estrogen alone regulates LH secretion, a control purchased at the expense of an elevated plasma estrogen concentration for a male. The hormonal changes in the infertile male syndrome are similar to those in the other receptor disorders but less marked. Some men with this syndrome do not have an elevation of plasma LH or plasma testosterone level.

Feminization in these disorders is the result of two interlocking phenomena. First, androgens and estrogens have antagonistic effects, and in the absence of androgen action, the cellular effect of estrogen is unopposed. Second, the testicular production of estradiol is greater than that of the normal male (although less than that of the normal female). Variable degrees of androgen resistance and enhanced estradiol production result in different degrees of defective virilization and enhanced feminization in the four clinical syndromes.

Each of these syndromes is the result of defects in the androgen receptor. In most families, the fundamental defect is due to point mutations in the coding sequence leading to premature termination codons or to amino acid substitutions in the hormone-binding domain. Such mutations impair receptor function to variable degrees. Some families with clinical syndromes typical of an androgen receptor disorder have normal androgen binding in fibroblasts; in most, point mutations in the DNA-binding domain of the androgen receptor are responsible for the androgen resistance.

TREATMENT

Individuals with 5a-reductase deficiency who are raised as females but elect at the time of expected puberty to change social sex to male or who are raised from the first as males should be monitored carefully and given supplemental androgens, preferably those such as nandrolone decanoate that do not require 5a-reduction for activation,

when virilization is incomplete. Fertility has been reported in such an individual. Individuals with 5a-reductase deficiency who continue to function as females should be gonadectomized, given feminizing doses of estrogens indefinitely, and receive surgical correction of the introitus when appropriate. The management of subjects with androgen receptor defects depends on the phenotypic manifestations. Women with testicular feminization should be castrated (preferably after the completion of the pubertal growth spurt and the feminization of the breasts) to prevent tumor development in the testes and receive estrogen replacement to maintain feminization, prevent hot flashes, and protect the bones; shallow vaginal depth can usually be treated medically with the Frank technique. Men with the Reifenstein phenotype should have surgical correction of the hypospadias and may require surgery for gynecomastia; supplemental androgen therapy in these men rarely improves the incomplete virilization.

Persistent Mullerian Duct Syndrome Men with this uncommon disorder have testes and male phenotypic development and in addition have fallopian tubes and a uterus. In some, one or both testes are descended and the uterus and ipsilateral fallopian tube are in the inguinal canal or scrotum; both testes and fallopian tubes may be present in the hernia sac or can be drawn into it. In others, the testes are located high in the abdomen and hernias are not present. In both types, the vasa deferentia are embedded in the wall of the uterus, a feature that complicates surgical procedures designed to preserve potential fertility. Most individuals have uninformative family histories, but in some the condition is inherited as an autosomal recessive trait. Because the external genitalia are well developed and the subjects masculinize normally at puberty, it is assumed that during the critical stage of embryonic differentiation the fetal testes produce a normal amount of androgen. However, mullerian regression does not occur. Two types of mutations have been described in this disorder. In one the gene that encodesAMH is defective, and blood levels of AMH are usually low or undetectable; in the other the AMH receptor is defective, and blood levels of AMH are elevated. To minimize the chance of tumor development and to maintain virilization, orchiopexy should be performed. Malignancy in the uterus or vagina has not been described, and because the vasa deferentia are closely associated with the broad ligaments, the uterus and vagina should be left in place to avoid disruption of the vasa deferentia during removal and consequently to preserve possible fertility.

Developmental Defects of the Male Genitalia

Hypospadias Hypospadias is a congenital anomaly in which the urethra terminates in an abnormal position along the ventral midline of the penis at some site between the normal urethral meatus and the perineum. This malformation occurs in 0.5 to 0.8% of male births in the United States and is often associated with ventral contraction and bowing of the penis (chordee). It is common to categorize hypospadias as glandular (involving the glans penis), penile, or perineoscrotal. Since androgens control penile development, hypospadias is generally assumed to result from some unidentified defect in androgen formation or action during embryogenesis. Indeed, hypospadias occurs in most disorders of male sexual differentiation, and a rare cause of hypospadias is maternal ingestion of progestational agents early in pregnancy. However, the known causes (single-gene defects, chromosomal abnormalities, and maternal drug ingestion) account for only about one-fourth of cases, and the etiology of most is unknown. The management is surgical.

Cryptorchidism The control of testicular descent is poorly understood, both in regard to the nature of the forces that cause the movement and to the hormonal factors that regulate the process. In anatomic terms, testicular descent can be divided into three phases: (1) transabdominal movement of the testis from its site of origin above the kidney to the inguinal ring, (2) formation of the opening in the inguinal canal (processus vaginalis) through which the testis exits the abdominal cavity, and (3) actual movement of the testis through the inguinal canal to its permanent site in the scrotum. This process occurs over a 6- to 7-month period during gestation, beginning at about the sixth week, and is not completed in some until after birth. Impairment at any stage in this process can impair descent of one or both testes. About 3% of full-term and 30% of premature male infants have at least one cryptorchid testis at birth, but descent is usually completed within the first few weeks of life, so that the incidence of failure of descent by 6 to 9 months of age is only 0.6 to 0.7%. It is this latter category of maldescent that requires intervention.

Permanent cryptorchidism can be classified as intraabdominal (10%), canalicular (in the inguinal canal) (20%), high scrotal (40%), or obstructed (30%), in which maldescent is due to a physical barrier between the inguinal pouch and the inlet of the scrotum. These disorders must be distinguished from the temporarily retracted normal testis.

The cryptorchid testis functions poorly after puberty, but the extent to which maldescent is the result of an abnormality of the testis or the cause of abnormal function is unknown. Two general theories have been advanced as to the etiology -- inadequate intraabdominal pressure and deficient endocrine function of the testis either because of deficient testosterone synthesis or inadequate formation of AMH. Indeed, defects that result in inadequate development of intraabdominal pressure or inadequate development of the testes can cause cryptorchidism. As in hypospadias, however, the known causes of cryptorchidism constitute only a small fraction of the cases, and the etiology in most remains to be identified. Two complications of cryptorchidism are important; spermatogenesis cannot occur at the temperature of the abdominal cavity, and it is necessary to correct the process as early as possible to allow possible fertility. The fact that infertility is common in men who have been treated for unilateral as well as bilateral cryptorchidism suggests that maldescent is usually the consequence rather than the cause of the testicular malfunction. There is also a greater frequency of malignancy in undescended testes, which should be surgically corrected for this reason (Chap. 96).

(Bibliography omitted in Palm version)

339. DISORDERS AFFECTING MULTIPLE ENDOCRINE SYSTEMS- *Steven I. Sherman, Robert F. Gagel*

NEOPLASTIC DISORDERS AFFECTING MULTIPLE ENDOCRINE ORGANS

Several distinct genetic disorders predispose to endocrine gland neoplasia and cause hormone excess syndromes (Table 339-1). DNA-based genetic testing is now available for these disorders, but effective management requires an understanding of endocrine neoplasia and the range of clinical features that may be manifest in an individual patient.

MULTIPLE ENDOCRINE NEOPLASIA (MEN) TYPE 1

Clinical Manifestations MEN 1, or Wermer's syndrome, is characterized by neoplasia of parathyroid, pituitary, and pancreatic islet cells (Table 339-1). The syndrome is inherited as an autosomal dominant trait, so that each child of an affected parent has a 50% chance of inheriting the predisposing gene.

Several features of the pathogenesis of MEN 1 have important implications for its management. Though each tumor is derived from a single cell (clonal in origin), any endocrine cell within the affected organs can become transformed. Hyperplasia is the initiating lesion, followed later by adenomatous or carcinomatous changes. Consequently, the disease process within a single organ is multicentric. Neoplasia in one organ may affect the progression of disease in another organ. For example, the ectopic production of hypothalamic releasing hormones by a pancreatic tumor may stimulate the growth of a pituitary tumor. Because MEN 1 generally evolves over a 30- to 40-year period, the manifestations depend in part on when the disorder is identified.

Hyperparathyroidism is the most common manifestation of MEN 1. Hypercalcemia may be present during the teenage years, and most individuals are affected by age 40 (Fig. 339-1). Screening for hyperparathyroidism involves measurement of either an albumin-adjusted or ionized serum calcium level. The diagnosis is established by demonstrating elevated levels of serum calcium and intact parathyroid hormone. Manifestations of hyperparathyroidism in MEN 1 do not differ substantially from those in sporadic hyperparathyroidism and include calcium-containing kidney stones, bone abnormalities, and gastrointestinal and musculoskeletal complaints (Chap. 341).

Other familial disorders associated with hypercalcemia include familial parathyroid hyperplasia, familial adenomatous hyperparathyroidism, and familial hypocalciuric hypercalcemia (FHH). Calcium excretion is usually elevated in the patient with MEN 1 or other forms of primary hyperparathyroidism and low in FHH. Another distinguishing feature is that the serum calcium level is rarely elevated at birth in patients with MEN 1 but is frequently elevated in newborns with FHH. Differentiation of hyperparathyroidism of MEN 1 from other forms of familial primary hyperparathyroidism is usually based on family history, histologic features of resected parathyroid tissue, and, sometimes, long-term observation to determine whether other manifestations of MEN 1 develop. FHH is due to inactivating mutations of the calcium sensor, a transmembrane G protein-coupled receptor found in parathyroid tissue and kidney (Chap. 341).

Parathyroid hyperplasia is the common cause of hyperparathyroidism in MEN 1,

although single and multiple adenomas have been described. Hyperplasia of one or more parathyroid glands is common in younger patients; adenomas are usually found in older patients or those with long-standing disease.

Neoplasia of the pancreatic islets is the second most common manifestation of MEN 1 and tends to occur in parallel with hyperparathyroidism (Fig. 339-1). Increased pancreatic islet cell hormones include pancreatic polypeptide (75 to 85%), gastrin [60%; Zollinger-Ellison syndrome (ZES)], insulin (25 to 35%), vasoactive intestinal peptide (VIP) (3 to 5%; Verner-Morrison or watery diarrhea syndrome), glucagon (5 to 10%), and somatostatin (1 to 5%). The tumors rarely produce adrenocorticotropin (ACTH), corticotropin-releasing hormone (CRH), growth hormone-releasing hormone (GHRH), calcitonin gene products, neurotensin, gastric inhibitory peptide, and others. Many of the tumors produce more than one peptide. The pancreatic neoplasms differ from the other components of MEN 1 in that approximately one-third of the tumors display malignant features, including hepatic metastases (Chap. 93).

Pancreatic islet cell tumors are diagnosed by identification of a characteristic clinical syndrome, hormonal assays with or without provocative stimuli, or radiographic techniques. One approach involves annual screening of people at risk with measurement of basal and meal-stimulated levels of pancreatic polypeptide to identify the tumors as early as possible; the rationale of this screening strategy is the concept that surgical removal of islet cell tumors at an early stage will be curative. Other approaches to screening include measurement of serum gastrin and pancreatic polypeptide levels every 2 to 3 years, with the rationale that pancreatic neoplasms will be detected at a later stage but can be managed medically, if possible, or by surgery. High-resolution, early-phase computed tomography (CT) scanning provides the best noninvasive technique for identification of these tumors, but intraoperative ultrasonography is the most sensitive method for detection of small tumors.

ZES is caused by excessive gastrin production and occurs in more than half of MEN 1 patients with pancreatic islet cell tumors (Fig. 339-1) (Chap. 93). Clinical features include increased gastric acid production, recurrent peptic ulcers, diarrhea, and esophagitis. The ulcer diathesis is refractory to conservative therapy such as antacids. The diagnosis is made by finding increased gastric acid secretion, elevated basal gastrin levels in serum [generally >115 pmol/L (200 pg/mL)], and an exaggerated response of serum gastrin to either secretin or calcium. Other causes of elevated serum gastrin levels, such as achlorhydria, treatment with H_2 receptor antagonists or omeprazole, retained gastric antrum, small-bowel resection, gastric outlet obstruction, and hypercalcemia, should be excluded. Gastrin-producing carcinoid-like tumors are frequently present in the duodenal wall.

Insulinoma causes hypoglycemia in about one-third of MEN 1 patients with pancreatic islet cell tumors (Fig. 339-1). The tumors may be benign or malignant (25%). The diagnosis can be established by documenting hypoglycemia during a short fast with simultaneous inappropriate elevation of serum insulin and C-peptide levels. More commonly, it is necessary to subject the patient to a supervised 72-h fast to provoke hypoglycemia (Chap. 334). Large insulinomas may be identified by CT scanning; small tumors not detected by radiographic techniques may be localized by selective arteriographic injection of calcium into each of the arteries that supply the pancreas and

sampling the hepatic vein for insulin to determine the anatomic region containing the tumor. Intraoperative ultrasonography can also be used to localize these tumors, but preoperative calcium injection data are helpful in guiding the subtotal pancreatectomy if multiple or no abnormalities are detected by intraoperative ultrasonography.

Glucagonoma in occasional MEN 1 patients causes a syndrome of hyperglycemia, skin rash (necrolytic migratory erythema), anorexia, glossitis, anemia, depression, diarrhea, and venous thrombosis. In about half of these patients the plasma glucagon level is high, leading to its designation as the *glucagonoma syndrome*, although elevation of plasma glucagon level in MEN 1 patients is not necessarily associated with these symptoms. The glucagonoma syndrome may represent a complex interaction between glucagon overproduction and the nutritional status of the patient.

The *Verner-Morrison* or *watery diarrhea syndrome* consists of watery diarrhea, hypokalemia, hypochlorhydria, and metabolic acidosis. The diarrhea can be voluminous and is almost always found in association with an islet cell tumor, prompting use of the term *pancreatic cholera*. However, the syndrome is not restricted to pancreatic islet tumors and has been observed with carcinoids or other tumors. This syndrome is believed to be due to overproduction of VIP, although plasma VIP levels may not be elevated. Hypercalcemia may be induced by the effects of VIP on bone as well as by hyperparathyroidism.

Pituitary tumors occur in more than half of patients with MEN 1 and tend to be multicentric, making them difficult to resect (Chap. 328). Prolactinomas are most common (Fig. 339-1) and are diagnosed by finding serum prolactin levels >200 ug/L, with or without a pituitary mass evident by magnetic resonance imaging (MRI). Values<200 ug/L may be due to a prolactin-secreting neoplasm or to compression of the pituitary stalk by a different type of pituitary tumor. Acromegaly due to excessive growth hormone (GH) production is the second most common syndrome caused by pituitary tumors in MEN 1 (Chap. 328) but can rarely be due to production of GHRH by an islet cell tumor. Cushing's disease can be caused by ACTH-producing pituitary tumors or by ectopic production of ACTH or CRH by other tumors in the MEN 1 syndrome. Diagnosis of pituitary Cushing's disease is generally best accomplished by a high-dose dexamethasone suppression test or by petrosal venous sinus sampling for ACTH after intravenous injection of CRH (Chap. 328). Differentiation of a primary pituitary tumor from an ectopic CRH-producing tumor may be difficult because the pituitary is the source of ACTH in both disorders; documentation of CRH production by a pancreatic islet or carcinoid tumor may be the only method of proving ectopic CRH production. Adrenal cortical tumors are found in almost one-half of gene carriers but are rarely functional; malignancy in the cortical adenomas is uncommon.

Unusual manifestations of MEN 1 The rare carcinoid tumors in MEN 1 are of the foregut type and are derived from thymus, lung, stomach, or duodenum; they may metastasize or be locally invasive. These tumors usually produce serotonin, calcitonin, or CRH; the typical carcinoid syndrome with flushing, diarrhea, and bronchospasm is rare (Chap. 93). Subcutaneous or visceral lipomas and cutaneous leiomyomas may also be present but rarely undergo malignant transformation. Skin angiofibromas or collagenomas are seen in most patients with MEN 1 when carefully sought.

GENETIC CONSIDERATIONS

MEN1 is transmitted as an autosomal dominant trait, reflecting the fact that the *MEN1* gene, located on chromosome 11q13, encodes a tumor suppressor protein termed *menin* (Fig. 339-2). Affected individuals typically harbor a germline mutation in *MEN1* and acquire a "second hit" in the normal gene as a result of another mutation or, more commonly, loss of the portion of chromosome 11 that contains the *MEN1* locus (Chap. 81). Though the function of menin is not well understood, it is a nuclear protein that interacts with a transcriptional factor, Jun D, suggesting a role in cell growth control. Several missense mutations in menin prevent its interaction with Jun D.

MEN1 gene mutations are found in>90% of families with the syndrome (Fig. 339-2). Genetic testing can be performed in individuals at risk for the development ofMEN 1, particularly when the specific mutation is known. The value of genetic testing for this disorder, in contrast to MEN 2 (see below), is debated because predisposed individuals must still be screened repeatedly using endocrine tests. A negative genetic analysis will, however, exclude disease with near 100% certainty in kindreds with a known mutation; for this reason, genetic testing is likely to gain favor as it becomes more widely available. A significant percentage of sporadic parathyroid, islet cell, and carcinoid tumors also have loss or mutation of *MEN1*. It is presumed that these mutations are somatic and occur in a single cell, leading to subsequent transformation.

TREATMENT

Almost everyone who inherits a mutant *MEN1* gene develops at least one clinical manifestation of the syndrome. Most develop hyperparathyroidism, 80% develop pancreatic islet cell tumors, and more than half develop pituitary tumors. For most of these tumors, initial surgery is not curative and patients frequently require multiple surgical procedures and surgery on two or more endocrine glands during a lifetime. For this reason, it is essential to establish clear goals for management of these patients rather than to recommend surgery casually each time a tumor is discovered. Ranges for acceptable management are discussed below.

Hyperparathyroidism Individuals with serum calcium levels >3.0 mmol/L (12 mg/dL), evidence of calcium nephrolithiasis or renal dysfunction, neuropathic or muscular symptoms, or bone involvement (including osteopenia) should undergo parathyroid exploration. InMEN 1 an additional criterion for parathyroid surgery is hypercalcemia associated with elevated gastrin levels, because elevated serum calcium may stimulate gastrin production andZES, a condition that may be improved by return of calcium levels to normal. There is less agreement regarding the necessity for parathyroid exploration in individuals who do not meet these criteria, and observation may be appropriate in the MEN 1 patient with asymptomatic hyperparathyroidism.

When parathyroid surgery is indicated inMEN 1, all parathyroid tissue should be identified and removed at the time of primary operation, and parathyroid tissue should be implanted in the nondominant forearm. Thymectomy should also be performed because of the potential for later development of malignant carcinoid tumors. If reoperation is necessary, transplanted tissue can be resected under local anesthesia with titration of tissue removal to return the serum calcium level to normal. A less

desirable approach is to remove 3 to $3\frac{1}{2}$ parathyroid glands from the neck, carefully marking the location of residual tissue so that the remaining tissue can be located easily during subsequent surgery.

Pancreatic Islet Tumors (See Chap. 93 for discussion of pancreatic islet tumors not associated with MEN 1.) Two features of pancreatic islet cell tumors in MEN 1 complicate the management. First, the pancreatic islet cell tumors are multicentric, malignant about a third of the time, and cause death in 10 to 20% of patients. Second, removal of all pancreatic islets to prevent malignancy causes diabetes mellitus, a disease with severe long-term complications. These features make it difficult to formulate clear-cut guidelines, but some general concepts appear to be valid. First, islet cell tumors producing insulin, glucagon, VIP, GHRH, or CRH should be resected because medical therapy is generally ineffective. Second, gastrin-producing islet cell tumors that cause ZES are frequently multicentric. Recent experience suggests that a high percentage of ZES in MEN 1 is caused by duodenal wall tumors and that resection of these tumors improves the cure rate. Treatment with H_2 receptor antagonists (cimetidine or ranitidine) and the H_+,K_+-ATPase inhibitors (omeprazole or lansoprazole) provides an alternative to surgery for control of ulcer disease in patients with multicentric tumors or with hepatic metastases. Third, in families in which there is a high incidence of malignant islet cell tumors that cause death, total pancreatectomy at an early age may be justified to prevent malignancy.

Management of metastatic islet cell carcinoma is unsatisfactory. Hormonal abnormalities can sometimes be controlled. For example, ZES can be treated with H_2 receptor antagonists or H_+,K_+-ATPase inhibitors; the somatostatin analogue, octreotide, is useful in the management of carcinoid and the watery diarrhea syndrome. Bilateral adrenalectomy may be required for ectopic ACTH syndrome if medical therapy is ineffective (Chap. 331). Islet cell carcinomas frequently metastasize to the liver but may grow slowly. Hepatic artery embolization or chemotherapy (5-fluorouracil, streptozocin, chlorozotocin, doxorubicin, or dacarbazine) may reduce tumor mass, control symptoms of hormone excess, and prolong life; however, these treatments are never curative.

Pituitary Tumors Treatment of prolactinomas with dopamine agonists (bromocriptine, cabergoline, or quinagolide) usually returns the serum prolactin level to normal and prevents further tumor growth (Chap. 328). Surgical resection of a prolactinoma is rarely curative but may relieve mass effects. Transsphenoidal resection is appropriate for neoplasms that secrete ACTH, GH, or the a-subunit of the pituitary glycoprotein hormones. Octreotide reduces tumor mass in one-third of GH-secreting tumors and reduces GH and insulin-like growth factor I levels in >75% of patients. Radiation therapy may be useful for large or recurrent tumors.

Improvements in the management of MEN 1, particularly islet cell and pituitary tumors, have improved outcome in these patients substantially. As a result, other neoplastic manifestations, such as carcinoid syndrome, are now seen with increased frequency.

MULTIPLE ENDOCRINE NEOPLASIA TYPE 2

Clinical Manifestations Medullary thyroid carcinoma (MTC) and pheochromocytoma

are associated in two major syndromes: MEN type 2A and MEN type 2B (Table 339-1). MEN 2A is the combination of MTC, hyperparathyroidism, and pheochromocytoma. Three subvariants of MEN 2A are familial medullary thyroid carcinoma (FMTC), MEN 2A with cutaneous lichen amyloidosis, and MEN 2A with Hirschsprung disease. MEN type 2B is the combination of MTC, pheochromocytoma, mucosal neuromas, intestinal ganglioneuromatosis, and marfanoid features.

Multiple Endocrine Neoplasia Type 2A MTC is the most common manifestation. This tumor usually develops in childhood, beginning as hyperplasia of the calcitonin-producing cells (C cells) of the thyroid. MTC is typically located at the junction of the upper one-third and lower two-thirds of each lobe of the thyroid, reflecting the high density of C cells in this location; tumors >1 cm in size are frequently associated with local or distant metastases. Measurement of the serum calcitonin level after calcium or pentagastrin injection makes it possible to diagnose this disorder when the likelihood of metastasis is low (see below).

Pheochromocytoma occurs in approximately 50% of patients with MEN 2A and causes palpitations, nervousness, headaches, and sometimes sweating (Chap. 332). About half the tumors are bilateral, and >50% of patients who have had unilateral adrenalectomy develop a pheochromocytoma in the contralateral gland within a decade. A second feature of these tumors is a disproportionate increase in the secretion of epinephrine relative to norepinephrine. Capsular invasion is common, but malignant behavior is uncommon.

Hyperparathyroidism occurs in 15 to 20% of patients, with the peak incidence in the third or fourth decade. The manifestations of hyperparathyroidism do not differ from those in other forms of primary hyperparathyroidism (Chap. 341), with nephrolithiasis being common. Diagnosis is established by finding hypercalcemia, hypophosphatemia, hypercalciuria, and an inappropriately high serum level of intact parathyroid hormone. Multiglandular parathyroid hyperplasia is the most common histologic finding, although with long-standing disease adenomatous changes may be superimposed on hyperplasia.

Multiple Endocrine Neoplasia Type 2B The association of MTC, pheochromocytoma, mucosal neuromas, and a marfanoid habitus is designated MEN 2B. MTC in MEN 2B develops earlier and is more aggressive than in MEN 2A. Metastatic disease has been described prior to 1 year of age, and death commonly occurs in the second or third decade of life. However, the prognosis is not invariably bad even in patients with metastatic disease, as evidenced by a number of multigenerational families with this disease.

Pheochromocytoma occurs in more than half of MEN 2B patients and does not differ from that in MEN 2A. Hypercalcemia is rare in MEN 2B, and there are no well-documented examples of hyperparathyroidism.

The mucosal neuromas and marfanoid body habitus are the most distinctive features and are recognizable in childhood. Neuromas are present on the tip of the tongue, under the eyelids, and throughout the gastrointestinal tract and are true neuromas, distinct from neurofibromas. Children may present with gastrointestinal symptoms,

including increased gas, intermittent obstruction, and diarrhea caused by neuromas.

GENETIC CONSIDERATIONS

Mutations of the *RET* proto-oncogene have been identified in 93 to 95% of patients with MEN 2 (Fig. 339-3). *RET* encodes a tyrosine kinase receptor that is normally activated by glial cell line-derived neurotropic factor. *RET* mutations induce constitutive activity of the receptor, explaining the autosomal dominant transmission of the disorder.

Naturally occurring mutations localize to two regions of the RET tyrosine kinase receptor. The first is a cysteine-rich extracellular domain; point mutations in the coding sequence for one of five cysteines (codons 609, 611, 618, 620, or 634) cause amino acid substitutions that induce receptor dimerization and activation in the absence of its ligand. Codon 634 mutations occur in 80% of MEN 2A kindreds and are most commonly associated with classic MEN 2A features (Figs. 339-3 and 339-2); an arginine substitution at this codon accounts for half of all MEN 2A mutations. All reported families with MEN 2A and cutaneous lichen amyloidosis are consistently associated with a codon 634 mutation. Mutations of codons 609, 611, 618, or 620 occur in 10 to 15% of MEN 2A kindreds and are more commonly associated with FMTC (Fig. 339-3). Mutations in codons 609, 618, and 620 have also been identified in MEN 2A and in the Hirschsprung variant (Fig. 339-3).

The second region of the RET tyrosine kinase that is mutated in MEN 2 is in the substrate recognition pocket at codon 918 (Fig. 339-3). This activating mutation is present in approximately 95% of patients with MEN 2B and accounts for 10 to 15% of all *RET* proto-oncogene mutations in MEN 2. Mutations of codon 883 and 22 have also been identified in a few patients with MEN 2B.

From 3 to 5% of kindreds with FMTC have no identifiable mutation of either of these regions. In a few such kindreds mutations of codons 768, 790, 791, 804, and 891 have been identified (Fig. 339-3).

Somatic mutations (found only in the tumor and not transmitted in the germline) of the *RET* proto-oncogene have been identified in sporadic MTC; 25 to 35% of sporadic tumors have codon 918 mutations, and somatic mutations in codons 630, 768, and 804 have also been identified (Fig. 339-3). Germline mutations of the *RET* proto-oncogene are present in about 6% of patients with apparent sporadic MTC, indicating that other family members may be at risk for the disease.

TREATMENT

Screening for Multiple Endocrine Neoplasia Type 2 Death from MTC can be prevented by early thyroidectomy. The identification of *RET* proto-oncogene mutations and the application of DNA-based molecular diagnostic techniques to identify these mutations has simplified the screening process. During the initial evaluation of a kindred, a *RET* proto-oncogene analysis should be performed on an individual with proven MEN 2A. Establishment of the specific mutation in a kindred facilitates the subsequent analysis of other family members. Each family member at risk should be tested twice for the presence of the specific mutation; the second analysis should be

performed on a new DNA sample and, ideally, in a second laboratory to exclude sample mix-up or technical error (see *endrcr06.mda.uth.tmc.edu* for a list of laboratory testing sites). Individuals in a kindred with a known mutation who have two normal analyses can be excluded from further screening.

There is general consensus that children with codon 883, 918, and 922 mutations, or those associated with MEN 2B, should have a total thyroidectomy and central lymph node dissection (level VI) performed during the first months of life or soon after identification of the syndrome. If local metastasis is discovered, a more extensive lymph node dissection (levels II to V) is generally indicated. In children with codon 611, 618, 620, 630, 634, and 891 mutations, thyroidectomy should be performed before the age of 6 years because of reports of local metastatic disease in children this age. Finally, there are kindreds with codon 609, 768, 790, 791, and 804 mutations where the phenotype of MTC appears to be less aggressive. In these kindreds, and in those with rare mutations, two management approaches have been suggested in association with genetic counseling: (1) perform a total thyroidectomy with or without central node dissection at some arbitrary age (perhaps 6 to 12 years of age), or (2) continue annual or biannual provocative testing for calcitonin release with performance of total thyroidectomy with or without central neck dissection when the test becomes abnormal. The pentagastrin test involves measurement of serum calcitonin basally and 2, 5, 10, and 15 min after a bolus injection of 5 ug pentagastrin per kilogram body weight. Patients should be warned before injection of epigastric tightness, nausea, warmth, and tingling of extremities and reassured that the symptoms will last approximately 2 min. The recent unavailability of pentagastrin in the United States has led to use of a short calcium infusion, performed by obtaining a baseline serum calcitonin and then infusing 150 mg calcium salt intravenously over 10 min with measurement of serum calcitonin at 5, 10, 15, 30 min after initiation of the infusion.

The *RET* proto-oncogene analysis should be performed in patients with suspected MEN 2B to detect codon 883, 918, and 922 mutations, especially in newborn children where the diagnosis is suspected but the clinical phenotype is not fully developed. Other family members at risk for MEN 2B should also be tested because the mucosal neuromas can be subtle and not always identified. In the rare families with proven germline transmission of MTC but no identifiable *RET* proto-oncogene mutation, annual pentagastrin or calcium-pentagastrin testing should be performed on members at risk.

Annual screening for pheochromocytoma in subjects with germline *RET* mutations should be performed by measuring basal plasma or 24-h urine catecholamines and metanephrines. The goal is to identify a pheochromocytoma before it causes significant symptoms or is likely to cause sudden death, an event most commonly associated with large tumors. Although there are kindreds with FMTC and specific *RET* mutations in which no pheochromocytomas have been identified (Fig. 339-3), it is not clear that a large enough experience has been gained to exclude pheochromocytoma screening in these individuals. Radiographic studies, such as MRI or CT scans, are generally reserved for individuals with abnormal screening tests or with symptoms suggestive of pheochromocytoma (Chap. 332). Women should be tested during pregnancy because undetected pheochromocytoma can cause maternal death during childbirth.

Measurement of serum calcium and parathyroid hormone levels every 2 to 3 years

provides an adequate screen for hyperparathyroidism, except in those families in which hyperparathyroidism is a prominent component, where measurements should be made annually.

Treatment of Medullary Thyroid Carcinoma MTC is a multicentric disorder. Total thyroidectomy with a central lymph node dissection should be performed in children who carry the mutant genes. Incomplete thyroidectomy leaves the possibility of later transformation of residual long-term C cells. The goal of early therapy is cure, and a strategy that does not accomplish this goal is short-sighted. Long-term follow-up studies indicate an excellent outcome with approximately 90% of children free of disease 15 to 20 years after surgery. In contrast, 15 to 25% of patients in whom the diagnosis is made on the basis of a palpable thyroid nodule die from the disease within 15 to 20 years.

In adults with MTC >1 cm in size, metastases to regional lymph nodes are common. Total thyroidectomy with central lymph node dissection and selective dissection of other regional chains provide the best chance for cure. In patients with extensive local metastatic disease in the neck, external radiation may prevent local recurrence or reduce tumor mass but is not curative. Chemotherapy with combinations of adriamycin, vincristine, cyclophosphamide, and dacarbazine may provide palliation.

Treatment of Pheochromocytoma The long-term goal for management of pheochromocytoma is to prevent death and cardiovascular complications. Improvements in radiographic imaging of the adrenals make direct examination of the apparently normal contralateral gland during surgery less important, and the rapid evolution of laparoscopic surgery has simplified management of early pheochromocytoma. The major question is whether to remove both adrenal glands or to remove only the affected adrenal at the time of primary surgery. Issues to be considered in making this decision include the possibility of malignancy (<15 reported cases), the likelihood of developing pheochromocytoma in the apparently unaffected gland over an 8- to 10-year period, and the risks of adrenal insufficiency caused by removal of both glands (at least two deaths related to adrenal insufficiency in MEN 2 patients). Most clinicians recommend removing only the affected gland. If both adrenals are removed, glucocorticoid and mineralocorticoid replacement is mandatory. An alternative approach is to remove the pheochromocytoma and adrenal medulla, leaving the adrenal cortex behind. This approach is usually successful and eliminates the necessity for steroid hormone replacement, although the pheochromocytoma recurs in some.

Treatment of Hyperparathyroidism Hyperparathyroidism has been managed by one of two approaches. Removal of $3\frac{1}{2}$ glands with maintenance of the remaining half gland in the neck is the usual procedure. In families in whom hyperparathyroidism is a prominent manifestation (almost always associated with a codon 634 *RET* mutation) and recurrence is common, total parathyroidectomy with transplantation of parathyroid tissue into the nondominant forearm is preferred. This approach is discussed above in the context of hyperparathyroidism associated with MEN1.

OTHER GENETIC TUMOR SYNDROMES

A number of mixed syndromes exist in which the neoplastic associations differ from those in MEN 1 or 2 (Table 339-1).

The cause of von Hippel-Lindau (VHL) syndrome, the association of central nervous system tumors, renal cell carcinoma, pheochromocytoma, and islet cell neoplasms, is mutations in the *VHL* tumor-suppressor gene. Germline-inactivating mutations of the *VHL* gene cause tumor formation when there is additional loss or somatic mutation of the normal *VHL* allele in brain, kidney, pancreatic islet, or adrenal medullary cells. A specific subset of mutations is more common in families with pheochromocytomas.

The molecular defect in type 1 neurofibromatosis inactivates neurofibromin, a cell membrane-associated protein that normally activates a GTPase. Inactivation of this protein impairs GTPase and causes continuous activation of p21 Ras and its downstream tyrosine kinase pathway. Endocrine tumors also form in less common neoplastic genetic syndromes. These include Cowden's disease, Carney complex, familial acromegaly, and familial carcinoid syndrome.

IMMUNOLOGIC SYNDROMES AFFECTING MULTIPLE ENDOCRINE ORGANS

When immune dysfunction affects two or more endocrine glands and other nonendocrine immune disorders are present, the *polyglandular autoimmune* (PGA) *syndromes* should be considered. The PGA syndromes are classified as two main types: the type I syndrome starts in childhood and is characterized by mucocutaneous candidiasis, hypoparathyroidism, and adrenal insufficiency; the type II, or *Schmidt syndrome*, is more likely to present in adults and most commonly comprises adrenal insufficiency, thyroiditis, and type 1 diabetes mellitus. However, the type II syndrome is heterogeneous and may consist of autoimmune thyroid disease along with a variety of other autoimmune endocrine disorders ([Table 339-2](#)).

POLYGLANDULAR AUTOIMMUNE SYNDROME TYPE I

[PGA](#)type I is usually recognized in the first decade of life and requires two of three components for diagnosis: mucocutaneous candidiasis, hypoparathyroidism, and adrenal insufficiency. Mineralocorticoids and glucocorticoids may be lost simultaneously or sequentially. This disorder is also called *autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy* (APECED). Other endocrine defects can include gonadal failure, hypothyroidism, anterior hypophysitis, and, less commonly, destruction of the bcells of the pancreatic islets and development of insulin-dependent (type 1) diabetes mellitus. Additional features include hypoplasia of the dental enamel, ungual dystrophy, tympanic membrane sclerosis, vitiligo, keratopathy, and gastric parietal cell dysfunction resulting in pernicious anemia. Some patients develop autoimmune hepatitis, malabsorption (variably attributed to intestinal lymphangiectasia, IgA deficiency, bacterial overgrowth, or hypoparathyroidism), asplenism, achalasia, and cholelithiasis ([Table 339-2](#)). At the outset, only one organ may be involved, but the number increases with time so that patients eventually manifest two to five components of the syndrome.

Most patients initially present with oral candidiasis in childhood; it is poorly responsive to treatment and relapses frequently. Chronic hypoparathyroidism usually occurs before adrenal insufficiency develops. More than 60% of postpubertal women develop premature hypogonadism. The endocrine components, including adrenal insufficiency

and hypoparathyroidism, may not develop until the fourth decade, making continued surveillance necessary.

Type I PGA syndrome allows no HLA associations and is inherited as an autosomal recessive trait. The responsible gene, designated as either *APECED* or *AIRE*, encodes a transcription factor that is expressed in thymus and lymph nodes; a variety of different mutations have been reported.

POLYGLANDULAR AUTOIMMUNE SYNDROME TYPE II

PGA type II is characterized by two or more of the endocrinopathies listed in Table 339-2. Most often these include primary adrenal insufficiency, Graves' disease or autoimmune hypothyroidism, type 1 diabetes mellitus, and primary hypogonadism. Because adrenal insufficiency is relatively rare, it is frequently used to define the presence of the syndrome. Among patients with adrenal insufficiency, type 1 diabetes mellitus coexists in 52% and autoimmune thyroid disease occurs in 69%. However, many patients with antimicrosomal and antithyroglobulin antibodies never develop abnormalities of thyroid function. Thus, increased antibody titers alone are poor predictors of future disease. Other associated conditions include hypophysitis, celiac disease, atrophic gastritis, and pernicious anemia. Vitiligo, caused by antibodies against the melanocyte (see Plate IIA-11), and alopecia are less common than in the type I syndrome. Mucocutaneous candidiasis does not occur. A few patients develop a late-onset, usually transient hypoparathyroidism caused by antibodies that compete with parathyroid hormone for binding to the parathyroid hormone receptor. Up to 25% of patients with myasthenia gravis, and an even higher percentage who have myasthenia and a thymoma, have PGA type II (Chap. 380).

The type II syndrome is familial in nature but does not exhibit a characteristic Mendelian pattern of transmission. Like many of the individual autoimmune endocrinopathies, certain HL-DR3 and HLA-DR4 alleles increase disease susceptibility; several different genes probably contribute to the expression of this syndrome.

A variety of autoantibodies are seen in PGA type II, including antibodies directed against: (1) thyroid antigens such as thyroid peroxidase, thyroglobulin, or the thyroid stimulating hormone (TSH) receptor; (2) adrenal side chain cleavage enzyme, steroid 21-hydroxylase, or ACTH receptor; and (3) pancreatic islet glutamic acid decarboxylase or the insulin receptor, among others. The roles of cytokines such as interferon and cell-mediated immunity are unclear.

DIAGNOSIS

The clinical manifestations of adrenal insufficiency often develop slowly, may be difficult to detect, and can be fatal if not diagnosed and treated appropriately. Thus, prospective screening should be performed routinely in all patients and family members at risk for PGA types I and II. The most effective screening test for adrenal disease is a cosyntropin stimulation test (Chap. 331). A fasting blood glucose level can be obtained to screen for hyperglycemia. Additional screening tests should include measurements of TSH, luteinizing hormone, follicle-stimulating hormone, and, in men, testosterone levels. In families with suspected type I PGA syndrome, calcium and phosphorus levels

should be measured. These screening studies should be performed every 1 to 2 years up to about age 50 in families with PGA type II syndrome and until about age 40 in patients with type I syndrome. Screening measurements of autoantibodies against potentially affected endocrine organs are of uncertain prognostic value. The differential diagnosis of PGA syndrome should include the DiGeorge syndrome (hypoparathyroidism due to glandular agenesis and mucocutaneous candidiasis), Kearns-Sayre syndrome (hypoparathyroidism, primary hypogonadism, type 1 diabetes mellitus, and panhypopituitarism), Wolfram's syndrome (congenital diabetes insipidus and diabetes mellitus), and congenital rubella (type 1 diabetes mellitus and hypothyroidism).

TREATMENT

With the exception of Graves' disease, the management of each of the endocrine components of the disease involves hormone replacement and is covered in detail in the chapters on adrenal, thyroid, gonadal, and parathyroid disease (Chaps. 330,331,335,336, and341). One aspect of therapy deserves special emphasis. Namely, primary hypothyroidism can mask adrenal insufficiency by prolonging the half-life of cortisol; consequently, administration of thyroid hormone to a patient with unsuspected adrenal insufficiency can precipitate adrenal crisis. Thus, all patients with hypothyroidism in the context ofPGAsyndrome should be screened for adrenal disease and, if it is present, be treated with glucocorticoids prior to or concurrently with thyroid hormone therapy.

OTHER AUTOIMMUNE ENDOCRINE SYNDROMES

Insulin Receptor Antibodies Rare insulin-resistance syndromes occur in patients who develop antibodies that block the interaction of insulin with its receptor. Conversely, other classes of anti-insulin receptor antibodies can activate the receptor and can cause hypoglycemia; this disorder should be considered in the differential diagnosis of fasting hypoglycemia (Chap. 334).

Patients with insulin receptor antibodies and acanthosis nigricans are often middle-aged women who acquire insulin resistance in association with other autoimmune disorders such as systemic lupus erythematosus or Sjogren's syndrome. Vitiligo, alopecia, Raynaud's phenomenon, and arthritis may also be seen. Other autoimmune endocrine disorders, including thyrotoxicosis, hypothyroidism, and hypogonadism, occur rarely. Acanthosis nigricans, a velvety, hyperpigmented, thickened skin lesion, is prominent on the dorsum of the neck and other skin fold areas in the axillae or groin and often heralds the diagnosis in these patients. However, acanthosis nigricans also occurs in patients with obesity or polycystic ovarian syndrome, in which insulin resistance appears to be due to a postreceptor defect; thus acanthosis nigricans itself is not diagnostic of the immunologic form of insulin resistance.

Some patients with acanthosis nigricans have mild glucose intolerance, with a compensatory increase in insulin secretion that is only detected when insulin levels are measured. Others have severe diabetes mellitus requiring massive doses of insulin (several thousand units per day) to lower the blood glucose levels. The nature of the antibodies determines the manifestations; though insulin resistance is more common,

fasting hypoglycemia can result from insulinomimetic antibodies.

Ataxia telangiectasia is an autosomal recessive disorder caused by mutations in *ATM*, a gene involved in cellular responses to ionizing radiation and oxidative damage ([Chap. 364](#)). This disorder is characterized by ataxia, telangiectasia, immune abnormalities, and an increased incidence of malignancies. Insulin-resistant diabetes mellitus occurs and is associated with anti-insulin antibodies.

Autoimmune Insulin Syndrome with Hypoglycemia This disorder typically occurs in patients with other autoimmune disorders and is caused by polyclonal insulin-binding autoantibodies that bind to endogenously synthesized insulin. If the insulin dissociates from the antibodies several hours or more after a meal, hypoglycemia can result. Most cases of the syndrome have been described from Japan, and there may be a genetic component. In plasma cell dyscrasias such as multiple myeloma, the plasma cells may produce monoclonal antibodies against insulin and cause hypoglycemia by a similar mechanism.

Antithyroxine Antibodies and Hypothyroidism Circulating autoantibodies against thyroid hormones in patients with both immune thyroid disease and plasma cell dyscrasias such as Waldenstrom's macroglobulinemia can bind thyroid hormones, decrease their biologic activity, and result in primary hypothyroidism. In other patients the antibodies simply interfere with thyroid hormone immunoassays and cause false elevations or decreases in measured hormone levels.

Crow-Fukase Syndrome The features of this syndrome are highlighted by an acronym that emphasizes its important features: *p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathy, *M*-proteins, and *s*kin changes (POEMS). The most important feature is a severe, progressive sensorimotor polyneuropathy associated with a plasma cell dyscrasia. Localized collections of plasma cells (plasmacytomas) can cause sclerotic bone lesions and produce monoclonal IgG or IgA proteins. Endocrine manifestations include amenorrhea in women and impotence and gynecomastia in men, hypogonadism, hyperprolactinemia, type 2 diabetes mellitus, primary hypothyroidism, and adrenal insufficiency. Skin changes include hyperpigmentation, thickening of the dermis, hirsutism, and hyperhidrosis. Hepatomegaly and lymphadenopathy occur in about two-thirds of patients, and splenomegaly is seen in about one-third. Other manifestations include increased cerebrospinal fluid pressure with papilledema, peripheral edema, ascites, pleural effusions, glomerulonephritis, and fever. Five-year survival is about 60%.

The systemic nature of the disorder may cause confusion with other connective tissue diseases. The endocrine manifestations suggest an autoimmune basis of the disorder, but circulating antibodies against endocrine cells have not been demonstrated. Increased serum and tissue levels of interleukin 6, interleukin 1b, vascular endothelial growth factor, and tumor necrosis factor a are present, but the pathophysiologic basis for the [POEMS](#) syndrome is uncertain. Therapy directed against the plasma cell dyscrasia such as local radiation of bony lesions, chemotherapy, plasmapheresis, and treatment with all-*trans* retinoic acid may result in endocrine improvement.

MISCELLANEOUS DISORDERS WITH ENDOCRINE MANIFESTATIONS

A variety of other clinical and genetic disorders are associated with multiple endocrine manifestations are summarized in Table 339-3.

(Bibliography omitted in Palm version)

SECTION 2 - DISORDERS OF BONE AND MINERAL METABOLISM

340. INTRODUCTION TO BONE AND MINERAL METABOLISM - *Michael F. Holick, Stephen M. Krane*

BONE STRUCTURE AND METABOLIM (See also Chap. 343)

Bone is a dynamic tissue that is remodeled constantly throughout life. The arrangement of compact and cancellous bone provides a strength and density suitable for mobility. In addition, bone provides a reservoir for calcium, magnesium, phosphorus, sodium, and other ions necessary for homeostatic functions. The skeleton is highly vascular and receives about 10% of the cardiac output.

The extracellular components of bone consist of a solid mineral phase in close association with an organic matrix, of which 90 to 95% is type I collagen (Chap. 351). The noncollagenous portion of the organic matrix contains proteins derived from serum (albumin and a_2-HS glycoproteins), proteins containing a-carboxyglutamic acid (GLA) [*bone* GLA protein (BGP), *osteocalcin*, and a matrix GLA protein], the glycoprotein *osteonectin*, the phosphoprotein *osteopontin*, sialoproteins, *thrombospondin*, and other less well characterized proteins. Some of these proteins may function in initiating mineralization and in binding of the mineral phase to the matrix. The mineral phase is made up of calcium and phosphate and is best characterized as a poorly crystalline hydroxyapatite. The mineral phase of bone is deposited initially in intimate relation to the collagen fibrils and is found in specific locations in the "holes" between the collagen fibrils. This architectural arrangement of mineral and matrix results in a two-phase material well suited to withstand mechanical stresses.

Osteoblasts synthesize and secrete the organic matrix. Mineralization of the matrix, both in trabecular bone and in osteones of compact cortical bone (haversian systems), begins soon after the matrix is secreted (primary mineralization) but is not completed until after several weeks (secondary mineralization). Osteoblasts are derived from cells of mesenchymal origin (Fig. 340-1*A*). Although relatively little is known about the controls of osteoblast development, two genes have been shown to be important: core-binding factor A1 (*CBFA1*) and Indian hedgehog (*Ihh*). CBFA1 is a transcription factor and a homologue of the Drosophila factor, runt, and is expressed specifically in osteoblast progenitors and regulates the expression of several osteoblast-specific genes including osteopontin, bone sialoprotein, type I collagen, osteocalcin, and receptor-activator of NFkB (RANK) ligand. CBFA1 expression is regulated, in part, by bone morphogenic proteins (BMPs). *Cbfa1*-deficient mice are devoid of osteoblasts. Mice with a functional deletion of *Cbfa1* (*Cbfa1-/-*) have a cartilaginous skeleton but no osteoblasts and no bone, whereas mice with a deletion of only one allele (*Cbfa1+/-*) do have an osseous skeleton but have a delay in intramembranous bone formation of some cranial bones and the clavicles. The latter abnormalities are similar to those in the human disorder cleidocranial dysplasia, which maps to the locus that corresponds to *Cbfa1*.

The growth factor Ihh also plays a critical role in osteoblast development, as evidenced by the fact that Ihh-deficient mice lack osteoblasts in bone formed by endochondral ossification. Numerous other growth-regulatory factors affect osteoblast function,

including transforming growth factor (TGF) b types I and II, acidic fibroblast growth factor (aFGF) and basic fibroblast growth factor (bFGF), platelet-derived growth factor (PDGF), and insulin-like growth factors (IGFs) I and II. Active osteoblasts are characterized by their location and morphology; the presence of a specific skeletal form of alkaline phosphatase; the presence of receptors for parathyroid hormone (PTH) and 1,25-dihydroxyvitamin D [1,25(OH)$_2$D]; and the ability to synthesize specific matrix proteins, such as type I collagen, osteocalcin, and osteopontin. As an osteoblast secretes matrix, which is then mineralized, the cell becomes an *osteocyte*, still connected with its blood supply through a series of canaliculi. Osteocytes are thought to be the mechanosensors in bone that communicate signals to surface osteoblasts and their progenitors through the canalicular network.

Resorption of bone is carried out mainly by *osteoclasts*, multinucleated cells that are formed by fusion of cells derived from hematopoietic stem cells related to the mononuclear phagocyte series. Multiple factors regulating osteoclast development have been identified (Fig. 340-1*B*). Macrophage colony stimulating factor (M-CSF) plays a critical role at several steps in the pathway that ultimately leads to fusion of osteoclast progenitor cells to form multinucleated, active osteoclasts. Discovery of the RANK signaling pathway provides new insight into the pathway that links osteoblast and osteoclast development (Fig. 343-2). RANK ligand is expressed on the surface of osteoblast progenitors and stromal fibroblasts. In a process involving cell-cell interactions, it binds to the RANK receptor on osteoclast progenitors, stimulating a signal transduction cascade that leads to osteoclast differentiation and activation. Alternatively, a soluble decoy receptor, referred to as *osteoprotegerin* (OPG), can bind RANK ligand and inhibit osteoclast differentiation. Several growth factors and cytokines, including interleukins (IL) 1, 6, and 11, tumor necrosis factor (TNF), interferong, and M-CSF, modulate the osteoclast differentiation and function. Osteoclasts are also regulated indirectly by osteoblasts and adjacent stromal fibroblasts in the marrow. For example, PTH receptors are not found on mature osteoclasts, and PTH increases osteoclastic bone resorption by first acting on osteoblasts or stromal fibroblasts. 1,25(OH)$_2$D receptors are found in precursor cells that can differentiate into monocytes or osteoclasts, and 1,25(OH)$_2$D also promotes differentiation along the osteoclast pathway.

In the embryo and in the growing child, bone develops by remodeling and replacing previously calcified cartilage (endochondral bone formation) or is formed without a cartilage matrix (intramembranous bone formation). The PTH/PTHrP receptor plays a central role in the control of chondrocyte differentiation at growth plates (Chap. 341). Ihh production by growth plate chondrocytes stimulates the production of PTH-related peptide (PTHrP), which slows the differentiation of chondrocytes. This pathway creates a local feedback system as Ihh is suppressed by the actions of PTHrP. Consistent with this mechanism, mice with null mutations of the PTH/PTHrP receptor or PTHrP exhibit growth plate chondrodysplasia that reflects accelerated differentiation of proliferating chondrocytes. In humans, homozygous inactivating mutations of the PTH/PTHrP receptor cause Blomstrand's chondrodysplasia.

New bone, whether formed in infants or in adults during repair, has a relatively high ratio of cells to matrix and is characterized by coarse fiber bundles of collagen that are interlaced and randomly dispersed (woven bone). In adults, the more mature bone is

organized with fiber bundles regularly arranged in parallel or concentric sheets (lamellar bone). In long bones, deposition of lamellar bone in a concentric arrangement around blood vessels forms the haversian systems. Growth in length of bones is dependent on proliferation of cartilage cells and on the endochondral sequence at the growth plate. Growth in width and thickness is accomplished by formation of bone at the periosteal surface and by resorption at the endosteal surface, with the rate of formation exceeding that of resorption. In adults, after the epiphyses close, growth in length and endochondral bone formation cease, except for some activity in the cartilage cells beneath the articular surface. Even in adults, however, remodeling of bone (remodeling of haversian systems as well as trabecular bone) continues throughout life. In adults, ~4% of the surface of trabecular bone (such as iliac crest) is involved in active resorption, whereas 10 to 15% of trabecular surfaces is covered with osteoid. Radioisotope studies indicate that as much as 18% of the total skeletal calcium is deposited and removed each year. Thus, bone is an active metabolizing tissue that requires an intact blood supply.

The response of bone to fractures, infection, and interruption of blood supply and to expanding lesions is relatively limited. Dead bone must be resorbed, and new bone must be formed, a process carried out in association with growth of new blood vessels into the involved area. In injuries that disrupt the organization of the tissue, such as a fracture in which apposition of fragments is poor or when motion exists at the fracture site, the progenitor stromal cells differentiate into cells with functional capacities different from those of osteoblasts, and varying amounts of fibrous tissue and cartilage are formed. When there is good apposition with fixation and little motion at the fracture site, repair occurs predominantly by formation of new bone without other scar tissue.

Remodeling of bone occurs along lines of force modulated by the mechanical stresses to which it is subjected. The signals from these mechanical stresses are sensed by osteocytes, which then transmit other signals to osteoclasts (or their precursors) or osteoblasts (or their precursors). A bowing deformity increases new bone formation at the concave surface and resorption at the convex surface, seemingly designed to produce the strongest mechanical structure. Expanding lesions in bone, such as tumors, induce resorption at the surface in contact with the tumor. Even in a disorder as architecturally disruptive as Paget's disease, remodeling is dictated by mechanical forces. Thus, the plasticity of bone is due to the interaction of cells with each other and with the environment.

The cycle of bone resorption and formation is a highly orchestrated process carried out by the basic multicellular unit (BMU), composed of a group of osteoclasts and osteoblasts (Fig. 340-2). Osteoclast-mediated resorption of bone takes place in scalloped spaces (*Howship's lacunae*) where the osteoclasts are attached through a specific av,b3 integrin to components of the bone matrix such as osteopontin. This clear zone contains contractile proteins. The resorbing end of the cell forms a specialized ruffled border, which is in contact with the bone. Proteins, including a specialized proton-pump ATPase, are found in the ruffled border membrane and contribute to the acid environment, which solubilizes the mineral phase. In addition to the proton pump, carbonic anhydrase (type II isoenzyme) is required to maintain the acid pH. Other features of active osteoclasts include expression of the proto-oncogene c-*src*, tartrate-resistant acid phosphatase, cell-surface receptors for calcitonin, sodium pumps

of the kidney type, a bicarbonate/chloride exchanger of the band 3 family, and an ability to resorb mineralized bone. The bone matrix is resorbed in the acid environment adjacent to the ruffled border by proteinases that act at low pH, such as cathepsin K, following solubilization of the mineral phase.

Bone formation involves deposition of an organic matrix by osteoblasts followed by mineralization. The mineral phase is composed of calcium and phosphorus, and the concentration of these ions in the plasma and extracellular fluid (ECF) influences the rate at which mineral is formed. In vitro, mineralization can proceed and crystals of hydroxyapatite can grow at concentrations of calcium and phosphorus similar to those in an ultrafiltrate of plasma. The calcium phosphate solid phase at the inception of mineralization is brushite ($CaHPO_4 \times 2H_2O$). As mineralization progresses, the solid phase is a poorly crystalline hydroxyapatite with a relatively low (~1.2) calcium/phosphate molar ratio. With age and maturation, the perfection of the crystals and the calcium/phosphate ratio increase. Fluoride ions, when incorporated into the mineral phase, decrease the proportion of amorphous calcium phosphate and enhance the crystal structure.

There is a limit for the concentration of calcium and phosphorus ions in the ECF below which mineralization does not occur. A "solubility product" for bone mineral is difficult to calculate because (1) the mineral phase itself is of variable composition, and (2) the various components in ECF that regulate this solubility product are not known. Nevertheless, when the concentrations of calcium and phosphorus in ECF are excessive, a mineral phase may form in areas (e.g., soft tissues) that are not normally mineralized.

Collagens from a variety of sources can catalyze the nucleation of a mineral phase of calcium and phosphorus from solutions of these ions. The organization of collagen probably influences the amount and type of mineral phase formed in bone. The primary structures of type I collagen in skin and bone tissues are similar. There are differences, however, in posttranslational modifications of type I collagen, such as hydroxylations; glycosylations; and the type, number, and distribution of intermolecular cross-links (Chap. 351). In addition, the holes in the packing structure of the collagen are larger in mineralized collagen of bone and dentin than in unmineralized collagens such as tendon. Single amino-acid substitutions in the helical portion of either the a1 or a2 chain of type I collagen due to mutations in the *COL*1A1 or *COL*1A2 genes in osteogenesis imperfecta disrupt the organization of bone, and this indicates the importance of the fibrillar matrix in the structure of bone. At the time of their discovery, it was thought that some of the non-collagenous bone proteins -- e.g., osteocalcin (bone-GLA protein), osteonectin, and osteopontin -- played a role in mineralization, although such a role has not been established. Osteocalcin, a product exclusively of osteoblasts, is measurable by immunoassays in normal human serum, and its levels correlate with other measurements of bone formation. Matrix-GLA protein (MGP), a component of bone as well as non-osseous tissues, acts to inhibit mineralization in the non-osseous tissues. Thus, functional deletion of the MGP gene in mice results in massive soft tissue calcification, particularly in arterial walls.

Alkaline phosphatase is a marker for osteoblasts, and cellular levels of this enzyme correlate with rates of bone formation. Although mineralization defects occur in

individuals with mutations that decrease alkaline phosphatase activity (hypophosphatasia), the function of alkaline phosphatase in mineralization is not completely understood. Other circulating markers of bone formation include osteocalcin and type I procollagen C-terminal peptides. Urinary markers for bone resorption are hydroxyproline, hydroxylysine and its glycosides, and the bone-specific hydroxypyridinium collagen cross-links (Chap. 342). Inorganic pyrophosphate is a potent inhibitor of mineralization at levels below those necessary to bind calcium ions.

CALCIUM METABOLISM

A total of 1 to 2 kg of calcium is present in the average adult,>98% of it in the skeleton. The calcium of the mineral phase at the surface of the crystals is in equilibrium with that in theECF, but only a minor fraction of the total pool (~0.5%) is exchangeable. In normal adults plasma levels range from 2.2 to 2.6 mmol/L (8.8 to 10.4 mg/dL). The calcium in plasma is present as three forms: free ions, ions bound to plasma proteins, and, to a small extent, diffusible complexes. The concentration of free calcium ions, averaging 1.2 mmol/L (4.8 mg/dL), influences many cellular functions and is subjected to tight hormonal control, especially throughPTH(Chap. 341). The concentration of serum proteins is an important determinant of calcium ion concentration; most calcium ion is bound to albumin. Ionized calcium can be measured directly with the use of calcium-specific electrodes. If ionized calcium cannot be measured, certain approximations can be used to estimate the protein-bound and ionized fractions. One formula that approximates the amount of calcium bound to protein is

A simplified correction is sometimes used to assess whether the total serum calcium concentration is abnormal when serum proteins are low. The correction is to add 1 mg/dL to the serum calcium level for every 1 g/dL by which the serum albumin level is below 4.0 g/dL. If the serum calcium level, for example, is 7.8 mg/dL (a subnormal value) and the serum albumin level is only 3.0 g/dL, then the stated serum calcium level is corrected by adding 1 mg/dL; the corrected value of 8.8 mg/dL is within the normal range.

The concentration of calcium ions in theECF is kept constant by processes that constantly add and remove calcium. Calcium enters the plasma via absorption from the intestinal tract and resorption of ions from the bone mineral. Calcium leaves the ECF via secretion into the gastrointestinal tract (~100 to 200 mg/d), urinary excretion (~50 to 300 mg/d), deposition in bone mineral, and losses in sweat (up to 100 mg/d). Bone resorption and formation are tightly coupled, with approximately 12 mmol (500 mg) calcium entering and leaving the skeleton daily (Fig. 340-3). Calcium ions inside the cell mediate a variety of cellular functions. The level of free calcium in the cell is very low, approximately 0.1 umol/L; thus, the gradient between plasma and intracellular free calcium is about 10,000 to 1. This gradient is tightly regulated by various channels and ion pumps.

The average dietary calcium intake for most adults in the United States is approximately 15 to 20 mmol/d (0.6 to 0.8 g/d). However, with heightened awareness of the role of adequate calcium intake for the prevention of osteoporosis, many adults on

supplements have an intake of 20 to 37 mmol/d (0.8 to 1.5 g/d). Less than half of dietary calcium is absorbed in adults. Calcium absorption increases during periods of rapid growth in children, in pregnancy, and in lactation and decreases with advancing age. Most of the calcium is absorbed in the proximal small intestine, and the efficiency of absorption decreases in the more distal intestinal segments. Both active transport and diffusion-limited absorption are involved; the former is more important in the upper intestine and the latter in the lower intestine. Both processes are influenced by vitamin D (see below). All forms of calcium in the diet are not equally absorbed; calcium as the chloride is probably absorbed more efficiently than that in other preparations. Secretion of calcium into the intestinal lumen is constant and independent of absorption. If calcium availability in the diet is low [<12 mmol/d (500 mg/d)], a positive calcium balance requires an efficiency of absorption>30 to 40%.

The urinary calcium excretion of normal adults having an average calcium intake ranges between 2.5 and 10 mmol/d (100 and 400 mg/d). When the dietary calcium level is<5 mmol/d (200 mg/d), urinary calcium excretion is usually <5 mmol/d (200 mg/d). However, in most normal individuals, wide variations in dietary intake have little effect on urinary calcium. Hence, when the diet is low in calcium, the relative inefficiency of renal calcium conservation leads to a negative calcium balance unless calcium absorption is maximal (Fig. 340-3).

The amount of calcium in the urine is small compared with that filtered by the glomerulus [~150 to 250 mmol/d (6 to 10 g/d)] because the rates of reabsorption of the filtered calcium are high. Reabsorption takes place predominantly in the proximal tubule (~60%) and in Henle's loop (~25%) and to a small extent in the distal tubule. The calcium-sensing receptor also plays a role in renal calcium excretion, though the mechanisms that regulate its function have not been fully defined (Chap. 341). The excretion of other electrolytes affects the urinary excretion of calcium. For example, urinary calcium is usually proportional to urinary sodium; sulfate also increases calcium excretion.

A deficiency ofPTH or vitamin D, intestinal disease, or severe dietary calcium deprivation may provide challenges to calcium homeostasis that cannot be compensated adequately by renal calcium conservation, resulting in a negative calcium balance. Increased bone resorption may protect againstECFcalcium depletion even in states of chronic negative calcium balance but only at the expense of progressive bone loss.

PATHOPHYSIOLOGY

A decrease in the concentration of free calcium ions in plasma results in increased neuromuscular irritability and tetany. This syndrome is characterized by peripheral and perioral paresthesia, carpal spasm, pedal spasm, anxiety, seizures, bronchospasm, laryngospasm, Chvostek's sign, Trousseau's sign, and Erb's sign, and lengthening of the QT interval of the electrocardiogram. In infants tetany may be manifested only by irritability and lethargy. The level of calcium ions that determines which features of tetany will be manifested varies among individuals. Tetany is also influenced by other components of theECF; e.g., hypomagnesemia and alkalosis lower whereas hypokalemia and acidosis raise the threshold for tetany.

Increases in total serum calcium concentration are usually accompanied by increases in free calcium levels and may be associated with anorexia, nausea, vomiting, constipation, hypotonia, depression, and occasionally lethargy and coma. Persistent hypercalcemia, especially when accompanied by normal or elevated levels of serum phosphate, may cause ectopic deposition of a solid phase of calcium and phosphate in walls of blood vessels, connective tissue about the joints, gastric mucosa, cornea, and renal parenchyma. Hypercalcemia per se alters renal function in addition to the pathologic effects of calcium phosphate deposition.

PHOSPHORUS METABOLISM

Phosphorus is a major component of bone and of all other tissues and in some form is involved in almost all metabolic processes, including energy storage, membrane transport, membrane composition, and signal transduction. About 600 g of phosphorus is present in the normal adult, of which 85% is present in the crystalline structure of the skeleton.

In plasma from fasting subjects, most of the phosphorus is present as inorganic orthophosphate in concentrations of approximately 0.75 to 1.45 mmol/L (2.5 to 4.5 mg/dL). In contrast to calcium, of which ~50% is bound, only ~12% of the phosphorus in plasma is bound to proteins. Free HPO_4^{2-} and $NaHPO_4^-$ normally account for ~75% of the total phosphorus, and free $H_2PO_4^-$ accounts for ~10%. Since so many species are present, depending on pH and other factors, concentrations are usually expressed in terms of elemental phosphorus, in units of mmol/L or mg/dL. The serum phosphorus, however, can vary based on age; young children have almost twice the serum phosphorus as adults due to the need for rapid skeletal mineralization. Postmenopausal women also have higher circulating phosphorus levels. After ingesting a meal containing carbohydrate, there is a decrease in serum phosphorus levels [by 0.3 to 0.5 mmol/dL (1.0 to 1.5 mg/dL)] in response to the increase in insulin secretion, which enhances cellular phosphorus uptake and utilization. An increase in serum pH will decrease serum phosphorus, whereas a decrease in pH increases phosphorus concentration. There is a circadian variation in phosphorus concentration even during a 24-h fast: the nadir occurs between 9:00 A.M. and noon followed by an increase to a plateau in the afternoon and another small peak after midnight.

Phosphorus is plentiful in the diet. Common sources include dairy products, meats, eggs, and carbonated beverages that contain phosphoric acid. Approximately 60 to 70%of phosphorus is passively absorbed in the small intestine (Fig. 340-4). 1,25(OH)$_2$D enhances phosphorus absorption along the entire small intestine, with the highest efficiency in the jejunum and ileum. Chronic low phosphorus intake (<2 mg/kg of body weight per day) decreases serum phosphorus levels. Low serum phosphorus stimulates the renal production of 1,25(OH)$_2$D, which, in turn, increases the efficiency intestinal absorption up to 80 to 90%. 1,25(OH)$_2$D also decreasesPTHsecretion and, thereby reduces renal tubular loss of phosphorus.

The major control of phosphorus balance is exerted by the kidney. Approximately 90% of phosphorus in the circulation is filtered through the glomerulus and is largely absorbed by the proximal tubule such that only 10 to 15% of the filtered load is normally

excreted. Urinary phosphorus excretion is reflective of dietary intake. Phosphorus absorption in the proximal tubule is coupled with sodium absorption. The primary regulation of phosphorus metabolism occurs in the distal convoluted tubule, and this mechanism is independent of sodium reabsorption. Volume expansion and decreased sodium reabsorption increase phosphorus clearance.

HYPOPHOSPHATEMIA

Causes Although there are many potential causes for hypophosphatemia ([Table 340-1](#)), the most common etiologies include: (1) decreased intestinal phosphorus absorption, either due to vitamin D deficiency or the presence of a phosphorus-binding antacid; (2) urinary losses that are PTH- or alcohol-mediated; and (3) a shift of phosphorus from extracellular to intracellular compartments due to exogenous administration of insulin or consumption of nutrients that stimulate insulin release (e.g., carbohydrates). Increased renal clearance of phosphorus occurs in primary hyperparathyroidism, vitamin D deficiency, vitamin D-resistant and D-dependent rickets, hyperglycemic states, and oncogenic osteomalacia. In vitamin D deficiency, serum phosphorus is low because of decreased intestinal absorption as well as secondary hyperparathyroidism, which increases phosphorus losses in the urine. In X-linked hypophosphatemic rickets, there is a genetic defect in the *PHEX* gene, which encodes a neutral endopeptidase presumed to degrade the phosphaturia hormone known as *phosphatonin*. The disorder is associated with a severe renal leak of phosphorus into the urine. In addition, there is a defect in hypophosphatemia-mediated stimulation of 25(OH)D-1a-hydroxylase, resulting in decreased intestinal phosphorus absorption. Acidosis and hyperglycemic states associated with polyuria also cause excessive phosphorus loss in the urine. Ketoacidosis enhances intracellular and organic phosphorus degradation, thereby releasing large amounts of inorganic phosphorus into the circulation that is cleared into the urine. In ketosis, the serum phosphorus is often normal because of the continuous shift of phosphorus from intracellular to extracellular pools. However, when the ketosis is corrected, hypophosphatemia is apparent because of the return of phosphorus into the intracellular compartment ([Chap. 333](#)). A severe, acquired form of hypophosphatemia, *oncogenic osteomalacia*, is associated with vascular, mesenchymal tumors such as hemangiopericytomas but occasionally also with small cell lung cancer, prostate cancer, and other malignant tumors. It is likely that these tumors secrete a substance similar or identical to phosphatonin. The phosphorus levels in these patients are usually extremely low [0.4 to 0.5 mmol/L (1.2 to 1.5 mg/dL)], and the $1,25(OH)_2D$ levels are low or undetectable. The disorder is associated with severe fatigue, muscle weakness, and unrelenting bone discomfort.

Alcohol abuse is the most common cause of severe hypophosphatemia, which is caused by poor dietary intake of phosphorus, ethanol-enhanced urinary excretion of inorganic phosphorus, the use of calcium- or aluminum-containing antacids, and vomiting. Hypophosphatemia may transiently worsen with refeeding. Alcoholics may also have associated calcium and vitamin D deficiency and secondary hyperparathyroidism, which enhances phosphorus-wasting in the urine. Alcoholic ketoacidosis induces marked phosphaturia. Intense hyperventilation for prolonged periods may depress serum phosphorus levels due to associated alkalosis. Rapid correction of chronic respiratory acidosis has also been associated with hypophosphatemia and can lead to diaphragm weakness and an exacerbation of

respiratory failure. Advanced leukemia with blast crisis (leukocyte counts usually>100,000) may cause severe hypophosphatemia; the likely cause is a rapid uptake of phosphorus into the rapidly dividing cells.

Laboratory and Clinical Findings Serum phosphorus levels should be determined in a fasting state. Mild hypophosphatemia is not usually associated with clinical symptoms. In severe hypophosphatemia [£0.3 mmol/L (£1.0 mg/dL)], multiple organ systems are affected. Patients become irritable, apprehensive, and hyperventilate, resulting in complaints of muscle weakness, numbness, and paresthesia. In the most severe form, they are confused or obtunded and suffer from seizures and coma, which can ultimately lead to death. This metabolic encephalopathy is often associated with slowing of the electroencephalogram.

Phosphorus is essential for muscle function because of the need for large amounts of ATP and creatine phosphate. Patients with severe hypophosphatemia often complain of fatigue, muscle weakness, myalgia, and myopathy. Hypophosphatemia can cause rhabdomyolysis, which is particularly common in chronic alcoholics or during alcohol withdrawal. Rhabdomyolysis can be precipitated during treatment for diabetic ketoacidosis or by hyperalimentation or refeeding in a malnourished patient. Cardiomyopathy can also occur, resulting in reduced cardiac output, impaired pressor responsiveness to catecholamines, hypotension, and ventricular arrhythmias. Restoration of phosphorus deficits can result in prompt reversal. Severe muscle weakness can lead to respiratory insufficiency.

Erythrocytes and leukocytes are highly dependent on phosphorus for their function. Chronic hypophosphatemia decreases 2,3-bisphosphoglycerate and ATP, enhancing oxygen dissociation from hemoglobin and leading to tissue hypoxia. Hypophosphatemia causes impaired phagocytosis and opsonization and, therefore, increases susceptibility to bacterial and fungal infections.

Chronic hypophosphatemia causes a mineralization defect of the skeleton. In children, this causes rickets. In adults, chronic hypophosphatemia (often due to vitamin D deficiency) causes osteomalacia (see below). Patients with severe renal phosphorus-wasting and severe chronic hypophosphatemia may have marked fatigue, muscle weakness, and severe bone pain, especially of their long bones and ribcage.

TREATMENT

Mild hypophosphatemia usually resolves spontaneously when the underlying cause is corrected. Oral phosphorus replacement is sufficient if serum phosphorus is>0.3 mmol/L (1 mg/dL) and the patient is asymptomatic. Milk is an excellent source of phosphorus as it contains 1 g of inorganic phosphorus per liter. Carbonated beverages that contain phosphoric acid provide another source of phosphorus, especially for patients with lactase deficiency. Pharmaceutical preparations of phosphorus, such as Neutraphos or KPhos, contain sodium and potassium salts of phosphate. Depending on the degree of hypophosphatemia, up to 3 g/d can be given in four to six divided doses per 24 h. These doses usually do not cause diarrhea; >5 g/d will induce diarrhea.

For severe hypophosphatemia, with serum phosphorus levels <0.2 to 0.3 mmol/L (<0.5

to 1.0 mg/dL),[3]3 g/d of phosphorus may be required over several days to replete body stores. In patients with severe symptomatic hypophosphatemia who are unable to eat, intravenous phosphorus can be given, up to 1 g in 1 L of fluid over 8 to 12 h. Some caution is necessary when giving phosphorus intravenously because of the potential for precipitating soft tissue calcification. A serum calcium ´ serum phosphorus product >70 markedly increases the risk of soft tissue calcification and nephrocalcinosis. Patients with chronic hypophosphatemia caused by inherited or acquired renal phosphorus leak require vigilance when receiving high doses of oral phosphorus. Transiently elevated serum phosphorus levels can decrease ionized calcium levels, resulting in chronic stimulation of the parathyroid gland and leading to autonomous, persistent hyperplasia of the parathyroid glands. Thus, it is best to give frequent divided doses of phosphorus (four to six times a day), equaling a total of 2 to 3 g/d.

Phosphorus should not be given intramuscularly or subcutaneously because it can cause soft tissue necrosis and severe discomfort. Intravenous sodium or potassium phosphate, 15 mmol (0.465 g) of elemental phosphorus given in 100 mL of 0.9% saline over 60 min, elevates serum phosphorus levels by an average of 0.6 to 1.2 mmol/L (1.75 to 3.8 mg/dL).

HYPERPHOSPHATEMIA

In adults, hyperphosphatemia is defined as a serum phosphorus level>1.6 mmol/L (5 mg/dL). In children, this level is much higher. The most common causes of hyperphosphatemia are acute and chronic renal failure (Table 340-2). In renal failure, the loss of tubular function impairs phosphorus excretion. This results in a cascade of events that can also affect calcium and phosphorus metabolism. The increase in serum phosphorus levels reduces serum calcium levels and the production of $1,25(OH)_2D$, leading to decreased intestinal calcium absorption and secondary hyperparathyroidism. Patients with pseudohypoparathyroidism and tumoral calcinosis also have decreased renal phosphorus clearance that results in hyperphosphatemia. Hypothyroidism reduces renal phosphorus excretion and may increase circulating concentrations of phosphorus. Vitamin D intoxication, due to excessive ingestion of either vitamin D or one of its analogues, can cause hyperphosphatemia along with hypercalcemia. Severe hypothermia, crush injuries, nontrauma rhabdomyolysis, tumoral calcinosis, and cytotoxic therapy of hematologic malignancies such as acute lymphoblastic leukemia can be associated with hyperphosphatemia. The serum phosphorus level can be artifactually elevated due to hemolysis of the blood sample. Thrombocytosis and multiple myeloma can cause spuriously elevated serum phosphorus levels due to thrombocytolysis.

Laboratory and Clinical Findings A rapid elevation of serum phosphorus can cause hypocalcemia and symptoms of neuromuscular irritability and tetany. Chronic hyperphosphatemia in association with normocalcemia can result in nephrocalcinosis and soft tissue calcification.

TREATMENT

In addition to treating the underlying disorder, dietary phosphorus intake should be limited by restricting carbonated beverages containing phosphoric acid and decreasing

milk and dairy product consumption. The dietary intake of phosphorus should be between 600 and 1000 mg a day with modest protein restriction. For control of chronic hyperphosphatemia, usually in patients with chronic renal failure, oral aluminum hydroxide or aluminum carbonate gels are indicated. Prolonged use of aluminum-containing compounds is not recommended because of aluminum toxicity causing adynamic bone disease, proximal myopathy, encephalopathy, and anemia. When hyperphosphatemia is due to vitamin D intoxication, calcium salts are contraindicated because the high efficiency of calcium absorption can lead to severe hypercalcemia, soft tissue calcification, and nephrocalcinosis.

MAGNESIUM METABOLISM

Magnesium is the most abundant intracellular divalent cation. It is an essential cofactor for a multitude of enzymatic reactions that are important for the generation of energy from ATP. Approximately 30% of magnesium in the serum is protein-bound, 55% is ionized, and the remaining 15% is complexed. Like calcium, magnesium is bound to albumin, and it is the ionized fraction that is important for physiologic processes including neuromuscular function and maintenance of cardiovascular tone.

The serum concentration of magnesium is tightly regulated within a narrow range of approximately 0.7 to 1.1 mmol/L (1.4 to 2.2 meq/L)(1.7 to 2.6 mg/dl) as a result of the efficient absorption of dietary magnesium by the small intestine and conservation of magnesium in the kidney. About 30% of dietary magnesium is absorbed in the small intestine, but this fraction increases markedly when intake is substantially reduced. Approximately 96% of filtered magnesium is reabsorbed along the nephron, and only 4%is excreted into the urine. Because there is no regulation of magnesium absorption in the distal tubule and because magnesium reabsorption is very efficient, an increase in distal delivery increases magnesium loss in the urine.

HYPOMAGNESEMIA

Although magnesium deficiency is a common clinical problem, serum magnesium levels are often overlooked or not measured in patients at risk for the disorder. Approximately 10% of patients admitted to city hospitals are hypomagnesemic, and up to 65% of patients in intensive care units may be magnesium-deficient. Hypomagnesemia is caused primarily by renal or gastrointestinal losses or decreased efficiency of intestinal magnesium absorption (Table 340-3). Reduced renal reabsorption due to loop diuretics and alcohol use is a common cause of hypomagnesemia. Because magnesium excretion is tightly coupled to sodium and calcium excretion, intravenous fluid therapy and volume-expanded states, such as primary hyperaldosteronism, may result in hypomagnesemia. Hypercalcemia and hypercalciuria decrease tubular reabsorption of magnesium. Osmotic diuresis in diabetes mellitus is one of the more common causes of hypomagnesemia.

Vomiting and nasogastric suctioning can cause severe magnesium depletion because intestinal tract fluids contain ~0.5 mmol/L (1.2 mg/dL)(1 meq/L). Fluid loss from diarrhea can contain as much as 7.4 mmol/L (18 mg/dL)(15 meq/L). Consequently, ulcerative colitis, Crohn's disease, and intestinal or biliary fistulas can result in magnesium depletion. Hypomagnesemia is prevalent in alcoholics. Ethanol causes a transient loss

of magnesium in the urine. In most alcoholics, however, the magnesium deficit is modest. A more profound fall in serum magnesium levels may occur during alcohol withdrawal, where the decrease is associated with falls in levels of serum phosphate and potassium, probably due to shifts of these ions into intracellular compartments. The use of loop diuretics, as well as aminoglycosides, cisplatin, cyclosporine, and amphotericin B can increase renal loss of magnesium.

The clinical manifestations of hypomagnesemia are similar to those of severe hypocalcemia. The signs and symptoms of hypomagnesemia include muscle weakness, prolonged PR and QT intervals, and cardiac arrhythmias. Positive Chvostek's and Trousseau's signs indicative of hypocalcemia are often positive in hypomagnesemic patients as well; carpopedal spasm can also occur with hypomagnesemia. Magnesium is important for effective PTH secretion as well as the renal and skeletal responsiveness to PTH; thus, hypomagnesemia is often associated with hypocalcemia due to impaired PTH secretion and function (Chap. 341).

Low serum magnesium levels <0.7 mmol/L (1.8 mg/dL)(1.5 meq/L) are indicative of magnesium deficiency. For mild deficiency, oral magnesium replacement is effective. The major side effect is diarrhea. Symptoms often occur when the serum magnesium is <0.5 mmol/L (1.2 mg/dL)(1.0 meq/L). This level is indicative of significantly depleted total-body magnesium stores. Because most magnesium resides in the intracellular space, the total-body magnesium deficit is often ~200 mmol (4800 mg) by the time serum levels fall to <0.5 mmol/L (1.2 mg/dL)(1.0 meq/L). Parenteral magnesium administration is usually needed under these circumstances. Two grams of magnesium sulfate [8.0 mmol (192 mg)(16.2 meq) of magnesium] can be given intravenously, with a cumulative dose up to 24 mmol (576 mg)(48 meq) over 24 h. Alternatively, a 50% solution of 2 g of magnesium sulfate can be given every 8 h intramuscularly although these injections can be painful. Patients with severe hypomagnesemia and associated seizures or acute arrhythmias can be given 4 to 8 mmol (96 to 192 mg)(8 to 16 meq) of magnesium as an intravenous injection over 5 to 10 min, followed by 24 mmol/d (576 mg/d)(48 meq/d).

A normal serum magnesium concentration attained after acute magnesium repletion is not necessarily indicative of repletion of the total-body magnesium stores. Restoration of urinary magnesium excretion is a better indicator of magnesium repletion. Once urinary magnesium excretion increases, the body stores are usually replenished. Patients who have chronic magnesium loss from intestinal or renal sources may require continued oral magnesium supplementation on a daily basis of up to 12.5 mmol/d (300 mg/d) in divided doses. Patients with renal failure need to be monitored carefully to prevent hypermagnesemia.

HYPERMAGNESEMIA

Hypermagnesemia is rare but can be seen in renal failure when patients are taking magnesium-containing antacids, laxatives, enemas, or infusions (Table 340-4). It can also be seen in acute rhabdomyolysis.

The most readily detected clinical sign of hypermagnesemia is the disappearance of deep tendon reflexes. Neuromuscular symptoms include depressed respiration and

apnea due to paralysis of the voluntary muscles, prolonged PR intervals, and increased QRS duration and QT interval; complete heart block and cardiac arrest can occur. Hypocalcemia may occur because hypermagnesemia depresses PTH secretion and induces an end-organ resistance to PTH similar to the effect seen in hypomagnesemia.

Treatment includes stopping the antacid or other preparations that contain large amounts of magnesium. The excess magnesium is quickly excreted by the kidney. Renal failure patients may require dialysis against a low magnesium bath. For severe hypermagnesemia with associated life-threatening complications, intravenous calcium in doses of 100 to 200 mg (elemental) over 5 to 10 min will antagonize the toxic effects of magnesium.

VITAMIN D

Vitamin D is a hormone rather than a classic vitamin, since with adequate exposure to sunlight, no dietary supplements are needed. Vitamin D exerts its physiologic effects on bone, intestine, kidney, and the parathyroid glands to modulate calcium and phosphorus metabolism. The active principle of vitamin D is synthesized under metabolic control via successive hydroxylations in the liver and kidney and is transported through the blood to its main target tissues (the small intestine and bone), where it regulates calcium homeostasis.

PHOTOBIOGENESIS

Vitamin D_3 is a derivative of 7-dehydrocholesterol (provitamin D_3), the immediate precursor of cholesterol. When skin is exposed to sunlight or certain artificial light sources, the ultraviolet radiation enters the epidermis and causes transformation of 7-dehydrocholesterol to vitamin D_3. Wavelengths between 290 and 315 nm are absorbed by the conjugated double bonds at C_5 and C_7 of 7-dehydrocholesterol to produce previtamin D_3 (Fig. 340-5). Vitamin D_3 is made in the skin from the previtamin for many hours after a single sun exposure (Fig. 340-5). Once vitamin D_3 is synthesized, it is translocated from the epidermis into the circulation by the vitamin D-binding protein. Melanin in the skin competes with 7-dehydrocholesterol for ultraviolet photons and thus can limit the synthesis of previtamin D_3. The photochemical isomerization of previtamin D_3 and vitamin D_3 to biologically inert products appears to be more important in preventing excessive production of previtamin D_3 and vitamin D_3 during prolonged exposure to the sun.

Aging decreases the capacity of the skin to produce vitamin D_3; this capacity is reduced more than fourfold after age 70. Topical sunscreens can reduce or prevent cutaneous production of vitamin D_3 by absorbing the solar radiation responsible for previtamin D_3 synthesis in the skin. Other factors that affect the cutaneous synthesis of vitamin D_3 include altitude, geographic location, time of day, and area exposed. Latitude has profound effects on the cutaneous synthesis of vitamin D_3. As the zenith angle of the sun increases with approaching winter, more of the high-energy ultraviolet photons responsible for formation of the previtamin are absorbed by the ozone layer. In an area such as Boston (42°N), the absorption of these photons is so complete that essentially no vitamin D_3 is made in the skin between the months of November through February.

When the entire body is exposed to sufficient sunlight to cause mild erythema, the increase in the blood vitamin D is approximately equivalent to consuming oral doses of 10,000 to 25,000 international units (1 IU = 0.025 ug) of vitamin D. Only when skin irradiation is insufficient to produce the required quantities of vitamin D_3 is dietary supplementation needed to prevent skeletal mineralization defects. The fortification of milk and some cereals with either crystalline vitamin D_2([Fig. 340-5](#)) or vitamin D_3should prevent rickets and osteomalacia. A survey of the vitamin D content in milk from the United States and western Canada revealed, however, that 71% did not contain 80 to 120% of the amount of vitamin D on the label and that ~15% of skim milk did not contain detectable vitamin D.

In 1997, the Food and Nutrition Board for the Institute of the Medicine recommended 200 IU/d as the adequate intake of vitamin D for neonates, children, and adults up to 50 years. For adults 51 to 70 and >71 years, the committee recommended 400 and 600 IU/d, respectively. In the absence of adequate sunlight exposure, all children and adults require at least 400 to 600 IU/d.

METABOLISM

In the liver, vitamin D is metabolized to 25-hydroxyvitamin D [25(OH)D] by hepatic mitochondrial and/or microsomal enzyme(s) ([Fig. 340-5](#)). 25(OH)D is one of the major circulating metabolites, and its half-life is about 21 days. The concentrations of 25(OH)D and some of its metabolites in the serum are measured using competitive binding assays. The normal serum 25(OH)D concentration varies among different laboratories from 20 to 200 nmol/L (8 to 80 ng/mL). Individuals exposed to excessive sunlight may have concentrations of 25(OH)D up to 250 nmol/L (100 ng/mL) without adverse effects on calcium metabolism. The serum 25(OH)D levels usually reflect both 25-hydroxyvitamin D_2[25(OH)D_2] and 25-hydroxyvitamin D_3[25(OH)D_3]. The ratio of these two 25-hydroxylated derivatives depends on the relative amounts of vitamins D_2 or D_3present in the diet and the amount of previtamin D_3produced by exposure to sunlight.

The hepatic 25-hydroxylation of vitamin D is regulated by a product feedback mechanism. This regulation, however, is not tight; an increase in dietary intake or endogenous production of vitamin D_3increases 25(OH)D levels in the serum. The levels can rise to>1200 nmol/L (500 ng/mL) when the intake of vitamin D is excessive. Serum 25(OH)D levels are reduced in severe chronic liver disease ([Table 340-5](#)). 25(OH)D is probably not biologically active at physiologic levels in vivo but is active in vitro at high concentrations.

After formation in the liver, 25(OH)D is bound by the vitamin D-binding protein and transported to the kidney for an additional stereospecific hydroxylation on either C_1 or C_{24}([Fig. 340-5](#)). The kidney plays a pivotal role in the metabolism of 25(OH)D to the biologically active metabolite. The renal mitochondrial 25(OH)D-1-hydroxylase activity is enhanced by hypocalcemia to increase the rate of conversion of 25(OH)D to 1,25(OH)$_2$D. Hypocalcemia may not control this hydroxylation directly, however. Any decrease in the serum concentration of calcium below normal is a stimulus for increased secretion of[PTH](#), which increases the synthesis of 1,25(OH)$_2$D in the renal proximal convoluted tubule. The renal production of 1,25(OH)$_2$D enhances the effects of PTH in

lowering circulating concentrations (and presumably renal intracellular concentrations) of phosphate (Fig. 340-6). 1,25(OH)$_2$D also influences the renal metabolism of 25(OH)D by diminishing 25(OH)D-1a-hydroxylase activity and enhancing the metabolism of 25(OH)D to 24R,25-dihydroxyvitamin D [24,25(OH)$_2$D].

24,25(OH)$_2$D is normally present in serum at a concentration of 1 to 10 nmol/L (0.5 to 5.0 ng/mL). 24,25(OH)$_2$D is also a substrate for renal 25(OH)D-1a-hydroxylase and is converted to 1a,24R,25-trihydroxyvitamin D [1,24,25(OH)$_3$D], which, in turn, is metabolized to the biologically inactive substance calcitroic acid (Fig. 340-5). Cultured cells that possess nuclear receptors for 1,25(OH)$_2$D, such as chondrocytes, skin keratinocytes and fibroblasts, and intestinal and melanoma cells, also metabolize 25(OH)D to 24,25(OH)$_2$D. Studies of the vitamin D-24-hydroxylase null mice indicate that the major role of 24-hydroxylation is in the regulation of levels of 1,25(OH)$_2$D.

PHYSIOLOGY

1,25(OH)$_2$D, produced by the kidney and the placenta, is the only known important metabolite of vitamin D; the potential roles of other metabolites have not been clarified. 1,25(OH)$_2$D bound to a vitamin D-binding protein is delivered to various target organs, where the free form is taken up by cells and transported to a specific nuclear receptor protein. The vitamin D receptor (VDR) belongs to the nuclear receptor superfamily of steroid-retinoid-thyroid hormone-vitamin D transcription regulatory factors (Chap. 327). The VDR interacts with the retinoic acid X receptor (RXR) to form a heterodimeric (RXR-VDR) complex that binds to specific DNA sequences, termed the *vitamin D response elements* (VDREs). After 1,25(OH)$_2$D binds to the receptor, it induces conformational changes that result in the recruitment of a multitude of transcriptional coactivators that stimulate the transcription of target genes. In the intestine, the activated VDR stimulates calcium-binding protein synthesis; in bone, it stimulates production of osteocalcin, osteopontin, and alkaline phosphatase. 1,25(OH)$_2$D also may have nonnuclear effects on its target tissues; 1,25(OH)$_2$D increases the transport of calcium from the extracellular to intracellular space, and it can mobilize calcium from intracellular calcium pools and enhance phosphatidylinositol metabolism. In the intestine, the net effect of 1,25(OH)$_2$D is to stimulate calcium and phosphate transport from the lumen of the small intestine into the circulation (Fig. 340-6). The effect of 1,25(OH)$_2$D on the enhancement of bone resorption is synergistic with that of PTH. Mature osteoclasts do not possess receptors for either PTH or 1,25(OH)$_2$D. Both PTH and 1,25(OH)$_2$D interact with their specific receptors on osteoblasts or stromal fibroblasts to induce the production of RANK ligand on the osteoblast's cell surface. As described above, the RANK ligand interacts with the RANK receptor on immature osteoclasts, stimulating immature osteoclastic precursors to differentiate into mature osteoclasts. The role of 1,25(OH)$_2$D in the renal handling of calcium and phosphorus remains uncertain. Whatever the role of extraintestinal VDRs may be, the compelling evidence is that the phenotype of VDR null mice is corrected in the setting of normal mineral ion homeostasis. Thus the skeletal consequences of VDR ablation are the result of impaired intestinal calcium absorption and/or the accompanying secondary hyperparathyroidism and hypophosphatemia.

Receptors for 1,25(OH)$_2$D are also present in cells not classically considered target organs for this hormone, including skin, breast, pituitary, parathyroids, pancreatic beta

cells, gonads, brain, skeletal muscle, circulating monocytes, and activated B and T lymphocytes. Although its physiologic role in these cells remains to be determined, 1,25(OH)$_2$D inhibits proliferation of keratinocytes and fibroblasts, stimulates terminal differentiation of keratinocytes, induces monocytes to produce interleukin (IL)1 and to differentiate into macrophages and osteoclast-like cells, inhibits the production of PTH, and inhibits the production of IL-2 and immunoglobulin by activated T and B lymphocytes, respectively.

In addition, a variety of tumor cell lines, including lines derived from breast carcinomas, melanomas, and promyeloblasts, possess receptors for 1,25(OH)$_2$D. Tumor cell lines that have 1,25(OH)$_2$D receptors respond to the hormone by decreasing the rate of proliferation and enhancing differentiation. For example, when malignant receptor-positive human promyelocytic cells (HL-60) are exposed to 1,25(OH)$_2$D, the cells mature into functioning macrophages within 1 week. Although calcitriol [1,25(OH)$_2$D] is not useful for the treatment of leukemia, the antiproliferative effects of calcitriol and its analogue calcipotriene provide the rationale for their use in the treatment of psoriasis.

1,25(OH)$_2$D regulates PTH synthesis by negative feedback (Fig. 340-6). This effect is the rationale for giving 1,25(OH)$_2$D$_3$, and its less calcemic-inducing analogue 19-nor-1,25-dihydroxyvitamin D$_3$ (Fig. 340-7), to lower circulating levels of PTH in patients with chronic renal failure (Chap. 341).

The principal physiologic mechanism regulating the production of 1,25(OH)$_2$D appears to involve changes in serum extracellular calcium concentrations that result in reciprocal changes in secretion of PTH, the latter controlling, possibly through actions on serum or tissue phosphorus levels, the rate of 1,25(OH)$_2$D production. Other factors that enhance 1,25(OH)$_2$D production include estrogen, prolactin, and growth hormone. Humans adapt to increased calcium requirements during growth, pregnancy, and lactation by increasing the efficiency of intestinal calcium absorption, possibly by enhancing 25(OH)D-1a-hydroxylase activity. During the first two trimesters of pregnancy, the levels of 1,25(OH)$_2$D increase in proportion to the concentration of the vitamin D-binding protein; levels of free 1,25(OH)$_2$D do not change. During the last trimester, the need for calcium for mineralization of the fetal skeleton is met by an increase in the concentrations of free 1,25(OH)$_2$D and enhanced maternal intestinal calcium absorption.

Most measurements of circulating 1,25(OH)$_2$D in various physiologic or pathologic states utilize a receptor/competitive binding assay. Serum levels of vitamin D and 25(OH)D vary with the season and with vitamin D intake, whereas levels of 1,25(OH)$_2$D appear to be unaltered by seasonal variation, by increases in dietary vitamin D, or by exposure to sunlight (Table 340-6); as long as vitamin D supplies and circulating concentrations of 25(OH)D are sufficient, metabolic influences control the renal 25(OH)D-1a-hydroxylase to ensure a closely regulated circulating concentration of 1,25(OH)$_2$D. The serum concentration of 1,25(OH)$_2$D ranges from 40 to 160 pmol/L (16 to 65 pg/mL), and its serum half-life is from 3 to 6 h.

PHARMACOLOGY

Casual exposure to sunlight provides most people with adequate vitamin D. In elderly

individuals, exposure of hands, face, and arms to a suberythemal dose of sunlight two to three times a week is usually adequate. A variety of over-the-counter vitamin preparations contain 400 IU of either vitamin D_2(ergocalciferol) or vitamin D_3(cholecalciferol). More potent preparations of vitamin D (calciferol) are available in capsule and tablet form (50,000 IU), as oil (500,000 IU/mL), and in oral solution (8000 IU/mL). A single oral dose of 50,000 IU of vitamin D_2increases the circulating concentration of vitamin D from<25 nmol/L (10 ng/mL) to 130 to 260 nmol/L (50 to 100 ng/mL) within 12 to 24 h; the plasma half-life is about 2 days. Serum concentrations of 25(OH)D and 1,25(OH)$_2$D are not changed by these doses of vitamin D. For treatment of vitamin D deficiency, 50,000 IU of vitamin D once a week for 8 weeks raises the circulating concentration of 25(OH)D into the normal range; in the presence of secondary hyperparathyroidism, the circulating concentrations of 1,25(OH)$_2$D can increase to supranormal levels [up to 600 pmol/L (250 pg/mL)]. 25(OH)D_3(calcifediol) available in capsules containing either 20 or 50 ug may be useful in treating vitamin D deficiency [low 25(OH)D concentrations] in patients with severe liver dysfunction. Pharmacologic doses are used to treat disorders of 25(OH)D metabolism; in pharmacologic doses, 25(OH)$D_3$is believed to act via interaction with theVDR. Calcitriol is available in capsules containing 0.25 or 0.5 ug and as a solution for intravenous use (1.0 and 2.0 ug/mL). Calcitriol is efficacious in a variety of disorders (Chap. 341), but even low doses can cause hypercalcemia, leading to attempts to develop analogues with less calcemic activity. Two such calcitriol analogues have been approved in the United States for the treatment of renal osteodystrophy; 19-nor-1,25-dihydroxyvitamin D_2, and 24-epi-1,25-dihydroxyvitamin D_2(Fig. 340-7). 1a-Hydroxyvitamin D_3[1(OH)D_3] is a potent 1,25(OH)$_2$$D_3$agonist that is used in Europe and Japan. The structure of this analogue is identical to that of the natural renal hormone with the exception that it lacks a C_{25}OH. In humans, this analogue is rapidly metabolized by the liver to 1,25(OH)$_2$$D_3$. Topical preparations of calcitriol (3 ug/g) in Europe and calcipotriene (50 ug/g) in Europe and the United States are used for the treatment of psoriasis. When applied over a large surface area, both can potentially cause hypercalcemia and hypercalciuria. Oral calcitriol is also effective for psoriasis and psoriatic arthritis.

When vitamin D is chemically manipulated to rotate the A ring through 180°, the C_3b-OH assumes a geometric position that mimics the C_1a-OH (Fig. 340-7). These compounds, called *pseudo-1a-hydroxyvitamin D analogues*, include the clinically useful dihydrotachysterol (DHT). This analogue is less effective in stimulating intestinal calcium transport on a weight basis than either vitamin D or 1,25(OH)$_2$D. Because it does not require 1a-hydroxylation to be active on intestinal calcium transport, it is 3 to 10 times more potent than vitamin D in disease states that impair renal 25(OH)D-1a-hydroxylase, such as hypoparathyroidism and chronic renal failure. Dihydrotachysterol is efficiently metabolized in the liver to 25-hydroxy-DHT, which is the biologically active form.

RICKETS AND OSTEOMALACIA

Rickets and osteomalacia are disorders in which mineralization of the organic matrix of the skeleton is defective. These disorders are caused by a number of different conditions associated with vitamin D deficiency or resistance (Table 340-7). In *rickets*, the growing skeleton is involved; defective mineralization occurs both in the bone and cartilaginous matrix of the growth plate. The term *osteomalacia* is usually used for this mineralization disorder in the adults in whom the epiphyseal growth plates are closed.

For normal skeletal mineralization, sufficient calcium and phosphate must be present at the mineralization sites. In addition, intact metabolic and transport functions of osteoblasts and chondrocytes and adequate production of cross-linked collagen matrix are required. In cartilage, the initial mineral phase is enclosed in membrane-bound extracellular vesicles. If the osteoblast continues to produce matrix components that cannot be mineralized adequately, rickets or osteomalacia results. A characteristic feature of these disorders is therefore an increase in osteoid volume and thickness (the latter being normally <12 to 14 um) and a decrease in calcification of the mineralization front. This can be detected in unmineralized sections by the fluorescence of previously ingested tetracyclines or by special stains. The inadequate mineralization of the matrix of cartilage in growing children leads to a widening of the epiphyseal plates of the long bones due to a disorganization of the otherwise highly ordered columns of hypertrophied cartilage cells. In addition, the poorly mineralized long bones are incapable of withstanding usual mechanical stresses and tend to undergo bowing deformities. Growth of the epiphyseal plates is diminished, stunting the growth of the long bones. Osteomalacia also compromises the architectural structure and strength of the skeleton in adults, causing an increase in fractures.

PATHOPHYSIOLOGY

A large number of disorders are associated with rickets or osteomalacia, primarily through alterations of vitamin D nutrition or metabolism or because of phosphate wasting (Table 340-7). *Hypovitaminosis D* results from inadequate endogenous production of vitamin D_3 in the skin, from insufficient dietary supplementation, and/or from an inability of the small intestine to absorb adequate amounts of the vitamin from the diet. Resistance to the effects of vitamin D can result from (1) use of drugs that antagonize vitamin D action, (2) alterations in the metabolism of vitamin D, or (3) deficient or defective receptors for 1,25(OH)$_2$D. The consequences of hypovitaminosis D include (1) disturbances of mineral ion metabolism and secretion of PTH, and (2) mineralization defects in the skeleton (e.g., rickets in children, osteomalacia in adults). With an adequate glomerular filtration rate (GFR), the main changes are hypophosphatemia, normal or near-normal serum calcium levels, increased levels of PTH, and low levels of 25(OH)D (Table 340-5).

With regard to calcium metabolism, lack of vitamin D action leads to insufficient intestinal calcium absorption and hypocalcemia. The latter stimulates the secretion of PTH (secondary hyperparathyroidism), which enhances calcium release from bone, decreases calcium clearance by the kidney, and tends to blunt the hypocalcemia; as a consequence, most patients have a normal or low-normal serum calcium level. (Late in the course of untreated hypovitaminosis D, severe hypocalcemia develops.) Hypophosphatemia is more marked than hypocalcemia, especially in early stages of the deficiency. The efficiency of intestinal phosphate absorption is also decreased. The increased secretion of PTH, although partially effective in minimizing hypocalcemia, leads to urinary phosphate wasting because of decreased renal tubular reabsorption. This latter effect is the most significant factor in causing hypophosphatemia. Aging decreases the responsiveness of the renal 25(OH)D-1-hydroxylase to PTH, decreasing circulating levels of 1,25(OH)$_2$D and contributing to decreased calcium absorption in the elderly.

Although the conversion of vitamin D to 25(OH)D is impaired in severe chronic liver disease, there is not a strong correlation between low serum 25(OH)D levels and osteopenia. Patients with nephrotic syndrome with >4 g/d of proteinuria may have low 25(OH)D levels owing to loss in the urine of the vitamin D-binding protein with its associated tightly bound 25(OH)D. Circulating levels of 25(OH)D may also be decreased when the metabolism of 25(OH)D to 1,25(OH)$_2$D is increased, as in sarcoidosis and hyperparathyroidism. Chronic anticonvulsant therapy can also cause the development of osteomalacia or rickets; mineralization defects are worse in patients receiving multiple drug therapy and when vitamin D intake or exposure to sunlight is inadequate. Anticonvulsant drugs have multiple effects on calcium metabolism. Phenobarbital induces hepatic microsomal enzymes, alters the kinetics of the vitamin D-25-hydroxylase, and stimulates bile secretion, resulting in decreased serum concentrations of vitamin D and 25(OH)D. Phenytoin and phenobarbital inhibit intestinal calcium transport and bone mineral mobilization, independent of effects on vitamin D metabolism.

Glucocorticoids in high doses cause osteoporosis but do not induce osteomalacia and rickets (Chap. 342). Glucocorticoids directly inhibit vitamin D-mediated intestinal calcium absorption and bone mineral mobilization. Patients receiving glucocorticoids chronically may have depressed circulating levels of 1,25(OH)$_2$D; the mechanism(s) is unknown. Glucocorticoids also exert direct effects to induce apoptosis of osteoblasts and osteocytes.

A genetic defect in the hepatic 25-hydroxylation of vitamin D has not been described, but in one inherited disorder of calcium and bone metabolism, renal production of 1,25(OH)$_2$D is defective because of recessive mutations 25(OH)D-1a-hydroxylase activity. In this syndrome of pseudovitamin D-deficient rickets (also known as vitamin D-dependent rickets type I), renal production of 1,25(OH)$_2$D is impaired, circulating levels of 1,25(OH)$_2$D are low, but the therapeutic response to physiologic doses of calcitriol (0.25 to 1.0 ug/d) is normal. In patients with a similar phenotype, pseudovitamin D-resistant rickets (vitamin D-dependent rickets type II), mutations in the vitamin D receptor impair its function by altering the binding of the hormone to the receptor or by altering the binding of the receptor heterodimer complex to DNA. Individuals with this disorder have high circulating levels of 1,25(OH)$_2$D, but the hormone is ineffective because of the receptor defect. Alopecia is another feature of this disorder, suggesting a role for the VDR in hair follicle development.

Inherited forms of phosphate wasting disorders include X-linked hypophosphatemic rickets and autosomal dominant hypophosphatemic rickets. The gene responsible for the autosomal form is unknown but has been mapped to chromosome 12p13. The X-linked disorder is caused by mutations in the *PHEX* (phosphate-regulating gene with homology to endopeptidases on the X-chromosome) gene. The *PHEX* gene is postulated to encode an osteoblast protein that inactivates phosphatonin, a phosphaturic factor. Consequently, inactivating mutations result in phosphate wasting. The gene is expressed in heterozygotes, suggesting that the disorder is caused by haploinsufficiency. In patients with X-linked hypophosphatemic rickets, serum concentrations of 1,25(OH)$_2$D are normal or low. Since hypophosphatemia is a potent stimulator of the renal 25(OH)D-1a-hydroxylase, levels of 1,25(OH)$_2$D should be high,

which suggests the existence of a functional defect in the 25(OH)D-1a-hydroxylase in this disorder. Therefore, the combination of calcitriol and phosphate supplements is more effective than therapy with phosphate supplements alone.

Patients with hypocalcemia due to hypoparathyroidism or pseudohypoparathyroidism have lower-than-normal mean serum concentrations of 1,25(OH)$_2$D, although individual values may be in the normal range. In these patients, small replacement doses of calcitriol (0.25 to 1.0 ug/d) are effective even when the serum 25(OH)D concentrations are elevated. These observations indicate that absent or ineffective action of PTH decreases the activity of renal 25(OH)D-1a-hydroxylase. It is not known whether serum 1,25(OH)$_2$D levels would be restored if the hyperphosphatemia were controlled.

Patients with tumor-induced (oncogenic) osteomalacia have low levels of serum phosphorus and 1,25(OH)$_2$D. Some of these tumors, especially malignant carcinomas, produce PTHrP (Chap. 341), leading to hypercalcemia as well as hypophosphatemia. Other tumors, particularly more benign neoplasms of vascular or mesenchymal origin, may be responsible for severe hypophosphatemia in the presence of normocalcemia. The mechanism for inhibition of 1,25(OH)$_2$D synthesis remains unknown; after removal of the tumor, however, the serum phosphorus and 1,25(OH)$_2$D levels return to normal.

It has been suggested that alteration of vitamin D receptor levels in target tissues (such as intestine) could affect calcium and bone metabolism, and bone mineral density appears to be associated with specific polymorphisms of the VDR gene. These polymorphisms affect the noncoding intervening DNA sequences (introns) or the coding sequence in a way that does not alter the amino acid sequence of the VDR. Some reports suggest that polymorphisms involving the endonucleases Bsm-I and Taq-I (bb, TT) are associated with higher bone mineral density; other studies have not confirmed these findings. The genetic contribution of VDR polymorphic variants to bone mineral density, as well as a number of other diseases with which they have been associated, require additional, larger-scale studies.

CLINICAL FEATURES

The clinical manifestations of rickets are the result of skeletal deformities, susceptibility to fractures, weakness and hypotonia, and disturbances in growth. As the disorder progresses, particularly that associated with vitamin D deficiency, children are unable to walk without support due to the skeletal deformities in the lower limbs and severe muscle weakness (Fig. 340-8). Abnormal parietal flattening and frontal bossing develops in the skull. The calvariae are softened (*craniotabes*) and sutures may be widened. Prominence of the costochondral junctions is called the *rachitic rosary* and the indentation of the lower ribs at the site of attachment of the diaphragm is known as *Harrison's groove*. If untreated, deformities of the pelvis and extremities progress, with bowing being particularly common in the tibia, femur, radius, and ulna. For women, the flattening of the pelvis increases the risk of maternal and infant morbidity and mortality during childbirth. Fractures are frequent, dental eruption is often delayed, and enamel defects are common.

The presentation of osteomalacia in adults is usually more insidious. Bone pain and muscle weakness are common complaints and may be overlooked as indicators of

vitamin D deficiency. It is estimated in the United States and Europe that>40% of the adult population over the age of 50 are vitamin D deficient. Although the standard lower limit of the normal range for 25(OH)D is 10 ng/mL, several studies suggest that the cutoff for vitamin D deficiency should be increased to 20 ng/mL. This suggestion is based on PTH responses to various serum levels of 25(OH)D. For example, elevated levels of PTH are frequently seen in individuals with 25(OH)D levels between 10 and 20 ng/mL, and administration of vitamin D (50,000 units of vitamin D once a week for 8 weeks) lowers PTH levels. Defects in skeletal mineralization may accompany these disturbances in vitamin D and mineral metabolism.

Pain in the hips may result in an antalgic gait. Muscle weakness is often associated with osteomalacia but is difficult to distinguish from hesitancy to move because of skeletal pain. Proximal weakness may mimic that of primary muscle disorders and contribute to the waddling gait. Many factors, including secondary hyperparathyroidism, hypophosphatemia, and vitamin D deficiency, contribute to the myopathy. Fractures of the involved bones may occur with minimal trauma. When the ribs are involved, severe deformities may develop in the thorax, and the collapse of the vertebral bodies may produce loss of height.

RADIOLOGIC FEATURES

In rickets, the most prominent radiologic alteration is evident at the growth plate (physis) which is increased in thickness, cupped, and hazy at the metaphyseal border owing to decreased calcification of the hypertrophic zone and inadequate mineralization of the primary spongiosa. The trabecular pattern of the metaphysis is abnormal, the cortices of the diaphysis may be thinned, and the shafts may be bowed.

In osteomalacia, a decrease in bone density is usually associated with loss of trabeculae and thinning of the cortices. Radiologic and bone densitometric changes are indistinguishable from those in osteoporosis (Chap. 342). Trabecular patterns may be blurred, producing a homogeneous "ground glass" appearance. Radiolucent bands, ranging from a few millimeters to several centimeters in length and usually oriented perpendicular to the surface of the bones, suggest the presence of osteomalacia. They are particularly common at the inner aspects of the femur (especially near the femoral neck), in the pelvis, in the outer edge of the scapula, in the upper fibula, and in the metatarsals. These radiolucent bands, called *pseudofractures* or *Looser's zones*, occur most often where major arteries cross the bones and are thought to be due to the pulsation of these vessels in the undermineralized area (Fig. 340-9). Increased rather than decreased density of bones may be observed in patients who have renal tubular disorders rather than vitamin D deficiency and may produce a striking thickening of the cortices and a trabecular pattern of the spongy bone. Despite the increase in bone mass per unit volume, the trabeculae are covered with thickened osteoid seams typical of osteomalacia. Similar findings may occur in patients with chronic renal failure. The reason for the hyperostosis is unknown; the bone is architecturally abnormal and is subject to fracture with minimal trauma.

LABORATORY FINDINGS

Changes in serum concentration of calcium, inorganic phosphorus, 25(OH)D, and

1,25(OH)$_2$D vary with the different disorders. In vitamin D deficiency, whether due to dietary lack, inadequate sunlight exposure, or intestinal malabsorption, serum calcium levels are normal or low, whereas phosphorus and 25(OH)D levels are consistently low, the latter usually <15 nmol/L (<10 ng/mL) depending on the assay used. In contrast, levels of 1,25(OH)$_2$D may be normal or elevated owing to secondary hyperparathyroidism and the fact that circulating levels of 1,25(OH)$_2$D are 1000-fold less than those of 25(OH)D. Only when vitamin D deficiency is chronic and severe is hypocalcemia observed. It may be sufficiently severe to produce tetany. Mild acidosis and generalized aminoaciduria result from secondary hyperparathyroidism. Patients with renal tubular disorders have normal serum calcium levels and hypophosphatemia. Other laboratory findings such as glucosuria, aminoaciduria, acidosis, and hypouricemia reflect variable degrees of disturbance of proximal tubular function or are features of underlying disease (e.g., low plasma ceruloplasmin in Wilson's disease or abnormalities of immunoglobulins in multiple myeloma). In chronic renal failure, hyperphosphatemia and hypocalcemia are usually accompanied by normal 25(OH)D and low 1,25(OH)$_2$D levels. In nephrotic syndrome, serum 25(OH)D levels can be low owing primarily to urinary losses of vitamin D binding protein-bound 25(OH)D. Serum phosphorus levels are also normal or elevated in hypophosphatasia. Markers of bone resorption increase when secondary hyperparathyroidism and excessive bone resorption are associated with the defect in mineralization. Alkaline phosphatase levels in serum are usually elevated in rickets and osteomalacia.

TREATMENT

In rickets and osteomalacia due to dietary absence of vitamin D or inadequate exposure to sunlight, vitamin D$_2$(ergocalciferol) or vitamin D$_3$(cholecalciferol) is given orally in doses of 800 to 4000 IU (0.02 to 0.1 mg) daily for 6 to 12 weeks, followed by daily supplements of 200 to 600 IU, which are adequate to prevent the development of the disorder in otherwise normal persons. In elderly persons with vitamin D deficiency, the administration of 50,000 IU vitamin D by mouth once each week for 8 weeks raises the serum levels of 25(OH)D into the mid-normal range. In infants and children, such treatment causes improvement in muscle tone and strength, an increase in serum calcium and phosphorus levels, and a decrease in alkaline phosphatase levels after several weeks. Radiologic evidence of healing appears within weeks, and healing may be complete by a few months. Calcium supplements and larger initial doses of vitamin D may be necessary in infants and children with tetany. In adults with nutritional osteomalacia, healing of pseudofractures may be evident within 3 to 4 weeks after therapy with as little as 2000 IU (0.5 mg) vitamin D daily. Healing is usually complete by 6 months.

Patients with osteomalacia due to intestinal malabsorption do not respond to small doses of vitamin D. In the presence of active steatorrhea, daily oral doses of vitamin D of 50,000 to 100,000 IU (1.25 to 2.5 mg) and large doses of calcium (e.g., 15 g calcium lactate or 4 g calcium carbonate orally per day) may be required. In some instances, oral vitamin D is ineffective, and the parenteral route is required (e.g., 10,000 IU/d intramuscularly). Another approach is the use of artificial ultraviolet B radiation or exposure to sunlight in addition to supplemental calcium. Small doses of calcitriol (0.5 to 1.0 ug daily) are also usually effective in this form of osteomalacia. Inorganic phosphate therapy is not indicated either in deficiency or in intestinal malabsorption of the vitamin,

since hypocalcemia will develop and intestinal calcium absorption will remain inadequate. In all patients in whom large doses of vitamin D are used, serum calcium and 25(OH)D levels should be monitored periodically. Semiquantitative urinary calcium measurements are inadequate.

In patients treated with multiple anticonvulsant agents, it is usually necessary to continue the drugs while adding 1000 IU/d of vitamin D and to monitor levels of serum calcium and serum 25(OH)D until a therapeutic response (evidence of radiologic healing, improvement in symptoms) is obtained.

Treatment of rickets and osteomalacia in the presence of renal tubular disorders is more difficult. Oral supplements of inorganic phosphate in divided doses of phosphorus (as elemental P), 1.0 to 3.6 g/d (50 mg/kg body weight per day for children), and calcitriol, 0.5 to 2.0 ug/d (30 ng/kg body weight per day for children), constitute the best regimen to restore skeletal growth and heal the bone disease. Patients with nephrotic syndrome and low serum 25(OH)D levels benefit from modest vitamin D supplementation (800 to 1000 IU/d). Small doses of calcitriol are equally effective in treating hypocalcemia and osteodystrophy resulting from chronic renal failure. The recommended initial dose of calcitriol is 0.25 ug/d. If after 2 to 4 weeks on this dose the biochemical parameters are unaltered, the dose is increased by 0.25 ug/d every 2 to 4 weeks until a satisfactory clinical biochemical response (including elevation of serum calcium levels and decrease in PTHlevels) is obtained. The usual dose is 0.5 to 1.0 ug/d. Calcitriol may also be administered intravenously (1.0 to 2.5 ug three times weekly) to patients on dialysis, particularly to treat refractory osteitis fibrosa. Serum calcium levels should be monitored frequently during the first 1 to 2 months of therapy and less frequently once a stable dose has been established.

In patients who have had rickets in childhood, the abnormal mechanical stress of severe deformities may contribute to the development of degenerative joint disease, particularly in the hips and knees. Osteotomies at the proper time after healing may prevent this complication and the requirement for more extensive arthroplasties later in life.

(Bibliography omitted in Palm version)

341. DISEASES OF THE PARATHYROID GLAND AND OTHER HYPER- AND HYPOCALCEMIC DISORDERS - *John T. Potts, Jr*

The four parathyroid glands are located posterior to the thyroid gland. They produce parathyroid hormone (PTH), which is the primary regulator of calcium physiology. PTH acts directly on bone, where it induces calcium resorption, and on the kidney, where it stimulates calcium reabsorption and synthesis of 1,25-dihydroxyvitamin D [1,25(OH)$_2$D], a hormone that stimulates gastrointestinal calcium absorption. Serum PTH levels are tightly regulated by a negative feedback loop. Calcium, acting through the calcium-sensing receptor, and vitamin D, acting through its nuclear receptor, inhibit PTH synthesis and release. Understanding the hormone pathways that regulate calcium levels and bone metabolism is essential for effective diagnosis and management of a wide array of hyper- and hypocalcemic disorders.

Hyperparathyroidism, characterized by excess production of PTH, is a common cause of hypercalcemia and is usually the result of autonomously functioning adenomas or hyperplasia. Surgery for this disorder is highly effective and has been shown recently to reverse some of the deleterious effects of long-standing PTH excess on bone density. Hypercalcemia of malignancy is also common and is usually due to the overproduction of parathyroid hormone-related peptide (PTHrP) by cancer cells. The similarities in the biochemical characteristics of hyperparathyroidism and hypercalcemia of malignancy, first noted by Albright in 1941, are now known to reflect the actions of PTH and PTHrP through the same G protein-coupled (GPC) PTH/PTHrP receptor.

Clarification over the past decade of genetic influences on parathyroid gland and bone cell function helps in constructing a logical approach to hyper- and hypocalcemic disorders. Advances that have occurred include elucidation of the genetic basis of multiple endocrine neoplasia (MEN) types 1 and 2, familial hypocalciuric hypercalcemia (FHH), the different forms of pseudohypoparathyroidism (PHP), Jansen's syndrome, disorders of vitamin D synthesis and action, and the molecular events associated with parathyroid gland neoplasia. The advent of new drugs, including bisphosphonates and selective estrogen receptor modulators (SERMs), offers new avenues for the treatment and prevention of metabolic bone disease. PTH analogues are promising therapeutic agents for the treatment of postmenopausal or senile osteoporosis, and calcimimetic agents, which act through the calcium-sensing receptor, may provide new approaches for PTH suppression.

PARATHYROID HORMONE

PHYSIOLOGY

The primary function of PTH is to maintain the extracellular fluid (ECF) calcium concentration within a narrow normal range. The hormone acts directly on bone and kidney and indirectly on intestine through its effects on synthesis of 1,25(OH)$_2$D to increase serum calcium concentrations; in turn, PTH production is closely regulated by the concentration of serum ionized calcium. This feedback system is the critical homeostatic mechanism for maintenance of ECF calcium. Any tendency toward hypocalcemia, as might be induced by calcium-deficient diets, is counteracted by an increased secretion of PTH. This in turn (1) acts to increase the rate of dissolution of

bone mineral, thereby increasing the flow of calcium from bone into blood; (2) reduces the renal clearance of calcium, returning more of the calcium filtered at the glomerulus into ECF; and (3) increases the efficiency of calcium absorption in the intestine. Immediate control of blood calcium is probably due to effects of the hormone on bone and, to a lesser extent, on renal calcium clearance. Maintenance of steady-state calcium balance, on the other hand, probably results from the effects of $1,25(OH)_2D$ on calcium absorption (Chap. 340). The renal actions of the hormone are exerted at multiple sites and include inhibition of phosphate transport (proximal tubule), increased reabsorption of calcium (distal tubule), and stimulation of the renal 25(OH)D-1a-hydroxylase. As much as 12 mmol (500 mg) calcium is transferred between the ECF and bone each day (a large amount in relation to the total ECF calcium pool), and PTH has a major effect on this transfer. The homeostatic role of the hormone can preserve calcium concentration in blood acutely at the cost of bone destruction.

PTH has multiple actions on bone, some direct and some indirect. It increases the rate of calcium release from bone into blood acutely; PTH-mediated changes in bone calcium release can be seen within minutes. The chronic effects of PTH are to increase the number of bone cells, both osteoblasts and osteoclasts, and to increase the remodeling of bone; these effects are apparent within hours after the hormone is given and persist for hours after PTH is withdrawn. Continuous exposure to elevated levels of PTH for days (as in hyperparathyroidism or long-term infusions in animals) leads to increased osteoclast-mediated bone resorption. However, the administration of PTH intermittently over days in animals or osteoporotic patients leads to a net stimulation of bone formation rather than bone breakdown. Striking increases, especially in trabecular bone in the spine and hip, have been reported with the use of PTH in combination with estrogen. PTH as monotherapy caused a highly significant reduction in fracture incidence in a worldwide placebo-controlled trial.

Osteoblasts (or stromal cell precursors), which have PTH receptors, are crucial to this bone-forming effect of PTH; osteoclasts, which appear to lack PTH receptors, mediate bone breakdown. PTH-mediated stimulation of osteoclasts is believed to be indirect, acting in part through cytokines released from osteoblasts to activate osteoclasts; in experimental studies of bone resorption in vitro, osteoblasts must be present for PTH to activate osteoclasts to resorb bone. The nature of the cytokines that stimulate osteoclasts is a subject of major interest. Insulin-like growth factor 1, interleukin 6, granulocyte-macrophage colony stimulating factor, and possibly other agents are candidates, but the definitive messenger(s) has not been determined. Direct cell-to-cell contact between osteoblasts (stromal cells) and osteoclast precursors is also key to osteoclast function. Cell-associated ligands and receptors, as well as soluble decoy receptors, are involved in these interactions (Chap. 340).

STRUCTURE

PTH is an 84-amino-acid single-chain peptide. The amino acid sequence of PTH has been characterized in multiple mammalian species, revealing marked conservation in the amino-terminal portion, which is critical for many biologic actions of the molecule. Synthetic fragments of the amino-terminal sequence as small as 1-14 residues are sufficient to activate the major receptor (see below). Biologic roles for the

carboxyl-terminal region of PTH are under investigation; a separate receptor may exist for this region of the molecule. Fragments shortened or modified at the amino terminus still bind to the PTH receptor but lose the capacity to stimulate biologic responses. For example, the peptide composed of sequences 7-34 is a competitive inhibitor of active hormone binding to receptors in vitro but is a weak inhibitor in vivo.

BIOSYNTHESIS, SECRETION, AND METABOLISM

Synthesis Parathyroid cells have multiple methods of adapting to increased needs for PTH production. Most rapid (within minutes) is secretion of preformed hormone in response to hypocalcemia. Second, within hours, changes in gene activity and increased PTH mRNA are induced by sustained hypocalcemia. Finally, protracted challenge leads within days to cellular replication to increase gland mass.

PTH is initially synthesized as a larger molecule (preproparathyroid hormone, consisting of 115 amino acids), which is then reduced in size by a second cleavage (proparathyroid hormone, 90 amino acids) before secretion as the 84-amino-acid peptide. The hydrophobic regions of the preproparathyroid hormone serve a role in guiding transport of the polypeptide from sites of synthesis on polyribosomes through the endoplasmic reticulum to secretory granules. In one kindred with hypoparathyroidism, a mutation in the preprotein region of the gene disrupts this critical hydrophobic sequence and interferes with hormone secretion.

Studies with cloned and expressed PTH genes in vitro have demonstrated DNA regions involved in transcriptional control, including sites for interaction and regulation by the vitamin D receptor, as well as sites through which ambient calcium regulates transcription. Suppression of PTH gene activity at the transcriptional level by calcium is nearly maximal at physiologic concentrations; hypercalcemia results in no significant change. Hypocalcemia, however, increases transcriptional activity within hours. $1,25(OH)_2D_3$ strongly suppresses PTH gene transcription, though not when chronic hypocalcemia is induced experimentally in animals. In patients with renal failure, however, intravenous administration of supraphysiologic levels of $1,25(OH)_2D_3$ or analogues of the active metabolite can dramatically suppress PTH overproduction, which is sometimes difficult to control due to severe secondary hyperparathyroidism. Control over hormone stores is exerted by variation in the rates of proteolytic destruction of preformed hormone under the control of ECF calcium; high calcium increases and low calcium inhibits the proteolytic destruction of hormone stores. Regulation of hormone precursor processing and proteolytic destruction of preformed hormone (posttranslational regulation of hormone production) is an important mechanism for mediating rapid (minutes) changes in hormone availability.

Regulation of PTH Secretion PTH secretion increases steeply to a maximum value of five times the basal rate of secretion as calcium concentration falls from normal to the range of 1.9 to 2.0 mmol/L (7.5 to 8.0 mg/dL) (measured as total calcium). The ionized fraction of blood calcium is the important determinant of hormone secretion. Magnesium may influence hormone secretion in the same direction as calcium. It is unlikely, however, that physiologic variations in magnesium concentration affect PTH secretion. Severe intracellular magnesium deficiency impairs PTH secretion (see below).

The level of ECF calcium controls PTH secretion by interaction with a calcium sensor, a GPCR for which Ca_{2+} ions act as the ligand (see below). This receptor is a member of a distinctive subfamily of the GPCR superfamily that is characterized by a large extracellular domain suitable for "clamping" the small-molecule ligand. Stimulation of the receptor by high calcium levels leads to suppression of PTH secretion. The intracellular signals generated by the active receptor appear to be inositol triphosphate (IP_3) and diacylglycerol (DAG) formed by activation of phospholipase. The receptor is present in parathyroid glands and the calcitonin-secreting cells (C cells) of the thyroid, brain, and kidney. Genetic evidence has revealed a key biologic role for the calcium-sensing receptor in parathyroid gland responsiveness to calcium and, unexpectedly, in renal calcium clearance. Point mutations associated with loss of function cause a syndrome resembling hyperparathyroidism (FHH) but with hypocalciuria. On the other hand, gain-of-function mutations cause a form of hypocalcemia resembling hypoparathyroidism (see below).

Metabolism The secreted form of PTH is indistinguishable by immunologic criteria and by molecular size from the 84-amino-acid peptide (PTH 1-84) extracted from glands. However, much of the immunoreactive material found in the circulation is smaller than the extracted or secreted hormone. The principal circulating fragments of immunoreactive hormone lack a portion of the critical amino-terminal sequence required for biologic activity and, hence, are biologically inactive fragments (so-called middle- and carboxyl-terminal fragments). Much of the proteolysis of hormone occurs in the liver and kidney. However, fragments corresponding to the middle- and carboxyl-terminal portions have also been detected in effluent blood from the parathyroids and in the peripheral circulation; there is no convincing evidence, however, for circulating amino-terminal fragments. Peripheral metabolism of PTH does not appear to be regulated by physiologic states (high versus low calcium, etc.); hence peripheral metabolism of hormone, although responsible for rapid clearance of secreted hormone, appears to be a high-capacity, metabolically invariant catabolic process.

The rate of clearance of the secreted 84-amino-acid peptide from blood is more rapid than the rate of clearance of the biologically inactive fragment(s) corresponding to the middle- and carboxyl-terminal regions of PTH. Consequently, the interpretation of PTH immunoassays is influenced by the nature of the peptide fragments detected by the antibodies. Before the introduction of double-antibody assays designed to detect intact, biologically active hormone, most immunoassays also measured biologically inert long-lived fragments. Changes in the rate of production or clearance of fragments therefore alter the concentration of immunoreactive hormone.

Although the problems inherent in PTH measurements have been largely circumvented by use of double-antibody assays that detect only the intact molecule, new evidence has revealed the existence of a hitherto unappreciated larger PTH fragment that may affect the interpretation of most currently available double-antibody assays as well. A large amino-terminally truncated form of PTH, possibly PTH(7-84), is present in normal and uremic individuals in addition to PTH(1-84). The concentration of the putative 7-84 fragment relative to that of intact PTH(1-84) is higher with induced hypercalcemia (e.g., in uremic patients) than in eucalcemic or hypocalcemic conditions. The large fragment almost certainly cannot have (on the basis of structure-activity studies with PTH discussed above) much, if any, of the biologic potency of PTH. The suggestion that the

PTH(7-84)-like fragment might act as an inhibitor of PTH action remains to be clarified. The identification of this fragment has clinical significance, particularly in renal failure, as efforts to prevent secondary hyperparathyroidism by a variety of measures (vitamin D analogues, higher calcium intake, and phosphate-lowering strategies) may have led to oversuppression of biologically active intact PTH when the presence of the amino-terminally truncated PTH was not appreciated. The role, if any, of excessive PTH suppression due to inaccurate measurement of PTH in adynamic bone disease in renal failure (see below) is unknown. Newer assays with extreme amino-terminal epitopes are being studied intensively.

PARATHYROID HORMONE-RELATED PROTEIN

The paracrine factor termed *PTHrP* is responsible for most instances of hypercalcemia of malignancy, a syndrome that resembles hyperparathyroidism. Many different cell types produce PTHrP, including brain, pancreas, heart, lung, mammary tissue, placenta, endothelial cells, and smooth muscle. In fetal animals, PTHrP directs transplacental calcium transfer, and high concentrations of PTHrP are produced in mammary tissue and secreted into milk. Human and bovine milk, for example, contain very high concentrations of the hormone; the biologic significance of the latter is unknown. PTHrP may also play a role in uterine contraction and other biologic functions, still being clarified in other tissue sites.

PTH and PTHrP, although distinctive products of different genes, exhibit considerable functional and structural homology (Fig. 341-1) and may have evolved from a shared ancestral gene. The structure of the gene for human PTHrP, however, is more complex than that of PTH, containing multiple exons and multiple sites for alternate splicing patterns during formation of the mRNA. Protein products of 141, 139, and 173 amino acids are produced, and other molecular forms may result from tissue-specific degradation at accessible internal cleavage sites. The biologic roles of these various molecular species and the nature of the circulating forms of PTHrP are unclear. It is uncertain whether PTHrP circulates at any significant level in normal human adults; as a paracrine factor, PTHrP may be produced, act, and be destroyed locally within tissues. In adults PTHrP appears to have little influence on calcium homeostasis, except in disease states, when large tumors, especially of the squamous cell type, lead to massive overproduction of the hormone (Fig. 341-1).

PTH AND PTHRP HORMONE ACTION

Because PTHrP shares a significant homology with PTH in the critical amino terminus, it binds to and activates the PTH/PTHrP receptor, indistinguishably from effects seen with PTH. The 500-amino-acid PTH/PTHrP receptor (also known as the PTH1 receptor) belongs to a subfamily of GPCR that includes those for glucagon, secretin, and vasoactive intestinal peptide. The extracellular regions are involved in hormone binding, and the intracellular domains, after hormone activation, bind G protein subunits to transduce hormone signaling into cellular responses through stimulation of second messengers (Fig. 341-2). A second PTH receptor (PTH2 receptor) is expressed in brain, pancreas, and several other tissues. Its amino acid sequence and the pattern of its binding and stimulatory response to PTH and PTHrP differ from those of the PTH1 receptor. The PTH/PTHrP receptor responds equivalently to PTH and PTHrP, whereas

the PTH2 receptor responds only to PTH. The endogenous ligand and the physiologic significance of this receptor are not completely defined.

The PTH1 and PTH2 receptors can be traced backward in evolutionary time to fish. The zebrafish PTH1 and PTH2 receptors exhibit the same selective responses to PTH and PTHrP as do the human PTH1 and PTH2 receptors. The evolutionary conservation of structure and function suggests unique biologic roles for these receptors. Recently, a 39-amino-acid hypothalamic peptide, tubular infundibular peptide (TIP-39), has been characterized and is a likely natural ligand of the PTH2 receptor.

G proteins of the G_s class link the PTH/PTHrP receptor to adenylate cyclase, an enzyme that generates cyclic AMP, leading to activation of protein kinase A. Coupling to G proteins of the G_q class links hormone action to phospholipase C, an enzyme that generates inositol phosphates (e.g., IP_3) and DAG, leading to activation of protein kinase C and intracellular calcium release (Fig. 341-2). Studies using the cloned PTH/PTHrP receptor confirm that it can be coupled to more than one G protein and second-messenger kinase pathway, apparently explaining the multiplicity of pathways stimulated by PTH. Incompletely characterized second-messenger responses may be independent of phospholipase C or adenylate cyclase stimulation (the latter, however, is the strongest and best characterized second messenger signaling pathway for PTH).

The details of the biochemical steps by which an increased intracellular concentration of cyclic AMP, IP_3, DAG, and intracellular Ca_{2+} lead to ultimate changes in ECF calcium and phosphate ion translocation or bone cell function are unknown. Stimulation of protein kinases (A and C) and calcium transport channels is associated with a variety of hormone-specific tissue responses. These responses include inhibition of phosphate and bicarbonate transport, stimulation of calcium transport, and activation of renal 1a-hydroxylase in the kidney. The responses in bone include effects on collagen synthesis; increased alkaline phosphatase, ornithine decarboxylase, citrate decarboxylase, and glucose-6-phosphate dehydrogenase activities; DNA, protein, and phospholipid synthesis; and calcium and phosphate transport. Ultimately, these biochemical events lead to an integrated hormonal response in bone turnover and calcium homeostasis.

PTH also activates Na_{+}/Ca_{2+} exchanges in renal distal tubular sites and stimulates translocation of preformed calcium transport channels, moving them from the interior to the apical surface to mediate increased tubular uptake of calcium. PTH-dependent stimulation of phosphate excretion (blocking reabsorption -- the opposite effect from actions on calcium in the kidney) involves the sodium-dependent phosphate cotransporter, NPT-2, lowering its apical membrane content (and therefore function). Similar shifts may be involved in other renal tubular transport effects of PTH.

PTHrP exerts important developmental influences on fetal bone development and in adult physiology. A homozygous knockout of the PTHrP gene (or the gene for the PTH receptor) in mice causes a lethal deformity in which animals are born with severe skeletal deformities resembling chondrodysplasia (Fig. 341-3).

CALCITONIN (See also Chap. 339)

Calcitonin is a hypocalcemic peptide hormone that in several mammalian species acts as the physiologic antagonist to PTH. Calcitonin seems to be of limited physiologic significance in humans, at least in calcium homeostasis, as contrasted with a clearly definable role in calcium metabolism in many other mammalian species. Calcitonin is of medical significance, however, because of its role as a tumor marker in sporadic and hereditary cases of medullary carcinoma and its medical use as an adjunctive treatment in severe hypercalcemia and in Paget's disease of bone.

The hypocalcemic activity of calcitonin is accounted for primarily by inhibition of osteoclast-mediated bone resorption and secondarily by stimulation of renal calcium clearance. These effects are mediated by receptors on osteoclasts and renal tubular cells. Calcitonin exerts additional effects through receptors present in brain, gastrointestinal tract, and the immune system. The hormone, for example, exerts analgesic effects directly on cells in the hypothalamus and related structures, possibly by interacting with receptors for related peptide hormones, such as calcitonin gene-related peptide (CGRP) or amylin. The latter ligands have specific high-affinity receptors and also can bind to and trigger calcitonin receptors. The calcitonin receptors are homologous in structure to the PTH/PTHrP receptor.

The thyroid is the major source of the hormone, and the cells involved in calcitonin synthesis arise from neural crest tissue. During embryogenesis, these cells migrate into the ultimobranchial body, derived from the last branchial pouch. In submammalian vertebrates, the ultimobranchial body constitutes a discrete organ, anatomically separate from the thyroid gland; in mammals, the ultimobranchial gland fuses with and is incorporated into the thyroid gland.

The naturally occurring calcitonins consist of a peptide chain of 32 amino acids. There is considerable sequence variability among species. Calcitonin from salmon is 10 to 100 times more potent than mammalian forms in lowering serum calcium in animals. Calcitonin is synthesized as a precursor molecule that is four times larger than calcitonin itself. Analysis of the sequence of the coding portions of the gene for rat calcitonin indicates that at least two peptides flank calcitonin. It is likely (by analogy with the common precursor for adrenocorticotropic hormone and endorphin) that these peptides, of still uncharacterized biologic function, are released along with calcitonin.

There are two calcitonin genes, a and b, located on chromosome 11 in the general region of the b-globulin and PTH genes; the transcriptional control of these genes is complex. Two different mRNA molecules are transcribed from the a gene; one is translated into the precursor for calcitonin, and the other message is translated into an alternative product, CGRP. CGRP is synthesized wherever the calcitonin mRNA is expressed, e.g., in medullary carcinoma of the thyroid. The b, or CGRP-2, gene is transcribed into the mRNA for CGRP in the central nervous system (CNS); this gene does not produce calcitonin, however. CGRP has cardiovascular actions and may serve as a neurotransmitter or play a developmental role in the CNS.

The secretion of calcitonin is under the direct control of blood calcium. The circulating level of calcitonin in humans is lower than that in many other species. In humans, changes in calcium and phosphate metabolism are not seen despite extreme variations in calcitonin production; no definite effects are attributable to calcitonin deficiency (totally

thyroidectomized patients receiving only replacement thyroxine) or excess (patients with medullary carcinoma of the thyroid, a calcitonin-secreting tumor) (Chap. 339). Although there are no obvious abnormalities in calcium metabolism in patients with elevated calcitonin levels, bone remodeling is chronically suppressed. Calcitonin has been a useful pharmacologic agent to suppress bone resorption in Paget's disease (Chap. 343), has had limited use in the treatment of osteoporosis (Chap. 342), and is useful in early phases of treatment of severe hypercalcemia (see below).

HYPERCALCEMIA

Hypercalcemia can be a manifestation of a serious illness such as malignancy or can be detected coincidentally by laboratory testing in a patient with no obvious illness. The number of patients recognized with asymptomatic hypercalcemia, usually hyperparathyroidism, increased in the late twentieth century but is now declining somewhat, perhaps due to decreased use of routine blood calcium measurements or for other unknown reasons.

Whenever hypercalcemia is confirmed, a definitive diagnosis must be established. Although hyperparathyroidism, a frequent cause of asymptomatic hypercalcemia, is a chronic disorder in which manifestations, if any, may be expressed only after months or years, hypercalcemia can also be the earliest manifestation of malignancy, the second most common cause of hypercalcemia in the adult. The causes of hypercalcemia are numerous (Table 341-1), but hyperparathyroidism and cancer account for 90% of cases.

Before undertaking a workup, it is essential to be sure that true hypercalcemia, not a false-positive laboratory test, is present. A false-positive diagnosis of hypercalcemia is usually the result of inadvertent hemoconcentration during blood collection or elevation in serum proteins such as albumin. Hypercalcemia is a chronic problem, and it is cost-effective to obtain several serum calcium measurements; these tests need not be in the fasting state.

Clinical features are helpful in differential diagnosis. Hypercalcemia in an adult who is asymptomatic is usually due to primary hyperparathyroidism. In malignancy-associated hypercalcemia the disease is usually not occult; rather, symptoms of malignancy bring the patient to the physician, and hypercalcemia is discovered during the workup. In such patients the interval between detection of hypercalcemia and death is often<6 months. Accordingly, if an asymptomatic individual has had hypercalcemia or some manifestation of hypercalcemia, such as kidney stones, for >1 or 2 years, it is unlikely that malignancy is the cause. Nevertheless, differentiating primary hyperparathyroidism from *occult* malignancy can occasionally be difficult, and careful evaluation is required, particularly when the duration of the hypercalcemia is unknown. Hypercalcemia not due to hyperparathyroidism or malignancy can result from excessive vitamin D action, high bone turnover from any of several causes, or from renal failure (Table 341-1). Dietary history and a history of ingestion of vitamins or drugs are often helpful in diagnosing some of the less frequent causes. PTH immunoassays based on double-antibody methods serve as the principal laboratory test in differential diagnosis.

Hypercalcemia from any cause can result in fatigue, depression, mental confusion,

anorexia, nausea, vomiting, constipation, reversible renal tubular defects, increased urination, a short QT interval in the electrocardiogram, and, in some patients, cardiac arrhythmias. There is a variable relation from one patient to the next between the severity of hypercalcemia and the symptoms. Generally, symptoms are more common at calcium levels >2.9 to 3 mmol/L (11.5 to 12.0 mg/dL), but some patients, even at this level, are asymptomatic. When the calcium level is >3.2 mmol/L (13 mg/dL), calcification in kidneys, skin, vessels, lungs, heart, and stomach occurs and renal insufficiency may develop, particularly if blood phosphate levels are normal or elevated due to impaired renal function. Severe hypercalcemia, usually defined as ³3.7 to 4.5 mmol/L (15 to 18 mg/dL) can be a medical emergency; coma and cardiac arrest can occur.

Except in malignancy-associated hypercalcemia, acute management of the hypercalcemia is usually successful prior to definitive therapy. The type of treatment is based on the severity of the hypercalcemia and the nature of associated symptoms.

PRIMARY HYPERPARATHYROIDISM

Natural History and Incidence Primary hyperparathyroidism is a generalized disorder of calcium, phosphate, and bone metabolism due to an increased secretion of PTH. The elevation of circulating hormone usually leads to hypercalcemia and hypophosphatemia. There is great variation in the manifestations. Patients may present with multiple signs and symptoms, including recurrent nephrolithiasis, peptic ulcers, mental changes, and, less frequently, extensive bone resorption. However, with greater awareness of the disease and wider use of multiphasic screening tests, including blood calcium assays, the diagnosis is frequently made in patients who have no symptoms and minimal, if any, signs of the disease other than hypercalcemia and elevated levels of PTH. The manifestations may be subtle, and the disease may have a benign course for many years or a lifetime. Rarely, hyperparathyroidism develops or worsens abruptly and causes severe complications, such as marked dehydration and coma, so-called hypercalcemic parathyroid crisis.

The annual incidence of the disease is estimated to be as high as 0.2% in patients>60, with an estimated prevalence, including undiscovered asymptomatic patients, of ³1%. The disease has a peak incidence between the third and fifth decades but occurs in young children and in the elderly.

Etiology

Solitary Adenomas The cause of hyperparathyroidism is one or more hyperfunctioning glands. The traditional view has been that a single abnormal gland is the cause in approximately 80% of patients; the abnormality in the gland is usually a benign neoplasm or adenoma and rarely a parathyroid carcinoma. Some surgeons and pathologists report that the enlargement of multiple glands is common; double adenomas are reported. In approximately 15% of patients, all glands are hyperfunctioning; *chief cell parathyroid hyperplasia* is usually hereditary and frequently associated with other endocrine abnormalities.

Multiple Endocrine Neoplasia Hereditary hyperparathyroidism can occur without other endocrine abnormalities but is usually part of a *multiple endocrine neoplasia* syndrome

([Chap. 339](#)).[MEN](#) 1 (Wermer's syndrome) consists of hyperparathyroidism and tumors of the pituitary and pancreas, often associated with gastric hypersecretion and peptic ulcer disease (Zollinger-Ellison syndrome). MEN 2A is characterized by pheochromocytoma and medullary carcinoma of the thyroid, as well as hyperparathyroidism; MEN 2B has additional associated features such as multiple neuromas but usually lacks hyperparathyroidism. Each of these MEN syndromes is transmitted in an autosomal dominant manner.

Pathology Adenomas are most often located in the inferior parathyroid glands, but in 6 to 10% of patients, parathyroid adenomas may be located in the thymus, the thyroid, the pericardium, or behind the esophagus. Adenomas are usually 0.5 to 5 g in size but may be as large as 10 to 20 g (normal glands weigh 25 mg on average). Chief cells are predominant in both hyperplasia and adenoma. The adenoma is sometimes encapsulated by a rim of normal tissue. With chief cell hyperplasia, the enlargement may be so asymmetric that some involved glands appear grossly normal. If generalized hyperplasia is present, however, histologic examination reveals a uniform pattern of chief cells and disappearance of fat even in the absence of an increase in gland weight. Thus, microscopic examination of biopsy specimens of several glands is essential to interpret findings at surgery. When an adenoma is present, the other glands are usually normal and contain a normal distribution of all cell types (rather than only chief cells) and normal amounts of fat.

Parathyroid carcinoma is usually not aggressive in character. Long-term survival without recurrence is common if at initial surgery the entire gland is removed without rupture of the capsule. Recurrent parathyroid carcinoma is usually slow-growing with local spread in the neck, and surgical correction of recurrent disease may be feasible. Occasionally, however, parathyroid carcinoma is more aggressive, with distant metastases (lung, liver, and bone) found at the time of initial operation. It may be difficult to appreciate initially that a primary tumor is carcinoma; increased numbers of mitotic figures and increased fibrosis of the gland stroma may precede invasion. The diagnosis of carcinoma is often made in retrospect. Hyperparathyroidism from a parathyroid carcinoma may be indistinguishable from other forms of primary hyperparathyroidism; a potential clue to the diagnosis, however, is provided by the degree of calcium elevation. Calcium values of 3.5 to 3.7 mmol/L (14 to 15 mg/dL) are frequent with carcinoma and may alert the surgeon to remove the abnormal gland with care to avoid capsular rupture.

GENETIC CONSIDERATIONS

Defects Associated with Hyperparathyroidism As in many other types of neoplasia, two fundamental types of genetic defects have been identified in parathyroid gland tumors: (1) overactivity of protooncogenes, and (2) loss of function of tumor suppressor genes. The former, by definition, can lead to uncontrolled cellular growth and function by activation (gain-of-function mutation) of a single allele of the responsible gene, whereas the latter requires loss of function of both allelic copies.

Mutations in the *MENIN* gene locus on chromosome 11q13 are responsible for causing[MEN](#) 1; the normal allele of this gene fits the definition of a tumor suppressor gene. A mutation of one allele is inherited; loss of the other allele via somatic cell mutation leads to monoclonal expansion and tumor development in tissues such as the

parathyroids. In approximately 20% of sporadic parathyroid adenomas, the *MENIN* locus on chromosome 11 appears to be deleted, implying that the same defect responsible for MEN 1 can also cause the sporadic disease ([Fig. 341-4](#)*A*). Consistent with the Knudson hypothesis for two-step neoplasia in certain inherited cancer syndromes ([Chap. 81](#)), the earlier onset of hyperparathyroidism in the hereditary syndromes reflects the statistical probability of only one mutational event triggering the monoclonal outgrowth. In sporadic adenomas, typically occurring later in life, two different somatic events must occur before the *MENIN* gene is silenced.

The *MENIN* gene codes for a novel protein consisting of 610 amino acids. The protein has a nuclear localization signal and appears to interact with the transcription factor Jun D. Most of the mutations are clearly of the inactivating type (nonsense, deletions); there is not, however, a good correlation between clinical features in different kindreds and the specific mutation detected (e.g., penetrance or age of onset of pituitary or pancreatic tumors). This is in contrast to the correlation between genotype and phenotype in [MEN](#) 2 (see below).

Other presumptive antioncogenes involved in hyperparathyroidism include a gene mapped to chromosome 1p seen in 40% of sporadic parathyroid adenomas and a gene mapped to chromosome Xp11 in patients with secondary hyperparathyroidism and renal failure, who progressed to "tertiary" hyperparathyroidism, now known to reflect monoclonal outgrowths within previously hyperplastic glands.

The *Rb* gene, a tumor suppressor gene located on chromosome 13q14, was initially associated with retinoblastomas but has since been implicated in many other forms of neoplasia including parathyroid carcinoma. Allelic deletion (with a presumed point mutation in the second allele) has been identified in all parathyroid carcinomas examined; there is also an abnormal staining pattern of the protein product of the gene. Allelic deletion is also seen in 10% of parathyroid adenomas, although the abnormal staining pattern of the Rb protein is not seen. Other gene loci on chromosome 13 may be involved in addition to the *Rb* locus.

There are two rare syndromes associated with hyperparathyroidism that involve one or more genes located on chromosome 1q. The hereditary hyperparathyroidism jaw tumor (HPT-JT) syndrome shows an autosomal dominant inheritance pattern; the jaw tumors are benign, but the parathyroid pathology may involve carcinoma as well as adenoma. Parathyroid carcinoma may also appear in the other syndrome, familial isolated primary hyperparathyroidism (FIPH). Both syndromes have been mapped through linkage studies to the chromosome 1q21-q31 region. Certain findings have led to speculation that this chromosome region might contain a protooncogene rather than an antioncogene.

In some parathyroid adenomas, activation of a protooncogene has been identified ([Fig. 341-4](#)*B*). A reciprocal translocation involving chromosome 11 has been identified that juxtaposes the[PTH](#) gene promoter upstream of a gene product termed *PRAD-1*, a cyclin D protein that plays a key role in normal cell division. This translocation is found in as many as 15% of parathyroid adenomas, usually in larger tumors. Targeted overexpression of cyclin D_1 in the parathyroid glands of transgenic mice causes the development of hyperparathyroidism, consistent with the role of this cell cycle control

protein in parathyroid neoplasia.

A mutated protooncogene, *RET*, is involved in each of the clinical variants of[MEN]2 ([Chap. 339](#)). *RET* encodes a tyrosine kinase-type receptor; specific mutations lead to constitutive activity of the receptor, thereby explaining the autosomal dominant mode of transmission and the relatively early onset of neoplasia.

Signs and Symptoms Half or more of patients with hyperparathyroidism are asymptomatic. In series in which patients are followed without operation, as many as 80% are classified as without symptoms. Manifestations of hyperparathyroidism involve primarily the kidneys and the skeletal system. Kidney involvement, due either to deposition of calcium in the renal parenchyma or to recurrent nephrolithiasis, was present in 60 to 70% of patients prior to 1970. With earlier detection, renal complications occur in<20% of patients in many large series. Renal stones are usually composed of either calcium oxalate or calcium phosphate. In occasional patients, repeated episodes of nephrolithiasis or the formation of large calculi may lead to urinary tract obstruction, infection, and loss of renal function. Nephrocalcinosis may also cause decreased renal function and phosphate retention.

The distinctive bone manifestation of hyperparathyroidism is *osteitis fibrosa cystica*, which in series reported 50 years ago occurred in 10 to 25% of patients. In recent years, osteitis fibrosa cystica is very rare in primary hyperparathyroidism, probably due to the increased incidence of mild disease. Histologically, the pathognomonic features are an increase in the giant multinucleated osteoclasts in scalloped areas on the surface of the bone (Howship's lacunae) and a replacement of the normal cellular and marrow elements by fibrous tissue. X-ray changes include resorption of the phalangeal tufts and replacement of the usually sharp cortical outline of the bone in the digits by an irregular outline (subperiosteal resorption).

With the use of multiple markers of bone turnover, such as formation indices (bone-specific alkaline phosphatase, osteocalcin, and type I procollagen peptides) and bone resorption indices (including hydroxypyridinium collagen cross-links and telopeptides of type I collagen), increased skeletal turnover is detected in essentially all patients with established hyperparathyroidism.

Computed tomography (CT) scan and dual-energy x-ray absorptiometry (DEXA) of the spine provide reproducible quantitative estimates (within a few percent) of spinal bone density ([Chap. 342](#)). Similarly, cortical bone density in the extremities can be quantified by single-photon densitometry, usually of the distal radius at a site chosen to be primarily cortical. Studies reveal that cortical bone density is reduced while cancellous bone density, especially in the spine, is relatively preserved. Serial studies in patients who choose to be followed without surgery have indicated that in the majority there is little further change over a number of years, consistent with laboratory data indicating relatively unchanged blood calcium and[PTH]levels. After an initial loss of bone mass in patients with mild asymptomatic hyperparathyroidism, a new equilibrium may be reached, with bone density and biochemical manifestations of the disease remaining relatively unchanged. This clinical course has led to the recommendations (discussed below) that asymptomatic patients may be safely followed with medical supervision. Certain recent findings have raised questions about the impact of asymptomatic

hyperparathyroidism on the skeleton, however. In one careful, long-term study, parathyroidectomy led to improvements in bone density in the spine and hip in 10 to 15% of patients; the improved bone density has been maintained for a number of years of follow-up. Another group compared fracture incidence in a large cohort of hyperparathyroid patients followed for years without surgery versus those seen in an age- and sex-matched control population. The incidence of fractures of the spine, wrist, and ribs was significantly increased in the hyperparathyroid group (although there were no data available on bone density).

In symptomatic patients, dysfunctions of the central nervous system, peripheral nerve and muscle, gastrointestinal tract, and joints also occur. An awareness of the signs and symptoms of hyperparathyroidism may give the initial clue to the diagnosis. It has been reported that severe neuropsychiatric manifestations may be reversed by parathyroidectomy; it remains unclear, in the absence of controlled studies, whether this improvement has a defined cause-and-effect relationship. Generally, the fact that hyperparathyroidism is common in elderly patients, in whom there are often other problems, suggests the possibility that such coexisting problems as hypertension, renal deterioration, and depression may not be parathyroid-related and suggests caution in recommending parathyroid surgery as a cure for these conditions.

Neuromuscular manifestations may include proximal muscle weakness, easy fatigability, and atrophy of muscles and may be so striking as to suggest a primary neuromuscular disorder. The distinguishing feature is the complete regression of neuromuscular disease after surgical correction of the hyperparathyroidism.

Gastrointestinal manifestations are sometimes subtle and include vague abdominal complaints and disorders of the stomach and pancreas. Again, cause and effect are unclear. In MEN 1 patients with hyperparathyroidism, duodenal ulcer may be the result of associated pancreatic tumors that secrete excessive quantities of gastrin (Zollinger-Ellison syndrome). Pancreatitis has been reported in association with hyperparathyroidism, but the incidence and the mechanism are not established.

Chondrocalcinosis and pseudogout are said to be sufficiently frequent in hyperparathyroidism that screening of such patients is warranted. Occasionally, pseudogout is the initial manifestation.

Diagnosis The diagnosis is typically made by detecting an elevated immunoreactive PTH level in a patient with asymptomatic hypercalcemia (see "Differential Diagnosis: Special Tests," below). Serum phosphate is usually low but may be normal, especially if renal failure has developed. Hypophosphatemia is a less specific diagnostic finding than hypercalcemia for two reasons: (1) phosphate levels are influenced by dietary intake, diurnal variations, and other factors; to be useful, samples must be obtained in the morning under fasting conditions; and (2) patients with severe hypercalcemia of any cause may have a low serum phosphate.

Many tests based on renal responses to excess PTH (renal calcium and phosphate clearance; blood phosphate, chloride, magnesium; urinary or nephrogenous cyclic AMP) were used in earlier decades. These tests have low specificity for hyperparathyroidism and are therefore not cost-effective; they have been replaced by PTH immunoassays.

TREATMENT

Medical Treatment Management of hyperparathyroidism involves two separate issues. The critical question is whether the disease should be treated surgically. If severe hypercalcemia [3.7 to 4.5 mmol/L (15 to 18 mg/dL)] is present, surgery is mandatory as soon as the diagnosis can be confirmed by a PTH immunoassay. However, in most patients with hyperparathyroidism, hypercalcemia is mild and does not require urgent surgical or medical treatment.

Several hundred patients have been closely followed without surgery in attempts to define the natural history of the disease and the benefits of surgery versus the risks of medical observation. Large-scale randomized, prospective clinical trials have not been undertaken, however. Rather, the long-term effects of hyperparathyroidism have been assessed in patients who do not have kidney stones, osteitis fibrosa cystica, or other clear-cut symptoms. Progressive loss of bone mass is a worrisome problem in women who face the problem of age-dependent and estrogen-deficient bone loss in the absence of hyperparathyroidism. The principal concern is that such patients, even though asymptomatic, will suffer sufficient bone loss due to PTH excess to make them more vulnerable to developing symptomatic osteoporosis.

The National Institutes of Health held a Consensus Conference on Management of Asymptomatic Hyperparathyroidism in 1991. *Asymptomatic hyperparathyroidism* was defined as documented (presumptive) hyperparathyroidism without signs or symptoms attributable to the disease. The consensus was that patients <50 should undergo surgery, given the long surveillance that would be required. Patients >50 are appropriate for medical monitoring if certain criteria are met and the patients wish to avoid surgery. Guidelines for recommending surgery in patients with asymptomatic hyperparathyroidism include the following:

1. Elevation of serum calcium, >0.25 to 0.40 mmol/L (1 to 1.6 mg/dL) above the upper limit of normal for the test laboratory.

2. History of life-threatening hypercalcemia, such as an episode induced by dehydration and recurring illness.

3. Reduction of age-matched creatinine clearance by >30% without a known cause. Presence of kidney stones detected by abdominal radiograph even if they are asymptomatic.

4. Elevation of 24-h urinary calcium excretion >400 mg.

5. Reduction of bone mass more than 2 standard deviations below normal using one of several noninvasive methods.

Other considerations that favor surgery include concern that consistent follow-up would be unlikely or that coexistent illness would complicate management. More recent data indicated that a subgroup of patients had selective vertebral osteopenia out of proportion to bone loss at other sites and responded to surgery with striking restoration

of bone mass (average >20%), suggesting that such patients might also be recommended for surgery. Asymptomatic patients should be monitored regularly. Surgical correction of hyperparathyroidism can always be undertaken when indicated, since the success rate is high (>90%), mortality is low, and morbidity is minimal. The goals of monitoring are early detection of worsening hypercalcemia, deteriorating bone or renal status, or other complications of hyperparathyroidism.

The consensus panel did not make a recommendation as to estrogen use in patients for whom surgery was not elected because there was insufficient cumulative experience with such therapy to balance theoretical risks (breast and endometrial cancer) versus benefits. New medical therapies may change the approach to the disease in the future. Raloxifene (Evista), the first of the SERMS, has been shown to have many of the bone protective effects of estrogen in osteoporotic subjects yet at the same time lowers the incidence of breast cancer; use of this agent has not yet been reported in a series of hyperparathyroid patients, however. Early experience with calcimimetics, drugs that selectively stimulate the calcium sensor and suppress PTH secretion, indicates that these agents decrease PTH levels for several hours after a single dose.

European investigators have reported serious cardiovascular complications in patients with hyperparathyroidism that reverse, at least in part, after surgery. They also found an increased incidence of malignancy and an absolute increase in age-adjusted mortality due to hyperparathyroidism. These reports clearly describe experiences in a group of patients with more advanced disease, at least based on laboratory tests such as blood PTH levels, than the patients followed in the United States. In fact, a recent long-term epidemiologic study of a large cohort of patients in the United States, with a milder form of the disease (fitting the criteria for medical surveillance listed above), had an age-adjusted mortality no different than that of euparathyroid patients and no increased incidence of malignancy.

Surgical Treatment Parathyroid exploration is challenging and is best undertaken by an experienced surgeon with the help of an experienced pathologist. Certain features help in predicting the pathology (e.g., multiple abnormal glands in familial cases). However, some critical decisions regarding management can be made only during the operation. The examination by frozen section of tissue removed at surgery helps direct the subsequent course of the operation.

As discussed above, there are many unresolved issues to consider in surgery for this disease. At the extreme of conservatism, the surgical approach is based on the view that typically only one gland (the adenoma) is abnormal. If an enlarged gland is found, a normal gland should be sought. In this view, if a biopsy of a normal-sized second gland confirms its histologic (and presumed functional) normality, no further exploration, biopsy, or excision is needed. At the other extreme is the minority viewpoint that all four glands be sought and that most of the total parathyroid tissue mass should be removed. The concern with the former approach is that the recurrence rate of hyperparathyroidism may be high if a second abnormal gland is missed; the latter approach could involve unnecessary surgery and an unacceptable rate of hypoparathyroidism. The majority viewpoint, judged by surgical reviews, is in favor of conservative surgery, i.e., removal of what is usually only one enlarged gland but only after four-gland exploration to eliminate the possibility that more than one gland is abnormal. When normal glands are found in

association with one enlarged gland, excision of the single adenoma usually leads to cure or symptom-free disease, although long-term follow-up studies are limited.

Surgical management has been enhanced recently by the use of preoperative ^{99m}Tc sestamibi scans to predict the location of an abnormal gland and intraoperative sampling of PTH before and at 5- to 15-min intervals after removal of a suspected adenoma to confirm a rapid fall (>50%) in PTH levels. In several centers, a combination of preoperative sestamibi imaging, cervical block anesthesia, minimal surgical incision, and intraoperative PTH measurements has allowed successful outpatient surgical management with a clear-cut cost benefit compared to general anesthesia and more extensive neck surgery. The use of these minimally invasive approaches requires clinical judgment to select patients unlikely to have multiple gland disease (e.g., MEN or secondary hyperparathyroidism).

Multiple gland hyperplasia, as predicted in familial cases, poses more difficult questions of surgical management. Once a diagnosis of hyperplasia is established, all the glands must be identified. Two schemes have been proposed for surgical management. One is that three glands be totally removed and the fourth gland be partially excised; care is taken to leave a good blood supply for the remaining gland. Other surgeons advocate total parathyroidectomy with immediate transplantation of a portion of a removed, minced parathyroid gland into the muscles of the forearm, with the view that surgical excision is easier from the ectopic site in the arm if there is recurrent hyperfunction. When parathyroid carcinoma is encountered, the tissue should be widely excised; care must be taken to avoid rupture of the capsule to prevent local seeding of tumor cells.

In a minority of cases, if no abnormal parathyroid glands are found in the neck, the issue of further exploration must be decided. There are documented cases of five or six parathyroid glands and of unusual locations for adenomas, such as in the mediastinum. A variety of techniques have been developed to aid in the preoperative localization of the abnormal parathyroid tissue. Usually these techniques are reserved for patients with initial unsuccessful neck explorations, since the combined success of the localization techniques is not better than that of an experienced parathyroid surgeon in finding the abnormal tissue at the first operation. Noninvasive or minimally invasive techniques include ultrasound, CT scan of the neck and mediastinum, and differential scanning after technetium-sestamibi administration.

When a second parathyroid exploration is indicated, the minimally invasive techniques such as ultrasound, CT scan, and isotope scanning should probably be combined with venous sampling and/or selective digital arteriography in one of the centers specializing in these techniques. Intraoperative monitoring of PTH levels by rapid PTH immunoassays may be useful in guiding the surgery, especially in patients who are reexplored after an initial unsuccessful operation. At one center, long-term cures have been achieved with selective embolization or injection of large amounts of contrast material into the end-arterial circulation feeding the parathyroid tumor.

A decline in serum calcium occurs within 24 h after successful surgery; usually blood calcium falls to low-normal values for 3 to 5 days until the remaining parathyroid tissue resumes hormone secretion. Severe postoperative hypocalcemia is likely only if osteitis fibrosa cystica is present or if injury to all the normal parathyroid glands occurs during

surgery.

In general, patients who do not have symptomatic bone disease or a large deficit in bone mineral and who have good renal and gastrointestinal function have few problems with postoperative hypocalcemia. The extent of postoperative hypocalcemia varies with the surgical approach. If all glands are biopsied, hypocalcemia may be transiently symptomatic and more prolonged. Hypocalcemia is more likely to be symptomatic after second parathyroid explorations, particularly when normal parathyroid tissue was removed at the initial operation and when the manipulation and/or biopsy of the remaining normal glands is more extensive in the search for the missing adenoma.

Patients with hyperparathyroidism have efficient intestinal calcium absorption due to the increased levels of 1,25(OH)$_2$D stimulated by PTH excess. Once hypocalcemia signifies successful surgery, patients can be put on a high-calcium intake or be given oral calcium supplements. Despite mild hypocalcemia, most patients do not require parenteral therapy. If the serum calcium falls to <2 mmol/L (8 mg/dL), *and if the phosphate level rises simultaneously*, the possibility that surgery has caused hypoparathyroidism must be considered. Coexistent hypomagnesemia should be checked for, as it interferes with PTH secretion and causes functional hypoparathyroidism (see below). Parenteral calcium replacement at a low level should be instituted when hypocalcemia is symptomatic. Such indications include a general sense of anxiety and positive Chvostek and Trousseau signs coupled with serum calcium consistently <2 mmol/L (8 mg/dL). For parenteral therapy, calcium (gluconate or chloride) solutions are prepared at a concentration of 1 mg/mL in 5% dextrose in water. The rate and duration of intravenous therapy are determined by the severity of the symptoms and the response of the serum calcium. An infusion of 0.5 to 2 (mg/kg)/h or 30 to 100 mL/h of a 1-mg/mL solution usually suffices to relieve symptoms. Usually, parenteral therapy is required for only a few days. If symptoms worsen or if parenteral calcium is needed for >2 to 3 days, therapy with a vitamin D analogue and/or oral calcium (2 to 4 g/d) should be started (see below). It is cost-effective to use calcitriol (doses of 0.5 to 1.0 ug/d) because of the rapidity of onset and rapidity of cessation of action, in comparison to other forms of vitamin D (see below). A rise in blood calcium after several months of vitamin D replacement may indicate restoration of parathyroid function to normal. It is also appropriate to monitor serum PTH serially to estimate gland function in such patients.

Magnesium deficiency may also complicate the postoperative course. Magnesium deficiency impairs the secretion of PTH, and so hypomagnesemia should be corrected whenever detected. Magnesium chloride is effective by mouth, but this compound is not widely available. Repletion is usually parenteral. Because the depressant effect of magnesium on central and peripheral nerve functions does not occur at levels<2 mmol/L (normal range 0.8 to 1.2 mmol/L), parenteral replacement can be given rapidly. A cumulative dose as great as 0.5 to 1 mmol/kg of body weight can be administered if severe hypomagnesemia is present; often, however, total doses of 20 to 40 mmol are sufficient. The magnesium is given either as an intravenous infusion over 8 to 12 h or in divided doses intramuscularly (magnesium sulfate, USP).

OTHER PARATHYROID-RELATED CAUSES OF HYPERCALCEMIA

Lithium Therapy Lithium, used in the management of bipolar depression and other psychiatric disorders, causes hypercalcemia in approximately 10% of treated patients. The hypercalcemia is dependent on continued lithium treatment, remitting and recurring when lithium is stopped and restarted. The parathyroid adenomas reported in some hypercalcemic patients with lithium therapy may reflect the presence of an independently occurring parathyroid tumor; a permanent effect of lithium on parathyroid gland growth need not be implicated as most patients have complete reversal of hypercalcemia when lithium is stopped. However, long-standing stimulation of parathyroid cell replication by lithium may predispose to development of adenomas (as is documented in secondary hyperparathyroidism and renal failure).

The presence of hypercalcemia does not correlate with plasma lithium level, but the frequency with which hypercalcemia occurs is sufficiently high to support a causal relationship between lithium and the hypercalcemia, particularly the dependence of the hypercalcemia on the continuation of the lithium. At the levels achieved in blood in treated patients, lithium can be shown in vitro to shift the PTH secretion curve in response to calcium to the right; i.e., higher calcium levels are required to lower PTH secretion, probably acting at the calcium sensor (see below). It is logical to assume that this effect can cause elevated PTH levels and consequent hypercalcemia in otherwise normal individuals. If persistent hypercalcemia is detected during lithium therapy, it may be necessary to try alternative medication for the underlying psychiatric illness. Parathyroid surgery should not be recommended unless hypercalcemia and elevated PTH levels persist after lithium is discontinued.

GENETIC DISORDERS CAUSING HYPERPARATHYROID-LIKE SYNDROMES

Familial Hypocalciuric Hypercalcemia FHH (also called *familial benign hypercalcemia*) is inherited as an autosomal dominant trait. Affected individuals are discovered because of asymptomatic hypercalcemia. This disorder and Jansen's disease (discussed below) are variants of hyperparathyroidism. FHH involves excessive secretion of PTH, whereas Jansen's disease is caused by excessive biologic activity of the PTH receptor in target tissues. Neither disorder, however, involves a primary growth disorder of the parathyroids.

The pathophysiology of FHH is now understood. The primary defect is abnormal sensing of the blood calcium by the parathyroid gland and renal tubule, causing inappropriate secretion of PTH and excessive renal reabsorption of calcium (Fig. 341-5). The calcium sensor is a member of the third family of GPCR (type C or III) and is located on chromosome 3. The receptor responds to the ECF calcium concentration, suppressing PTH secretion through second messenger signaling, thereby providing negative-feedback regulation of PTH secretion. More than 20 different mutations in the calcium-sensing receptor have been identified in patients with FHH (Fig. 341-6). These mutations lower the capacity of the sensor to bind calcium, and the mutant receptors function as though blood calcium levels are low; excessive secretion of PTH occurs from an otherwise normal gland. Approximately two-thirds of patients with FHH have mutations within the protein-coding region of the gene. The remaining one-third of kindreds may have mutations in the gene promoter or in other regions of the genome identified through mapping studies (e.g., chromosome 19).

Even before elucidation of the pathophysiology of FHH, abundant clinical evidence served to separate the disorder from primary hyperparathyroidism. Patients with primary hyperparathyroidism have <99% renal calcium reabsorption, whereas most patients with FHH have >99% reabsorption. The hypercalcemia in FHH is often detectable in affected members of the kindreds in the first decade of life, whereas hypercalcemia rarely occurs in patients with primary hyperparathyroidism or the MEN syndromes who are <10. PTH may be elevated in FHH, but the values are usually normal or lower for the same degree of calcium elevation than in patients with primary hyperparathyroidism. Parathyroid surgery in a few patients with FHH led to permanent hypoparathyroidism, but hypocalciuria persisted nevertheless, establishing that hypocalciuria, therefore, is not PTH-dependent (now known to be due to the abnormal calcium sensor in the kidney).

Few clinical signs or symptoms are present in patients with FHH, and other endocrine abnormalities are not present. Most patients are detected as a result of family screening after hypercalcemia is detected in a proband. In those patients inadvertently operated upon, the parathyroids appeared normal or moderately hyperplastic. Parathyroid surgery is not appropriate, nor, in view of the lack of symptoms, does medical treatment seem needed to lower the calcium. Calcimimetic agents that bind to the calcium sensor and elevate the set point are under investigation.

One striking exception to the rule against parathyroid surgery in this syndrome is the occurrence, usually in consanguinous marriages (due to the rarity of gene), of a homozygous or compound heterozygote state, resulting in complete loss of the calcium sensor function. In this condition, neonatal severe hypercalcemia, total parathyroidectomy is mandatory.

Jansen's Disease Mutations in the PTH/PTHrP receptor have been identified as responsible for this rare autosomal dominant syndrome (Fig. 341-7). Because the mutations lead to constitutive receptor function, one abnormal copy of the mutant receptor is sufficient to cause the disease, thereby accounting for its dominant mode of transmission. The disorder leads to short-limbed dwarfism due to abnormal regulation of the bone growth plate. In adult life, there are numerous abnormalities in bone, including multiple cystic resorptive areas resembling those seen in severe hyperparathyroidism. Hypercalcemia and hypophosphatemia with undetectable or low PTH levels are typically seen. The pathogenesis of the disease has been confirmed by transgenic experiments in which targeted expression of the mutant receptor to the growth plate emulated several features of the disorder.

MALIGNANCY-RELATED HYPERCALCEMIA

Clinical Syndromes and Mechanisms of Hypercalcemia Hypercalcemia due to malignancy is common (occurring with 10 to 15% of certain types of tumor, such as lung carcinoma), often severe and difficult to manage, confusing as to etiology, and sometimes difficult to distinguish from primary hyperparathyroidism. Although malignancy is often clinically obvious, hypercalcemia can occasionally be due to an occult tumor. Previously, hypercalcemia associated with malignancy was thought to be due to local invasion and destruction of bone by tumor cells; many cases are now known to result from the elaboration by the malignant cells of humoral mediators of hypercalcemia. PTHrP is the responsible humoral agent in most cases.

The histologic character of the tumor is more important than the extent of skeletal metastases in predicting hypercalcemia. Small cell carcinoma (oat cell) and adenocarcinoma of the lung, although the most common lung tumors associated with skeletal metastases, rarely cause hypercalcemia. By contrast, as many as 10% of patients with squamous cell carcinoma of the lung develop hypercalcemia. Histologic studies of bone in patients with squamous cell or epidermoid carcinoma of the lung, in sites invaded by tumor as well as areas remote from tumor invasion, reveal bone remodeling, including osteoclastic and osteoblastic activity. In contrast, minimal skeletal metabolic activation occurs even if there are extensive skeletal metastases of small cell (oat cell) carcinoma.

Two main mechanisms of hypercalcemia are operative in cancer hypercalcemia. Many solid tumors associated with hypercalcemia, particularly squamous cell and renal tumors, produce and secrete humoral factors that cause increased bone resorption and mediate the hypercalcemia through systemic actions on the skeleton. Alternatively, direct bone marrow invasion occurs with hematologic malignancies such as leukemia, lymphoma, and multiple myeloma. Lymphokines and cytokines produced by cells involved in the marrow response to the tumors promote resorption of bone through local destruction. Several hormones, hormone analogues, cytokines, and growth factors have been implicated as the result of clinical assays, in vitro tests, or chemical isolation. In some lymphomas, typically B cell lymphomas, there is an increased blood level of $1,25(OH)_2D$, which is probably produced by lymphocytes. The etiologic factor produced by activated normal lymphocytes and by myeloma and lymphoma cells, termed *osteoclast activation factor*, now appears to represent the biologic action of several different cytokines, probably interleukin 1 and lymphotoxin or tumor necrosis factor.

The more common mechanism, humoral hypercalcemia of malignancy, occurs with cancers of the lung and kidney, in particular, in which bone metastases are absent, minimal, or not detectable clinically. The clinical picture resembles primary hyperparathyroidism (hypophosphatemia accompanies hypercalcemia), and elimination or regression of the primary tumor leads to disappearance of the hypercalcemia. The disorder is due to secretion by the tumors of the PTH-like factor, PTHrP, that activates the PTH/PTHrP receptor (see above).

As in hyperparathyroidism, patients with the humoral hypercalcemia of malignancy have elevated urinary nephrogenous cyclic AMP excretion, hypophosphatemia, and increased urinary phosphate clearance. However, in humoral hypercalcemia of malignancy, immunoreactive PTH is undetectable or suppressed, making the differential diagnosis easier. Other features of the disorder differ from those of true hyperparathyroidism. Patients may have high, rather than low, renal calcium clearance (relative to serum calcium when compared to true hyperparathyroidism, unlike the expected elevation) and low to normal levels of $1,25(OH)_2D$. The reason that the humoral syndrome differs from hyperparathyroidism in these parameters is unclear since the biologic actions of PTH and PTHrP are presumably exerted through the same receptor. Other cytokines elaborated by the malignancy may be responsible for these variations from hyperparathyroidism. In some patients with the humoral hypercalcemia of malignancy, osteoclastic resorption is unaccompanied by an osteoblastic or bone-forming response, implying inhibition of the normal coupling of bone formation and

resorption. Thus the interaction of more than one substance may determine whether hypercalcemia develops in a particular patient.

Several different assays (single- or double-antibody, different epitopes) have been developed to detect PTHrP. Most data indicate that circulating PTHrP levels are undetectable (or low) in normal individuals, elevated in most cancer patients with the humoral syndrome, and high in human milk. Despite the discovery of PTHrP, identifying the etiologic mechanisms in cancer hypercalcemia is often complex. For example, in breast carcinoma (metastatic to bone) and in a distinctive type of T cell lymphoma/leukemia initiated by human T cell lymphotropic virus I, hypercalcemia is caused by direct local lysis of bone as well as by a humoral mechanism involving excess production of PTHrP.

Diagnostic Issues Levels of PTH measured by the double-antibody technique are undetectable or extremely low in tumor hypercalcemia, as would be expected with the mediation of the hypercalcemia by a factor other than PTH (the hypercalcemia suppresses the normal parathyroid glands). In a patient with minimal symptoms referred for hypercalcemia, low or undetectable PTH and elevated PTHrP levels would focus attention on occult malignancy.

Ordinarily, the diagnosis of cancer hypercalcemia is not difficult because tumor symptoms are prominent when hypercalcemia is detected. Indeed, hypercalcemia may be noted incidentally during the workup of a patient with known or suspected malignancy. Clinical suspicion that malignancy is the cause of the hypercalcemia is heightened when there are other paraneoplastic signs or symptoms, such as weight loss, fatigue, muscle weakness, or unexplained skin rash, or when symptoms specific for a particular tumor are present. Squamous cell tumors are most frequently associated with hypercalcemia, particularly tumors of the lung, kidney, head and neck, and urogenital tract. Radiologic examinations can focus on these areas when clinical evidence is unclear. Bone scans with technetium-labeled bisphosphonate are useful for detection of osteolytic metastases; the sensitivity is high, but specificity is low; results must be confirmed by conventional x-rays to be certain that areas of increased uptake are due to osteolytic metastases per se. Bone marrow biopsies are helpful in patients with anemia or abnormal peripheral blood smears.

TREATMENT

Treatment of the hypercalcemia of malignancy is first directed to control of the tumor; reduction of tumor mass usually corrects hypercalcemia. If a patient has severe hypercalcemia yet has a good chance for effective tumor therapy, treatment of the hypercalcemia should be vigorous while awaiting the results of definitive therapy. If hypercalcemia occurs in the late stages of a tumor that is resistant to therapy, the treatment of the hypercalcemia should be judicious as high calcium levels can have a mild sedating effect. Standard therapies for hypercalcemia (discussed below) are applicable to patients with malignancy.

VITAMIN D-RELATED HYPERCALCEMIA

Hypercalcemia caused by vitamin D can be due to excessive ingestion or abnormal

metabolism of the vitamin. Abnormal metabolism of the vitamin is usually acquired in association with a widespread granulomatous disorder. Vitamin D metabolism is carefully regulated, particularly the activity of renal 1a-hydroxylase, the enzyme responsible for the production of 1,25(OH)$_2$D (Chap. 340). The regulation of 1a-hydroxylase and the normal feedback suppression by 1,25(OH)$_2$D seem to work less well in infants than in adults and to operate poorly, if at all, in sites other than the renal tubule; these phenomena explain the occurrence of hypercalcemia secondary to excessive 1,25(OH)$_2$D$_3$production in infants with Williams' syndrome (see below) and in adults with sarcoidosis or lymphoma.

Vitamin D Intoxication Chronic ingestion of 50 to 100 times the normal physiologic requirement of vitamin D (amounts >50,000 to 100,000 U/d) is usually required to produce significant hypercalcemia in normal individuals. An upper limit of dietary intake of 2000 U/d (50 ug/d) in adults is now recommended because of concerns about potential toxic effects of cumulative supraphysiologic doses. Vitamin D excess increases intestinal calcium absorption and, if severe, also increases bone resorption.

Hypercalcemia in vitamin D intoxication is due to an excessive biologic action of the vitamin, perhaps the consequence of increased levels of 25(OH)D rather than increased levels of the usual active metabolite 1,25(OH)$_2$D (the latter is frequently not elevated in vitamin D intoxication). 25(OH)D has definite, if low, biologic activity in intestine and bone. The production of 25(OH)D is less tightly regulated than is the production of 1,25(OH)$_2$D. Hence concentrations of 25(OH)D are elevated several-fold in patients with excess vitamin D intake.

The diagnosis is substantiated by documenting elevated levels of 25(OH)D>100 ng/mL. Hypercalcemia is usually controlled by restriction of dietary calcium intake and appropriate attention to hydration. These measures, plus discontinuation of vitamin D, usually lead to resolution of hypercalcemia. However, vitamin D stores in fat may be substantial, and vitamin D intoxication may persist for weeks after vitamin D ingestion is terminated. Such patients are responsive to glucocorticoids, which in doses of 100 mg/d of hydrocortisone or its equivalent, usually return serum calcium levels to normal over several days; severe intoxication may require intensive therapy.

Sarcoidosis and Other Granulomatous Diseases In patients with sarcoidosis and other granulomatous diseases, such as tuberculosis and fungal infections, excess 1,25(OH)$_2$D is synthesized in macrophages or other cells in the granulomas. Indeed, increased 1,25(OH)$_2$D levels have been reported in anephric patients with sarcoidosis and hypercalcemia. Macrophages obtained from granulomatous tissue convert 25(OH)D to 1,25(OH)$_2$D at an increased rate. There is a positive correlation in patients with sarcoidosis between 25(OH)D levels (reflecting vitamin D intake) and the circulating concentrations of 1,25(OH)$_2$D, whereas normally there is no increase in 1,25(OH)$_2$D with increasing 25(OH)D levels due to multiple feedback controls on renal 1a-hydroxylase (Chap. 340). The usual regulation of active metabolite production by calcium orPTH does not operate in these patients; hypercalcemia does not lead to a reduction in the blood levels of 1,25(OH)$_2$D in patients with sarcoidosis. Clearance of 1,25(OH)$_2$D from blood may be decreased in sarcoidosis as well. PTH levels are usually low and 1,25(OH)$_2$D levels are elevated, but primary hyperparathyroidism and sarcoidosis may coexist in some patients.

Management of the hypercalcemia can often be accomplished by avoiding excessive sunlight exposure and limiting vitamin D and calcium intake. Presumably, however, the abnormal sensitivity to vitamin D and abnormal regulation of 1,25(OH)$_2$D synthesis will persist as long as the disease is active. Alternatively, glucocorticoids in the equivalent of £100 mg/d of hydrocortisone control hypercalcemia. Glucocorticoids appear to act by blocking excessive production of 1,25(OH)$_2$D as well as the response to it in target organs.

Idiopathic Hypercalcemia of Infancy This rare disorder, usually referred to as *Williams' syndrome*, is an autosomal dominant disorder characterized by multiple congenital development defects, including supravalvular aortic stenosis, mental retardation, and an elfin facies, in association with hypercalcemia due to abnormal sensitivity to vitamin D. The syndrome was first recognized in England after the fortification of milk with vitamin D. Levels of 1,25(OH)$_2$D are elevated, ranging from 46 to 120 nmol/L (150 to 500 pg/mL). The mechanism of the abnormal sensitivity to vitamin D and of the increased circulating levels of 1,25(OH)$_2$D is still unclear. Studies suggest that mutations involving the elastin locus and perhaps other genes on chromosome 7 may play a role in the pathogenesis.

HYPERCALCEMIA ASSOCIATED WITH HIGH BONE TURNOVER

Hyperthyroidism As many as 20% of hyperthyroid patients have high-normal or mildly elevated serum calcium concentrations; hypercalciuria is even more common. The hypercalcemia is due to increased bone turnover, with bone resorption exceeding bone formation. Severe calcium elevations are not typical, and the presence of such suggests a concomitant disease such as hyperparathyroidism. Usually, the diagnosis is obvious, but signs of hyperthyroidism may occasionally be occult, particularly in the elderly (Chap. 330). Hypercalcemia is managed by treatment of the hyperthyroidism.

Immobilization Immobilization is a rare cause of hypercalcemia in adults in the absence of an associated disease but may cause hypercalcemia in children and adolescents, particularly after spinal cord injury and paraplegia or quadriplegia. With resumption of ambulation, the hypercalcemia in children usually returns to normal.

The mechanism appears to involve a disproportion between bone formation and bone resorption. Hypercalciuria and increased mobilization of skeletal calcium can develop in normal volunteers subjected to extensive bed rest, although hypercalcemia is unusual. Immobilization of an adult with a disease associated with high bone turnover, such as Paget's disease, may cause hypercalcemia.

Thiazides Administration of benzothiadiazines (thiazides) can cause hypercalcemia in patients with high rates of bone turnover, such as patients with hypoparathyroidism treated with high doses of vitamin D. Traditionally, thiazides are associated with aggravation of hypercalcemia in primary hyperparathyroidism, but this effect can be seen in other high-bone-turnover states as well. The mechanism of thiazide action is complex. Chronic thiazide administration leads to reduction in urinary calcium; the hypocalciuric effect appears to reflect the enhancement of proximal tubular resorption of sodium and calcium in response to sodium depletion. Some of this renal effect is due to

augmentation of PTH action and is more pronounced in individuals with intact PTH secretion. However, thiazides cause hypocalciuria in hypoparathyroid patients on high-dose vitamin D and oral calcium replacement if sodium intake is restricted. This finding is the rationale for the use of thiazides as an adjunct to therapy in hypoparathyroid patients, as discussed below. Thiazide administration to normal individuals causes a transient increase in blood calcium (usually within the high-normal range) that reverts to preexisting levels after a week or more of continued administration. If hormonal function and calcium and bone metabolism are normal, homeostatic controls are reset to counteract the calcium-elevating effect of the thiazides. In the presence of hyperparathyroidism or increased bone turnover from another cause, homeostatic mechanisms are ineffective. The abnormal effects of the thiazide on calcium metabolism disappear within days of cessation of the drug.

Vitamin A Intoxication Vitamin A intoxication is a rare cause of hypercalcemia and is most commonly a side effect of dietary faddism (Chap. 75). Calcium levels can be elevated into the 3 to 3.5 mmol/L (12 to 14 mg/dL) range after the ingestion of 50,000 to 100,000 units of vitamin A daily (10 to 20 times the minimum daily requirement). Typical features of severe hypercalcemia include fatigue, anorexia, and, in some, severe muscle and bone pain. Excess Vitamin A intake is presumed to increase bone resorption.

The diagnosis can be established by history and by measurement of vitamin A levels in serum, which may be severalfold above normal. Occasionally, skeletal x-rays reveal periosteal calcifications, particularly in the hands. Withdrawal of the vitamin is usually associated with prompt disappearance of the hypercalcemia and reversal of the skeletal changes. As in vitamin D intoxication, administration of 100 mg/d hydrocortisone or its equivalent leads to a rapid return of the serum calcium to normal.

HYPERCALCEMIA ASSOCIATED WITH RENAL FAILURE

Severe Secondary Hyperparathyroidism Secondary hyperparathyroidism occurs when partial resistance to the metabolic actions of PTH leads to excessive production of the hormone. Parathyroid gland hyperplasia occurs because resistance to the normal level of PTH leads to hypocalcemia, which, in turn, is a stimulus to parathyroid gland enlargement. This concept is supported by studies of the treatment of patients treated with bisphosphonates, which block the skeletal resorptive response. Because a portion of PTH secretion by each parathyroid cell is not suppressible by any degree of elevation of blood calcium, larger glands (more cells) have a higher hormone output at the hypercalcemic end of the dose-response curve.

Secondary hyperparathyroidism occurs not only in patients with renal failure but also in those with osteomalacia due to multiple causes (Chap. 340), including deficiency of vitamin D action, and PHP (deficient response to PTH at the level of the receptor). Hypocalcemia seems to be the common denominator in initiating secondary hyperparathyroidism. Only in patients with renal failure, however, is hypercalcemia sometimes encountered despite appropriate medical management regimens (see below). Primary and secondary hyperparathyroidism can be distinguished conceptually by the autonomous growth of the parathyroid glands in primary hyperparathyroidism (presumably irreversible) and the adaptive response of the parathyroids in secondary

hyperparathyroidism (typically reversible). In fact, reversal over weeks from an abnormal pattern of secretion, presumably accompanied by involution of parathyroid gland mass to normal, occurs in patients who have been treated effectively to reverse the resistance to PTH (such as with calcium and vitamin D in osteomalacia).

Patients with secondary hyperparathyroidism may develop bone pain, ectopic calcification, and pruritus. The bone disease seen in patients with secondary hyperparathyroidism and renal failure is termed *renal osteodystrophy*. Osteomalacia (predominantly due to vitamin D and calcium deficiency) and/or osteitis fibrosa cystica (excessive PTH action on bone) may occur.

Two other skeletal disorders are associated with long-term dialysis in patients with renal failure. Aluminum deposition (see below) is associated with an osteomalacia-like picture. The other entity is a low-bone-turnover state termed "aplastic" or "adynamic" bone disease; PTH levels are lower than in typical secondary hyperparathyroidism. It is believed that the condition is caused, at least in part, by excessive PTH suppression, which may be even greater than previously appreciated in light of evidence that some of the immunoreactive PTH detected by most commercially available PTH assays is not the full-length biologically active molecule (as discussed above).

TREATMENT

Medical therapy to reverse secondary hyperparathyroidism includes reduction of excessive blood phosphate by restriction of dietary phosphate, the use of nonabsorbable antacids, and careful, selective addition of calcitriol (0.25 to 2.0 ug/d); calcium carbonate is preferred over aluminum-containing antacids to prevent aluminum toxicity. Intravenous calcitriol, administered as several pulses each week, helps control secondary hyperparathyroidism. Aggressive but carefully administered medical therapy can often, but not always, reverse hyperparathyroidism and its symptoms and manifestations.

Occasional patients develop severe manifestations of secondary hyperparathyroidism, including hypercalcemia, pruritus, extraskeletal calcifications, and painful bones, despite aggressive medical efforts to suppress the hyperparathyroidism. PTH hypersecretion no longer responsive to medical therapy, a state of severe hyperparathyroidism in patients with renal failure that requires surgery, has been referred to as *tertiary hyperparathyroidism*. Parathyroid surgery is necessary to control this condition. Based on genetic evidence from examination of tumor samples in these patients, the emergence of autonomous parathyroid function is due to a monoclonal outgrowth of one or more previously hyperplastic parathyroid glands.

Aluminum Intoxication Aluminum intoxication (and often hypercalcemia as a complication of medical treatment) may occur in patients on chronic dialysis; manifestations include acute dementia and unresponsive and severe osteomalacia. Bone pain, multiple nonhealing fractures, particularly of the ribs and pelvis, and a proximal myopathy may occur. Hypercalcemia develops when these patients are treated with vitamin D or calcitriol because of impaired skeletal responsiveness. Aluminum is present at the site of osteoid mineralization, osteoblastic activity is minimal, and calcium incorporation into the skeleton is impaired. Prevention is accomplished by avoidance of

aluminum excess in the dialysis regimen; treatment of established disease involves mobilizing aluminum through the use of the chelating agent deferoxamine (Chap. 348).

Milk-Alkali Syndrome The milk-alkali syndrome is due to excessive ingestion of calcium and absorbable antacids such as milk or calcium carbonate. It is much less frequent since nonabsorbable antacids and other treatments became available for peptic ulcer disease. However, the increased use of calcium carbonate in the management of osteoporosis has led to reappearance of the syndrome. Several clinical presentations -- acute, subacute, and chronic -- have been described, all of which feature hypercalcemia, alkalosis, and renal failure. The chronic form of the disease, termed *Burnett's syndrome*, is associated with irreversible renal damage. The acute syndromes reverse if the excess calcium and absorbable alkali are stopped.

Individual susceptibility is important in the pathogenesis, as many patients are treated with calcium carbonate alkali regimens without developing the syndrome. One variable is the fractional calcium absorption as a function of calcium intake. Some individuals absorb a high fraction of calcium, even with intakes as high as 2 g or more of elemental calcium per day, instead of reducing calcium absorption with high intake, as occurs in most normal individuals. Resultant mild hypercalcemia after meals in such patients is postulated to contribute to the generation of alkalosis. Development of hypercalcemia causes increased sodium excretion and some depletion of total-body water. These phenomena and perhaps some suppression of endogenousPTHsecretion due to mild hypercalcemia lead to increased bicarbonate resorption and to alkalosis in the face of continued calcium carbonate ingestion. Alkalosis per se selectively enhances calcium resorption in the distal nephron, thus aggravating the hypercalcemia. The cycle of mild hypercalcemia® bicarbonate retention® alkalosis ® renal calcium retention ® severe hypercalcemia perpetuates and aggravates hypercalcemia and alkalosis as long as calcium and absorbable alkali are ingested.

DIFFERENTIAL DIAGNOSIS: SPECIAL TESTS

Differential diagnosis of hypercalcemia is best achieved by using clinical criteria, but the immunoassay forPTH is especially useful in distinguishing among major causes (Fig. 341-8). The clinical features that deserve emphasis are the presence or absence of symptoms or signs of disease and evidence of chronicity. If one discounts fatigue or depression,>90% of patients with primary hyperparathyroidism have *asymptomatic hypercalcemia*; symptoms of malignancy are usually present in cancer-associated hypercalcemia. Disorders other than hyperparathyroidism and malignancy cause <10% of cases of hypercalcemia, and some of the nonparathyroid causes are associated with clear-cut manifestations such as renal failure.

Hyperparathyroidism is the likely diagnosis in patients with *chronic hypercalcemia*. If hypercalcemia has been manifest for >1 year, malignancy can usually be excluded as the cause. A striking feature of malignancy-associated hypercalcemia is the rapidity of the course, whereby signs and symptoms of the underlying malignancy are evident within months of the detection of hypercalcemia. A careful history of dietary supplements and drug use may suggest intoxication with vitamins D or vitamin A or the use of thiazides.

Although clinical considerations are helpful in arriving at the correct diagnosis of the cause of hypercalcemia, appropriate laboratory testing is essential for definitive diagnosis. The immunoassay for PTH should separate hyperparathyroidism from all other causes of hypercalcemia. Patients with hyperparathyroidism have elevated PTH levels despite hypercalcemia, whereas patients with malignancy and the other causes of hypercalcemia (except for disorders mediated by PTH such as lithium-induced hypercalcemia) have levels of hormone below normal or undetectable. Assays based on the double-antibody method for PTH exhibit very high sensitivity (especially if serum calcium is simultaneously evaluated) and specificity for the diagnosis of primary hyperparathyroidism (Fig. 341-9).

In summary, PTH values are elevated in >90% of parathyroid-related causes of hypercalcemia, undetectable or low in malignancy-related hypercalcemia, and undetectable or normal in vitamin D-related and high-bone-turnover causes of hypercalcemia. In view of the specificity of the PTH immunoassay and the high frequency of hyperparathyroidism in hypercalcemic patients, it is cost-effective to measure the PTH level in all hypercalcemic patients unless malignancy or a specific nonparathyroid disease is obvious. False-positive PTH assay results are rare. There are very rare reports of ectopic production of excess PTH by nonparathyroid tumors. Immunoassays for PTHrP are helpful in diagnosing certain types of malignancy-associated hypercalcemia. Although FHH is parathyroid-related, the disease should be managed distinctively from hyperparathyroidism. Clinical features and the low urinary calcium excretion can help make the distinction. Because the incidence of malignancy and hyperparathyroidism both increase with age, they can coexist as two independent causes of hypercalcemia.

$1,25(OH)_2D$ levels are elevated in many (but not all) patients with primary hyperparathyroidism. In other disorders associated with hypercalcemia, concentrations of $1,25(OH)_2D$ are low or, at the most, normal. However, this test is of low specificity and is not cost-effective, as not all patients with hyperparathyroidism have elevated $1,25(OH)_2D$ levels, and not all nonparathyroid hypercalcemic patients have suppressed $1,25(OH)_2D$. Measurement of $1,25(OH)_2D$ is, however, critically valuable in establishing the cause of hypercalcemia in sarcoidosis and certain B cell lymphomas.

A useful general approach is outlined in Fig. 341-8. If the patient is *asymptomatic* and there is evidence of *chronicity* to the hypercalcemia, hyperparathyroidism is almost certainly the cause. If PTH levels (usually measured at least twice) are elevated, the clinical impression is confirmed and little additional evaluation is necessary. If there is only a short history or no data as to the duration of the hypercalcemia, *occult malignancy* must be considered; if the PTH levels are not elevated, then a thorough workup must be undertaken for malignancy, including chest x-ray, CT of chest and abdomen, and bone scan. Immunoassays for PTHrP may be especially useful in such situations. Attention should also be paid to clues for underlying hematologic disorders such as anemia, increased plasma globulin, and abnormal serum immunoelectrophoresis; bone scans can be negative in some patients with metastases, such as in multiple myeloma. Finally, if a patient with chronic hypercalcemia is asymptomatic and malignancy therefore seems unlikely on clinical grounds, but PTH values are not elevated, it is useful to search for other chronic causes of hypercalcemia, such as occult sarcoidosis.

TREATMENT

Hypercalcemic States The approach to medical treatment of hypercalcemia varies with its severity. Mild hypercalcemia, <3.0 mmol/L (12 mg/dL), can be managed by hydration. More severe hypercalcemia [levels of 3.2 to 3.7 mmol/L (13 to 15 mg/dL)] must be managed aggressively; above that level, hypercalcemia can be life-threatening and requires emergency measures. By using a combination of approaches, the serum calcium concentration can be decreased by 0.7 to 2.2 mmol/L (3 to 9 mg/dL) within 24 to 48 h in most patients, enough to relieve acute symptoms, prevent death from hypercalcemic crisis, and permit diagnostic evaluation. Therapy can then be directed at the underlying disorder -- the second priority.

Hypercalcemia develops because of excessive skeletal calcium release, increased intestinal calcium absorption, or inadequate renal calcium excretion. Understanding the particular pathogenesis helps guide therapy. For example, hypercalcemia in patients with malignancy is primarily due to excessive skeletal calcium release and is, therefore, minimally improved by restriction of dietary calcium. On the other hand, patients with vitamin D hypersensitivity or vitamin D intoxication have excessive intestinal calcium absorption, and restriction of dietary calcium is beneficial. Decreased renal function orECFdepletion decreases urinary calcium excretion. In such situations, rehydration may rapidly reduce or reverse the hypercalcemia, even though increased bone resorption persists. As outlined below, the more severe the hypercalcemia, the greater the number of combined therapies that should be used. Rapid acting (hours) approaches -- rehydration, forced diuresis, and calcitonin -- can be used with the most effective antiresorptive agents, such as bisphosphonates (since severe hypercalcemia usually involves excessive bone resorption).

Hydration, Increased Salt Intake, Mild and Forced Diuresis The first principle of treatment is to restore normal hydration. Many hypercalcemic patients are dehydrated because of vomiting, inanition, and/or hypercalcemia-induced defects in urinary concentrating ability. The resultant drop in glomerular filtration rate is accompanied by an additional decrease in renal tubular sodium and calcium clearance. Restoring a normalECFvolume corrects these abnormalities and increases urine calcium excretion by 2.5 to 7.5 mmol/d (100 to 300 mg/d). Increasing urinary sodium excretion to 400 to 500 mmol/d increases urinary calcium excretion even further than simple rehydration. After rehydration has been achieved, saline can be administered or furosemide or ethacrynic acid can be given twice daily to depress the tubular reabsorptive mechanism for calcium (care must be taken to prevent dehydration). The combined use of these therapies can increase urinary calcium excretion to³12.5 mmol/d (500 mg/d) in most hypercalcemic patients. Since this is a substantial percentage of the exchangeable calcium pool, the serum calcium concentration usually falls 0.25 to 0.75 mmol/L (1 to 3 mg/dL) within 24 h. Precautions should be taken to prevent potassium and magnesium depletion; calcium-containing renal calculi are a potential complication.

Under life-threatening circumstances, the preceding approach can be pursued more aggressively, giving as much as 6 L isotonic saline (900 mmol sodium) daily plus furosemide or equivalent in doses up to 100 mg every 1 to 2 h or ethacrynic acid in doses to 40 mg every 1 to 2 h. Urinary calcium excretion may exceed 25 mmol/d (1000

mg/d), and the serum calcium may decrease by ³1 mmol/L (4 mg/dL) within 24 h. Depletion of potassium and magnesium is inevitable unless replacements are given; pulmonary edema can be precipitated. The potential complications can be reduced by careful monitoring of central venous pressure and plasma or urine electrolytes; catheterization of the bladder may be necessary. This treatment approach should be supplemented with agents to block bone resorption. Though these agents do not become effective for several days, forced diuresis is difficult to sustain even in patients with good cardiopulmonary and renal function.

Bisphosphonates The bisphosphonates are analogues of pyrophosphate, with high affinity for bone, especially in areas of increased bone turnover, where they are powerful inhibitors of bone resorption. These bone-seeking compounds are stable in vivo because phosphatase enzymes cannot hydrolyze the central carbon-phosphorus-carbon bond. The bisphosphonates are concentrated in areas of high bone turnover and are taken up by and inhibit osteoclast action; the mechanism of action is complex. Bisphosphonates alter osteoclast proton pump function or impair the release of acid hydrolases into the extracellular lysosomes contiguous with mineralized bone. They may also inhibit the differentiation of monocyte-macrophage precursors into osteoclasts and possibly have effects on osteoblasts as well. The bisphosphonate molecules that contain amino groups in the side chain structure (see below) interfere with prenylation of proteins and can lead to cellular apoptosis. The highly active non-amino group-containing bisphosphonates are also metabolized to cytotoxic products.

The initial bisphosphonate widely used in clinical practice, etidronate, was effective but had several disadvantages, including the capacity to inhibit bone formation as well as blocking resorption. Subsequently, a number of second-generation compounds have become the mainstays of antiresorptive therapy for treatment of hypercalcemia. The most widely used, pamidronate, is a potent inhibitor of osteoclast-mediated skeletal resorption yet does not cause mineralization defects at ordinary doses. Several additional bisphosphonates (alendronate, tiludronate, and risedronate) are potent and also have a highly favorable ratio of blocking resorption versus inhibiting bone formation. Though the bisphosphonates have similar structures, the routes of administration, efficacy, toxicity, and side effects vary. The potency of the compounds for inhibition of bone resorption varies a thousandfold in the order of etidronate, tiludronate, pamidronate, alendronate, and risedronate. Oral alendronate is approved for the therapy of osteoporosis in the United States, and in Europe oral preparations of these bisphosphonates are used in the chronic treatment of hypercalcemia. Only the intravenous use of pamidronate is approved for this purpose in the United States; between 30 and 90 mg pamidronate, given as a single intravenous dose over a few hours, returns serum calcium to normal within 24 to 48 h with an effect that lasts for weeks in 80 to 100% of patients.

Pamidronate causes low-grade fever in as many as 20% of patients, likely related to release of cytokines from osteoclasts, monocytes, and macrophages (Table 341-2). This effect is usually seen only with the initial doses. Etidronate causes hyperphosphatemia through a direct renal mechanism, whereas hypophosphatemia is seen after therapy with other bisphosphonates. Overall, second-generation bisphosphonates are now the agents of choice in severe hypercalcemia, particularly that

associated with malignancy. Zoledronate, a third-generation bisphosphonate, is claimed to be 100 to 800 times more potent than pamidronate and to normalize calcium more quickly and for longer periods.

Calcitonin Calcitonin acts within a few hours of its administration, through receptors on osteoclasts, to block bone resorption and, in addition, to increase urinary calcium excretion by inhibition of renal tubular calcium reabsorption. However, calcitonin leads to variable and usually minimal lowering of calcium. Tachyphylaxis, a known phenomenon with this drug, may explain the variable results. However, in life-threatening hypercalcemia, calcitonin can be used effectively within the first 24 h in combination with rehydration and saline diuresis while waiting for more sustained effects from a simultaneously administered bisphosphonate such as pamidronate. Usual doses of calcitonin are 2 to 8 U/kg of body weight intravenously, subcutaneously, or intramuscularly every 6 to 12 h.

Other Therapies *Plicamycin* (mithramycin), which inhibits bone resorption, has been a useful therapeutic agent but is now little used because of the effectiveness of bisphosphonates. Plicamycin must be given intravenously, either as a bolus injection or by slow infusion. The usual dose is 25 ug/kg body weight. Major side effects include thrombocytopenia, hepatocellular necrosis with increased lactic acid dehydrogenase (LDH) and aspartate aminotransferase (AST) levels, and decreased levels of clotting factors with resultant epistaxis, bruising, hemorrhage, and bleeding gums.

Gallium nitrate exerts a hypocalcemic action by inhibiting bone resorption and altering the structure of bone crystals. Major disadvantages include the 5-day duration of infusion and the potential for nephrotoxicity and relatively shorter duration of action than bisphosphonates. Accordingly it is not often used because of superior alternatives.

Glucocorticoids increase urinary calcium excretion and decrease intestinal calcium absorption when given in pharmacologic doses, but they also cause negative skeletal calcium balance. In normal individuals and in patients with primary hyperparathyroidism, glucocorticoids neither increase nor decrease the serum calcium concentration. In patients with hypercalcemia due to certain osteolytic malignancies, however, glucocorticoids may be effective as a result of antitumor effects. The malignancies in which hypercalcemia responds to glucocorticoids include multiple myeloma, leukemia, Hodgkin's disease, other lymphomas, and carcinoma of the breast, at least early in the course of the disease. Glucocorticoids are also effective in treating hypercalcemia due to vitamin D intoxication and sarcoidosis. In all the preceding situations, the hypocalcemic effect develops over several days, and the usual glucocorticoid dosage is 40 to 100 mg prednisone (or its equivalent) daily in four divided doses. The side effects of chronic glucocorticoid therapy may be acceptable in some circumstances.

Dialysis is often the treatment of choice for hypercalcemia complicated by renal failure, which is difficult to manage. Peritoneal dialysis with calcium-free dialysis fluid can remove 5 to 12.5 mmol (200 to 500 mg) of calcium in 24 to 48 h and lower the serum calcium concentration by 0.7 to 3 mmol/L (3 to 12 mg/dL). Large quantities of phosphate are lost during dialysis, and serum inorganic phosphate concentrations usually fall, thus aggravating hypercalcemia. Therefore, the serum inorganic phosphate concentration should be measured after dialysis, and phosphate supplements should be added to the

diet or to dialysis fluids if necessary.

Phosphate therapy, oral or intravenous, has a limited role in certain circumstances. Patients with primary hyperparathyroidism are frequently hypophosphatemic, and hypercalcemia of other causes also may be complicated by hypophosphatemia. Hypophosphatemia decreases the rate of calcium uptake into bone, increases intestinal calcium absorption, and directly and indirectly stimulates bone breakdown. These effects aggravate hypercalcemia, and correcting hypophosphatemia lowers the serum calcium concentration. The usual treatment is 1 to 1.5 g phosphorus per day for several days, given in four divided doses to minimize the chances of developing hyperphosphatemia. Such therapy has been administered for prolonged periods in selected patients. It is generally believed, but not established, that toxicity does not occur if therapy is limited to restoring serum inorganic phosphate concentrations to normal.

Raising the serum inorganic phosphate concentration above normal decreases serum calcium levels, sometimes strikingly. Intravenous phosphate is one of the most dramatically effective treatments available for severe hypercalcemia but is toxic and even dangerous (fatal hypocalcemia). For these reasons, it is used rarely and only in severely hypercalcemic patients with cardiac or renal failure. A phosphate phosphorus dose of ³1500 mg intravenously over 6 to 8 h leads to a prompt decrease in serum calcium of as much as 1.2 to 2.5 mmol/L (5 to 10 mg/dL) in patients with initially normal serum inorganic phosphate concentrations. This therapy should be employed only in extreme emergencies. Inorganic phosphate is commercially available for oral use in liquid, powder, and capsule form and as a liquid for intravenous use. It is important to calculate doses in terms of phosphate phosphorus.

Summary The various therapies for hypercalcemia are listed in Table 341-2. The choice depends on the underlying disease, the severity of the hypercalcemia, the serum inorganic phosphate level, and the renal, hepatic, and bone marrow function. Mild hypercalcemia [£3 mmol/L (12 mg/dL)] can usually be managed by hydration. Severe hypercalcemia [3.7 mmol/L (15 mg/dL)] requires rapid correction. Calcitonin should be given for its rapid, albeit short-lived, blockade of bone resorption, and intravenous pamidronate should be administered, although its onset of action is delayed for 1 to 2 days. In addition, for the first 24 to 48 h, aggressive sodium-calcium diuresis with intravenous saline and large doses of furosemide and ethacrynic acid following initial hydration should be initiated, but only if appropriate monitoring is available and cardiac and renal function are adequate. Otherwise, dialysis may be necessary. Intermediate degrees of hypercalcemia between 3.0 and 3.7 mmol/L (12 and 15 mg/dL) should be approached with vigorous hydration and then the most appropriate selection for the patient of the combinations used with severe hypercalcemia.

HYPOCALCEMIA

PATHOPHYSIOLOGY OF HYPOCALCEMIA: CLASSIFICATION BASED ON MECHANISM

Chronic hypocalcemia is less common than hypercalcemia; causes include chronic renal failure, hereditary and acquired hypoparathyroidism, vitamin D deficiency, PHP,

and hypomagnesemia.

Critically ill patients may have transient hypocalcemia with severe sepsis, burns, acute renal failure, and extensive transfusions with citrated blood. Acute hypocalcemia with certain medications is usually transient and may produce no symptoms. Although as many as half of patients in an intensive care setting are reported to have calcium concentrations <2.1 mmol/L (8.5 mg/dL),<10% have a reduction in ionized calcium. Patients with severe sepsis may have a decrease in ionized calcium (true hypocalcemia), but in other severely ill individuals, hypoalbuminemia is the primary cause of the reduced total calcium concentration. Alkalosis increases calcium binding to proteins, and in this setting direct measurements of ionized calcium should be made.

Medications such as protamine, heparin, and glucagon may cause transient hypocalcemia. These forms of hypocalcemia are usually not associated with tetany and resolve with improvement in the overall medical condition. The hypocalcemia after repeated transfusions of citrated blood usually resolves quickly.

Patients with *acute pancreatitis* have hypocalcemia that persists during the acute inflammation and varies in degree with the severity of the pancreatitis. The cause of hypocalcemia remains unclear.PTHvalues are reported to be low, normal, or elevated, and both resistance to PTH and impaired PTH secretion have been postulated. Occasionally, a chronic low total calcium and low ionized calcium concentration are detected in an elderly patient without obvious cause and with a paucity of symptoms; the pathogenesis is unclear.

Chronic hypocalcemia, however, is usually symptomatic and requires treatment. Neuromuscular and neurologic manifestations of chronic hypocalcemia include muscle spasms, carpopedal spasm, facial grimacing, and, in extreme cases, laryngeal spasm and convulsions. Respiratory arrest may occur. Increased intracranial pressure occurs in some patients with long-standing hypocalcemia, often in association with papilledema. Mental changes include irritability, depression, and psychosis. The QT interval on the electrocardiogram is prolonged, in contrast to its shortening with hypercalcemia. Arrhythmias occur, and digitalis effectiveness may be reduced. Intestinal cramps and chronic malabsorption may occur. Chvostek's or Trousseau's sign can be used to confirm latent tetany.

The classification of hypocalcemia shown inTable 341-3 is based on the premise thatPTH is responsible for minute-to-minute regulation of plasma calcium concentration and, therefore, that the occurrence of hypocalcemia must mean a failure of the homeostatic action of PTH. Failure of the PTH response can occur due to hereditary or acquired parathyroid gland failure, if PTH is ineffective in target organs, or if the action of the hormone is overwhelmed by the loss of calcium from theECF at a rate faster than it can be replaced.

PTH ABSENT

Whether hereditary or acquired, hypoparathyroidism has a number of common components. Acute and chronic symptoms of untreated hypocalcemia are shared by both types of hypoparathyroidism, although the onset of hereditary hypoparathyroidism

is more gradual and is often associated with other developmental defects. Basal ganglia calcification and extrapyramidal syndromes are more common and earlier in onset in hereditary hypoparathyroidism. In earlier decades, acquired hypoparathyroidism secondary to surgery in the neck was more common than hereditary hypoparathyroidism, but the frequency of surgically induced parathyroid failure has diminished as a result of improved surgical techniques that spare the parathyroid glands and increased use of nonsurgical therapy for hyperthyroidism.PHP, an example of ineffective PTH action rather than a failure of parathyroid gland production, may share several features with hypoparathyroidism, including extraosseous calcification and extrapyramidal manifestations such as choreoathetotic movements and dystonia.

Papilledema and raised intracranial pressure may occur in both hereditary or acquired hypoparathyroidism, as do chronic changes in fingernails and hair and lenticular cataracts, the latter usually reversible with treatment of hypocalcemia. Certain skin manifestations, including alopecia and candidiasis, are characteristic of hereditary hypoparathyroidism associated with autoimmune polyglandular failure (Chap. 339).

Hypocalcemia associated with hypomagnesemia is associated with both deficientPTHrelease and impaired responsiveness to the hormone. Patients with hypocalcemia secondary to hypomagnesemia have absent or low levels of circulating PTH, indicative of diminished hormone release despite maximum physiologic stimulus by hypocalcemia. Plasma PTH levels return to normal with correction of the hypomagnesemia. Thus hypoparathyroidism with low levels of PTH in blood can be due to hereditary gland failure, acquired gland failure, or acute but reversible gland dysfunction (hypomagnesemia).

Genetic Abnormalities and Hereditary Hypoparathyroidism Hereditary hypoparathyroidism can occur as an isolated entity without other endocrine or dermatologic manifestations (idiopathic hypoparathyroidism); more typically, it occurs in association with other abnormalities such as defective development of the thymus or failure of function of other endocrine organs such as the adrenal, thyroid, or ovary (Chap. 339). Idiopathic and hereditary hypoparathyroidism are often manifest within the first decade but may appear later.

A rare form of hypoparathyroidism associated with defective development of both the thymus and the parathyroid glands is termed the *DiGeorge syndrome* (DGS), or the *velocardiofacial syndrome* (VCFS). Congenital cardiovascular, facial, and other developmental defects are present, and most patients die in early childhood with severe infections, hypocalcemia and seizures, or cardiovascular complications. Some survive into adulthood, and milder, incomplete forms occur. Most cases are sporadic, but an autosomal dominant form involving microdeletions of chromosome 22q11.2 has been described. Smaller deletions in this region are seen in incomplete forms of the DGS syndrome, appearing in childhood or adolescence, that are manifest primarily by parathyroid gland failure.

Hypoparathyroidism can occur in association with a complex hereditary autoimmune syndrome involving failure of the adrenals, the ovaries, the immune system, and the parathyroids in association with recurrent mucocutaneous candidiasis, alopecia, vitiligo, and pernicious anemia (Chap. 339). The responsible gene on chromosome 21q22.3 has

been identified. The protein product, which resembles a transcription factor, has been termed the *AIRE* (autoimmune regulator). A stop codon mutation occurs in many Finnish families with the disorder, commonly referred to as *polyglandular autoimmune type 1 deficiency*.

Gain-of-function mutations in the calcium-sensing receptor cause *autosomal dominant hypocalcemia* (ADH). These mutations induce constitutive receptor functions that lead to features that are the inverse of FHH. The activated receptor suppresses PTH, leading to hypocalcemia; receptor activation in the kidney results in excessive renal calcium excretion. Recognition of the syndrome is important because efforts to treat the hypocalcemia of these patients with vitamin D analogues and increased oral calcium exacerbate the already excessive urinary calcium secretion (several grams or more per 24 h), leading to irreversible renal damage from stones and ectopic calcification.

Hypoparathyroidism is seen in two disorders associated with mitochondrial dysfunction and myopathy, one termed the *Kearns-Sayre syndrome* (KSS), with ophthalmaplegia and pigmentary retinopathy, and the other termed the *MELAS syndrome*, *m*itochondrial *e*ncephalopathy, *l*actic *a*cidosis, and *s*troke-like episodes. Mutations or deletions in mitochondrial genes have been identified (Chap. 67).

The two other rare forms of hypoparathyroidism with other multisystem developmental abnormalities follow either an autosomal dominant pattern, with deafness and/or renal dysplasia, or an autosomal recessive pattern, with growth retardation and dysmorphic features.

Hereditary hypoparathyroidism occurs also as an isolated entity without any other defects. The pattern of inheritance varies and includes autosomal dominant, autosomal recessive, and X-linked inheritance patterns. In one family in which the disorder is transmitted as an autosomal dominant trait, a structural abnormality in the PTH gene has been identified. A defect in the signal sequence needed for processing of the hormone impairs PTH secretion. In another kindred with autosomal recessive inheritance, the mutant allele in the first intron of the PTH gene causes a splicing defect in mRNA production. An X-linked recessive form of hypoparathyroidism has been described in males from two kindreds that are probably related. The locus of the defect has been located to chromosome Xq26-q27.

Acquired Hypoparathyroidism *Acquired chronic hypoparathyroidism* is usually the result of inadvertent surgical removal of all the parathyroid glands; in some instances, not all the tissue is removed, but the remainder undergoes vascular supply compromise secondary to fibrotic changes in the neck after surgery. In the past, the most frequent cause of acquired hypoparathyroidism was surgery for hyperthyroidism. Hypoparathyroidism now usually occurs after surgery for hyperparathyroidism when the surgeon, facing the dilemma of removing too little tissue and thus not curing the hyperparathyroidism, removes too much. Parathyroid function may not be totally absent in all patients with postoperative hypoparathyroidism.

Even rarer causes of acquired chronic hypoparathyroidism include radiation-induced damage subsequent to radioiodine therapy of hyperthyroidism and glandular damage in patients with hemochromatosis or hemosiderosis after repeated blood transfusions.

Infection may involve one or more of the parathyroids but usually does not cause hypoparathyroidism because all four glands are rarely involved.

Transient hypoparathyroidism is frequent following surgery for hyperparathyroidism. After a variable period of hypoparathyroidism, normal parathyroid function may return due to hyperplasia or recovery of remaining tissue. Occasionally, recovery occurs months after surgery.

TREATMENT

Treatment of acquired and hereditary hypoparathyroidism involves replacement with vitamin D or $1,25(OH)_2D_3$(calcitriol) combined with a high oral calcium intake. In most patients, blood calcium and phosphate levels are satisfactorily regulated, but some patients show resistance and a brittleness with a tendency to alternate between hypocalcemia and an overshoot hypercalcemia. For many patients, vitamin D in doses of 1 to 3 mg/d (40,000 to 120,000 U/d) combined with³1 g elemental calcium is satisfactory. The wide dosage range reflects the variation encountered from patient to patient; precise regulation of each patient is required. Compared to typical daily requirements in euparathyroid patients of 200 U/d, the high dose of vitamin D reflects the reduced conversion of vitamin D to $1,25(OH)_2D$. Many physicians now use 0.5 to 1.0 ug of calcitriol in management of such patients, especially if they are difficult to control. When vitamin D (because of storage in fat) is withdrawn, weeks are required for the disappearance of the biologic effects, compared with a few days for calcitriol, which has a rapid turnover.

Oral calcium and vitamin D restore the overall calcium-phosphate balance but do not reverse the lowered urinary calcium reabsorption typical of hypoparathyroidism. Therefore, care must be taken to avoid excessive urinary calcium excretion after vitamin D and calcium replacement therapy; otherwise, kidney stones can develop. Thiazide diuretics lower urine calcium by as much as 100 mg/d in hypoparathyroid patients on vitamin D, provided they are maintained on a low-sodium diet. Use of thiazides seems to be of benefit in mitigating hypercalciuria and easing the daily management of these patients.

Hypomagnesemia Severe hypomagnesemia is associated with hypocalcemia (Chap. 340). Restoration of the total-body magnesium deficit leads to rapid reversal of hypocalcemia. There are at least two causes of the hypocalcemia -- impairedPTHsecretion and reduced responsiveness to PTH.

Hypomagnesemia is generally classified as primary or secondary; primary hypomagnesemia is due to hereditary defects in intestinal absorption or renal reabsorption of magnesium. Secondary hypomagnesemia, a more common condition, occurs on a nutritional basis or as a result of acquired intestinal or renal disorders. The most common causes of the secondary disorder are chronic alcoholism with poor nutritional intake, intestinal malabsorption syndromes, and parenteral nutrition when magnesium replacement is omitted.

The effects of magnesium onPTHsecretion are similar to those of calcium; hypermagnesemia suppresses and hypomagnesemia stimulates PTH secretion. The

effects of magnesium on PTH secretion are normally of little significance, however, because the calcium effects dominate. Greater change in magnesium than in calcium is needed to influence hormone secretion. Nonetheless, hypomagnesemia might be expected to increase hormone secretion. It is therefore surprising to find that severe hypomagnesemia is associated with blunted secretion of PTH. The explanation for the paradox is that severe, chronic hypomagnesemia leads to intracellular magnesium deficiency, which interferes with secretion and peripheral responses to PTH. The mechanism of the cellular abnormalities caused by hypomagnesemia is unknown, although effects on adenylate cyclase (for which magnesium is a cofactor) have been proposed.

Serum magnesium must usually fall below 0.4 mmol/L (1.0 mg/dL) to cause hypocalcemia. PTH levels are undetectable or inappropriately low despite the stimulus of severe hypocalcemia, and acute repletion of magnesium leads to a rapid increase in PTH level. Serum phosphate levels are often not elevated, in contrast to the situation with acquired or idiopathic hypoparathyroidism, probably because phosphate deficiency is a frequent accompaniment of hypomagnesemia.

Diminished peripheral responsiveness to PTH also occurs in some patients, as documented by subnormal response in urinary phosphorus and urinary cyclic AMP excretion after administration of exogenous PTH to patients who are hypocalcemic and hypomagnesemic. Both blunted PTH secretion and lack of renal response to administered PTH can occur in the same patient. When acute magnesium repletion is undertaken, the restoration of PTH levels to normal or supranormal may precede restoration of normal serum calcium by several days.

TREATMENT

Repletion of magnesium cures the condition, and attention must be given to restoring the intracellular deficiency, which may be considerable. Repletion should be parenteral. After intravenous magnesium administration, serum magnesium may return transiently to the normal range, but unless replacement therapy is adequate serum magnesium will again fall. If renal function is normal, urinary magnesium excretion is a useful indicator of magnesium repletion, as magnesium is retained by the kidney until the deficiency is corrected. Intracellular deficits can be ³50 mmol. Parenteral administration of 10 to 14 mmol magnesium usually reverses the signs of magnesium deficiency, but greater amounts may occasionally be required if the deficit is large. If the cause of the hypomagnesemia is renal magnesium wasting, magnesium may have to be given chronically to prevent recurrence (Chap. 340).

PTH INEFFECTIVE

PTH is ineffective when the hormone receptor-guanyl nucleotide-binding protein complex is defective (PHP, discussed below), when PTH action to promote calcium absorption from the diet is impaired because of vitamin D deficiency or because vitamin D is ineffective (receptor or synthesis defects), or in chronic renal failure in which the calcium-elevating action of PTH is impaired.

Typically, hypophosphatemia is more severe than hypocalcemia in vitamin D deficiency

states because of the increased secretion of PTH, which, although only partly effective in elevating blood calcium, is capable of promoting phosphaturia.

PHP, on the other hand, has a pathophysiology different from the other disorders of ineffective PTH action. PHP resembles hypoparathyroidism (in which PTH synthesis is deficient) and is manifested by hypocalcemia and hyperphosphatemia. The cause of the disorder is defective hormone activation of guanyl nucleotide-binding proteins, resulting in failure of PTH to increase intracellular cyclic AMP (see below).

Chronic Renal Failure Improved medical management of chronic renal failure and/or a more indolent course of the renal disease now allow many patients to survive long enough to develop features of renal osteodystrophy. Phosphate retention and impaired production of 1,25(OH)$_2$D are the principal factors that cause calcium deficiency, secondary hyperparathyroidism, and bone disease. The uremic state also causes impairment of intestinal absorption by mechanisms other than defects in vitamin D metabolism. Nonetheless, treatment with supraphysiologic amounts of vitamin D or calcitriol corrects the impaired calcium absorption.

Hyperphosphatemia in renal failure lowers blood calcium levels by several mechanisms, including extraosseous deposition of calcium and phosphate, impairment of the bone-resorbing action of PTH, and reduction in 1,25(OH)$_2$D production by remaining renal tissue. Low levels of 1,25(OH)$_2$D due to hyperphosphatemia and destruction of renal tissue and are critical in the development of hypocalcemia.

TREATMENT

Therapy of chronic renal failure (Chap. 270) involves appropriate management of patients prior to dialysis and adjustment of regimens once dialysis is initiated. Attention should be paid to restriction of phosphate in the diet; use of calcium-containing salts as phosphate-binding antacids is preferable, rather than aluminum, to avoid the problem of aluminum intoxication; provision of an adequate calcium intake by mouth, usually 1 to 2 g/d; and supplementation with 0.25 to 1.0 ug/d calcitriol. Each patient must be monitored closely. The aim of therapy is to restore normal calcium balance to prevent osteomalacia and secondary hyperparathyroidism. Reduction of hyperphosphatemia and restoration of normal intestinal calcium absorption by calcitriol can improve blood calcium levels and reduce the manifestations of secondary hyperparathyroidism. Since adynamic bone disease can occur in association with low PTH levels, it is important to avoid excessive suppression of the parathyroid glands while recognizing the beneficial effects of controlling the secondary hyperparathyroidism. These patients should probably be closely monitored with PTH assays that detect only the full-length PTH 1-84 to avoid interference by biologically inactive amino-terminally truncated PTH.

Vitamin D Deficiency due to Inadequate Diet and/or Sunlight Vitamin D deficiency due to inadequate intake of dairy products enriched with vitamin D, lack of vitamin supplementation, and reduced sunlight exposure in the elderly, particularly during winter in northern latitudes, is more common in the United States than previously recognized. Biopsies of bone in elderly patients with hip fracture (documenting osteomalacia) and abnormal levels of vitamin D metabolites, PTH, calcium, and phosphate indicate that vitamin D deficiency may occur in as many as 25% of elderly patients, particularly in

areas where there is little ambient sunlight. Concentrations of 25(OH)D are low or low-normal in these patients. Quantitative histomorphometry of bone biopsy specimens reveals widened osteoid seams consistent with osteomalacia. PTH hypersecretion compensates for the tendency for the blood calcium to fall but also induces renal phosphate wasting and results in osteomalacia.

Treatment involves adequate replacement with vitamin D and calcium until the deficiencies are corrected. Severe hypocalcemia rarely occurs in moderately severe vitamin D deficiency of the elderly, but vitamin D deficiency must be considered in the differential diagnosis of mild hypocalcemia.

Defective Vitamin D Metabolism

Anticonvulsant therapy Anticonvulsant therapy with any of several agents induces acquired vitamin D deficiency by increasing the conversion of vitamin D to inactive compounds. The more marginal the vitamin D intake in the diet, the more likely that anticonvulsant therapy will lead to abnormal mineral and bone metabolism (Chap. 340).

Although $1,25(OH)_2D$ levels are lower in patients treated with chronic anticonvulsants than in the normal population, there is a great deal of variation. The greater prevalence of the disorder in some European populations and in the mentally retarded may reflect the lower vitamin D intake of those groups. Restoration of bone mineral mass and reversal of hypocalcemia can be accomplished with vitamin D replacement plus oral calcium. Administration of 50,000 units of vitamin D monthly may be preventive if anticonvulsant therapy needs to be given chronically.

Vitamin D-dependent rickets type I Rickets can be due to *resistance to the action* of vitamin D as well as to vitamin D deficiency. Vitamin D-dependent rickets type I, previously termed *pseudo-vitamin D-resistant rickets*, differs from true vitamin D-resistant rickets (vitamin D-dependent rickets type II, see below) in that it is less severe and the biochemical and radiographic abnormalities can be reversed with appropriate doses of the vitamin or the active metabolite, $1,25(OH)_2D_3$.

Clinical features include hypocalcemia, often with tetany or convulsions, hypophosphatemia, secondary hyperparathyroidism, and osteomalacia, often associated with skeletal deformities and increased alkaline phosphatase. Physiologic amounts of calcitriol cure the disease (Chap. 340). This finding fits with the pathophysiology as the disorder, which is autosomal recessive, is now known to be caused by a series of mutations in the gene for the 25(OH)D-1a-hydroxylase. Over 20 different mutations have been identified. All patients have both alleles inactivated, but often the genetic pattern is that of a compound heterozygote. Response to high doses of vitamin D or calcifediol, as noted in prior years, is probably due to direct effects of 25(OH)D at high levels. Treatment begins with 1 to 2 ug/d calcitriol, but maintenance is satisfactory with physiologic doses of calcitriol (0.5 to 1.0 ug/d). Careful adjustment of calcitriol dose is required, particularly during growth periods.

Vitamin D Ineffective

Intestinal Malabsorption Mild hypocalcemia, secondary hyperparathyroidism, severe

hypophosphatemia, and a variety of nutritional deficiencies occur with gastrointestinal diseases. Hepatocellular dysfunction can lead to reduction in 25(OH)D levels, as in portal or biliary cirrhosis of the liver, and malabsorption of vitamin D and its metabolites, including 1,25(OH)$_2$D, may occur in a variety of bowel diseases, hereditary or acquired. Hypocalcemia itself can lead to steatorrhea, due to deficient production of pancreatic enzymes and bile salts. Depending on the disorder, vitamin D or its metabolites can be given parenterally, guaranteeing adequate blood levels of active metabolites.

Vitamin D-dependent rickets type II Vitamin D-dependent rickets type II results from end-organ resistance to the active metabolite 1,25(OH)$_2$D$_3$. The clinical features resemble those of the type I disorder and include hypocalcemia, hypophosphatemia, secondary hyperparathyroidism, and rickets. A clear distinction is partial or total alopecia in type II. Plasma levels of 1,25(OH)$_2$D are at least three times normal, in keeping with the refractoriness of the end organs. Some patients respond to very high doses of vitamin D or vitamin D metabolites (e.g., 17 to 20 ug/d calcitriol). Earlier suggestions that there were both receptor and postreceptor defects are incorrect. All of the genetically characterized phenotypes have mutations in the gene for the vitamin D receptor. Nineteen mutations have been identified that affect different regions of the receptor primarily in the DNA binding domain (with normal ligand binding: these were the so-called postreceptor defects detected by earlier indirect methods) or in the ligand-binding domain (classified previously as receptor negative).

Pseudohypoparathyroidism PHP is a hereditary disorder characterized by symptoms and signs of hypoparathyroidism, typically in association with distinctive skeletal and developmental defects. The hypoparathyroidism is due to a deficient end-organ response to PTH. Hyperplasia of the parathyroids, a response to hormone resistance, causes elevation of PTH levels. Studies, both clinical and basic, have clarified some aspects of this syndrome, including the variable clinical spectrum, the pathophysiology, the genetic defects, and the inheritance.

A working classification of the various forms of PHP is given in Table 341-4. The classification scheme is based on the signs of ineffective PTH action (low calcium and high phosphate), urinary cyclic AMP response to exogenous PTH, the presence or absence of *Albright's hereditary osteodystrophy* (AHO), and assays of the concentration of the G$_s$a subunit of the adenylate cyclase enzyme. Using these criteria, there are four types: PHP type I, subdivided into a and b categories; PHP-II; and pseudopseudohypoparathyroidism (PPHP).

PHP-Ia and PHP-Ib Individuals with PHP-I, the most common of the disorders, show a deficient urinary cyclic AMP response to administration of exogenous PTH. Patients with PHP-I are divided into type a, who have reduced amounts of G$_s$a in vitro assays with erythrocytes, and type b, with normal amounts of G$_s$a in erythrocytes. There is a third type (PHP-Ic, reported in a few patients) that differs from PHP-Ia only in having normal erythroycte levels of G$_s$a despite having AHO, hypocalcemia, and decreased urinary cyclic AMP responses to PTH (presumably with a post-G$_s$a defect in adenyl cyclase stimulation).

Most patients show characteristic features of AHO, consisting of short stature, round face, skeletal anomalies (brachydactyly), and heterotopic calcification. Patients have low

calcium and high phosphate levels, as with true hypoparathyroidism. PTH levels, however, are elevated, reflecting resistance to hormone action.

Amorphous deposits of calcium and phosphate are found in the basal ganglia in about half of patients. The defects in metacarpal and metatarsal bones are sometimes accompanied by short phalanges as well, possibly reflecting premature closing of the epiphyses. The typical findings are short fourth and fifth metacarpals and metatarsals. The defects are usually bilateral. Exostoses and radius curvus are frequent. Impairments in olfaction and taste and unusual dermatoglyphic abnormalities have been reported.

PPHP The initial view that the defect responsible for PHP-Ia was simply the deficiency of $G_s\alpha$ subunits was temporarily confounded by the subsequent discovery that the same 50% reduction in $G_s\alpha$ subunits was seen in patients with PPHP, who have typical features of the hereditary osteodystrophy syndrome despite normal serum calcium levels and normal response of urinary cyclic AMP to exogenous PTH.

Multiple defects have now been identified in the *GNAS-1* gene in PHP-Ia and PPHP patients. This gene, which is located on chromosome 20q13, encodes the stimulatory G protein subunit $G_s\alpha$, among other products (see below). Mutations include abnormalities in splice junctions associated with deficient mRNA production and point mutations that result in a protein with defective function as well as the 50% reduction in $G_s\alpha$ levels in erythrocytes.

Detailed analyses of disease transmission in affected kindreds have clarified many features of PHP-Ia, PPHP, and PHP-Ib (Fig. 341-10). The former two entities, traced through multiple kindreds, have an inheritance pattern consistent with gene imprinting -- only females, not males, can transmit the full disease with hypocalcemia -- and PHP and PPHP do not coexist in the same generation. The phenomenon of gene imprinting involves selective inactivation of either the maternal or the paternal allele (Chap. 65). In the case of the $G_s\alpha$ gene, it is paternally imprinted (silenced) so that the disease PHP-Ia is never inherited from the father carrying the defective allele but only from the mother. On the other hand, the defective allele is not imprinted or silenced in all tissues. It seems possible, therefore, that the AHO phenotype recognized in PPHP as well as PHP-Ia reflects haplotype insufficiency. In the renal cortex, however, it is postulated that only the maternal allele is normally active, such that lack of activity from a defective paternal allele is not of consequence. This explains the occurrence in PHP-Ia of hypocalcemia, hyperphosphatemia, and other stigmata such as variable resistance to other hormones (if similar tissue-specific imprinting occurs in other organs). Strong evidence favoring this overall hypothesis comes from gene knockout studies in the mouse (ablating exon 2 of the gene). Mice inheriting the mutant allele from the female had undetectable $G_s\alpha$ protein in renal cortex and were hypocalcemic and resistant to renal actions of PTH. Offspring inheriting the mutant allele from the male showed no evidence of PTH resistance or hypercalcemia.

The complex mechanisms that control the GNAS-1 gene also contribute to challenges involved in unraveling the pathogenesis of these disorders. Alternative splicing patterns produce three different transcripts that encode distinct proteins. In addition to $G_s\alpha$, this gene encodes a second protein product with a unique NH_2-terminus (the XL exon); $XL_\alpha s$

includes exons 2-13. It is unknown whether this protein can function as a stimulatory G protein, but the mRNA encoding it is expressed in numerous endocrine tissues and is transcribed from only the paternal allele. A third transcript is transcribed from only the maternal allele and encodes the protein product, NESP55, which contains no homology with XL_as or G_sa.

PHP-Ib, lacking the AHO phenotype, shares with PHP-Ia the resistance to PTH action and a blunted urinary cyclic AMP response to administered PTH, a standard test for hormone resistance (Table 341-4). PHP-Ib patients, however, show normal levels of G_sa in erythrocytes. Bone responsiveness may be excessive rather than blunted in PHP-Ib compared to PHP-Ia patients, based on case reports that have emphasized an osteitis fibrosa-like pattern in some PHP patients who lack the AHO phenotype. The inheritance patterns in PHP-Ib kindreds are clearly consistent with paternal imprinting and lack male transmission of symptomatic disease; gene cloning studies have narrowed the responsible region to chromosome 20, close to -- if not within -- the *GNAS-1* gene locus. Elucidation of the responsible genetic and pathogenetic mechanisms in this disorder may further illuminate the function of the complex *GNAS-1* gene and the role of its products in hormonal signaling.

PHP-II refers to patients with hypocalcemia and hyperphosphatemia who have a normal urinary cyclic AMP response to PTH. These patients are assumed to have a defect in the response to PTH at a locus distal to cyclic AMP production, although at least some patients may instead have occult vitamin D deficiency.

The diagnosis of these hormone-resistant states can usually be made without difficulty when there is a positive family history for developmental defects and/or the presence of developmental anomalies, including brachydactyly, in association with the signs and symptoms of hypoparathyroidism. In all categories -- PHP-Ia, -Ib, and -II -- serum PTH levels are elevated, particularly when patients are hypocalcemic. However, patients with PHP-Ib or PHP-II do not have phenotypic abnormalities, only hypocalcemia with high PTH levels, confirming hormone resistance. In PHP-Ib, the response of urinary cyclic AMP to the administration of exogenous PTH is blunted. Levels of G_sa subunits in erythrocyte membranes are, however, normal in those with PHP-Ib. The diagnosis of PHP-II is more complex, in that cyclic AMP responses in urine are, by definition, normal. Since vitamin D deficiency itself can dissociate phosphaturic and urinary cyclic AMP responses to exogenous PTH, vitamin D deficiency must be excluded before the diagnosis of PHP-II can be entertained.

TREATMENT

Treatment of PHP is similar to that of hypoparathyroidism, except that the doses of vitamin D and calcium are usually lower than those required in true hypoparathyroidism, presumably because the defect in PHP is only partial because of imprinting in specific tissues (renal cortex vs. renal medulla). Variability in response makes it necessary to establish the optimal regimen for each patient, based on maintaining the appropriate blood calcium level and urinary calcium excretion.

PTHOverwhelmed Occasionally, loss of calcium from the ECF is so severe that PTH cannot compensate. Such situations include acute pancreatitis and severe, acute

hyperphosphatemia, often in association with renal failure, conditions in which there is rapid efflux of calcium from extracellular fluid. Severe hypocalcemia can occur quickly; PTH rises in response to hypocalcemia but does not return blood calcium to normal.

Severe, Acute Hyperphosphatemia Severe hyperphosphatemia is associated with extensive tissue damage or cell destruction (Chap. 340). The combination of increased release of phosphate from muscle and impaired ability to excrete phosphorus because of renal failure causes moderate to severe hyperphosphatemia, the latter causing calcium loss from the blood and mild to moderate hypocalcemia. Hypocalcemia is usually reversed with tissue repair and restoration of renal function as phosphorus and creatinine values return to normal. There may even be a mild hypercalcemic period in the oliguric phase of renal function recovery. This sequence, severe hypocalcemia followed by mild hypercalcemia, reflects widespread deposition of calcium in muscle and subsequent redistribution of some of the calcium to the ECF after return of phosphate levels to normal.

Other causes of hyperphosphatemia include hypothermia, massive hepatic failure, and hematologic malignancies, either because of high cell turnover of malignancy or because of cell destruction by chemotherapy.

TREATMENT

Treatment is directed toward lowering of blood phosphate by the administration of phosphate-binding antacids or dialysis, often needed for the management of renal failure. Although calcium replacement may be necessary if hypocalcemia is severe and symptomatic, calcium administration during the hyperphosphatemic period tends to increase extraosseous calcium deposition and aggravate tissue damage. The levels of 1,25(OH)$_2$D may be low during the hyperphosphatemic phase and return to normal during the oliguric phase of recovery.

Osteitis Fibrosis after Parathyroidectomy Severe hypocalcemia after parathyroid surgery is less common now that osteitis fibrosa cystica is an infrequent manifestation of hyperparathyroidism. When osteitis fibrosa cystica is severe, however, bone mineral deficits can be large. After parathyroidectomy, hypocalcemia can persist for days if calcium replacement is inadequate. Treatment may require parenteral administration of calcium; addition of calcitriol and oral calcium supplementation is sometimes needed for weeks to a month or two until bone defects are filled (which, of course, is of therapeutic benefit in the skeleton), making it possible to discontinue parenteral calcium and/or reduce the amount.

DIFFERENTIAL DIAGNOSIS OF HYPOCALCEMIA

Care must be taken to ensure that true hypocalcemia is present; in addition, acute transient hypocalcemia can be a manifestation of a variety of severe, acute illnesses, as discussed above. *Chronic hypocalcemia*, however, can usually be ascribed to a few disorders associated with absent or ineffective PTH. Important clinical criteria include the duration of the illness, signs or symptoms of associated disorders, and the presence of features that suggest a hereditary abnormality. A nutritional history can be helpful in recognizing a low intake of vitamin D and calcium in the elderly, and a history of

excessive alcohol intake may suggest magnesium deficiency.

Hypoparathyroidism and PHP are typically lifelong illnesses, usually (but not always) appearing by adolescence; hence a recent onset of hypocalcemia in an adult is more likely due to nutritional deficiencies, renal failure, or intestinal disorders that result in deficient or ineffective vitamin D. Neck surgery, even long past, however, can be associated with a delayed onset of postoperative hypoparathyroidism. A history of seizure disorder raises the issue of anticonvulsive medication. Developmental defects, particularly in childhood and adolescence, may point to the diagnosis of PHP. Rickets and a variety of neuromuscular syndromes and deformities may indicate ineffective vitamin D action, either due to defects in vitamin D metabolism or to vitamin D deficiency.

A pattern of *low calcium with high phosphorus* in the absence of renal failure or massive tissue destruction almost invariably means hypoparathyroidism or PHP. A *low calcium and low phosphorus* points to absent or ineffective vitamin D, thereby impairing the action of PTH on calcium metabolism (but not phosphate clearance). The relative ineffectiveness of PTH in vitamin D deficiency, anticonvulsant therapy, gastrointestinal disorders, and hereditary defects in vitamin D metabolism leads to secondary hyperparathyroidism as a compensation. The relatively unopposed action of the excess PTH on renal tubule phosphate transport, which is less dependent on vitamin D than calcium transport, accounts for renal phosphate wasting and hypophosphatemia.

Exceptions to these patterns may occur. Most forms of hypomagnesemia are due to long-standing nutritional deficiency as seen in chronic alcoholics. Despite the fact that the hypocalcemia is principally due to an acute absence of PTH, phosphate levels are usually low, rather than elevated as in hypoparathyroidism. Chronic renal failure is often associated with hypocalcemia and hyperphosphatemia, despite secondary hyperparathyroidism.

Diagnosis is usually established by application of the PTH immunoassay, tests for vitamin D metabolites, and measurements of the urinary cyclic AMP response to exogenous PTH. In hereditary and acquired hypoparathyroidism and in severe hypomagnesemia, PTH is either undetectable or in the normal range. This finding in a hypocalcemic patient is supportive of hypoparathyroidism, as distinct from ineffective PTH action, in which even mild hypocalcemia is associated with elevated PTH levels. Hence a failure to detect elevated PTH levels establishes the diagnosis of hypoparathyroidism; elevated levels suggest the presence of secondary hyperparathyroidism, as found in many of the situations in which the hormone is ineffective due to associated abnormalities in vitamin D action. Assays for 25(OH)D and 1,25(OH)$_2$D can be helpful. Low or low normal 25(OH)D indicates vitamin D deficiency due to lack of sunlight, inadequate vitamin D intake, or intestinal malabsorption. A low level of 1,25(OH)$_2$D in the presence of elevated concentrations of PTH suggests ineffective PTH action in disorders such as chronic renal failure, severe vitamin D deficiency, vitamin D-dependent rickets type I, and PHP. Recognition that mild hypocalcemia, rickets, and hypophosphatemia are due to anticonvulsant therapy is made by history.

TREATMENT

Hypocalcemic States The management of hypoparathyroidism,<u>PHP</u>, chronic renal failure, and hereditary defects in vitamin D metabolism involves the use of vitamin D or vitamin D metabolites and calcium supplementation. Vitamin D itself is the least expensive form of vitamin D replacement and is frequently used in the management of uncomplicated hypoparathyroidism and some disorders associated with ineffective vitamin D action. When vitamin D is used prophylactically, as in the elderly or in those with chronic anticonvulsant therapy, there is a wider margin of safety than with the more potent metabolites. However, most of the conditions in which vitamin D is administered chronically for hypocalcemia require amounts 50 to 100 times the daily replacement dose because the formation of $1,25(OH)_2D$ is deficient. In such situations, vitamin D is no safer than the active metabolite because intoxication can occur with high-dose therapy (because of storage in fat). Calcitriol is more rapid in onset of action and also has a short biologic half-life.

Vitamin D (5 ug/d) or calcifediol and lower doses of calcitriol (0.25 to 1.0 ug/d) are required to prevent rickets in normal individuals. In contrast, 1 to 3 mg (1000 to 3000 ug of vitamin D_2 or D_3 is typically required in hypoparathyroidism; doses of calcifediol are also high (several hundred micrograms per day). The dose of calcitriol is unchanged in hypoparathyroidism, since the defect is in hydroxylation by the 25(OH)D-1a-hydroxylase.

Patients with hypoparathyroidism should be given 2 to 3 g elemental calcium by mouth each day. The two agents, vitamin D or calcitriol and oral calcium, can be varied independently. If hypocalcemia alternates with episodes of hypercalcemia in more brittle patients with hypoparathyroidism, administration of calcitriol and use of thiazides, as discussed above, may make management easier.

(Bibliography omitted in Palm version)

342. OSTEOPOROSIS - *Robert Lindsay, Felicia Cosman*

Osteoporosis, characterized by decreased bone strength, is prevalent among postmenopausal women but also occurs in men and women with underlying conditions or major risk factors associated with bone demineralization. Its chief clinical manifestations are vertebral and hip fractures. Osteoporosis affects>10 million individuals in the United States, but only 10 to 20% are diagnosed and treated.

DEFINITION

Osteoporosis is defined as a reduction of bone mass (or density) or the presence of a fragility fracture. This reduction in bone tissue is accompanied by deterioration in the architecture of the skeleton, leading to a markedly increased risk of fracture. Osteoporosis is defined operationally as a bone density that falls 2.5 standard deviations (SD) below the mean -- also referred to as a *T-score* of-2.5. Those who fall at the lower end of the young normal range (a T-score of >1 SD below the mean) have low bone density and are considered to be at increased risk of osteoporosis.

EPIDEMIOLOGY

In the United States, as many as 8 million women and 2 million men have osteoporosis (T-score <-2.5), and an additional 18 million individuals have bone mass levels that put them at increased risk of developing osteoporosis (e.g., bone mass T-score <-1.0). Osteoporosis occurs more frequently with increasing age as bone tissue is progressively lost. In women, the loss of ovarian function at menopause (typically after age 50) precipitates rapid bone loss such that most women meet the criteria for osteoporosis by age 70.

The epidemiology of fractures follows similar trends as the loss of bone density. Fractures of the distal radius increase in frequency before age 50 and plateau by age 60, with only a modest age-related increase thereafter. In contrast, incidence rates for hip fractures double every 5 years after age 70 (Fig. 342-1). This distinct epidemiology may be related to the way people fall as they age, with fewer falls on an outstretched hand. At least 1.5 million fractures occur each year in the United States as a consequence of osteoporosis. As the population continues to age, the total number of fractures will continue to escalate.

About 300,000 hip fractures occur each year in the United States, most of which require hospital admission and surgical intervention. The probability that a 50-year-old white individual will have a hip fracture during his or her lifetime is 14% for women and 5% for men; the risk for African Americans is much lower (about half these rates). Hip fractures are associated with a high incidence of deep vein thrombosis and pulmonary embolism (20 to 50%) and a mortality rate between 5 and 20% during the few months after surgery.

There are about 500,000 vertebral crush fractures per year in the United States. Only a fraction of these are recognized clinically, since many are relatively asymptomatic and are identified incidentally during radiography for other purposes (Fig. 342-2). Vertebral fractures rarely require hospitalization but are associated with long-term morbidity and a

slight increase in mortality. Multiple fractures lead to height loss (often of several inches), kyphosis, and secondary pain and discomfort related to altered biomechanics of the back. Thoracic fractures can be associated with restrictive lung disease, whereas lumbar fractures are associated with abdominal symptoms including distention, early satiety, and constipation.

Approximately 200,000 wrist fractures occur in the United States each year. Fractures of other bones also occur with osteoporosis, which is not surprising given that bone loss is a systemic phenomenon. Fractures of the pelvis and proximal humerus are clearly associated with osteoporosis. Although some fractures are clearly the result of major trauma, the threshold for fracture is reduced for an osteoporotic bone (Fig. 342-3). A list of common risk factors for osteoporotic fractures is summarized inTable 342-1. Prior fractures, a family history of osteoporotic fractures, and low body weight are each independent predictors of fracture. Chronic diseases that increase the risk of falling or frailty, including dementias, Parkinson's disease, and multiple sclerosis, also increase fracture risk.

In the United States and Europe, osteoporosis-related fractures are more common among women than men, presumably due to a lower peak bone mass as well as postmenopausal bone loss in women. However, this gender difference in bone density and age-related increase in hip fractures is not as apparent in some other cultures, possibly due to genetics, physical activity level, or diet.

PATHOPHYSIOLOGY

Bone Remodeling Osteoporosis results from bone loss due to normal age-related changes in bone remodeling as well as extrinsic and intrinsic factors that exaggerate this process. These changes may be superimposed on a low peak bone mass. Consequently, the bone remodeling process is fundamental for understanding the pathophysiology of osteoporosis (Chap. 340). The skeleton increases in size by linear growth and by apposition of new bone tissue on the outer surfaces of the cortex (Fig. 342-4). This latter process is the phenomenon of modeling, which also allows the long bones to adapt in shape to the stresses placed upon them. Increased sex hormone production at puberty is required for maximum skeletal maturation, which reaches maximum mass and density in early adulthood. Nutrition and lifestyle also play an important role in growth, though genetic factors are the major determinants of peak skeletal mass and density. Numerous genes control skeletal growth, peak bone mass, and body size, but it is likely that separate genes control skeletal structure and density. Heritability estimates of 50 to 80% for bone density and size have been derived based on twin studies. Though peak bone mass is often lower among individuals with a family history of osteoporosis, association studies of candidate genes [vitamin D receptor; Type I collagen, the estrogen receptor (ER), interleukin (IL) 6; and insulin-like growth factor (IGF) I] have not been consistently replicated. Linkage studies suggest that a genetic locus on chromosome 11 is associated with high bone mass.

Once peak skeletal mass has been attained, the process of remodeling becomes the principal metabolic activity of the skeleton. This process has three primary functions: (1) to repair microdamage within the skeleton, (2) to maintain skeletal strength, and (3) to supply calcium from the skeleton to maintain serum calcium. Acute demands for calcium

involve osteoclast-mediated resorption as well as calcium transport by osteocytes. The activation of remodeling may be induced by microdamage to bone due to excessive or accumulated stress.

Bone remodeling is also regulated by several circulating hormones, including estrogens, androgens, vitamin D, and parathyroid hormone (PTH), as well as locally produced growth factors such as IGF-I and -II, transforming growth factor (TGF) b, parathyroid hormone-related peptide (PTHrP), ILs, prostaglandins, tumor necrosis factor (TNF), and osteoprotegrin ligand (Fig. 342-5). Additional influences include include nutrition (particularly calcium intake) and physical activity level. The end result of this remodeling process is that the resorbed bone is replaced by an equal amount of new bone tissue. Thus, the mass of the skeleton remains constant after peak bone mass is achieved in adulthood. After age 30 to 45, however, the resorption and formation processes become imbalanced, and resorption exceeds formation. This imbalance may begin at different ages and varies at different skeletal sites; it becomes exaggerated in women after menopause. Excessive bone loss can be due to an increase in osteoclastic activity and/or a decrease in osteoblastic activity. In addition, an increase in remodeling activation frequency can magnify the small imbalance seen at each remodeling unit.

In trabecular bone, if the osteoclasts are sufficiently aggressive to penetrate trabeculae, they leave no template for new bone formation to occur and, consequently, may cause rapid bone loss. In cortical bone, increased activation of remodeling creates more porous bone. The effect of this increased porosity on cortical bone strength may be modest if the overall diameter of the bone is not changed. However, decreased apposition of new bone on the periosteal surface coupled with increased endocortical resorption of bone decreases the biomechanical strength of long bones. Even a slight exaggeration in normal bone loss patterns increases the risk of osteoporotic fracture.

Calcium Nutrition Peak bone mass may be impaired by inadequate calcium intake during growth, thereby leading to increased risk of osteoporosis in later life. During the adult phase of life, calcium deprivation induces secondary hyperparathyroidism and an increase in the rate of remodeling. PTH stimulates the hydroxylation of vitamin D in the kidney, leading to increased levels of 1,25-dihydroxyvitamin D [1,25(OH)$_2$D] and enhanced gastrointestinal calcium absorption. PTH also reduces renal calcium loss. Though these are appropriate short-term homeostatic responses for improving calcium economy, the long-term effects are detrimental to the skeleton because of the ongoing imbalance at remodeling sites.

Total daily calcium intakes of <400 mg are likely to be detrimental to the skeleton, but there is more doubt about intakes in the 600- to 800-mg range, which is the average intake among adults in the United States. The recommended daily required intake of 1000 to 1200 mg for adults accomodates population heterogeneity in controlling calcium balance (Chap. 73).

Vitamin D (See also Chap. 340) Severe vitamin D deficiency causes rickets in children or osteomalacia in adults. There is accumulating evidence that vitamin D deficiency may be more prevalent than previously thought, particularly among individuals at increased risk, such as the elderly; those living in northern latitudes; and in individuals with poor nutrition, malabsorption, or chronic liver or renal disease. Modest vitamin D deficiency

leads to compensatory secondary hyperparathyroidism and is an important risk factor for osteoporosis and fractures. Some studies have shown that>50% of inpatients on a general medical service exhibit biochemical features of vitamin D deficiency, including increased levels of PTH and alkaline phosphatase and lower levels of ionized calcium. In women living in northern latitudes, it has been shown that vitamin D levels decline during the winter months. This is associated with a striking seasonal bone loss, reflecting increased bone turnover. Treatment with vitamin D and calcium supplementation prevents this seasonal effect on bone metabolism. Reduced fracture rates have also been documented among individuals in northern latitudes who have greater vitamin D intake and have higher 25-hydroxyvitamin D [25(OH)D] levels (see below).

Estrogen Status Estrogen deficiency probably causes bone loss by two distinct but interrelated mechanisms: (1) activation of new bone remodeling sites, and (2) exaggeration of the imbalance between bone formation and resorption. The change in activation frequency causes a transient bone loss until a new steady state between resorption and formation is achieved. The remodeling imbalance, however, results in a permanent decrement in mass that can only be corrected by a remodeling event during which bone formation exceeds resorption. In addition, the very presence of more remodeling sites in the skeleton increases the probability that trabeculae will be penetrated, thereby eliminating the template upon which new bone can be formed and accelerating the loss of bony tissue.

The most frequent estrogen-deficient state is the cessation of ovarian function at the time of menopause, which occurs on average at the age of 51. Thus, with current life expectancy, an average woman will spend about 30 years without ovarian supply of estrogen. The mechanism by which estrogen deficiency causes bone loss is summarized inFig. 342-5. Marrow cells (macrophages, monocytes, osteoclast precursors, mast cells) as well as bone cells (osteoblasts, osteocytes, osteoclasts) expressERsa and b. The net effect of estrogen deficiency is increased osteoclast recruitment and perhaps activity. Estrogen may also play an important role in determining the life span of bone cells by controlling the rate of apoptosis. Thus, in situations of estrogen deprivation, the life span of osteoblasts may be decreased whereas the longevity of osteoclasts is increased.

Since remodeling is initiated at the surface of bone, it follows that trabecular bone -- which has a considerably larger surface area (80% of the total) than cortical bone -- will be preferentially affected by estrogen deficiency. Fractures occur earliest at sites where trabecular bone contributes most to bone strength; consequently, vertebral fractures are the most common early consequence of estrogen deficiency.

Physical Activity Inactivity, such as prolonged bed rest or paralysis, results in significant bone loss. Concordantly, athletes have higher bone mass than the general population. These changes in skeletal mass are most marked when the stimulus begins during growth and before the age of puberty. Adults are less capable than children of increasing bone mass following restoration of physical activity. Epidemiologic data support the beneficial effects on the skeleton of chronic high levels of physical activity. Fracture risk is lower in rural communities and in countries where physical activity is maintained into old age. However, when exercise is initiated during adult life, the effects

of moderate exercise are modest, with a bone mass increase of 1 to 2%. It is argued that more active individuals are less likely to fall and are more capable of protecting themselves upon falling, thereby reducing fracture risk.

Chronic Disease Various genetic and acquired diseases are associated with an increase in the risk of osteoporosis ([Table 342-2](#)). Mechanisms that contribute to bone loss are unique for each disease and typically result from multiple factors including nutrition, reduced physical activity levels, and factors that affect bone-remodeling rates.

Medications A large number of medications used in clinical practice have potentially detrimental effects on the skeleton ([Table 342-3](#)). *Glucocorticoids* are a common cause of medication-induced osteoporosis. It is often not possible to determine the extent to which osteoporosis is related to the glucocorticoid or to other factors, as treatment is superimposed on the effects of the primary disease, which may itself be associated with bone loss (e.g., rheumatoid arthritis). Excessive doses of thyroid hormone can accelerate bone remodeling and result in bone loss.

Other medications have less detrimental effects upon the skeleton than pharmacologic doses of glucocorticoids. *Anticonvulsants* are thought to increase the risk of osteoporosis, although many affected individuals have concomitant vitamin D insufficiency, as anticonvulsants that induce the cytochrome P450 system alter vitamin D metabolism. Patients undergoing transplantation are at high risk for rapid bone loss and fracture not only from glucocorticoids but also from treatment with other *immunosuppressants*, such as cyclosporine and tacrolimus (FK506). In addition, these patients often have underlying metabolic abnormalities, such as hepatic or renal failure, that predispose to osteopenia.

Cigarette Consumption The use of cigarettes over a long period has detrimental effects on bone mass. These effects may be mediated directly, by toxic effects on osteoblasts, or indirectly by modifying estrogen metabolism. On average, cigarette smokers reach menopause 1 to 2 years earlier than the general population. Cigarette smoking also has secondary effects on bone growth, such as illness, frailty, decreased exercise, and the need for additional medications (e.g., glucocorticoids for lung disease).

MEASUREMENT OF BONE MASS

Several noninvasive techniques are now available for estimating skeletal mass or density. These include dual-energy x-ray absorptiometry (DXA), single-energy xray absorptiometry (SXA), quantitative computed tomography (CT), and ultrasound.

[DXA](#)is a highly accurate x-ray technique that has become the standard for measuring bone density in most centers. Though it can be used for measurements of any skeletal site, clinical determinations are usually made of the lumbar spine and hip. Portable DXA machines have been developed that measure the heel (calcaneus), forearm (radius and ulna), or finger (phalanges), and DXA can also be used to measure body composition. In the DXA technique, two x-ray energies are used to estimate the area of mineralized tissue, and the mineral content is divided by the area, which partially corrects for body size. However, this correction is only partial since DXA is a two-dimensional scanning

technique and cannot estimate the depths or posteroanterior length of the bone. Thus, small people tend to have lower-than-average bone mineral density (BMD). Bone spurs, which are frequent in osteoarthritis, tend to falsely increase bone density of the spine. Because DXA instrumentation is provided by several different manufacturers, the output varies in absolute terms. Consequently, it has become standard practice to relate the results to "normal" values using T-scores, which compare individual results to those in a young population that is matched for race and gender. Alternatively, Z-scores compare individual results to those of an age-matched population that is also matched for race and gender. Thus, a 60-year-old woman with a Z-score of -1 (1 SD below mean for age) could have a T-score of -2.5 (2.5 SD below mean for a young control group) (Fig. 342-6).

CTis used primarily to measure the spine, and peripheral CT is used to measure bone in the forearm or tibia. Research into the use of CT for measurement of the hip is ongoing. The results obtained from CT are different from all others currently available since this technique specifically analyzes trabecular bone and can provide a true density (mass of bone per unit volume) measurement. However, CT remains expensive, involves greater radiation exposure, and is less reproducible.

Ultrasound is used to measure bone mass by calculating the attenuation of the signal as it passes through bone or the speed with which it traverses the bone. It is unclear whether ultrasound assesses bone quality, but this may be an advantage of the technique. Because of its relatively low cost and mobility, ultrasound is amenable for use as a screening procedure.

All of these techniques for measuringBMD have been approved by the U.S. Food and Drug Administration (FDA) based upon their capacity to predict fracture risk. The hip is the preferred site of measurement in most individuals, since it directly assesses bone mass at an important fracture site. When hip measurements are performed byDXA, the spine can be measured at the same time. In younger individuals, such as perimenopausal women, spine measurements may be the most sensitive indicator of bone loss.

When to Measure Bone Mass Clinical guidelines developed by the National Osteoporosis Foundation recommend bone mass measurements in postmenopausal women, assuming they have risk factors for osteoporosis in addition to age, gender, and estrogen deficiency. The guidelines further recommend that bone mass measurement be considered in *all* women by age 60 to 65. Criteria approved for Medicare reimbursement of BMDare summarized in Table 342-4.

When to Treat Based Upon Bone Mass Results The guidelines developed by the National Osteoporosis Foundation suggest that patients be considered for treatment whenBMD> 2.5 SD below the mean value for young adults (T-score £ -2.5). Treatment should also be considered in women with risk factors in addition to menopause, if measurement of BMD of the hip gives a T score< -2.0. Because the fracture risk increases continuously as T-scores decline, there is no critical threshold and treatment decisions must be individualized. Cost-benefit analyses in this area are changing rapidly because of the availability of new drugs [e.g., bisphosphonates, selective estrogen receptor modulators (SERMs)] and the results of trials [e.g., the Heart and

Estrogen-Progestin Replacement Study (HERS)] examining the long-term cardiovascular effects of hormone replacement therapy (HRT).

Approach to the Patient

The perimenopausal transition is a good opportunity to initiate discussion about risk factors for osteoporosis and to consider indications for a BMD test. A careful history and physical examination should be performed to identify risk factors for osteoporosis. As noted above, a low Z-score increases the suspicion of a secondary disease. Height loss >2.5 to 3.8 cm (1 to 1.5 in.) is an indication for radiography to rule out asymptomatic vertebral fractures, as is the presence of significant kyphosis or back pain, particularly if it began after menopause. For patients who present with fractures, it is important to ensure that the fractures are truly due to trauma or osteoporosis and not to secondary underlying malignancy. Usually this is clear on routine radiography, but on occasion, CT, magnetic resonance imaging, or radionuclide scans may be helpful. Severe unremitting back pain also raises the suspicion of other causes such as malignancy (especially myeloma).

Routine Laboratory Evaluation There is no established algorithm for the evaluation of women presenting with osteoporosis. A general evaluation that includes complete blood count, serum calcium, and perhaps urine calcium is helpful for identifying selected secondary causes of low bone mass, particularly for women with fractures or very low Z-scores. An elevated serum calcium level suggests hyperparathyroidism or malignancy, whereas a reduced serum calcium level may reflect malnutrition and osteomalacia. In the presence of hypercalcemia, a serum PTH level differentiates between hyperparathyroidism (PTH-) and malignancy (PTH⁻), and a high PTHrP level can help document the presence of humoral hypercalcemia of malignancy (Chap. 341). A low urine calcium (<50 mg/24 h) suggests osteomalacia, malnutrition, or malabsorption; a high urine calcium (>300 mg/24 h) is indicative of hypercalciuria and must be investigated further. Hypercalciuria occurs primarily in three situations: (1) a renal calcium leak, which is more frequent in males with osteoporosis; (2) absorptive hypercalciuria, which can be idiopathic or associated with increased $1,25(OH)_2D$ in granulomatous disease; or (3) hematologic malignancies or conditions associated with excessive bone turnover such as Paget's disease, hyperparathyroidism, and hyperthyroidism.

When there is clinical suspicion of hyperthyroidism or Cushing's syndrome, thyroid stimulating hormone (TSH) or urinary free cortisol levels should be measured. When bowel disease, malabsorption, or malnutrition is suspected, serum albumin, cholesterol, and a complete blood count should be checked. Asymptomatic malabsorption might be suspected if there is anemia (macrocytic -- vitamin B_{12} or folate deficiency; or microcytic -- iron deficiency), or low serum cholesterol or urinary calcium levels. If these or other features suggest malabsorption, further evaluation is required. Asymptomatic celiac sprue with selective malabsorption is not uncommon; the diagnosis requires antigliadin and antiendomysial antibody tests and often a small-bowel biopsy. A trial of a gluten-free diet may be confirmatory (Chap. 286).

Myeloma can masquerade as generalized osteoporosis, although it more commonly presents with bone pain and characteristic "punched-out" lesions on radiography.

Serum and urine electrophoresis and evaluation for light chains in urine are required to exclude this diagnosis. A bone marrow biopsy may be required to rule out myeloma (in patients with equivocal electrophoretic results) and can also be used to exclude mastocytosis, leukemia, and other marrow infiltrative disorders, such as Gaucher's disease.

Bone Biopsy Although the use of bone biopsy is rarely required today, it remains an important tool in clinical research. Tetracycline labeling of the skeleton allows determination of the rate of remodeling as well as evaluation for other metabolic bone diseases. The current use of BMD tests, in combination with hormonal evaluation and biochemical markers of bone remodeling, has largely replaced bone biopsy.

Biochemical Markers Several biochemical tests are now available that provide an index of the overall rate of bone remodeling (Table 342-5). Biochemical markers are usually characterized as those related primarily to *bone formation* or *bone resorption*. These tests measure the overall state of bone remodeling at a single point in time. Clinical use of these tests has been hampered by biologic variability (in part related to circadian rhythm) as well as to analytical variability.

For the most part, remodeling markers do not predict rates of bone loss well enough to use this information clinically. However, markers of bone resorption may help in the prediction of fracture risk, particularly in older individuals. In women ³65 years, when bone density results are greater than the usual treatment thresholds noted above, a high level of bone resorption should prompt consideration of treatment. The primary use of biochemical markers is for monitoring the response to treatment. With the introduction of antiresorptive therapeutic agents, bone remodeling declines rapidly, with the fall in resorption occurring earlier than the fall in formation. Inhibition of bone resorption is maximal within 3 to 6 months. Thus, measurement of bone resorption prior to initiating therapy and 4 to 6 months after starting therapy provides an earlier estimate of patient response than does bone densitometry. A decline in resorptive markers can be ascertained after treatment with bisphosphonates and HRT; this effect is less marked after treatment with either raloxifene or intranasal calcitonin. A biochemical marker response to therapy is particularly useful for asymptomatic patients and helps to ensure long-term compliance. When agents that stimulate bone formation become available, bone remodeling markers may be useful to help select therapy. However, since all current therapeutic approaches reduce bone turnover, this strategy currently has little value.

TREATMENT

Management of Osteoporotic Fractures Treatment of the patient with osteoporosis frequently involves management of acute fractures as well as treatment of the underlying disease. Hip fractures almost always require surgical repair if the patient is to become ambulatory again. Depending on the location and severity of the fracture, condition of the neighboring joint, and general status of the patient, procedures may include open reduction and internal fixation with pins and plates, hemiarthroplasties, and total arthroplasties. These surgical procedures are followed by intense rehabilitation in an attempt to return patients to their prefracture functional level. Long bone fractures often require either external or internal fixation. Other fractures (e.g., vertebral, rib, and

pelvic fractures) are usually managed with only supportive care, requiring no specific orthopedic treatment.

Only ~25 to 30% of vertebral compression fractures present with sudden-onset back pain. For acutely symptomatic fractures, treatment with analgesics is required, including nonsteroidal anti-inflammatory agents and/or acetaminophen, sometimes with the addition of a narcotic agent (codeine or oxycodone). A few small, randomized clinical trials have demonstrated that calcitonin may reduce pain related to acute vertebral compression fracture. A recently developed, but still experimental, technique involves percutaneous injection of artificial cement (polymethylmethacrylate) into the vertebral body (vertebroplasty or kyphoplasty); this has been reported to offer significant immediate pain relief in the majority of patients. Short periods of bed rest may be helpful for pain management, but, in general, early mobilization is recommended as it helps prevent further bone loss associated with immobilization. Occasionally, use of a soft elastic-style brace may facilitate earlier mobilization. Muscle spasms often occur with acute compression fractures and can be treated with muscle relaxants and heat treatments.

Severe pain usually resolves within 6 to 10 weeks. Chronic pain is probably not bony in origin; instead, it is related to abnormal strain on muscles, ligaments, and tendons and to secondary facet-joint arthritis associated with alterations in thoracic and/or abdominal shape. Chronic pain is difficult to treat effectively and may require analgesics, sometimes including narcotic analgesics. Frequent intermittent rest in a supine or semireclining position is often required to allow the soft tissues, which are under tension, to relax. Back-strengthening exercises (paraspinal) may be beneficial. Heat treatments help relax muscles and reduce the muscular component of discomfort. Various physical modalities, such as ultrasound and transcutaneous nerve stimulation, may be beneficial in some patients. Pain also occurs in the neck region, not as a result of compression fractures (which almost never occur in the cervical spine as a result of osteoporosis) but because of chronic strain associated from trying to elevate the head in a person with a severe thoracic kyphosis.

Multiple vertebral fractures are often associated with psychological symptoms, not always commonly appreciated. The changes in body configuration and back pain can lead to marked loss of self-image and a secondary depression. Altered balance, precipitated by the kyphosis and the anterior movement of the body's center of gravity, leads to a fear of falling, a consequent tendency to remain indoors, and the onset of social isolation. These symptoms can sometimes be alleviated by family support and/or psychotherapy. Medication may be necessary when depressive features are present.

Management of the Underlying Disease

Risk Factor Reduction Patients should be thoroughly educated to reduce the likelihood of any risk factors associated with bone loss and falling. Medications should be reviewed to ensure that any glucocorticoid medication is truly indicated and is being given in doses as low as possible. For those on thyroid hormone replacement, TSH testing should be performed to ensure that an adequate, but not excessive, dose is being used, as excess can be associated with increased bone loss. In patients who smoke, efforts should be made to facilitate smoking cessation. Reducing

risk factors for falling also includes alcohol abuse treatment and a review of the medical regimen for any drugs that might be associated with orthostatic hypotension and/or sedation, including hypnotics and anxiolytics. If nocturia occurs, the frequency should be reduced, if possible (e.g., by decreasing or modifying diuretic use), as arising in the middle of sleep is a common precipitant of a fall. Patients should be instructed about environmental safety with regard to eliminating exposed wires, curtain strings, slippery rugs, and mobile tables. Avoiding stocking feet on wood floors, checking carpet condition (particularly on stairs), and providing good light in paths to bathrooms and outside the home are good preventive measures. Treatment for impaired vision is recommended, particularly a problem with depth perception, which is specifically associated with increased falling risk. Elderly patients with neurologic impairment (e.g., stroke, Parkinson's disease, Alzheimer's disease) are particularly at risk of falling and require specialized supervision and care.

Nutritional Recommendations

CALCIUM A large body of data indicates that optimal calcium intake reduces bone loss and suppresses bone turnover. Recommended intakes from a recent report from the Institute of Medicine are shown in Table 342-6. The National Health and Nutritional Evaluation Studies (NHANES) have consistently documented that average calcium intakes fall considerably short of these recommendations. The preferred source of calcium is from dairy products and other foods, but many patients require additional calcium supplementation. Food sources of calcium are dairy products (milk, yogurt, and cheese) and fortified foods such as certain cereals, waffles, snacks, juices, and crackers. Some of these fortified foods contain as much calcium per serving as milk.

If a calcium supplement is required, it should be taken in doses£600 mg at a time, as the calcium absorption fraction decreases at higher doses. Calcium supplements should be calculated based on the elemental calcium content of the supplement, not the weight of the calcium salt (Table 342-7). Calcium supplements containing carbonate are best taken with food since they require acid for solubility. Calcium citrate supplements can be taken at any time.

Several controlled clinical trials of calcium plus vitamin D have confirmed reductions in clinical fractures, including fractures of the hip (~20 to 30% risk reduction). All recent studies of pharmacologic agents have been conducted in the context of calcium replacement (± vitamin D). Thus, it is standard practice to ensure an adequate calcium and vitamin D intake in patients with osteoporosis, whether they are receiving additional pharmacologic therapy or not.

Although side effects from supplemental calcium are minimal, individuals with a history of kidney stones should have a 24-h urine calcium determination before starting increased calcium to avoid hypercalciuria. Furthermore, a thiazide-containing diuretic might be indicated in some patients to increase renal tubular calcium reabsorption and to reduce urine calcium levels.

VITAMIN D Vitamin D is synthesized in skin under the influence of heat and ultraviolet light (Chap. 340). However, large segments of the population do not obtain sufficient vitamin D to maintain what is now considered an adequate supply [serum 25(OH)D

consistently >15 to 20 ng/mL]. Since vitamin D supplementation at doses that would achieve these serum levels is safe and inexpensive, it is now routine to recommend supplemental vitamin D. The Institute of Medicine recommends daily intakes of 200 IU for adults <50 years of age, 400 IU for those from 50 to 70 years, and 600 IU for those >70 years. Multivitamin tablets usually contain 400 IU, and many calcium supplements also contain vitamin D.

OTHER NUTRIENTS Other nutrients such as salt and caffeine may have modest effects on calcium excretion or absorption. Adequate vitamin K states is required for optimal carboxylation of osteocalcin. States in which vitamin K nutrition or metabolism is impaired, such as with long-term coumadin therapy, have been associated with reduced bone mass.

Magnesium is abundant in foods, and magnesium deficiency is quite rare in the absence of a serious chronic disease. Magnesium supplementation may be warranted in patients with inflammatory bowel disease, celiac sprue, chemotherapy, severe diarrhea, malnutrition, or alcoholism. Phytoestrogens may impact skeletal health, although the degree of this effect is unclear. Dietary phytoestrogens, which are derived primarily from soy products and legumes (e.g., garbanzo beans, chickpeas, and lentils), are insufficiently potent to justify their use in place of a pharmacologic agent in the treatment of osteoporosis.

Patients with hip fracture are often frail and relatively malnourished. Some data suggest an improved outcome in such patients when they are provided calorie and protein supplementation.

Exercise Exercise in young individuals increases the likelihood that they will attain the maximal genetically determined peak bone mass. Meta-analyses of studies performed in postmenopausal women indicate that weight-bearing exercise prevents bone loss but does not appear to result in substantial bone gain. When the exercise is discontinued, any effects on bone mass wane. It is important to note, however, that exercise also has beneficial effects on neuromuscular function. Exercise can improve coordination, balance, and strength and thereby reduce the risk of falling, as well as the severity of injury upon a fall. Therefore, the beneficial effects of exercise on muscle mass and reduced risk of falling justify its recommendation for all age groups. A walking program is a practical way to start. Other activities such as dancing, racquet sports, cross-country skiing, and use of gym equipment are also recommended, depending on the patient's personal preference. Even women who cannot walk benefit from swimming or water exercises, not so much for the effects on bone, which are quite minimal, but because of effects on muscle. Exercise habits should be consistent, optimally at least three times a week.

Pharmacologic Therapies Until fairly recently, estrogen treatment, either by itself or in concert with a progestin, was the primary therapeutic agent for prevention or treatment of osteoporosis. Over the past 5 years, a number of new drugs have appeared, and more are expected in the near future. Some are agents that specifically treat osteoporosis (bisphosphonates, calcitonin); others, such as tissue-selective estrogens (or SERMs), have broader effects. The availability of these drugs allows therapy to be tailored to the needs of an individual patient. The evidence supporting the effectiveness

of each remedy is variable, in part because these treatments are new.

Estrogens A large body of clinical trial data indicates that various types of estrogens (conjugated equine estrogens, estradiol, estrone, esterified estrogens, ethinyl estradiol, and mestranol) reduce bone turnover, prevent bone loss, and induce small increases in bone mass of the spine, hip, and total body. The effects of estrogen are seen in women with natural or surgical menopause and in late postmenopausal women with or without established osteoporosis. Estrogens are efficacious when administered orally, buccally, vaginally, percutaneously, subcutaneously, and transdermally. For both oral and transdermal routes of administration, combined estrogen/progestin preparations are now available in many countries, obviating the problem of taking two tablets or using a patch and oral progestin. One large study, referred to as PEPI (Postmenopausal Estrogen/ Progestin Intervention Trial), indicated that C-21 progestins alone do not augment the effect of estrogen on bone mass ([Fig. 342-7](#)).

DOSE OF ESTROGEN For oral estrogens, the recommended dose is 0.3 mg/d for esterified estrogens, 0.625 mg/d for conjugated equine estrogens, and 5 ug/d for ethinyl estradiol. For transdermal estrogen, the commonly used dose supplies 50 ug estradiol per day, but a lower dose may be appropriate for some individuals. Dose-response data are not available for other routes of administration.

FRACTURE DATA In contrast to the body of clinical trial data evaluating the effects of estrogen on bone mass, its effects on fracture occurrence have been less well studied. Epidemiologic databases indicate that women who take estrogen replacement have a 50% reduction, on average, of osteoporotic fractures, including hip fractures. The beneficial effect of estrogen is greatest among those who start replacement early and continue the treatment; the benefit wanes after discontinuation such that there is no residual protective effect against fracture by 10 years after discontinuation. There are no clinical trial data confirming that estrogen administration reduces the risk of hip fracture. In fact, the[HERS](#)trial of women with established coronary artery disease showed no reduction in the risk of hip or clinical fractures in the estrogen-progestin arm relative to the placebo group. These data may not be definitive, however, since the women were not chosen for osteoporosis risk and were at unknown risk of osteoporotic fracture. Furthermore, radiographic vertebral fractures were not assessed in this study. One clinical study which looked at all nonvertebral fractures suggested a reduction in HRT-treated women.

A few clinical trials have evaluated spine fracture occurrence as an outcome with estrogen therapy. One that used high doses of estrogen (2.5 mg conjugated equine estrogen per day) indicated marked vertebral fracture reduction in estrogen-treated women. Several other small studies, using lower estrogen doses, have consistently shown that estrogen treatment reduces the incidence of vertebral compression fracture. The ongoing Women's Health Initiative will provide additional data about the effects of estrogen on the risk of other osteoporosis-related fractures.

Long-term estrogen use may be associated with an increase in the risk of venous thromboembolism and gallbladder, uterine, and breast cancer; in observational studies, estrogens have been associated with a significant reduction in myocardial infarction, although this was not so in HERS. The WHI will provide further information.

MODE OF ACTION Two subtypes of ERs, a and b, have been identified in bone and other tissues. Cells of monocyte lineage express both ERa and -b, as do osteoblasts. Estrogen-mediated effects vary depending on the receptor type. Using ER knockout mouse models, elimination of ERa produces a modest reduction in bone mass, whereas ERb null animals had very little abnormality, except greater cortical bone mass. A male patient with a homozygous mutation of ERa had markedly decreased bone density as well as abnormalities in epiphyseal closure, confirming the important role of ERa in bone biology. The mechanism of estrogen action in bone is an area of active investigation (Fig. 342-5). Though data are conflicting, estrogens appear to inhibit osteoclasts directly. However, the majority of estrogen (and androgen) effects on bone resorption are mediated indirectly through paracrine factors produced by osteoblasts. These actions include: (1) increasing IGF-1 and TGF-b, and (2) suppressing IL-1 (a and b), IL-6, TNF-a, and osteocalcin synthesis. The consequence of these effects is primarily to decrease bone resorption.

Progestins In women with a uterus, daily progestin or cyclical progestins at least 12 days per month are prescribed in combination with estrogens to reduce the risk of uterine cancer. Medroxyprogesterone acetate and norethindrone acetate blunt the high-density lipoprotein response to estrogen, but micronized progesterone does not. Neither medroxyprogesterone acetate nor micronized progesterone appears to have an independent effect on bone; at lower doses, norethindrone acetate might have an additive benefit. On breast tissue, progestins may increase the risk of breast cancer, though this is by no means definite.

Tissue-Selective Estrogens, or SERMs Two SERMs are currently being used in postmenopausal women: raloxifene, which is approved for prevention and treatment of osteoporosis, and tamoxifen, which is approved for the prevention and treatment of breast cancer.

Tamoxifen reduces bone turnover and bone loss in postmenopausal women compared to placebo groups. These findings support the concept that tamoxifen acts as an estrogenic agent in bone. There are limited data on the effect of tamoxifen on fracture risk, but the Breast Cancer Prevention study indicated a possible reduction in clinical vertebral, hip, and Colles' fractures. The major benefit of tamoxifen is on breast cancer occurrence. The breast cancer prevention trial indicated that tamoxifen administration over 4 to 5 years reduced the incidence of new invasive and noninvasive breast cancer by approximately 45% in women at increased risk of breast cancer. The incidence of ER-positive breast cancers was reduced by 65%.

Raloxifene (60 mg/d) has effects on bone turnover and bone mass that are very similar to those of tamoxifen, indicating that this agent is also estrogenic on the skeleton. The effect of raloxifene on bone density (+1.4 to 2.8% versus placebo in the spine, hip, and total body) is somewhat less than that seen with standard doses of estrogens. Raloxifene reduces the occurrence of vertebral fracture by 30 to 50%, depending on the subpopulation.

Raloxifene, like tamoxifen and estrogen, has effects throughout other organ systems. The most positive effect appears to be a reduction in invasive breast cancer (mainly

decreasedER-positive) occurrence of about 70% in women who take raloxifene compared to placebo. In contrast to tamoxifen, raloxifene is not associated with an increase in the risk of uterine cancer or benign uterine disease. Raloxifene increases the occurrence of hot flashes. Although raloxifene reduces serum total and low-density lipoprotein cholesterol, lipoprotein(a), and fibrinogen, no studies including cardiovascular disease or cerebrovascular disease endpoints are available.

MODE OF ACTION OF SERMS AllSERMsbind to theER, but each agent produces a unique receptor conformation. As a result, specific coactivator or corepressor proteins are bound to the receptor (Chap. 327), resulting in differential effects on gene transcription that vary according to other transcription factors present in the cell. Another aspect of selectivity is the affinity of each SERM for the different ERa and bsubtypes, which are expressed differentially in various tissues. These tissue-selective effects of SERMs offer the possibility of tailoring estrogen therapy to best meet the needs and risk factor profile of an individual patient.

Bisphosphonates Both alendronate and risendronate are approved for the prevention and treatment of postmenopausal osteoporosis and treatment of steroid-induced osteoporosis. Risedronate is also approved for the prevention of steroid-induced osteoporosis.

Alendronate has been shown to have dramatic effects in patients with osteoporosis, decreasing bone turnover and increasing bone mass in the spine by up to 8% versus placebo and by 6% versus placebo in the hip (Fig. 342-8). Multiple trials have evaluated the effect of alendronate on fracture occurrence. The Fracture Intervention Trial provided evidence in over 2000 women with prevalent vertebral fractures that daily alendronate treatment (5 mg/d for 2 years and 10 mg/d for 9 months afterwards) reduces vertebral fracture risk by about 50%, multiple vertebral fractures by up to 90%, and hip fractures by up to 50% (Fig. 342-9). Several subsequent trials have confirmed these findings. For example, in a study of >1900 women with low bone mass treated with alendronate (10 mg/d) versus placebo, the incidence of all nonvertebral fractures was reduced by ~47% after just 1 year.

Alendronate (5 to 10 mg/d) should be given with a full glass of water before breakfast, as bisphosphonates are poorly absorbed. Because of the potential for esophageal irritation, alendronate is contraindicated in patients who have stricture or inadequate emptying of the esophagus. It is recommended that patients remain upright for at least 30 min after taking the medication to avoid esophageal irritation. Cases of esophagitis, esophageal ulcer, and esophageal stricture have been described, but the incidence appears to be low. In clinical trials, overall gastrointestinal symptomatology was no different with alendronate compared to placebo.

Risedronate produces a dramatic reduction in bone turnover and an increase in bone mass. Controlled clinical trials have demonstrated >40% reduction in vertebral fracture risk over 3 years, accompanied by a 33% reduction in clinical nonspine fractures. Reports from several studies show a 40% reduction in hip fracture in patients with osteoporosis, with a somewhat greater effect in patients with prevalent vertebral fractures. Patients should take risedronate (5.0 mg orally) with a full glass of plain water [0.18 to 0.25 L (6 to 8 oz)], to facilitate delivery to the stomach, and should not lie down

for 30 min after taking the drug. The incidence of gastrointestinal side effects in these trials with risedronate was similar to that of placebo.

Etidronate was the first bisphosphonate to be approved, initially for use in Paget's disease and hypercalcemia. This agent has also been used in osteoporosis trials of smaller magnitude than those performed for alendronate and risedronate. Etidronate probably has some efficacy against vertebral fracture when given as an intermittent cyclical regimen (2 weeks on, 2 1/2 months off).

MODE OF ACTION Bisphosphonates are structurally related to pyrophosphates, compounds that are incorporated into bone matrix. Through mechanisms that remain to be fully elucidated, bisphosphonates specifically impair osteoclast function and reduce osteoclast number, in part by the induction of apoptosis. Recent evidence suggests that the nitrogen-containing bisphosphonates also inhibit protein prenylation, one of the end products in the mevalonic acid pathway. This effect disrupts intracellular protein trafficking and may ultimately lead to apoptosis. Bisphosphonates have very long retention in the skeleton and may exert long-term effects.

Calcitonin Calcitonin is a polypeptide hormone produced by the thyroid gland (Chap. 341). Its physiologic role is unclear as no skeletal disease has been described in association with calcitonin deficiency or calcitonin excess. Calcitonins are approved by the FDA approved for Paget's disease, hypercalcemia, and osteoporosis in women>5 years past menopause.

Injectable calcitonin produces small increments in bone mass of the lumbar spine. However, difficulty of administration and frequent reactions, including nausea and facial flushing, make general use limited. In 1995, a nasal spray containing calcitonin (200 IU/d) was approved for treatment of osteoporosis in postmenopausal women. Several studies indicate that nasal calcitonin produces small increments in bone mass and a small reduction in new vertebral fractures in calcitonin-treated patients versus those on calcium alone.

Calcitonin is not indicated for prevention of osteoporosis and is not sufficiently potent to prevent bone loss in early postmenopausal women. As mentioned above, calcitonin might have an analgesic effect on bone pain, both in the subcutaneous and possibly the nasal form.

MODE OF ACTION Calcitonin suppresses osteoclast activity by direct action on the osteoclast calcitonin receptor. Osteoclasts exposed to calcitonin cannot maintain their active ruffled border, which normally maintains close contact with underlying bone. Calcitonin also affects osteoclast mobility and the movement of enzyme-containing cytoplasmic granules.

Experimental Agents

PARATHYROID HORMONE Endogenous PTH is an 84-amino-acid peptide that is largely responsible for calcium homeostasis (Chap. 341). Although chronic elevation of PTH, as occurs in hyperparathyroidism, is associated with bone loss (particularly cortical bone), PTH can also exert anabolic effects on bone. Consistent with this, some

observational studies have indicated that mild elevations in PTH are associated with maintenance of trabecular bone mass. On the basis of these findings, preclinical and early clinical studies have been performed using an exogenous PTH analogue (1-34 PTH). The first randomized controlled trial in postmenopausal women showed that PTH, when superimposed on ongoing estrogen therapy, produced substantial increments in bone mass (13% over a 3-year period compared to estrogen alone). This increment in bone mass was also associated with a reduction in risk of vertebral compression deformity (Fig. 342-10). More recent studies have confirmed the ability of combined treatment with estrogen and PTH to induce striking increases in bone mass.

PTHuse may be limited by its mode of administration, which currently requires subcutaneous injection. Alternative modes of delivery are being investigated, including transdermal and inhalation routes. The optimal frequency of administration also remains to be established, and it is possible that PTH might also be effective when used in high doses, 1 month out of every 3.

MODE OF ACTION Exogenously administeredPTHappears to have direct actions on osteoblast activity, with biochemical and histomorphometric evidence of de novo bone formation early in response to PTH, prior to activation of bone resorption. Subsequently, PTH activates bone remodeling but still appears to favor bone formation over bone resorption. PTH stimulatesIGF-I and collagen production and appears to increase osteoblast number by inhibiting apoptosis and stimulating replication.

FLUORIDE Fluoride has been available for many years and is a potent stimulator of osteoprogenitor cells when studied in vitro. It has been used in multiple osteoporosis studies with conflicting results, in part related to use of varying doses and preparations. Fluoride produces marked effects on bone mass, especially in the spine, where gains of around 10% per year have been observed. However, despite increments in bone mass, there is no consistent effect of fluoride on vertebral or nonvertebral fracture, which might actually increase when high doses of fluoride are used. Furthermore, animal data suggest that there is reduced biomechanical strength when fluoride is incorporated into bone as fluoroapatite, with excess osteoid accumulation and evidence of woven rather than lamellar bone formation, especially at high doses. For these reasons, fluoride remains an experimental agent, despite its long history and multiple studies.

OTHER POTENTIAL ANABOLIC AGENTS Several small studies of growth hormone (GH), alone or in combination with other agents, have not shown consistent or substantial positive effects on skeletal mass. Many of these studies are relatively short-term, and the effects of GH and the IGFsare still under investigation. Anabolic steroids, mostly derivatives of testosterone, act primarily as antiresorptive agents to reduce bone turnover but may also stimulate osteoblastic activity. Effects on bone mass remain unclear but appear weak, in general, and use is limited by masculinizing side effects. Several recent observational studies suggest that the statin drugs, currently used to treat hypercholesterolemia, may be associated with increased bone mass and reduced fractures, but there are not clinical trial data.

Nonpharmacologic Approaches Protective pads worn around the outer thigh, which cover the trochanteric region of the hip can prevent hip fractures in elderly residents in nursing homes. The use of hip protectors is limited largely by compliance and comfort,

but new devices are being developed that may circumvent these problems and provide adjunctive treatments.

Treatment Monitoring There are currently no well-accepted guidelines for monitoring treatment of osteoporosis. Because most osteoporosis treatments produce small or moderate bone mass increments on average, it is reasonable to consider BMD as a monitoring tool. As with any biologic or assay determination, there is precision error with repeated measurements. Changes must exceed ~4% in the spine and 6% in the hip to be considered significant in any individual. The hip is the preferred site due to larger surface area and greater reproducibility. Medication-induced increments may require several years to produce changes of this magnitude (if they do at all). Consequently, it can be argued that BMD should not be repeated at intervals <2 years. Only significant BMD reductions should prompt a change in medical regimen, as it is expected that many individuals will not show responses greater than the detection limits of the current measurement techniques.

Biochemical markers of bone turnover may prove useful for treatment monitoring, but there is currently little hard evidence to support this concept; it remains unclear which endpoint is most useful. If bone turnover markers are used, a determination should be made before starting therapy and repeated ³4 months after therapy is initiated. In general, a change in bone turnover markers must be 30 to 40% lower than the baseline to be significant because of the biologic and technical variability in these tests. A positive change in biochemical markers and/or bone density can be useful to help patients adhere to treatment regimens.

GLUCOCORTICOID-INDUCED OSTEOPOROSIS

Osteoporotic fractures are a well-characterized consequence of the hypercortisolism associated with Cushing's syndrome. However, the therapeutic use of glucocorticoids is by far the most common form of glucocorticoid-induced osteoporosis. Glucocorticoids are widely used in the treatment of a variety of disorders, including chronic lung disorders, rheumatoid arthritis and other connective tissue diseases, inflammatory bowel disease, and posttransplantation. Osteoporosis and related fractures are serious side effects of chronic glucocorticoid therapy. Because the effects of glucocorticoids on the skeleton are often superimposed upon the consequences of aging and menopause, it is not surprising that women and the elderly are most frequently affected. The skeletal response to steroids is remarkably heterogeneous, however, and even young, growing individuals treated with glucocorticoids can present with fractures.

The risk of fractures depends on the dose and duration of glucocorticoid therapy. Thus, cumulative dose is an important determinant of fracture risk. Bone loss is more rapid during the early months of treatment, and trabecular bone is more severely affected than cortical bone. Bone loss has been documented with the use of oral prednisone at doses that are generally considered to be less than replacement levels, and the lower threshold is not known. High-dose inhaled glucocorticoids can produce systemic effects on the skeleton, as can intraarticular injections. Alternate-day delivery does not appear to ameliorate the skeletal effects of glucocorticoids. The prevalence of vertebral fractures in asthmatic patients treated for 1 year with glucocorticoids is 11%, and increased risk of fractures has been demonstrated in most other disease states treated

with glucocorticoids.

Pathophysiology Glucocorticoids increase bone loss by multiple mechanisms including: (1) inhibition of osteoblast function and potential increase in osteoblast apoptosis, resulting in impaired synthesis of new bone; (2) stimulation of bone resorption, probably as a secondary effect; (3) impairment of the absorption of calcium across the intestine, probably by a vitamin D-independent effect; (4) increase of urinary calcium loss and induction of some degree of secondary hyperparathyroidism; (5) reduction of adrenal androgens and suppression of ovarian and testicular secretion of estrogens and androgens; and (6) potential induction of glucocorticoid myopathy, which may exacerbate effects on skeletal and calcium homeostasis, as well as increase the risk of falls.

Evaluation of the Patient Because of the prevalence of glucocorticoid-induced osteopenia, it is important to evaluate the status of the skeleton in all patients being initiated on or already receiving long-term glucocorticoid therapy. Modifiable risk factors should be identified, including those for falls. Examination should include height and muscle strength testing. Laboratory evaluation should include an assessment of 24-h urinary calcium. A task force of the American College of Rheumatology recommends that all patients who are being initiated on glucocorticoids and patients already on long-term (>6 months) glucocorticoids have measurement of bone mass at both the spine and hip using DXA. If only one skeletal site can be measured, it is best to assess the spine in individuals <60 years and the hip for those >60 years.

Prevention Bone loss caused by glucocorticoids can be prevented, and the risk of fractures significantly reduced. Strategies must include using the lowest dose of glucocorticoid for disease management. Topical and inhaled routes of administration are preferred, where appropriate. Risk factor reduction is important, including smoking cessation, limitation of alcohol consumption, and participation in weight-bearing exercise, where appropriate. All patients should receive an adequate calcium and vitamin D intake from the diet or from supplements.

TREATMENT

Only bisphosphonates have been demonstrated in large clinical trials to reduce the risk of fractures in patients being treated with glucocorticoids. Risedronate has been shown to prevent bone loss and to reduce vertebral fracture risk by about 70%. Similar beneficial effects are observed in studies of alendronate and etidronate. Controlled trials of HRT have shown bone-sparing effects, and calcitonin also has some protective effect. Thiazides reduce urine calcium loss, but their role in prevention of fractures is unclear.

(Bibliography omitted in Palm version)

Back to Table of Contents

343. PAGET'S DISEASE AND OTHER DYSPLASIAS OF BONE - *Stephen M. Krane, Alan L. Schiller*

PAGET'S DISEASE OF BONE

Paget's disease of bone (osteitis deformans) is characterized by excessive resorption of bone by osteoclasts, followed by the replacement of normal marrow by vascular, fibrous connective tissue. At some stage and to a variable degree, the resorbed bone is replaced by coarse-fibered, dense trabecular bone organized in haphazard fashion. The irregular and often rapid deposition of this new bone, to a great extent still lamellar, results in an increase in the number of prominent, irregular cement lines that give the bone its characteristic "mosaic" pattern. Most lesions show both excessive resorption and chaotic new bone formation. The disorder is usually focal but may be widespread.

INCIDENCE

The prevalence is difficult to determine because the disease is often asymptomatic. It is frequently detected when roentgenograms are obtained for other reasons or because of a high level of alkaline phosphatase on routine blood screening. On the basis of autopsy examination, the incidence is estimated to be about 3% in individuals over age 40; the likelihood of occurrence increases with age. The incidence varies in different parts of the world. Radiologic surveys indicate that the frequency in adults is <1% in the United States, Great Britain, and Australia. In India, Japan, the Middle East, and Scandinavia, the disease is rare.

ETIOLOGY

The cause is unknown. Some of the manifestations can be suppressed with glucocorticoids, salicylates, and cytotoxic drugs, but there is no convincing evidence that the fundamental lesion is inflammatory. Intranuclear inclusions have been found by electron microscopy in osteoclasts in pagetic bone. Some of the inclusions resemble nucleocapsids of viruses belonging to the measles group. Indirect immunofluorescence studies using antibodies to measles virus suggest that the inclusions are indeed measles virus nucleocapsids. The presence of mutations in specific regions of the viral genome is consistent with persistent infection. In some individuals with Paget's disease, osteoclasts and bone marrow mononuclear cells contain nucleocapsids of respiratory syncytial virus alone or in addition to nucleocapsids of measles virus; in some areas of Britain, canine distemper virus sequences have been identified in pagetic bone cells. Thus, different paramyxoviruses may have roles in the initiation or propagation of Paget's disease. Further evidence supporting the potential role of measles virus in the pathogenesis of the excessive bone resorption in Paget's disease has been obtained from studies of osteoclast precursors in vitro. Normal bone marrow-derived CD34+ cells transduced with a measles virus nucleocapsid gene differentiate into abnormal multinucleated osteoclasts that can resorb bone.

There is renewed interest in genetic factors that might be important in the predisposition to and pathogenesis of Paget's disease. Several large kindreds have been identified in which Paget's disease affects two or more generations with a pattern of inheritance consistent with autosomal dominant transmission. A rare Paget's disease-like disorder,

familial expansile osteolysis, has been mapped to chromosome 18q21-22; Paget's disease was mapped to the same locus in several families. However, in other families with Paget's disease, there is no linkage to 18q21-22, indicating genetic heterogeneity. In some sporadic osteosarcomas, there is constitutional loss of heterozygosity mapped to the same region of chromosome 18, and the rare sarcomas associated with Paget's disease (see below) also exhibit loss of heterozygosity in this region. It is of great interest that the gene within the 18q21-22 locus (*TNFRS-F11A* gene) responsible for familial expansile osteolysis encodes the receptor activator of NfkB (RANK), a member of the tumor necrosis factor (TNF) superfamily crucial for osteoclast differentiation (see below). The mutation results in constitutive activation of RANK. It is unknown whether similar mutations occur in some individuals with Paget's disease.

PATHOPHYSIOLOGY

The characteristic feature is increased resorption of bone accompanied by an increase in bone formation. In the early phase, bone resorption predominates (e.g., in the variant *osteoporosis circumscripta*) and the bones are very vascular. This has been termed the *osteoporotic*, *osteolytic*, or *destructive phase* of disease. Body calcium balance may be negative. Commonly, the excessive resorption is followed closely by formation of new pagetic bone. In this so-called mixed phase of the disease, the rate of bone formation is so geared to that of bone resorption that the magnitude of the increase in bone turnover is not reflected in the overall calcium balance. As the activity decreases, the resorptive rate may decline progressively relative to formation, eventually leading to development of hard, dense, less vascular bone (the so-called *osteoplastic* or *sclerotic phase*) and a positive calcium balance. The rates of bone turnover may be increased enormously in the early phases of the disease, occasionally more than 20 times normal.

Increased generation and overactivity of osteoclasts are considered the major abnormality. The osteoclasts are larger than normal and contain multiple pleomorphic nuclei. Increased numbers of osteoclast-like multinucleated cells are generated from hematopoietic precursors in long-term marrow cultures from individuals with Paget's disease. Production of interleukin (IL) 6 by the pagetic bone (marrow) cells is increased, and the cells are more sensitive than normal to the pro-resorptive effects of $1,25(OH)_2D_3$.

The calcification rate is characteristically increased in pagetic bone. Bone turnover correlates with the increased plasma level of bone alkaline phosphatase, which is higher in Paget's disease than in any other condition except for hereditary hyperphosphatasia. Although increased bone resorption enhances release of calcium and phosphate ions from bone, the plasma concentrations of these ions are usually normal, presumably because of mineral deposition in new bone and because of feedback control of parathyroid hormone secretion. The concentration of phosphate in the plasma is normal or slightly elevated. When the imbalance between bone formation and resorption favors resorption, as after prolonged immobilization or fractures, urinary calcium excretion may be increased and on occasion hypercalcemia may occur. If, on the other hand, bone formation exceeds resorption (relatively uncommon), circulating levels of parathyroid hormone may be increased. Significant increases in trabecular bone resorption and osteoid surfaces in normal bone from patients with Paget's disease may be due to compensatory, secondary hyperparathyroidism. Resorption involves both the organic

and mineral phases of bone. While the inorganic ions of the mineral phase are reutilized for bone formation, amino acids such as hydroxyproline and hydroxylysine and the hydroxypyridinium cross-link compounds are released during resorption of the collagen matrix of bone and are not reutilized for collagen biosynthesis. The increased urinary excretion of small peptides containing hydroxyproline reflects increased bone resorption. The pyridinium cross-link compounds pyridinoline (Pyr) and deoxypyridinoline (D-Pyr) are released from bone collagen during osteoclastic bone resorption and can be measured in urine by several commercial assays. The C- and N-telopeptide measurements are also useful for monitoring therapy, but measurements of serum alkaline phosphatase activity alone are usually sufficient.

RADIOLOGIC CHANGES

The pelvic bones are most commonly involved, followed by the femur, skull, tibia, lumbosacral spine, dorsal spine, clavicles, and ribs; small bones are not as frequently diseased. The lytic phase of the disease may be overlooked except when it occurs in the skull as osteoporosis circumscripta, with areas of sharply demarcated radiolucency in the frontal, parietal, and occipital bones. In the long bones, the lytic areas are usually first seen at one end and progress toward the other end with a V-shaped advancing edge. The lesion may cause expansion of the cortex and exhibit features suggesting malignancy. Usually lysis is followed by a zone of increased density, representing the new bone formation of the mixed phase of the disease. In general, the bone enlarges with an irregularly widened cortex in a coarse, striated pattern and with increased density, occasionally focal in distribution. Perpendicular lines of radiolucency (cortical infractions) are frequent and occur on the convex side of bowed long bones, particularly the femur and tibia. The remodeling of the pagetic bone usually follows the lines of stress produced by muscle pull or gravity, accounting for the characteristic lateral bowing of the femur or anterior bowing of the tibia and the tendency for most of the dense bone to be deposited on the concave side of the bowed bone. In the mixed stage, there is enlargement and thickening of the skull, especially of the outer table, with irregular, spotty areas of increased density (Fig. 343-1). The changes in the pelvis reflect the varying degrees of bone resorption and new bone formation and are frequently accompanied by a characteristic thickening of the pelvic brim. In the sclerotic phase of the disease, the bone may show uniform increase in density, often in the absence of striations. This feature is common in the facial bones but occasionally occurs in the vertebrae, where a homogeneous, dense pattern gives an "ivory" appearance similar to that typical of Hodgkin's disease, although the involved vertebrae in Hodgkin's disease are not enlarged. Computed tomography (CT) and magnetic resonance imaging (MRI) are useful in defining atypical lesions, particularly when neoplastic involvement is suspected. Technetium 99m diphosphonate bone scans are useful in documenting the extent of disease when therapy is contemplated or to confirm the diagnosis when radiologic findings are inconclusive.

CLINICAL MANIFESTATIONS

The clinical presentation is a function of the extent of the disease, the particular bones involved, and the presence of complications. Many patients are asymptomatic. In these individuals the disorder is discovered during radiologic examination of the pelvis or spine for another problem or because of the finding of an elevated level of plasma alkaline

phosphatase. Other individuals may gradually become aware of a swelling or deformity of a long bone or develop a disturbance in gait due to unequal length of the lower extremities. Enlargement of the skull is often not noticed by the patients, except by awareness of increasing hat size. Facial pain and headache are initial complaints in some; backache and leg pain are common. The pain is usually dull but may be shooting or knifelike. Back pain is most common in the lumbar region and may radiate into the buttocks or lower extremities. This pain can be due to the pagetic process itself, to distortion of articular facets, or to secondary osteoarthritis. Pain in the lower extremities may be associated with the transverse cortical infractions along the convex lateral surface of the femur or the anterior surface of the tibia. New lytic lesions detected on bone scan may be the most painful. Pain may also be due to involvement of the hip joint resembling degenerative joint disease, which is characterized by narrowing of the joint space, bony lipping at the margin of the acetabulum, and deepening of the acetabulum. Angioid streaks may be present in the retina. Hearing loss can be due to direct involvement of the ossicles of the inner ear, involvement of bone in the region of the cochlea, or impingement on the eighth cranial nerve in the auditory foramen. More serious neurologic complications can result from overgrowth of bone at the base of the skull (platybasia) and compression of the brainstem. Compression of the spinal cord can cause paraplegia, particularly with involvement of the mid-dorsal spine. Pathologic fractures of vertebrae may also produce spinal cord lesions.

COMPLICATIONS

Blood flow may be markedly increased in extremities involved with Paget's disease. There is proliferation of blood vessels in pagetic bone, but anatomic and functional studies have not confirmed the presence of arteriovenous fistulas. Although blood flow is increased in bone, cutaneous vasodilatation in the pagetic extremities accounts for the increased warmth noted clinically. When the disease is widespread, involving one-third or more of the skeleton, the increased blood flow raises *cardiac output* and rarely leads to high-output heart failure. However, heart disease in individuals with Paget's disease is usually due to the same conditions that occur in other patients of similar age. *Pathologic fracture* may occur at any stage but is more common in the destructive phase of the disease. In the weight-bearing bones fractures are often incomplete, multiple, and on the convex side of the bone. They may occur spontaneously or after slight trauma. Though many lesions heal with no major disability, more serious fractures also occur. Complete fractures are often transverse as if the bone was snapped like a piece of chalk.

There is no characteristic level of urinary calcium excretion, but it tends to be higher when the resorptive phase predominates, possibly accounting for the somewhat higher incidence of *urinary stones* in these patients. *Hyperuricemia* and gout are common in men with Paget's disease, and calcific periarthritis may occur.

Sarcoma is the dread complication. The incidence is probably £1%, although higher incidence has been noted in series that include many patients with polyostotic involvement. The sarcomas most frequently arise in the femur, humerus, skull, facial bones, and pelvis and rarely in the vertebrae. Pagetic osteosarcomas are lytic in appearance on radiographs, in contrast to the sclerotic appearance of radiation-induced osteosarcomas. The tumors are multicentric in about 20% of patients. Fibrosarcomas

and chondrosarcomas have also been found. Pain and swelling are the common complaints that lead to recognition of the sarcomas. The extent and character of the neoplastic involvement are established by CT and/or MRI. In occasional patients, an "explosive rise" of the phosphatase level may accompany the growth of the sarcoma, whereas in patients with limited Paget's disease, phosphatase levels may be only slightly elevated and give no clue to the development of the malignant lesion. The prognosis is poor following the development of sarcomas, and ablative surgery is rarely successful. In contrast to the successful treatment of some osteosarcomas in children, chemotherapy has little effect on survival of patients with pagetic osteosarcomas. Reparative granulomas resembling giant cell tumors may cause local destruction, but they do not metastasize.

TREATMENT

Most patients require no treatment, since the disease is localized and does not cause symptoms. Indications for therapy include persistent pain in involved bones, neural compression, rapidly progressive deformity resulting in disabling disturbance of posture and/or gait, high-output congestive heart failure, hypercalcemia, severe hypercalciuria with or without formation of renal stones, repeated fractures or nonunion, and preparation for major orthopedic surgery. Nonsteroidal anti-inflamatory drugs, such as one of the COX2 inhibitors or acetaminophen, may be useful to relieve pain, especially that involving the hip joints. Patients with severe hip or knee pain, unrelieved by analgesics and not responsive to therapy with agents that inhibit bone resorption, are candidates for total joint replacement. Results of joint replacement are often excellent, although patients with Paget's disease have an increased risk of ectopic bone formation around the operative site. Osteotomies are also useful in patients with bowing deformities of the tibia. In patients who have undergone surgical procedures, early ambulation and adequate fluid intake are important to prevent the development of hypercalciuria and hypercalcemia.

Potent bisphosphonates can inhibit bone resorption and are usually well tolerated. They appear to act by adsorbing to the surface of the calcium/phosphate mineral phase of bone and inhibiting osteoclast function. Bisphosphates are chemically stable analogues of inorganic pyrophosphate and are available in two classes: nitrogen-containing and non-nitrogen-containing. The non-nitrogen-containing bisphosphates (e.g., clodronate and etidronate) are metabolically incorporated into nonhydrolyzable analogues of ATP that may inhibit ATP-mediated reactions. The nitrogen-containing bisphosphates (e.g., pamidronate, alendronate, and risedronate) do not form ATP compounds, but they do inhibit enzymes in the mevalonate pathway, particularly enzymes involved in the synthesis of farnesyl pyrophosphate and geranylgeranyl pyrophosphate. The latter compounds are involved in protein prenylation reactions.

The first bisphosphate available for treatment of Paget's disease in the United States, editronate, was moderately effective in alleviating symptoms but did not decrease biochemical indices to the normal range. Editronate also inhibits mineralization of bone and produces osteomalacia. The newer bisphosphonates such as alendronate, pamidronate, risedronate, and tiludronate are more potent than etidronate and do not produce mineralization defects. Consequently, editronate is no longer indicated for treatment of Paget's disease. In the United States, pamidronate is approved for

intravenous use, and alendronate and risedronate are approved for oral administration.

The bisphosphonates as a class are poorly absorbed from the gastrointestinal tract. Alendronate should be given orally with water after an overnight fast 30 to 60 min before breakfast; the dose is 40 mg/d for 6 months. Risedronate is administered at 30 mg/d for 2 to 3 months. Gastric irritability and rarely esophageal ulcerations may occur. Several regimens are advocated for the intravenous administration of pamidronate. For example, pamidronate is used intravenously as an infusion of 30 mg/d over 3 to 4 h in 5% glucose in water or normal saline on three or four successive days. Responses are usually rapid, with decreases in urinary excretion of hydroxyproline and pyridinium cross-link compounds within days to weeks, followed by a fall in levels of serum alkaline phosphatase. Flulike symptoms accompanied by fever may occur but usually subside rapidly.

Patients given bisphosphonates should also be given daily calcium supplements of 1 to 1.5 g and approximately 400 IU of vitamin D. Clinical and biochemical improvement often lasts for more than a year after bisphosphonate therapy. Clinical evaluation and assessment of alkaline phosphatase levels at 3-month intervals are useful for assessing the need for retreatment. Radiographs at 6-month intervals may be indicated for evaluation of lytic lesions, which usually heal with these agents.

Calcitonin therapy has largely been replaced by bisphosphonates for primary treatment of severe disease, but calcitonin may still be useful in patients who cannot tolerate alendronate or residronate because of gastrointestinal side effects or who prefer to avoid intravenous therapy with pamidronate. The administration of porcine, salmon, and human *calcitonins* for prolonged periods decreases plasma alkaline phosphatase and urinary hydroxyproline excretion. Treatment with calcitonin variably decreases bone pain due to suppression of the pagetic lesion as well as to an independent, centrally mediated analgesic effect. The calcitonins are probably most useful in patients with pain in areas of pagetic involvement not due to associated joint disease. The dose of salmon calcitonin is 50 to 100 MRC units daily given subcutaneously. In most cases, it is possible to reduce the dose to three times weekly. Some patients develop a sensation of warmth and/or nausea 30 min to several hours after injection. Nasal spray formulations of calcitonin can be administered at doses of 200 IU/d. Cytotoxic drugs such as plicamycin and dactinomycin no longer have a place in therapy.

Although the bisphosphonates and calcitonins act primarily to decrease bone resorption, the rate of new bone formation subsequently falls. As a result, the state of high bone turnover is shifted to a state of lower turnover, where rates of formation and resorption are still apparently geared to each other. In this lower turnover state, collagen fibers of the bone matrix are deposited in a more orderly fashion similar to normal bone.

HYPEROSTOSIS

A number of disease states have in common an increase in the mass of bone per unit volume (*hyperostosis*) (Table 343-1). This increase in bone mass is often associated with disturbance in the architecture of the tissue. The additional bone may be located at the periosteum, within the compact bone of the cortex, or in the trabeculae of the cancellous regions. In some diseases, the increase in bone mass may be spotty, as in

osteopoikilosis, whereas in others, most of the skeleton may be involved, as in the malignant form of osteopetrosis in children. The increase in mass is usually not due to an excessive amount of mineral relative to matrix, except in disorders where islands of calcified cartilage may persist such as osteopetrosis. In some diseases, such as the osteosclerosis of untreated renal insufficiency, bone mass and radiodensity may be increased, even though the new bone formed is poorly mineralized and contains widened osteoid seams.

Although hyperostosis is usually due to decreased numbers of osteoclasts or altered osteoclast function, dysfunction of osteoblasts can also occur. For example, an engineered null mutation of the osteocalcin gene in mice results in a higher bone mass due to increased bone formation without change in bone resorption. Infection of newborn mice also produces an osteopetrosis-like phenotype in which osteoblast progenitors appear to induce increased bone formation. In human osteopetrosis of the relatively benign and sporadic type, viral nucleocapsid particles have been found in osteoclasts, and it is possible that viral infection accounts for the excessive bone mass.

OSTEOPETROSIS

Also known as Albers-Schonberg or marble bone disease, osteopetrosis is clinically, biochemically, and genetically heterogeneous. Although osteopetrosis has many causes, a defect in bone resorption is always the underlying mechanism.

Several inherited forms of osteopetrosis occur in rodents, some of which can be cured by bone marrow transplantation from a normal littermate and are probably due to stem cell defects. The osteopetrosis in *op/op* mice and in *tl/tl* toothless rats is not cured by bone marrow transplantation, however. These animals have few osteoclasts, and those that are present appear to be defective. The *op/op* mice have a defect in the coding region of the gene for macrophage colony stimulating factor (M-CSF). The skeletal defects in these animals and in *tl/tl* rats can be reversed by treatment with M-CSF. Another form of osteopetrosis has been produced in mice by targeted disruption of the c-*src* gene, which is normally expressed at high levels in osteoclasts. These *src-/-*mice still have osteoclasts on bone surfaces, but they fail to form a ruffled border at the bone-resorbing surface. Disruption of the c-*fos* gene results in osteopetrosis in which osteoclasts are absent.

Important advances have also been made in understanding the interactions between osteoblasts/stromal cells and hemopoietic osteoclast precursor cells that lead to osteoclastogenesis (Fig. 343-2). A novel member of theTNFreceptor superfamily, referred to as *osteoprotegerin* (OPG; also known as *osteoclastogenesis inhibitory factor*, OCIF) functions as a soluble decoy receptor that binds, and presumably neutralizes,RANKligand, a transmembrane ligand expressed on osteoblasts/stromal cells. RANK ligand binds to RANK, a transmembrane receptor on hemopoietic osteoclast precursor cells, to activate osteoclast differentiation and function (Chap. 340). The RANK receptor binds to intracellular signaling molecules called *TNF receptor-associated factors* (TRAFs) that activate NFkB, a transcription factor known to be required for normal osteoclast function. Genetic models in mice are beginning to unravel this complex signaling pathway. Expression of a soluble version of RANK ligand stimulates osteoclast differentiation. Overexpression of OPG in transgenic mice leads to

osteopetrosis, apparently by blocking RANK ligand. Mice deficient in RANK lack osteoclasts and develop severe osteopetrosis (as well as T cell immunologic defects). TRAF6-deficient mice also develop osteopetrotic features because of defective osteoclast function. It is clear, based on these and other lines of evidence, that the RANK ligand/OPG/RANK/TRAF/NFkB pathway plays a pivotal role in the control of osteoclast development and function.

In humans the infantile autosomal recessive form of osteopetrosis, until recently, has been of unknown cause. It is a severe bone disease that is usually fatal within the first decade of life. Osteoclasts are usually present in normal or increased numbers. In addition, since bone resorption is markedly suppressed, it has been assumed that the defect is not in genes responsible for osteoclast differentiation but in those responsible for osteoclast function. This form of osteopetrosis is manifested in utero and progresses after birth with anemia, hepatosplenomegaly, hydrocephalus, cranial nerve involvement, and death, often due to infections. Transplantation of bone marrow from allogeneic donors to provide normal osteoclast precursor cells has been successful in several patients, in whom osteopetrotic bone was repopulated with donor osteoclasts that produced radiologic and/or bone-biopsy evidence of bone resorption. Nearly 100 bone marrow transplantations have been reported over the past 15 years. If the transplants are successful, markers of donor cells can be found in bone resorbing areas and skeletal improvement persists for years. Although restoration of visual acuity usually does not occur with successful transplantation, this is the only means of approaching cure even in mild cases. The genetic defect in about half of the subjects studied has now been identified. The gene in the human disease was mapped to chromosome 11q13, a region that contains several potential candidate genes. In mice, introduction of a null mutation in one of the genes in this region, *Tcirg1*, that encodes the osteoclast-specific (*OC116*) subunit of the vacuolar proton pump ([V]-type H+-ATPase) responsible for acidification of the extracellular compartment adjacent to the brush border, results in osteopetrosis with abundant osteoclasts. Furthermore a deletion of the 5¢ portion of the gene is the cause for the defect in *oc/oc* mice with spontaneous osteopetrosis. In approximately half of the human subjects with the autosomal recessive form of osteopetrosis so far examined, missense, frameshift, or potential splicing mutations have just been identified in the homologous gene, *TC1RG1*. Thus, mutations in *TG1RG1* are a frequent, although not the sole, cause of this form of osteopetrosis.

Less fulminant forms of osteopetrosis occur in older children and adults. In some the disorder appears to be sporadic, and in others the osteopetrosis is inherited as an autosomal dominant trait and progresses with age; anemia is not as severe, neurologic abnormalities are not as frequent, and recurrent pathologic fractures are the main feature. Although the disorder is most common in infants and children, the diagnosis may be made in adults when roentgenograms are obtained because of fractures or unrelated diseases.

An "intermediate" form of autosomal recessive osteopetrosis has been described in kindreds in which the skeletal abnormality is associated with renal tubular acidosis and cerebral calcification. This form is compatible with long survival and is associated with profound impairment of the activity of one of the isoenzymes of carbonic anhydrase (carbonic anhydrase II). Carbonic anhydrase II is a major component of the system that generates the acid environment adjacent to the ruffled border of the osteoclast.

Deficiency of the enzyme impairs bone resorption. The defect in remodeling results in disorganization of bone structure, with thickened cortices and lack of funnelization of metaphyses. Despite increased density, the bone may be abnormal mechanically and can fracture readily. Osteomalacia or rickets is sometimes a component of osteopetrosis in children.

Roentgenograms reveal uniformly dense, sclerotic bone, often with no distinction between the cortical and cancellous regions (Fig. 343-3). In the severe infantile form, there is persistence of the primary spongiosa with central calcified cartilage cores surrounded by woven bone. Osteoclasts may be increased in number but do not function normally due to the acidification defect that results from the mutated vacuolar proton pump. In other forms of osteopetrosis, there may be different morphologic abnormalities such as loss of ruffled borders. The variability may reflect heterogeneity in this syndrome, as in the osteopetrosis in rodents. The long bones are usually involved, with increased density along the entire shaft. The metaphyses have a characteristic clubbed or splayed appearance. Alternating horizontal bands of increased and decreased density in the long bones and vertebrae suggest that the defect is intermittent during periods of growth. The skull, pelvis, ribs, and other bones may be involved. The phalanges and the distal humerus are usually spared.

Encroachment of bone on the marrow cavity, particularly in the severe infantile disorder, is associated with anemia of the myelophthisic type with extramedullary hematopoiesis in liver, spleen, and lymph nodes and enlargement of these organs. Neurologic abnormalities caused by encroachment on cranial nerves include optic atrophy, nystagmus, papilledema, exophthalmos, and impairment of extraocular muscles. Facial paralysis and deafness are frequent; trigeminal lesions and anosmia are less common. In infants, macrocephaly, hydrocephalus, and convulsions may occur, and infections such as osteomyelitis are frequent. Renal tubular acidosis is a feature of the osteopetrosis associated with carbonic anhydrase II deficiency.

In the less severe forms, about half of patients have no symptoms, and the disorder is discovered incidentally on roentgenograms. Others present with fractures, bone pain, osteomyelitis, and cranial nerve palsies.

Fractures may occur with trivial trauma. Healing of such fractures is usually slow but satisfactory. When the disease is manifested first in adult life, fractures may be the only clinical problem. Levels of calcium and alkaline phosphatase in the plasma are usually normal in adults, but hypophosphatemia and moderate hypocalcemia may occur in children. Serum acid phosphatase levels are usually increased.

As mentioned, in children with severe osteopetrosis, bone marrow transplantation from allogeneic donors or HLA-identical siblings has resulted in histologic and radiologic increases in bone resorption and variable improvement in anemia, vision, hearing, and growth and development. Unfortunately, it is not always possible to find appropriate donors, or patients may not be good candidates for bone marrow transplantation. In some patients with the lethal forms of the disorder, calcitriol therapy is associated with the appearance of osteoclasts with normal ruffled borders and other evidence of increased bone resorption.

PYKNODYSOSTOSIS

Pyknodysostosis is an autosomal recessive form of osteosclerosis that superficially resembles osteopetrosis. It is a form of short-limbed dwarfism associated with bone fragility and a tendency to fracture with minimal trauma. Nevertheless, life span is usually normal. In addition to a generalized increase in bone density, features include short stature; separated cranial sutures; hypoplasia of the mandible; kyphoscoliosis and deformities of the trunk; persistence of deciduous teeth; progressive acroosteolysis of the terminal phalanges; high, arched palate; proptosis; blue sclerae; and a pointed, beaked nose. Patients usually present because of frequent fractures. The disorder is caused by mutations in a gene on chromosome 1q21 that encodes cathepsin K, a cysteine protease that is expressed in normal osteoclasts. Null mutations in the cathepsin K gene in mice result in a phenotype with many features of pyknodysostosis. Osteoclasts are present but do not function normally since there is no proteinase secreted into the area adjacent to the ruffled border where bone collagen resorption normally takes place.

OSTEOMYELOSCLEROSIS

In osteomyelosclerosis, the marrow cells are replaced by diffuse fibroplasia, occasionally accompanied by osseous metaplasia and increased skeletal density on roentgenograms. In early stages woven bone may be found in intratrabecular locations, whereas in more advanced stages, woven bone is observed in the medulla. The disorder is probably a phase in the course of the myeloproliferative disorders and is characteristically accompanied by extramedullary hematopoiesis.

Hyperostosis corticalis generalisata (van Buchem's disease) is characterized by osteosclerosis of the skull (base and calvaria), lower jaw, clavicles, and ribs and thickening of the diaphyseal cortices of the long and short bones. Alkaline phosphatase levels in the serum are elevated, and the disorder may be due to increased formation of bone of normal structure. The major manifestations are due to neural compression and consist of optic atrophy, facial paralysis, and perception deafness. In *hyperostosis generalisata with pachydermia* (Uehlinger), the sclerosis is due to increased formation of subperiosteal spongy bone and involves the epiphyses, metaphyses, and diaphyses. Pain, swelling of joints, and thickening of the skin of the lower arms are common.

HEREDITARY HYPERPHOSPHATASIA

This disorder is characterized by structural deformities of the skeleton, with increased thickness of the calvaria, increased density at the base of the skull, and widening and loss of normal architecture of the shafts and the epiphyses of the long and short bones. The failure to deposit normal bone and the haphazard orientation of lamellae suggest active remodeling that resembles that of Paget's disease. Osteoclasts with multiple nuclei characteristic of Paget's disease and the typical "mosaic" pattern of faceted units of lamellar bone are not found, however. Levels of plasma alkaline phosphatase and urinary excretion of hydroxyproline peptides and other collagen-degradation products are increased. The disorder is apparently inherited as an autosomal recessive trait. Treatment with bisphosphonates or calcitonin therapy may be of value.

PROGRESSIVE DIAPHYSEAL DYSPLASIA (CAMURATI-ENGELMANN DISEASE)

This is an autosomal dominant disorder in which a symmetric thickening and increased diameter of the diaphyses of long bones occurs, particularly in the femur, tibia, fibula, radius, and ulna. Pain over affected areas, fatigue, abnormal gait, and muscle wasting are the major manifestations. Serum alkaline phosphatase levels may be elevated, and, on occasion, hypocalcemia and hyperphosphatemia may be found. Other abnormalities include anemia, leukopenia, and an elevated erythrocyte sedimentation rate. Linkage studies have localized a candidate gene (*DPD1*) to chromosome 19q13.2. Clinical and biochemical improvement may result from the use of glucocorticoids.

MELORHEOSTOSIS

This rare, sporadic condition usually begins in childhood and is characterized by a slowly progressive linear hyperostosis in one or more bones of one limb, usually in a lower extremity. All segments of the bone may be involved, with sclerotic areas that have a "flowing" distribution. The involved limb is often extremely painful. Soft tissue masses, not connected to bone, are often mineralized and are composed of osseous or cartilaginous tissue. Other types of soft tissue masses are associated with joint contractures or consist of fibrofatty, lymphatic, or vascular tissue.

OSTEOPOIKILOSIS

This is a benign autosomal dominant trait usually discovered by chance. In some families, the occurrence of melorheostosis suggests that these disorders may involve the same genetic locus. Osteopoikilosis is characterized by dense spots of trabecular bone <1 cm in diameter, usually of uniform density, located in the epiphyses and adjacent parts of the metaphyses. All bones may be involved except the skull, ribs, and vertebrae.

HYPEROSTOSIS FRONTALIS INTERNA

This is an abnormality of the inner table of the frontal bones of the skull consisting of smooth, rounded enostoses covered by dura and projecting into the cranial cavity. These enostoses are usually <1 cm at their greatest diameter and do not extend posteriorly beyond the coronal suture. The abnormality is found almost exclusively in women who are frequently obese, hirsute, and may have a variety of neuropsychiatric complaints (Morgagni-Stewart-Morel syndrome). The disorder also occurs in women with no obvious illness or particular associated disease. The finding in the skull may be a manifestation of a generalized metabolic disorder.

FIBROUS DYSPLASIA (MCCUNE-ALBRIGHT SYNDROME)

The bony lesions of fibrous dysplasia are characterized by proliferation of fibroblast-like cells that in some areas have features of osteoblasts, with production of an extracellular matrix that may be calcified and have the appearance of woven bone. In other areas the cells have features of chondrocytes and produce a cartilage-like extracellular matrix. The lesions of fibrous dysplasia are usually focal and have a radiolucent appearance; they may be monostotic or polyostotic. The disorder occurs with equal frequency in both

sexes. Some individuals have distinctive areas of skin pigmentation and precocious puberty (McCune-Albright syndrome) (Chap. 336). These diverse manifestations are the consequence of postzygotic mutations in the gene encoding the regulatory G_{sa} proteins.

INCIDENCE

The monostotic form is the most common type of fibrous dysplasia. The lesions can be asymptomatic, can be associated with local pain, or predispose to pathologic fracture. Most of the lesions are in the ribs or in the craniofacial bones, especially the maxillae. Other bones that may be affected include metaphyseal or diaphyseal portions of the proximal femurs or tibias. Monostotic fibrous dysplasia is most often diagnosed in patients between 20 and 30 years of age. There are usually no associated skin lesions. Approximately one-quarter of the individuals with the polyostotic form have more than half the skeleton involved by disease. One side of the body may be affected, and the lesions may be distributed segmentally in a limb, particularly in the lower extremities. Craniofacial lesions are present in approximately half of patients with the polyostotic form. Whereas the monostotic form is usually detected in young adults, fractures and skeletal deformities occur in childhood in the polyostotic form; early-onset disease is generally more severe. Lesions, especially monostotic lesions, can become quiescent at puberty and worsen during pregnancy. McCune-Albright syndrome (polyostotic fibrous dysplasia, multiple cafe au lait spots, and sexual precocity) is more common (10:1) in females. Short stature is due to premature closure of the epiphyses.

PATHOPHYSIOLOGY

Histologically, the lesions contain benign-appearing fibroblastic tissue arranged in a loose whorled pattern (Fig. 343-4). Malignant transformation of either monostotic or polyostotic fibrous dysplasia occurs with a frequency of<1%. The malignant change is usually detected in the third or fourth decade in individuals who have had lesions first identified in childhood. In about one-third of the cases the neoplasms arise in previously irradiated lesions. Ossifying fibroma of long bones is a peculiar fibroosseous cortical lesion that may be a variant of fibrous dysplasia. It is most common in the tibial shaft in teenagers. Although benign, the lesion has a tendency to recur if not adequately excised.

Fibrous dysplasia and McCune-Albright syndrome represent a phenotypic spectrum of disorders caused by activating mutations in the *GNAS1* gene, which encodes the G_{sa} protein. Because these postzygotic mutations occur at different stages in early development, the extent and type of tissues affected by the mutations are variable, explaining the mosaic pattern of skin and bone changes. The mutations occur in regions (e.g., Arg 201) of G_{sa} that selectively inhibit its GTPase activity. Because the GTP-bound form of the regulatory protein confers its active state (Chap. 341), the mutations confer constitutive stimulation of the cyclic AMP-protein kinase A signal transduction pathway. The mutations in *GNAS1* are also found in patients with fibrous dysplasia without manifestations of the McCune-Albright syndrome. Thus, in the boney lesions these mutations result in abnormalities in osteoblastic differentiation and the production of abnormal bone. In addition, there is an associated increase in osteoclastic bone resorption that provides a rationale for therapy with the bisphosphonate, pamidronate. Other tissues in which growth control and function are strongly regulated by

G$_{sa}$protein-coupled receptors are particularly susceptible to the mutations. In addition to bone (parathyroid hormone receptor) and skin (melanocyte-stimulating hormone receptor), various endocrine glands, including the ovary (follicle-stimulating hormone receptor), thyroid (thyroid-stimulating hormone receptor), adrenal (ACTH-receptor), and pituitary (growth hormone-releasing hormone receptor), are commonly affected by the G$_{sa}$mutations. It is of interest that the genetic abnormality in Albright's hereditary osteodystrophy (pseudohypoparathyroidism) is the opposite of that found in the McCune-Albright syndrome. In the former, alterations in G$_{sa}$function or expression result in *deficient* activity and decreased responsiveness to hormones that function through cyclic AMP-mediated signal transduction pathways (Chap. 341).

RADIOLOGIC CHANGES

The roentgenographic appearance of the lesions is that of a radiolucent area with a well-delineated, smooth or scalloped border, typically associated with focal thinning of the cortex of the bone (Fig. 343-5). Fibrous dysplasia can cause bones to become larger than normal, a feature characteristic of Paget's disease as well. The "ground-glass" appearance is due to the thin spicules of calcified woven bone. Deformities can include coxa vara, shepherd's-crook deformity of the femur, bowing of the tibia, Harrison's grooves, and protrusio acetabuli. Involvement of facial bones, usually with lesions of increased radiodensity, may create a leonine appearance (*leontiasis ossea*). Fibrous dysplasia of the temporal bones can cause progressive loss of hearing and obliteration of the external ear canal. Advanced skeletal age in girls is correlated with sexual precocity but can occur in boys without sexual precocity. Occasionally, a focus of fibrous dysplasia may undergo cystic degeneration, with an enormous distortion of the shape of the bone, and mimic the so-called aneurysmal bone cyst.

CLINICAL MANIFESTATIONS

The clinical course is variable. Skeletal lesions are usually detected because of localized pain, deformities, or fractures. Other symptoms ascribable to bone involvement are headache, seizures, cranial nerve abnormalities, hearing loss, narrowing of the external ear canal, or even spontaneous scalp hemorrhages if there is craniofacial bone disease. On rare occasions the onset of sexual precocity is the first clinical manifestation of the McCune-Albright syndrome. Serum calcium and phosphorus values are usually normal. In approximately one-third of patients with the polyostotic form, bone turnover is increased, as reflected by high levels of serum alkaline phosphatase and increased urinary excretion of collagen breakdown products. In some patients high cardiac output resembles that in extensive Paget's disease. Widespread disease does not usually develop when the disease is mild at the outset.

The cutaneous pigmentation in most patients with McCune-Albright syndrome consists of isolated dark-brown to light-brown macules that tend to be located on one side of the midline (Fig. 343-6). The border is usually, although not always, irregular or jagged ("coast of Maine"), in contrast to the smooth borders of the pigmented macules of neurofibromatosis ("coast of California"). As a rule, there are fewer than six of the lesions, which range in size from 1 cm to very large lesions, covering areas such as the back, buttocks, or sacral regions. When the lesions are in the scalp, the overlying hair may be more deeply pigmented. Localized alopecia is associated with osteomas of the

skin, and such lesions tend to overly skeletal lesions. The pigmentation also tends to be on the same side as the skeletal lesions and actually to overlie them. Occasionally, neurofibromatosis and fibrous dysplasia coexist.

Sexual precocity occurs more commonly in girls than in boys. Premature vaginal bleeding, breast development, and growth of axillary and pubic hair are the usual features. Sexual precocity is due to autonomous end-organ activity, not to pituitary or hypothalamic dysfunction. Thus, girls have high estrogen levels and low or undetectable gonadotropins. The characteristic pigmented macules are usual but not invariable. Hyperthyroidism occurs with increased frequency, and rare associations include Cushing's syndrome, acromegaly, pulmonary lesions, and soft tissue myxomas. Hypophosphatemic osteomalacia may also accompany fibrous dysplasia and resembles the disorder associated with other skeletal and nonskeletal tumors.

Although the lytic lesions of fibrous dysplasia resemble the brown tumors of hyperparathyroidism, the age of the patient, normal calcium levels, increased density of bone in the skull, and areas of cutaneous pigmentation identify the former condition. Fibrous dysplasia and hyperparathyroidism may coexist, however. Neurofibromas may involve bone and produce cutaneous pigmentation as well as nodules in the skin. The pigmented macules of neurofibromatosis are more numerous and more widely distributed than in fibrous dysplasia, usually have smooth borders, and tend to involve areas such as the axillary folds. Other lesions that have roentgenographic features similar to those of isolated fibrous dysplasia are unicameral bone cysts, aneurysmal bone cysts, and nonossifying fibromas. Leontiasis ossea is most often due to fibrous dysplasia, although other disorders may also produce this appearance, such as craniometaphyseal dysplasia, hyperphosphatasia, and, in adults, Paget's disease.

TREATMENT

Fibrous dysplasia is not curable. The skeletal lesions, however, can be improved by orthopedic procedures such as casting, osteotomy with internal fixation, curettage, and bone grafting, depending on the lesion and the age of the patient. Indications for such procedures include progressive deformity, nonunion of fractures, and pain unresponsive to conservative treatment. Calcitonin may be effective in treatment of widespread disease associated with bone pain and high serum alkaline phosphatase levels. Pamidronate (0.5 to 1 mg/kg per day intravenously for 2 to 3 days) at 6-month or yearly intervals has been shown to reduce bone pain with refilling of osteolytic lesions in about half of patients and to decrease elevated levels of serum alkaline phosphatase and urinary hydroxyproline excretion. Precocious puberty does not respond to long-acting gonadotropin-releasing hormone (GnRH) analogues, consistent with the autonomous function of the gonads. Aromatase inhibitors, such as testolactone (22 mg/kg per day), have been used to block estrogen production, but with limited efficacy. Promising initial results have been seen with estrogen antagonists, such as tamoxifen. In addition to preventing pubertal progression, blockade of estrogen action is helpful to prevent early epiphyseal closure and short stature.

OTHER DYSPLASIAS OF BONE AND CARTILAGE

A variety of diseases of bone and cartilage have been called *dystrophies* or *dysplasias*.

The *osteochondrodysplasias* are heritable disorders of connective tissue characterized by primary abnormalities of cartilage that lead to disturbances in cartilage and bone growth and development. They comprise several hundred distinct entities, which can be distinguished on the basis of clinical, genetic, and radiologic features. The molecular defects in a number of these disorders have been identified utilizing positional cloning and screening of candidate genes. Several of the disorders are due to mutations in collagen genes. *For discussion of chondrodysplasias, see Chap. 351.*

SPONDYLOEPIPHYSEAL DYSPLASIA

The *spondyloepiphyseal dysplasias* are disorders in which abnormalities of growth occur in various bones, including the vertebrae, pelvis, carpal and tarsal bones, and the epiphyses of tabular bones. On the basis of roentgenographic findings, this group can be divided into (1) those with generalized platyspondyly, (2) those with multiple epiphyseal dysplasias, and (3) those with epiphysometaphyseal dysplasias. *Morquio's syndrome*, in which there is a defect in degradation of glycosaminoglycans (therefore, a "mucopolysaccharidosis"), is inherited as an autosomal recessive trait and is associated with corneal opacities, dental defects, variable disturbances in intellect, and increased urinary excretion of keratosulfate; it belongs in the first group (Chap. 349). Other forms of spondyloepiphyseal dysplasia, some of which are accounted for by defects in type II collagen, may not be recognized until late in childhood or young adult life. Flat vertebral bodies are associated with other abnormalities in shape and alignment. The disordered development of the capital femoral epiphyses leads to irregularities in shape and flattening of the femoral heads and early onset of osteoarthritis of the hips.

ACHONDROPLASIA

This disorder is among the more common types of dwarfism (1 in 15,000 to 1 in 40,000 live births). It is inherited as an autosomal dominant trait, although most cases are sporadic and due to new fibroblast growth factor receptor 3 (FGFR3) mutations (see below). The appearance of short limbs, particularly the proximal portions, with a normal trunk is characteristically accompanied by a large head, a saddle nose, and an exaggerated lumbar lordosis. The length of the spine is almost always normal. Features of the disorder are usually recognizable at birth. Those who survive infancy usually have normal mental and sexual development, and life span may be normal. Spinal deformity nevertheless may lead to a cord compression and nerve root encroachment, especially in those with kyphoscoliosis. Homozygous achondroplasia is a more serious disorder and a cause of neonatal death.

The most common mutation responsible for achondroplasia substitutes an arginine for glycine in the transmembrane domain of FGFR3 and causes a gain-of-function, implying that fibroblast growth factor normally acts via the FGFR3 to inhibit chondrocyte proliferation in the growth plate. The abnormal proliferation at the growth plate, leaving other areas relatively unaffected in the tubular bones, causes production of short bones that are proportionately thick. Formation and maturation of the secondary ossification centers and articular cartilage are not disturbed. Appositional growth at the metaphysis continues, with resulting flare in this region of the bone; intramembranous bone formation at the periosteum is normal. Consistent with the inhibitory role of the FGFR3, a null mutation of the *Fgfr3* gene in mice causes increased growth in the physis.

Mutations in other domains of the *FGFR3* gene have been described in thanatophoric dysplasia, the most severe and lethal dysplasia. In several types of the so-called craniosynostosis syndromes (Pfeiffer, Crouzon, Jackson-Weiss, and Apert syndromes), mutations have been identified in the *FGFR1* or *FGFR2* genes.

Pseudoachondroplasia clinically resembles achondroplasia with respect to the limb deformities but there are no skull abnormalities. Affected individuals have mutations in the gene encoding a non-collagenous component of cartilage called cartilate oligomeric matrix protein (COMP). Mutations in the *COMP* gene have also been described in one of the less severe forms of multiple epiphyseal dysplasia (EDM1).

ENCHONDROMATOSIS (DYSCHONDROPLASIA, OLLIER'S DISEASE)

This is also a disorder of the growth plate in which the hypertrophic cartilage is not resorbed and ossified normally. It results in masses of cartilage with disorderly arrangement of the chondrocytes showing variable proliferative and hypertrophic changes. These masses are located in the metaphyses in close association with the growth plate in children but may be diaphyseal in teenagers and young adults. The disorder is usually recognized in childhood by the appearance of deformities or retardation in growth. The most common sites of involvement are the ends of long bones, usually in the region where rate of growth is most marked. The pelvis is often involved, but ribs, sternum, and skull are seldom affected. There is a tendency toward unilateral involvement. Chondrosarcoma develops occasionally in the enchondromata. The association of enchondromatosis and cavernous hemangiomata in the soft tissues including the skin is known as *Maffucci's syndrome*. Both Ollier's disease and Maffuci's syndrome have been associated with other primary malignancies as diverse as granulosa cell tumor of the ovary and cerebral gliomas.

MULTIPLE EXOSTOSES (DIAPHYSEAL ACLASIS OR OSTEOCHONDROMATOSIS)

This is a disorder of the metaphysis, transmitted in an autosomal dominant manner, in which areas of the growth plate become displaced, presumably by growing through a defect in the perichondrium, or so-called ring of Ranvier. The spongiosa forms within the mass as vessels invade the cartilage. Therefore, the diagnostic radiographic finding is the direct continuity of the mass to the marrow cavity of the parent bone with absence of underlying cortex. Usually the growth of these exostoses ceases when growth of the adjacent plate ceases. The lesions may be solitary or multiple and are usually located in the metaphyseal areas of long bones, with the apex of the exostosis directed toward the diaphysis. Often the lesions produce no symptoms, but occasionally, interference with the function of a joint or tendon or compression of nerves may result. Dwarfism may occur. The metacarpals may be shortened, resembling those seen in Albright's hereditary osteodystrophy. Multiple exostoses are sometimes seen in patients with pseudohypoparathyroidism.

An exostosis may suddenly begin to enlarge long after growth should have ceased, and rarely, chondrosarcomas may develop from the cartilage cap of an exostosis. Although pregnancy may stimulate growth of an exostosis that clinically may mimic malignancy, the lesion merely undergoes exuberant endochondral ossification and cartilage hyperplasia without malignant changes.

Multiple inactivating mutations or deletions have been identified in the *EXT1* or *EXT2* genes in patients with hereditary multiple exostoses. The *EXT* genes probably function normally as a tumor suppressor, and mutations in *EXT* could contribute both to the development of the exostoses and to the malignant transformation to chondrosarcoma that sometimes occurs. Mutations in *EXT* genes apparently cause abnormal processing the cytoskeletal proteins in chondrocytes.

(Bibliography omitted in Palm version)

344. DISORDERS OF LIPOPROTEIN METABOLISM - *Henry N. Ginsberg,, Ira J. Goldberg*

Lipoproteins are macromolecular complexes that carry hydrophobic plasma lipids, particularly cholesterol and triglyceride, in the plasma. More than half of the coronary heart disease (CHD) in the United States is attributable to abnormalities in the levels and metabolism of plasma lipids and lipoproteins. Some premature CHD is due to mutations in major genes involved in lipoprotein metabolism. However, elevated lipoprotein levels in most patients with CHD reflect the adverse impact of a sedentary lifestyle, excess body weight, and diets high in total and saturated fat superimposed on a genetic background that confers susceptibility to increased circulating lipids. A large body of evidence indicates that lifestyle changes and drug treatment strategies that correct hyperlipidemias reduce CHD risk (Chap 242). More than 70 clinical trials examining the effects of cholesterol reduction have been reported, including several large-scale studies using the potent cholesterol-lowering HMG-CoA reductase inhibitors (also known as statins). These studies unequivocally demonstrate that lowering low-density lipoprotein (LDL) cholesterol reduces fatal and nonfatal heart attacks (Table 344-1).

This chapter focuses on the major lipid disorders, including both the dyslipoproteinemias caused by single-gene defects and the disorders that are likely to be multifactorial in origin. A practical approach is provided to assist in the identification, evaluation, and treatment of patients with increased risk ofCHD.

LIPID AND LIPOPROTEIN TRANSPORT

LIPOPROTEIN STRUCTURE

Lipoproteins are spherical particles made up of hundreds of lipid and protein molecules. They are smaller than red blood cells and visible only by electron microscopy. However, when the larger, triglyceride-rich lipoproteins are present in high concentration, plasma can appear turbid or milky to the naked eye. The major lipids of the lipoproteins are cholesterol, triglycerides, and phospholipids. Triglycerides and the esterified form of cholesterol (cholesteryl esters) are nonpolar lipids that are insoluble in aqueous environments (hydrophobic) and comprise the core of the lipoproteins. Phospholipids and a small quantity of free (unesterified) cholesterol, which are soluble in both lipid and aqueous environments (amphipathic), cover the surface of the particles, where they act as the interface between the plasma and core components. A family of proteins, the apolipoproteins, also occupies the surface of the lipoproteins; the apolipoproteins play crucial roles in the regulation of lipid transport and lipoprotein metabolism.

Lipoproteins have been classified on the basis of their densities into five major classes:(1) chylomicrons, (2) very low density lipoproteins (VLDL), (3) intermediate-density lipoproteins (IDL), (4)LDL, and (5) high-density lipoproteins (HDL). The physical-chemical characteristics of the major lipoprotein classes are presented inTable 344-2.

APOLIPOPROTEINS

The apolipoproteins (apos) provide structural stability to the lipoproteins and determine the metabolic fate of the particles upon which they reside. They were named in an arbitrary alphabetical order and, for the purposes of this discussion, will be described in relation to their association with lipoprotein classes (Table 344-3).

There are two forms of apo B -- apo B100 and apo B48. Apo B100 is the major apolipoprotein of VLDL, IDL, and LDL, comprising approximately 30, 60, and 95% of the protein in these lipoproteins, respectively. Apo B100 has a molecular mass of about 545 kDa and is synthesized in the liver. It is essential for the assembly and secretion of VLDL from the liver and is the ligand for the removal of LDL by the LDL receptor. The LDL receptor is a cell-surface protein that binds and internalizes lipoproteins that contain apo B100 or apo E. The LDL receptor binding domain of apo B100 is the sequence between amino acids 3200 and 3600, a region that is absent in apo B48.

Apo B48 is essential for the assembly and secretion of chylomicrons. Apo B48 is encoded by the same gene and messenger ribonucleic acid (mRNA) as Apo B100. However, the mRNA is edited in an unusual way: A cytidine deaminase in the intestine changes a cytidine to a uridine in base 6666 of the apo B100 mRNA to produce a stop codon so that apo B48 contains only the N-terminal 48% of the full-length apo B100. In contrast, the apo B100 mRNA in human liver is not edited. The role of apo B48 in the metabolism of chylomicrons in plasma is unclear. Individuals with mutations that interfere with the normal synthesis of apo B have absent or very low levels of chylomicrons, VLDL, IDL and LDL.

The apolipoproteins of the C series are synthesized in the liver and are present in all plasma lipoproteins (trace amounts in LDL). Individual apo Cs have different metabolic roles, but all inhibit the removal of plasma chylomicrons and VLDL remnants by the liver. Overexpression of apo CI in transgenic mice inhibits the uptake of chylomicron and VLDL remnants by the liver. Apo CI under- or overexpression has not been described in humans. Apo CII is an essential activator of the enzyme lipoprotein lipase (LPL), which hydrolyzes triglycerides in chylomicrons and VLDL; individuals lacking apo CII have severe hypertriglyceridemia. Apo CIII inhibits LPL, and apo CIII overexpression in transgenic mice causes severe hypertriglyceridemia. Humans who lack apo CIII have accelerated rates of VLDL triglyceride lipolysis.

Apo E is synthesized mainly in hepatocytes but is also made in other cells, including macrophages, neurons, and glial cells. It is found in chylomicrons, IDL, VLDL, and HDL and mediates the uptake of these lipoproteins in the liver by both the LDL receptor and the LDL receptor-related protein (LRP). Apo E also binds to heparin-like proteoglycan molecules on the surface of all cells. There are three major apo E alleles: E2, E3, and E4; these isoforms differ in sequence at two positions and have frequencies of about 0.12, 0.75, and 0.13, respectively, in the general population. Apo E2 binds to the LDL receptor with lower affinity than apo E3 or E4. Individuals who are homozygous for apo E2 may develop severe hyperlipidemia (type III dysbetalipoproteinemia); complete absence of apo E increases plasma levels of chylomicron and VLDL remnants and causes early atherosclerosis.

Apo AI, apo AII, and apo AIV are found primarily on HDL. Apo AI and apo AII are synthesized in the small intestine and the liver; apo AIV is made only in the intestine. Apo AI comprises about 70 to 80% of the protein of HDL and plays a critical structural role in HDL particles. Individuals with a profound deficiency of apo AI also lack HDL. Apo AI activates the enzyme lecithin:cholesterol acyltransferase (LCAT), which esterifies free cholesterol in plasma. Plasma levels of HDL cholesterol and apo AI are inversely related to risk for CHD, and some patients with apo AI deficiency develop early, severe atherosclerosis. Transgenic mice that overexpress human apo AI are resistant to atherosclerosis. Apo AII is the second most abundant apoprotein in HDL, but its function has not been determined; transgenic mice that overexpress apo AII have high plasma levels of both HDL cholesterol and triglycerides but may be susceptible to atherosclerosis. Apo AII knockout mice have low levels of HDL, indicating that apo AII is also necessary for the integrity of HDL particles. Apo AIV, a minor component of HDL and chylomicrons may play a role in the activation of LCAT.

Apoprotein(a), a large glycoprotein that shares a high degree of sequence homology with plasminogen, is made by hepatocytes and is secreted into plasma where it forms a covalent linkage with the apo B100 of LDL to form lipoprotein(a). The physiologic role of lipoprotein(a) is not known, but elevated levels are associated with an increased risk for atherosclerosis.

ENZYMES INVOLVED IN LIPOPROTEIN METABOLISM

LPL is synthesized in fat and muscle, secreted into the interstitial space, transported across endothelial cells, and bound to proteoglycans on the luminal surfaces in the adjacent capillary beds. LPL mediates the hydrolysis of the triglycerides of chylomicrons and VLDL to generate free fatty acids and glycerol. The free fatty acids diffuse into adjacent tissues to be burned for energy or stored as fat. Most circulating LPL is associated with LDL. Insulin stimulates the synthesis and secretion of LPL; reduced LPL activity in diabetes mellitus can lead to impaired triglyceride clearance. Homozygotes for mutations that impair LPL have severe hypertriglyceridemia that usually manifests in childhood (type I hyperlipidemia). Heterozygotes for LPL defects have mild to moderate fasting hypertriglyceridemia but may have marked hypertriglyceridemia after consuming a high-fat meal. LPL is also expressed in macrophages, including cholesteryl ester-laden macrophages (foam cells) in atherosclerotic lesions. In this setting, secreted LPL may associate with LDL, causing retention of the lipoprotein in the subendothelial space.

Hepatic triglyceride lipase (HTGL), a member of a family of enzymes that includes LPL and pancreatic lipase, is synthesized in the liver and interacts with lipoproteins in hepatic sinusoids. HTGL removes triglycerides from VLDL remnants (IDL), thus promoting the conversion of VLDL to LDL. It may also play a role in the clearance of chylomicron remnants and in the conversion of HDL_2 to HDL_3 in the liver by hydrolyzing the triglyceride and phospholipid in HDL (see below). Severe hypertriglyceridemia in individuals with genetic deficiency of HTGL is due to the accumulation of chylomicron and VLDL remnants in plasma. In contrast to most patients with hypertriglyceridemia, however, individuals with HTGL deficiency have normal levels of HDL.

LCAT is synthesized in the liver and secreted into plasma where it is bound

predominantly to HDL. LCAT mediates the transfer of linoleate from lecithin to free cholesterol on the surface of HDL to form cholesteryl esters that are then transferred to VLDL and eventually LDL. Apo AI is a cofactor for esterification of free cholesterol by LCAT. Deficiency of LCAT can be caused by mutations in the enzyme or in Apo A1. LCAT deficiency causes low levels of cholesteryl esters and HDL, and it can lead to corneal clouding and renal insufficiency.

Cholesteryl ester transfer protein (CETP) is synthesized primarily in the liver and circulates in plasma in association with HDL. CETP mediates the exchange of cholesteryl esters from HDL with triglyceride from chylomicrons or VLDL. This exchange can explain much of the inverse relationship between plasma levels of triglycerides and HDL cholesterol. LDL cholesteryl ester can also be exchanged with triglyceride from chylomicrons and VLDL, leading to the generation of small, dense LDL. Individuals who are homozygotes for mutations in the CETP gene have marked elevations of HDL cholesterol and apo AI. Heterozygotes for these mutations have slight elevations of HDL, indicating that CETP plays an important role in the removal of cholesteryl esters from HDL.

Phospholipid transfer protein (PLTP) is synthesized in the liver and lung. The production of mature HDL particles depends on PLTP, which provides phospholipid to the enlarging particles.

TRANSPORT OF EXOGENOUS (DIETARY) LIPIDS

Exogenous lipid transport in chylomicrons and chylomicron remnants is depicted in Fig. 344-1A. In western societies, where individuals ordinarily consume 50 to 100 g of fat and 0.5 g of cholesterol during three or four meals, transport of dietary fats is essentially continual. Normolipidemic individuals dispose of most dietary fat in the bloodstream within 8 h of the last meal, but some individuals with dyslipidemia, particularly those with elevated fasting levels of VLDL triglyceride, have measurable levels of intestinally derived lipoproteins in the circulation as long as 24 h after the last meal.

In the intestinal mucosa dietary triglyceride and cholesterol are incorporated into the core of nascent chylomicrons. The surface coat of the chylomicron is composed of phospholipid, free cholesterol, apo B48, apo AI, apo AII, and apo AIV. The chylomicron, essentially a fat droplet containing 80 to 95% triglycerides, is secreted into lacteals and transported to the circulation via the thoracic duct. In the plasma, apo C proteins are transferred to the chylomicron from HDL. Apo CII mediates hydrolysis of triglycerides by activating LPL on capillary endothelial cells in fat and muscle. After the triglyceride core has been hydrolyzed, apo CII and apo CIII recirculate back to HDL. The addition of apo E allows the chylomicron remnant to bind first to heparan sulfate proteoglycans within the space of Disse and then to hepatic LDL receptors and/or LDL receptor-related protein. As a consequence, dietary triglyceride is delivered to adipocytes and muscle cells as fatty acids, and dietary cholesterol is taken up by the liver where it can be used for bile acid formation, incorporated into membranes, resecreted as lipoprotein cholesterol back into the circulation, or excreted as cholesterol into bile. Dietary cholesterol also regulates endogenous hepatic cholesterol synthesis.

Abnormal transport and metabolism of chylomicrons may predispose to atherosclerosis,

and postprandial hyperlipidemia may be a risk factor for CHD. Chylomicrons and their remnants can be taken up by cells of the vessel wall, including monocyte-derived macrophages that migrate into the vessel wall from plasma. Cholesteryl ester accumulation by these macrophages transforms them into foam cells, the earliest cellular lesion of the atherosclerotic plaque (Chap. 241). If the postprandial levels of chylomicrons or their remnants are elevated, or if their removal from plasma is prolonged, cholesterol delivery to the artery wall may be increased.

TRANSPORT OF ENDOGENOUS LIPIDS

The endogenous lipid transport system, which conveys lipids from the liver to peripheral tissues and from peripheral tissues back to the liver, can be separated into two subsystems: the apo B100 lipoprotein system (VLDL, IDL, and LDL) and the apo AI lipoprotein system (HDL).

The Apo B100 Lipoprotein System (See Fig. 344-1*B*) In the liver, triglycerides are made from fatty acids that are either taken up from plasma or synthesized de novo within the liver. Cholesterol can also be synthesized by the liver or delivered to the liver via lipoproteins, particularly chylomicron remnants. These core lipids are packaged together with apo B100 and phospholipids into VLDL and secreted into plasma where apos CI, CII, CIII, and E are added to the nascent VLDL particles. Triglycerides make up the bulk of the VLDL (55 to 80% by weight), and the size of the VLDL is determined by the amount of triglyceride available. Hence, very large triglyceride-rich VLDL are secreted in situations where excess triglycerides are synthesized, such as in states of caloric excess, in diabetes mellitus, and after alcohol consumption. Small VLDL are secreted when fewer triglycerides are available. Although VLDL are the principal hepatic lipoprotein secreted by most individuals, VLDL and cholesteryl ester-enriched IDL and/or LDL-like particles may be secreted by the liver in individuals with combined hyperlipidemia (see below).

In the plasma, triglycerides are hydrolyzed by LPL and VLDL particles are converted to VLDL remnants (IDL). In contrast to chylomicron remnants, VLDL remnants can either enter the liver or give rise to LDL. Larger VLDL particles carry more triglycerides and are likely to be removed directly from plasma without being converted to LDL; apo E in the VLDL remnants binds to the LDL receptor to mediate removal from the plasma. Smaller, more dense VLDL particles are more efficiently converted to LDL; apo E and HTGL play important roles in this process. Individuals with deficiency of either apo E2 or HTGL accumulate IDL in plasma. Apo B100 is the only protein that remains on the surface of the LDL particle.

The half-life of LDL in plasma is determined principally by the availability (or "activity") of LDL receptors. Most plasma LDL is taken up by the liver, and the remainder is delivered to peripheral tissues, including the adrenals and gonads, which utilize cholesterol as a precursor for steroid hormone synthesis. The adrenals have the highest concentration of LDL receptors per cell in the body. Overall, about 70 to 80% of LDL catabolism occurs via LDL receptors, and the remainder is removed by fluid endocytosis and possibly by other receptors.

The LDL receptor, a glycoprotein with a molecular mass of approximately 160 kDa, is

present on the surfaces of nearly all cells in the body. Goldstein and Brown characterized the molecular genetics and cell biology of the LDL receptor and defined its role in cholesterol metabolism. They showed that cholesterol delivered to the cytoplasm by LDL regulates both the rate of cholesterol synthesis in the liver and the number of LDL receptors on the surface of hepatocytes. LDL receptor synthesis is mediated by sterol response element regulatory proteins (SREBPs). These transcription factors are activated in the absence of cholesterol, proteolytically cleaved, and transferred from the endoplasmic reticulum into the nucleus where they stimulate LDL receptor gene expression. Though the LDL receptor is a major factor in determining plasma LDL cholesterol levels, the rates of entry of VLDL into plasma and the efficiency with which VLDL is converted to LDL also influence steady-state LDL concentrations in plasma.

Increased levels of plasma LDL cholesterol and apo B100 are risk factors for atherosclerosis. Normal LDL does not cause foam cell formation when incubated with cultured macrophages or smooth-muscle cells. But, when LDL undergoes lipid peroxidation, it becomes a ligand for alternative, scavenger receptors that are present on endothelial cells and macrophages. Uptake of modified (oxidized) lipoproteins by these receptors in macrophages results in formation of cholesterol-laden foam cells. In addition to inducing foam cell formation, oxidized LDL acts in the vessel wall to stimulate the secretion of cytokines and growth factors by endothelial cells, smooth-muscle cells, and monocyte-derived macrophages (Chap. 242). The consequence is recruitment of more monocytes to the lesion and proliferation of smooth-muscle cells, which synthesize and secrete increased amounts of extracellular matrix, such as collagen. The critical role of LDL in atherosclerosis has been confirmed in genetically altered mice. Although mice are normally resistant to atherosclerosis, increased plasma levels of remnant lipoproteins or LDL lead to atherosclerosis in this species.

The role of VLDL in atherogenesis is less certain. The major reason for this uncertainty derives from the inverse relationship between elevated levels of triglyceride-rich lipoproteins and reduced levels of the antiatherogenic HDL cholesterol. It is possible, for example, that hypertriglyceridemia may not be directly atherogenic but rather the surrogate of other lipoprotein abnormalities. If postprandial hyperlipidemia is a risk factor for CHD, individuals who have normal fasting plasma triglyceride levels but develop postprandial hypertriglyceridemia after consumption of a fat load would be misclassified as "normal" in studies in which only fasting blood samples are analyzed. It is clear that cholesteryl ester-enriched VLDL, isolated from cholesterol-fed animals, can be taken up by receptors on macrophages and smooth-muscle cells and cause foam cell formation. These cholesteryl ester-laden VLDLs are enriched in apo E and are probably representative of VLDL remnants. Thus, the risk of atherosclerosis from hypertriglyceridemia and elevated VLDL levels may be determined by the level of cholesteryl ester-enriched VLDL remnants. The atherogenic potential of IDL is probably similar to that of VLDL remnants.

Apo AI-Containing Lipoproteins (See Fig. 344-1*C*) In contrast to atherogenic apo B lipoproteins, the apo AI-containing HDL appear to be antiatherogenic. In fact, in some studies, HDL cholesterol levels are as strong an indicator of protection from CHD as LDL cholesterol levels are an indicator of risk. Although a great deal is known about the HDL transport system, the mechanism by which these lipoproteins protect against

atherosclerosis is poorly defined.

HDLparticles are formed in plasma from the coalescence of individual phospholipid-apolipoprotein complexes. Apo AI appears to be the crucial, structural apoprotein for HDL, and apo AI/phospholipid complexes probably fuse with other phospholipid vesicles that contain apo AII and apo AIV to form the various types of HDL. The C apoproteins can be added to HDL after their secretion as phospholipid complexes or by their transfer from triglyceride-rich lipoproteins. This may involve the action ofPLTP. These small, cholesterol-poor HDL particles are heterogeneous in size and content and are referred to as HDL_3. Free cholesterol is transferred from cell membranes to HDL_3; a cholesterol transporter called ABC1 mediates this important first step in reverse cholesterol transport. Free cholesterol in HDL_3 is converted to cholesteryl ester byLCAT, and the cholesteryl ester moves into the core of the HDL. Formation of cholesteryl ester increases the capacity of the HDL_3 to accept more free cholesterol and enlarge to form the more buoyant class of HDL particles termed *HDL_2*. HDL_2 can be metabolized by two pathways: (1) cholesteryl esters can be transferred from HDL_2 to apo B lipoproteins or cells, or (2) the entire HDL_2particle can be removed from plasma. The transfer of cholesteryl ester from HDL to triglyceride-rich apo B lipoproteins (chylomicrons andVLDL in the fed and fasted states, respectively) is mediated byCETP. Triglyceride is transferred to HDL in this process and is a substrate for lipolysis byLPLand/orHTGL. As a result, HDL_2 is converted back into HDL_3. When the apo B lipoproteins are removed by the liver, reverse cholesterol transfer is complete. HDL cholesteryl ester may also be transferred selectively to cells via interaction of HDL with the scavenger receptor B-1, a receptor expressed by hepatocytes and steroid-producing cells. HDL-mediated reverse cholesterol transport (from peripheral tissues to the liver) is thought to be the primary mechanism by which HDL protects against atherosclerosis.

Rarely, low plasmaHDL is due to a genetic deficiency of one of the structural components of HDL (such as apo AI). However, low HDL cholesterol levels are usually the secondary consequence of increased plasma levels ofVLDL andIDL (or chylomicrons and their remnants). Mutations in the *ABC1* gene (see above) are associated with Tangier's disease, a rare form of low HDL. Low levels of HDL cholesterol and apo AI may increase atherosclerosis risk by any of several mechanisms. HDL could remove cholesterol from foam cells in atherosclerotic lesions or protectLDL from oxidative modification. Alternatively, the atherosclerotic risk of low HDL may be due to the commonly associated elevations of apo B-containing lipoproteins, which accept HDL cholesteryl esters and deliver cholesteryl esters to the vessel wall.

THE HYPERLIPOPROTEINEMIAS (See Table 344-4)

HYPERCHOLESTEROLEMIA

Elevated levels of fasting plasma total cholesterol in the presence of normal levels of triglycerides are almost always associated with increased concentrations of plasmaLDLcholesterol (type IIa), as LDL carries about 65 to 75% of total plasma cholesterol. A rare individual with markedly elevatedHDLcholesterol may also have increased plasma total cholesterol levels. Elevations of LDL cholesterol can result from single-gene defects, polygenic disorders, or from the secondary effects of other disease

states.

Familial Hypercholesterolemia (FH) FH is a codominant genetic disorder that occurs in the heterozygous form in approximately 1 in 500 individuals. FH is due to mutations in the gene for theLDLreceptor and is genetically heterogeneous, >200 different mutations in the gene having been described. Plasma levels of LDL cholesterol are elevated at birth and remain so throughout life. In untreated adults, total cholesterol levels range from 7 to 13 mmol/L (275 to 500 mg/dL). Plasma triglyceride levels are typically normal, andHDLcholesterol levels are normal or reduced. As would be expected of a disorder with decreased numbers of LDL receptors, the fractional clearance of LDL apo B is reduced. LDL production is increased because the liver secretes moreVLDL andIDL and more IDL particles are converted to LDL rather than taken up by the hepatic LDL receptors. FH heterozygotes usually develop severe atherosclerosis in early or middle age. *Tendon xanthomas*, which are due to both intracellular and extracellular deposits of cholesterol, most commonly involve the Achilles tendons and the extensor tendons of the knuckles; they are found in about 75% of adults with FH (Fig. 344-CD1). *Tuberous xanthomas*, which are softer, painless nodules on the elbows and buttocks, and *xanthelasmas*, which are barely elevated deposits of cholesterol on the eyelids, are common in heterozygous FH (Figs. 344-CD2 and344-CD3).CHDdevelops in men by the fourth decade of life or earlier.

The homozygous form ofFHoccurs in 1 out of 1 million individuals and is associated with a marked increase of plasma cholesterol levels (>13 mmol/L;>500 mg/dL), large xanthelesmas, and prominent tendon and planar xanthomas. These individuals have severe, prematureCHD that can be manifested in childhood.

Familial Defective Apo B100 This autosomal dominant disorder is a phenocopy ofFH and is due to a missense mutation at amino acid 3500 that reduces the affinity ofLDL for the LDL receptor and, thus, impairs LDL catabolism. The prevalence and manifestations of both the heterozygous and homozygous forms are similar to those produced by mutations of the LDL receptor.

Polygenic Hypercholesterolemia Most moderate hypercholesterolemia [plasma cholesterol levels between 6.5 and 9 mmol/L (240 and 350 mg/dL)] is polygenic in origin. Multiple genes interact with environmental factors to contribute to the hypercholesterolemia, and both overproduction and reduced catabolism ofLDL are thought to play roles in the pathophysiology. The severity is probably affected by the consumption of saturated fat and cholesterol, age, and the level of physical activity. Plasma triglyceride andHDLcholesterol levels are usually normal. These individuals are at increased risk of atherosclerosis. Tendon xanthomas are not present. Genes involved in cholesterol and bile acid metabolism may be involved in the pathogenesis.

HYPERTRIGLYCERIDEMIA

The diagnosis of hypertriglyceridemia is made by determining plasma lipids after an overnight fast. Because of the less certain association of triglycerides withCHD(compared toLDLcholesterol), plasma concentrations greater than the 90th or 95th percentile for age and sex have been used to define hypertriglyceridemia. Some studies show, however, that plasma triglyceride levels >130 to 150 mg/dL are

associated with low HDL cholesterol levels and small, dense LDL particles. Furthermore, a meta-analysis of several prospective population studies confirms that triglyceride concentrations are independent predictors of CHD risk. Isolated elevations of plasma triglycerides can be due to increased levels of VLDL (type IV) or combinations of VLDL and chylomicrons (type V). Rarely, only chylomicron levels are elevated (type I). Plasma is usually clear when triglyceride levels are <4.5 mmol/L (<400 mg/dL) and cloudy when levels are higher and VLDL (and/or chylomicron) particles become large enough to scatter light. When chylomicrons are present, a creamy layer floats to the top of plasma after refrigeration for several hours. Tendon xanthomas and xanthelasmas do not occur with isolated hypertriglyceridemia, but eruptive xanthomas (small orange-red papules) (Fig. 344-CD4) can appear on the trunk and extremities when triglyceride levels are >11.5 mmol/L (>1000 mg/dL) (i.e., when chylomicronemia is present). At these high levels of triglycerides, the retinal vessels can appear to be orange-yellow in color (lipemia retinalis). Pancreatitis is the major risk associated with plasma triglyceride concentrations >11 mmol/L (>1000 mg/dL).

Elevations in plasma triglycerides are usually associated with increased synthesis and secretion of VLDL triglycerides by the liver. Hepatic triglyceride synthesis is regulated by substrate flow (the availability of free fatty acids), energy balance (the level of glycogen stores in the liver), and hormonal status (the balance between insulin and glucagon). Obesity, excessive consumption of simple sugars and saturated fats, inactivity, alcohol consumption, and insulin resistance are commonly associated with hypertriglyceridemia. In most of these situations, increased free fatty acid flux from adipose tissue to the liver stimulates the assembly and secretion of VLDL. When VLDL triglyceride levels are markedly elevated [>11.5 mmol/L (>1000 mg/dL)], LPL may be saturated so that an acquired LPL deficiency develops during the postprandial period even if there is no underlying genetic disorder. The addition of chylomicrons to the circulation may cause dramatic increases in plasma triglycerides.

Familial Hypertriglyceridemia Familial hypertriglyceridemia appears to be transmitted as an autosomal dominant disorder, though the underlying mutation(s) have not been identified. The pathophysiology is complex: both reduced catabolism of triglyceride-rich lipoproteins and overproduction of VLDL have been reported. Elevated levels of fasting plasma triglycerides in the range of 2.3 to 8.5 mmol/L (200 to 750 mg/dL) are usually associated with increased levels of VLDL triglycerides only. When VLDL triglyceride levels are markedly elevated (regardless of etiology), chylomicron triglycerides can also be present, even after a 14-h fast. A 20-year follow-up of individuals with familial hypertriglyceridemia demonstrated a moderate increase in CHD risk.

Familial Lipoprotein Lipase Deficiency This autosomal recessive disorder is due to the severe impairment or absence of LPL, leading to massive accumulation of chylomicrons in plasma. Manifestations begin in infancy and include pancreatitis, eruptive xanthomas, hepatomegaly, splenomegaly, foam cell infiltration of the bone marrow, and, when the level of triglycerides is >11 mmol/L (1000 mg/dL), lipemia retinalis. Atherosclerosis is not accelerated. The diagnosis is suspected by finding a creamy layer (chylomicrons) at the top of plasma that has incubated overnight at 4°C; it is confirmed by demonstrating that LPL levels in plasma do not increase after the administration of heparin (which normally releases LPL from endothelial surfaces). Manifestations recede dramatically when patients are placed on fat-free diets.

LPLlevels are within the normal range in most patients with moderate hypertriglyceridemia [2.8 to 5.6 mmol/L (250 to 500 mg/dL)]. Heterozygous mutations in the LPL gene are present in 5 to 10% of hypertriglyceridemic individuals; LPL activity may be reduced by 20 to 50% in these individuals. Heterozygotes for LPL deficiency may also present with severe hypertriglyceridemia if they have poorly controlled diabetes, are pregnant, consume excessive quantities of alcohol, take exogenous estrogen, or are obese.

Familial Apoprotein CII Deficiency This rare autosomal recessive disorder causes a functional deficiency ofLPL and clinical manifestations similar to those of familial LPL deficiency. Deficiency of apoprotein CII impairs hydrolysis of chylomicrons andVLDL so that either, or both, lipoproteins accumulate in blood. The diagnosis is suspected in children or adults with recurrent attacks of pancreatitis and confirmed by demonstrating the absence of apo CII on gel electrophoresis and that plasma transfusion (which contains abundant apo CII) causes a dramatic fall in plasma triglycerides. Heterozygotes have half-normal levels of apo CII, may have mild elevations of triglycerides, and are asymptomatic. Dietary fat restriction should be life-long.

Hepatic Lipase Deficiency Total deficiency ofHTGL is a rare autosomal recessive disorder that impairs the final catabolism and/or remodeling of smallVLDL andIDL. Subjects with HTGL deficiency often have elevated levels of VLDL remnants;HDL$_2$levels may be elevated because HTGL participates in the conversion of HDL$_2$ to HDL$_3$. HTGL activity is frequently increased in hypertriglyceridemic individuals, but the meaning of this association is unclear.

HYPERCHOLESTEROLEMIA WITH HYPERTRIGLYCERIDEMIA

Concomitant hypercholesterolemia and hypertriglyceridemia occurs in two disorders -- familial combined hyperlipidemia (FCHL) and dysbetalipoproteinemia.

Familial Combined Hyperlipidemia FCHL is transmitted as an autosomal dominant disorder. Probands (the initial case discovered within a family) typically have combined hyperlipidemia, isolated hypertriglyceridemia, or isolated elevated levels ofLDLcholesterol. The diagnosis requires documentation at some time of combined hyperlipidemia in the proband or, if the proband has isolated hypercholesterolemia or hypertriglyceridemia, the various lipid phenotypes in first-degree relatives at risk. The lipoprotein phenotype in affected individuals may change over time. The underlying defect in this disorder is not known, though mutations or polymorphisms in theLPL gene and in the gene cluster for apo AI, apo CIII, and apo AIV may contribute to the disorder in some families. Insulin resistance is present in many individuals with FCHL; the link may result from increased free fatty acid flux driving assembly and secretion of apo B100 lipoproteins.

FCHLis associated with increased secretion ofVLDLparticles, as determined by the flux of VLDL apo B. The lipoprotein patterns associated with the disorder are most likely determined by genetic polymorphisms in genes that regulate the metabolism of VLDL. For example, if the affected individual also has a defect inLPL, hypertriglyceridemia will be present. Since the hydrolysis of VLDL triglycerides also regulates the generation

of LDL in plasma, individuals with FCHL who have inefficient catabolism of VLDL may also have reduced levels of LDL cholesterol and high VLDL cholesterol. Finally, individuals with FCHL who synthesize normal quantities of triglycerides and secrete VLDL that carries normal amounts of triglyceride generate increased numbers of LDL particles and present with isolated elevations of plasma LDL cholesterol. These variations in VLDL catabolism, together with additional genetic heterogeneity and environmental variability, form the basis for the variable phenotype in this disorder. FCHL may occur in as many as 0.5 to 1.0% of Americans and is the most common familial lipid disorder in survivors of myocardial infarction. The increased risk for atherosclerosis is due to the presence of increased numbers of small, atherogenic VLDLs and the conversion of VLDL to the more atherogenic IDL and LDL. Persons with FCHL usually have clear plasma and do not have xanthomas or xanthelasma.

Dysbetalipoproteinemia This rare disorder affects 1 in 10,000 persons and is due to homozygosity for apo E2, the binding-defective form of apo E. Because apo E plays a crucial role in the catabolism of chylomicron and VLDL remnants, affected individuals have elevations in both VLDL triglyceride and VLDL cholesterol, and chylomicron remnants are present in fasting plasma. The ratio of total cholesterol to triglyceride approximates 1.0, and the ratio of VLDL cholesterol to triglyceride is greater than 0.25. LDL and HDL cholesterol levels are usually low. Although 1% of the population is homozygous for apo E2, most have normal plasma triglyceride and cholesterol levels. Thus, a second defect in lipid metabolism must be present in the 0.01% of individuals with dysbetalipoproteinemia. These individuals may have tuberous xanthomas and deposits of cholesterol in the palmar creases (striae palmaris); the latter, appearing as yellow-orange lines, are specific for dysbetalipoproteinemia. The risk for atherosclerosis and its complications is increased, with onset in the fourth and fifth decades. The incidence of peripheral vascular disease is higher than in FH.

REDUCED HDL CHOLESTEROL

Low levels of HDL cholesterol can be defined as <0.9 mmol/L (<35 mg/dL) in men and <1 to 1.2 mmol/L (<40 to 45 mg/dL) in women. Low concentrations of HDL cholesterol are usually associated with coexistent hypertriglyceridemia, though "primary hypoalphalipoproteinemia" has been identified in both individuals and families. The relationship between hypertriglyceridemia and low HDL levels probably derives from: (1) CETP-mediated transfer of cholesteryl ester from the core of HDL to VLDL; (2) shift of surface components, particularly phospholipids apo CII, and apo CIII, from HDL to VLDL; and (3) increased fractional catabolism of the cholesteryl ester-poor apoAI that results from the first two processes. The complexity of the relationship between HDL and triglyceride levels is highlighted by the fact that HDL levels do not return to normal when fasting plasma triglycerides are reduced in most persons with hypertriglyceridemia and low HDL cholesterol levels. Low HDL is clinically silent, and the plasma is usually clear (it can be cloudy or creamy if there is concomitant hypertriglyceridemia).

Primary hypoalphalipoproteinemia refers to the state where HDL cholesterol concentrations are markedly reduced but plasma triglyceride concentrations are normal. Many individuals with this phenotype have had hypertriglyceridemia in the past or have an older (or more obese) first-degree relative who has both low HDL and increased triglyceride levels. Hence, both family studies and long-term follow-up may be required

to identify individuals with primary reductions in HDL cholesterol. Rare mutations have been described in the apo AI gene that lead to reductions in apo AI synthesis or increases in catabolism. One mutation that is common in Italy, apo AI-Milano, is associated with a high fractional clearance rate of apo AI but is not associated with increased risk for atherosclerosis.

Some rare genetic disorders of lipid metabolism are summarized in Table 344-5.

SECONDARY CAUSES OF HYPERLIPOPROTEINEMIA (See Table 344-6)

Diabetes Mellitus Diabetes can affect lipid and lipoprotein metabolism through several mechanisms (Chap. 333). In type 1 diabetes mellitus (DM) (formerly called insulin-dependent diabetes mellitus), plasma lipids are usually normal when control of diabetes with insulin is adequate. In diabetic ketoacidosis, hypertriglyceridemia can be severe due to increases in both VLDL and chylomicrons. These abnormalities are associated with overproduction of VLDL and LPL deficiency secondary to insulinopenia. They usually improve with tight control of the diabetes. In type 2 DM (formerly called non-insulin-dependent diabetes mellitus), insulin resistance and obesity combine to cause mild to moderate hypertriglyceridemia and low HDL cholesterol levels. In general, this pattern of dyslipidemia is due to overproduction of VLDL. LDL cholesterol is usually normal in type 2 DM, though the LDLs are small, dense, and perhaps more atherogenic. Treatment of type 2 DM and weight reduction improve, but usually do not completely correct, the dyslipidemia (particularly the low HDL cholesterol levels). Therapy of hyperlipidemia should not be delayed in patients with type 2 DM, as they are at increased risk for CHD. It is recommended that patients with diabetes should be treated as if they already have CHD, i.e., the treatment goal is to reduce their LDL to <2.6 mmol/L (<100 mg/dL) (Fig. 344-2).

Hypothyroidism Hypothyroidism accounts for about 2% of all cases of hyperlipidemia and is second only to DM as a cause of secondary hyperlipidemia. Levels of LDL cholesterol can be elevated, even in patients with subclinical disease in whom thyroid-stimulating hormone (TSH) levels are elevated but other thyroid function tests are normal. Hypertriglyceridemia can occur if obesity is present. Hypothyroidism is also associated with increased levels of HDL cholesterol, probably because of reduced HTGL activity. Correction of hypothyroidism reverses the lipid abnormalities.

Renal Disease Renal disease causes a wide range of lipid abnormalities. The nephrotic syndrome can be accompanied by elevations in LDL, VLDL, or both. The severity of the hyperlipidemia correlates with the degree of hypoproteinemia. Renal failure is associated with hypertriglyceridemia and low HDL cholesterol concentrations.

Ethanol The metabolism of ethanol enhances the level of NADH in the liver which, in turn, stimulates the synthesis of fatty acids and their incorporation into triglycerides. Moderate ethanol consumption raises plasma VLDL levels, with the degree of elevation dependent on the baseline level. Severe hypertriglyceridemia and pancreatitis usually develop on the background of a genetic hyperlipidemia and heavy alcohol intake. Because ethanol also stimulates the synthesis of apo AI and inhibits CETP, ethanol-associated hypertriglyceridemia is usually accompanied by normal or elevated levels of HDL cholesterol.

Liver Disease Primary biliary cirrhosis and extrahepatic biliary obstruction can cause hypercholesterolemia and elevated levels of plasma phospholipids associated with increased levels of an abnormal lipoprotein (lipoprotein X;Chap. 299) andLDL. Severe liver injury often leads to a decrease in levels of both cholesterol and triglyceride. Acute hepatitis can cause elevated levels ofVLDL and impairment ofLCATformation.

AIDS Use of protease inhibitor therapies has been associated with a generalized metabolic syndrome that includes hypertriglyceridemia, alterations in fat distribution, and occasionally type 2DM(Chap. 309).

DIAGNOSIS

Although the initial indication of an abnormality in lipoprotein metabolism is via blood measurements of triglyceride and cholesterol, the disorders are due to abnormalities of specific lipoproteins. Thus, lipoprotein analysis should assessVLDL,LDL, andHDLlevels. Direct measurements of plasma LDL require laborious centrifugation techniques. However, LDL cholesterol concentrations can be estimated indirectly in individuals with triglyceride levels <4.5 mmol/L (<400 mg/dL) by subtracting the HDL and VLDL cholesterol from the total plasma cholesterol. HDL cholesterol is determined after chemical precipitation of VLDL and LDL. VLDL cholesterol is estimated to be the plasma triglyceride level divided by five. Therefore

where all values are measured in milligrams per deciliter.

In persons with triglyceride levels >4.5 mmol/L (>400 mg/dL), the ratio of triglyceride to cholesterol inVLDL is>5, and this equation cannot be used to calculate the plasmaLDLcholesterol level. The other disorder that is not detected with this method is dysbetalipoproteinemia because the ratio of triglyceride to cholesterol in the VLDL is <<5. In these two situations, direct measurement of LDL cholesterol must be performed in ultracentrifuged plasma. Commercial methods for the measurement of "direct LDL" are available. Although these methods appear to be precise and accurate, the measured values for LDL cholesterol are 0.06 to 0.17 mmol/L (5 to 15 mg/dL) less than estimated LDL because the estimated value is actually the combination ofIDL and LDL. If a "direct LDL" measurement is used, the National Cholesterol Education Program (NCEP) guidelines (based on estimated LDL) must be adjusted before therapeutic decisions are made.

Because plasma triglyceride levels rise and bothHDL andLDLcholesterol levels fall modestly after a fat-containing meal (due to the action ofCETP), it is preferable to measure plasma lipids after a 12-h fast. Measuring cholesterol levels alone will not detect individuals with isolated low HDL; screening forCHDshould therefore include measurement of HDL. Because serum lipid levels vary from day to day, at least two to three measurements should be made days or weeks apart before initiating therapy. Some experts advocate the use of total cholesterol/HDL ratios as a better assessment of individual risk. This is a reasonable approach provided both the patient and physician are aware that the treatment goal is to reduce LDL. In addition, rare patients with very

high or very low levels of both LDL and HDL have ratios that are not interpretable on the basis of population studies. Although some laboratories offer measurements of individual apoproteins (e.g., apo B100 and apo AI), or size estimates of LDL, this information is not generally helpful in decision-making. Measurement of lipoprotein (a) levels can provide an indication of risk that cannot be gleaned from lipid measurements. Lipoprotein electrophoresis is not useful except for the diagnosis of dysbetalipoproteinemia, a diagnosis that otherwise requires ultracentrifugation methods. Apo E genotyping is also helpful in the diagnosis of dysbetalipoproteinemia (although rarely the disorder can be due to other defects in the apo E gene).

BothLDL andHDLcholesterol levels are temporarily decreased for several weeks after myocardial infarction or acute inflammatory states but can be accurately measured if blood is obtained within 8 h of the event.

Approach to the Patient

Elevated LDL Cholesterol Treatment of elevated LDL cholesterol can have either of two aims -- *primary prevention* of the complications of atherosclerosis or *secondary treatment* after complications have occurred. The rationale for primary prevention is based on the large body of data linking elevated levels of LDL cholesterol with increasedCHD risk and an impressive body of clinical and experimental data demonstrating that reducing LDL cholesterol slows progression and may actually induce regression of atherosclerotic lesions (Chap. 242). Both primary and secondary intervention trials indicate that total mortality can be reduced when the LDL cholesterol is lowered (Table 344-1). A meta-analysis of four randomized trials (4S, CARE, AFCAPS/TexCAPS, LIPID) comparing HMG-CoA reductase inhibitors to control included 30,817 participants and found that HMG-CoA reductase inhibitor treatment was associated with: (1) a 20% decrease in total cholesterol, a 28% decrease in LDL cholesterol, a 13% decrease in triglycerides, and a 5% increase in HDLcholesterol; (2) a 31% decrease in major coronary events and a 21% decrease in all-cause mortality; (3) similar risk reduction in women and men; and (4) no effect on noncardiovascular mortality. Unexpectedly, the risk of stroke was also reduced 19 to 32% by HMG-CoA reductase inhibitor treatment, even though previous observational studies show a relatively weak association between cholesterol level and stroke risk.

Dietary Alterations A fundamental starting point for both primary prevention and secondary treatment involves counseling to modify diet, exercise, smoking, and other life-style factors that increase the risk of CHD. The typical American diet derives about 35% of its calories from fat (14 to 15% from saturated fat) and contains 400 to 500 mg/d of cholesterol. Individuals with hyperlipidemia should be encouraged to eat a diet lower in cholesterol and saturated fat. TheNCEPStep 1 diet, which is recommended for all Americans above age 2, provides 30% of calories from fat, <10% of calories from saturated fat, and <300 mg/d of cholesterol (Table 344-7). Carbohydrate is the typical nutrient used to replace fat in patients with isolated hypercholesterolemia. In general, whole-milk dairy products, egg yolks, meats, palm oil, and coconut oil should be replaced with fresh fruits and vegetables, complex carbohydrates (especially whole-grain products), and low-fat dairy products. Shellfish are low in fat content and, except for shrimp, also have low cholesterol levels; shrimp, in moderation, is acceptable. Portion size needs to be stressed; the protein and fat-rich portion of meat in

a given meal should be<115 g (4 oz), the size of a deck of cards. Substitutions with any food low in saturated fat such as bran, nuts, and olive oil will have positive effects on LDL. Hydrogenation of vegetable oils increases the saturation of the fatty acids. In particular, trans-fatty acids, mainly found in commercially hydrogenated vegetable oils, raise LDL and can lower HDL cholesterol levels. Use of stanol-containing margarines has, by contrast, lowered LDL cholesterol about 5 to 10% by blocking cholesterol absorption in the small intestine. When further diet therapy is indicated, the NCEP Step 2 diet provides 30% of calories as fat but <7% of calories from saturated fat, and <200 mg/d of cholesterol. After changing from the average American diet to the Step 1 diet, the LDL cholesterol usually drops 8 to 10%; an additional reduction of 5 to 7% can be achieved by advancing to the Step 2 diet. There is, however, great individual variability in diet responsiveness, and several values should be obtained before judging the efficacy of any diet treatment.

Primary Prevention The NCEP Adult Treatment Panel recommends measuring plasma cholesterol in all adults older than age 20 at least every 5 years. Ideally, this testing involves a lipoprotein profile to allow better risk stratification. Primary prevention goals include LDL cholesterol <3.36 mmol/L (<130 mg/dL), triglycerides <1.7 mmol/L (<150 mg/dL), and HDL cholesterol >1.03 mmol/L (>40 mg/dL) for men and >1.29 mmol/L (>50 mg/dL) for women. In individuals without DM or known CHD, treatment recommendations for primary prevention are outlined in Fig. 344-2. Assessment of risk factors in addition to LDL cholesterol is an essential part of this decision-making process. Risk factors include: (1) family history of premature CHD (<55 years in a male parent or sibling or <65 in female relatives), (2) hypertension (even if it is controlled with medications), (3) cigarette smoking (>10 cigarettes per day), (4) DM, and (5) low HDL [<0.9 mmol/L (<35 mg/dL)]. In addition, because CHD is more prevalent in older individuals, age (men >45 years, women >55 years, or younger women with premature menopause without estrogen replacement) is also an important risk factor. HDL cholesterol >1.6 mmol/L (>60 mg/dL) is a negative risk factor, i.e., one other risk factor can be negated by a high HDL cholesterol level.

In individuals with fewer than two risk factors, life-style modifications alone and follow-up testing may be used if LDL is <4.14 mmol/L (<160 mg/dL). For those with LDL >4.91 mmol/L (>190 mg/dL), drug treatment is indicated. If two or more risk factors are present, drug treatment in addition to life-style modifications should be instituted if LDL cholesterol is >3.36 mmol/L (>130 mg/dL). HMG-CoA reductase inhibitors are first-line medications for most patients; niacin and resins are second-line treatments (see below).

Secondary Prevention The NCEP guidelines are stringent for the secondary treatment of patients with CHD. Patients with CHD should be screened for lipid abnormalities during and after their initial diagnoses. A goal of lowering plasma LDL concentrations to <2.6 mmol/L (<100 mg/dL) is advocated for such individuals as well as for patients with DM (Fig. 344-2). As described below, this requires modifications of diet in addition to the use of one or more medications.

If a patient with CHD has only a modestly elevated LDL cholesterol level [e.g., <3.36 mol/L (<130 mg/dL)], a 4- to 6-week period of Step 1 diet therapy can precede the addition of drugs. In such a patient, moving to the Step 2 diet, which provides the same total fat but <7% of calories from saturated fat, can be useful. If, however, the LDL

cholesterol is >3.36 mmol/L (>130 mg/dL), drug therapy should be instituted along with diet therapy (Fig. 344-2).

***High Triglycerides and Low*HDL**The evidence that treatment to reduce plasma triglyceride levels or increase levels of HDL cholesterol leads to long-term health benefits is less compelling than that for treatment of highLDLlevels. Two recent clinical trials have shown, however, that lowering triglycerides (using fibric acids) or lowering LDL (HMG-CoA reductase inhibitors) decreasesCHDevents in these patients. There have been no intervention trials in which only increases in HDL cholesterol concentrations have been achieved. Beneficial effects of niacin have been attributed, in part, to its HDL-raising effect and its action to reduce triglycerides and LDL. Even with drugs that primarily affect LDL cholesterol levels, such as bile acid-binding resins and HMG-CoA reductase inhibitors, some of the benefits achieved may be related to increases in HDL cholesterol levels.

In patients with isolated elevations of triglyceride levels or with hypertriglyceridemia and highLDLcholesterol, life-style modifications should be introduced as described above, and weight reduction should be strongly encouraged if obesity is present. Fat intake should be decreased, but the concomitant increase in carbohydrate intake may raise triglyceride and lowerHDLcholesterol levels. If this occurs, replacing some of the saturated fat with monounsaturated fat, which will not raise LDL cholesterol, may be valuable. Severe hypertriglyceridemia and hyperchylomicronemia require very low fat diets, avoidance of free sugars, and decreased alcohol intake. Patients with geneticLPLdeficiency are instructed to prepare their food using medium-chain triglycerides, which are not incorporated into chylomicrons. Fish oils decrease triglyceride synthesis, and high doses may be used for severe hypertriglyceridemia.

The management of hypertriglyceridemia focuses on the associatedLDL andHDLconcentrations as guidelines for therapy. Thus, the overall risk profile can be used to set goals for LDL cholesterol, using a low HDL level (commonly associated with hypertriglyceridemia) as a concomitant major risk factor for atherosclerosis. However, when triglyceride levels are>5.6 mmol/L (>500 mg/dL), the risk of developing pancreatitis increases, and a direct focus on lowering triglycerides is recommended. Thus, triglyceride levels>5.6 mmol/L (>500 mg/dL) are generally treated with drugs, whereas lower levels [2.2 to 5.6 mmol/L (200 to 500 mg/dL)] are not treated unless otherCHD risk factors are present (Fig. 344-2).

TREATMENT

Three classes of lipid-lowering agents are recommended as first-line therapy against hypercholesterolemia: (1) the HMG-CoA reductase inhibitors; (2) niacin; and (3) the bile acid-binding resins (Table 344-8). Fibric acid derivatives are second-line agents for hypercholesterolemia and are most effective for lowering triglycerides.

HMG-CoA Reductase Inhibitors This class of drugs, which include lovastatin, simvastatin, pravastatin, fluvastatin, atorvastatin, and cerivastatin, inhibits the rate-limiting step in hepatic cholesterol biosynthesis (the conversion of HMG-CoA to mevalonate), causing an increase inLDLreceptor levels in hepatocytes and enhanced receptor-mediated clearance of LDL cholesterol from the circulation. At usual doses, the

HMG-CoA reductase inhibitors decrease total cholesterol by 20 to 30% and LDL cholesterol by 25 to 40%. Larger reductions may be achieved with higher doses. Treatment with reductase inhibitors often reduces triglycerides by 10 to 20%, possibly due to reduced secretion of VLDL by the liver. Higher doses of more potent reductase inhibitors, which can lower LDL cholesterol by 45 to 60%, can lower triglycerides by 30 to 45%. HDL cholesterol levels rise about 5 to 10%. In comparison with other lipid-lowering agents, HMG-CoA reductase inhibitors are relatively free of side effects. Mild, transient elevations in liver enzymes occur with all of the agents at the highest doses, but elevations in serum aminotransferases to more than three times the upper limits of normal occur in <2% of patients. Therapy should be discontinued when elevations of this magnitude occur. A rare but potentially serious adverse effect of HMG-CoA reductase inhibitors is myopathy, manifest by muscle pain with elevation of serum creatine phosphokinase (CPK). This occurs in <1% of patients treated with reductase inhibitors alone but is more common (about 2 to 3%) when used in combination with gemfibrozil, niacin, or cyclosporine.

Niacin The mechanism of action of niacin is not fully understood, but it appears to inhibit the secretion of lipoproteins containing apo B100 from the liver. Niacin decreases both total and LDL cholesterol approximately 15 to 25%, reduces VLDL levels by 25 to 35%, and raises HDL cholesterol levels by as much as 15 to 25%. Thus, niacin exerts favorable changes on the three major lipoproteins (VLDL, LDL, and HDL). Efficacy of monotherapy was confirmed in a long-term secondary prevention trial in which niacin significantly reduced the incidence of myocardial infarction. An even longer-term follow-up of that study (15 years total) showed an 11% decrease in all-cause mortality among patients randomized to niacin. Because of its ability to reduce VLDL synthesis, niacin is also a first-line drug for treatment of hypertriglyceridemia.

Niacin is safe, having been in use for almost 30 years, but unpleasant side effects, including cutaneous flushing with or without pruritus, may limit patient acceptability. The cutaneous symptoms tend to subside after several weeks and may be minimized by initiating therapy at low doses or by administering aspirin 30 min before the niacin dose. Less common adverse effects include elevations of liver enzymes, gastrointestinal distress, impaired glucose tolerance, and elevated serum uric acid levels with or without gouty arthritis. Liver enzymes may be elevated in 3 to 5% of patients on full doses of niacin (>2 g/d). Because of its propensity to worsen the control of blood sugar, niacin should be used with caution in patients with DM. Niaspan, an intermediate-release form of niacin, appears to exhibit lipid-altering activity similar to regular niacin.

Bile Acid-Binding Resins Cholestyramine and colestipol have been in use as lipid-lowering agents for almost three decades. These drugs interfere with reabsorption of bile acids in the intestine, resulting in a compensatory increase in bile acid synthesis and upregulation of LDL receptors in hepatocytes. The bile acid sequestrants are useful in the treatment of patients with elevated levels of LDL cholesterol and normal triglycerides. Sequestrants produce dose-dependent decreases on the order of 15 to 25% in total cholesterol and of 20 to 35% in LDL cholesterol. The agents cause modest increases in HDL cholesterol. A limitation of the sequestrants is their tendency to raise triglyceride levels through compensatory increases in hepatic synthesis of VLDL; they should not be given to hypertriglyceridemic individuals. Bile acid-binding resins are efficacious and safe and are recommended for young adult men and premenopausal

women with moderate cholesterol elevations. Patient compliance is low, in part because of the need to dissolve these powdered agents in fluid; the availability of colestipol as a tablet may alleviate this problem. Gastrointestinal side effects include constipation, bloating, and gas.

Combination Therapy Combinations of bile acid-binding resins and reductase inhibitors are effective for the treatment of severe, isolated elevations of LDL cholesterol. Combinations of reductase inhibitors and niacin, or resins and niacin, are useful for the treatment of high LDL and low HDL cholesterol levels, though the former combination carries an increased risk of myositis (2 to 3%). If triglyceride and LDL levels are both elevated (HDL is usually reduced as well), resins and niacin are an excellent combination, with resins and gemfibrozil (see below) as an alternative. The combination of a reductase inhibitor and gemfibrozil can be useful when LDL cholesterol is very high in the face of concomitant hypertriglyceridemia, but the risk of myositis (about 2 to 3%) must be considered. Combinations of reductase inhibitors with either niacin or gemfibrozil might best be reserved for patients with CHD and combined hyperlipidemia.

LDL Apheresis In patients with homozygous FH and in ordinary FH patients who respond poorly to diet and drug therapy or who cannot tolerate drugs, apheresis at 7- to 14-day intervals can cause profound lowering of LDL cholesterol levels. Diet and drug regimens are continued during treatment. This approach should be considered for patients with few therapeutic options.

Fibric Acids Gemfibrozil and fenofibrate stimulate the activity of a liver transcription factor termed *PPARa* that increases LPL activity and production of apo AI. Moreover, these drugs reduce VLDL triglyceride entry into plasma and reduce synthesis of apo CIII, which might improve LPL-induced lipolysis or reduce VLDL secretion. Stimulation of peroxisomal fatty acid oxidation by fibrates may also contribute to the triglyceride-lowering actions. Gemfibrozil and fenofibrate treatment is associated with 25 to 40% reductions in plasma triglyceride levels. Postprandial triglyceride levels, which are linked to fasting concentrations, are also reduced. HDL cholesterol levels increase 5 to 15% with fibrate treatment. Fibric acids and a low-fat diet are particularly useful in the treatment of dysbetalipoproteinemia and are first-line therapy for this disorder except in postmenopausal women, who should initially be given estrogen replacement (if not contraindicated).

Significant increases in LDL cholesterol can accompany otherwise potentially beneficial falls in triglycerides and increases in HDL cholesterol during fibrate therapy. Such rises may require a change to another drug or addition of a second agent.

In the short term, these drugs are well tolerated; mild gastrointestinal distress in the form of epigastric pain is the major side effect. Elevations of liver enzymes occur in 2 to 3% of patients but do not usually require cessation of treatment. Rarely, hepatitis can occur. Fibrates appear to make the bile more lithogenic, and long-term use is probably associated with a twofold increase in gallstone formation. Myopathy with myositis is a rare occurrence with the fibrates, either alone or in combination with HMG CoA reductase inhibitors.

Fish Oils Large doses of omega-3 fatty acids reduce triglyceride levels by diminishing

their production. In the United States, omega-3 fatty acid capsules contain 40 to 60% omega-3 fatty acids; the rest of the fatty acids are omega-6. Therefore, to consume 2 to 4 g of omega-3 fatty acids, an individual must take 5 to 10 capsules per day.

HYPOCHOLESTEROLEMIA

A low total cholesterol concentration [<2.6 mmol/L (<100 mg/dL)] in an adult can be due to rare hereditary traits or secondary to a number of diseases. As described earlier, mutations in the gene for apo B100 that disrupt synthesis or produce truncated forms of apo B100 are associated with hypobetalipoproteinemia. These mutations are inherited as codominant traits; heterozygotes have plasma cholesterol levels in the range of 1.3 to 2.6 mmol/L (50 to 100 mg/dL), with reduced LDL cholesterol levels but normal plasma HDL cholesterol levels. Heterozygotes are asymptomatic, whereas hypolipoproteinemia homozygotes (or compound heterozygotes) have even lower total and LDL cholesterol concentrations and may have malabsorption of fats and fat-soluble vitamins similar to that in abetalipoproteinemia.

Abetalipoproteinemia (Table 344-5) is a rare, autosomal recessive disorder in which there are mutations in the microsomal triglyceride transfer protein (MTP) gene. Individuals who are homozygous for this disorder have total cholesterol levels <1.3 mmol/L (<50 mg/dL) and essentially no VLDL, IDL, LDL, or chylomicrons. Because dietary fat and vitamins A and E are transported from the intestine in chylomicrons, these patients may have malabsorption of fat and fat-soluble vitamins. Vitamin E deficiency in infancy and early childhood can result in neurologic problems (Chap. 75). If vitamin replacement is adequate, individuals with abetalipoproteinemia can live normal, healthy lives.

Moderately low levels of total cholesterol may also be associated with extreme reductions in HDL cholesterol. As noted above, these are almost always secondary to mutations in the gene for apo AI and a lack of apo AI in plasma.

A number of systemic diseases can cause low cholesterol concentrations. Malnutrition, often associated with alcoholism or gastrointestinal disease, can cause low levels of total and LDL cholesterol. Hyperthyroidism, particularly when severe, can reduce cholesterol levels. Patients with uncontrolled AIDS may have total cholesterol levels<2.1 mmol/L (<80 mg/dL), usually associated with severe wasting, diarrhea, and a poor prognosis. Several neoplasms, particularly those involving the hematopoietic system, are associated with hypocholesterolemia. Patients with acute and chronic myelogenous leukemia and myeloid metaplasia with splenomegaly can have severe reductions in both LDL and HDL levels. Other diseases with concomitant splenomegaly, including lipid storage diseases such as Gaucher's disease and Niemann-Pick disease, can cause very low LDL and HDL cholesterol concentrations due to increased lipoprotein catabolism.

(Bibliography omitted in Palm version)

345. HEMOCHROMATOSIS - *Lawrie W. Powell, Kurt J. Isselbacher*

DEFINITION

Hemochromatosis is a common disorder of iron storage in which an appropriate increase in intestinal iron absorption results in deposition of excessive amounts of iron in parenchymal cells with eventual tissue damage and impaired function of organs, especially the liver, pancreas, heart, joints, and pituitary. The disease was termed *hemochromatosis* and the iron-storage pigment was called *hemosiderin* because it was believed that the pigment was derived from the blood. The terms *hemosiderosis* and *siderosis* are often used to describe the presence of stainable iron in tissues, but tissue iron must be quantified to assess body iron status (see below and Chap. 105). *Hemochromatosis* implies potentially severe progressive iron overload leading to fibrosis and organ failure. Cirrhosis of the liver, diabetes mellitus, arthritis, cardiomyopathy, and hypogonadotrophic hypogonadism are common manifestations.

Although there is debate about definitions, it seems logical to use the following terminology:

1. *Hereditary* or *genetic hemochromatosis*: This disorder is most often caused by inheritance of a mutant *HFE* gene, which is tightly linked to the HLA-A locus on chromosome 6p (see below). The genetic disease can be recognized during its early stages when iron overload and organ damage are minimal. At this stage the disease is best referred to as *early* or *precirrhotic hemochromatosis* (Fig. 345-1).

2. *Secondary iron overload*: Tissue injury usually occurs secondary to an iron-loading anemia such as thalassemia or sideroblastic anemia, in which increased erythropoiesis is ineffective. In the acquired iron-loading disorders, massive iron deposits in parenchymal tissues can lead to the same clinical and pathologic features as in hemochromatosis.

PREVALENCE

Hemochromatosis is one of the most common genetic diseases, although its prevalence varies in different ethnic groups. It is most common in populations of northern European extraction in whom approximately 1 in 10 persons are heterozygous carriers and 0.3 to 0.5% are homozygotes. However, expression of the disease is modified by several factors, especially dietary iron intake, blood loss associated with menstruation and pregnancy, and blood donation. The clinical expression of the disease is 5 to 10 times more frequent in men than in women. Nearly 70% of affected patients develop the first symptoms between ages 40 and 60. The disease is rarely evident before age 20, although with family screening (see below) and periodic health examinations, asymptomatic subjects with iron overload can be identified, including young menstruating women. A recent study in a European non-blood bank population revealed that 30% of homozygous individuals did not have evidence of iron overload. Thus, the penetrance of the mutation is variable.

GENETIC BASIS AND MODE OF INHERITANCE

The gene involved in the most common form of hemochromatosis was cloned in 1996 and is termed *HFE.* A homozygous G®A mutation resulting in a cysteine to tyrosine substitution at position 282 (C282Y) is the most common mutation. It was identified in 85 to 100% of patients with hereditary hemochromatosis in populations of northern European descent but was found in only 60% of cases from Mediterranean populations (e.g., southern Italy). A second, relatively common *HFE* mutation has also been identified. This results in an amino acid substitution of histidine to aspartic acid at position 63 (H63D). Some compound heterozygotes (e.g., one copy each of C282Y and H63D) have increased body iron stores. Thus, *HFE*-associated hemochromatosis is inherited as an autosomal recessive trait; heterozygotes have no, or minimal, increase in iron stores. However, in some cases this slight increase in hepatic iron acts as a cofactor that aggravates other diseases such as porphyria cutanea tarda (PCT) and nonalcoholic steatohepatitis.

Mutations in other genes, currently unidentified, are responsible for non-*HFE* associated hemochromatosis, including juvenile hemochromatosis, which affects subjects in the second and third decade of life (Table 345-1).

PATHOGENESIS

Normally, the body iron content of 3 to 4 g is maintained such that intestinal mucosal absorption of iron is equal to iron loss. This amount is approximately 1 mg/d in men and 1.5 mg/d in menstruating women. In hemochromatosis, mucosal absorption is inappropriate to body needs and amounts to 4 mg/d or more. The progressive accumulation of iron causes an early elevation in plasma iron, an increased saturation of transferrin, and progressive elevation of plasma ferritin level (Fig. 345-1).

The *HFE* gene encodes a 343 amino acid protein that is structurally related to MHC class I proteins. The basic defect in hemochromatosis is a lack of cell surface expression of HFE (due to the C282Y mutation). The normal (wild type) HFE protein forms a complex withb_2-microglobulin and transferrin, and the C282Y mutation completely abrogates this interaction. As a result, the mutant HFE protein remains trapped intracellularly, reducing transferrin receptor-mediated iron uptake by the intestinal crypt-cell. This is postulated to upregulate the divalent metal transporter (DMT-1) on the brush border of the villous cells, leading to inappropriately increased intestinal iron absorption. In advanced disease, the body may contain 20 g or more of iron that is deposited mainly in parenchymal cells of the liver, pancreas, and heart. Iron may be increased 50- to 100-fold in the liver and pancreas and 5- to 25-fold in the heart. Iron deposition in the pituitary causes hypogonadotropic hypogonadism in both men and women. Tissue injury may result from disruption of iron-laden lysosomes, from lipid peroxidation of subcellular organelles by excess iron, or from stimulation of collagen synthesis by activated stellate cells.

Secondary iron overload with deposition in parenchymal cells occurs in chronic disorders of erythropoiesis, particularly in those due to defects in hemoglobin synthesis or ineffective erythropoiesis such as sideroblastic anemia and thalassemia (Chap. 106). In these disorders, the absorption of iron is increased. Moreover, these patients require blood transfusions and are also frequently treated inappropriately with iron.PCT, a disorder characterized by a defect in porphyrin biosynthesis (Chap. 346), is also

sometimes associated with excessive parenchymal iron deposits. The magnitude of the iron load in PCT is usually insufficient to produce tissue damage. However, recent reports have found that many patients with PCT also have mutations in the *HFE* gene, and some have associated hepatitis C infection. Although the relationship among these disorders remains to be clarified, iron overload accentuates the inherited enzyme deficiency in PCT and should be avoided along with other agents (alcohol, estrogens, haloaromatic compounds) that may exacerbate PCT. Another cause of hepatic parenchymal iron overload is hereditary aceruloplasminemia. In this disorder, impairment of iron mobilization due to deficiency of ceruloplasmin (a ferroxidase) causes iron overload in hepatocytes.

Alcoholic subjects with end-stage chronic liver disease may have increased tissue iron stores of the degree seen in hemochromatosis. The increased iron may be caused by cell death and uptake of the released iron, as well as by hemolysis associated with spur-cell anemia (Chap. 108). Hemochromatosis in a heavy drinker can be distinguished from alcoholic liver disease by the presence of the C282Y mutation.

Excessive iron ingestion over many years rarely results in hemochromatosis. An important exception has been reported in South Africa among groups who brew fermented beverages in vessels made of iron. Hemochromatosis has on occasion been described in apparently normal subjects who have taken medicinal iron over many years, but such individuals probably have a genetic disorder.

The common denominator in all patients with hemochromatosis is *excessive amounts of iron in parenchymal tissues*. Parenteral administration of iron in the form of blood transfusions or iron preparations results predominantly in reticuloendothelial cell iron overload. This appears to lead to less tissue damage than iron loading of parenchymal cells.

PATHOLOGY

At autopsy, the enlarged nodular liver and pancreas are rusty in color. Histologically, iron is increased in amount in many organs, particularly in the liver, heart, and pancreas, and to a lesser extent in the endocrine glands. The epidermis of the skin is thin, and melanin is increased in the cells of the basal layer. Deposits of iron are present around the synovial lining cells of the joints.

In the liver of patients with hemochromatosis, parenchymal iron is in the form of ferritin and hemosiderin. In the early stages these deposits are seen in the periportal parenchymal cells, especially within lysosomes in the pericanalicular cytoplasm of the hepatocytes. This stage progresses to perilobular fibrosis and eventually to deposition of iron in bile duct epithelium, Kupffer cells, and fibrous septa. In the advanced stage, a macronodular or mixed macro- and micronodular cirrhosis develops.

CLINICAL MANIFESTATIONS

Initial symptoms include weakness, lassitude, weight loss, change in skin color, abdominal pain, loss of libido, and symptoms of diabetes mellitus. Hepatomegaly, increased pigmentation, spider angiomas, splenomegaly, arthropathy, ascites, cardiac

arrhythmias, congestive heart failure, loss of body hair, testicular atrophy, and jaundice are prominent in advanced disease.

The *liver* is usually the first organ to be affected, and hepatomegaly is present in more than 95% of symptomatic patients. Hepatic enlargement may exist in the absence of symptoms or of abnormal liver function tests. Indeed, over half of patients with symptomatic hemochromatosis have little laboratory evidence of functional impairment of the liver, in spite of hepatomegaly and fibrosis. Loss of body hair, palmar erythema, testicular atrophy, and gynecomastia are common. Manifestations of portal hypertension and esophageal varices occur less commonly than in cirrhosis from other causes. Hepatocellular carcinoma develops in about 30% of patients with cirrhosis, and it is the most common cause of death in treated patients; hence the importance of early diagnosis and therapy. Its incidence increases with age, is more common in men, and occurs almost exclusively in cirrhotic patients. Splenomegaly occurs in approximately half of symptomatic cases.

Excessive skin pigmentation is present in over 90% of symptomatic patients at the time of diagnosis. The characteristic metallic or slate gray hue is sometimes referred to as bronzing and results from increased melanin and iron in the dermis. Pigmentation usually is diffuse and generalized, but it may be more pronounced on the face, neck, extensor aspects of the lower forearms, dorsa of the hands, lower legs, genital regions, and in scars.

Diabetes mellitus occurs in about 65% of patients and is more likely to develop in those with a family history of diabetes, suggesting that direct damage to the pancreatic islets by iron deposition occurs in combination with a genetic predisposition. The management is similar to that of other forms of diabetes, although pronounced insulin resistance is more common in association with hemochromatosis. Late complications are the same as seen in other causes of diabetes mellitus.

Arthropathy develops in 25 to 50% of patients. It usually occurs after age 50, but may occur as a first manifestation, or long after therapy. The joints of the hands, especially the second and third metacarpophalangeal joints, are usually the first joints involved, a feature that helps to distinguish the chondrocalcinosis associated with hemochromatosis from the idiopathic form (Chap. 322). A progressive polyarthritis involving wrists, hips, ankles, and knees also may ensue. Acute brief attacks of synovitis may be associated with deposition of calcium pyrophosphate (chondrocalcinosis or pseudogout), mainly in the knees. Radiologic manifestations include cystic changes of the subchondral bones, loss of articular cartilage with narrowing of the joint space, diffuse demineralization, hypertrophic bone proliferation, and calcification of the synovium. The arthropathy tends to progress despite removal of iron by phlebotomy. Although the relation of these abnormalities to iron metabolism is not known, the fact that similar changes occur in other forms of iron overload suggests that iron is directly involved.

Cardiac involvement is the presenting manifestation in about 15% of patients. The most common manifestation is congestive heart failure, which occurs in about 10% of young adults with the disease, especially those with juvenile hemochromatosis. Symptoms of congestive failure may develop suddenly, with rapid progression to death if untreated. The heart is diffusely enlarged and may be misdiagnosed as idiopathic cardiomyopathy

if other overt manifestations are absent. Cardiac arrhythmias include premature supraventricular beats, paroxysmal tachyarrhythmias, atrial flutter, atrial fibrillation, and varying degrees of atrioventricular block.

Hypogonadism occurs in both sexes and may antedate other clinical features. Manifestations include loss of libido, impotence, amenorrhea, testicular atrophy, gynecomastia, and sparse body hair. These changes are primarily the result of decreased production of gonadotropins due to impairment of hypothalamic-pituitary function by iron deposition; however, primary testicular dysfunction may be seen in some cases. Adrenal insufficiency, hypothyroidism, and hypoparathyroidism may also occur.

DIAGNOSIS

The association of (1) hepatomegaly, (2) skin pigmentation, (3) diabetes mellitus, (4) heart disease, (5) arthritis, and (6) hypogonadism should suggest the diagnosis. However, a parenchymal iron overload of comparatively short duration or modest degree may exist with none or only some of these manifestations [e.g., in young subjects (Fig. 345-1)]. Therefore, a high index of suspicion is needed to make the diagnosis early. This is particularly important because treatment before there is permanent organ damage can reverse the iron toxicity and restore life expectancy to normal (see below).

The history should be particularly detailed in regard to disease in other family members, alcohol ingestion, iron intake, and ingestion of large doses of ascorbic acid, which promotes iron absorption (Chap. 75). Appropriate tests should be performed to exclude iron deposition due to hematologic disease. The presence of liver, pancreatic, cardiac, and joint disease should be confirmed by physical examination, roentgenography, and standard function tests of these organs. The degree of increase in total-body iron stores should be assessed with particular attention to an increase in parenchymal iron concentration, with or without tissue damage.

The methods available for assessing parenchymal iron stores include (1) measurement of serum iron and the percent saturation of transferrin (or the unsaturated iron-binding capacity); (2) measurement of serum ferritin concentration; (3) liver biopsy with measurement of the iron concentration and calculation of the hepatic iron index (Table 345-2), (4) estimation of chelatable iron stores following the administration of deferoxamine; and (5) computed tomography (CT) and/or magnetic resonance imaging (MRI) of the liver. Each has its advantages and limitations. The serum iron level and percent saturation of transferrin are elevated early in the course, but their specificity is reduced by significant false-positive and false-negative rates. For example, serum iron concentration may be increased in patients with alcoholic liver disease without iron overload; in this situation, however, the hepatic iron index is usually not increased as in hemochromatosis (Table 345-1). In otherwise healthy persons, a fasting serum transferrin saturation greater than 50% is abnormal and suggests homozygosity for hemochromatosis.

The serum ferritin concentration is usually a good index of body iron stores, whether decreased or increased. In fact, an increase of 1 ug/L in serum ferritin level reflects an

increase of about 65 mg in body stores. In most untreated patients with hemochromatosis, the serum ferritin level is greatly increased (Fig. 345-1 andTable 345-1). However, in patients with inflammation and hepatocellular necrosis, serum ferritin levels may be elevated out of proportion to body iron stores due to increased release from tissues. A repeat determination of serum ferritin should therefore be carried out after acute hepatocellular damage has subsided, e.g., in alcoholic liver disease. Ordinarily, the combined measurements of the percent transferrin saturation and serum ferritin level provide a simple and reliable screening test for hemochromatosis, including the precirrhotic phase of the disease. If either of these tests is abnormal, genetic testing for hemochromatosis should be performed (Fig. 345-2).

The role of liver biopsy in the diagnosis and management of hemochromatosis is being reassessed as a result of the widespread availability of genetic testing for the C282Y mutation. The absence of severe fibrosis can be accurately predicted in most patients using clinical and biochemical variables. Thus, there is virtually no risk of severe fibrosis in a C282Y homozygous subject with: (1) serum ferritin level less than 1000 ug/L; (2) normal serum alanine amino transaminase values; (3) no hepatomegaly; and (4) no excess alcohol intake. However, it should be emphasized that liver biopsy is the only reliable method for establishing or excluding the presence of hepatic cirrhosis, which is the critical factor determining prognosis and the risk of developing hepatocellular carcinoma. Biopsy also permits histochemical estimation of tissue iron and measurement of hepatic iron concentration. Increased density of the liver due to iron deposition can be demonstrated byCT orMRI. A retrospective assessment of body iron storage is also provided by performing weekly phlebotomy and calculating the amount of iron removed before iron stores are exhausted (1 mL blood = approximately 0.5 mg iron).

SCREENING FOR HEMOCHROMATOSIS

When the diagnosis of hemochromatosis is established, it is important to counsel and screen other family members (Chap. 68). Asymptomatic as well as symptomatic family members with the disease usually have an increased saturation of transferrin and an increased serum ferritin concentration. These changes occur even before the iron stores are greatly increased (Fig. 345-1). All first-degree relatives of patients with hemochromatosis should be tested for the C282Y and H63D mutations and advised appropriately. In affected individuals, it is important to confirm or exclude the presence of cirrhosis, and begin therapy as early as possible. When children of a proband are affected, a homozygote-heterozygote mating is most likely.

The role of population screening for hemochromatosis is controversial. Hemochromatosis fulfills the criteria established by the World Health Organization for population screening, and DNA testing could, in principle, be performed along with other neonatal tests. However, because iron overload does not develop until the second, third, or fourth decades, and the degree of penetrance is still uncertain, screening by phenotypic expression is more practical at present (Fig. 345-2).

TREATMENT

The therapy of hemochromatosis involves removal of the excess body iron and

supportive treatment of damaged organs. Iron removal is best begun by weekly or twice-weekly phlebotomy of 500 mL. Although there is an initial modest decline in the volume of packed red blood cells to about 35 mL/dL, the level stabilizes after several weeks. The plasma transferrin saturation remains increased until the available iron stores are depleted. In contrast, the plasma ferritin concentration falls progressively, reflecting the gradual decrease in body iron stores. Since one 500-mL unit of blood contains 200 to 250 mg iron and about 25 g iron should be removed, weekly phlebotomy may be required for 1 or 2 years. When the transferrin saturation and ferritin level become normal, phlebotomies are performed at appropriate intervals to maintain levels within the normal range. The measurements promptly become abnormal with iron reaccumulation. Usually one phlebotomy every 3 months will suffice.

Chelating agents such as deferoxamine, when given parenterally, remove 10 to 20 mg iron per day, which is much less than that mobilized by once-weekly phlebotomy. Phlebotomy is also less expensive, more convenient, and safer for most patients. However, chelating agents are indicated when anemia or hypoproteinemia is severe enough to preclude phlebotomy. Subcutaneous infusion of deferoxamine using a portable pump is the most effective means of administration.

The management of hepatic failure, cardiac failure, and diabetes mellitus is similar to conventional therapy for these conditions. Loss of libido and change in secondary sex characteristics are partially relieved by parenteral testosterone or gonadotropin therapy (Chap. 335).

PROGNOSIS

The principal causes of death in untreated patients are cardiac failure (30%), hepatocellular failure or portal hypertension (25%), and hepatocellular carcinoma (30%).

Life expectancy is improved by removal of the excessive stores of iron and maintenance of these stores at near-normal levels. The 5-year survival rate with therapy increases from 33 to 89%. With repeated phlebotomy, the liver and spleen decrease in size, liver function improves, pigmentation of skin decreases, and cardiac failure may be reversed. Diabetes improves in about 40%, but removal of excess iron has little effect on hypogonadism or arthropathy. Hepatic fibrosis may decrease, but cirrhosis is irreversible. End-stage liver disease can be treated with orthotopic liver transplantation, but the results are suboptimal unless excess iron stores are first corrected. Hepatocellular carcinoma usually occurs as a late sequela in patients who are cirrhotic at presentation. The apparent increase in its incidence in treated patients is probably related to their increased life span. Hepatocellular carcinoma does not appear to develop if the disease is treated in the precirrhotic stage. Indeed, the life expectancy of homozygotes treated before the development of cirrhosis is normal.

The importance of family screening and early therapy cannot be emphasized too strongly. Asymptomatic subjects detected by family studies should have phlebotomy therapy if iron stores are moderately to severely increased. Assessment of iron stores at appropriate intervals is also important. With this management approach, most manifestations of the disease can be prevented.

(Bibliography omitted in Palm version)

346. THE PORPHYRIAS - *Robert J. Desnick*

The porphyrias are inherited or acquired disorders of specific enzymes in the heme biosynthetic pathway (Fig. 346-1). These disorders are classified as either *hepatic* or *erythropoietic* depending on the primary site of overproduction and accumulation of the porphyrin precursor or porphyrin (Tables 346-1 and 346-2), although some have overlapping features. The major manifestations of the hepatic porphyrias are neurologic, including neuropathic abdominal pain, neuropathy, and mental disturbances, whereas the erythropoietic porphyrias characteristically cause cutaneous photosensitivity. The reason for neurologic involvement in the hepatic porphyrias is poorly understood. Cutaneous sensitivity to sunlight is due to the fact that excitation of excess porphyrins in the skin by long-wave ultraviolet light leads to cell damage, scarring, and deformation. Steroid hormones, drugs, and nutrition influence the production of porphyrin precursors and porphyrins, thereby precipitating or increasing the severity of some porphyrias. Thus, the porphyrias are multifactorial genetic disorders, in which environmental, physiologic, and genetic factors interact to cause disease (Chap. 68).

Because many symptoms of the porphyrias are nonspecific, diagnosis is often delayed. Laboratory testing is required to confirm or exclude the various types of porphyria (Table 346-2). Urinary d-aminolevulinic acid (ALA) and porphobilinogen (PBG) are easily quantitated by chemical methods; the urinary porphyrin isomers can be separated and quantitated by high-performance liquid chromatography. The diagnostic profile of accumulated precursors and/or porphyrins in each disorder can also be defined by extraction and thin-layer chromatography of fecal porphyrins. However, a definite diagnosis requires demonstration of the specific enzyme deficiency or gene defect. The isolation and characterization of the cDNAs encoding the heme biosynthetic enzymes have permitted identification of the genetic basis of each porphyria. Molecular genetic analyses now make it possible to provide precise heterozygote identification and prenatal diagnoses in families with known mutations or with informative polymorphisms.

HEME BIOSYNTHESIS

The first and last three enzymes in the heme biosynthetic pathway are located in the mitochondrion, whereas the other four are in the cytosol (Fig. 346-1). The first enzyme, d-aminolevulinate synthase (ALA synthase), catalyzes the condensation of glycine, activated by pyridoxal phosphate and succinyl coenzyme A, to form ALA. In the liver, this rate-limiting enzyme can be induced by a variety of drugs, steroids, and other chemicals. Distinct erythroid-specific and nonerythroid (e.g., housekeeping) forms of ALA synthase are encoded by separate genes; defects in the erythroid form cause X-linked sideroblastic anemia (XLSA).

The second enzyme, d-aminolevulinate dehydratase (ALA dehydratase), catalyzes the condensation of two molecules of ALA to form PBG. Four molecules of PBG condense to form the tetrapyrrole uroporphyrinogen (URO) III by a two-step process catalyzed by hydroxymethylbilane (HMB) synthase (also known as PBG deaminase or URO I synthase) and URO III synthase. HMB synthase catalyzes the head-to-tail condensation of four PBG molecules by a series of deaminations to form the linear tetrapyrrole HMB. URO synthase catalyzes the rearrangement and rapid cyclization of HMB to form the asymmetric, physiologic, octacarboxylate porphyrinogen URO III isomer.

The fifth enzyme in the pathway, UROdecarboxylase, catalyzes the sequential removal of the four carboxyl groups from the acetic acid side chains of URO III to form coproporphyrinogen (COPRO) III, a tetracarboxylate porphyrinogen. This compound then enters the mitochondrion, where COPRO oxidase, the sixth enzyme, catalyzes the decarboxylation of two of the four propionic acid groups to form the two vinyl groups of protoporphyrinogen (PROTO) IX, a dicarboxylate porphyrinogen. Next, PROTO oxidase oxidizes PROTO IX to protoporphyrin IX by the removal of six hydrogen atoms. The product of the reaction is a porphyrin (oxidized form), in contrast to the preceding tetrapyrrole intermediates, which are porphyrinogens (reduced forms). Finally, ferrous iron is inserted into protoporphyrin IX to form heme, a reaction catalyzed by the eighth enzyme in the pathway, ferrochelatase (also known as heme synthetase or protoheme ferrolyase).

Each of the heme biosynthetic enzymes is encoded by a separate gene. Full-length human cDNAs for each of the enzymes, including those for both forms of ALAsynthase, have been isolated and sequenced, and the chromosomal locations of the genes have been identified (Table 346-3).

REGULATION OF HEME BIOSYNTHESIS

About 85% of the heme produced in the body is synthesized in erythroid cells to provide heme for hemoglobin; most of the remainder is produced in the liver, where the biosynthetic pathway is under negative feedback control (Chap. 104). "Free" heme in the liver regulates the synthesis and mitochondrial translocation of the housekeeping form of ALAsynthase. Heme represses the synthesis of the ALA synthase mRNA and interferes with the transport of the enzyme from the cysotol into mitochondria. ALA synthase is increased by many of the same chemicals that induce the cytochrome P450 enzymes in the endoplasmic reticulum of the liver. Because most of the heme in the liver is used for the synthesis of cytochrome P450 enzymes, hepatic ALA synthase and the cytochrome P450s are regulated in a coordinated fashion.

Different regulatory mechanisms control production of heme for hemoglobin. The erythroid-specific ALAsynthase encoded on the X chromosome is expressed at higher levels than the hepatic enzyme, and an erythroid-specific control mechanism regulates iron transport into erythroid cells. During erythroid differentiation, the activities of the heme biosynthetic enzymes are increased.

THE HEPATIC PORPHYRIAS

The acute hepatic porphyrias are characterized by the rapid onset of neurologic manifestations. During the acute attack, individuals have markedly elevated plasma and urinary concentrations of the porphyrin precursors ALA and PBG, which originate from the liver.

ALA DEHYDRATASE-DEFICIENT PORPHYRIA

This is a rare autosomal recessive disorder that has been described in only a few patients. Onset and severity of the disease are variable, presumably depending on the

amount of residual ALA dehydratase activity. Treatment and prevention of the neurologic complications are the same as for other acute porphyrias (see below).

Clinical Features The clinical presentation is variable. The first reported cases were in two unrelated German men who had clinical onset during adolescence of abdominal pain and neuropathy, resembling acute intermittent porphyria (AIP; see below). A Swedish infant presented with failure to thrive and required transfusions and parenteral nutrition. Presumably, the earlier age of onset and more severe manifestations reflect a more complete enzyme deficiency. A Belgian man developed an acute motor polyneuropathy and polycythemia at age 63. Recently a Japanese woman was described who had her first acute attack and the syndrome of inappropriate secretion of antidiuretic hormone at age 69.

Diagnosis Patients have increased urinary levels of ALA and coproporphyrin. ALA dehydratase activity in erythrocytes is <5% of normal. Because either succinylacetone (which accumulates in hereditary tyrosinemia and is structurally similar to ALA) or lead can inhibit ALA dehydratase, increase urinary excretion of ALA, and cause manifestations that resemble those of the acute porphyrias, lead intoxication and hereditary tyrosinemia (fumarylacetoacetase deficiency) should be considered in the differential diagnosis of ALA dehydratase-deficient porphyria. Immunologic studies in the reported cases demonstrated the presence of nonfunctional enzyme proteins that cross-reacted with anti-ALA dehydratase antibodies. DNA analysis revealed different missense mutations that resulted in the amino acid substitutions G133R and V275M in the infantile-onset patient, and R240W and A274T in a juvenile-onset patient.

Heterozygotes are clinically asymptomatic and do not excrete increased levels of ALA, but they can be detected by demonstration of intermediate levels of erythrocyte ALA dehydratase activity or by demonstrating a specific mutation in the *ALA dehydratase* gene. Prenatal diagnosis of this disorder has not been achieved but should be possible by determination of the ALA dehydratase activity in cultured chorionic villi or amniocytes.

TREATMENT

Treatment is similar to that of AIP (see below). The severely affected infant was supported by hyperalimentation and periodic blood transfusions. Continued failure to thrive led to liver transplantation, which did not improve the hematologic manifestations.

ACUTE INTERMITTENT PORPHYRIA

This hepatic porphyria is an autosomal dominant condition resulting from the half-normal level of HMB synthase (also termed PBG deaminase) activity. The disease is widespread but is especially common in Scandinavia and perhaps Great Britain. The enzyme deficiency can be demonstrated in most heterozygous individuals, but clinical expression is highly variable. Activation of the disease is related to environmental or hormonal factors, such as drugs, diet, and steroid hormones, which can precipitate the manifestations. Attacks can be prevented by avoiding known precipitating factors.

Clinical Features Most heterozygotes remain clinically asymptomatic (latent) unless

exposed to factors that increase the production of porphyrins. Endogenous and exogenous gonadal steroids, porphyrinogenic drugs, alcohol ingestion, and low-calorie diets, usually instituted for weight loss, are common precipitating factors. Table 346-4 lists the major drugs that are harmful in AIP [and also in hereditary coproporphyria (HCP) and variegate porphyria (VP)] and some drugs and anesthetic agents known to be safe. More extensive lists of drugs considered harmful or safe are available (see the bibliography), but information is incomplete for many of them. Attacks also can be provoked by infections and by surgery.

Because the neurovisceral symptoms rarely occur before puberty and are often nonspecific, a high index of suspicion is required to make the diagnosis. The disease can be disabling but is rarely fatal. Abdominal pain, the most common symptom, is usually steady and poorly localized but may be cramping. Ileus, abdominal distention, and decreased bowel sounds are common. However, increased bowel sounds and diarrhea may occur. Abdominal tenderness, fever, and leukocytosis are usually absent or mild because the symptoms are neurologic rather than inflammatory. Nausea, vomiting, constipation, tachycardia, hypertension, mental symptoms, pain in the limbs, head, neck, or chest, muscle weakness, sensory loss, dysuria, and urinary retention are characteristic. Tachycardia, hypertension, restlessness, tremors, and excess sweating are due to sympathetic overactivity.

The peripheral neuropathy is due to axonal degeneration (rather than demyelinization) and primarily affects motor neurons. Significant neuropathy does not occur with all acute attacks; abdominal symptoms are usually more prominent. Motor neuropathy affects the proximal muscles initially, more often in the shoulders and arms. The course and degree of involvement are variable. Deep tendon reflexes may be normal or hyperactive but are usually decreased or absent with advanced neuropathy. Motor weakness can be asymmetric and focal and may involve cranial nerves. Sensory changes such as paresthesia and loss of sensation are less prominent. Progressive muscle weakness can lead to respiratory and bulbar paralysis and death when diagnosis and treatment are delayed. Sudden death may result from sympathetic overactivity and cardiac arrhythmia.

Mental symptoms such as anxiety, insomnia, depression, disorientation, hallucinations, and paranoia can occur in acute attacks. Seizures can be due to neurologic effects or to hyponatremia. Treatment of seizures is difficult because virtually all antiseizure drugs (except bromides) may exacerbate AIP (clonazepam may be safer than phenytoin or barbiturates). Hyponatremia results from hypothalamic involvement and inappropriate vasopressin secretion or from electrolyte depletion due to vomiting, diarrhea, poor intake, or excess renal sodium loss. Persistent hypertension and impaired renal function may occur. When an attack resolves, abdominal pain may disappear within hours, and paresis begins to improve within days and may continue to improve over several years.

Diagnosis ALA and PBG levels are increased in plasma and urine during acute attacks. Urinary PBG excretion is usually 220 to 880 umol/d (50 to 200 mg/d) [normal, 0 to 18 umol/d (0 to 4 mg/d)], and urinary ALA excretion is 150 to 760 umol/d (20 to 100 mg/d) [normal, 8 to 53 umol/d (1 to 7 mg/d)]. The excretion of these compounds generally decreases with clinical improvement, particularly after hematin infusions (see below). A normal urinary PBG level effectively excludes AIP as a cause for current symptoms.

Fecal porphyrins are usually normal or minimally increased in AIP, in contrast to HCP and VP. Most asymptomatic ("latent") heterozygotes with HMB synthase deficiency have normal urinary excretion of ALA and PBG. Therefore, measurement of HMB synthase in erythrocytes is useful to confirm the diagnosis and to screen asymptomatic family members.

The enzyme deficiency is detectable in erythrocytes from most AIP heterozygotes (*classic AIP*). Note that the activity is higher in young erythrocytes and may increase into the normal range in AIP when erythropoiesis is increased due to a concurrent condition. However, patients with the rare erythroid form of AIP (*erythroid, or variant, AIP*) have normal enzyme levels in erythrocytes and deficient activity in nonerythroid tissues (see below). The erythroid and housekeeping forms of HMB synthase are encoded by a single gene, which has two promoters: one promoter generates the ubiquitously expressed housekeeping mRNA; the other promoter transcribes the erythroid-specific mRNA. Several deletions and over 150 different point mutations have been found in the coding region of the gene in unrelated AIP families (Fig. 346-2). These mutations alter the kinetic properties and/or stability of the mutant enzymes or create premature termination codons. Mutations that cause erythroid AIP variants with half-normal enzyme in nonerythroid tissues, but normal activity in erythrocytes, include point mutations in the initiation methionine codon (which prevent translation) or in the 5¢ donor splice site of intron 1 (which cause abnormal splicing of the HMB synthase transcript).

Heterozygotes can be identified using various polymorphic sites in the *HMB synthase* gene. Efforts are now under way to identify the specific mutations in the *HMB synthase* gene in all AIP families; this information will make it possible to identify all heterozygotes in affected families and to advise them to avoid the factors that cause acute attacks. The prenatal diagnosis of a fetus at risk can be made with cultured amniotic cells or chorionic villi.

TREATMENT

During acute attacks, narcotic analgesics may be required for abdominal pain, and phenothiazines are useful for nausea, vomiting, anxiety, and restlessness. Chloral hydrate can be given for insomnia, and benzodiazepines are probably safe in low doses, if a minor tranquilizer is required. Although intravenous glucose (at least 300 g/d) can be effective in acute attacks of porphyria, a more complete parenteral nutritional regimen may be beneficial if oral feeding is not possible for a prolonged period. However, intravenous heme is more effective than glucose in reducing porphyrin precursor excretion and probably leads to more rapid recovery. The response to heme therapy is reduced if therapy is delayed. Therefore, 3 to 4 mg of heme, in the form of hematin (Abbott Laboratories), heme albumin, or heme arginate (Leiras Oy, Turku, Finland), may be infused daily for 4 days beginning as soon as possible after onset of an attack. Heme arginate and heme albumin are chemically stable and are less likely than hematin to produce phlebitis or an anticoagulant effect. The rate of recovery from an acute attack depends on the degree of neuronal damage and may be rapid (1 to 2 days) with prompt therapy. Recovery from severe motor neuropathy may require months or years. Identification and avoidance of inciting factors can hasten recovery from an attack and prevent future attacks. Multiple inciting factors may contribute to a

symptomatic episode. Frequent clear-cut cyclical attacks occur in some women and can be prevented with a long-acting gonadotropin-releasing hormone analogue (this indication is not approved by the U.S. Food and Drug Administration) ([Chap. 336](#)).

PORPHYRIA CUTANEA TARDA

Porphyria cutanea tarda (PCT), the most common of the porphyrias, can be sporadic (type I) or familial (types II and III) and can also develop after exposure to halogenated aromatic hydrocarbons. Hepatic URO decarboxylase is deficient in all types of PCT. In type I PCT, URO decarboxylase activity is normal in erythrocytes. In type II PCT, an autosomal dominant disorder, the enzyme is deficient in erythrocytes and other tissues. In type III PCT, deficiency of the enzyme is limited to the liver. Deficient hepatic URO decarboxylase and a porphyrin pattern resembling PCT can be produced by exposure of normal individuals to a number of halogenated aromatic hydrocarbons. Hepatoerythropoietic porphyria (HEP) is an autosomal recessive form of porphyria that results from marked systemic deficiency of URO decarboxylase activity.

Clinical Features Cutaneous photosensitivity is the major clinical feature. Neurologic manifestations are not observed. Fluid-filled vesicles and bullae develop on sun-exposed areas such as the face, the dorsa of the hands and feet, the forearms, and the legs. The skin in these areas is friable, and minor trauma may lead to the formation of bullae. The appearance of small white plaques, termed *milia*, may precede or follow vesicle formation. Bullae and denuded skin heal slowly and are subject to infection. Other features include hypertrichosis and hyperpigmentation, especially of the face, and thickening, scarring, and calcification resembling the cutaneous changes of systemic sclerosis.

A number of factors contribute to the development of hepatic URO decarboxylase deficiency, including excess alcohol, iron, and estrogens. The importance of excess hepatic iron as a precipitating factor is underscored by the finding that the incidence of the common hemochromatosis-causing mutations, *HFE* C282Y and H63D, are increased in patients with types I and II PCT ([Chap. 345](#)). Various chemicals can also induce PCT; an epidemic of PCT occurred in eastern Turkey in the 1950s as a consequence of wheat contaminated with the fungicide hexachlorobenzene. PCT also occurs after exposure to other chemicals, including di- and trichlorophenols and 2,3,7,8-tetrachlorodibenzo-(*p*)-dioxin (TCDD, dioxin). Patients with PCT characteristically have liver damage and are at risk for hepatocellular carcinoma. These carcinomas do not produce porphyrins.

HEP resembles congenital erythropoietic porphyria (CEP) and usually presents with blistering skin lesions, hypertrichosis, scarring, and red urine in infancy or childhood.

Diagnosis Porphyrins are increased in the liver, plasma, urine, and stool. The urinary ALA level may be slightly increased, but the PBG level is normal. Urinary porphyrins consist mostly of uroporphyrin and 7-carboxylate porphyrin, with lesser amounts of coproporphyrin and 5- and 6-carboxylate porphyrins. Plasma porphyrins are also increased in a pattern that resembles that in urine. Isocoproporphyrins are increased in feces and sometimes in plasma and urine. The finding of increased isocoproporphyrins is diagnostic for a deficiency of hepatic URO decarboxylase.

Type II PCT and HEP can be distinguished by finding decreased URO decarboxylase in erythrocytes. URO decarboxylase activity in liver, erythrocytes, and cultured skin fibroblasts in type II PCT is approximately 50% of normal in affected individuals and in family members with latent disease. In HEP, the URO decarboxylase activity is markedly deficient, with typical levels of 3 to 10% of normal. Several point mutations have been identified in the coding region of the *URO decarboxylase* gene from unrelated type II PCT and HEP patients (Fig. 346-3). Excess hepatic iron contributes to development of sporadic and familial forms of PCT. As noted above, coinheritance of *HFE* mutations that cause hemochromatosis increases susceptibility to PCT-precipitating factor. In the familial forms (types II and III), iron inhibits the residual normal enzyme, so that enzymatic activity in liver is<50% of normal. In type I PCT the decreased hepatic URO decarboxylase activity is not accompanied by a decrease in the amount of enzyme protein, suggesting that the enzyme is present in an inactive form; hepatic URO decarboxylase activity gradually increases after a remission is induced by phlebotomy.

TREATMENT

Alcohol, estrogens, iron supplements, and, if possible, any drugs that may exacerbate the disease should be discontinued, but this step does not always lead to improvement. A complete response can almost always be achieved by repeated phlebotomy to reduce hepatic iron. A unit (450 mL) of blood can be removed every 1 to 2 weeks. Because iron overload is not marked in most cases, remission may occur after only five or six phlebotomies. Hemoglobin levels or hematocrits and serum ferritin should be followed closely to prevent development of iron deficiency and anemia. After remission, continued phlebotomy may not be needed even if ferritin levels return to normal. Relapses are treated by additional phlebotomy.

PCT can also be treated with chloroquine or hydroxychloroquine, both of which complex with the excess porphyrins and promote their excretion. Small doses (e.g., 125 mg chloroquine phosphate twice weekly) should be given, because standard doses can induce transient, sometimes marked increases in photosensitivity and hepatocellular damage. Hepatic imaging can diagnose or exclude complicating hepatocellular carcinoma. Treatment of PCT in patients with end-stage renal disease is facilitated by administration of erythropoietin.

HEREDITARY COPROPORPHYRIA

HCP is an autosomal dominant form of hepatic porphyria that results from half-normal levels of COPRO oxidase activity. Photosensitivity may occur. A few cases of homozygous HCP have been reported.

Clinical Features HCP is influenced by the same factors that cause attacks in AIP. The disease is latent before puberty, and symptoms are more common in women. Neurovisceral symptoms and other manifestations are virtually identical to those of AIP. Photosensitivity may resemble that in PCT and VP. Cutaneous lesions may begin in childhood in rare homozygous cases.

Diagnosis Coproporphyrin is markedly increased in the urine and feces in symptomatic disease and sometimes when there are no symptoms. Urinary ALA and PBG levels are increased during acute attacks but may return to normal when symptoms resolve. Although the diagnosis can be confirmed by measuring COPRO oxidase activity, these assays are not widely available and require cells other than erythrocytes.

TREATMENT

Neurologic symptoms are treated as in AIP (see above). Phlebotomy and chloroquine are ineffective when cutaneous lesions are present.

VARIEGATE PORPHYRIA

VP, a hepatic porphyria that results from the deficient activity of PROTO oxidase, is transmitted in an autosomal dominant manner and can present with neurologic symptoms, photosensitivity, or both.

Clinical Features Neurovisceral signs and symptoms develop after puberty and are similar to those of AIP or HCP (see above). Attacks are provoked by the same drugs, steroids, and nutritional factors that are detrimental in AIP. Skin manifestations are more common than in HCP but usually occur apart from the neurovisceral symptoms. Because the skin lesions in VP, HCP, and PCT are not distinguishable by clinical examination or biopsy, these conditions must be diagnosed by assay of porphyrins and porphyrin precursors in blood, urine, and feces.

VP is particularly common in South Africa, where 3 of every 1000 whites have the disorder. Most are descendants of a couple who emigrated from Holland to South Africa in 1688. Homozygous VP is associated with photosensitivity, neurologic symptoms, and developmental disturbances, including growth retardation, in infancy or childhood; all cases had increased erythrocyte levels of zinc protoporphyrin, a characteristic finding in all homozygous porphyrias so far described.

Dual porphyria, the simultaneous occurrence of VP and familial PCT, has been documented in several kindreds. *Chester porphyria* was described in a large British family in which individuals had acute porphyric attacks and deficiency of both PROTO oxidase and HMB synthase. Photosensitivity was not observed. It is unclear whether Chester porphyria is a variant of VP or AIP.

Diagnosis When VP is symptomatic, levels of fecal protoporphyrin and coproporphyrin and of urinary coproporphyrin are increased. Urinary ALA and PBG levels are increased during acute attacks. Plasma levels of porphyrins are increased, particularly when there are cutaneous lesions. VP can be distinguished rapidly from all other porphyrias by examining the fluorescence emission spectrum of porphyrins in plasma at neutral pH. This test is particularly useful for differentiating VP from PCT.

Assays of PROTO oxidase activity in cultured fibroblasts or lymphocytes are not widely available. Some latent cases of VP can be diagnosed by measurement of fecal porphyrins in relatives of VP patients.

TREATMENT

Acute attacks are treated with hematin as in AIP. Other than avoiding sun exposure, there are few effective measures for treating the skin lesions. b-Carotene, phlebotomy, and chloroquine are not helpful.

THE ERYTHROPOIETIC PORPHYRIAS

In the erythropoietic porphyrias, porphyrins from bone marrow erythrocytes and plasma are deposited in the skin and lead to cutaneous photosensitivity.

X-LINKED SIDEROBLASTIC ANEMIA

XLSA results from the deficient activity of the erythroid form of ALA synthase and is associated with ineffective erythropoiesis, weakness, and pallor.

Clinical Features Typically, males with XLSA develop refractory hemolytic anemia, pallor, and weakness during infancy. They have secondary hypersplenism, become iron overloaded, and can develop hemosiderosis. The severity depends on the level of residual erythroid ALA synthase activity and on the responsiveness of the specific mutation to pyridoxal 5¢-phosphate supplementation (see below). Peripheral blood smears reveal a hypochromic, microcytic anemia with striking anisocytosis, poikilocytosis, and polychromasia; the leukocytes and platelets appear normal. Hemoglobin content is reduced, and the mean corpuscular volume and mean corpuscular hemoglobin concentration are decreased. Patients with milder, late-onset disease have been reported recently.

Diagnosis Bone marrow examination reveals hypercellularity with a left shift and megaloblastic erythropoiesis with an abnormal maturation. A variety of Prussian blue-staining sideroblasts are observed. Levels of urinary porphyrin precursors and of both urinary and fecal porphyrins are normal. The level of erythroid ALA synthase is decreased in bone marrow, but this enzyme is difficult to measure in the presence of the normal ALA synthase housekeeping enzyme. Definitive diagnosis requires the demonstration of mutations in the *erythroid ALA synthase* gene.

TREATMENT

The severe anemia may respond to pyridoxine supplementation. This cofactor is essential for ALA synthase activity, and mutations in the pyridoxine binding site of the enzyme have been found in several responsive patients. Cofactor supplementation may make it possible to eliminate or reduce the frequency of transfusion. Unresponsive patients may be transfusion-dependent and require chelation therapy.

CONGENITAL ERYTHROPOIETIC PORPHYRIA

CEP is an autosomal recessive disorder, also known as *Gunther's disease*, that is due to the markedly deficient activity of URO synthase; it is associated with hemolytic anemia and cutaneous lesions. CEP is characterized by accumulation of uroporphyrin I and coproporphyrin I isomers.

Clinical Features Severe cutaneous photosensitivity begins in early infancy. The skin over sun-exposed areas is friable, and bullae and vesicles are prone to rupture and infection. Skin thickening, focal hypo- and hyperpigmentation, and hypertrichosis of the face and extremities are characteristic. Secondary infection of the cutaneous lesions can lead to disfigurement of the face and hands. Porphyrins are deposited in teeth and in bones. As a result, the teeth are reddish brown and fluoresce on exposure to long-wave ultraviolet light. Hemolysis is probably due to the marked increase in erythrocyte porphyrins and leads to splenomegaly. Adults with a milder form of the disease have been described.

Diagnosis Uroporphyrin and coproporphyrin (mostly type I isomers) accumulate in the bone marrow, erythrocytes, plasma, urine, and feces. The diagnosis should be confirmed by demonstration of markedly deficientUROsynthase activity. The disease can be detected in utero by measuring porphyrins in amniotic fluid and URO synthase activity in cultured amniotic cells or chorionic villi. Molecular analyses of the mutant alleles from over 20 unrelated patients have revealed the presence of gene rearrangements, an mRNA processing defect, and several point mutations that cause amino acid substitutions.

TREATMENT

The transfusion of sufficient blood to suppress erythropoiesis is effective but results in iron overload. Splenectomy may reduce hemolysis and decrease transfusion requirements. Protection from sunlight and from minor skin trauma is important. b-Carotene may be of some value. Complicating bacterial infections should be treated promptly. Recently, bone marrow transplantation has proven effective in several transfusion-dependent children, providing the rationale for stem-cell gene therapy.

ERYTHROPOIETIC PROTOPORPHYRIA

Erythropoietic protoporphyria (EPP) is an autosomal dominant disorder due to the partial deficiency of ferrochelatase activity. Protoporphyrin accumulates in erythroid cells and plasma and is excreted in bile and feces. EPP is the most common erythropoietic porphyria and, afterPCT, the second most common porphyria.

Clinical Features Skin photosensitivity usually begins in childhood. The skin manifestations differ from those of other porphyrias. Vesicular lesions are uncommon. Redness, swelling, burning, and itching can develop within minutes of sun exposure and resemble angioedema. Symptoms may seem out of proportion to the visible skin lesions. Sparse vesicles and bullae occur in 10% of cases. Chronic skin changes may include lichenification, leathery pseudovesicles, labial grooving, and nail changes. Severe scarring is rare, as are pigment changes, friability, and hirsutism.

The primary source of excess protoporphyrin is the bone marrow reticulocyte. Erythrocyte protoporphyrin is free (not complexed with zinc) and is mostly bound to hemoglobin. In plasma, protoporphyrin is bound to albumin. Hemolysis and anemia are usually absent or mild.

Liver function is usually normal, but in some patients accumulation of protoporphyrin causes chronic liver disease that can progress to liver failure and death. The hepatic complications are often preceded by increasing levels of erythrocyte and plasma protoporphyrin and probably result, in part, from protoporphyrin accumulation in the liver. Protoporphyrin is insoluble, forms crystalline structures in liver cells, and can decrease hepatic bile flow. Gallstones composed at least in part of protoporphyrin occur in some patients.

Some obligate heterozygotes are asymptomatic and have little or no increase in erythrocyte protoporphyrin. Thus there is phenotypic variation in this disease.

Diagnosis Protoporphyrin levels are increased in bone marrow, circulating erythrocytes, plasma, bile, and feces. Urinary levels of porphyrin and porphyrin precursors are normal. Ferrochelatase activity in cultured lymphocytes or fibroblasts is decreased.

TREATMENT

Oralb-carotene (120 to 180 mg/d) improves tolerance to sunlight in many patients. The dosage may need to be adjusted to maintain serum carotene levels in the recommended range of 10 to 15 umol/L (600 to 800 ug/dL). Mild skin discoloration due to carotenemia is the only significant side effect. The beneficial effects of b-carotene may involve quenching of singlet oxygen or free radicals. Unfortunately, this drug is less effective in other forms of porphyria associated with photosensitivity.

Treatment of hepatic complications is difficult. However, cholestyramine and other porphyrin absorbents such as activated charcoal may interrupt the enterohepatic circulation of protoporphyrin and promote its fecal excretion, leading to some improvement. Splenectomy may be helpful when the disease is accompanied by hemolysis and significant splenomegaly. Caloric restriction and drugs or hormones that may induce the heme pathway or impair hepatic excretory function should be avoided. Iron deficiency should be prevented or treated. Transfusions or intravenous heme therapy may suppress erythroid and hepatic protoporphyrin production and are sometimes beneficial. Liver transplantation has been carried out in some patients with severe liver complications.

(Bibliography omitted in Palm version)

347. DISORDERS OF PURINE AND PYRIMIDINE METABOLISM - *Robert L. Wortmann*

Purines and pyrimidines are the bases that, when linked to sugars (ribose or deoxyribose) and phosphate groups, create the nucleic acids that comprise the building blocks of RNA and DNA. The main purine bases are adenine and guanine; the pyrimidine bases include cytosine, thymine, and uracil. In addition, purines participate in diverse cellular functions, including intracellular energy metabolism (e.g., ATP), cell signaling pathways (e.g., GTP), and intercellular communication (e.g., adenosine). The nucleotides therefore serve fundamental roles in the replication of genetic material, gene transcription, protein synthesis, and cellular metabolism. Disorders that involve abnormalities of nucleotide metabolism range from relatively common diseases such as hyperuricemia and gout, in which there is increased production or impaired excretion of a metabolic end product of purine metabolism (uric acid), to rare enzyme deficiencies that affect purine and pyrimidine synthesis or degradation. Understanding these biochemical pathways has led, in some instances, to the development of specific forms of treatment, such as the use of allopurinol to reduce uric acid production.

URIC ACID METABOLISM

Uric acid is the final breakdown product of purine degradation in humans. It is a weak acid with pK_as of 5.75 and 10.3. Urates, the ionized forms of uric acid, predominate in plasma extracellular fluid and synovial fluid, with approximately 98% existing as monosodium urate at pH 7.4. Monosodium urate is easily dialyzed from plasma. Binding of urate to plasma proteins has little physiologic significance.

Plasma is saturated with monosodium urate at a concentration of 415 umol/L (6.8 mg/dL) at 37°C. At higher concentrations, plasma is therefore supersaturated, creating the potential for urate crystal precipitation. However, precipitation sometimes does not occur even at plasma urate concentrations as high as 4800 umol/L (80 mg/dL), perhaps because of the presence of solubilizing substances in plasma.

Uric acid is more soluble in urine than in water, possibly because of the presence of urea, proteins, and mucopolysaccharides. The pH of urine greatly influences its solubility. At pH 5.0, urine is saturated with uric acid at concentrations ranging from 360 to 900 umol/L (6 to 15 mg/dL). At pH 7.0, saturation is reached at concentrations between 9480 and 12,000 umol/L (158 and 200 mg/dL). Ionized forms of uric acid in urine include mono- and disodium, potassium, ammonium, and calcium urates.

Although purine nucleotides are synthesized and degraded in all tissues, urate is produced only in tissues that contain xanthine oxidase, primarily the liver and small intestine. The amount of urate in the body is the net result of the amount produced and the amount excreted (Fig. 347-1). Urate production varies with the purine content of the diet and the rates of purine biosynthesis, degradation, and salvage. Normally, two-thirds to three-fourths of urate is excreted by the kidneys, and most of the remainder is eliminated through the intestines. A four-component model describes the renal handling of uric acid in humans and includes: (1) glomerular filtration, (2) tubular reabsorption, (3) secretion, and (4) postsecretory reabsorption (Fig. 347-2). Approximately 8 to 12% of urate filtered by the glomeruli is excreted in the urine as uric acid. After filtration, 98 to

100% of the urate is reabsorbed; about half the reabsorbed urate is secreted back into the proximal tubule, and about 40% of that is again reabsorbed.

Serum urate levels vary with age and sex. Most children have serum urate concentrations of 180 to 240 umol/L (3.0 to 4.0 mg/dL). Levels begin to rise during puberty in males but remain low in females until menopause. Although the cause of this gender variation is not completely understood, it is due in part to a higher excretion of urate in females. Mean serum urate values of adult men and premenopausal women are 415 and 360 umol/L (6.8 and 6.0 mg/dL), respectively. After menopause, values for women increase to approximate those of men. In adulthood, concentrations rise steadily over time and vary with height, body weight, blood pressure, renal function, and alcohol intake.

HYPERURICEMIA

Hyperuricemia can result from increased production or decreased excretion of uric acid or from a combination of the two processes. When sustained hyperuricemia exists, plasma and extracellular fluids are supersaturated with respect to urate, and total body urate is increased. Sustained hyperuricemia predisposes some individuals to develop clinical manifestations including gouty arthritis (Chap. 322) and renal dysfunction (see below).

Hyperuricemia may be defined as a plasma (or serum) urate concentration >420 umol/L (7.0 mg/dL). This definition is based on physicochemical, epidemiologic, and disease-related criteria. Physicochemically, hyperuricemia is the concentration of urate in the blood that exceeds the solubility limits of monosodium urate in plasma, 415 umol/L (6.8 mg/dL). In epidemiologic studies, hyperuricemia is defined as the mean plus 2 standard deviations of values determined from a randomly selected healthy population. When measured in unselected individuals, 95% have serum urate concentrations <420 umol/L (7.0 mg/dL). Finally, hyperuricemia can be defined in relation to the risk of disease. The risk of developing gouty arthritis or urolithiasis increases with urate levels >420 umol/L (7.0 mg/dL) and escalates in proportion to the degree of elevation. Hyperuricemia is present in between 2.0 and 13.2% of ambulatory adults and somewhat more frequently in hospitalized individuals.

CAUSES OF HYPERURICEMIA

Hyperuricemia may be classified as primary or secondary depending on whether the cause is innate or is the result of an acquired disorder (Table 347-1). However, it is more useful to classify hyperuricemia in relation to the underlying pathophysiology, i.e., whether it results from increased production, decreased excretion, or a combination of the two (Fig. 347-1,Table 347-2).

Increased Urate Production Diet provides an exogenous source of purines and, accordingly, contributes to the serum urate in proportion to its purine content. Strict restriction of purine intake reduces the mean serum urate level by about 60 umol/L (1.0 mg/dL) and urinary uric acid excretion by approximately 1.2 mmol/d (200 mg/d). Because about 50% of ingested RNA purine and 25% of ingested DNA purine appear in the urine as uric acid, foods high in nucleic acid content have a significant effect on the

serum urate level. Such foods include liver, "sweetbreads" (i.e., thymus and pancreas), kidney, and anchovy.

Endogenous sources of purine production also influence the serum urate level (Fig. 347-3). De novo purine biosynthesis, the formation of a purine ring from nonring structures, is an 11-step process that results in formation of inosine monophosphate (IMP). The first step combines phosphoribosylpyrophosphate (PRPP) and glutamine and is catalyzed by amidophosphoribosyltransferase (amidoPRT). The rates of purine biosynthesis and urate production are determined, for the most part, by this enzyme. AmidoPRT is regulated by the substrate PRPP, which drives the reaction forward, and by the end products of biosynthesis (IMP and other ribonucleotides), which provide feedback inhibition. A secondary regulatory pathway is the salvage of purine bases by hypoxanthine phosphoribosyltransferase (HPRT). HPRT catalyzes the combination of the purine bases hypoxanthine and guanine with PRPP to form the respective ribonucleotides IMP and guanosine monophosphate (GMP). Increased salvage activity thus retards de novo synthesis by reducing PRPP levels and increasing concentrations of inhibitory ribonucleotides.

Serum urate levels are closely coupled to the rates of de novo purine biosynthesis, which is driven in part by the level of PRPP, as evidenced by two inborn errors of purine metabolism. Both increased PRPP synthetase activity and HPRT deficiency are associated with overproduction of purines, hyperuricemia, and hyperuricaciduria (see below for clinical descriptions). An X-linked disorder that causes an increase in activity of the enzyme PRPP synthetase leads to increased PRPP production and accelerated de novo biosynthesis. PRPP is a substrate and allosteric activator of amidoPRT, the first enzyme in the de novo pathway. HPRT deficiency is also X-linked and enhances urate biosynthesis in two ways. PRPP is accumulated as a result of decreased utilization in the salvage pathway and, in turn, provides increased substrate for amidoPRT and de novo biosynthesis. In addition, decreased formation of the nucleoside monophosphates, IMP and GMP, via the salvage pathway impairs feedback inhibition on amidoPRT, further enhancing de novo biosynthesis.

Accelerated purine nucleotide degradation can also cause hyperuricemia, i.e., with conditions of rapid cell turnover, proliferation, or cell death, as in leukemic blast crises, cytotoxic therapy for malignancy, hemolysis, or rhabdomyolysis. Nucleic acids released from cells are hydrolyzed by the sequential activities of nucleases and phosphodiesterases, forming nucleoside monophosphates, which in turn are degraded to nucleosides, bases, and urate. Hyperuricemia can result from excessive degradation of skeletal muscle ATP after strenuous physical exercise or status epilepticus and in glycogen storage diseases types III, V, and VII (Chap. 350). The hyperuricemia of myocardial infarction, smoke inhalation, and acute respiratory failure may also be related to accelerated breakdown of ATP.

Decreased Uric Acid Excretion Over 90% of individuals with sustained hyperuricemia have a defect in the renal handling of uric acid. In hyperuricemia with gout the renal defect is evidenced by a lower than normal ratio of urate clearance to glomerular filtration rate (or urate to insulin clearance rate) over a wide range of filtered loads. As a result, gouty individuals excrete approximately 40% less uric acid than nongouty individuals for any given plasma urate concentration. Uric acid excretion increases in

gouty and nongouty individuals when plasma urate levels are raised by purine ingestion or infusion, but in those with gout, plasma urate concentrations must be 60 to 120 umol/L (1 to 2 mg/dL) higher than normal to achieve equivalent uric acid excretion rates.

Altered uric acid excretion could theoretically result from decreased glomerular filtration, decreased tubular secretion, or enhanced tubular reabsorption. Decreased urate filtration does not appear to cause primary hyperuricemia but does contribute to the hyperuricemia of renal insufficiency. Although hyperuricemia is invariably present in chronic renal disease, the correlation between serum creatinine, urea nitrogen, and urate concentration is poor. Uric acid excretion per unit of glomerular filtration rate increases progressively with chronic renal insufficiency, but tubular secretory capacity tends to be preserved, tubular reabsorptive capacity is reduced, and extrarenal clearance of uric acid increases as renal damage becomes more severe.

Decreased tubular secretion of urate causes the secondary hyperuricemia of acidosis. Diabetic ketoacidosis, starvation, ethanol intoxication, lactic acidosis, and salicylate intoxication are accompanied by accumulations of organic acids (b-hydroxybutyrate, acetoacetate, lactate, or salicylates) that compete with urate for tubular secretion. Hyperuricemia may be due to enhanced reabsorption of uric acid distal to the site of secretion. This mechanism is known to be responsible for the hyperuricemia of extracellular volume depletion that occurs with diabetes insipidus or diuretic therapy.

Combined Mechanisms Both increased urate production and decreased uric acid excretion may contribute to hyperuricemia. Individuals with a deficiency of glucose-6-phosphatase, the enzyme that hydrolyzes glucose-6-phosphate to glucose, are hyperuricemic from infancy and develop gout early in life (Chap. 350). Increased urate production results from accelerated ATP degradation during fasting or hypoglycemia. In addition, the lower levels of nucleoside monophosphates decrease feedback inhibition ofamidoPRT, thereby accelerating de novo biosynthesis. Glucose-6-phosphatase-deficient individuals may also develop hyperlacticacidemia, which blocks uric acid excretion by decreasing tubular secretion.

Patients with hereditary fructose intolerance caused by fructose-1-phosphate aldolase deficiency also develop hyperuricemia by both mechanisms. In homozygotes, vomiting and hypoglycemia after fructose ingestion can lead to hepatic failure and proximal renal tubular dysfunction. Ingestion of fructose, the substrate for the enzyme, causes accumulation of fructose-1-phosphate. This action results in ATP depletion, accelerated purine nucleotide catabolism, and hyperuricemia. Both lactic acidosis and renal tubular acidosis contribute to urate retention. Heterozygous carriers develop hyperuricemia, and perhaps one-third develop gout. The heterozygous state has a prevalence of 0.5 to 1.5%, suggesting that fructose-1-phosphate aldolase deficiency may be a relatively common cause of familial gout.

Alcohol also promotes hyperuricemia by both mechanisms. Excessive alcohol consumption accelerates hepatic breakdown of ATP and increases urate production. Alcohol consumption can also induce hyperlacticacidemia, which blocks uric acid secretion. The higher purine content in some alcoholic beverages such as beer may also be a factor.

EVALUATION OF HYPERURICEMIA

Hyperuricemia does not necessarily represent a disease, nor is it a specific indication for therapy. Rather, the finding of hyperuricemia is an indication to determine its cause. The decision to treat depends on the cause and the potential consequences of the hyperuricemia in each individual.

Quantification of uric acid excretion can be used to determine whether hyperuricemia is caused by overproduction or decreased excretion. On a purine-free diet, men with normal renal function excrete <3.6 mmol/d (600 mg/d). Thus, the hyperuricemia of individuals who excrete uric acid above this level while on a purine-free diet is due to purine overproduction, whereas it is due to decreased excretion in those who excrete lower amounts on the purine-free diet. If the assessment is performed while the patient is on a regular diet, the level of 4.2 mmol/d (800 mg/d) can be used as the discriminating value. With renal insufficiency, less urate is filtered in the glomeruli and less uric acid appears in the urine. Consequently, a lower 24-h urinary uric acid value in the presence of renal insufficiency does not necessarily rule out urate overproduction, but an elevated value provides strong evidence of urate overproduction. Spuriously high values can occur if a uricosuric agent is being taken at the time of urine collection. Glucocorticoids, ascorbic acid, salicylates in doses >2 g/d, and other agents that promote urate excretion interfere with the interpretation of results.

Assessment of the ratio of uric acid to creatinine (or the ratio of uric acid clearance to creatinine clearance) in spot or random urine samples is not a reliable method to screen for urate overproduction. However, this is a useful tool for evaluating individuals with acute renal failure suspected of having acute uric acid nephropathy (see below).

Pyrazinamide, which has a suppressive action on tubular secretion, can be used to investigate presecretory reabsorption of uric acid. Probenecid, an agent that inhibits postsecretory reabsorption, can be used to evaluate tubular secretion and postsecretory reabsorption.

COMPLICATIONS OF HYPERURICEMIA

The most recognized complication of hyperuricemia is *gouty arthritis.* In the general population the prevalence of hyperuricemia ranges between 2.0 and 13.2%, and the prevalence of gout is between 1.3 and 3.7%. The higher the serum urate level, the more likely an individual is to develop gout. In one study, the incidence of gout was 4.9% for individuals with serum urate concentrations>540 umol/L (9.0 mg/dL) compared with 0.5% for those with values between 415 and 535 umol/L (7.0 and 8.9 mg/dL). The complications of gout correlate with both the duration and severity of hyperuricemia. *For further discussion of gout, see Chap. 322.*

Hyperuricemia also causes several renal problems: (1) nephrolithiasis; (2) urate nephropathy, a rare cause of renal insufficiency attributed to monosodium urate crystal deposition in the renal interstitium; and (3) uric acid nephropathy, a reversible cause of acute renal failure resulting from deposition of large amounts of uric acid crystals in the renal collecting ducts, pelvis, and ureters.

Nephrolithiasis Uric acid nephrolithiasis occurs most commonly, but not exclusively, in individuals with gout. In gout, the prevalence of nephrolithiasis correlates with the serum and urinary uric acid levels, reaching approximately 50% with serum urate levels of 770 umol/L (13 mg/dL) or urinary uric acid excretion>6.5 mmol/d (1100 mg/d).

Uric acid stones can develop in individuals with no evidence of arthritis, only 20% of whom are hyperuricemic. Uric acid can also play a role in other types of kidney stones. Some nongouty individuals with calcium oxalate or calcium phosphate stones have hyperuricemia or hyperuricaciduria. Uric acid may act as a nidus on which calcium oxalate can precipitate or lower the formation product for calcium oxalate crystallization.

Urate Nephropathy Urate nephropathy, sometimes referred to as *urate nephrosis*, is a late manifestation of severe gout and is characterized histologically by deposits of monosodium urate crystals surrounded by a giant cell inflammatory reaction in the medullary interstitium and pyramids. The disorder is now rare and cannot be diagnosed in the absence of gouty arthritis. The lesions may be clinically silent or cause proteinuria, hypertension, and renal insufficiency.

Uric Acid Nephropathy This reversible cause of acute renal failure is due to precipitation of uric acid in renal tubules and collecting ducts that causes obstruction to urine flow. Uric acid nephropathy develops following sudden urate overproduction and marked hyperuricaciduria. Factors that favor uric acid crystal formation include dehydration and acidosis. This form of acute renal failure occurs most often during an aggressive "blastic" phase of leukemia or lymphoma prior to or coincident with cytolytic therapy but has also been observed in individuals with other neoplasms, following epileptic seizures, and after vigorous exercise with heat stress. Autopsy studies have demonstrated intraluminal precipitates of uric acid, dilated proximal tubules, and normal glomeruli. The initial pathogenic events are believed to include obstruction of collecting ducts with uric acid and obstruction of distal renal vasculature.

If recognized, uric acid nephropathy is potentially reversible. Appropriate therapy has reduced the mortality from about 50% to practically nil. Serum levels cannot be relied on for diagnosis because this condition has developed in the presence of urate concentrations varying from 720 to 4800 umol/L (12 to 80 mg/dL). The distinctive feature is the urinary uric acid concentration. In most forms of acute renal failure with decreased urine output, urinary uric acid content is either normal or reduced, and the ratio of uric acid to creatinine is<1. In acute uric acid nephropathy the ratio of uric acid to creatinine in a random urine sample or 24-h specimen is>1, and a value that high is essentially diagnostic.

TREATMENT

Asymptomatic Hyperuricemia Hyperuricemia is present in approximately 5% of the population and in up to 25% of hospitalized individuals. The vast majority are asymptomatic with regard to their hyperuricemia and are at no clinical risk because of it. Elevated serum urate concentrations have been associated with insulin resistance, obesity, hypertension, dyslipidemia (sometimes referred to as syndrome X), and atherosclerotic disease. However, urate does not appear to have a causal role in the development of coronary heart disease or death from cardiovascular disease. In the

past, the association of hyperuricemia with cardiovascular disease and renal failure led to the use of urate-lowering agents for people with asymptomatic hyperuricemia. This practice is no longer recommended with the exception of individuals receiving cytolytic therapy for neoplastic disease, in which treatment is given in an effort to prevent uric acid nephropathy.

Hyperuricemic individuals are at risk to develop gouty arthritis, especially those with higher serum urate levels. However, treatment of asymptomatic hyperuricemia to prevent the first attack of gouty arthritis is not indicated because most hyperuricemic people never develop gout. Furthermore, neither structural kidney damage nor tophi are identifiable before the first attack. Reduced renal function cannot be attributed to asymptomatic hyperuricemia, and treatment of asymptomatic hyperuricemia does not alter the progression of renal dysfunction in patients with renal disease. Although nephrolithiasis is common in gouty patients, and a number of individuals with nephrolithiasis are hyperuricemic, increased risk of stone formation in people with asymptomatic hyperuricemia is not established.

Thus, because treatment with antihyperuricemic agents entails inconvenience, cost, and potential toxicity, routine treatment of asymptomatic hyperuricemia cannot be justified other than for prevention of acute uric acid nephropathy. In addition, routine screening for asymptomatic hyperuricemia is not recommended. If hyperuricemia is diagnosed, however, the cause should be determined. Causal factors should be corrected if the condition is secondary, and associated problems such as hypertension, hypercholesterolemia, diabetes mellitus, and obesity should be treated.

Symptomatic Hyperuricemia (See Chap. 322 for treatment of gout)

Nephrolithiasis Antihyperuricemic therapy is recommended for the individual who has both gouty arthritis and either uric acid- or calcium-containing stones, both of which may occur in association with hyperuricaciduria. Regardless of the nature of the calculi, fluid ingestion should be sufficient to produce a daily urine volume>2 L. Alkalinization of the urine with sodium bicarbonate or acetazolamide may be justified to increase the solubility of uric acid. Specific treatment of uric acid calculi requires reducing the urine uric acid concentration with allopurinol. Allopurinol administration decreases the serum urate concentration and the urinary excretion of uric acid in the first 24 h, with a maximum reduction occurring within 2 weeks. The average effective dose of allopurinol is 300 mg/d. Allopurinol can be given once a day because of the long half-life (18 h) of its active metabolite oxypurinol. The drug is effective in patients with renal insufficiency, but the dose should be reduced. Allopurinol is also useful in reducing the recurrence of calcium oxalate stones in gouty patients and in nongouty individuals with hyperuricemia or hyperuricaciduria. Potassium citrate (30 to 80 mmol/d orally in divided doses) is an alternative therapy for patients with uric acid stones alone or mixed calcium/uric acid stones. Allopurinol is also indicated for the treatment of 2,8-dihydroxyadenine kidney stones.

Uric Acid Nephropathy Uric acid nephropathy is often preventable, and immediate, appropriate therapy has greatly reduced the mortality rate. Vigorous intravenous hydration and diuresis with furosemide dilute the uric acid in the tubules and promote urine flow to ³100 mL/h. The administration of acetazolamide, 240 to 500 mg every 6 to

8 h, and sodium bicarbonate, 89 mmol/L, intravenously enhances urine alkalinity and thereby solubilizes more uric acid. It is important to ensure that the urine pH remains>7.0 and to watch for circulatory overload. In addition, antihyperuricemic therapy in the form of allopurinol in a single dose of 8 mg/kg is administered to reduce the amount of urate that reaches the kidney. If renal insufficiency persists, subsequent daily doses should be reduced to 100 to 200 mg because oxypurinol, the active metabolite of allopurinol, accumulates in renal failure. Despite these measures, hemodialysis may be required.

HYPOURICEMIA

Hypouricemia, defined as a serum urate concentration <120 umol/L (2.0 mg/dL) can result from decreased production of urate, increased excretion of uric acid, or a combination of both mechanisms. It occurs in<0.2% of the general population and<0.8% of hospitalized individuals. Hypouricemia causes no symptoms or pathology and therefore requires no therapy. It is, however, a sign of potential pathology, and its cause should be determined.

Most hypouricemia results from increased renal uric acid excretion. The finding of normal amounts of uric acid in a 24-h urine collection in an individual with hypouricemia is evidence for a renal cause. Medications with uricosuric properties (Table 347-3) include aspirin (at doses>2.0 g/d), x-ray contrast materials, and glycerylguaiacholate. Total parenteral hyperalimentation can also cause hypouricemia, possibly a result of the high glycine content of the infusion formula. Other causes of increased urate clearance include conditions such as neoplastic disease, hepatic cirrhosis, diabetes mellitus, and inappropriate secretion of vasopressin; defects in renal tubular transport such as primary Fanconi syndrome and Fanconi syndromes caused by Wilson's disease, cystinosis, multiple myeloma, and heavy metal toxicity; and isolated congenital defects in the bidirectional transport of uric acid.

Hypouricemia from decreased production of urate is accompanied by very low urinary uric acid levels. Accumulation of other purine nucleosides and bases may occur depending on the specific defect. Individuals treated with allopurinol and some patients with neoplastic disease or severe hepatic dysfunction are hypouricemic and excrete increased quantities of hypoxanthine and xanthine in the urine. Xanthine oxidase deficiency can be inherited or acquired. Inherited forms include isolated xanthine oxidase deficiency and combined xanthine oxidase and sulfite oxidase deficiencies. Both cause hypouricemia and xanthinuria. Affected individuals excrete essentially no uric acid and may develop xanthine nephrolithiasis. Individuals with purine nucleoside phosphorylase deficiency, an inborn error of metabolism causing T cell-deficient immune dysfunction, are hypouricemic and excrete increased quantities of guanosine, deoxyguanosine, inosine, and deoxyinosine in the urine.

INBORN ERRORS OF PURINE METABOLISM (See also Table 347-4, Fig. 347-3)

HPRTDEFICIENCY

A complete deficiency of HPRT, the Lesch-Nyhan syndrome, is characterized by hyperuricemia, self-mutilative behavior, choreoathetosis, spasticity, and mental

retardation. A partial deficiency of HPRT, the Kelley-Seegmiller syndrome, is associated with hyperuricemia but no central nervous system manifestations. In both disorders, the hyperuricemia results from urate overproduction and can cause uric acid crystalluria, nephrolithiasis, obstructive uropathy, and gouty arthritis. Early diagnosis and appropriate therapy with allopurinol can prevent or eliminate all the problems attributable to hyperuricemia but have no effect on the behavioral or neurologic abnormalities.

HPRT catalyzes the reaction that combines PRPP and the purine bases hypoxanthine and guanine to form the respective nucleoside monophosphate IMP or GMP and pyrophosphate. The enzyme is encoded by a single gene located on the X chromosome in region q26-q27. Consequently, affected males are hemizygous for the trait and inherit the mutant allele from their asymptomatic mother, who is a carrier, or are the result of spontaneous gene mutations. The deficiency state is generally the result of point mutations, small deletions or insertions, or endoduplication of exons rather than major gene alterations.

INCREASED PRPP SYNTHETASE ACTIVITY

Cells from individuals with increased PRPP synthetase activity contain elevated levels of PRPP. The high substrate content drives de novo purine synthesis, causing overproduction of uric acid. Like the HPRT deficiency states, PRPP synthetase overactivity is X-linked and results in gouty arthritis and uric acid nephrolithiasis. Nerve deafness occurs in some families.

ADENINE PHOSPHORIBOSYLTRANSFERASE (APRT) DEFICIENCY

Individuals with a deficiency of APRT develop kidney stones composed of 2,8-dihydroxyadenine. APRT catalyzes the conversion of adenine to adenosine monophosphate (AMP). In the absence of APRT, adenine is converted by xanthine oxidase to 2,8-dihydroxyadenine, which is insoluble in urine. Reports of 2,8-dihydroxyadenine stones are rare, most likely because of its chemical similarity to uric acid. Analysis by x-ray powder diffraction is necessary for correct identification. Because this technique is rarely employed, many 2,8-dihydroxyadenine stones are incorrectly called uric acid. The consequence of this misidentification is not deleterious, as allopurinol therapy is the correct treatment for each type of stone.

APRT deficiency is inherited as an autosomal recessive trait. Caucasians with the disorder have a complete deficiency (type I), whereas Japanese subjects have some measurable enzyme activity (type II). Expression of the defect is similar in the two populations, as is the frequency of the heterozygous state (0.4 to 1.1 per 100).

HEREDITARY XANTHINURIA

A deficiency of xanthine oxidase causes all purine in the urine to occur in the form of hypoxanthine and xanthine. About two-thirds of deficient individuals are asymptomatic. The remainder develop kidney stones composed of xanthine. A very small number of symptomatic individuals also have myopathy or recurrent polyarteritis. Xanthinuria appears to be inherited in an autosomal recessive pattern.

In a second form of inherited xanthinuria, the deficiency of xanthine oxidase is associated with a deficiency of sulfite oxidase. Neurologic symptoms attributable to the sulfite oxidase deficiency predominate over those of xanthinuria in individuals with the combined deficiency.

MYOADENYLATE DEAMINASE DEFICIENCY

Adenylate deaminase (AMPdeaminase) catalyzes the conversion of AMP toIMP with the release of ammonia and is an integral component of the purine nucleotide cycle, which plays an important role in skeletal muscle energy metabolism. Both primary (inherited) and secondary (acquired) forms of myoadenylate deaminase deficiency have been described. Myoadenylate deaminase is the only activity affected in the inherited form, whereas other muscle enzymes (creatine kinase and myokinase) are also decreased in the acquired deficiencies. In contrast, mRNA abundance is low in muscle from patients with acquired deficiencies, suggesting a different molecular basis for this form.

The primary form is inherited as an autosomal recessive trait. Clinically, this form does not appear to cause disease, and most individuals with this defect may be asymptomatic. Another explanation for the myopathy should be sought in symptomatic patients with this deficiency. The acquired deficiency occurs in association with a wide variety of neuromuscular disease, including muscular dystrophies, neuropathies, inflammatory myopathies, and collagen vascular diseases.

ADENYLOSUCCINATE LYASE DEFICIENCY

Adenylosuccinate lyase participates in the synthesis of purine nucleotides in two ways. It catalyzes the conversion of succinylaminoimidazole carboxamide ribotide (SAICAR) to aminoimidazole carboxamide ribotide (AICAR) in the de novo pathway and in the conversion ofAMPsuccinate to AMP in the purine nucleotide cycle. Deficiency of this enzyme is due to an autosomal recessive trait and causes profound psychomotor retardation, seizures, and other movement disorders. All individuals with this deficiency are mentally retarded, and most are autistic.

ADENOSINE DEAMINASE DEFICIENCY AND PURINE NUCLEOSIDE PHOSPHORYLASE DEFICIENCY SeeChap. 308.

PYRIMIDINE DISORDERS

The pyrimidine, cytidine, is found in both DNA and RNA; it is a complementary base pair for guanine. Thymidine is found only in DNA where it is paired with adenine. Uridine is found only in RNA and can pair with either adenine or guanine in RNA secondary structures. Pyrimidines can be synthesized by a de novo pathway (Fig. 347-4) or reused in a salvage pathway. More than 25 different enzymes are involved in pyrimidine synthesis, salvage, and degradation pathways. Nonetheless, disorders of pyrimidine metabolism are rare. They are more difficult to recognize than purine disorders because of heterogeneous phenotypes and the absence of readily detected biochemical markers. Three disorders of pyrimidine metabolism are discussed below.

OROTIC ACIDURIA

Hereditary orotic aciduria is an autosomal recessive disorder caused by mutations in a bifunctional enzyme, uridine-5¢-monophosphate (UMP) synthase, which converts orotic acid to UMP in the de novo synthesis pathway (Fig. 347-4). The same protein encodes two distinct enzymatic activities. The disorder is characterized by hypochromic megaloblastic anemia that is unresponsive to vitamin B_{12} and folic acid, growth retardation, and neurologic abnormalities. Increased excretion of orotic acid causes crystalluria and obstructive uropathy. Replacement of uridine (100 to 200 mg/kg per day) corrects the anemia, reduces orotic acid excretion, and improves the other sequelae of the disorder.

PYRIMIDINE 5¢-NUCLEOTIDASE DEFICIENCY

Pyrimidine 5¢-nucleotidase catalyzes the removal of the phosphate group from pyrimidine ribonucleoside monophosphates (cytidine-5¢-monophosphate orUMP) (Fig. 347-4). Deficiency of this enzyme is transmitted as a recessive trait and causes hemolytic anemia with prominent basophilic stippling of erythrocytes. The accumulation of pyrimidines or cytidine diphosphate choline (CDPC) is thought to induce hemolysis. The enzyme deficiency alters nucleoside composition and thereby generates a characteristic ultraviolet spectrum in deproteinized erythrocytes. There is no specific treatment.

DIHYDROPYRIMIDINE DEHYDROGENASE DEFICIENCY

Dihydropyrimidine dehydrogenase (DPD) is the rate-limiting enzyme in the pathway of uracil and thymine degradation (Fig. 347-4). Deficiency of this enzyme causes excessive urinary excretion of uracil and thymine. DPD deficiency is transmitted in a recessive manner and causes nonspecific cerebral dysfunction with convulsive disorders, motor retardation, and mental retardation. A splice donor site mutation, which causes deletion of exon 14, accounts for 52% of mutant alleles. No specific treatment is available. DPD is also involved in the degradation of 5-fluorouracil (5-FU), a chemotherapeutic agent that inhibits thymidylate synthase. Consequently, deficiency of this enzyme is associated with 5-FU neurotoxicity.

MEDICATION EFFECTS ON PYRIMIDINE METABOLISM

In addition to the role ofDPD in5-FUdegradation (see above), other medications can influence pyrimidine metabolism. Leflunomide, which is used to treat rheumatoid arthritis, inhibits de novo pyrimidine synthesis by inhibiting dihydroorotate dehydrogenase, resulting in an antiproliferative effect on T cells. Allopurinol, an inhibitor of xanthine oxidase and purine synthesis, also inhibits orotidine-5¢-phosphate decarboxylase, a step inUMPsynthesis. Consequently, allopurinol use is associated with increased excretion of orotidine and orotic acid; there are no known clinical effects of this inhibition.

(Bibliography omitted in Palm version)

348. WILSON'S DISEASE - *I. Herbert Scheinberg*

Wilson's disease is an inherited disorder of copper metabolism in individuals with two mutant *ATP7B* genes. Impairment of the normal excretion of hepatic copper results in toxic accumulation of the metal in liver, brain, and other organs. The disease occurs in every ethnic and geographic population, with a worldwide prevalence of ~1 in 30,000, and a heterozygous carrier frequency of ~1 in 90.

NATURAL HISTORY

The average concentrations of ceruloplasmin and hepatic copper are indistinguishable in normal neonates and in patients with Wilson's disease. In normal infants, however, the ceruloplasmin concentration increases and the hepatic copper concentration falls to adult levels during the first 3 months of life. In infants with Wilson's disease, the neonatal deficiency of ceruloplasmin and excess of hepatic copper persist indefinitely. Clinical manifestations are rare before age 6, occur most frequently in mid-adolescence, and eventually develop in all untreated patients.

In about half of patients any of four types of hepatic disturbances may herald the clinical onset. *Acute hepatitis* is usually self-limited, is often mistaken for viral hepatitis or infectious mononucleosis, and may be forgotten later in life. *Parenchymal liver disease* may persist after acute hepatitis or may develop insidiously without prior acute disease into a histologic and clinical picture indistinguishable from chronic active hepatitis and cirrhosis. In other patients *cirrhosis* may develop insidiously after a lapse of decades with no prior sign or symptom of liver disease. *Fulminant hepatitis*, generally fatal, is characterized by progressive jaundice, ascites, encephalopathy, hypoalbuminemia, hypoprothrombinemia, moderately elevated plasma levels of liver enzymes, and Coombs-negative hemolytic anemia.

In most other patients neurologic or psychiatric disturbances are the first clinical signs and are always accompanied by Kayser-Fleischer rings ([Plate III-16](#)). These golden deposits of copper in Descemet's membrane of the cornea do not interfere with vision but indicate that copper has been released from the liver and has probably caused brain damage. If a patient with frank neurologic or psychiatric disease does not have Kayser-Fleischer rings when examined by a trained observer using a slit lamp, the diagnosis of Wilson's disease can be excluded. Rarely, Kayser-Fleischer rings may be accompanied by sunflower cataracts.

The neurologic manifestations include resting and intention tremors, spasticity, rigidity, chorea, drooling, dysphagia, and dysarthria. Babinski responses may be present, and abdominal reflexes are often absent. Inexplicably -- in view of the ubiquity of copper excess in the brain -- sensory changes never occur, except for headache.

Psychiatric disturbances are present in most patients with neurologic symptoms. Schizophrenia, manic-depressive psychoses, and classic neuroses may occur, but the commonest disturbances are bizarre behavioral patterns that defy classification. Improvement in the psychiatric state can occur with pharmacologic reduction of the copper excess, but psychotherapy and additional pharmacotherapy may be required.

In about 5% of patients the clinical onset reflects neither a hepatic nor a central nervous system disturbance. The first manifestation may be primary or secondary amenorrhea or repeated and unexplained spontaneous abortions, perhaps due to excess free copper in intrauterine secretions. Kayser-Fleischer rings may occasionally first be discovered during routine ophthalmologic examination.

PATHOGENESIS

The metabolic defect in Wilson's disease is an inability to maintain a near-zero balance of copper. Dietary copper is generally in excess of the small amount that is essential to life. Normally, any excess absorbed copper is excreted by the liver; in patients with Wilson's disease, copper accumulates in the liver, reaching a mean level of about 1000 ug/g dry weight -- 40 times normal.

Fatty infiltration of the hepatic parenchyma and nuclear glycogen deposits are the earliest findings by light microscopy (Fig. 348-1). With electron microscopy, characteristic mitochondrial abnormalities appear to be specific for Wilson's disease. Later, necrosis, inflammation, fibrosis, bile duct proliferation, and cirrhosis ensue. Abnormalities in liver chemistries, particularly elevations in aminotransferases, may be seen at any stage. The capacity of hepatocytes to store copper is eventually exceeded, and copper is released into blood and taken up into extrahepatic tissues with disastrous effects in the brain (Table 348-1).

With magnetic resonance imaging, the effects of copper toxicity in the brain are seen most frequently in the lenticular nuclei and less commonly in the pons, medulla, thalamus, cerebellum, and cerebral cortex. Opalski and Alzheimer type II cells are present early in the course, although neither is specific for Wilson's disease, and neuronal necrosis and cavitation develop later.

An increased copper concentration in the kidney produces little, if any, structural change and usually does not alter renal function. Microscopic hematuria and/or minimal proteinuria occur occasionally; and nephrocalcinosis, renal calculi, and renal tubular acidosis are rare. Pathologic effects in other organs and tissues are minor.

GENETIC CONSIDERATIONS

The autosomal recessive Wilson disease gene, *ATP7B*, and the X-linked Menkes disease gene, *ATP7A*, are membrane-bound, P-type, copper-transporting ATPases containing 6 copper-binding sites (Chap. 353). The amino acid sequences of these genes are 54% identical. In liver the Wilson protein incorporates copper ions into apoceruloplasmin to form ceruloplasmin whose catabolism is accompanied by biliary excretion of its copper ions. In fetal liver the Menkes protein incorporates copper ions into apoceruloplasmin to form fetal ceruloplasmin whose catabolism is accompanied by hepatic retention of its copper ions. In Wilson's disease, ceruloplasmin that is synthesized in the presence of the *ATP7B* mutant is catabolized like fetal ceruloplasmin, leading to hepatic retention of its copper ions.

DIAGNOSIS

The diagnosis is easy provided it is suspected. Wilson's disease should be considered in any patient younger than 40 years with an unexplained disorder of the central nervous system, signs or symptoms of hepatitis, chronic active hepatitis, unexplained persistent elevations of serum aminotransferase, hemolytic anemia in the presence of hepatitis, or unexplained cirrhosis and in any patient who has a relative with Wilson's disease.

The diagnosis is confirmed by the demonstration of either (1) a serum ceruloplasmin level<20 mg/dL *and* Kayser-Fleischer rings or (2) a serum ceruloplasmin level <20 mg/dL *and* a concentration of copper in a liver biopsy sample>250 ug/g dry weight. Most symptomatic patients excrete >100 ug copper per day in urine and have histologic abnormalities on liver biopsy.

TREATMENT

Treatment consists of removing and detoxifying the deposits of copper as rapidly as possible and must be instituted once the diagnosis is secure whether the patient is ill or asymptomatic. Penicillamine is administered orally in an initial dose of 1 g daily in a single or divided doses at least 30 min before and 2 h after eating. Because penicillamine has an antipyridoxine effect, 25 mg/d of pyridoxine is also given. In ~10% of patients sensitivity to penicillamine develops early, making it necessary to monitor the body temperature and skin daily. White blood cell and platelet counts should be assessed and urinalysis performed several times during the first month of treatment. Penicillamine should be discontinued and replaced by trientine, if rash, fever, leukopenia, thrombocytopenia, lymphadenopathy, or proteinuria develops, or if neurologic worsening accompanies the institution of penicillamine and persists for a week or more.

After therapy with penicillamine has been successfully instituted, the patient should be seen at 1- to 3-month intervals to assess the effectiveness of therapy and monitor for late drug toxicity. The history and the physical examination should focus on hepatic, neurologic, and psychiatric signs and symptoms. Slit-lamp examination of the corneas should be performed by an ophthalmologist if neurologic or psychiatric disturbances appear or worsen. White blood cell and platelet counts, transaminase levels, albumin, bilirubin, and free serum copper (total serum copper minus ceruloplasmin-bound copper) should be measured, the aim being a concentration of free copper less than ~0.2 umol/dL (10 ug/dL). A persistent concentration >0.4 umol/dL (20 ug/dL) indicates that the dose of penicillamine is too low or that the patient is noncompliant. For patients who are asymptomatic or who have improved maximally after several years on 1 g/d of penicillamine, the usual effective maintenance dose is 0.75 g/d taken 45 min before breakfast.

At any time, even after years of uneventful penicillamine administration, granulocytopenia, thrombocytopenia, the nephrotic syndrome, Goodpasture's syndrome, systemic lupus erythematosus, severe arthralgias, myasthenia, mammary gigantism, or elastosis perforans serpiginosa may supervene. Except for transient thrombocytopenia or granulocytopenia, these reactions mandate the replacement of penicillamine with trientine.

The dose of trientine is 1g/d on an empty stomach. Most patients find it convenient to take four 250-mg capsules, delaying breakfast for about an hour. Pyridoxine need not be given. Although the only reported toxic reaction to trientine is sideroblastic anemia, the same clinical procedures and laboratory determinations should be performed during its administration as are used during penicillamine therapy. Except for systemic lupus erythematosus and, occasionally, elastosis perforans serpiginosa, the other late penicillamine-induced toxic reactions disappear or improve with trientine therapy. Moreover, trientine is as effective therapeutically as penicillamine.

Zinc acetate or gluconate are effective as maintenance therapy, at doses of 150 mg/d of elemental zinc, for patients who are asymptomatic or have improved maximally on penicillamine or trientine. Zinc must *not*, however, be given together with penicillamine or trientine, both of which can chelate zinc and form complexes that are therapeutically ineffective.

Treatment must be continued for life. Inadequate treatment or interruption of therapy can be fatal or cause irreversible relapse. Indeed, of 11 patients who voluntarily discontinued penicillamine after years of successful treatment, 8 died after an average of 2.6 years of noncompliance. In contrast, of 13 patients in whom trientine was substituted because of an adverse reaction to penicillamine, 1 died accidentally, 5 were lost to follow-up, and 7 are alive and well 11 to 23 years later.

Prophylactic treatment of more than 100 asymptomatic patients with a documented diagnosis of Wilson's disease has shown that continual therapy with penicillamine or trientine can maintain the asymptomatic state indefinitely. Several such patients have been treated with penicillamine for >30 years.

Patients with severe neurologic disease who do not improve with penicillamine or trientine therapy that has reduced serum free copper to <0.3 umol/dL (15 ug/dL) may benefit significantly from treatment with dimercaprol. Intramuscular injections of 3 mL (containing 300 mg of dimercaprol) are given on five successive weekdays for 4 weeks. Treatment is interrupted for 1 week, and a second 4-week course is given. If there is neurologic improvement, additional courses are given as long as improvement continues. If no improvement is seen after two courses, it is unlikely that additional courses will be effective.

The simultaneous occurrence of fulminant hepatitis and Coombs-negative hemolytic anemia may be the initial clinical manifestation of Wilson's disease or may occur in a noncompliant patient. The syndrome is almost always fatal, usually within a week or two, unless liver transplantation is performed. Transplantation is also indicated if progressive hepatic insufficiency occurs despite adequate treatment with penicillamine or trientine.

More than 150 women with Wilson's disease treated with penicillamine and more than 20 women treated with trientine have had successful and uneventful pregnancies.

(Bibliography omitted in Palm version)

349. LYSOSOMAL STORAGE DISEASES - *Gregory A. Grabowski*

GENERAL FEATURES

Lysosomes are heterogeneous subcellular organelles containing specific hydrolyases that allow targeted processing or degradation of proteins, nucleic acids, carbohydrates, and lipids. There are >30 different lysosomal storage diseases, and they vary greatly in the types of metabolic abnormalities that occur as well as in their clinical manifestations. Nonetheless, these disorders are considered together because they share a related pathophysiology that involves the accumulation of specific macromolecules within cells that normally process large amounts of these substrates. The disorders are classified based on the nature of the stored material and include mucopolysaccharidoses (MPS), gangliosidoses, glycosphingolipidoses, glycoproteinoses, mucolipidoses, leukodystrophies, and lipid storage disorders (Table 349-1). The most prevalent lysosomal storage diseases in adults are Fabry disease, Gaucher disease, and Niemann Pick disease (NPD).

Lysosomal storage diseases should be considered in the differential diagnosis of patients with neurologic or muscular degeneration, unexplained hepatomegaly or splenomegaly, or skeletal dysplasias and deformations. Physical findings are disease-specific, and definitive diagnosis is made by enzyme assays.

PHYSIOLOGY OF LYSOSOMES

Lysosomal biogenesis is a continuous process that involves ongoing synthesis of lysosomal hydrolases, membrane constitutive proteins, and new membranes. Lysosomes originate from the fusion of *trans*-golgi network (TGN) vesicles with late endosomes. Progressive vesicular acidification accompanies the maturation of TGN vesicles, which contain various hydrolases, into lysosomes. Early endosomes have an internal pH of ~6.0 to 6.2; late endosomes and lysosomes have a pH of ~5.5 to 6.0 and £5, respectively. This gradient facilitates the pH-dependent dissociation of receptors and ligands [e.g., the mannose-6-phosphate (M6P) receptor and M6P-containing oligosaccharides] as well as activating lysosomal hydrolase function. This dynamic system, which has supplanted the static view of the lysosome, is consistent with the presence of heterogeneous populations of similar organelles whose contents differ significantly in time and location within the cell.

The accurate sorting, targeting, and activation of lysosomal enzymes is essential for maintaining normal cellular function. Abnormalities at several steps along the biosynthetic pathway can impair enzyme activation and lead to a lysosomal storage disorder. After cleavage of the hydrophobic signal peptide in the endoplasmic reticulum (ER) membrane, *N*-glycosylation occurs cotranslationally in the lumen of the ER. Complex oligosaccharide modifications occur during transit through the Golgi. The M6P modification of high-mannose oligosaccharide chains of many soluble lysosomal hydrolases occurs early in this process. Defects of this modification result in inappropriate extracellular secretion of most soluble lysosomal hydrolases, leading to severe phenotypes (I-cell disease). Lysosomal integral or associated membrane proteins (LIMPS or LAMPS) are sorted to the membrane or interior of the lysosome by several different signals. Phosphorylation, sulfation, additional proteolytic processing,

and macromolecular assembly of heteromers occur concurrently and are critical to enzyme function. Defects in the latter can result in multiple enzyme/protein deficiencies.

PATHOGENESIS OF LYSOSOMAL STORAGE DISEASES

The final common pathway for lysosomal storage diseases is the accumulation of specific macromolecules within tissues and cells that normally have a high flux of these substrates. The majority of lysosomal enzyme deficiencies result from point mutations or genetic rearrangements at a locus that encodes a single lysosomal hydrolase. However, some mutations cause deficiencies of several different lysosomal hydrolases by altering the enzymes/proteins involved in targeting, active site modifications, or macromolecular association or trafficking. All are inherited as autosomal recessive disorders, except Hunter (MPS II) and Fabry diseases, which are X-linked. Lysosomal distortion, which is caused by substrate accumulation, probably has significant pathologic consequences. However, abnormal amounts of metabolites may also have pharmacologic effects important to disease pathophysiology.

For many lysosomal diseases, the accumulated substrates are endogenously synthesized within particular tissue sites of pathology. Other diseases have greater exogenous substrate supplies; that is, they are delivered by low-density lipoprotein receptor-mediated uptake in Fabry and cholesteryl ester storage diseases or by phagocytosis in Gaucher disease type 1.

The concept of the *threshold hypothesis* has important implications for disease classification, pathophysiology, and treatment. It implies a threshold of enzyme activity below which disease develops. Consequently, small changes in enzyme activity near the threshold can lead to or prevent disease. A critical element of this model is that enzymatic activity can be challenged by changes in substrate flux based on genetic background, cell turnover, recycling, or metabolic demands. Thus, a set level of residual enzyme may be adequate for substrate in some tissues or cells, but not in others. For some lysosomal storage diseases [e.g., metachromatic leukodystrophy (MLD), Tay-Sachs disease, Gaucher disease], genotype-phenotype correlations may help to predict the severity of clinical consequences associated with certain levels of enzyme activity. Defining enzyme activity thresholds may also be useful for predicting dose-response relationships for treatment and for the evaluation of exogenous enzyme replacement therapy.

SPECIFIC DISORDERS

MUCOPOLYSACCHARIDOSES (MPS)

The various forms of MPS result from deficiencies of lysosomal enzymes needed for glycosaminoglycan (GAG) catabolism. GAGs are long-chain, complex carbohydrates that are linked to proteins in connective tissue. They are components of proteoglycans and include: chondroitin-4-sulfate, chondroitin-6-sulfate, heparan sulfate, dermatan sulfate, keratan sulfate, and hyaluronic acid. The particular accumulated GAG is determined by the specific enzyme deficiency. GAG accumulation in various tissues produces a spectrum of clinical features that can include skeletal abnormalities, corneal clouding, organomegaly, joint stiffness, hernias, short stature, and, in some disorders,

mental retardation. Vacuolated lymphocytes in the peripheral smear and excessive urinary GAGs are typical findings. Clinical and laboratory features overlap among the MPS diseases and are not diagnostic. The MPS diseases occur with individual frequencies of about 1/50,000 to 1/100,000 in most populations.

The diagnosis is established by specific enzyme assays or, when known, DNA mutation analysis. Prenatal diagnosis is conducted most frequently with cultured amniotic cells. Mutation studies for carriers can be performed, if the mutation is known.

Treatment of the MPS diseases requires comprehensive, multisystem evaluations. Symptomatic therapies currently include corneal transplantation, correction of nerve entrapment, and heart valve replacement. Physical therapy for joint contractures is needed in all but MPS IV. In MPS III variants, psychotropic drugs are used to control behavior. For patients with MPS IH and IV, cervical myelopathy can be prevented by prospective cervical spinal fusion. The efficacy of bone marrow transplantation and enzyme replacement is under investigation.

MPS IH (Hurler Disease) This a severe autosomal recessive disorder that results from numerous different mutations of a-L-iduronidase. Progressive mental retardation, hepatosplenomegaly, skeletal malformations, and cardiopulmonary compromise typically lead to death during the first decade. Affected individuals appear normal at birth but exhibit accelerated growth and mild coarsening of facial features in the first year. Subsequently, there is slowing of growth, leading to short stature. In the first 2 years, clinical diagnosis is suggested by hepatosplenomegaly, corneal clouding, coarse features, large tongue, joint stiffness, and characteristic dysostosis multiplex on skeletal x-rays. Instability of the cervical vertebral bodies can lead to paralysis, particularly with subluxation on hyperextension. Developmental delay is apparent by 12 to 28 months, with subsequent slow mental regression. Additional features of this multisystem disease include hearing loss, chronic respiratory infections, valvular heart disease, and brain ventricular enlargement. The latter occurs from involvement of the arachnoid granulations.

MPS (Scheie Disease) and MPS I H/S (Hurler-Scheie Disease) These MPS variants are less severe than MPS IH (Hurler disease). They result from allelic mutations in the a-L-iduronidase gene, presumably with a less severe effect on enzyme function. Patients with MPS IS can survive into late adulthood with normal intelligence, though with severe progressive skeletal disease that resembles osteoarthritis. Bone marrow transplantation in MPS IH, if instituted before substantial central nervous system CNS involvement, has shown therapeutic promise. Preliminary intravenous enzyme administration has led to improvement in hepatosplenomegaly and connective tissue involvement.

MPS II (Hunter Syndrome) This is an X-linked recessive disorder that results from deletions and point mutations in the gene encoding iduronate sulfatase. Clinically, MPS IH and II are similar, though corneal clouding is absent in MPS II. Clinical manifestations range from severe CNS and visceral involvement with death in late childhood to milder forms with normal CNS function and survival into adulthood. Bone marrow transplantation has not been successful for treating the severe variants, and experience in the less severe variants is too limited to permit conclusions. Enzyme therapy trials are

imminent.

MPS IIIA, IIIB, IIIC, and IIID (the Sanfilippo Syndromes) These autosomal recessive disorders are caused by various enzymatic deficiencies as summarized in Table 349-1. Skeletal defects and hepatosplenomegaly are less pronounced in this group of MPS variants, though progressive behavioral problems, mental retardation, and seizures are present. Affected patients can survive into the third or fourth decade with progressive CNS disease.

MPS IV (Morquio Syndrome) These MPS variants are autosomal recessive disorders characterized by severe skeletal diseases that resemble the spondyloepiphyseal dysplasias. There is extreme shortening of the trunk due to multiple vertebral collapses. The long bones are relatively spared. Joint laxity can lead to osteoarthritis-like destruction of the joints. Upper cervical spinal cord compression due to atlantoaxial instability predisposes to subluxation and paralysis. Many patients have mitral valve insufficiency that can be functionally significant. The A and B variants are distinguished clinically by more severe skeletal disease, *N*-acetylgalactosamine-6-sulfate sulfatase deficiency in A, than in b-galactosidase defects in B. Enzyme therapy trials will begin in the near future.

MPS VI (Maroteaux-Lamy Disease) Mutations in the arylsulfatase B gene cause this autosomal recessive disorder. Although clinically variable, the general phenotype resembles Hurler disease. Intelligence is normal, and the life span can extend beyond three decades. Cardiac valvular disease and progressive pulmonary hypertension are frequent causes of death. Bone marrow transplantation may be useful in diminishing these manifestations.

GM$_2$GANGLIOSIDOSES

The Tay-Sachs and Sandhoff disease variants are caused by defects in b-hexosaminidase (Hex) A and/or B. Hex A is a heteromeric protein with a and b chains, whereas Hex B contains only b chains. The a and b chains are encoded by different genes. Infantile, juvenile, and adult-onset variants are distinguished by age at onset and rate of progression.

In addition to other clinical manifestations described below, specific neurologic features, such as the retinal cherry red spot, suggest the diagnosis of Tay-Sachs and Sandhoff diseases. Diagnosis is confirmed by Hex A and/or B levels in blood plasma or nucleated cells. Screening for Tay-Sachs disease carriers in the Ashkenazi Jewish population is recommended.

Tay-Sachs Disease About 1 in 30 Ashkenazi Jews is a carrier for Tay-Sachs disease, which is caused by total Hex A deficiency. The infantile form is a fatal neurodegenerative disease that is characterized by macrocephaly, loss of motor skills, increased startle reaction, and macular pallor with cherry red spot on retinal examination. The juvenile-onset form presents with ataxia and dementia, with death by age 10 to 15 years. The adult-onset disorder is characterized by clumsiness in childhood; progressive motor weakness in adolescence; and additional spinocerebellar, lower motor neuron symptoms, and dysarthria in adulthood. Intelligence declines slowly, and psychosis is

also common.

Sandhoff Disease Sandhoff disease is nearly identical to Tay-Sachs disease, though hepatosplenomegaly and bony dysplasias are present in the former. The later onset variants are characterized by progressive visceral and CNS disease. Treatment is supportive.

NEUTRAL GLYCOSPHINGOLIPID LIPID STORAGE DISORDERS

Fabry Disease This is an X-linked disorder that results from a variety of mutations in the a-galactosidase gene. This enzyme cleaves the terminal a-galactosyl moiety from globotriaosylceramide (trihexosylceramide, THC), a key step in glycosphingolipid metabolism. Clinically, the disease manifests with angiokeratomas (telangiectatic skin lesions); hypohidrosis; corneal and lenticular opacities; acroparesthesia; and small-vessel disease of the kidney, heart, and brain. The estimated prevalence of hemizygous males with Fabry disease is 1/40,000.

The angiokeratomas and acroparesthesia may appear in childhood and lead to early diagnosis, if suspected. Angiokeratomas are punctate, dark red to blue-black, flat or slightly raised, usually symmetric, and do not blanch with pressure. They range from barely visible to several millimeters in diameter and have a tendency to increase in size and number with age. Characteristically, they are most dense between the umbilicus and knees -- "the bathing suit area" -- but may occur anywhere, including the mucosal surfaces. Corneal and lenticular lesions, detectable on slit-lamp examination, are present in affected men and ~70% of heterozygous women. Tortuosity of the conjunctival and retinal vessels is common. The acroparesthesia can be debilitating in childhood and adolescence, with a tendency to decrease after the third decade. Episodic agonizing, burning pain of the hands, feet, and proximal extremities can last from minutes to days and can be precipitated by exercise, fatigue, or fever. Abdominal pain can resemble that from appendicitis or renal colic.

Casts and microscopic hematuria can occur early, whereas proteinuria, isosthenuria, and progressive renal dysfunction occur in the second to fourth decades. Progressive renal failure occurs and requires transplantation. Hypertension, left ventricular hypertrophy, anginal chest pain with or without myocardial ischemia or infarction, and congestive heart failure can occur in the third to fourth decades. Leg lymphedema without hypoproteinemia and episodic diarrhea also occur. Death is due to renal failure or cardiovascular or cerebrovascular disease in untreated patients. Variants with residual a-galactosidase activity may have late-onset manifestations limited to the cardiovascular system that resemble hypertrophic myocardiopathies. Heterozygous females may exhibit some of these clinical manifestations but usually not the severe organ involvement.

Acroparesthesia, hypohidrosis or anhidrosis, angiokeratomas, and the typical corneal and lenticular lesions provide a presumptive diagnosis in males. Angiokeratomas are not diagnostic, however, and also occur in Fordyce scrotal angiokeratoma and several other lysosomal storage diseases.

Phenytoin and carbamazepine diminish the chronic and episodic acroparesthesia.

Chronic hemodialysis or kidney transplantation can be lifesaving in patients with renal failure. Initial enzyme therapy results are promising.

Gaucher Disease This is an autosomal recessive disorder that results from defective activity of acidb-glucosidase; >175 mutations have been described. This enzyme cleaves glucosylceramide, the parent compound of many glycosphingolipids and related glucolipids. Disease variants are classified based on the absence or presence and severity of neuronopathic involvement.

Type 2 Gaucher disease is a rare, severeCNSdisease that leads to death by 2 years of age; it will not be addressed here.

Type 3 Gaucher disease has highly variable manifestations in theCNS and viscera. It can present in early childhood with rapidly progressive, massive visceral disease and slowly progressive to static CNS involvement; in adolescence with dementia; or in early adulthood with rapidly progressive, uncontrollable myoclonic seizures and mild visceral disease. Variants that span this spectrum also occur. Visceral disease in type 3 is nearly identical to that in type 1 but is generally more severe (see below). Early CNS findings may be limited to defects in lateral gaze tracking, which may remain static for decades. Mental retardation can be slowly progressive or static. This variant is most frequent among individuals of Swedish descent.

Type 1 Gaucher disease is a highly variable nonneuronopathic disease that can present in childhood to adulthood with slowly to rapidly progressive visceral disease. There is marked variability in age at onset and degree and progression of visceral involvement. In general, earlier diagnoses are associated with worse prognosis. The average age at diagnosis is ~20 years in Caucasian populations and somewhat younger in other groups. This pattern of presentation is distinctly bimodal, however, with peaks at <10 to 15 years and at ~25 years. Younger patients tend to have a greater degree of hepatosplenomegaly and accompanying blood cytopenias. In contrast, the older group has a greater tendency for chronic bone disease. Hepatosplenomegaly occurs in virtually all symptomatic patients and can be minor or massive. Accompanying anemia and thrombocytopenia are variable and are not linearly related to liver or spleen volume. Severe liver dysfunction is unusual, though minor liver function abnormalities are common. Splenic infarctions can resemble an acute abdomen. Pulmonary hypertension and alveolar Gaucher cell accumulation are uncommon, but life-threatening, and can occur at any age.

Though it is more common in adult patients, clinically evident skeletal disease in children can be devastating, resulting in massive destruction of the axial and peripheral skeleton. All patients with Gaucher disease have nonuniform infiltration of bone marrow by lipid-laden macrophages, termed *Gaucher cells*. This can lead to marrow packing with subsequent infarction, ischemia, necrosis, and cortical bone destruction. Bone marrow involvement spreads from proximal to distal in the limbs and can involve the axial skeleton extensively, causing vertebral collapse. In addition to bone marrow involvement, bone remodeling is defective, with loss of total bone calcium leading to osteopenia, osteonecrosis, avascular infarction, and vertebral compression fractures and spinal cord involvement. Aseptic necrosis of the femoral head is common, as is fracture of the femoral neck. The mechanism by which diseased bone marrow

macrophages interact with osteoclasts and/or osteoblasts to lead to this complex bone disease is not well understood.

Affected patients experience chronic, ill-defined bone pain that can be debilitating and poorly correlated with radiographic findings. These are treated symptomatically. Some patients have one or more "bone crises" in their lifetimes that are associated with localized, excruciating pain, and, on occasion, local erythema, fever, and leukocytosis. Some patients have frequent crises, whereas other patients experience only one. Any bone can be involved, though the femurs and vertebral bodies are affected most often. These crises represent acute infarctions of bone, as evidenced in nuclear scans by localized absent uptake of pyrophosphate agents. X-rays are usually negative initially but may show lytic lesions 4 to 6 months after the acute phase. Osteomyelitis should be excluded by appropriate cultures. Bone cultures should be obtained only under sterile operating room conditions to minimize the chance of seeding an infection.

The diagnosis of Gaucher disease is established by demonstrating decreased acidb-glucosidase activity (0 to 20% of normal) in nucleated cells. The enzyme is not present in bodily fluids. The sensitivity of enzyme testing is poor for detecting heterozygous carriers; molecular testing is preferred when the mutations are known. The disease frequency varies from about 1 in 1000 in Ashkenazi Jews to <1 in 100,000 in some other populations. About 1 in 12 to 15 Ashkenazi Jews carries a Gaucher disease allele. Four common mutations account for ~90 to 95%of the mutations in affected patients: N370S (1226G), 84GG (a G insertion at cDNA position 84), L444P (1448C), and IVS-2 (an intron 2 splice junction mutation).

Genotype/phenotype studies indicate a significant correlation, though not absolute, between disease type and severity and the acid b-glucosidase genotype. For example, the most common mutation in the Ashkenazi Jewish population (N370S) shares a 100% association (to date) with nonneuronopathic, type 1 Gaucher disease, possibly because the N370S enzyme retains significant activity. The N370S/N370S and N370S/other mutant allele genotypes are associated with later onset/less severe and earlier onset/severe disease, respectively. The other alleles are L444P (very low activity), 84GG (null), or IVS-2 (null), and rare/private or uncharacterized alleles. As many as 50 to 60% of N370S/N370S patients are discovered as asymptomatic family members. The N370S/other mutant allele genotypes have disease onset about 2 decades earlier than those with N370S/N370S. The L444P/L444P patients almost always have life-threatening to very severe/early-onset disease, and many, though not all, developCNSinvolvement in the first 2 decades of life. Some patients with this genotype have lethal neuronopathic disease at <1 year (type 2). Thus, this genotype is prognostic of very severe disease, with or without obvious CNS involvement.

Symptomatic management of the blood cytopenias and joint replacement surgeries continue to have important roles in the treatment of affected patients. However, enzyme therapy is currently the treatment of choice in significantly affected patients. Cerezyme, a recombinantly produced mannose-terminated (macrophage-targeted) acidb-glucosidase, has proved highly efficacious and safe in diminishing the hepatosplenomegaly and improving bone marrow involvement and hematologic findings.

Niemann-Pick Disease NPD is an autosomal recessive trait that occurs as several variants. Types A and B result from defects in acid sphingomyelinase; various mutations have been detected at this locus. Other variants, including NPD C, result from defective transport of cholesterol across the lysosomal membrane. Only NPD variants A and B will be considered here.

NPDA and B are distinguished primarily by an early age of onset and progressiveCNSdisease in A. NPD A typically has onset in the first 6 months, with rapidly progressive CNS deterioration, spasticity, failure to thrive, and massive hepatosplenomegaly. In contrast, NPD B has a later, more variable onset and progression of hepatosplenomegaly, with eventual development of cirrhosis and hepatic replacement by foam cells. Affected patients develop progressive pulmonary disease with dyspnea, hypoxemia, and a reticular infiltrative pattern on chest x-ray. Foam cells are present in alveoli, lymphatic vessels, and pulmonary arteries. Progressive hepatic or lung disease with associated bronchopneumonia, pulmonary hypertension, cor pulmonale, and decreased diffusion capacities lead to demise in adolescence to early adulthood.

The diagnosis is established by markedly decreased (1 to 10% of normal) sphingomyelinase activity in nucleated cells. Enzyme assays to detectNPD A or B carriers are unreliable. In families with known mutations, the molecular defect in heterozygotes can be identified by DNA analysis.

There is no specific treatment forNPD. The efficacy of hepatic or bone marrow transplantation has not been proven. Clinical trials using enzyme therapy are anticipated to begin soon.

THE LEUKODYSTROPHIES

Globoid cell leukodystrophy and metachromatic leukodystrophy variants are autosomal recessive disorders that primarily involveCNSwhite matter and myelinated peripheral nervous system tracts.

Globoid Cell Leukodystrophy (GCL) The GCL variants are due to mutations in the galactosylceramidase gene. The term *globoid cell* is derived from the presence of characteristic multinucleated cells filled with galactosylceramide in the brains of affected patients. Infantile GCL (Krabbe disease) is a rapidly progressive, fatal disorder; patients succumb in the first 2 years. Juvenile and adult variants present with more slowly progressive dementia. For all variants, manifestations are confined to theCNS and peripheral nervous system. Diagnosis is confirmed by demonstrating defective galactosylceramidase activity in nucleated cells. Treatment is supportive, but bone marrow transplantation has shown some promise in the later onset variants.

Metachromatic Leukodystrophy MLD is due to defects in arylsulfatase A and accumulation of its substrate, galactosylceramide sulfate or sulfatide. The late-infantile form presents in the second year with progressive regression of developmental milestones and intellectual development. The disease is fatal in the first decade. The juvenile and adult forms have variable manifestations and present with gait disturbances, ataxia, mental regression, peripheral neuropathy, and/or seizures. In

adults, behavioral disturbances, psychosis, and dementia tend to predominate. These later onset diseases may respond to bone marrow transplantation.

The diagnosis is established by demonstrating a deficiency of arylsulfatase A in nucleated cells. Homozygosity for a null allele (splicing defect in intron 2) produces severe infantile disease, whereas P426L homozygotes develop adult-onset disease. Compound heterozygotes, such as null and P426L alleles, have juvenile-onset variants. The diagnosis of MLD is complicated by the presence of a very frequent (10 to 15%) "pseudodeficiency allele" for arylsulfatase A. Although in vitro activity with synthetic substrates is deficient, cleavage of sulfatide is low-normal in vivo. Compound heterozygotes for the pseudodeficiency allele and true MLD alleles therefore appear to have deficient arylsulfatase A activity, diagnostic of MLD, but these individuals are not affected by the disease. MLD-causing mutations also occur on the background of the pseudodeficient allele, emphasizing the importance of careful genetic testing.

GLYCOGEN STORAGE DISEASE TYPE II (See also Chap. 350)

Glycogen storage disease type II is an autosomal recessive disorder due to defects in acid a-glucosidase that lead to lysosomal glycogen accumulation. Numerous mutations have been found at this locus in affected patients. Skeletal and cardiac muscles are primarily involved. The infantile form (Pompe disease) is a fatal disorder characterized by hypertrophic cardiomegaly, macroglossia, and hypotonia due to glycogen accumulation in muscle. The juvenile form has progressive proximal muscle weakness, including impairment of respiratory function. Adult patients have phenotypes resembling slowly progressive muscular dystrophies. Prenatal diagnosis can be performed. Treatment is supportive, and enzyme trials are underway.

MUCOLIPIDOSES

Mucolipidoses II (I-cell) and III (pseudo-Hurler polydystrophy) are rare autosomal recessive diseases. Both are caused by defective targeting of lysosomal hydrolases that require the M6P signal for sorting to the lysosome. As a result, >20 enzymes are secreted out of the cell and their substrates accumulated in specific cell types. *N*-acetylglucosamine-1-phosphotransferase activity, which is necessary for developing the M6P signal, is defective in these diseases. I-cell disease has a phenotype similar to MPS IH, whereas mucolipidosis-III is more similar to MPS IH/S; mental retardation is a feature of both. Diagnosis is suspected based on the characteristic phenotype and is established by demonstrating greatly elevated serum levels of lysosomal enzymes, as well as their deficiency in cells. Specific enzyme assays can also be performed. Carrier detection is possible but is not straightforward. Prenatal diagnosis can be performed using amniotic fluid and cells in at-risk families. Treatment is supportive.

(Bibliography omitted in Palm version)

Back to Table of Contents

350. GLYCOGEN STORAGE DISEASES AND OTHER INHERITED DISORDERS OF CARBOHYDRATE METABOLISM - *Yuan-Tsong Chen*

Carbohydrate synthesis and degradation play a vital role in cellular function by providing the energy required for most metabolic processes. The carbohydrates to be discussed include three monosaccharides: glucose, galactose, and fructose, and a polysaccharide, glycogen; the relevant biochemical pathways involved in the metabolism of these carbohydrates are shown in Fig. 350-1. Glucose is the principle substrate of energy metabolism in humans. Metabolism of glucose generates ATP via glycolysis and mitochondrial oxidative phosphorylation. A continuous source of glucose from dietary intake, gluconeogenesis, and degradation of glycogen maintain normal blood glucose levels. Sources of glucose in our diet are obtained by ingesting polysaccharides, primarily starch, and disaccharides including lactose, maltose, and sucrose. Galactose and fructose are two other monosaccharides that provide fuel for cellular metabolism; however, their role as fuel sources is much less significant than that of glucose. Galactose is derived from lactose (galactose+ glucose), which is found in milk and milk products. If necessary, galactose can be incorporated into glycogen and thus becomes a source of glucose. Galactose is also an important component for certain glycolipids, glycoproteins, and glycosaminoglycans. The two dietary sources of fructose are sucrose (fructose + glucose), a commonly used sweetener, and fructose itself, which is found in fruits, vegetables, and honey.

This chapter is devoted to the inherited disorders of carbohydrate metabolism caused by defects in enzymes or transport proteins involved in glycogen metabolism, gluconeogenesis, and glycolysis (Table 350-1). Defects in glycogen metabolism typically cause an accumulation of glycogen in the tissues; hence, the name *glycogen storage diseases*. The defects in gluconeogenesis or glycolytic pathways including galactose and fructose metabolism do not usually result in glycogen accumulation.

Clinical manifestations of the various disorders of carbohydrate metabolism differ markedly. The symptoms range from harmless to lethal. Unlike disorders of lipid metabolism, mucopolysaccharidoses, or other storage diseases, dietary therapy has been effective in many of the carbohydrate disorders. Almost all the genes responsible for the inherited defects of carbohydrate metabolism have been cloned, and mutations have been identified. Advances in our understanding of the molecular basis of these diseases are being used to improve diagnosis and management, and some of these disorders are candidates for early trials of gene therapy.

Glycogen, the storage form of glucose in animal cells, is composed of glucose residues joined in straight chains bya1-4 linkages and branched at intervals of 4 to 10 residues with a1-6 linkages. The treelike molecule can have a molecular weight of many millions and may aggregate to form structures recognizable by electron microscopy. In muscle, glycogen formsb particles, which are spherical and contain up to 60,000 glucose residues. Each b particle contains a covalently linked protein called *glycogenin*. Liver containsb particles and rosettes of glycogen calleda particles, which appear to be aggregatedb particles.

The primary function of glycogen varies in different tissues. In skeletal muscle, stored glycogen is a source of fuel that is used for short-term, high-energy consumption during

muscle activity; in the brain, the small amount of stored glycogen is used during brief periods of hypoglycemia or hypoxia as an emergency supply of energy. In contrast, the liver takes up glucose from the bloodstream after a meal and stores it as glycogen. When blood glucose levels start to fall, the liver converts glycogen back into glucose and releases it into the blood for use by tissues such as brain and erythrocytes that cannot store significant amounts of glycogen.

Glycogen storage diseases are inherited disorders that affect glycogen metabolism. Disorders in virtually every enzyme involved in the synthesis or degradation of glycogen and its regulation cause some type of glycogen storage disease (Fig. 350-1) in which glycogen is abnormal in quantity, quality, or both. Excluded from this chapter are those conditions in which tissue glycogen accumulation is secondary, such as overtreatment of diabetes mellitus with insulin or administration of pharmacologic amounts of glucocorticoids.

Historically, the glycogen storage diseases were categorized numerically in the order in which the enzymatic defects were identified. They can also be classified by the organs involved and clinical manifestations, the system followed in this chapter (Table 350-1).

Because liver and muscle have abundant glycogen, they are the most commonly and seriously affected tissues. The hepatic glycogen storage diseases can be divided into two groups, with some overlap. The first is characterized by hepatomegaly and hypoglycemia. Because carbohydrate metabolism in the liver controls plasma glucose levels, the disorders of hepatic glycogen degradation and glucose release cause fasting hypoglycemia. Diseases in this group include glucose-6-phosphatase deficiency (type I), debranching enzyme deficiency (type III), liver phosphorylase deficiency (type VI), phosphorylase kinase deficiency (type IX), glycogen synthase deficiency (type 0), and glucose transporter-2 defects (type XI). The second group, characterized by cirrhosis of the liver and hepatomegaly, is associated with accumulation of abnormal forms of glycogen, which may be the cause of the hepatocellular injury. This group is represented by branching enzyme deficiency (type IV).

The role of glycogen in muscle is to provide substrates for the generation of sufficient ATP for muscle contraction. The muscle glycogen storage diseases can also be divided into two groups. The first is a muscle-energy disorder characterized by muscle pain, exercise intolerance, myoglobinuria, and susceptibility to fatigue. This group includes type V (McArdle disease), a muscle phosphorylase deficiency, and deficiencies of phosphofructokinase (type VII), phosphoglycerate kinase, phosphoglycerate mutase, lactase dehydrogenase, fructose 1,6-biphosphate aldolase A, and pyruvate kinase. Some of these latter enzyme deficiencies are associated with a compensated hemolysis, suggesting a more generalized defect in glucose metabolism. The second group of muscle disorders is characterized by progressive skeletal muscle weakness and atrophy and/or cardiomyopathy; it includes a lysosomal enzyme deficiency (acida-glucosidase, type II) and deficiency of cardiac-specific phosphorylase kinase. Some glycogen storage diseases such as debranching enzyme deficiency (type IIIa) and branching enzyme deficiency (type IV) involve both muscle and liver.

The overall frequency of all forms of glycogen storage disease is approximately 1 in 20,000 live births; most are inherited as autosomal recessive traits, but

phosphoglycerate kinase deficiency and one form of phosphorylase kinase deficiency are X-linked disorders. The most common childhood disorders are glucose-6-phosphatase deficiency (type I), lysosomal acida-glucosidase deficiency (type II), debrancher deficiency (type III), and liver phosphorylase kinase deficiency (type IX). The most common adult disorder is myophosphorylase deficiency (type V, or McArdle disease). In the past, the prognosis for many glycogen storage diseases was guarded. However, early diagnosis and better management have improved the survival rates, and many affected children are now adults.

GLYCOGEN STORAGE DISEASES: LIVER GLYCOGENOSES

DISORDERS WITH HEPATOMEGALY AND HYPOGLYCEMIA

Type I Glycogen Storage Disease (Glucose-6-Phosphatase or Translocase Deficiency, von Gierke Disease) Type I glycogen storage disease is due to a defect in glucose-6-phosphatase in liver, kidney, and intestinal mucosa. It can be divided into two subtypes: type Ia, in which the glucose-6-phosphatase enzyme is defective, and type Ib, which is due to a defect in the translocase that transports glucose-6-phosphate across the microsomal membrane. The defects in both subtypes lead to inadequate conversion in the liver of glucose-6-phosphate to glucose and thus make affected individuals susceptible to fasting hypoglycemia.

GENETIC CONSIDERATIONS

Type I glycogen storage disease is an autosomal recessive disorder. Both types Ia and Ib disease have been reported in many ethnic groups, but type Ia is rarely seen in blacks. The structural gene for glucose-6-phosphatase is located on chromosome 17q21; three common mutations (R83C, 130X, Q347X) are responsible for 70% of the known disease alleles. The structural gene for glucose 6-phosphate translocase is located on chromosome 11q23; two mutations, G339C and 1211delCT, appear to be prevalent in Caucasian patients, while W118R appears to be most common in Japanese patients. Carrier detection and prenatal diagnosis are possible with the use of molecular techniques.

Clinical and laboratory findings Persons with type I disease may develop hypoglycemia and lactic acidosis during the neonatal period, but, more commonly, they present at 3 to 4 months of age with hepatomegaly and/or hypoglycemia. These children often have doll-like faces with fat cheeks, relatively thin extremities, short stature, and a protuberant abdomen that is due to massive hepatomegaly; the kidneys are enlarged, but the spleen and heart are of normal size.

The hallmarks of the disease are hypoglycemia, lactic acidosis, hyperuricemia, and hyperlipidemia. Hypoglycemia and lactic acidosis can develop after a short fast. Hyperuricemia is present in young children, but gout rarely develops before puberty. Despite hepatomegaly, liver enzymes are usually normal or near normal. Intermittent diarrhea may occur (the mechanism is not known). Easy bruising and epistaxis are associated with a prolonged bleeding time as a result of impaired platelet aggregation/adhesion.

Hypertriglyceridemia may cause the plasma to appear "milky," and cholesterol and phospholipids are also elevated. The lipid abnormality resembles type IV hyperlipidemia and is characterized by increased levels of very low-density lipoprotein (VLDL); low-density lipoprotein (LDL); increased levels of apoliproteins B, C, and E; and normal or reduced levels of apoliproteins A and D. The hepatocytes are distended by glycogen and fat with large and prominent lipid vacuoles. There is little associated fibrosis.

All these findings apply to both types Ia and Ib disease, but type Ib has the additional feature of recurrent bacterial infections due to neutropenia and impaired neutrophil function. Oral and intestinal mucosa ulcerations are common, and inflammatory bowel disease may occur.

Long-term complications Although type I glycogen storage disease mainly affects the liver, multiple organ systems are also involved. Gout usually becomes symptomatic around puberty as a result of the long-term hyperuricemia. Puberty is often delayed, but fertility appears to be normal. Hypertriglyceridemia causes an increased risk of pancreatitis, but premature atherosclerosis has not been documented. Impaired platelet aggregation may reduce the risk of atherosclerosis.

By the second or third decade of life, most patients with type I glycogen storage disease develop hepatic adenomas that can hemorrhage and, in rare cases, may become malignant. Other complications include pulmonary hypertension and osteoporosis.

Renal disease is a late complication, and almost all patients older than 20 years have proteinuria. Many have hypertension, kidney stones, nephrocalcinosis, and altered creatinine clearance. Glomerular hyperfiltration, increased renal plasma flow, and microalbuminuria can occur before the onset of gross proteinuria. In young patients, hyperfiltration and hyperperfusion may be the only signs of renal abnormalities. With advanced renal disease, focal segmental glomerulosclerosis and interstitial fibrosis are evident on biopsy. In some patients, renal function deteriorates and progresses to failure, requiring dialysis or transplantation. Other abnormalities in renal function include amyloidosis, Fanconi-like syndrome, and distal renal tubular acidification defect. The increases in renal perfusion and maternal blood volume that normally occur in pregnancy can exacerbate renal problems. In addition, hypoglycemia may also become more difficult to control.

Diagnosis The diagnosis of type I disease can be suspected on the basis of clinical presentation and abnormal plasma lactate and lipid values. In addition, administration of glucagon or epinephrine causes little or no rise in blood glucose but increases lactate levels significantly. Before the glucose 6-phosphatase and glucose 6-phosphate translocase genes were cloned, a definitive diagnosis required a liver biopsy to demonstrate a deficiency. Gene-based mutation analysis now provides a noninvasive way of diagnosis for most patients with types Ia and Ib disease.

TREATMENT

Treatment is designed to maintain normal blood glucose levels and is achieved by continuous nasogastric infusion of glucose or oral administration of uncooked cornstarch. Nasogastric drip feeding in early infancy may consist of an elemental enteral

formula or may contain only glucose to maintain normoglycemia during the night; frequent feedings with a high-carbohydrate content are given during the day.

Uncooked cornstarch acts as a slow-release form of glucose and can be given at a dose of 1.6 g/kg every 4 h for infants younger than 2 years. As the child grows older, the cornstarch regimen can be changed to every 6 h, and it can be given by mouth as a liquid (1:2, weight:volume) at a dose of 1.75 to 2.5 g/kg of body weight. Because fructose and galactose cannot be converted to free glucose, their dietary intake should be restricted, and dietary supplements of multivitamins and calcium are required. Allopurinol is given to lower the levels of uric acid. In patients with type Ib disease, granulocyte and granulocyte-macrophage colony stimulating factors have been used successfully to correct the neutropenia, decrease the severity of bacterial infection, and improve the chronic inflammatory bowel disease.

Before surgery, the bleeding status of the patient should be evaluated, and good metabolic control should be established. Prolonged bleeding time can be corrected by the administration of a constant intravenous glucose infusion for 24 to 48 h before surgery. Vasopressin can be given during surgery to reduce bleeding complications, and normal glucose levels should be maintained throughout surgery.

Prognosis In the past, many patients with type I glycogen storage disease died, and the prognosis was guarded for those who survived. The long-term complications discussed above occur mostly in adults whose disease was not adequately treated during childhood. Early diagnosis and initiation of effective treatment have improved the outcome, but it is not known if all long-term complications can be avoided through good metabolic control.

Type III Glycogen Storage Disease (Debrancher Deficiency, Limit Dextrinosis)
Type III glycogen storage disease is caused by a deficiency of glycogen debranching enzyme. Debranching enzyme and phosphorylase are responsible for complete degradation of glycogen; when debranching enzyme is defective, glycogen breakdown is incomplete, and an abnormal glycogen accumulates that has short outer chains and resembles limit dextrin.

GENETIC CONSIDERATIONS

The type III glycogenoses are inherited as autosomal recessive traits. The disease has been reported in many different ethnic groups, and the frequency is relatively high in non-Ashkenazi Jews of North African descent. The gene for debranching enzyme is located on chromosome 1p21. At least 20 different mutations that cause type III disease have been identified. Two mutations (17delAG and Q6X), both located in exon 3 at amino acid codon 6, are exclusively found in the subtype IIIb. Carrier detection and prenatal diagnosis are possible with DNA-based linkage or mutation analysis.

Clinical and laboratory findings Deficiency of glycogen debranching enzyme causes hepatomegaly, hypoglycemia, short stature, variable skeletal myopathy, and cardiomyopathy. The disorder usually involves both liver and muscle and is termed *type IIIa glycogen storage disease*. However, in about 15% of patients, the disease appears to involve only the liver and is classified as *type IIIb*.

During infancy and childhood, the disease may be almost indistinguishable from type I disease because hepatomegaly, hypoglycemia, hyperlipidemia, and growth retardation are common features of both. Splenomegaly may be present, but the kidneys are not enlarged in type III disease. Remarkably, hepatomegaly and hepatic symptoms in most patients with type III disease improve with age and usually disappear after puberty. However, progressive liver cirrhosis with failure may occur and seems especially common in Japanese patients.

In patients with muscle involvement (type IIIa), muscle weakness is usually minimal during childhood but can become severe during the third or fourth decade of life, as evidenced by slowly progressive weakness and muscle wasting. Electromyographic (EMG) changes are consistent with a widespread myopathy, and nerve conduction may be abnormal. Ventricular hypertrophy is frequent, but overt cardiac dysfunction is rare. Hepatic symptoms may be so mild that the diagnosis is not made until adulthood, when neuromuscular disease becomes manifest. Polycystic ovaries appear to be a common finding in female patients; fertility, however, does not seem to be affected.

Hypoglycemia, hyperlipidemia, and elevated liver transaminases occur in children. In contrast to type I disease, fasting ketosis is prominent, and blood lactate and uric acid concentrations are usually normal. The administration of glucagon 2 h after a carbohydrate meal causes a normal rise of blood glucose, but after an overnight fast glucagon may provoke no change in blood glucose. Serum creatine kinase levels can sometimes be used to identify patients with muscle involvement, but normal levels do not rule out muscle enzyme deficiency.

The histology of the liver is characterized by a universal distention of hepatocytes by glycogen and by the presence of fibrous septa. The fibrosis and the paucity of fat distinguish type III from type I glycogenosis. The fibrosis can range from minimal periportal fibrosis to micronodular cirrhosis.

Diagnosis In type IIIa glycogen storage disease, deficient debranching enzyme activity can be demonstrated in liver, skeletal muscle, and heart. In contrast, patients with type IIIb have debranching enzyme deficiency in the liver but not in muscle. In the past, definitive assignment of subtype required enzyme assays in both liver and muscle. DNA-based analyses now provide a noninvasive way of subtyping these disorders in most patients.

TREATMENT

Dietary management of type III disease is less demanding than that of type I. If hypoglycemia is present, frequent high-carbohydrate meals with cornstarch supplements or nocturnal gastric drip feedings are usually effective. A high-protein diet during the day plus overnight protein enteral infusion may be tried in patients with myopathy, but it is not established whether such a regimen is effective. Patients do not need to restrict dietary intake of fructose and galactose, as do those with type I disease.

Prognosis Liver symptoms improve with age and usually disappear after puberty. Cirrhosis of the liver may occur later in life. In type IIIa disease, muscle weakness and

atrophy worsen during adulthood.

Type VI Glycogen Storage Disease [Liver Phosphorylase Deficiency (Hers Disease)] The number of patients with enzymatically documented liver phosphorylase deficiency is small. It appears that patients with liver phosphorylase deficiency have a benign course. These patients present with hepatomegaly and growth retardation early in childhood. Hypoglycemia, hyperlipidemia, and hyperketosis are usually mild if present. Plasma lactic acid and uric acid levels are normal. The heart and skeletal muscles are not involved. The hepatomegaly and growth retardation improve with age and usually disappear at puberty. Treatment is symptomatic. A high-carbohydrate diet and frequent feeding are effective in preventing hypoglycemia, but most patients require no specific treatment. The liver phosphorylase gene is located on chromosome 14q21. A splicing site mutation in intron 13 has been identified in a large Mennonite kindred, and four other mutations have been found in patients with different ethnic backgrounds.

Type IX Glycogen Storage Disease (Liver Phosphorylase Kinase Deficiency) Defects of phosphorylase kinase cause a heterogeneous group of glycogenoses. The heterogeneity is due to the complexity of the phosphorylase kinase enzyme complex. It consists of four subunits (a, b,g, and d), each encoded by different genes (X chromosome as well as autosomes) that are differentially expressed in various tissues. Phosphorylase kinase deficiency can be divided into several subtypes on the basis of the gene/subunit involved, the tissues that are primarily affected, and the mode of inheritance.

Subtypes of Phosphorylase Kinase Deficiency

X-LINKED LIVER PHOSPHORYLASE KINASE DEFICIENCY X-linked liver phosphorylase kinase deficiency is one of the most common liver glycogenoses. Phosphorylase kinase activity may also be deficient in erythrocytes and leukocytes but is normal in muscle. Typically, a child between the ages of 1 and 5 years presents with growth retardation and hepatomegaly. Levels of cholesterol, triglycerides, and liver enzymes are mildly elevated. Ketosis may occur after fasting. Lactic and uric acid levels are normal. Hypoglycemia is mild, if present. The rise in the blood glucose level after the administration of glucagon is normal. Hepatomegaly and abnormal blood chemistries gradually return to normal with age. Most adults achieve a normal final height and are practically asymptomatic, despite a persistent phosphorylase kinase deficiency.

Liver histology shows glycogen-distended hepatocytes. The accumulated glycogen (a particles, rosette form) has a frayed or burst appearance and is less compact than in type I or type III disease. Fibrous septa and low-grade inflammatory changes may be present.

The structural gene for the liver isoform of the phosphorylase kinase a-subunit is located on chromosome Xp22, and mutations of this gene have been found in the disorder. Subtle mutations tend to retain the phosphorylase kinase activity in blood cells, whereas nonsense mutations cause enzyme deficiency in both liver and blood cells.

AUTOSOMAL LIVER AND MUSCLE PHOSPHORYLASE KINASE DEFICIENCY An autosomal recessive form of liver and muscle phosphorylase kinase deficiency has

been reported in several patients. As in the X-linked form of the disorder, hepatomegaly and growth retardation are the predominant symptoms in early childhood. Some patients also exhibit muscle hypotonia and have reduced activity of phosphorylase kinase in muscle. This form of the phosphorylase kinase deficiency is caused by mutations in theb-subunit of the gene located on chromosome 16q12-13.

AUTOSOMAL LIVER PHOSPHORYLASE KINASE DEFICIENCY In contrast to the benign course of X-linked phosphorylase kinase deficiency, patients with the autosomal recessive form of liver phosphorylase kinase deficiency have more severe phenotypes and often develop liver cirrhosis. This form of phosphorylase kinase deficiency is due to mutations in the testis/liver isoform of the g-subunit gene located on chromosome 16p.

MUSCLE-SPECIFIC PHOSPHORYLASE KINASE DEFICIENCY Muscle-specific phosphorylase kinase deficiency causes cramps and myoglobinuria on exercise or progressive muscle weakness and atrophy. The activity of the enzyme is decreased in muscle but normal (when determined) in liver and blood cells. There is no hepatomegaly or cardiomegaly. The disorder may be due to mutation in the muscle isoform of thea-subunit located on the X chromosome.

CARDIAC-SPECIFIC PHOSPHORYLASE KINASE DEFICIENCY Several sporadic cases of cardiac-specific phosphorylase kinase deficiency have been reported. All patients died during infancy from cardiac failure due to massive glycogen deposition in the myocardium. The molecular basis has not been defined.

Diagnosis Definitive diagnosis of phosphorylase kinase deficiency requires demonstration of the enzymatic defect in affected tissues. Although phosphorylase kinase can be measured in leukocytes and erythrocytes, the enzyme has many tissue-specific isozymes, and the diagnosis can be missed without studies of the liver, muscle, or heart.

TREATMENT

The treatment for liver phosphorylase or phosphorylase kinase deficiency is based on symptoms. A high-carbohydrate diet and frequent feedings are effective in preventing hypoglycemia, but most patients require no specific treatment. Prognosis is usually good; adult patients have normal stature and minimal hepatomegaly. There is no treatment for the fatal form of isolated cardiac phosphorylase kinase deficiency.

Type 0 Glycogen Storage Disease (Glycogen Synthase Deficiency) Strictly speaking, type 0 is not a glycogen storage disease, as the deficiency of the enzyme leads to decreased glycogen stores. The patients present in early infancy with morning drowsiness and fatigue and sometimes convulsions associated with hypoglycemia and hyperketonemia. There is no hepatomegaly or hyperlipidemia. Prolonged hyperglycemia and elevated lactate levels with normal insulin levels after administration of glucose suggest a possible diagnosis of glycogen synthase deficiency. Definitive diagnosis requires a liver biopsy to measure the enzyme activity. Treatment is symptomatic and involves frequent feedings rich in protein and nighttime supplements of uncooked cornstarch to alleviate hypoglycemia. Prognosis is good as patients survive to adulthood with a resolution of hypoglycemia except during pregnancy. Mutations in the liver

glycogen synthase gene (located on chromosome 12p12.2) that cause glycogen synthase deficiency have been identified.

Type XI Glycogen Storage Disease (Hepatic Glycogenosis with Renal Fanconi Syndrome, Fanconi-Bickel Syndrome) This rare autosomal recessive disease is caused by defects in the facilitative glucose transporter 2 (GLUT-2) which transports glucose in and out of hepatocytes, pancreatic cells, and the baso-lateral membranes of intestinal and renal epithelial cells. The disease is characterized by proximal renal tubular dysfunction, impaired glucose and galactose utilization, and accumulation of glycogen in liver and kidney.

GENETIC CONSIDERATIONS

The low prevalence of Fanconi-Beckel syndrome (fewer than 100 cases reported worldwide) is underscored by the fact that consanguinity is found in 70% of the patients with a detectable GLUT-2 mutation. The gene for GLUT-2 is located on chromosome 3q26 and most mutations detected so far cause premature termination of translation.

Clinical and laboratory findings The affected child presents in the first year of life with failure to thrive, rickets, and a protuberant abdomen due to liver and kidney enlargement. Laboratory findings include glucosuria, phosphaturia, generalized aminoaciduria, bicarbonate wasting, hypophosphatemia, increased serum alkaline phosphatase levels, and radiologic findings of rickets. Mild fasting hypoglycemia and hyperlipidemia may be present. Liver transaminases, plasma lactate, and uric acid levels are usually normal. Oral galactose or glucose tolerance tests show intolerance to these sugars, which may be caused by the functional loss of GLUT-2, which prevents liver uptake of these sugars. Tissue biopsies show marked accumulation of glycogen in hepatocytes and proximal renal tubular cells, presumably due to the altered transport of glucose out of these organs.

TREATMENT

There is no specific therapy. Growth retardation persists through adulthood. Symptomatic replacement of water, electrolytes, and vitamin D, restriction of galactose intake, and a diabetes mellitus-like diet, presented in frequent and small meals with a cornstarch supplement, may improve growth.

DISORDERS ASSOCIATED WITH LIVER CIRRHOSIS

Type IV Glycogen Storage Disease (Branching Enzyme Deficiency, Amylopectinosis, or Andersen Disease) Deficiency of branching enzyme activity results in accumulation of an abnormal glycogen with poor solubility. The disease is referred to as *type IV glycogen storage disease or amylopectinosis*, because the abnormal glycogen has fewer branch points, more 1-4 linked glucose units, and longer outer chains, resulting in a structure resembling amylopectin.

GENETIC CONSIDERATIONS

Type IV glycogen storage disease is a rare autosomal recessive disease. Prenatal

diagnosis is available with use of cultured amniocytes or chorionic villi to measure the level of enzymatic activity. The glycogen branching enzyme gene is located on chromosome 3p12. Both hepatic and neuromuscular forms of the disease are caused by mutations in the same branching enzyme gene; its characterization in individual patients may be useful in predicting the clinical course.

Clinical and laboratory findings This disorder is clinically variable. The most common form is characterized by progressive cirrhosis of the liver and is manifest in the first 18 months of life as hepatosplenomegaly and failure to thrive. The cirrhosis progresses to cause portal hypertension, ascites, esophageal varices, and liver failure that leads to death by age 5. Less frequently, patients survive without progression of liver disease.

Tissue deposition of amylopectin-like materials can be demonstrated in liver, heart, muscle, skin, intestine, brain, spinal cord, and peripheral nerve. The histologic findings in the liver are characterized by both micronodular cirrhosis and faintly stained basophilic inclusions in the hepatocytes. The inclusions consist of coarsely clumped, stored material that is periodic acid Schiff-positive and partially resistant to diastase digestion. Electron microscopy shows, in addition to the conventional c and bglycogen particles, an accumulation of fibrillar aggregations typical of amylopectin. Definitive diagnosis requires demonstration that branching enzyme activity is deficient in liver, muscle, cultured skin fibroblasts, or leukocytes.

A neuromuscular form of type IV glycogen storage disease has also been reported. Patients with this disease may present (1) at birth with severe hypotonia, muscle atrophy, and neuronal involvement and die during the neonatal period; (2) in late childhood with myopathy or cardiomyopathy; or (3) as adults with diffuse central and peripheral nervous system dysfunction accompanied by accumulation of polyglucosan bodies in the nervous system (so-called adult polyglucosan body disease). Definitive diagnosis of the adult disease requires assay of the branching enzyme in leukocytes or nerve biopsy, as the deficiency is limited to those tissues.

TREATMENT

There is no specific treatment for type IV glycogen storage disease. For progressive hepatic failure, liver transplantation has been performed. However, caution should be taken in selecting patients for liver transplantation because a nonprogressive hepatic form of the disease exists, and extra hepatic manifestations of the disease may occur after transplantation.

GLYCOGEN STORAGE DISEASES: MUSCLE GLYCOGENOSES

DISORDERS WITH MUSCLE-ENERGY IMPAIRMENT

Type V Glycogen Storage Disease (Muscle Phosphorylase Deficiency, McArdle Disease) Deficiency of muscle phosphorylase is the prototype muscle-energy disorder. Deficiency of this enzyme in muscle limits ATP generation by glycogenolysis and results in glycogen accumulation.

GENETIC CONSIDERATIONS

Type V glycogen storage disease is an autosomal recessive disorder that does not appear to have ethnic predilection. The gene for muscle phosphorylase is located on chromosome 11q13. The most common mutation in patients in the United States is a nonsense mutation that changes an arginine to a stop codon (R49X), and the most common mutation in the Japanese is deletion of a single codon (F708). These features allow DNA-based diagnosis and carrier detection in these two populations.

Clinical and laboratory findings Symptoms usually develop first in adulthood and are characterized by exercise intolerance with muscle cramps. Two types of activity tend to cause symptoms: (1) brief exercise of great intensity, such as sprinting or carrying heavy loads; and (2) less intense but sustained activity, such as climbing stairs or walking uphill. Moderate exercise, such as walking on level ground, can be performed by most patients for long periods. Many patients experience a characteristic "second wind" phenomenon; if they rest briefly at the first appearance of muscle pain, they can resume exercise with more ease. About half of patients report burgundy-colored urine after exercise, the consequence of myoglobinuria secondary to rhabdomyolysis. Intense myoglobinuria after vigorous exercise may cause renal failure. Although most patients are diagnosed in the second or third decade, many report weakness and lack of endurance since childhood. In rare cases, EMG findings may suggest an inflammatory myopathy, and the diagnosis can be confused with polymyositis.

The level of serum creatine kinase is usually elevated at rest and increases more after exercise. Exercise also increases the levels of blood ammonia, inosine, hypoxanthine, and uric acid. The latter abnormalities are attributed to accelerated recycling of muscle purine nucleotides in the face of insufficient ATP production.

Diagnosis Lack of an increase in blood lactate and exaggerated blood ammonia elevations after an ischemic exercise test are indicative of muscle glycogenosis and suggest a defect in the conversion of glycogen or glucose to lactate. The abnormal exercise response, however, is not limited to type V disease and can occur with other defects in glycogenolysis or glycolysis, such as deficiencies of muscle phosphofructokinase or debranching enzyme (when the test is done after fasting). Definitive diagnosis is made by enzymatic assay in muscle tissue or by mutation analysis of the myophosphorylase gene.

TREATMENT

In general, avoidance of strenuous exercise can prevent major episodes of rhabdomyolysis. Exercise tolerance can be augmented by aerobic training or by ingestion of glucose or fructose. A high-protein diet may increase exercise endurance in some patients. Longevity does not appear to be affected.

Type VII Glycogen Storage Disease (Muscle Phosphofructokinase Deficiency, Tarui Disease) Type VII disease is caused by a deficiency of muscle phosphofructokinase, which catalyzes the conversion of fructose-6-phosphate to fructose-1,6-diphosphate and is a key regulatory enzyme of glycolysis.

Phosphofructokinase is composed of three isozyme subunits (M, muscle; L, liver; and P,

platelet), which are encoded by different genes and are differentially expressed in tissues. Skeletal muscle contains only M isozyme, and red blood cells contain a hybrid of L and M forms. Type VII disease is due to defective M isoenzyme, which causes complete enzyme deficiency in muscle and partial deficiency in red blood cells.

GENETIC CONSIDERATIONS

Type VII glycogen storage disease is inherited as an autosomal recessive trait. The disease appears to be rare, and most reported patients are either Ashkenazi Jews or Japanese. The gene for the M isoenzyme is located on chromosome 12q13.3. In Ashkenazi Jews, 95% of mutant alleles are either a splicing defect or a nucleotide deletion.

Clinical and laboratory findings The features are similar to those in type V disease, namely, early onset of fatigue and pain with exercise. Vigorous exercise causes severe muscle cramps and myoglobinuria. However, several features of type VII disease are distinctive. (1) Exercise intolerance is usually evident in childhood, is more severe than in type V disease, and may be associated with nausea and vomiting. (2) A compensated hemolysis occurs as evidenced by an increased level of serum bilirubin and reticulocyte count. (3) Hyperuricemia is common and becomes more marked after exercise. (4) An abnormal glycogen resembling amylopectin is present in muscle fibers; it is periodic acid Schiff-positive and resistant to diastase digestion. (5) Exercise intolerance is particularly acute after meals rich in carbohydrate because glucose cannot be utilized in muscle and because the ingested glucose inhibits lipolysis and thus deprives muscle of fatty acid and ketone substrates. In contrast, patients with type V disease can metabolize glucose derived from either liver glycogenolysis or exogenous glucose. Indeed, glucose infusion improves exercise tolerance in patients with type V disease.

Two rare type VII variants have been reported. One begins in infancy with hypotonia and limb weakness, and a rapidly progressive myopathy leads to death by age 4. The other occurs in adults and is characterized by a slowly progressive, fixed muscle weakness rather than by cramps and myoglobinuria.

Diagnosis The M isoenzyme defect must be demonstrated in muscle, red blood cells, or cultured skin fibroblasts by biochemical or histochemical techniques.

TREATMENT

There is no specific treatment. Avoidance of strenuous exercise prevents acute attacks of muscle cramps and myoglobinuria.

Other Muscle Glycogenoses with Muscle-Energy Impairment Five additional enzyme defects produce muscle glycogenoses, namely, deficiencies in phosphoglycerate kinase, phosphoglycerate mutase, lactate dehydrogenase, fructose 1,6-bisphosphate aldolase A, and pyruvate kinase. All five enzymes affect terminal glycolysis, and deficiency causes muscle-energy impairment similar to that in type V and type VII disease. The failure of blood lactate to increase in response to exercise can be used to separate muscle glycogenoses from disorders of lipid metabolism, such as carnitine palmitoyl transferase II deficiency and very long chain acyl-coenzyme A

dehydrogenase deficiency, which also cause muscle cramps and myoglobinuria. Muscle glycogen levels may be normal in the disorders affecting terminal glycolysis, and definitive diagnosis is made by assaying the enzymatic activity in muscle.

DISORDERS WITH PROGRESSIVE SKELETAL MUSCLE MYOPATHY AND/OR CARDIOMYOPATHY

Type II Glycogen Storage Disease (Acid a-1,4 Glucosidase Deficiency, Pompe Disease) (See also[Chap. 349](#)) Type II disease is caused by a deficiency of lysosomal acid a-1,4 glucosidase (acid maltase), an enzyme responsible for the degradation of glycogen in lysosomal vacuoles. It is characterized by the accumulation of glycogen in lysosomes as opposed to its accumulation in cytoplasm in the other glycogenoses.

GENETIC CONSIDERATIONS

Pompe disease is an autosomal recessive disorder and does not appear to have an ethnic predilection. The gene for acid a-glucosidase is on chromosome 17q25. A splice site mutation (IVS1-13TG) is common in patients with adult-onset disease. Prenatal diagnosis with amniocytes or chorionic villi is available.

Clinical and laboratory findings The disorder encompasses a range of phenotypes, each including myopathy but differing in age of onset, organ involvement, and clinical severity. The most severe is the infantile-onset disease with cardiomegaly, hypotonia, and death before 1 year of age. Infants appear normal at birth but soon develop generalized muscle weakness with feeding difficulties, macroglossia, hepatomegaly, and congestive heart failure due to a hypertrophic cardiomyopathy. Electrocardiographic findings include a high-voltage QRS complex and a shortened PR interval. Death usually occurs from cardiorespiratory failure.

The juvenile or late-childhood form is characterized by skeletal muscle manifestations, usually without cardiac involvement, and a slowly progressive course. The juvenile form typically presents as delayed motor milestones (if age of onset is early enough) and difficulty in walking; these manifestations are followed by swallowing difficulties, proximal muscle weakness, and respiratory muscle involvement. This form can cause death before the end of the second decade.

An adult form of type II disease presents as a slowly progressive myopathy without cardiac involvement and has its onset between the second and seventh decades. The clinical picture is dominated by slowly progressive proximal muscle weakness with truncal involvement. The pelvic girdle, paraspinal muscles, and the diaphragm are most seriously affected. The initial symptoms may be respiratory insufficiency manifested by somnolence, morning headache, orthopnea, and exertional dyspnea.

Laboratory findings include elevated levels of serum creatine kinase, aspartate transaminase, and lactate dehydrogenase, particularly in infants. Muscle biopsy shows the presence of vacuoles that stain positively for glycogen, and muscle acid phosphatase is increased, presumably from a compensatory increase of lysosomal enzymes. Electron microscopy reveals the glycogen accumulation. EMG reveals myopathic features with irritability of muscle fibers and pseudomyotonic discharges.

Serum creatine kinase is not always elevated in adults and, depending on the muscle biopsied or tested, muscle histology or EMG may not be abnormal. It is prudent to examine affected muscle.

Diagnosis Diagnosis can be established by demonstration of the absence or reduced levels of acid a-glucosidase activity in muscle or cultured skin fibroblasts. Deficiency is usually more severe in the infantile form than in the juvenile and adult disorders.

TREATMENT

No effective treatment for the infantile form is currently available. Enzyme replacement is a promising therapy for this fatal lysosomal storage disease, and clinical trials are ongoing to test its safety and efficacy. A high-protein diet may be useful for the juvenile and adult forms. Nocturnal ventilatory support can improve the quality of life.

DISORDERS OF GALACTOSE METABOLISM

GALACTOSE 1-PHOSPHATE URIDYL TRANSFERASE DEFICIENCY GALACTOSEMIA

"Classic" galactosemia is a serious disease with an early onset of symptoms; the incidence is 1 in 60,000. The newborn infant normally receives up to 20% of caloric intake as lactose, which consists of glucose and galactose. Without the transferase, the infant is unable to metabolize galactose 1-phosphate (Fig. 350-1), the accumulation of which results in injury to parenchymal cells of the kidney, liver, and brain.

GENETIC CONSIDERATIONS

Galactosemia caused by transferase deficiency is inherited as an autosomal recessive trait; there are several enzymatic variants. The Duarte variant exhibits diminished red cell enzyme activity that is usually of no clinical significance. Some African-American patients have milder symptoms despite the absence of measurable transferase activity in erythrocytes; these patients retain 10% of the enzyme activity in liver and intestinal mucosa. Most Caucasian patients have no detectable enzyme activity in any of these tissues. The gene for galactose-1 phosphate uridyl transferase is located on chromosome 9p13. In African Americans, 48% of the mutant alleles are represented by the S135L substitution, perhaps accounting for the milder phenotype. In the Caucasian population, 70% of mutant alleles are represented by the Q188R substitution. Carrier testing and prenatal diagnosis can be carried out by direct enzyme analysis of amniocytes or chorionic villi. DNA-based testing can also be performed.

Clinical and laboratory findings The clinical manifestations of uridyl transferase deficiency are myriad, necessitating a high index of suspicion if the diagnosis is to be made. Clinical features may include: jaundice, hepatomegaly, vomiting, hypoglycemia, convulsions, lethargy, irritability, feeding difficulties, poor weight gain, aminoaciduria, cataracts, vitreous hemorrhage, hepatic cirrhosis, ascites, splenomegaly, or mental retardation. Patients with galactosemia are at increased risk for *Escherichia coli* neonatal sepsis; the onset of sepsis often precedes the diagnosis of galactosemia. When the diagnosis is not made at birth, damage to the liver (cirrhosis) and brain

(mental retardation) becomes increasingly severe and irreversible. For this reason, routine neonatal screening tests for galactosemia have been instituted in many parts of the world.

Diagnosis The preliminary diagnosis of galactosemia is made by demonstrating a reducing substance (by Clinitest) in urine specimens collected while the patient is receiving human or cow's milk or another formula containing lactose. The reducing substance found in urine, which is negative in a glucose oxidase test, can be identified as galactose with chromatography or an enzymatic test specific for galactose. Definitive diagnosis requires the demonstration of deficient activity of galactose-1-phosphate uridyl transferase in erythrocytes or other tissues, which also exhibit increased concentrations of galactose-1-phosphate.

TREATMENT

Because of widespread newborn screening for galactosemia, patients are now being identified and treated early. Elimination of galactose from the diet reverses growth failure, renal, and hepatic dysfunction. Cataracts regress and most patients have no impairment of eyesight. Early diagnosis and treatment have improved the prognosis of galactosemia; on long-term follow-up, however, patients still have ovarian failure manifest as primary or secondary amenorrhea, as well as developmental delay and learning disabilities, which increase in severity with age. In addition, most patients have speech disorders and a smaller number demonstrate poor growth and impaired motor function and balance (with or without overt ataxia). The relative control of galactose-1-phosphate levels does not always correlate with long-term outcome, suggesting that other factors, such as uridine diphosphate (UDP)-galactose deficiency (a donor for galacto-lipids and proteins), may be responsible for some of the metabolic consequences of the disease.

GALACTOKINASE DEFICIENCY

In contrast to the multiple systems that are affected in uridyl transferase deficiency, cataracts are usually the sole manifestation of galactokinase deficiency. The affected infant is otherwise asymptomatic. This disorder is characterized by elevated blood galactose levels with normal uridyl transferase activity and an absence of galactokinase activity in erythrocytes. Treatment is dietary restriction of galactose intake. Mutations leading to galactokinase deficiency have been identified in the gene coding for galactokinase (located on chromosome 17q24).

URIDINE DIPHOSPHATE GALACTOSE 4-EPIMERASE (UDP GAL 4-EPIMERASE) DEFICIENCY

The abnormally accumulated metabolites in this disorder are very much like those seen in uridyl transferase deficiency; however, there is also an increase in cellular UDP galactose. There are two distinct forms of epimerase deficiency. A benign form was discovered incidentally as a result of the neonatal screening program. Affected persons in this case are healthy; the enzyme deficiency is limited to leukocytes and erythrocytes, without deranged metabolism in other tissues, and no treatment is required. The second form of epimerase deficiency is severe with clinical manifestations resembling uridyl

transferase deficiency. Additional features include hypotonia and nerve deafness. The enzyme deficiency is generalized, and clinical symptoms respond to restriction of dietary galactose. Although this form of galactosemia is very rare, it must be considered in a symptomatic patient who has normal transferase activity. The gene for epimerase is located on chromosome 1p35-36; mutations responsible for both forms of the epimerase deficiency have been identified.

DISORDERS OF FRUCTOSE METABOLISM

DEFICIENCY OF FRUCTOKINASE (BENIGN FRUCTOSURIA)

This condition is not associated with any clinical manifestations. It is an incidental finding, usually made through the detection of fructose as a reducing substance in the urine. No treatment is necessary.

DEFICIENCY OF FRUCTOSE 1,6-BISPHOSPHATE ALDOLASE (ALDOLASE B) (HEREDITARY FRUCTOSE INTOLERANCE)

This is a severe disease of infants that appears with the ingestion of fructose-containing food. It is caused by deficiency of fructose 1,6-bisphosphate aldolase B activity in the liver, kidney, and intestine. The enzyme catalyzes the hydrolysis of fructose 1-phosphate and fructose 1,6-bisphosphate into the 3-carbon sugars, dihydroxyacetone phosphate, glyceraldehyde-3-phosphate, and glyceraldehyde (Fig. 350-1). Deficiency of this enzyme activity causes rapid accumulation of fructose 1-phosphate and initiates severe toxic symptoms when exposed to fructose.

GENETIC CONSIDERATIONS

The true incidence of hereditary fructose intolerance is not known but may be as high as 1 in 23,000. The gene for aldolase B is on chromosome 9q22.3. Several mutations causing heredity fructose intolerance have been identified. A single missense mutation, which results in substitution of proline for alanine at position 149, is the most common mutation identified in northern Europeans. This mutation plus two other point mutations account for approximately 80 to 85% of hereditary fructose intolerance in Europe and the United States. The diagnosis of hereditary fructose intolerance can thus be made in most cases by direct DNA analysis. Prenatal diagnosis should be possible from either amniocentesis or chorionic villi with the use of DNA for mutational or linkage analysis.

Clinical and laboratory findings Patients with fructose intolerance are healthy and asymptomatic until fructose or sucrose (table sugar) is ingested (usually from fruit, fruit juice, or sweetened cereal). Clinical manifestations may resemble those of galactosemia and include jaundice, hepatomegaly, vomiting, lethargy, irritability, and convulsions. Laboratory findings include prolonged clotting time, hypoalbuminemia, elevation of bilirubin and transaminases, and proximal renal tubular dysfunction. If the disease is not diagnosed and intake of the noxious sugar persists, hypoglycemic episodes recur, and liver and kidney failure progress, eventually leading to death.

Diagnosis Suspicion of the enzyme deficiency is suggested by the presence of a reducing substance in the urine during an attack. The diagnosis is supported by an

intravenous fructose tolerance test, which will cause a rapid decline of serum phosphate, followed by blood glucose, and a subsequent rise of uric acid and magnesium. An oral tolerance test should not be performed as patients may become acutely ill. Definitive diagnosis is made by assay of fructaldoase B activity in the liver.

TREATMENT

Treatment consists of the complete elimination of all sources of sucrose, fructose and sorbitol from the diet. With this treatment, liver and kidney dysfunction improve, and catch-up growth is common. Intellectual development is usually unimpaired. As the patient matures, symptoms become milder, even after fructose ingestion, and the long-term prognosis is good. Owing to dietary avoidance of sucrose, affected patients have few dental caries.

FRUCTOSE 1,6-DIPHOSPHATASE DEFICIENCY

Fructose 1,6-diphosphatase deficiency is a defect in gluconeogenesis. The disease is characterized by life-threatening episodes of acidosis, hypoglycemia, hyperventilation, convulsions, and coma. These episodes are triggered by febrile infections and gastroenteritis when oral food intake decreases. Laboratory findings include low blood glucose, high lactate and uric acid levels, and a blood gas picture of metabolic acidosis. In contrast to hereditary fructose intolerance, there is usually no aversion to sweets, and renal tubular and liver functions are normal. Treatment of acute attacks consists of correction of hypoglycemia and acidosis by intravenous infusion; the response is usually rapid. Later, avoidance of fasting and elimination of fructose and sucrose from the diet prevent further episodes. For long-term prevention of hypoglycemia, a slowly released carbohydrate such as cornstarch is useful. Patients who survive childhood seem to develop normally.

Diagnosis The diagnosis is established by demonstrating enzyme deficiency in either the liver or an intestinal biopsy specimen. The enzyme defect may also be demonstrated in leukocytes in some cases. The gene coding for fructose 1,6-diphosphatase is located on chromosome 9q22. In patients with known mutations, carrier detection and prenatal diagnosis are possible by DNA-based testing.

(Bibliography omitted in Palm version)

351. INHERITED DISORDERS OF CONNECTIVE TISSUE - *Darwin J. Prockop, Helena Kuivaniemi, Gerard Tromp, Leena Ala-Kokko*

Heritable disorders that involve the major connective tissues of the body such as bone, skin, cartilage, blood vessels, and basement membranes are among the most common genetic diseases in human beings. Here we will focus primarily on those disorders that can have severe manifestations, are relatively common, and are sufficiently understood at the molecular level to provide useful paradigms: osteogenesis imperfecta (OI), the Ehlers-Danlos syndrome (EDS), chondrodysplasias (CDs), the Marfan syndrome (MFS), epidermolysis bullosa (EB), and the Alport syndrome (AS).

THE CHALLENGE OF CLASSIFYING THE DISEASES

The original classification of connective tissue diseases was based on the pattern of inheritance, the cluster of signs and symptoms, the histologic changes in tissues, and limited information about the molecular defects involved. This classification included about a dozen types and subtypes for OI, about the same number for the EDS, and over 150 for the CDs. Several limitations in these original classifications are now apparent. One is that the same mutation does not always produce the same disease phenotype in terms of severity of the condition or its clinical course. Such phenotypic variation occurs in many genetic diseases, including the connective tissue disorders, in which some members of a family are severely affected, whereas others with the same mutation have a mild disorder.

Most patients with classic features of a severe connective tissue disease have a mutation in a gene or genes coding for a single protein. For example, the majority of patients with OI have a mutation in one of the two genes coding for type I procollagen. Similarly, most patients with MFS have mutations in a gene for fibrillin. For other disease categories, the situation is more complex. In EDS, for example, the type IV variant is usually caused by mutations in the gene for type III procollagen, the type VI variant by defects in the gene for lysyl hydroxylase, and the type VII variant by defects that impair the processing of type I procollagen to type I collagen.

Classifications of these disorders also tend to overemphasize the etiologic differences between severe genetic diseases that are apparent in infants and the more common diseases that appear much later in life. Single-gene defects can cause subsets of late-onset diseases such as osteoporosis, aneurysms, and osteoarthritis. For example, a small subset of patients with postmenopausal osteoporosis have mutations in the genes for procollagen I similar to the mutations in the same genes that produce lethal variants of OI. Likewise, a subset of patients with familial aortic aneurysms have mutations in the gene for procollagen III similar to the mutations in the same gene that cause lethal variants of type IV EDS, and occasional patients with osteoarthritis have mutations in the gene for procollagen II similar to the mutations in the same gene that cause lethal CDs. There is disagreement as to the best diagnosis for such patients in that some investigators feel that after a mutation similar to those seen in the early-onset diseases is identified, the patients should be reclassified as having mild forms of OI, EDS, or CD, even though they do not have definitive evidence of the early-onset diseases or seek medical attention until adulthood. Therefore, the category of diseases referred to as *inherited disorders of connective tissue* may have to be expanded as we

obtain additional information about more common diseases.

DEFINITION AND COMPOSITION OF CONNECTIVE TISSUES

Connective tissues are composed of specific macromolecules, many of which are also constituents of the lung, the kidney, the walls of blood vessels, the vitreous gel of the eye, and the synovial fluid. Indeed, most organs and tissues contain small amounts of the same macromolecules assembled into membranes and septa. Therefore, virtually all structures contain some connective tissue.

The distinguishing feature of connective tissues is that the component macromolecules are assembled into an insoluble extracellular matrix (Table 351-1). The macromolecules include at least 19 different types of collagens, the related fibrous proteins known as *elastin* and *fibrillin*, a series of proteoglycans, and components whose structure and function are only partially defined.

Differences in the connective tissues of bone, skin, and cartilage are in part explained by differences in the content of specific components (Table 351-1). For example, tendons and ligaments consist primarily of type I collagen fibrils and small amounts of other components that help organize the type I fibrils into fibers and fiber bundles. Cartilage consists primarily of fibrils of type II collagen in the form of arcades that are distended by highly charged proteoglycans. The extracellular matrix of the aorta contains collagens that provide tensile strength and elastin that provides elasticity. Differences among the connective tissues also depend on the three-dimensional organization of the molecular components. The type I collagen fibrils in tendon are packed into thick, parallel bundles of fibers, whereas type I collagen fibrils in skin are randomly oriented. In cortical bone, helical arrays of type I collagen fibrils are deposited around haversian canals.

BIOSYNTHESIS OF CONNECTIVE TISSUE

Connective tissues form primarily by a process of self-assembly, in which a molecule of the correct size, shape, and surface properties binds to other molecules with the same or similar structure in a spontaneous but ordered manner. The molecular mechanisms and driving forces are similar to those involved in crystal formation.

Collagen Synthesis The self-assembly of connective tissue is illustrated by the assembly of collagen into fibrils. The collagen molecule that forms fibrils is a long, thin rod consisting of three polypeptide a chains wrapped into a rigid, ropelike triple helix (Fig. 351-1). The molecule has a triple-helical conformation, because each of the three a chains has a simple, repetitive amino acid sequence of about 1000 amino acids in which glycine (Gly) appears as every third amino acid. Therefore, the sequence of eacha chain can be designated as $(-Gly-X-Y-)_{333}$, where X and Y represent amino acids other than glycine. To fold into a triple helix, every third amino acid in an a chain must be glycine, the smallest amino acid, since this residue must fit in a sterically restricted space where the three chains of the triple helix come together. Many of the X- and Y-position amino acids are proline and hydroxyproline, which provide rigidity to the triple helix. The remaining amino acids form clusters of hydrophobic and charged regions on the surface of the molecule that direct how one molecule spontaneously binds to other

collagen molecules and thereby self-assembles into the large collagen fibrils in tissues (see Fig. 351-1).

More than 19 different collagens have been identified. Most are minor constituents that probably have highly specialized functions. The fibrillar collagens are found in tissues as long, highly ordered fibrils with a characteristic banding pattern by electron microscopy. Type I collagen, the most abundant, is found as cross-striated fibrils in a large number of tissues (Table 351-1). It is composed of two identicala chains called a1(I) and a third called a2(I). Type II collagen, another fibrillar collagen of cartilage, is composed of three identicala chains called a1(II). Type III collagen is found in small amounts in many tissues that contain type I collagen and in large amounts in large blood vessels; it is composed of three identical chains called a1(III). The nonfibrillar collagens are similar to the fibrillar collagens in that they contain -Gly-X-Y- sequences of amino acids that form triple-helical domains, but they also contain large globular domains. Self-assembly of most of the nonfibrillar collagens usually involves binding between the globular domains to form networks. For example, the type IV collagen in basement membranes self-assembles into a complex three-dimensional network that provides a diffusion barrier in the renal glomerulus and pulmonary alveolus. The network also provides support for epithelial and endothelial cells in these tissues and in skin, the gastrointestinal tract, and blood vessels. Some nonfibrillar collagens bind to the surface of fibrils formed by the more abundant collagens and alter the lateral growth of the fibrillar collagens or prevent the fibrils from coalescing into fiber bundles.

Because the molecules or monomers fibrillar collagens spontaneously self-assemble into fibrils, they are first synthesized as larger and more soluble precursors called *procollagens* and composed of proa chains. As the proa chains of procollagen are synthesized on ribosomes, the free ends move into the cisternae of the rough endoplasmic reticulum (Fig. 351-1). Hydrophobic signal peptides at the N termini are cleaved, and additional posttranslational reactions begin. Proline residues in the Y position of the repeating -Gly-X-Y- sequences are converted to hydroxyproline by prolyl hydroxylase in a reaction requiring ascorbic acid. Lysine residues in the Y position are similarly hydroxylated to hydroxylysine by lysyl hydroxylase. Many of the hydroxylysine residues are glycosylated with galactose or with galactose and glucose. A large mannose-rich oligosaccharide is assembled on the C-terminal propeptide of each chain. The association of the proa chains is directed by the structure and the surface properties of the globular C-propeptides. After the C-propeptides assemble correctly, the structure is locked in place by the formation of interchain disulfide bonds. Posttranslational modifications of the proa chains continue until each chain acquires about 100 hydroxyproline residues. Then a few of the hydroxyproline-rich -Gly-X-Y- sequences at the C terminus of the protein fold into a triple helix. The short region of triple helix becomes a nucleus for self-assembly of the triple helix of the whole protein, much like a nucleus for crystallization, in that the triple-helical conformation in one -Gly-X-Y- sequence induces the next -Gly-X-Y- sequence to fold into the same conformation. As a result, the conformation is propagated in a zipper-like fashion from the C terminus to the N terminus of the molecule, and the entire a-chain domain becomes a continuous triple helix. The protein then passes from the rough endoplasmic reticulum to other compartments and is secreted. The requirement for ascorbic acid in the hydroxylation of prolyl residues explains why wounds fail to heal in scurvy (Chap. 75). If sufficient proline residues are not converted to hydroxyproline, collagen cannot

fold into a triple helix that is stable at body temperature. The abnormal protein accumulates in the cisternae of the rough endoplasmic reticulum and is slowly degraded.

After secretion, procollagen is processed to collagen by cleavage of the N-propeptides by procollagen N-proteinase and of the C-propeptides by procollagen C-proteinase. The processing of type I procollagen by the two proteinases converts the precursor to type I collagen and thereby decreases the solubility of the protein about 1000-fold. The 1000-fold decrease in solubility provides the entropic energy that drives the spontaneous self-assembly of the collagen into fibrils. Collagen monomers first assemble into a nucleus that grows by addition of monomers as determined by the structure of the nucleus. The nucleus and the final fibril can be assembled spontaneously from a single kind of collagen such as type I or type II. Fibrils can also be assembled as copolymers in which two or more collagens are incorporated into the same fibril simultaneously. Alternatively, a second collagen or a proteoglycan can bind to the surface of a growing fibril or to a fully formed fibril and thereby influence the final structure and the functional properties of the fibrils. The final structure of the fibrils in tissues is also influenced by the pressure and tensions on the fibrils, particularly after their tips are inserted into muscle and bone. The tension on tendons, for example, probably makes the initial thin fibrils coalesce into large fiber bundles.

Self-assembled collagen fibers have considerable tensile strength, which is increased by cross-linking reactions that form covalent bonds between a chains in one molecule and a chains in adjacent molecules. The first step in cross-linking is oxidation by lysyl oxidase of amino groups on a few lysine or hydroxylysine residues to form aldehydes that interact to form stable covalent bonds.

During growth and development, the collagen fibrils in all tissues undergo repeated synthesis, degradation, and resynthesis. The degradation of collagen fibers in tissues is initiated by specific collagenases found in leukocytes, fibroblasts, synovial cells, or related cell types. The collagenases cleave the collagen molecule at a point about three-quarters of the distance from its N terminus. The cleavage apparently triggers unfolding of the molecules on the surface of a fibril and further degradation by other proteinases.

Collagen fibers in most tissues of normal adults undergo very little metabolic turnover. One exception to this is the collagen fibrils that are degraded and resynthesized as part of the continual remodeling of bone. Although the collagen in many adult tissues is metabolically stable, the rate of turnover changes under some circumstances. In starvation, a large fraction of the collagen in skin and other connective tissues is degraded, thus providing amino acids for gluconeogenesis. Large losses of collagen also occur in most connective tissues during immobilization or prolonged periods of low-gravitational stress. In rheumatoid arthritis, pannus invasion causes a rapid degradation of collagen in the articular cartilage. Glucocorticoids decrease the collagen content of most connective tissues, including bone, by decreasing the rate of collagen synthesis. Decreases in collagen weaken tissues. In many pathologic states, however, collagen is deposited in excess. With injury to tissue, inflammation is usually followed by increased deposition primarily of type I collagen fibrils in the form of fibrotic tissue and scars. The deposition of collagen fibrils during the repair process is largely irreversible

and is a major feature of the pathologic changes in hepatic cirrhosis, pulmonary fibrosis, atherosclerosis, and nephrosclerosis and in the scarring in skin and ligaments after surgery or trauma.

Elastin Synthesis Elastin assembly appears to be closely related to that of collagen, since a few of the prolines in the protein are hydroxylated to hydroxyproline by prolyl hydroxylase. The elastin monomer, however, is a single polypeptide that does not fold into a defined three-dimensional structure and is not synthesized as a larger precursor molecule. Instead, it is slowly secreted from cells into extracellular compartments, where it forms amorphous deposits around previously deposited microfibrils. The elastin deposits then become covalently cross-linked through oxidation of lysine residues to aldehydes by the same lysyl oxidase that initiates the cross-linking of collagen. The microfibrils in elastin deposits are largely composed of fibrillin, a large protein that forms beadlike strands.

Proteoglycan Synthesis The synthesis of proteoglycans begins in the cisternae of the rough endoplasmic reticulum with assembly of a core protein that then undergoes modification by sugar and sulfate transferases that generate large side chains of glycosaminoglycans. At least 30 proteoglycans have been identified by differences in the structures of their core proteins. The major proteoglycan of cartilage, called *aggrecan*, has a core protein of about 2000 amino acids to which are bound multiple side chains of chondroitin sulfate and keratin sulfate, polysaccharides consisting of highly charged and repetitive disaccharide sequences. After secretion from cells, the aggrecan monomer binds to a smaller protein called a *link protein*. The complex of core protein and link protein then spontaneously binds to a long chain of hyaluronic acid to form a huge copolymer called a *proteoglycan aggregate*. The highly charged proteoglycan aggregate binds water and small ions and thereby provides a large swelling pressure and resiliency to cartilage. Smaller proteoglycans such as decorin, biglycan, and fibromodulin have smaller core proteins with fewer and different polysaccharide side chains. They do not form large aggregates with hyaluronate but bind to fibrils of collagen or fibronectin and may thereby help regulate the assembly or spatial orientation of fibrils. One group of small proteoglycans known as *syndecans* binds to the plasma membranes of cells and may have a role in cell migration along fibrils or in signal transduction.

The assembly of bone follows much the same principles as the assembly of other connective tissues (Chap. 340). The first step is deposition of osteoid tissue that consists largely of type I collagen fibrils. Mineralization of osteoid occurs by steps that are still incompletely defined; proteins such as osteopontin and osteocalcin probably bind to the collagen fibrils and chelate calcium to initiate mineralization. Small proteoglycans such as decorin or fibromodulin may also play a role.

MUTATIONS THAT PRODUCE DISEASES OF CONNECTIVE TISSUES

Because of the large number of tissue-specific macromolecules present in connective tissues, a large number of gene-protein systems are candidates for mutations that might cause disease of the tissues.

The most complete data on mutations causing heritable disorders of connective tissue

are available on OI. Most patients with severe OI (types II and III) have mutations in either the gene for the proa1(I) chain or the gene for the proa2(I) chain of type I procollagen (the COL1A1 and COL1A2 genes). In patients with mild disease, many of the mutations decrease expression of protein from one allele of the genes. Most of the mutations in patients with severe OI cause synthesis of a structurally abnormal but partially functional proa chain. Mutations that cause synthesis of structurally abnormal proa chains include partial gene deletions, partial gene duplications, and RNA splicing mutations. The most common mutations, however, cause the substitution of single amino acids with bulky side chains for the glycine residues that appear as every third amino acid in the triple helix of a proa chain (Fig. 351-2). The structurally abnormal proa chains exert their effects primarily through one of three mechanisms (Fig. 351-1). First, the presence of an abnormal proa chain in a procollagen molecule containing two normal proa chains can prevent folding of the protein into a triple-helical conformation and lead to degradation of the whole molecule in a process called *procollagen suicide*. Similar dominant negative mutations are seen with other multisubunit proteins. The net result of procollagen suicide is accumulation of the unfolded protein in the rough endoplasmic reticulum of cells and a reduction in the amount of collagen available for fibril assembly. Second, the presence of one abnormal proa chain in a procollagen molecule can interfere with cleavage of the N-propeptide from the protein. The persistence of the N-propeptide on a fraction of the molecules interferes with the self-assembly of normal collagen so that thin and irregular collagen fibrils are formed. Third, the substitution of a bulkier amino acid for glycine can produce a change in the conformation of the molecule and result in the assembly of collagen fibrils that are abnormally branched or abnormally thick and short. Also, copolymerization of the mutated collagen with normal collagen can slow fibril assembly and decrease the total amount of collagen incorporated into fibrils.

Over 300 mutations in the two genes for type I procollagen have been found in patients with OI (Fig. 351-2). Initially, there was concern that many of the mutations might be neutral variations in the structure of the genes and not the cause of the disease phenotypes. However, the causal relationship between most of the mutations and the disease has been established by several kinds of evidence: (1) DNA linkage studies in families with mild variants of OI showed that specific mutated alleles were coinherited with the disease phenotypes. (2) Probands with lethal variants of OI were shown to have new mutations not found in the normal parents or in only a few cells from a mosaic parent (see below). (3) Studies with cultured skin fibroblasts from patients demonstrated that the mutations either produced specific disruptions in the biosynthesis of type I procollagen or caused synthesis of a type I procollagen that formed abnormal collagen fibrils (Fig. 351-1). (4) The mutations in probands with OI were not found in normal alleles for the genes. (5) Expression of several of the mutated genes for type I procollagen in transgenic mice generated disease phenotypes similar to those seen in patients who inherited the mutated genes (Fig. 351-3).

The data on mutations in type I procollagen that cause OI have been used as a paradigm for defining other mutations in other collagen and procollagen genes that cause other disorders of connective tissue. For example, similar mutations in the gene for type III procollagen occur in patients with the type IV variant of EDS, which causes early death because of rupture of the aorta or other hollow organs (Fig. 351-4). Also, similar mutations in the gene for type II procollagen (COL2A1) are found in a number of

patients with CDs (Fig. 351-5). In addition, transgenic mice expressing mutated genes for type II procollagen develop phenotypes resembling the CDs. Similar mutations in the gene for type VII collagen (COL7A1) are found in patients with the dystrophic form of EB, and mutations in the genes for type IV collagen are found in many patients with AS. As discussed below, the paradigm for defining the consequences of mutations in procollagen genes also helps explain findings on mutations in a fibrillin gene that cause MFS and mutations in keratin genes that cause the simplex variant of EB.

Several generalizations can be made about mutations in collagen genes. One is that unrelated patients rarely have the same mutation in the same gene. Another is that mutations that cause the most severe disease are usually new mutations in one allele that occur either during the generation of the germline in one of the parents or during meiosis in the fertilized egg (Chap. 65). Still another generalization is that most mild variants are caused by mutations that are specific or "private" to a given family. Indeed, the number of recurrent mutations in structural genes are so infrequent that there are, in effect, no common mutations responsible for the disorders in unrelated patients and no "hot spots" that contain most of the mutations.

Another general trend is that similar mutations in the same gene can produce different disease syndromes in terms of both severity and the major tissues involved. One reason for heterogeneity in pathologic manifestations is that different regions of a large molecule may be more important for its function in some connective tissues than in others. For example, some regions of the type I collagen molecule may be essential for the binding of mineralizing proteins in bone so that mutations in these regions cause fragile bones but do not impair function in skin and other nonmineralizing tissues. It is more difficult, however, to explain how the same mutation can produce a severe phenotype in some and a mild phenotype in other members of the same family. Such phenotypic variation appears to be characteristic of OI, where some subjects are short and have multiple fractures from minor trauma, whereas others in the same family can be of normal stature and free of fractures. In the past, such phenotypic variation was explained by undefined variations in the genetic background of different family members. Studies in transgenic mice, however, demonstrated similar phenotypic variation with expression of a mutated collagen gene in an inbred strain of mice in whom the genetic background is uniform. Therefore, the phenotypic variation is probably caused by undefined stochastic or chance events during embryonic and fetal development. Although dramatic phenotypic variations are relatively rare in OI and related disorders, it is important to consider in counseling families about the consequences of inherited mutations.

OSTEOGENESIS IMPERFECTA

OI is an inherited disorder that causes a generalized decrease in bone mass (osteopenia) and makes the bones brittle. The disorder is frequently associated with blue sclerae, dental abnormalities (dentinogenesis imperfecta), progressive hearing loss, and a positive family history. The most severe forms cause death in utero, at birth, or shortly thereafter. The course of mild and moderate forms is more variable. Some patients appear normal at birth and become progressively worse. Some have multiple fractures in infancy and childhood, improve after puberty, and fracture more frequently later in life. Women are particularly prone to fracture during pregnancy and after

menopause. A few women from families with mild variants of OI do not develop fractures until after menopause, and their disease may be difficult to distinguish from postmenopausal osteoporosis.

Classification The most common classification for OI was developed by Sillence (Table 351-2). Type I, the mildest form, is inherited as an autosomal dominant trait. Most patients have distinctly blue sclerae. Type I is subdivided into types IA and IB depending on whether or not dentinogenesis imperfecta is present. Type II is lethal in utero or shortly after birth. Radiographic criteria can be used to subdivide type II into five groups, with subgroup 1 showing the most severe changes and subgroup 5 the least. Types III and IV OI are intermediate in severity between types I and II. They differ from type I because of lesser severity and because the sclerae are only slightly bluish in infancy and white in adulthood. Type III differs from type IV in that it tends to become more severe with age. Also, type III can be inherited either as an autosomal recessive or autosomal dominant trait, whereas type IV is always dominant. The clinical courses are variable, and the mode of inheritance in types III and IV may be difficult to ascertain because many patients have sporadic mutations and because many couples with one child severely affected by OI do not have additional children. For these and related reasons, the distinction between type IV OI and other severe variants of OI may not be helpful. Therefore, it may be sufficient to classify patients simply as mild (type I), lethal (type II), and moderately severe (type III).

Incidence Type I OI has a frequency of about 1 in 30,000. Type II OI has a reported incidence at birth of about 1 in 60,000, but the incidence of the three severe forms recognizable at birth (types II, III, and IV) may be as high as 1 in 20,000.

Skeletal Changes In type I OI, the fragility of bones may be severe enough to limit physical activity or so mild that individuals are unaware of any disability. Radiographs of the skull of patients with mild disease may show a mottled appearance because of small islands of irregular ossification. In type II OI, bones and other connective tissues are so fragile that massive injuries can occur in utero or during delivery (Fig. 351-3). Ossification of many bones is frequently incomplete. Continuously beaded or broken ribs and crumpled long bones (accordina femora) may be present. For unclear reasons, the long bones may be either thick or thin. In types III and IV, multiple fractures from minor physical stress can produce severe deformities. Kyphoscoliosis can impair respiration, cause cor pulmonale, and predispose to pulmonary infections. The appearance on radiographs of "popcorn-like" deposits of mineral on the ends of long bones is an ominous sign. Progressive neurologic symptoms may result from basilar compression and communicating hydrocephalus.

In all forms of OI, bone mineral density in unfractured bone is decreased. However, the degree of osteopenia may be difficult to evaluate because recurrent fractures limit exercise and thereby worsen the decrease in bone mass. Surprisingly, fractures appear to heal normally.

Ocular Changes The sclerae can be normal, slightly bluish, or bright blue. The color is probably caused by a thinness of the collagen layers of the sclerae that allows the choroid layers to be seen. Blue sclerae, however, are an inherited trait in some families who do not have increased bone fragility.

Dentinogenesis Imperfecta The teeth may be normal, moderately discolored, or grossly abnormal. The enamel generally appears normal, but the teeth may have a characteristic amber, yellowish brown, or translucent bluish gray color because of improper deposition or deficiency of dentin. The deciduous teeth are usually smaller than normal, whereas permanent teeth are frequently bell-shaped and restricted at the base. In some patients, the teeth readily fracture and need to be extracted. The defect in dentin is directly attributable to the fact that normal dentin is rich in type I collagen. Similar tooth defects, however, can be inherited without any evidence of OI.

Hearing Loss Hearing loss usually begins during the second decade of life and occurs in over 50% of subjects over age 30. The loss can be conductive, sensorineural, or mixed and varies in severity. The middle ear usually exhibits maldevelopment, deficient ossification, persistence of cartilage in areas that are normally ossified, and abnormal calcium deposits.

Associated Features Other connective tissue involvement can include thin skin that scars extensively, joint laxity with permanent dislocations indistinguishable from those of EDS, and, occasionally, cardiovascular manifestations such as aortic regurgitation, floppy mitral valves, mitral incompetence, and fragility of large blood vessels. For unknown reasons, some patients develop a hypermetabolic state with elevated serum thyroxine levels, hyperthermia, and excessive sweating.

Molecular Defects Most patients with OI have mutations in one of the two genes that encode type I procollagen. Over 90% of patients with type I OI and blue sclerae have mutations in the proa1(I) gene that decrease the steady-state levels of the mRNA for proa1(I) chains and decrease the rates of synthesis of proa1(I) chains relative to those for proa2(I) chains. In more severe forms (types II, III, and IV), the effects of mutations that cause synthesis of abnormal proa chains are amplified by the three mechanisms discussed above (Fig. 351-1). Mutations that change the structure of the protein near the N-proteinase cleavage site cause accumulation of a partially processed procollagen and produce lax joints similar to those in type VII EDS that is caused by mutations in the gene for the N-proteinase. Mutations that change the structure in the middle or near the C terminus of the molecule tend to cause severe or lethal variants of OI. It is difficult, however, to correlate the site or nature of the mutation and the clinical phenotype (Fig. 351-2). Most patients are heterozygotes with mutations in a single allele, but rare patients are homozygotes with two mutated alleles for proa1(I) or proa2(I) chains.

Mosaicism in Germ-Line Cells and in Somatic Cells Most lethal OI is the result of new autosomal dominant mutations. The frequency of a second child with lethal OI in the same family, however, is about 7% because of germ-line mosaicism in one of the parents. The presence of germ-line mosaicism has been demonstrated in several fathers of patients with type II OI by demonstrating the mutated gene in a fraction of their sperm. Apparently normal parents of children with severe OI may also have somatic cell mosaicism in which the mutated allele is present in a fraction of somatic cells such as fibroblasts, leukocytes, and hair root cells. Because of the possibility of germ-line mosaicism, asymptomatic parents of a child with severe OI should be counseled that recurrence can occur.

Diagnosis The diagnosis is usually made on the basis of clinical criteria. The presence of fractures together with blue sclerae, dentinogenesis imperfecta, or family history of the disease is usually sufficient to make the diagnosis. Other causes of pathologic fractures must be excluded, including the battered child syndrome, nutritional deficiencies, malignancies, and other inherited disorders such as CDs and hypophosphatasia (Table 351-3). X-rays usually reveal a decrease in bone density that can be verified by photon or x-ray absorptiometry. There is no consensus, however, as to whether the diagnosis can be made by microscopy of bone. A molecular defect in type I procollagen can be demonstrated in half or more of patients by incubating skin fibroblasts with radioactive amino acids and then analyzing the proa chains by polyacrylamide gel electrophoresis. The analysis detects decreases in the rate of synthesis of proa1(I) chains relative to proa2(I) chains, abnormally long proa chains, abnormally short proa chains, and proa chains with abnormal posttranslational modification because of an amino acid substitution that impairs folding of the triple helix. The mutations themselves can be defined in most patients by sequencing of genomic DNA. Because each proband and family usually has a "private" mutation, extensive analysis of about 10,000 bases in each of the two genes is required to identify the exact mutation. After a mutation in a type I procollagen gene is identified, a test based on the polymerase chain reaction can be used to screen family members at risk and for prenatal diagnosis.

TREATMENT

Many patients with OI have successful careers despite severe deformities. Those with mild disorder may need little treatment when fractures decrease after puberty, but women require special attention during pregnancy and after menopause, when fractures again increase. More severely affected children require a comprehensive program of physical therapy, surgical management of fractures and skeletal deformities, and vocational education.

Many of the fractures are only slightly displaced and have little soft tissue swelling. Therefore, they can be treated with minimal support or traction for a week or two followed by a light cast. If fractures are relatively painless, physical therapy can be initiated early. A judicious amount of exercise prevents loss of bone mass secondary to physical inactivity. Some physicians advocate insertion of steel rods into long bones to correct limb deformities; the risk/benefits and cost/benefits of such procedures are difficult to evaluate. Aggressive conventional intervention is usually warranted for pneumonia and cor pulmonale. For severe hearing loss, stapedectomy or replacement of the stapes with a prosthesis may be successful. Moderately to severely affected patients should be evaluated periodically to anticipate possible neurologic problems. About half of children have a substantial increase in growth when given growth hormone. Treatment with bisphosphonates to decrease bone loss has been introduced on an experimental basis. Initial results are promising, but the long-term effects of decreasing bone resorption are unknown. Also, a clinical trial has been initiated to use stromal cells from bone marrow that can differentiate into osteoblasts after systemic infusion. In the first phase of the trial, three children with severe OI (type III) showed clinical improvement after marrow ablation and transplantation of whole bone marrow from an HLA-compatible sibling.

A program for careful orthotic management developed by Bleck and a program for compressive management developed by Marini are useful. Counseling and emotional support are important for patients and parents, and lay organizations in some countries provide help in these areas. Prenatal ultrasonography will detect severely affected fetuses at about 16 weeks of pregnancy. Diagnosis by demonstrating synthesis of abnormal proa chains or by DNA sequencing can be carried out in chorionic villa biopsies at 8 to 12 weeks of pregnancy.

EHLERS-DANLOS SYNDROME

EDS is characterized by hyperelasticity of the skin and hypermobile joints.

Classification Five types of EDS were initially defined based primarily on the extent to which the skin, joints, and other tissues are involved, but the classification has now been extended (Table 351-4). Type I is the classic, severe form of the disease, with both severe joint hypermobility and skin that is velvety in texture, hyperextensible, and easily scarred. Type II is similar to type I but milder. In type III joint hypermobility is more prominent than skin changes. In type IV the skin changes are more prominent than joint changes. However, type IV patients are predisposed to sudden death from rupture of large blood vessels or other hollow organs. Type V is similar to type II but is inherited as an X-linked trait. Type VI is characterized by scoliosis, ocular fragility, and a cone-shaped deformity of the cornea (keratoconus). Type VII is characterized by marked joint hypermobility that is difficult to distinguish from type III except by the specific molecular defects in the processing of type I procollagen to collagen. Type VIII is distinguished by periodontal changes. Types IX, X, and XI were defined on the basis of preliminary biochemical and clinical data, but these classifications have not proven useful. Because of overlapping signs and symptoms, many patients and families with some of the features of EDS cannot be assigned to any of the defined types.

Incidence The incidence of EDS is difficult to establish, largely because patients with mild skin or joint symptoms rarely seek medical attention. It is also difficult to define the normal range of variation for joint mobility or skin elasticity. The incidence may be about 1 in 5000 births, although a higher value has been reported for blacks. Types I, II, and III account for most diagnoses.

Skin The changes vary from thin and velvety skin to skin that is either dramatically hyperextensible ("rubber man" syndrome) or easily torn or scarred. Type I patients develop characteristic "cigarette-paper" scars. In type IV extensive scars and hyperpigmentation develop over bony prominences, and the skin may be so thin that subcutaneous blood vessels are visible. In type VIII the skin is more fragile than hyperextensible, and it heals with atrophic, pigmented scars. Easy bruisability occurs in several types of EDS.

Ligament and Joint Changes Laxity and hypermobility of joints vary from mild to unreducible dislocations of hips and other large joints. In mild forms patients learn to reduce dislocations themselves and to avoid them by limiting physical activity. In more severe forms, surgical repair may be required. Some patients have progressive difficulty with age, but severe joint laxity is compatible with a normal life span.

Associated Changes Mitral valve prolapse and hernias occur, particularly in type I. Pes planus and mild to moderate scoliosis are common. Extreme joint laxity and repeated dislocations may lead to degenerative arthritis. In type VI the eye may rupture with minimal trauma, and kyphoscoliosis can cause respiratory impairment. Sclerae may be blue in type VI.

Molecular Defects Mutations in two of the three genes for type V collagen have been found in patients with types I and II EDS. The mutations include glycine substitutions in the triple-helical domain, RNA splicing mutations, exon skipping mutations, a small deletion of 7 bp, and a substitution of serine for cysteine in the C-propeptide. Mutations in both thea1(V) and a2(V) chain are found in patients with type I EDS, but to date only mutations in thea1(V) chain have been found in patients with type II EDS. Electron microscopy of skin from some patients with types I, II, or III EDS are consistent with mutations in a low-abundance collagen such as types III or V that either copolymerize with or bind to the surface of type I fibrils. However, irregular fibrils are not seen in all patients, and similar irregular fibrils can be seen in normal skin.

Most patients with type IV EDS have a defect either in the synthesis or structure of type III procollagen, a finding consistent with the fact that these patients are prone to spontaneous rupture of the aorta and intestines, tissues rich in type III collagen. The thinness and scarring of skin are more difficult to explain, since type III constitutes a small fraction of the collagen in skin (Table 351-1). The >50 mutations identified in the type III procollagen gene include partial gene deletions, RNA splicing mutations, and single-base mutations that cause substitution of glycine by amino acids with bulkier side chains (Fig. 351-4). In effect, most of the mutations lead to synthesis of abnormal but partially functional proa1(III) chains that produce procollagen suicide or alter fibril formation by the same mechanisms that amplify the effects of mutations in the genes for type I procollagen. Similar mutations in type III procollagen can cause aortic aneurysms in some individuals without other evidence of EDS type IV, MFS, or other inherited disorders of connective tissue.

Type VI EDS is caused by mutations in the gene that encodes lysyl hydrolase. In one series, all the patients were homozygous or compound heterozygotes for the mutated genes, and all the mutations caused profound deficiency of lysyl hydroxylase, a decrease in the hydroxylysine content of collagen, and a decrease in the cross-links in collagen fibers. The decrease in cross-links is explained by the observation that some cross-links are less stable if formed from lysine instead of hydroxylysine.

Type VII EDS is due to a defect in the conversion of procollagen to collagen caused either by mutations that make type I procollagen resistant to cleavage by procollagen N-proteinase or by mutations that decrease the activity of the enzyme. The type VIIA mutations alter the cleavage site in the proa1(I) chain, and the type VIIB mutations alter the cleavage site in the proa2(I) chain. Both types are dominantly inherited. Type VIIC is caused by mutations that decrease the activity of procollagen N-proteinase and is inherited as an autosomal recessive trait. In all three forms of type VII EDS, the persistence of the N-propeptide causes the formation of collagen fibrils that are thin and irregular. Since most patients do not have clinical osteopenia, the thin and irregular fibrils apparently suffice for the mineralization of bone but do not provide the necessary tensile strength for ligaments and joint capsules. However, some patients fracture easily

and are difficult to distinguish from variants of OI.

The cause of type VIIIEDS is unknown. Type IX is a disorder of copper transport. The syndrome, also referred to as *Menkes' syndrome*, is due to an X-linked defect and is associated with cutis laxa, hypopigmentation, unusual hair ("kinky"), vascular aneurysms, neurologic degeneration, and mental retardation. Mutations in a gene coding for a copper-transporting ATPase cause the disease (Chaps. 348 and353). Type X EDS may be caused by defects in fibronectin, but no specific mutations have been defined.

Diagnosis The diagnosis is based on clinical criteria. Biochemical assays and gene analyses for known molecular defects inEDS are difficult and time-consuming, but specific diagnostic tests should be available in the future for families in which the genes at fault have been defined.

TREATMENT

There is no specific therapy. Surgical repair and tightening of joint ligaments require careful evaluation of individual patients, as the ligaments frequently do not hold sutures. Patients with easy bruisability should be evaluated for other bleeding disorders. Patients with type IVEDS and members of their families should probably be evaluated at regular intervals by sonography and related techniques for early detection of aneurysms. Surgical repair of aneurysms may be difficult because of increased friability of tissues, and there is limited experience with elective surgery in such patients. Also, women with type IV EDS should be counseled about the increased risk of uterine rupture, bleeding, and other complications of pregnancy.

CHONDRODYSPLASIAS (See alsoChap. 343)

TheCDs are inherited skeletal disorders that cause dwarfism and abnormal body proportions. The category also includes some individuals with normal stature and body proportions who have features such as ocular changes or cleft palate that are common in more severe CDs. Many patients develop degenerative joint changes, and mild CD in adults may be difficult to differentiate from primary generalized osteoarthritis. Some authors refer to the disorders as "skeletal dysplasias," but CD is a more widely used term.

Classification Over 150 distinct types and subtypes have been defined (Table 351-5) based on criteria such as "bringing death" (thanatophoric), causing "twisted" bones (diastrophic), affecting metaphyses (metaphyseal), affecting epiphyses (epiphyseal), and producing histologic changes such as an apparent increase in the fibrous material in the epiphyses (fibrochondrogenesis). Also, a number of eponyms are based on the first or most comprehensive case reports. Severe forms of the diseases produce gross distortions of most cartilaginous structures and of the eye. Mild forms are more difficult to classify. Among the features are cataracts, degeneration of the vitreous and retinal detachment, high forehead, hypoplastic facies, cleft palate, short extremities, and gross distortions of the epiphyses, metaphyses, and joint surfaces.

Incidence Data on the frequency of mostCDs are not available, but the incidence of the

Stickler syndrome may be as high as 1 in 10,000. Therefore, the diseases are probably among the more common heritable disorders of connective tissue.

Molecular Defects The first mutations shown to cause CDs were in the COL2A1 gene for type II collagen, the most abundant protein in cartilage. A number of mutations in this gene have now been reported in variants of CD ranging from mild to lethal (Fig. 351-5). A large fraction of patients with lethal CDs, a smaller number of patients with moderately severe CDs, and about 2% of families with early-onset generalized osteoarthritis have mutations in the same gene. However, similar phenotypes can also be caused by mutations in other genes, including genes for three other collagens, additional components of the cartilage matrix, growth factors, growth factor receptors, and transcription factors (see Table 351-6 for selected examples). The number of mutated genes reported does not necessarily reflect the incidence of such mutations in the diseases themselves but rather the complexity of the genes and the technical difficulties in searching the complete gene for mutations. Also, it reflects the availability of large families for DNA linkage analysis and the vigor with which investigators have pursued their interest in a given gene. It is likely that mutations in additional genes will be found.

Mutations in the COL2A1 gene were first found in patients with severe CDs characterized by gross deformities of bones and cartilage such as spondyloepiphyseal dysplasia congenita, hypochondrogenesis/achondrogenesis II, and the Kniest syndrome. However, mutations in the COL2A1 gene have been found in a few families in which few if any symptoms are present in childhood but in which joint stiffness, joint pain, and degenerative changes of osteoarthritis develop in midlife. The mutations in the COL2A1 gene are similar to the mutations in the genes for types I and III procollagens (Fig. 351-5), and the correlations between genotype and the severity of the phenotype are equally difficult. In addition, mutations that change a codon for a Y-position amino acid in the -Gly-X-Y- repeat sequence from an arginine to cystine were found in families with early-onset osteoarthritis and minimal evidence of CDs. Stickler syndrome and related syndromes are caused by mutations in three different genes: the COL2AI gene for type II collagen and the COL11A1 and COL11A2 genes for type XI collagen. A series of mutations that introduce premature terminal signals in the COL2A1 gene lead to classic Stickler syndrome. However, some patients with classic Stickler syndrome have glycine substitutions in COL11A1. RNA splicing mutations in the COL11A1 gene are found in patients with the Marshall syndrome, which is similar to classic Stickler syndrome but with milder eye changes and more severe hearing loss. Patients classified as having nonocular Stickler syndrome have RNA splicing mutations in the COL11A2 gene.

Many individuals with the Schmid metaphyseal CD, characterized by short stature, *coxa vara*, flaring metaphyses, and waddling gait, have mutations in the gene for the type X collagen, a short, network-forming collagen found primarily in the hypertrophic zone of endochondral cartilage.

Mutations in the receptor for fibroblast growth factor 3 (FGFR-3) are present in most patients with achondroplasia, the most common cause of short-limbed dwarfism accompanied by macrocephaly and dysplasias of the metaphyses of long bones (Table 351-6). The same single-base mutation in the gene that converts glycine to arginine at position 380 is present in>90% of patients. Most patients represent sporadic new mutations, and this nucleotide change must be one of the most common recurring

mutations in the human genome. The mutation causes unregulated signal transduction through the receptor and inappropriate development of cartilage. Mutations that alter other domains of FGFR-3 have been found in patients with the more severe disorders hypochondroplasia and thanatophoric dysplasia and in a few families with a variant of craniosynostosis. However, most patients with craniosynostosis appear to have mutations in the related gene FGFR-2 gene.

Mutations in the gene for the cartilage oligomeric matrix protein (COMP) have been found in patients with multiple epiphyseal dysplasia or pseudoachondroplasia, and in related syndromes characterized by short limbs and degenerative arthritis. However, some families with multiple epiphyseal dysplasia had a mutation in the gene for thea2(IX) or a3(IX) chain of type IX collagen (COL9A2 and COL9A3). All the known mutations in these two type IX collagen genes in patients with multiple epiphyseal dysplasia cause splicing out of the codons of exon 3. A mutation in the COL9A2 gene was also found in patients with the common condition of sciatica and herniations of vertebral discs. About 4% of Finnish probands with the phenotype had a single base substitution that converted a codon for glutamate to tryptophan in thea2(IX) chain of type IX collagen.

Diagnosis The diagnosis of severe forms ofCD is made on the basis of the physical appearance, x-ray findings, histologic changes, and clinical course (Table 351-5).

TREATMENT

No definitive therapy is available. Symptomatic treatment is directed to secondary features such as degenerative arthritis. Many patients require joint replacement surgery and corrective surgery for cleft palate. The eyes should be monitored carefully for the development of cataracts and for the need for laser therapy to prevent retinal detachment. Patients should probably be advised to avoid obesity and contact sports. Counseling for the psychological problems of short stature is critical, and support groups have formed in many countries. Ultrasonography is sometimes successful for prenatal diagnosis but less frequently than withOI. Specific DNA tests are available for theCDscaused by mutations in the genes for types II, X, and XI collagens.

MARFAN SYNDROME

SevereMFS is characterized by a triad of features: (1) long, thin extremities frequently associated with other skeletal changes, such as loose joints and arachnodactyly; (2) reduced vision as the result of dislocations of the lenses (ectopia lentis); and (3) aortic aneurysms that typically begin at the base of the aorta.

Classification The clinical diagnosis is frequently problematic because some affected members of families with MFSpresent with only one or two features of the typical clinical trial. Also, many patients present with one or two of the features of MFS without a family history, apparently because they represent sporadic mutations. Therefore, it is frequently difficult to determine on the basis of clinical data alone whether a patient with ectopia lentis or the characteristic body habitus of MFS is at risk for developing a life-threatening aortic aneurysm. The new DNA diagnostic tests for mutations in the fibrillin-1 and fibrillin-2 genes can probably resolve most, but not all, of these problems.

Most patients who are prone to develop an aortic aneurysm as a component of MFS can be identified by detection of mutations in the fibrillin-1 gene. Some of these patients develop aortic aneurysms because of a mutation in the fibrillin-1 gene without the skeletal or ocular changes characteristic of MFS. Patients with the rarer form of MFS that is characterized by contractural arachnodactyly instead of loose joints can usually be identified by detection of a mutation in the fibrillin-2 gene that is similar in structure to the gene for fibrillin-1. Preliminary data suggest that patients with mutations in the fibrillin-2 gene are not prone to develop aneurysms. However, affected members of some rare families with a mutation in the fibrillin-1 gene also do not develop aortic aneurysms, even though they may show the skeletal or ocular changes. Therefore, the DNA tests are most helpful if: (1) a mutation is detected in either of the two genes, and (2) informative data are available on the clinical symptoms that the same mutation produces in the patient's family or in other families with similar clinical features.

Incidence and Inheritance MFS has an incidence of about 1 in 10,000 in most racial and ethnic groups. The disorder is inherited as an autosomal dominant trait; at least one-fourth of patients do not have an affected parent, and their cases are probably due to new mutations.

Skeletal Changes Patients are usually tall compared with other members of the same family and have long limbs. The ratio of the upper segment (top of the head to top of the pubic ramus) to the lower segment (top of the pubic ramus to the floor) is usually 2 SDs below mean for age, race, and sex. The fingers and hands are long and slender and have a spider-like appearance (arachnodactyly). Many patients have severe chest deformities, including depression (pectus excavatum), protrusion (pectus carinatum), or asymmetry. Scoliosis is frequent and usually accompanied by kyphosis. High-arched palate and high pedal arches or pes planus are common. A few patients have severe joint hypermobility similar to EDS.

Cardiovascular Changes Cardiovascular abnormalities are the major source of morbidity and mortality (Chap. 247). Mitral valve prolapse develops early in life and in about one-quarter progresses to mitral valve regurgitation of increasing severity because of redundancy of the leaflets, stretching of the chordae tendineae, and dilatation of the valvulae annulus. Dilatation of the root of the aorta and the sinuses of Valsalva are characteristic and ominous features of the disease that can develop at any age and in rare instances may be detected by echocardiography in utero. The rate of dilatation is unpredictable, but it can lead to aortic regurgitation, dissection of the aorta, and rupture. Dilatation is probably accelerated by physical and emotional stress, as well as by pregnancy.

Ocular Changes Dislocations of the lens may be readily apparent, but diagnosis usually requires pupillary dilatation and slit-lamp examination. The displacement is usually not progressive but may contribute to the formation of cataracts. The ocular globe is frequently elongated, most patients are myopic, and some develop retinal detachment. A few patients have lattice degeneration and retinal tears; most have adequate vision.

Associated Changes Striae may occur over the shoulders and buttocks. Otherwise the skin is normal. A number of patients develop spontaneous pneumothorax. Inguinal and

incisional hernias are common. Marked dilatation of the dural sac is seen frequently in computed tomography scans, but the condition is usually asymptomatic. Patients are typically thin with little subcutaneous fat, but adults may develop centripetal obesity.

Molecular Defects Most patients with the classic features of MFS are heterozygotes for mutations in a gene on chromosome 15 that encodes fibrillin-1, a glycoprotein of 350 kDa that is a major component of elastin-associated microfibrils. These microfibrils are abundant in large blood vessels and the suspensory ligaments of the lens. Mutations in the fibrillin-1 gene include missense, nonsense, in-frame deletions, and RNA splicing mutations. Many of the mutations are single amino acid substitutions in the epidemal growth factor-like domains of the molecule that may be involved with calcium binding. Mutations in the fibrillin-2 gene that cause the MFS variant characterized by contractures appear to follow a similar pattern. As with most genetic diseases, the nature and location of mutations in the genes are only an approximate guide to the severity of the phenotype unless the same mutation has been seen in other members of the same family or in similar unrelated patients. However, there is a clustering of mutations in the middle portion of the molecule of fibrillin-1 encoded by exons 23 to 32 that causes the most severe phenotype, referred to as *neonatal lethal MFS*. The function of fibrillin has not been defined, but the data suggest that fibrillin self-assembles into a fibrillar structure and that the conformation and surface properties of the entire molecule are critical for normal assembly. Therefore, the functional consequences of mutations that change the amino acid sequence of fibrillin may be similar to the effects of mutations that change the conformation of a fibrillar collagen.

Diagnosis The diagnosis is easily established if the patient and other members of the family have dislocated lenses, aortic dilatation, and long and thin extremities together with kyphoscoliosis or other chest deformities. The diagnosis is frequently made if ectopia lentis and an aneurysm of the ascending aorta occur in the absence of a Marfan habitus or a positive family history. All patients in whom the diagnosis is suspected should have a slit-lamp examination and an echocardiogram. Also, homocystinuria (Table 351-3) should be ruled out by a negative cyanide-nitroprusside test for disulfides in the urine (Chap. 352). A few patients with types I, II, and III EDS have ectopia lentis but lack the Marfan habitus and instead have characteristic skin changes not present in MFS. Patients with familial aortic aneurysms tend to develop aneurysms at the base of the abdominal aorta. The location of the aneurysms, however, is somewhat variable, and the high incidence of aortic aneurysms in the general population (1 in 100) makes the differential diagnosis difficult unless other features of MFS are clearly present. A few families with familial aortic aneurysms have mutations in the gene for type III procollagen (Fig. 351-4).

TREATMENT

There is no established treatment, but several investigators have recommended use of propranolol or other b-adrenergic blocking agents to lower blood pressure and thereby delay or prevent aortic dilatation. Surgical replacement of the aorta, aortic valve, and mitral valve has been successful in some patients, and all patients should be followed carefully with echocardiography and other techniques for evaluation of cardiovascular changes (Chap. 247). Patients should probably be advised of the risks of severe physical and emotional stress and of pregnancy.

The scoliosis tends to be progressive and should be treated by mechanical bracing and physical therapy if>20° or by surgery if it progresses to>45°. Estrogen has been tried in girls with scoliosis, but the results are inconclusive. Dislocated lenses rarely require surgical removal, but patients should be followed closely for retinal detachment.

Diagnostic tests based on detection of fibrillin defects in cultured skin fibroblasts or DNA analysis of the gene are now available from several laboratories.

DISEASES RELATED TO ELASTIN

As may be expected from the role of elastin in maintaining the elasticity of skin, mutations in the elastin gene cause *cutis laxa*, a rare and heterogeneous group of disorders characterized by skin that is both lax and inelastic. Three different frame-shift mutations were found in three unrelated families with dominant forms of the disease. Surprisingly, other mutations in the elastin gene produce phenotypes that primarily involve the aorta, whose elasticity also depends on the presence of elastin. A large deletion that includes the elastin gene and probably several adjacent genes causes the *Williams syndrome*, characterized by supravalvular aortic stenosis, growth retardation, characteristic facies, and an unusual mental phenotype of low intelligence quotient together with a high degree of sociability.

EPIDERMOLYSIS BULLOSA

EBconsists of a group of similar disorders in which the skin and related epithelial tissues break and blister as the result of minor trauma. As with most heritable disorders of connective tissues, the clinical manifestations range from lethal to mild.

Classification Four types of EB are defined on the basis of the level at which blistering occurs:EBsimplex for blistering in the epidermis, EB hemidesmosomal for fissures between kerotinocytes and between kerotinocytes and the basal lamina, EB junctional for blistering in the dermal-epidermal junction, and EB dystrophica for blistering in the dermis (Table 351-7).

Incidence The incidence ofEB in the United States is estimated to be 1 in 50,000.

Molecular Defects The molecular basis of several specific variants ofEB has been defined. A series of patients with EB simplex were found to have mutations in either keratin 14 or keratin 5, two of the major keratins in basal epithelial cells. Patients with the related syndrome, epidermolytic icthyosis, have mutations in keratin 1 and keratin 10. The new disease phenotype of hemidesmosomal EB has three clinical variants that are caused by mutations in one of four genes (Table 351-7): (1) A generalized atrophic and benign form of EB is caused by mutations in the COL17A1 gene for type XVII collagen; (2) a variant with EB associated with pyloric atresia and other intestinal abnormalities is caused by mutations in either the gene for thea6 integrin (ITG A6) or the gene for theb4 integrin (ITG B4); and (3) another variant characterized by relatively mild blistering at birth but associated with late-onset muscular dystrophy is caused by mutations in the gene for plectin (PLEC-1). Junctional EB is caused by mutations in any one of three genes for laminin (LAMA-3, LAMB-3, LAMC-2). The most severe dystrophic

form of EB is caused by mutations in the gene for type VII collagen (COL7A1).

Diagnosis The diagnosis is based on skin that readily breaks and forms blisters.EBsimplex and EB hemidesmosomal are generally milder than EB junctional or EB dystrophica. EB dystrophica variants usually cause large and prominent scars. Precise classification within subtypes usually requires electron microscopy. The treatment is symptomatic.

ALPORT SYNDROME (See also Chap. 275)

ASis an inherited disorder characterized by hematuria. Four forms of the disease are now recognized: (1) classic AS, which is inherited as an X-linked disorder with hematuria, sensorineural deafness, and conical deformation of the anterior surface of the lens (lenticonus); (2) a subtype of the X-linked form associated with diffuse leiomyomatosis; (3) an autosomal recessive form; and (4) an autosomal dominant form. The two autosomal forms can cause renal disease without deafness or lenticonus.

Incidence The incidence ofAS is about 1 in 10,000 in the general population and as high as 1 in 5000 in some ethnic groups. About 80% of AS patients have the X-linked variant.

Molecular Defects Electron microscopy of kidneys from patients with classicASdemonstrates that the glomerular basement membrane is up to five times thicker than normal and that the lamina densa is distorted and split. The X-linked and autosomal recessive forms are caused primarily by mutations in genes for thea3(IV), a4(IV),a5(IV), or a6(IV) chains of type IV collagen, a major component of basement membranes. The type IV collagen in most membranes consists primarily of a1(IV) anda2(IV) chains folded into a large, rodlike molecule with globular ends and a long triple-helical domain that is interrupted by short sequences that do not form triple helices. The molecules self-assemble through both the globular ends and the long triple-helical domain to form a complex three-dimensional network. The four additional achains of type IV collagen are minor components of basement membranes, similar in structure, and are probably incorporated into the same or similar molecules. The six genes for the proteins are arranged in tandem pairs on different chromosomes in a head-to-head orientation and with overlapping promoters, i.e., the a1(IV) and a2(IV) genes are head-to-head on chromosome 13q34, the a3(IV) anda4(IV) genes are on chromosome 2q35-37, and thea5(IV) and a6(IV) genes are on chromosome Xq22. An X-linked variant is caused by mutations in the COL4A5 gene, and the X-linked variant associated with leiomyomatosis is caused by deletions that involve both the COL4A5 gene and the nearby COL4A6 gene. The autosomal recessive variants are caused by mutations in either the COL4A3 or COL4A4 genes. The mutations responsible for the autosomal dominant variants are still unknown, but they have been mapped to the same locus as the COL4A3 and COL4A4 genes.

Diagnosis The diagnosis of classicAS is based on X-linked inheritance of hematuria, sensorineural deafness, and lenticonus. Because of the X-linked transmission, women are usually less severely affected than men and are generally underdiagnosed. The hematuria progresses to nephritis and may cause renal failure in late adolescence in affected males and at older ages in some women. The sensorineural deafness is

primarily in the high-tone range. It can frequently be detected only by an audiogram and is usually not progressive. The lenticonus can occur without nephritis but is generally considered to be pathognomonic of classic AS.

TREATMENT

There is no known treatment, but renal transplantation is usually successful.

(Bibliography omitted in Palm version)

Back to Table of Contents

352. INHERITED DISORDERS OF AMINO ACID METABOLISM AND STORAGE - *Nicola Longo*

All polypeptides and proteins are polymers of amino acids. Eight amino acids, referred to as *essential*, cannot be synthesized by humans and must be obtained from dietary sources. The others are formed endogenously. Although most of the body's amino acids are "tied up" in proteins, small intracellular pools of *free* amino acids are in equilibrium with extracellular reservoirs in plasma, cerebrospinal fluid, and the lumina of the gut and kidney. Physiologically, amino acids are more than mere "building blocks" of proteins. Some (glycine, glutamate,g-aminobutyric acid) are neurotransmitters. Others (phenylalanine, tyrosine, tryptophan, glycine) are precursors of hormones, coenzymes, pigments, purines, or pyrimidines. Each has a unique degradative pathway by which its nitrogen and carbon components are used for the synthesis of other amino acids, carbohydrates, and lipids.

More than 70 disorders of amino acid metabolism are now known. The catabolic and storage defects (approximately 60) discussed in this chapter far outnumber the transport abnormalities (approximately 10) considered in Chap. 353. Each of these disorders is rare -- the incidences range from 1 in 10,000 for cystinuria or phenylketonuria to 1 in 200,000 for homocystinuria or alkaptonuria. Collectively, however, they occur in perhaps 1 in 500 to 1 in 1000 live births. Almost all are transmitted as autosomal recessive traits.

The features of inherited disorders of amino acid catabolism are summarized inTable 352-1. In general, these disorders are named for the compound that accumulates to highest concentration in blood (-*emias*) or urine (-*urias*). For many conditions (often called *aminoacidopathies*), the parent amino acid is found in excess; for others, generally referred to as *organic acidemias*, products in the catabolic pathway accumulate. Which compound(s) accumulates depends, of course, on the site of the enzymatic block, the reversibility of the reactions proximal to the lesion, and the availability of alternative pathways of metabolic "runoff." For some amino acids, such as the sulfur-containing or branched-chain molecules, defects have been described at nearly each step in the catabolic pathway. For others, only small numbers of defective reactions have been described. Biochemical and genetic heterogeneity is common. Four distinct forms of hyperphenylalaninemia, seven forms of homocystinuria, and seven types of methylmalonic acidemia are recognized. Such heterogeneity reflects the presence of an even larger array of molecular defects.

The manifestations of these conditions differ widely (Table 352-1). Some, such as sarcosinemia or hyperprolinemia, produce no clinical consequences. At the other extreme, complete deficiency of ornithine transcarbamylase or of branched-chain keto acid dehydrogenase is lethal in the untreated neonate. Central nervous system (CNS) dysfunction, in the form of developmental retardation, seizures, alterations in sensorium, or behavioral disturbances, is present in more than half the disorders. Protein-induced vomiting, neurologic dysfunction, and hyperammonemia occur in many disorders of urea cycle intermediates. Metabolic ketoacidosis, often accompanied by hyperammonemia, is a frequent presenting finding in the disorders of branched-chain amino acid metabolism. Occasional disorders produce focal tissue or organ involvement such as liver disease, renal failure, cutaneous abnormalities, or ocular lesions.

The clinical manifestations in many of these conditions can be prevented or mitigated if diagnosis is achieved early and appropriate treatment (i.e., dietary protein or amino acid restriction or vitamin supplementation) is instituted promptly. For this reason, several aminoacidopathies and organic acidemias are routinely screened in newborns using an array of chemical and microbiologic techniques. Once a presumptive diagnosis is made, confirmation can be provided by direct enzyme assay on extracts of leukocytes, erythrocytes, or cultured fibroblasts. DNA-based testing is possible for several disorders including phenylketonuria, ornithine transcarbamylase deficiency, citrullinemia, gyrate atrophy of the retina, propionic acidemia, and methylmalonic acidemia. As additional mutations are defined, DNA-based analysis may allow better predictions of outcome and improved therapeutic plans.

Several of these disorders (including branched-chain ketoaciduria, isovaleric acidemia, propionic acidemia, methylmalonic acidemia, homocystinuria, cystinosis, phenylketonuria, ornithine transcarbamylase deficiency, citrullinemia, argininosuccinic aciduria) can be diagnosed prenatally by chemical analysis of amniotic fluid or by chemical, enzymatic, or DNA-based studies of fresh or cultured amniotic fluid cells. In addition to predicting genotype and alleviating parental anxiety, prenatal diagnosis has led to improved treatment of affected newborns.

The focus of this chapter is on selected disorders that illustrate the principles, properties, and problems presented by the disorders of amino acid metabolism.

THE HYPERPHENYLALANINEMIAS

DEFINITION

The hyperphenylalaninemias (Table 352-1) result from impaired conversion of phenylalanine to tyrosine. The most common and clinically important is *phenylketonuria*, which is characterized by an increased concentration of phenylalanine in blood, increased concentrations of phenylalanine and its by-products (notably phenylpyruvate, phenylacetate, phenyllactate, and phenylacetylglutamine) in urine, and severe mental retardation if untreated in infancy.

GENETIC CONSIDERATIONS

Each of the hyperphenylalaninemias results from reduced activity of phenylalanine hydroxylase. In humans, this complete enzyme system is expressed only in liver. Phenylalanine and molecular oxygen are substrates, and a reduced pteridine, tetrahydrobiopterin, is a cofactor (Fig. 352-1). Tyrosine and dihydrobiopterin are the products of this catalytic system, the latter being reconverted to tetrahydrobiopterin by two enzymes, pterin-4a-carbinolamine dehydratase and dihydropteridine reductase.

Abnormalities in phenylalanine metabolism are autosomal recessive traits that occur in about 1 in 10,000 births. Phenylketonuria type I is widely distributed among whites and Orientals. It is rare in blacks. Phenylalanine hydroxylase activity in obligate heterozygotes is low but higher than in affected homozygotes. Adult heterozygous carriers are clinically well but can be identified by an increased ratio of phenylalanine/tyrosine in plasma in the semifasting state. They may have transient

cognitive impairment after phenylalanine loads. Hyperphenylalaninemias are caused by mutations in the gene encoding phenylalanine hydroxylase (*PAH*) or in genes encoding enzymes involved in tetrahydrobiopterin synthesis or recycling. In the vast majority of patients, mutations occur in the *PAH* gene (causing phenylketonuria type I), and>300 mutations have been identified. Mutations causing a complete impairment of enzyme activity, such as the R408W, are associated with a more severe outcome requiring stringent dietary restriction of phenylalanine. Mutations causing a less complete deficiency of the enzyme, such as the I65T, are associated with milder forms of phenylketonuria.

In fewer than 2% of patients with phenylketonuria, mutations occur in other genes, including dihydrobiopterin reductase (*DHPR*) (30%), 6-pyruvoyl-tetrahydropterin synthase (*6-PTS*) (60%), GTP cyclohydrolase I (*GTP-CH*) (5%), and pterin-4a-carbinolamine dehydratase (*PCD*) (5%). In these cases, the impairment in phenylalanine hydroxylation results from tetrahydrobiopterin deficiency due to blocks in the pathway by which tetrahydrobiopterin is synthesized from GTP (phenylketonuria type III and malignant hyperphenylalaninemia) or deficiency of dihydropterine reductase (phenylketonuria type II), the enzyme that regenerates tetrahydrobiopterin from dihydrobiopterin (Fig. 352-1). Tyrosine hydroxylase and tryptophan hydroxylase also require tetrahydrobiopterin. Their products (L-dopa and 5-hydroxytryptophan) are essential for the synthesis of neurotransmitters. Heterozygotes for these conditions do not have hyperphenylalaninemia, but carriers of mutations in GTP cyclohydrolase have a peculiar form of dystonia, transmitted as a dominant trait with variable expressivity and higher penetrance in females; it is exquisitely responsive to levodopa. Neurotransmitter levels are not altered in transient hyperphenylalaninemia (sometimes called *transient phenylketonuria*), which has been described in some patients with pterin-4a-carbinolamine dehydratase deficiency.

Etiology and Pathogenesis Phenylalanine accumulation in blood and urine and reduced tyrosine formation are direct consequences of the impaired hydroxylation. In untreated phenylketonuria and in its tetrahydrobiopterin-deficient variants, plasma concentrations of phenylalanine become sufficiently high [1 mmol/L (16 mg/dL)] to activate alternative pathways of metabolism and lead to formation of phenylpyruvate, phenylacetate, phenyllactate, and other derivatives that are rapidly cleared by the kidney and excreted in urine. The severe brain damage is due to several consequences of phenylalanine accumulation: competitive inhibition of transport of other amino acids required for protein synthesis, impaired polyribosome formation or stabilization, reduced synthesis and increased degradation of myelin, and inadequate formation of norepinephrine and serotonin. Phenylalanine is a competitive inhibitor of tyrosinase, a key enzyme in the pathway of melanin synthesis. This block, plus reduced availability of the melanin precursor tyrosine, accounts for the hypopigmentation of hair and skin.

Clinical Manifestations No abnormalities are apparent at birth, but untreated children with classic phenylketonuria fail to attain early developmental milestones, develop microcephaly, and demonstrate progressive impairment of cerebral function. Hyperactivity, seizures, and severe mental retardation are major clinical problems later in life. Electroencephalographic abnormalities; "mousy" odor of skin, hair, and urine (due to phenylacetate accumulation); and a tendency to hypopigmentation and eczema complete the devastating clinical picture. In contrast, affected children who are detected

at birth and treated promptly show none of these abnormalities. Children with tetrahydrobiopterin deficiency, however, suffer a worse clinical course. Seizures appear early, followed by progressive cerebral and basal ganglia dysfunction (rigidity, chorea, spasms, hypotonia). Most succumb to secondary infection within a few years despite early diagnosis and vigorous treatment.

A number of women with phenylketonuria who have been treated since infancy have reached adulthood and become pregnant. If maternal phenylalanine levels are not strictly controlled before and during pregnancy, their offspring are at risk, even if they are heterozygous, for *maternal phenylketonuria*. Affected children have microcephaly and an increased risk of congenital defects. After birth, these children have severe neurodevelopmental delay and growth retardation.

Diagnosis Plasma phenylalanine concentrations are usually normal at birth in the hyperphenylalaninemias but rise rapidly after institution of protein feedings. To prevent mental retardation, diagnosis and initiation of dietary treatment of classic phenylketonuria must occur before the child is 3 weeks of age. For this reason, most newborns in North America and Europe are screened by determinations of blood phenylalanine concentration using the Guthrie bacterial inhibition assay. Abnormal values are confirmed using quantitative analysis of plasma amino acids. Prenatal diagnosis of type I phenylketonuria is now feasible using DNA-based tests that detect specific mutations or polymorphic markers that are linked to the *PAH* gene.

In newborns with type I phenylketonuria, plasma levels depend on the amount of phenylalanine in the diet and the degree of impairment of phenylalanine hydroxylase. Dietary phenylalanine restriction is usually instituted if blood phenylalanine levels are >250 umol/L (4 mg/dL). Careful monitoring of these infants reveals the degree of phenylalanine hydroxylase impairment and dictates the degree of dietary phenylalanine restriction.

Deficiency of tetrahydrobiopterin, which occurs in 1 to 2% of newborns with increased blood phenylalanine, is excluded by screening the urinary pteridine profile and by assay of dihydropteridine reductase activity on dried blood specimens. Dihydropteridine reductase deficiency and the blocks in tetrahydrobiopterin synthesis can be detected in utero using assays on cultured amniocytes. Tetrahydrobiopterin deficiency manifests with hyperphenylalaninemia and progressive neurologic impairment despite prompt dietary restriction of phenylalanine.

TREATMENT

Phenylketonuria is the first inherited metabolic disease in which a strategy of reducing the accumulation of the offending metabolite prevented the dire clinical consequences. This was accomplished by a special diet low in phenylalanine and supplemented with tyrosine. Tyrosine becomes an essential amino acid in phenylalanine hydroxylase deficiency. Sufficient phenylalanine is provided for new protein synthesis and normal growth. This amount varies with age and requires frequent adjustments, especially early in life. Ordinarily, plasma phenylalanine concentrations are maintained between 120 and 360 umol/L (2 and 6 mg/dL). Such diet therapy must be instituted during the first 3 weeks of life. Even then, modest CNS dysfunction may occur with more deleterious

mutations or after excess protein intake. Because uncontrolled hyperphenylalaninemia results in brain damage, dietary restriction should be continued and monitored indefinitely, recognizing that transient phenylketonuria may not require lifelong therapy. An enteric-coated formulation of phenylalanine ammonia lyase, which degrades phenylalanine in the gut, is under investigation for its potential to reduce the strain imposed by the special diet.

Children with tetrahydrobiopterin deficiency deteriorate despite dietary phenylalanine restriction. Such patients may be helped, however, by a regimen in which dietary phenylalanine restriction is combined with tetrahydrobiopterin supplements, levodopa, 5-hydroxytryptophan, and carbidopa. Finally, the deleterious consequences of maternal phenylketonuria can be minimized by continuing lifelong phenylalanine-restricted diets in females with phenylketonuria and assuring strict phenylalanine restriction prior to conception and throughout gestation.

THE HOMOCYSTINURIAS (HYPERHOMOCYSTEINEMIAS)

The homocystinurias are seven biochemically and clinically distinct disorders (Table 352-1), each characterized by increased concentration of the sulfur-containing amino acid homocystine in blood and urine. The most common form results from reduced activity of cystathionineb-synthase (CBS), an enzyme in the transsulfuration pathway that converts methionine to cysteine (Fig. 352-2). The other forms are the result of impaired conversion of homocysteine to methionine, a reaction catalyzed by homocysteine:methyltetrahydrofolate methyltransferase (also known as *methionine synthase*) and two essential cofactors, methyltetrahydrofolate and methylcobalamin (methyl-vitamin B_{12}). Depending on the underlying disorder, some patients show chemical and, in some instances, clinical improvement following administration of specific vitamin supplements (pyridoxine, folate, or cobalamin) (Chap. 75). In classic homocystinuria, the levels of free homocystine in plasma increase and result in homocystinuria. *Hyperhomocysteinemia* refers to increased total plasma concentration of homocysteine in the sulfhydryl and disulfide form, free and protein-bound. Hyperhomocysteinemia, in the absence of significant homocystinuria, is found in individuals who are heterozygous or homozygous for certain genetic defects that impair folate or vitamin B_{12} metabolism or cause cystathionine synthase deficiency. Changes of homocysteine levels are also observed with increasing age; in postmenopausal women; in patients with renal failure, hypothyroidism, leukemias, or psoriasis; and during therapy with drugs such as methotrexate, nitrous oxide, isoniazid, and some antiepileptic agents.

Homocysteine acts as an atherogenic and thrombophilic agent. An increase in total plasma homocysteine represents an independent risk factor for coronary, cerebrovascular, and peripheral arterial disease as well as for deep-vein thrombosis (Chap. 241). Homocysteine is synergistic with hypertension and smoking, and it is additive with other risk factors that predispose to peripheral arterial disease. In addition, hyperhomocysteinemia and folate and vitamin B_{12} deficiency have been associated with an increased risk of neural tube defects in pregnant women.

CYSTATHIONINEb-SYNTHASE DEFICIENCY

Definition Deficiency of this enzyme leads to increased concentrations of methionine and homocystine in body fluids and to decreased concentrations of cysteine and cystine. Clinical hallmarks include dislocation of optic lenses (usually downward and medially), mental retardation, marfanoid habitus, osteoporosis, and thrombotic vascular disease.

GENETIC CONSIDERATIONS

The sulfur atom of the essential amino acid methionine is transferred to cysteine by the transsulfuration pathway (Fig. 352-2). In one of these steps, homocysteine condenses with serine to form cystathionine. This reaction is catalyzed by the pyridoxal phosphate-dependent enzyme CBS. Heterogenous mutations in the *CBS* gene are present in different families. The G307S mutation is associated with lack of response to pyridoxine, whereas the I278T mutation correlates with pyridoxine-responsiveness and a milder clinical phenotype. Homocystinuria is common in Ireland (1 in 60,000 births) but rare elsewhere (<1 in 200,000 births).

Etiology and Pathogenesis Homocysteine and methionine accumulate in cells and body fluids; cysteine synthesis is impaired, resulting in reduced concentrations of this amino acid and its disulfide form, cystine. In approximately half of patients, synthase activity in liver, brain, leukocytes, and cultured fibroblasts is undetectable. In the remaining patients, tissues retain 1 to 5% of normal activity, and this residual activity can often be stimulated by pyridoxine supplementation.

Homocysteine interferes with the normal cross-linking of collagen, an effect that likely plays an important role in the ocular, skeletal, and vascular complications. Altered collagen in the suspensory ligament of the optic lens and in bone matrix may account for the dislocated lenses and osteoporosis. Similarly, interference with normal ground substance metabolism in vascular walls may predispose to the arterial and venous thrombotic diathesis. Increased platelet adhesiveness may result from homocysteine accumulation, thereby contributing to the thrombotic occlusive disease so often observed. Recurrent cerebrovascular accidents secondary to thrombotic disease may account for the mental retardation, but direct chemical effects on cerebral cell metabolism have not been excluded.

Clinical Manifestations More than 80% of homozygotes for complete CBS deficiency develop dislocated optic lenses. This abnormality usually appears by 3 to 4 years of age and often results in glaucoma and impaired visual acuity. Mental retardation occurs in about half of such patients, often accompanied by ill-defined behavioral disturbances. Osteoporosis is a common radiologic finding (seen in two-thirds of patients by age 15) but rarely causes clinical disease. Life-threatening vascular complications, probably initiated by damage to vascular endothelium, are the major cause of morbidity and mortality. Occlusion of coronary, renal, and cerebral arteries with attendant tissue infarction can occur during the first decade of life. Nearly a fourth of patients die of vascular disease before age 30. These vascular complications seem to be exacerbated by angiographic procedures. Importantly, pyridoxine-responsive patients have milder clinical manifestations in all regards and may escape newborn screening and present with ectopia lentis or premature vascular occlusion. Heterozygous carriers for *CBS* deficiency (about 1 in 70 in the population) may have hyperhomocysteinemia, with an

increased risk for premature coronary, peripheral, and cerebral occlusive vascular disease.

Diagnosis The cyanide-nitroprusside test is a simple way of demonstrating increased excretion of sulfhydryl-containing compounds in urine. This is confirmed by measurement of free plasma methionine and homocystine. Plasma methionine tends to be increased in synthase-deficient patients and normal or low in those with other causes of homocystinuria and impaired methionine formation (see below). Diagnostic confirmation depends on measurements of CBS activity in tissue extracts or cells cultured from patients. Heterozygotes can be identified by measurement of peak serum homocystine after an oral methionine load (100 mg/kg) and by measurement of tissue synthase activity.

TREATMENT

As with classic phenylketonuria, effective treatment depends on early diagnosis. A number of infants diagnosed in the newborn period have been treated successfully with methionine-restricted, cystine-supplemented diets. In approximately half of patients, oral pyridoxine (25 to 500 mg/d) produces a fall in plasma and urinary methionine and homocystine and an increase in cystine concentration in body fluids. This effect probably reflects a modest increase in CBS activity in cells of patients in whom the defect is characterized either by reduced affinity for cofactor or by accelerated degradation of mutant enzyme. Vitamin supplementation at these doses is apparently harmless and should be tried in all patients. Folate deficiency should be prevented by adequate supplementation. Betaine has also been effective in reducing homocystine levels in pyridoxine-unresponsive patients.

5,10-METHYLENETETRAHYDROFOLATE REDUCTASE (MTFR) DEFICIENCY

Definition Hyperhomocysteinemia with normal or decreased methionine levels is caused by deficiency of MTFR, the enzyme involved in the synthesis of 5-methyltetrahydrofolate, a cofactor in the enzymatic formation of methionine from homocysteine (Fig. 352-2). CNS dysfunction and premature vascular occlusion may occur.

Genetic Basis and Pathogenesis 5-Methyltetrahydrofolate:homocysteine methyltransferase (methionine synthase) catalyzes the conversion of homocysteine to methionine. A primary defect in MTFR activity results, secondarily, in deficient methyltransferase activity and impaired conversion of homocysteine to methionine. This series of reactions is critical to normal DNA and RNA synthesis. Methionine deficiency and impaired nucleic acid synthesis may contribute to CNS dysfunction, while homocystine accumulation may predispose to thrombosis.

Hyperhomocysteinemia is inherited as an autosomal recessive trait and is caused by mutations in the MTFR (*MTHFR*) gene, which is located on chromosome 1p36.3. A thermolabile variant form of this enzyme, which has reduced activity, may be a common cause of hyperhomocysteinemia associated with increased risk of vascular disease in young adults.

Clinical Manifestations More than 30 patients with homocystinuria due to MTFR deficiency have been reported. The most severely affected have developmental retardation and cerebral atrophy early in life. Others have behavioral disturbances (catatonia) during the second decade or mild retardation. The severity of the clinical manifestations reflects the severity of the reductase deficiency.

Diagnosis Increased concentrations of free homocystine in body fluids with normal or decreased concentrations of methionine suggest severe MTFR deficiency. Total plasma homocysteine levels slightly above the normal range suggest milder dysfunction of this enzyme. Serum folate concentration is low in some patients. Confirmation requires direct MTFR assays in cultured fibroblasts.

TREATMENT

Therapeutic experience is limited. Folate, vitamin B_{12}, methionine, or betaine supplementation decrease homocystine urinary excretion and improve the clinical manifestations in some patients.

DEFICIENCY OF COBALAMIN (VITAMIN B_{12}) COENZYME SYNTHESIS

Definition Five other forms of homocystinuria also reflect impaired conversion of homocysteine to methionine. The primary defects in these entities, however, are in the synthesis of methylcobalamin, a cobalamin (vitamin B_{12}) coenzyme required by methionine synthase (MS) (Fig. 352-2). In some, methylmalonic acid accumulates in body fluids because of impaired synthesis of a second coenzyme, adenosylcobalamin, required for isomerization of methylmalonyl coenzyme A (CoA) to succinyl CoA. These disorders are designated cblC, -D, -E, -F, and -G.

Etiology and Pathogenesis As with MTFR deficiency, each disorder impairs remethylation of homocysteine. Since methylcobalamin is required for methyl-group transfer from methyltetrahydrofolate to homocysteine, impaired cobalamin metabolism leads to deficient methyltransferase activity. The defects responsible for impaired synthesis of methylcobalamin involve one of several steps in lysosomal or cytosolic activation of the vitamin precursor (Fig. 352-2). Somatic cell genetic studies indicate that each of the five abnormalities (cblC to -G) is distinct and imply that all are inherited as autosomal recessive traits.

Clinical Manifestations More than 45 patients -- mostly children -- with these defects in cobalamin metabolism have been described. Although clinical manifestations vary, abnormalities include developmental delay, dementia, spasticity, megaloblastic anemia, and pancytopenia. It is not possible to define a specific clinical syndrome for each of the defects in cobalamin metabolism.

Diagnosis Homocystinuria, homocysteinemia, and hypomethioninemia are the chemical hallmarks. Methylmalonic acidemia, too, has been noted in those defects resulting from defective synthesis of both cobalamin coenzymes. These findings may also be present in juvenile- or adult-onset pernicious anemia in which intestinal cobalamin absorption is impaired. Measurement of serum cobalamin concentrations, low in pernicious anemia and normal in patients with defective conversion of cobalamin vitamin to coenzymes,

helps in the differential diagnosis. Definitive diagnosis depends on demonstrating impaired coenzyme synthesis in cultured cells.

TREATMENT

Treatment of affected children with hydroxycobalamin injections (1 to 2 mg/QD) and betaine supplements decreases homocystine and methylmalonate excretion; the hematologic and neurologic deficits have also been diminished to variable degrees in some patients. Intervention early in life seems to offer the best long-term prognosis.

ALKAPTONURIA

Definition Alkaptonuria is a rare disorder of tyrosine catabolism in which deficiency of homogentisate 1,2-dioxygenase (also known as *homogentisic acid oxidase*) leads to excretion of large amounts of homogentisic acid in urine and accumulation of oxidized homogentisic acid pigment in connective tissues (*ochronosis*). After many years, ochronosis produces a distinctive form of degenerative arthritis.

Genetic Basis and Pathogenesis Alkaptonuria was the first human disease shown to be inherited as an autosomal recessive trait. Affected homozygotes occur with a frequency of about 1 in 200,000. Heterozygous carriers are clinically well and excrete no homogentisic acid in urine, even after loading doses of tyrosine. The gene for homogentisate 1,2-dioxygenase (*HGD*) maps to chromosome 3q21-q23 and encodes a 445 amino acid protein expressed not only in liver and kidney but also in small intestine, colon, and prostate. Expression in this latter organ is consistent with accumulation of black calculi of homogentisic acid in the prostate of patients with alkaptonuria, sometimes requiring prostatectomy.

Patients have minimally increased concentrations of homogentisic acid in blood because it is rapidly cleared by the kidney. However, homogentisic acid accumulates in cells and body fluids. Its oxidized polymers bind to collagen, leading to the progressive deposition of a gray to bluish-black pigment. The mechanism(s) by which this deposition causes degenerative changes in cartilage, intervertebral disk, and other connective tissues is unknown but may involve direct chemical irritation, impaired collagen cross-linking, disturbed articular chondrocyte metabolism, or some combination of factors.

Clinical Manifestations Alkaptonuria may go unrecognized until middle life when degenerative joint disease develops. Prior to this time, the tendency of the patient's urine to darken on standing may go unnoticed, as may slight pigmentation of the sclerae and ears. The latter manifestations are generally the earliest external evidence of the disorder and develop after age 20 to 30. Foci of gray-brown scleral pigment and generalized darkening of the concha, anthelix, and finally, helix of the ear are typical. Ear cartilages may be irregular and thickened. *Ochronotic arthritis* is heralded by pain, stiffness, and some limitation of motion of the hips, knees, and shoulders. Acute arthritis may resemble rheumatoid arthritis, but small joints are usually spared. Limitation of motion and ankylosis of the lumbosacral spine are common late manifestations. Pigmentation of heart valves, larynx, tympanic membranes, and skin occurs, and occasional patients develop pigmented renal or prostatic calculi. Degenerative

cardiovascular disease may be increased in older patients.

Diagnosis A patient whose urine darkens to blackness on standing must be suspected of having alkaptonuria, but this may not be observed with the use of modern plumbing conditions. The diagnosis is usually made from the triad of degenerative arthritis, ochronotic pigmentation, and urine that turns black upon alkalinization. Homogentisic acid in urine may be identified presumptively by other tests: after addition of ferric chloride, a purple-black color is observed; treatment with Benedict's reagent yields a brown color; addition of a saturated silver nitrate solution produces an intermediate black color. These screening tests can be confirmed by chromatographic, enzymatic, or spectrophotometric determinations of homogentisic acid. X-rays of the lumbar spine show degeneration and dense calcification of the intervertebral disks and narrowing of the intervertebral spaces (bamboo-like appearance).

TREATMENT

There is no specific treatment for ochronotic arthritis. Joint manifestations might be mitigated if homogentisic acid accumulation and deposition could be curbed by dietary restriction of phenylalanine and tyrosine, but the long-term nature of the disease discourages such therapeutic attempts. Ascorbic acid impedes oxidation and polymerization of homogentisic acid in vitro, but the efficacy of this form of treatment has not been established. Symptomatic treatment is similar to that for osteoarthritis (Chap. 321).

CYSTINOSIS

Definition Cystinosis is a rare disorder characterized by the intralysosomal accumulation of free cystine in body tissues. This results in the appearance of cystine crystals in the cornea, conjunctiva, bone marrow, lymph nodes, leukocytes, and internal organs. Three variants have been identified: an infantile (nephropathic) form leading to the Fanconi syndrome and renal insufficiency in the first decade, a juvenile (intermediate) form in which renal disease is manifest during the second decade, and an adult (benign) form characterized by deposition of cystine in the cornea but not in the kidney.

GENETIC CONSIDERATIONS

All types are inherited as autosomal recessive traits. The gene for the infantile form of nephropathic cystinosis encodes an integral membrane protein, which is a putative lysosomal cystine transporter. The gene is located on chromosome 17p13 and is designated *CTNS*. It is highly expressed in the pancreas, kidney, skeletal muscle, placenta, and heart. A common 65-kb deletion accounts for the majority of patients of European ancestry with infantile cystinosis. The juvenile- and adult-onset forms of nephropathic cystinosis are allelic with the infantile form. Obligate heterozygotes have intracellular cystine concentrations intermediate between those of normal persons and affected patients, but they are free of clinical abnormalities.

The basic defect involves impaired efflux of cystine from lysosomes rather than an abnormality in cystine catabolism. Lysosomal cystine efflux is an active, ATP-dependent

process. The cystine content of tissues may be more than 100 times normal in the infantile form and more than 30 times normal in the adult form. Intracellular cystine in lysosomes does not exchange with other intra- or extracellular pools of this amino acid. Neither plasma nor urinary concentrations of cystine are particularly elevated. Cystine accumulation in the kidney causes renal insufficiency in the infantile and juvenile forms. Patchy depigmentation and degeneration of the peripheral retina occur in the infantile and juvenile forms. Cystine crystals may also be deposited in the cornea, ocular conjunctiva, or uvea.

Clinical Manifestations In the infantile form, abnormalities are usually apparent by 6 to 10 months of age. Growth retardation, vomiting, fever, vitamin D-resistant rickets, polyuria, dehydration, and metabolic acidosis are prominent. Generalized proximal tubular dysfunction (the Fanconi syndrome) leads to hyperphosphaturia and hypophosphatemia, renal glycosuria, generalized amino aciduria, low plasma carnitine, hypouricemia, and often hypokalemia. Death due to uremia or intercurrent infection usually occurs before age 10. Ocular manifestations are prominent. Photophobia is usually demonstrable within the first years of life due to cystine deposits in the cornea, and retinal degeneration may appear even earlier. Hypothyroidism, insulin-dependent diabetes mellitus, and delayed puberty are often observed in older patients.

In contrast, patients with the adult form have only ocular abnormalities. Photophobia, headache, and burning or itching of the eyes are major complaints. Glomerular and tubular function and the integrity of the retina are preserved. The findings in the juvenile variant fall between these extremes. Ocular and renal manifestations do not become significant until the second decade. The renal lesion, albeit milder than that in the infantile form, eventually leads to renal insufficiency.

Diagnosis Cystinosis must be considered in any child with vitamin D-resistant rickets, the Fanconi syndrome, or glomerular insufficiency. Hexagonal or rectangular cystine crystals can be detected in the cornea (by slit-lamp examination), in leukocytes from peripheral blood or bone marrow, or in biopsies of rectal mucosa. Diagnosis is confirmed by measurement of cystine in leukocytes. The infantile form has been diagnosed prenatally by the demonstration of increased cystine content in cultured amniotic fluid cells.

TREATMENT

The adult form is benign and requires no treatment. Symptomatic treatment of renal disease in the infantile or juvenile form includes maintenance of adequate fluid intake to prevent dehydration; correction of the metabolic acidosis; administration of supplementary calcium, phosphate, and vitamin D to heal the rickets; and carnitine supplements (100 mg/kg per day) to correct the increased urinary losses. Specific therapy with the free thiol cysteamine slows the progression of renal dysfunction and improves growth. This compound acts in lysosomes by forming a mixed disulfide with cysteine, allowing it to be transported out of the organelle by an unrelated transporter not affected by the disease. Treatment is more effective if initiated before the patient is 2 years of age. Eye drops containing cysteamine can remove corneal crystals but requires frequent applications (10 to 14 times a day).

Children with nephropathic cystinosis and end-stage renal disease benefit from kidney transplantation. Patients who tolerate the procedure and do not develop immunologic problems have return of kidney function toward normal. The transplanted kidneys have not developed the functional abnormalities typical of cystinosis (i.e., the Fanconi syndrome or glomerular insufficiency). Patients may, however, continue to accumulate cystine in the cornea and other organs (thyroid, brain, and muscle).

PRIMARY HYPEROXALURIA

Definition Primary hyperoxaluria is the designation for two rare autosomal recessive disorders characterized by chronic excessive urinary excretion of oxalic acid and by calcium oxalate nephrolithiasis and nephrocalcinosis. Typically, patients with both forms develop renal insufficiency early in life and die of uremia. Calcium oxalate deposits are widespread in renal and extrarenal tissues, causing a condition referred to as *oxalosis*.

GENETIC CONSIDERATIONS

The metabolic basis for the primary hyperoxalurias involves pathways of glyoxylate metabolism. In type I hyperoxaluria, urinary excretion of oxalate and of the oxidized and reduced forms of glyoxylate is increased. The excessive synthesis of these substances results from a block in glyoxylate metabolism. The primary defect in most patients is deficiency of the hepatic peroxisomal enzyme alanine:glyoxylate amino transferase. The gene (*AGXT*) for this enzyme maps to 2q36-37, and several distinct mutations have been defined in patients with type I hyperoxaluria. Some of these mutations misdirect the enzyme to mitochondria and render it nonfunctional.

In type II hyperoxaluria, L-glyceric acid is excreted in excess along with oxalate. In this condition, activity of D-glyceric acid dehydrogenase, which catalyzes the reduction of hydroxypyruvate to D-glyceric acid in the catabolic pathway of serine metabolism, is absent in leukocytes (and presumably other tissues). The accumulated hydroxypyruvate is instead reduced by lactic dehydrogenase to the L-isomer of glycerate, which is excreted in the urine. The same enzyme possesses glyoxylate reductase activity, and its deficiency promotes the oxidation of glyoxylate to oxalate, thus causing the formation of increased oxalate.

Stone formation, nephrocalcinosis, and oxalosis are due to insolubility of calcium oxalate. Extrarenal deposits of oxalate are prominent in the heart, walls of arteries and veins, male urogenital tract, and bone, particularly in type I hyperoxaluria.

Clinical Manifestations Nephrolithiasis and oxalosis may be manifest during the first year of life. Most patients experience renal colic or hematuria between ages 2 and 10 and succumb to uremia before age 20. With the onset of uremia, patients may develop severe peripheral arterial spasm and necrosis with resulting vascular insufficiency. Oxalate excretion falls as renal failure worsens. In patients with delayed onset of symptoms, survival to age 50 or 60 has been reported, despite recurrent nephrolithiasis. Type II hyperoxaluria is a milder disease with less involvement of extrarenal organs and delayed impairment of kidney function.

Diagnosis Oxalate excretion in normal children or adults is <0.5 mmol (60 mg) per 1.73

$m_2$surface area per day. Patients with type I or type II hyperoxaluria excrete two to four times this amount. Distinction between the two types depends on the identification of the other organic acids that identify them: glycolic acid in type I and L-glyceric acid in type II. Since patients with pyridoxine deficiency or chronic ileal disease may excrete excessive amounts of oxalate, these conditions must be excluded.

TREATMENT

There is no satisfactory treatment. Increasing the volume of urine can transiently reduce urinary oxalate concentration. Large doses of pyridoxine (100 mg/d) may reduce urinary oxalate in some patients, but long-term effects are not dramatic. A diet high in phosphate content seems to reduce the frequency of attacks of renal colic, but oxalate excretion is unaffected. Combined liver-kidney transplantation can correct the enzyme deficiency and replace the damaged organs. Liver transplantation seems promising in patients diagnosed before the onset of kidney failure.

ACKNOWLEDGEMENT
This chapter includes the contributions of Dr. Leon E. Rosenberg and Dr. Louis J. Elsas from previous editions.

(Bibliography omitted in Palm version)

353. INHERITED DEFECTS OF MEMBRANE TRANSPORT - *Nicola Longo*

Specific membrane transporters mediate the passage of a wide variety of substances across plasma cell membranes. Classes of substrates represented include amino acids, sugars, cations, anions, vitamins, and water. The disorders considered in this chapter have three features in common: each is characterized by a specific defect in the transport of one or more compounds; each is inherited as a dominant or recessive trait, implying that variation in a single genetic locus is involved; and each is presumed to reflect a primary alteration in a specific membrane protein. Many of these defects have been well characterized physiologically and genetically. Inherited defects impairing the transport of amino acids, hexoses, and chloride are discussed here as examples of the abnormalities encountered; others are considered elsewhere in this text.

The number of inherited disorders of membrane transport continues to increase with the identification of new transporters and the clarification of the molecular basis of diseases with previously unknown pathophysiology. The first transport disorders identified affected the gut or the kidney, but transport processes are now proving essential for the normal function of every organ. Mutations in transporter molecules have been demonstrated in disorders of the heart, muscle, brain, and endocrine and sensory organs (see examples in Table 353-1). In some cases, the same phenotype can be caused by mutations in different genes (*nonallelic heterogeneity*), often because they encode interacting proteins. In other cases, distinct mutations in the same gene (*allelic heterogeneity*) can cause different diseases depending on the degree of functional inactivation caused by the mutation, the presence of dominant-negative effects, or paradoxical activation of function.

DISORDERS OF AMINO ACID TRANSPORT

As listed in Table 353-1, 10 disorders of amino acid transport have been described. Five (cystinuria, dibasic aminoaciduria, Hartnup disease, iminoglycinuria, and dicarboxylic aminoaciduria) show transport abnormalities for structurally related amino acids, thereby implying the existence of group-specific membrane receptors or carriers. With the exception of iminoglycinuria and dicarboxylic aminoaciduria, the defects have important clinical consequences. The remaining five disorders affect the transport of only one amino acid, implying the existence of substrate-specific transport systems. Each of these conditions affects transport in the kidney, gut, or both; none has been shown to alter transport in other tissues.

CYSTINURIA

Definition Cystinuria, the most common inborn error of amino acid transport, is characterized by impaired renal tubular reabsorption and excessive urinary excretion of the dibasic amino acids lysine, arginine, ornithine, and cystine. A similar transport defect exists in the intestinal mucosa. Because cystine is the least soluble of the naturally occurring amino acids, its overexcretion predisposes to the formation of renal, ureteral, and bladder stones. Such stones are responsible for the signs and symptoms of the disorder.

GENETIC CONSIDERATIONS

Cystinuria is among the most common inborn errors, with a frequency of 1 in 10,000 to 1 in 15,000 in many ethnic groups. The disorder is transmitted as an autosomal recessive trait and results from impaired function of membrane carrier proteins in the apical brush border of proximal renal tubule and small intestinal cells.

There are three genetic variants of cystinuria. The urinary excretion patterns and renal clearance abnormalities in each type are similar in homozygotes, but the three variants are distinguished by studies of intestinal transport in homozygotes and urinary excretion patterns in heterozygotes. Type I homozygotes lack mediated intestinal transport of cystine, lysine, arginine, and ornithine; heterozygotes have normal urinary amino acid excretion patterns. Type II homozygotes lack mediated lysine transport in the gut but retain some capacity for cystine transport; heterozygotes have moderately increased urinary excretion of each of the four amino acids. Type III homozygotes retain some capacity for mediated intestinal transport of the four involved substrates; heterozygotes have modestly increased urinary lysine and cystine. The gene for type I cystinuria (*SLC3A1*) encodes solute carrier family 3 and maps to chromosome 2p16.3. The other two types of cystinuria (types II and III) are caused by mutations in *SLC7A9,* which maps to chromosome 19q13 and encodes the light subunit needed for the correct processing of SLC3A1.

Clinical Manifestations Massive excretion of cystine and the other disbasic amino acids occurs in homozygotes with classic cystinuria. Cystine stones account for 1 to 2% of all urinary tract calculi but are the most common cause of stones in children. The maximum solubility of cystine in the physiologic urinary pH range of 4.5 to 7.0 is about 1200 umol/L (300 mg/L). Since affected homozygotes regularly excrete 2400 to 7200 umol (600 to 1800 mg) daily, crystalluria and stone formation are a constant threat. Stone formation usually becomes manifest in the second or third decade but may occur in the first year of life or as bladder calculi at birth. Symptoms and signs are those typical of urolithiasis: hematuria, flank pain, renal colic, obstructive uropathy, and infection ([Chap. 279](#)). Recurrent urolithiasis may lead to progressive renal insufficiency.

Diagnosis The presence of cystine in a urinary tract stone is pathognomonic of cystinuria. However, because half the stones in cystinuric individuals are of mixed composition, and because the cystine core in as many as 10% may not be detected, a urinary nitroprusside test should be performed on all patients with urolithiasis to exclude this diagnosis. The nitroprusside test is also positive (appearance of a cherry red color) in some heterozygotes for cystinuria and in patients with hypercystinuria, homocystinuria, and mercaptolactate-cysteine disulfiduria. When cystine content exceeds 1000 umol/L (250 mg/L), cystine crystals may be seen in the sediment of acidified, concentrated, chilled urine. These hexagonal crystals are pathognomonic of cystine overexcretion in patients not taking sulfonamides.

Diagnostic confirmation of cystinuria depends on demonstration of the characteristic amino acid excretion pattern in the urine. Selective overexcretion of cystine, lysine, arginine, and ornithine can be demonstrated by qualitative and quantitative chromatography. Quantitation is important for differentiating heterozygotes from homozygotes and for following free cystine excretion during therapy.

TREATMENT

Management is aimed at preventing cystine crystal formation by reducing the concentration of cystine in urine. This aim is accomplished by increasing urinary volume and by maintaining an alkaline urine pH. Fluid ingestion in excess of 4 L/d is essential, and 5 to 7 L/d is optimal. Urinary cystine concentration should be<1000 to 1200 umol/L (250 to 300 mg/L). The daily fluid ingestion necessary to maintain this dilution of excreted cystine should be spaced over 24 h, with one-third of the total volume ingested between bedtime and 2 to 3 A.M. Stones can be prevented and even dissolved by such hydration. It must be made clear to individuals with cystinuria that water is a necessary drug for them. Solubility of cystine in urine rises sharply above pH 7.5, and urinary alkalinization can be therapeutic in some situations. Vigorous administration of sodium bicarbonate, acetazolamide, and polycitrates is required to maintain a persistently alkaline pH, but this measure introduces the danger of inducing formation of calcium oxalate, calcium phosphate, and magnesium ammonium phosphate stones and of producing nephrocalcinosis.

Another treatment involves administration of penicillamine, which undergoes sulfhydryl-disulfide exchange with cystine to form the mixed disulfide of penicillamine and cysteine. Since this disulfide is more than 50 times as soluble as cystine, penicillamine (1 to 3 g/d) reduces free cystine excretion markedly, thereby preventing new stone formation and promoting dissolution of existing calculi. Unfortunately, side effects include acute serum sickness, agranulocytosis, pancytopenia, immune glomerulitis, and the Goodpasture syndrome. Thus its use should be reserved for patients who fail to respond to hydration alone or who are in a high-risk category (one remaining kidney, renal insufficiency). Tiopronin (a-mercaptopropionylglycine, 800 to 1200 mg/d in four divided doses) has a mechanism of action similar to that of penicillamine, has lower toxicity, and is a suitable alternative. Captopril, a sulfhydryl-containing antihypertensive agent, has limited efficacy to reduce cystine excretion. When medical management fails, urologic surgery is required, but it should be a last resort as cystine stones reform more easily in scarred epithelium. Small (<1.5 cm) cystine stones can be treated with extracorporeal shock wave lithotripsy. Ureteroscopic removal is effective for ureteral stones, while larger or branched cystine stones require percutaneous nephrostolithotomy, sometimes associated with other procedures. All these procedures may produce smaller fragments, which can cause severe renal colic. Occasional patients progress to renal failure and require kidney transplantation.

DIBASIC AMINOACIDURIA

This disorder is characterized by a defect in renal tubular reabsorption of the three dibasic amino acids lysine, arginine, and ornithine but *not* cystine. There are two variants, transmitted as autosomal recessive traits. In the common form of dibasic aminoaciduria (type II), also known as *lysinuric protein intolerance*, homozygotes show defective intestinal transport of dibasic amino acids as well as exaggerated renal losses. It is most common in Finland (1 in 60,000) and is rare elsewhere. The transport defect affects basolateral rather than luminal membrane transport and is associated with impairment of the urea cycle. The defective gene (*SLC7A7*) in this condition maps to chromosome 14q11.2 and encodes a unique membrane transporter, y+LAT, that

associates with the cell-surface glycoprotein 4F2 heavy chain to form the complete sodium-independent transporter y_+L. The requirement for multiple gene products in the formation of this dimeric transporter probably explains part of the intrafamilial variability observed in lysinuric protein intolerance.

Manifestations are related to the losses of ornithine, arginine, and lysine. Affected patients present in childhood with hepatosplenomegaly, protein intolerance, and episodic ammonia intoxication. Older patients may present with severe osteoporosis, impairment of kidney function, or interstitial changes in the lungs. Plasma concentrations of lysine, arginine, and ornithine are reduced, whereas urinary excretion of lysine and orotic acid are increased. Hyperammonemia may develop after the ingestion of protein loads or with infections, probably due to insufficient amounts of arginine and ornithine to maintain proper function of the urea cycle. The clinical features have been attributed to the hyperammonemia and to insufficient amounts of lysine to support protein synthesis during growth.

Type I dibasic aminoaciduria has been described in a large French-Canadian kindred. Type I patients have profound mental retardation without hyperammonemia and protein intolerance. The condition also differs from type II by the presence of a modest excess of dibasic amino acids in the urine of asymptomatic heterozygotes. Type I disease may involve the same transport system as that impaired in the more common type II dibasic aminoaciduria.

TREATMENT

Dietary protein should be restricted in conjunction with supplementation of citrulline (2 to 8 g/d), a neutral amino acid that fuels the urea cycle when metabolized to arginine and ornithine. Carnitine supplements may improve growth by sparing lysine and by enhancing fatty acid oxidation. Pulmonary disease responds to glycocorticoids in some patients.

HARTNUP DISEASE

Pellagra-like skin lesions, variable neurologic manifestations, and neutral or aromatic aminoaciduria characterize this disease. Alanine, serine, threonine, valine, leucine, isoleucine, phenylalanine, tyrosine, tryptophan, glutamine, asparagine, and histidine are excreted in urine in quantities 5 to 10 times normal, and intestinal transport of these same amino acids is defective. The clinical manifestations result from nutritional deficiency of the essential amino acid tryptophan, caused by its intestinal and renal malabsorption. Manifestations are episodic, related in part to metabolic demands for tryptophan. Only a small fraction of patients with the chemical findings of this disorder develop a pellagra-like syndrome, implying that manifestations depend on other factors in addition to the transport defect.

Hartnup disease is inherited as an autosomal recessive trait, and the gene has been mapped to chromosome 11q13. Homozygotes occur with a frequency of about 1 in 24,000 births. Heterozygotes exhibit no clinical or chemical abnormalities. In patients with Hartnup disease, the renal and intestinal transport defect for tryptophan leads to niacin deficiency. Tryptophan metabolism leads to the synthesis of niacin and

nicotinamide-adenine dinucleotide and supplies about half the daily niacin needs. The transport defect likely reflects abnormalities of a group-specific system for neutral amino acids. Some residual reabsorptive capacity persists for each involved amino acid. This suggests that they are transported by other carrier systems as well, a conclusion supported by the identification of patients with substrate-specific transport defects for tryptophan, methionine, and histidine.

The diagnosis of Hartnup disease should be suspected in any patient with clinical features of pellagra without a history of dietary niacin deficiency (Chap. 75). The neurologic and psychiatric manifestations range from attacks of cerebellar ataxia to mild emotional lability to frank delirium and are usually accompanied by exacerbations of the erythematous, eczematoid skin rash. Fever, sunlight, stress, and sulfonamide therapy provoke clinical relapses. Diagnosis is made by detection of the neutral aminoaciduria, which does not occur in dietary niacin deficiency. Treatment is directed at niacin repletion and includes a high-protein diet and daily nicotinamide supplementation (50 to 250 mg). Tryptophan ethyl esters can also bypass the absorption defect.

IMINOGLYCINURIA

This benign autosomal recessive trait is characterized by excessive urinary excretion of glycine and the imino acids proline and hydroxyproline. Homozygotes occur with a frequency of about 1 in 16,000. The enhanced excretion of glycine, proline, and hydroxyproline reflects a defect in the tubular transport system shared by these three compounds. An intestinal transport defect may also be present. This suggests that more than one mutation may lead to iminoglycinuria, a thesis corroborated by the demonstration that obligate heterozygotes from some, but not all, families have glycinuria. No consistent clinical abnormalities have been reported in homozygotes that are usually detected by urinary amino acid screening programs.

DICARBOXYLIC AMINOACIDURIA

Selective urinary loss and exaggerated endogenous renal clearance of glutamic and aspartic acids have been described in two unrelated children. Intestinal absorption of these dicarboxylic amino acids was impaired in one. This patient suffered from recurrent hypoglycemia; the other was asymptomatic.

SUBSTRATE-SPECIFIC DEFECTS IN AMINO ACID TRANSPORT

Rare pedigrees exist in which individuals have defective renal tubular reabsorption and/or impaired intestinal absorption of a single free amino acid (Table 353-1). These disorders, each apparently inherited as an autosomal recessive trait, suggest that transport of amino acids is catalyzed by substrate-specific as well as group-specific transport mechanisms. Examples include hypercystinuria, lysinuria, histidinuria, and selective malabsorption of methionine or tryptophan.

DISORDERS OF HEXOSE TRANSPORT

D-Glucose is the major carbohydrate used by the cell for energy production and many other anabolic purposes. A number of transporter proteins work together to maintain

glucose homeostasis in the intact organism by coordinating absorption and utilization of D-glucose by all cells in the body. Two main classes of glucose transporters have been identified in humans: active Na_+-glucose cotransporters (SGLT) and the facilitative glucose transporters (GLUT). Na_+-glucose cotransporters actively concentrate glucose inside intestinal and renal cells using the electrochemical potential of Na_+ as their energy source. Defects in this class of transporters cause renal glycosuria (SGLT2) and intestinal glucose-galactose malabsorption (SGLT1). Facilitative glucose transporters allow glucose to enter cells using its own concentration gradient. This process is essential for the delivery of glucose to the cell for energy production. Defects in facilitative transporters cause the glucose-transporter protein syndrome (GLUT1) and the Fanconi-Bickel syndrome (GLUT2) and may be involved in one subtype of glycogen storage disorder (GLUT7).

DISORDERS OF CONCENTRATIVE GLUCOSE TRANSPORTERS

Renal glycosuria is characterized by the urinary excretion of glucose at normal concentrations of blood glucose. The renal tubular malabsorption is specific for glucose. Unlike generalized tubular dysfunction, other compounds such as phosphate and amino acids are transported normally. The genetic basis for this condition is not known at present. The condition is benign, but occasionally glycosuria may be severe enough to cause polyuria and polydipsia. Even more rarely, dehydration or ketosis may develop under conditions of stress such as pregnancy or starvation.

In normal persons, glucose is present in the glomerular filtrate at a concentration equal to that in plasma water and is reabsorbed throughout the proximal renal tubule by a sodium-dependent, phlorizin-inhibitable transport process. Reabsorptive capacity exceeds normal plasma glucose concentration. The plasma concentration at which filtered glucose begins to escape proximal tubular reabsorption is usually around 10 mmol/L (200 mg/dL). Maximal renal absorptive capacity is exceeded at a filtered load of around 2 mmol (325 mg)/min per 1.73 m_2 body surface area, and this value is defined as the tubular maximum for glucose (TmG).

Two patterns of glycosuria are recognized: type A, characterized by a reduced tubular maximum reabsorptive capacity, and type B, showing a reduced threshold for glycosuria, an increased "splay" in the titration curve, and a normalTmG. Renal glycosuria occurs in homozygotes with either of these recessively inherited mutations and in compound heterozygotes with these presumably allelic mutations. Modest reduction in renal threshold or TmG is present in heterozygotes in some families; modest glycosuria occurs in such family members when plasma glucose is elevated. In the few patients studied, renal glycosuria was not associated with impaired intestinal transport.

Glucose-galactose malabsorption is characterized by profuse, watery diarrhea in infants fed milk or foods containing lactose, sucrose, glucose, or galactose. The primary defect involves the sodium-hexose cotransporter in the intestinal and renal brush border. A specific defect in intestinal absorption of glucose and galactose can be demonstrated by oral tolerance tests that produce little or no increase in plasma glucose or galactose. Active D-glucose and D-galactose transport is absent in affected children, and intermediate transport capacity is present in their parents. Fructose-containing or

carbohydrate-free formulas are well tolerated. Treatment with a glucose- and galactose-free diet leads to resolution of symptoms in childhood. Although the basic transport defect is present throughout life, most patients show an improved tolerance for glucose and galactose with age.

A number of these patients have renal glycosuria at normal plasma glucose concentrations. Renal titration studies generally demonstrate a reduced threshold for glucose reabsorption (type B renal glycosuria) and a normal TmG. Urinary glucose loss is not as severe as in isolated renal glycosuria. This finding suggests the presence of multiple glucose transport proteins in the kidney. One, whose gene remains to be identified, is responsible for the bulk of glucose reabsorption in the proximal convoluted tubule and is believed to be abnormal in renal glycosuria. Another transporter, SGLT1, is shared by glucose and galactose and is responsible for the reabsorption of the least traces of glucose in the late proximal straight tubule. Its function is abnormal in glucose-galactose malabsorption, and heterogeneous mutations have been found in the *SGLT1* gene on chromosome 22q. In glucose-galactose malabsorption, as in renal glycosuria, transport of sugars in other tissues is normal, reflecting the multiplicity and tissue specificity of hexose transporters.

DISORDERS OF FACILITATIVE GLUCOSE TRANSPORTERS

At least five different facilitative glucose transporters (GLUT1, -2, -3, -4, and -7) mediate the influx and efflux of glucose in mammalian cells. Disease-causing mutations have been identified in two of these transporters (GLUT1 and GLUT2). The tissue specificity and redundancy of facilitative glucose transporters in different tissues helps to explain the clinical manifestations that result from their defective function. De novo mutations in the gene encoding the ubiquitous GLUT1 transporter cause the glucose-transporter protein syndrome. Patients present with seizures, developmental delay, and acquired microcephaly. These patients have normal blood glucose concentration but markedly decreased concentration of glucose in their cerebrospinal fluid. GLUT1 is the predominant glucose transporter in the blood-brain barrier. Haploinsufficiency reduces the transfer of glucose through the blood-brain barrier, restricting the energy supply to the brain. This defect is expressed in other cells, such as erythrocytes and fibroblasts, that have less stringent energy requirements or express additional glucose transporters, preventing cellular damage and clinical sequelae. Therapy in the glucose-transporter protein syndrome consists of using a ketogenic diet to deliver alternative fuels to the brain.

The GLUT2 transporter is expressed mainly in liver, pancreaticb cells, and in the basolateral membrane of gut and renal tubular cells. Truncating mutations in the *GLUT2* transporter gene have been identified in the autosomal recessive disorder Fanconi-Bickel syndrome. Patients present early in life with failure to thrive and polydipsia, with prominent glycosuria and aminoaciduria, rickets, fasting hypoglycemia with ketonuria, and prolonged postprandial hyperglycemia. Glycogen is accumulated in the liver and kidney, reflecting the inability to release glucose through the GLUT2 transporter. This results in fasting hypoglycemia and ketonuria and in generalized renal tubular dysfunction. The prolonged postprandial hyperglycemia is due to decreased sugar uptake by the liver and by the pancreaticb cell, the latter resulting in defective insulin synthesis and release. Affected patients do not develop diabetes because

human pancreatic b cells have alternative glucose transporters (GLUT1 and GLUT3) that can partially compensate for the absence of GLUT2 transporters. Therapy consists of symptomatic replacement of the renal losses of water and electrolytes, vitamin D replacement, and a diet plan consisting of frequent meals rich in complex carbohydrates to prevent hypoglycemia, analogous to the treatment of patients with glycogen storage diseases (Chap. 350).

DEFECTIVE ANION TRANSPORT: CHLORIDORRHEA

This rare, autosomal recessive disease results from impairment of active transport of chloride in the ileum and colon. Absence of chloride-bicarbonate ion exchange causes profound symptoms even before birth (polyhydramnios and absence of meconium). Massive watery diarrhea is apparent from the first days of life. This fluid loss, with its attendant impairment of electrolyte homeostasis, is life-threatening. A hypokalemic, hypochloremic, hyponatremic metabolic alkalosis develops with dehydration and secondary hyperaldosteronism. Fecal fluid contains an excess of chloride ion over the sum of the accompanying cations sodium and potassium. Fecal chloride concentration always exceeds 90 mmol/L when volume and serum electrolyte disturbances are corrected, and this chloridorrhea is diagnostic. Renal chloride transport is normal. Decreased urine chloride results from the kidney's attempts to conserve salt and water. The defective gene in this condition, called *DRA* (for downregulated in adenoma), maps to chromosome 7q and encodes an anion transporter, which is expressed only in the gastrointestinal tract. A deletion of the Val 317 codon in the *DRA* gene is responsible for the Finnish form of congenital chloride diarrhea.

Treatment requires adequate, lifelong repletion of electrolyte and fluid losses. Exact replacement of water, sodium chloride, and potassium chloride can prevent the growth and psychomotor retardation and the development of progressive renal damage. The renal lesion, with hyalinized glomeruli, juxtaglomerular hyperplasia, calcifications, and arteriolar changes, is probably a result of chronic volume depletion. Treatment of hyperreninemia and hypokalemia with prostaglandin inhibitors may reduce renal damage but does not alter intestinal symptoms or the need for chronic sodium chloride repletion. Omeprazole, while not decreasing the need for adequate oral replacement of electrolytes, may decrease stool output and improve the social life of patients.

(Bibliography omitted in Palm version)

354. THE LIPODYSTROPHIES AND OTHER PRIMARY DISORDERS OF ADIPOSE TISSUE - *Abhimanyu Garg*

The distribution and quantity of adipose tissue are controlled by multiple factors, including genetic background, diet, hormones, and exercise. The lipodystrophies are a heterogeneous group of adipose tissue disorders characterized by a selective loss of body fat (Tables 354-1 and 354-2). Patients with lipodystrophies have a propensity to develop insulin resistance, hypertriglyceridemia, diabetes mellitus, and fatty liver.

FAMILIAL OR GENETIC LIPODYSTROPHIES

CONGENITAL GENERALIZED LIPODYSTROPHY (BERARDINELLI-SEIP SYNDROME)

Clinical Features The primary diagnostic features of congenital generalized lipodystrophy (CGL) include a near-total lack of body fat and a marked muscular appearance from birth (Fig. 354-1*A*). On careful physical examination, however, fat can be detected in the palms and soles; with magnetic resonance imaging (MRI), normal amounts of fat can also be visualized in the orbits, scalp, perineum, and juxtaarticular and epidural regions (where the cushioning or protective functions of adipose tissue are critical). MRI studies also reveal a near-complete absence of metabolically active adipose tissue from most subcutaneous areas, intraabdominal and intrathoracic regions, and bone marrow.

Children exhibit accelerated linear growth and advanced bone age, but plasma levels of growth hormone or insulin-like growth factor I (IGF-I) are normal. Their basal metabolic rate is relatively high. It is unclear, however, whether hypermetabolism results from a primary increase in sympathetic nervous system activity or whether it is a compensatory response to protect against excessive heat loss due to extreme lack of body fat. Other features include acanthosis nigricans, prominent umbilicus or hernia, and an acromegalic appearance with coarse facial features and large hands and feet. Occasionally, excess body hair and hyperhidrosis have been noted. Fatty liver has been noted during infancy and can lead to cirrhosis and its complications. Liver, spleen, and kidney enlargement can cause abdominal protuberance. A few patients develop hypertrophic cardiomyopathy, but it rarely leads to heart failure.

Postpubertal women may have clitoromegaly, mild hirsutism, polycystic ovaries, and oligomenorrhea. Successful pregnancy in affected women is rare, whereas affected males have normal reproductive potential. Penile enlargement may be noted in childhood. After puberty, the skeleton appears sclerotic and focal lytic lesions develop in the appendicular bones. Some patients develop goiter.

Metabolic Abnormalities Patients with CGL have markedly elevated fasting serum insulin and C-peptide concentrations, as well as extreme insulin resistance. Diabetes mellitus appears during the pubertal years, and pancreatic pathology reveals severe amyloidosis of the pancreatic islets with loss of b cells. Plasma leptin concentrations are low, as expected in view of the reduced adipose tissue. Fasting plasma free fatty acid concentrations are normal. Hypertriglyceridemia may be observed during childhood, and patients can develop chylomicronemia, eruptive xanthomas, and acute pancreatitis. Low

concentrations of high-density lipoprotein (HDL) cholesterol are also common.

Course Patients with CGL are at risk of early mortality from cirrhosis and its complications, acute pancreatitis, or diabetic nephropathy. They also may develop diabetic retinopathy. Despite long-standing hypertriglyceridemia and hyperglycemia, atherosclerotic vascular complications are rare.

GENETIC CONSIDERATIONS

Approximately 120 patients of various ethnic backgrounds have been reported with this autosomal recessive form of lipodystrophy. The absence of fat could result from agenesis, failure of differentiation of preadipocytes, or an inability of mature adipocytes to synthesize and/or store triglycerides. Numerous candidate genes have been excluded, including the insulin receptor, b_3-adrenergic receptor, fatty acid binding protein 2, IGF-I receptor, insulin receptor substrate-1, hormone-sensitive lipase, leptin, and peroxisome proliferator-activated receptor g. Genome-wide linkage analysis of 17 families revealed genetic heterogeneity, but a candidate *CGL1* gene was localized to chromosome 9q34; the defective *CGL* gene(s) have yet to be identified.

FAMILIAL PARTIAL LIPODYSTROPHY

Dunnigan Variety

Clinical Features Patients with familial partial lipodystrophy, Dunnigan variety (FPLD) appear normal during childhood. During puberty, however, these patients begin to lose subcutaneous fat from the limbs and trunk, while exhibiting "increased muscularity" (Fig. 354-1 *B*). Many patients accumulate excess fat in the face and neck, often resulting in a double chin, supraclavicular humps, and round face. Occasionally, fat accumulates in the axillae. Labia majora appear prominent in women. MRI demonstrates excess fat inside the abdomen and in the intermuscular fasciae. Bone marrow and fat in certain mechanical locations, such as the orbits and joints, are normal. Acanthosis nigricans, hirsutism, menstrual abnormalities, and polycystic ovaries are infrequent. Hepatomegaly due to fatty liver is common, but progression to cirrhosis has not been reported.

Metabolic Abnormalities Patients with FPLD have mild to moderate insulin resistance, and diabetes mellitus develops, usually after the second decade. Patients have low serum HDL cholesterol and can develop severe hypertriglyceridemia. Fasting plasma free fatty acid concentrations may be elevated.

Course Major causes of morbidity and mortality in patients with FPLD include coronary heart disease, other atherosclerotic vascular complications, and acute pancreatitis.

GENETIC CONSIDERATIONS

This rare autosomal dominant disorder has been reported in 35 Caucasian families and one Indian family comprising approximately 200 affected individuals. The *FPLD* locus has been mapped to chromosome 1q21-22. Recently, several missense mutations in the gene encoding the nuclear envelope protein lamin A/C (*LMNA*) have been found to be responsible for FPLD. Alternative splicing of *LMNA* produces Lamins A and C,

members of the intermediate filament multigene family. All mutations causing typical FPLD cluster in exon 8 of *LMNA*, affecting the globular C-terminal tail of the Lamin A/C protein except one in exon 11 that only affects Lamin A and causes an atypical mild FPLD. The loss of subcutaneous adipose tissue in the limbs and trunk in FPLD may be due to adipocyte apoptosis and degeneration. Fat accumulation in the face and neck may be a secondary phenomenon, since it is not always present.

Kobberling Variety Characteristic features include loss of fat from the limbs with preservation of facial fat; truncal subcutaneous fat may be excessive. Most patients have hypertriglyceridemia and diabetes mellitus. Still unknown are the pattern of inheritance, age of onset, and whether this disorder is a distinct entity or a variant of the Dunnigan variety. Only a few women from two small pedigrees and four sporadic cases have been reported; it is not yet clear whether men can also be affected.

Mandibuloacral Dysplasia Variety This autosomal recessive disorder is characterized by short stature, high-pitched voice, mandibular and clavicular hypoplasia, dental abnormalities, acroosteolysis, stiff joints, and ectodermal defects. A few patients have also exhibited loss of limb fat. Insulin resistance and diabetes mellitus are rare.

OTHER TYPES

An autosomal dominant type of generalized lipodystrophy with acromegaloid features has been reported in a pedigree from Brazil. The onset of lipodystrophy occurred after 18 years of age. In another form of lipodystrophy, marked loss of subcutaneous fat from the limbs, face, palms, and soles, but excess subcutaneous fat in the neck and trunk has been noted.

ACQUIRED LIPODYSTROPHIES

ACQUIRED GENERALIZED LIPODYSTROPHY (LAWRENCE SYNDROME)

This form of lipodystrophy has been reported in approximately 50 patients and is characterized by a generalized disappearance of fat, mostly during childhood or adolescence. It is three times more common in females than males.

Clinical Features Fat loss affects the face, neck, trunk, and extremities and usually occurs over several months or years; superficial veins and muscles become prominent (Fig. 354-1*C*). Fat loss can include the palms and soles. In some, the onset of the disorder was reported after infections such as varicella, measles, pertussis, diphtheria, pneumonia, osteomyelitis, parotitis, infectious mononucleosis, or hepatitis. In others, lipodystrophy starts with painful, purple-brown subcutaneous nodules that leave depressed areas with loss of subcutaneous fat. Adipose tissue may show infiltration with lymphocytes, mononuclear macrophages, fat-cell necrosis, and fat-filled macrophages; this infiltration is consistent with a type of acute panniculitis. Almost one-third of these patients develop acanthosis nigricans, and some have mild hirsutism. Hepatomegaly due to fatty infiltration is a consistent finding and can lead to cirrhosis. Splenomegaly has also been reported.

Metabolic Abnormalities Ketosis-resistant diabetes mellitus usually occurs after the

onset of lipodystrophy, and metabolic abnormalities are similar to those in CGL. Severely hyperglycemic patients may have elevated plasma free fatty acids.

Pathogenesis It is not yet known whether preceding infections play a causal role in this disorder. Some patients reportedly develop autoimmune diseases, including childhood dermatomyositis, juvenile rheumatoid arthritis, Hashimoto's thyroiditis, vitiligo, hemolytic anemia, or chronic active hepatitis. Autoantibodies against adipocyte membranes have been reported; it seems likely that antibody- and/or cell-mediated adipocyte lysis causes fat loss in these patients.

ACQUIRED PARTIAL LIPODYSTROPHY (BARRAQUER-SIMONS SYNDROME)

This form of lipodystrophy affects females three times more often than males and has been reported in about 200 patients. The onset usually occurs during childhood or adolescence. Fat loss typically affects the face, neck, upper limbs, thorax, and upper abdomen, and there is increased fat deposition in the hips and lower extremities (Fig. 354-1*D*).

Clinical Features Fat loss occurs gradually over 1 to 2 years and initially affects the face; other areas are affected later. In most cases, the lower abdomen, hips, and lower extremities are spared. In general, patients do not develop insulin resistance and other metabolic abnormalities, acanthosis nigricans, hirsutism, or menstrual problems. Approximately one-third of patients develop mesangiocapillary glomerulonephritis, usually 10 years after disease onset. Systemic lupus erythematosus and other autoimmune diseases have also been reported, including childhood dermatomyositis, thyroiditis, pernicious anemia, celiac disease, dermatitis herpetiformis, rheumatoid arthritis, Sjogren's syndrome, temporal arteritis, and leukocytoclastic vasculitis. In addition, many patients have serum antinuclear and anti-double-stranded DNA antibodies.

Pathogenesis C3 nephritic factor (C3NeF), a polyclonal IgG immunoglobulin, can be detected in the serum of up to 90% of these patients. Serum C3 is universally low, but C1q, C4, C5, C6, factor B, and properdin concentrations are usually normal (which suggests activation of the alternative complement pathway). Loss of fat may be due to C3NeF-induced lysis of adipocytes that express factor D. C3NeF also binds and inactivates factor H, which can induce glomerulonephritis by mechanisms similar to those seen in genetic factor H deficiency.

HIV-1 PROTEASE INHIBITOR-INDUCED LIPODYSTROPHY

Highly active antiretroviral therapy (HAART) therapy for HIV, a combination which includes HIV-1 protease inhibitors, is associated with the development of lipodystrophy in the majority of patients after 18 months to 2 years of treatment (Chap. 309). It is characterized by marked reduction in subcutaneous fat from the face, trunk, and limbs, resulting in an appearance of "increased muscularity." Excess fat may also accumulate around the neck (double chin and buffalo hump) and inside the abdomen. Patients are prone to develop insulin resistance, diabetes mellitus, and hypertriglyceridemia. It is unclear whether this disorder is caused by a side effect of one or more of the drugs (most likely a protease inhibitor) or by a metabolic response to dramatic reduction of

viral load. Hormonal causes, such as hypercortisolism, have been excluded.

LOCALIZED LIPODYSTROPHIES

These disorders are characterized by a loss of subcutaneous adipose tissue from small areas or parts of a limb. Fat loss may occur secondary to injections of insulin, glucocorticoids, antibiotics, iron dextrans, or diphtheria/pertussis/tetanus vaccine. Repeated pressure against any body part, such as the thigh or chin, can cause lipodystrophy. In some patients, acute panniculitis causes localized lipodystrophy without progressing further. *Centrifugal lipodystrophy* begins in the abdomen, groin, and axillae of children under the age of 3, and eventually spreads to involve the entire abdomen. The surrounding areas show slightly erythematous and scaly changes with an accumulation of lymphocytes and histiocytes on histology. Complete or partial improvement occurs spontaneously after 8 to 10 years.

TREATMENT

Patients with lipodystrophies have cosmetic problems that warrant judicious treatment. Facial reconstruction can be accomplished with free flaps, transposition of facial muscle, and silicone or other implants. In acquired partial lipodystrophy, adipose tissue transplantation from the thigh to face lasts for only 2 to 5 years. In FPLD, excess fat in the face and neck may require liposuction or lipectomy. Etretinate and fish oil have improved acanthosis nigricans in some patients with generalized lipodystrophy.

Dietary fat should be restricted for patients with severe hypertriglyceridemia. Reduced energy intake and increased physical activity can mitigate insulin resistance. In children, however, enough energy should be provided to allow for normal growth and development. Medium chain triglycerides have been reported to benefit some patients with acquired generalized lipodystrophy.

In patients with CGL, diabetes control may require extremely high doses of insulin. Oral hypoglycemic agents may also be used (Chap. 333). Glycemic control can mitigate dyslipidemia and prevent diabetic complications. Severe hypertriglyceridemia should be treated with fibrates and/or omega-3 polyunsaturated fatty acids. Niacin worsens glycemic control and should not be used. Estrogens should be avoided because they may accentuate hypertriglyceridemia and can cause acute pancreatitis.

LIPOMATOSIS

MULTIPLE SYMMETRIC LIPOMATOSIS (MADELUNG DISEASE)

This type of lipomatosis affects men 4 to 15 times more frequently than women. It is characterized by a symmetric, progressive growth of nonencapsulated subcutaneous adipose tissue, primarily in the neck (bull neck with buffalo hump and double chin) and supraclavicular and shoulder regions. Fat may also accumulate in the trunk and proximal limbs, though the distal arms and legs are spared. Rarely, laryngeal, tracheal, or vena caval compression may occur from deep lipomatous infiltration in the neck and mediastinum. Many patients also have peripheral neuropathy; hypertriglyceridemia and hyperuricemia are uncommon. Serum HDL cholesterol levels are usually elevated, and

diabetes mellitus has not been reported.

Most of these patients have a preceding history of heavy ethanol intake. The underlying mechanisms and predisposing factors for the disorder, however, remain unknown. The lipomatous tissue contains small adipocytes (but not brown fat) with increased lipoprotein lipase activity and reduced catecholamine-stimulated lipolysis. In several families and in some sporadic cases, mitochondrial DNA mutations A-to-G and G-to-A transitions at nucleotides 8344 and 8363, respectively, or deletions in the tRNA$_{Lys}$gene have been reported. Most of these patients had multiple, discrete, and encapsulated lipomas in the neck and trunk, which is distinct from the features of typical patients with multiple symmetric lipomatosis. These patients also have associated peripheral neuropathy, myopathy, cerebellar ataxia, myoclonus, or hearing loss. Mitochondrial DNA mutations have not been found in many patients with typical multiple symmetric lipomatosis.

Surgical resection may be required to relieve compression or for cosmetic reasons. Cessation of alcohol intake does not result in regression but may slow growth rate.

OTHER FORMS OF LIPOMATOSIS

Mediastinal lipomatosis is characterized by local overgrowth of adipose tissue in the mediastinum. It occurs in patients with Cushing's syndrome and can occasionally cause tracheal compression.

Pelvic lipomatosis is characterized by overgrowth of pelvic fat, causing bladder dysfunction (frequency, dysuria, and nocturia), constipation, and lower abdominal pain. Bilateral ureteral obstruction may also occur. The male:female ratio is 18:1. The etiology is not known, but the condition may result from a localized manifestation of obesity. Surgery may be needed to relieve urinary tract obstruction.

Epidural lipomatosis occurs in obese patients or in those receiving exogenous steroid therapy. Fat deposition most often occurs in the thoracic or lumbar spine, causing back pain, radicular pain, or spinal cord compression. Laminectomy may be indicated for cord compression. Weight loss or discontinuation of steroid therapy may also be helpful.

ADIPOSIS DOLOROSA (DERCUM DISEASE)

This is a rare disease of unknown etiology that mainly affects obese postmenopausal women (female:male ratio, 30:1). It is characterized by the presence of multiple circumscribed or diffuse painful subcutaneous fat deposits on the trunk and limbs, particularly near the knees. Patients also report weakness, fatigue, and emotional lability. Relief of pain is difficult; intravenous lidocaine, glucocorticoids, surgical excision, and liposuction are sometimes helpful.

ACUTE PANNICULITIS

A variety of systemic diseases including collagen vascular diseases such as systemic lupus erythematosus and scleroderma are associated with *acute panniculitis*, or *nodular fat necrosis* (Chap. 311). Panniculitis may also occur as a manifestation of

lymphoproliferative disorders ([Chap. 112](#)).

Disseminated fat necrosis is usually associated with acute pancreatitis or pancreatic carcinoma. It may be caused by the release of pancreatic enzymes into the circulation ([Chap. 304](#)).

HORMONAL EFFECTS ON ADIPOSE DISTRIBUTION

A variety of hormones influence the distribution of adipose tissue. Growth hormone, for example, reduces truncal fat but can increase fat in the palms and soles. Insulin enhances lipogenesis and fat storage. Thyroid hormones increase metabolic rate, including energy expenditure by fat tissue. Estrogens induce fat accumulation in the hips, legs, breasts and other subcutaneous regions. Glucocorticoids redistribute adipose tissue from peripheral to central locations. In Cushing's syndrome, characteristic features include buffalo hump, increased supraclavicular and truncal fat.

(Bibliography omitted in Palm version)

PART FOURTEEN - NEUROLOGIC DISORDERS

SECTION 1 - DIAGNOSIS OF NEUROLOGIC DISORDERS

355. NEUROBIOLOGY OF DISEASE - *Stephen L. Hauser, M. Flint Beal*

The human nervous system is the organ of consciousness, cognition, ethics, and behavior; as such, it is the most intricate structure known to exist. One-third of the 100,000 genes encoded in the human genome is expressed in the nervous system. Each mature brain is composed of 100 billion neurons, several million miles of axons and dendrites, and more than 10^{15} synapses. Neurons exist within a dense parenchyma of multifunctional glial cells that synthesize myelin, preserve homeostasis, and regulate immune responses. Measured against this background of complexity, the achievements of molecular neuroscience have been extraordinary. Advances in cell biology and genetics have provided new tools to explore the pathophysiology of nervous system diseases, clarifying their underlying causes, revealing new unanticipated groupings, and raising realistic hope that novel therapies and prevention strategies will be possible. This chapter reviews selected themes in neurobiology that provide a context for understanding fundamental mechanisms underlying neurologic disorders. *The reader is also referred to related discussions of neurogenetic disorders (Chap. 359) and the neurobiology of addiction (Chap. 386), and to the individual chapters on specific disorders.*

ION CHANNELS AND CHANNELOPATHIES

The resting potential of neurons and the action potentials responsible for impulse conduction are generated by ion currents and ion channels. Most ion channels are gated, meaning that they can transition between conformations that are open or closed to ion conductance. Individual ion channels are distinguished by the specific ions they conduct; by their kinetics; and by whether they directly sense voltage, are linked to receptors for neurotransmitters or other ligands such as neurotrophins, or are activated by second messengers. The diverse characteristics of different ion channels provide a means by which neuronal excitability can be exquisitely modulated at both the cellular and the subcellular levels. Mutations in ion channels -- channelopathies -- are responsible for a growing list of human neurologic disorders (Table 355-1). One example is epilepsy, a syndrome of diverse causes characterized by repetitive, synchronous firing of neuronal action potentials. Action potentials are normally generated by the opening of sodium channels and the inward movement of sodium ions down the intracellular concentration gradient. Depolarization of the neuronal membrane opens potassium channels, resulting in outward movement of potassium ions, repolarization, closure of the sodium channel, and hyperpolarization. Sodium or potassium channel subunit genes have long been considered candidate disease genes in inherited epilepsy syndromes, and recently such mutations have been identified (Chap. 360). These mutations appear to alter the normal gating function of these channels, increasing the inherent excitability of neuronal membranes in regions where the abnormal channels are expressed.

Whereas the specific clinical manifestations of channelopathies are quite variable, one common feature is that manifestations tend to be intermittent or paroxysmal, such as

occurs in epilepsy, migraine, ataxia, myotonia, or periodic paralysis. Exceptions are clinically progressive channel disorders such as spinocerebellar ataxia type 6 (SCA6) and autosomal dominant hearing impairment. The neurologic channelopathies identified to date are all uncommon disorders caused by obvious mutations in channel genes. As the full repertoire of human ion channels and related proteins are identified, it is likely that additional channelopathies will be discovered. In addition to rare disorders that result from obvious mutations, it is possible that subtle allelic variations in channel genes or in their pattern of expression might underlie susceptibility to some common forms of epilepsy, migraine, or other disorders.

NEUROTRANSMITTERS AND NEUROTRANSMITTER RECEPTORS

Synaptic neurotransmission is the predominant means by which neurons communicate with each other. Classic neurotransmitters are synthesized in the presynaptic region of the nerve terminal; stored in vesicles; and released into the synaptic cleft, where they bind to receptors on the postsynaptic cell. Secreted neurotransmitters are eliminated by reuptake into the presynaptic neuron (or glia), by diffusion away from the synaptic cleft, and/or by specific inactivation. In addition to the classic neurotransmitters, many neuropeptides have been identified as definite or probable neurotransmitters; these include substance P, neurotensin, enkephalins, b-endorphin, histamine, vasoactive intestinal polypeptide, cholecystokinin, neuropeptide Y, and somatostatin. Peptide neurotransmitters are synthesized in the cell body rather than the nerve terminal and may colocalize with classic neurotransmitters in single neurons. Nitric oxide and carbon monoxide are gases that appear also to function as neurotransmitters, in part by signaling in a retrograde fashion from the postsynaptic to the presynaptic cell.

Neurotransmitters modulate the function of postsynaptic cells by binding to specific neurotransmitter receptors, of which there are two major types. *Ionotropic receptors* are direct ion channels that open after engagement by the neurotransmitter. *Metabotropic receptors* interact with G proteins, stimulating production of second messengers and activating protein kinases, which modulate a variety of cellular events. Ionotropic receptors are multiple subunit structures, whereas metabotropic receptors are composed of single subunits only. One important difference between ionotropic and metabotropic receptors is that the kinetics of ionotropic receptor effects are fast (generally less than a millisecond) because neurotransmitter binding directly alters the electrical properties of the postsynaptic cell, whereas metabotropic receptors function over longer time periods. These different properties contribute to the potential for selective and finely modulated signaling by neurotransmitters.

Individual neurotransmitter systems are perturbed in a large number of clinical disorders, examples of which are highlighted in Table 355-2. One example is the involvement of dopaminergic neurons originating in the substantia nigra of the midbrain and projecting to the striatum (nigrostriatal pathway) in Parkinson's disease and in heroin addicts after the ingestion of the toxin MPTP (1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine) (Chap. 363). A second important dopaminergic system arising in the subtantia nigra is the mesolimbic pathway, which influences behavior and appears to be important in the pathogenesis of addiction. Addictive drugs share the property of increasing dopamine release, and blockade of dopamine in the nucleus accumbens (a part of the mesolimbic pathway) terminates the

rewarding effects of addictive drugs (Chap. 386).

CELL TO CELL COMMUNICATION THROUGH GAP JUNCTIONS

Not all cell-to-cell communication in the nervous system occurs via neurotransmission. Gap junctions provide for direct neuron-neuron electrical conduction and also create openings for the diffusion of ions and metabolites between cells. In addition to neurons, gap junctions are also widespread in glia, creating a syncytium that protects neurons by removing glutamate and potassium from the extracellular environment. Gap junctions consist of membrane-spanning proteins termed connexins that pair across adjacent cells. Mechanisms that involve gap junctions have been related to a variety of neurologic disorders. Mutations in connexin 32, a gap junction protein expressed by Schwann cells, are responsible for the X-linked form of Charcot-Marie-Tooth disease (Chap. 379). Mutations in either of two gap junction proteins expressed in the inner ear -- connexin 26 and connexin 31 -- result in autosomal dominant progressive hearing loss (Chap. 29). Glial calcium waves mediated through gap junctions also appear to explain the phenomenon of spreading depression associated with migraine auras and the march of epileptic discharges. Spreading depression is a neural response that follows a variety of different stimuli and is characterized by a circumferentially expanding negative potential that propagates at a characteristic speed of 20 um/s and is associated with an increase in extracellular potassium.

SIGNALING PATHWAYS AND GENE TRANSCRIPTION

The fundamental issue of how memory, learning, and thinking are encoded in the nervous system is likely to be clarified by identifying the signaling pathways involved in neuronal differentiation, axon guidance, and synapse formation, and by understanding how these pathways are modulated by experience. Many families of transcription factors, each comprising multiple individual components, are expressed in the nervous system. Elucidation of these signaling pathways has already begun to provide insights into the cause of a variety of neurologic disorders, including inherited disorders of cognition such as X-linked mental retardation. This syndrome affects approximately 1 in 500 males, and linkage studies in different families suggest that as many as 60 different X-chromosome encoded genes may be responsible. A number of disease genes have now been identified. Three encode proteins that regulate members of the ras family of GTP-binding proteins thought to have roles in regulation of the actin cytoskeleton and in neurite outgrowth (*OPHN1*, *PAK3*) or synaptic vesicle transport and neurotransmitter release (*GDI1*); one (*IL1RAPL*) has homology to an interleukin (IL)1 receptor accessory protein involved in IL-1 signaling; and one (*FMR2*) functions as a nuclear transcriptional regulatory protein. Rett syndrome, a common cause of (dominant) X-linked progressive mental retardation in females, is also due to a mutation in a gene (*MECP2*) encoding a DNA-binding protein involved in transcriptional repression. As the X chromosome comprises only approximately 3% of germline DNA, then by extrapolation the number of genes that potentially contribute to clinical disorders affecting intelligence in humans must be potentially very large.

MYELIN

Myelin is the multilayered insulating substance that surrounds axons and speeds

impulse conduction by permitting action potentials to jump between naked regions of axons (nodes of Ranvier) and across myelinated segments. A single oligodendrocyte usually ensheaths multiple axons in the central nervous system (CNS), whereas in the peripheral nervous system (PNS) each Schwann cell typically myelinates a single axon. Myelin is a lipid-rich material formed by a spiraling process of the membrane of the myelinating cell around the axon, creating multiple membrane bilayers that are tightly apposed (compact myelin) by charged protein interactions. A number of clinically important neurologic disorders are caused by inherited mutations in myelin proteins of the CNS or PNS. Constituents of myelin also have a propensity to be targeted as autoantigens in autoimmune demyelinating disorders (Fig. 355-1).

NEUROTROPHIC FACTORS

Neurotrophic factors (Table 355-3) are secreted proteins that modulate neuronal growth, differentiation, repair, and survival; some have additional functions, including roles in neurotransmission and in the synaptic reorganization involved in learning and memory. Because of their survival promoting and anti-apoptotic effects, neurotrophic factors are in theory outstanding candidates for therapy of disorders characterized by premature death of neurons such as occurs in amyotrophic lateral sclerosis (ALS) and other degenerative motor neuron disorders. Knockout mice lacking receptors for ciliary neurotrophic factor (CNTF) receptor or brain-derived neurotrophic factor (BDNF) show loss of motor neurons, and experimental motor neuron death can be rescued by treatment with various neurotrophic factors including CNTF and BDNF. However, in phase 3 clinical trials both CNTF and BDNF were ineffective in human ALS, and two other trials of insulin-like growth factor 1 yielded conflicting results with little evidence of clinically significant efficacy. Current understanding of the redundancy and diversity of neurotrophic factor activities at different stages in the life and health of individual neurons is extremely limited, and data obtained in rodent systems are not always applicable to humans. For example, CNTF knockout mice show a partial loss of motoneurons, yet humans who have homozygous mutations that inactivate the gene for CNTF gene are asymptomatic.

STEM CELLS AND TRANSPLANTATION

The nervous system is traditionally considered to be a nonmitotic organ, in particular with respect to neurons. These concepts have been challenged by the finding that neural progenitor or stem cells exist in the adultCNS that are capable of differentiation, migration over long distances, and extensive axonal arborization and synapse formation with appropriate targets. These capabilities also indicate that the repertoire of factors required for growth, survival, differentiation, and migration of these cells exist in the mature nervous system. The poor outcome associated with many neurologic disorders, however, clearly indicates that any potential for functional neuronal reconstitution after injury must be extremely limited in most clinical contexts. In rodents, neural stem cells, defined as progenitor cells capable of differentiating into mature cells of neural or glial lineage, have been experimentally propagated from fetal CNS and neuroectodermal tissues, and also from adult germinal matrix and ependyma regions. Human fetal CNS tissue is also capable of differentiation into cells with neuronal, astrocyte, and oligodendrocyte morphology when cultured in the presence of particular growth factors. Impressively, such cells could be stably engrafted into mouse CNS tissue, creating

neural chimeras. Once the repertoire of signals required for cell type specification are better understood, differentiation into specific neural or glial subpopulations can be directed in vitro; such cells could also be engineered to express therapeutic molecules.

Experimental transplantation of human fetal dopaminergic neurons in patients with Parkinson's disease has shown that these transplanted cells can survive within the host striatum. Studies of transplantation for patients with Huntington's disease have also reported encouraging, although very preliminary, results. Oligodendrocyte precurser cells transplanted into mice with a dysmyelinating disorder effectively migrated in the new environment, interacted with axons, and mediated myelination; such experiments raise hope that similar transplantation strategies may be feasible in human disorders of myelin such as multiple sclerosis. Enthusiasm for transplantation therapy must be tempered by unresolved concerns over safety (including the theoretical risk of malignant transformation of transplanted cells), ethics (particularly with respect to use of fetal tissue), and efficacy.

CELL DEATH -- EXCITOTOXICITY AND APOPTOSIS

Excitotoxicity refers to neuronal cell death caused by activation of excitatory amino acid receptors (Fig. 355-2). Compelling evidence for a role of excitotoxicity, especially in ischemic neuronal injury, is derived from experiments in animal models. Experimental models of stroke are associated with increased extracellular concentrations of the excitatory amino acid neurotransmitter glutamate, and neuronal damage is attenuated by denervation of glutamine-containing neurons or the administration of glutamate receptor antagonists. The distribution of cells sensitive to ischemia corresponds closely with that of *N*-methyl-D-aspartate (NMDA) receptors (except for cerebellar Purkinje cells, which are vulnerable to hypoxia-ischemia but lack NMDA receptors); and competitive and noncompetitive NMDA antagonists are effective in preventing focal ischemia. In global cerebral ischemia, non-NMDA receptors (kainic acid and AMPA) are activated, and antagonists to these receptors are protective. Experimental brain damage induced by hypoglycemia is also attenuated by NMDA antagonists.

Excitotoxicity is not a single event but rather a cascade of cell injury. Excitotoxicity causes influx of calcium into cells and much of the calcium is sequestered in mitochondria rather than in the cytoplasm. Increased mitochondrial calcium causes metabolic dysfunction and free radical generation; activates protein kinases, phospholipases, nitric oxide synthase, proteases, and endonucleases; and inhibits protein synthesis. Activation of nitric oxide synthase generates nitric oxide (NO_x), which can react with superoxide (O_{x2}) to generate peroxynitrite ($ONOO_-$), which may play a direct role in neuronal injury. Another critical pathway is activation of poly-ADP-ribose polymerase, which occurs in response to free radical-mediated DNA damage. Experimentally, mice with knockout mutations of neuronal nitric oxide synthase or poly-ADP-ribose polymerase, or those that overexpress superoxide dismutase, are resistant to focal ischemia.

Apoptosis, or programmed cell death, plays an important role in both physiologic and pathologic conditions. During embryogenesis, apoptotic pathways operate to destroy neurons that fail to differentiate appropriately or reach their intended targets. There is mounting evidence for an increased rate of apoptotic cell death in a variety of acute and

chronic neurologic diseases. Apoptosis is characterized by neuronal shrinkage, chromatin condensation, and DNA fragmentation, whereas necrotic cell death is associated with cytoplasmic and mitochondrial swelling followed by dissolution of the cell membrane. Apoptotic and necrotic cell death can coexist or be sequential events depending on the severity of the initiating insult. Cellular energy reserves appear to have an important role in these two forms of cell death, with apoptosis favored under conditions in which ATP levels are preserved. Evidence of DNA fragmentation has been found in a number of degenerative neurologic disorders, including Alzheimer's disease, Huntington's disease, and ALS. The best characterized genetic neurologic disorder related to apoptosis is infantile spinal muscular atrophy (Werdnig-Hoffmann disease), in which two genes thought to be involved in the apoptosis pathways are causative.

Mitochondria are essential in controlling specific apoptosis pathways. The redistribution of cytochrome c from mitochondria during apoptosis leads to the activation of a cascade of intracellular proteases known as caspases. Redistribution of cytochrome c is prevented by overproduction of the apoptotic protein BCL2 and is promoted by the proapoptotic protein BAX. These pathways may be triggered by activation of a large pore in the mitochondrial inner membrane known as the permeability transition pore. Recent studies suggest that blocking this pore reduces both hypoglycemic and ischemic cell death.

PROTEIN AGGREGATION AND NEURODEGENERATION

The possibility that protein aggregation plays a role in the pathogenesis of neurodegenerative diseases is a major focus of current research. Protein aggregation is a major histopathologic hallmark of neurodegenerative diseases. Deposition of b-amyloid is strongly implicated in the pathogenesis of Alzheimer's disease. Genetic mutations in familial Alzheimer's disease produce increased amounts of b-amyloid with 42 amino acids, which has an increased propensity to aggregate, as compared to b-amyloid with 40 amino acids. Mutations in genes encoding the microtubule associated protein tau lead to altered splicing of tau and the production of neurofibrillary tangles in frontotemporal dementia and progressive supranuclear palsy. Familial Parkinson's disease is associated with mutations in a-synuclein and the ubiquitin carboxy-terminal hydrolase. The characteristic histopathologic feature of Parkinson's disease is the Lewy body, an eosinophilic cytoplasmic inclusion that contains both neurofilaments and a-synuclein. Huntington's disease and cerebellar degenerations are associated with expansions of polyglutamine repeats in proteins, which aggregate to produce neuronal intranuclear inclusions. Familial ALS is associated with superoxide dismutase mutations and cytoplasmic inclusions containing superoxide dismutase. In autosomal dominant neurohypophyseal diabetes insipidus, mutations in vasopressin result in abnormal protein processing, accumulation in the endoplasmic reticulum, and cell death (Chap. 329).

The major scientific question presently is whether protein aggregates contribute to neuronal death or whether they are merely a secondary bystander. Protein aggregates are usually ubiquinated, which targets them for degradation by the 26S component of the proteosome. An inability to degrade protein aggregates could lead to cellular dysfunction, impaired axonal transport, and cell death by apoptotic mechanisms.

In experimental models of Huntington's disease and cerebellar degeneration, protein aggregates are not well correlated with neuronal death. A number of compounds have been developed to block b-amyloid production and/or aggregation, and these agents are being studied in early clinical trials in humans.

NEUROIMMUNOLOGY

The nervous system is traditionally considered to be an immunologically privileged organ, a concept originally derived from observations that tissue grafts implanted in the brain were not rejected efficiently. In this context, immune privilege of the CNS may be maintained by a variety of mechanisms including: the lack of an efficient surveillance function by T cells; the absence of a traditional lymphoid system; limited expression of major histocompatibility complex (MHC) molecules required for T cell recognition of antigen; effects of regulatory cytokines secreted spontaneously or in response to mediators such as nerve growth factor (NGF), creating an immunosuppressive milieu; and also from expression of fas ligand that can induce apoptosis of fas-expressing immune cells that enter the brain. The blood-brain barrier (BBB) partially isolates the brain from the peripheral environment and contributes to immune privilege. Anatomically, the barrier is created by the presence of impermeable tight junctions between endothelial cells, and by a relative absence of transendothelial conduits for the passive diffusion of soluble molecules. The BBB serves to preserveCNShomeostasis by excluding neuroactive substances present in the serum, such as neurotransmitters and neurotrophic factors. Because of the BBB, lipid insoluble molecules must utilize either ion channels or specific transport systems (for glucose or various amino acids) to gain entry to the CNS. Astrocyte foot processes that encircle the subendothelial basal surface of small blood vessels in the brain contribute to development and maintenance of the BBB.

The concept of immune privilege is at odds with clinical experience that vigorous immune reactions readily occur in the nervous system in response to infections and that autoimmune diseases of the nervous system are relatively common. Although primary (sensitizing) immune responses are not easily generated in theCNS for the reasons outlined above, this is not the case for secondary immune responses. When sensitization to nervous system antigens occurs *outside* the nervous system (e.g., in a regional lymph node), activated autoreactive T lymphocytes are easily generated, and these cells readily cross theBBB and induce immune mediated injury. The paradigm for this mechanism of T cell-mediated CNS disease is experimental allergic encephalomyelitis (EAE), a laboratory model for the human autoimmune demyelinating disorders multiple sclerosis (MS) and acute disseminated encephalomyelitis; the sequence of events in EAE is illustrated in Fig. 355-3.

Under normal circumstances theBBB is impermeable to antibodies. For autoantibodies to reach theCNS, the BBB must first be disrupted. In inflammatory conditions it is thought that this disruption most often occurs via actions of proinflammatory cytokines elaborated within the brain consequent to interactions between pathogenic T cells and antigen-presenting cells (APCs). In contrast to the BBB, in thePNS the blood-nerve barrier is incomplete. Endothelial tight junctions are lacking, and the capacity of charged molecules, including antibodies, to cross the barrier appears to be greatest in two regions of the PNS: proximally in the spinal roots and distally at neuromuscular

junctions. This anatomic feature is likely to contribute to the propensity of antibody-mediated autoimmune disorders of the PNS to target proximal nerves (Guillain-Barre syndrome) or the neuromuscular junction (myasthenia gravis, Eaton-Lambert syndrome).

The major APCs in the CNS are microglial cells and macrophages; both cell types express MHC class 2 molecules as well as costimulatory molecules required for antigen presentation. Neurons do not express MHC class 2 molecules; however, some neurons express MHC class 1 proteins, which may be further increased in response to neuronal activity. Neuronal MHC class 1 molecules may function as retrograde postsynaptic signaling molecules that interact with presynaptic CD3z molecules to stabilize active synapses and transynaptically modulate neuronal function. Studies in mice also indicate that MHC class 1 molecules influence the mating behavior of females; a hierarchichal pattern of preference is determined by the specific class 1 alleles expressed by potential male suitors. This behavior appears to be mediated by distinctive odors imparted either by the class 1 molecules themselves or by other families of molecules controlled by class 1 alleles. Thus, it appears likely that MHC molecules subserve a variety of signaling and adhesion functions that influence nervous system function far beyond their well-established roles as mediators of APC-T lymphocyte interactions.

(Bibliography omitted in Palm version)

356. APPROACH TO THE PATIENT WITH NEUROLOGIC DISEASE- *Joseph B. Martin, Stephen L. Hauser*

Neurologic disorders are common and costly. According to one recent estimate, 180 million Americans suffer from a nervous system disorder, resulting in annual cost of 634 billion dollars (Table 356-1). Most patients with neurologic symptoms seek care from internists and other generalists rather than from neurologists, and this situation is likely to continue as primary care-based health care systems become increasingly prevalent and access to specialists is reduced. Because useful therapies now exist for many neurologic disorders, a skillfull approach to their diagnosis is important. Many errors result from an over-reliance on neuroimaging and other laboratory tests at the expense of a primary focus on the history and examination. These errors can be avoided by adherence to an approach in which the patient's illness is defined first in *anatomic* and then in *pathophysiologic* terms; only then should a specific diagnosis be entertained. Arrival at a diagnosis permits the physician to institute therapy and to inform and counsel patients and their families about the expected disease course.

THE NEUROLOGIC METHOD OF CLINICAL EVALUATION

LOCATE THE LESION(S)

The first priority is to define the anatomic substrate responsible for the patient's illness by seeking to determine what part of the neural axis is likely to be involved in causing the neurologic symptoms. Can the disorder be mapped to one specific site in the nervous system, is it multifocal, or is there evidence of a more diffuse neurologic disease? Is the disorder restricted to the nervous system, or does it arise in the context of a systemic illness? Is it in the central nervous system (CNS), the peripheral nervous system (PNS), or both? If in the CNS, is the process restricted to the cerebral cortex, or is there evidence of basal ganglia, brainstem, cerebellum, and/or spinal cord involvement? Are the pain-sensitive meninges involved? If in the PNS, could the disorder be located in peripheral nerves and, if so, are motor or sensory nerves primarily affected, or is a lesion in the neuromuscular junction or muscle more likely?

The first clues to defining the anatomic area of involvement appear in the history, and the examination is then directed to confirm or rule out these impressions and to clarify uncertainties suggested by the history. A more detailed examination of a particular region of the CNSor PNSis often indicated. For example, the examination of a patient who presents with a history of ascending paresthesias and weakness should be directed toward deciding, among other things, if the location of the lesion is in the spinal cord or peripheral nerves. Focal back pain, a spinal cord sensory level, and incontinence suggest a spinal cord origin, whereas a stocking-glove pattern of sensory loss suggests peripheral nerve disease; areflexia usually indicates peripheral neuropathy but may also be present with spinal shock in acute spinal cord disorders.

Deciding "where the lesion is" accomplishes the task of limiting the possible etiologies to a manageable, finite number. In addition, this strategy safeguards against making tragic errors. Symptoms of recurrent vertigo, diplopia, and nystagmus should not trigger "multiple sclerosis" as an answer (etiology) but "brainstem" or "pons" (location); then a diagnosis of brainstem arteriovenous malformation will not be missed for lack of

consideration. Similarly, the combination of optic neuritis and spastic ataxic paraparesis should initially suggest optic nerve and spinal cord disease; multiple sclerosis, CNS syphilis, and vitamin B_{12} deficiency are treatable disorders that can produce this syndrome. Once the question, "Where is the lesion?" is answered, then the question, "What is the lesion?" can be addressed.

DEFINE THE PATHOPHYSIOLOGY

Clues to the pathophysiology of the disease process may also be present in the history. Primary neuronal (gray matter) disorders may present as early cognitive disturbances, movement disorders, or seizures, whereas white matter involvement produces predominantly "long tract" disorders of motor, sensory, visual and cerebellar pathways. Progressive and symmetric symptoms often have a metabolic or degenerative origin; in such cases lesions are usually not sharply circumscribed. Thus, a patient with paraparesis and a clear spinal cord sensory level is unlikely to have vitamin B_{12} deficiency as the explanation. A Lhermitte symptom (electric shock-like sensations evoked by neck flexion) is due to ectopic impulse generation in white matter pathways and occurs with demyelination in the cervical spinal cord. Symptoms that worsen after exposure to heat or exercise may indicate conduction block in demyelinated axons and suggest a diagnosis of multiple sclerosis. Slowly advancing visual scotoma with luminous edges, termed fortification spectra, are diagnostic of spreading cortical depression, such as occurs in migraine.

ESTABLISH AN ETIOLOGIC DIAGNOSIS

The clinical data obtained from the history and the examination are assembled into one of the known syndromes and are interpreted and translated in terms of neuroanatomy and neurophysiology (Table 356-2). From the syndrome the physician should be able to determine the anatomic localization(s) that best explains the clinical findings. The proper selection of laboratory tests is important to arrive at an anatomic, but more particularly an etiologic, diagnosis. The laboratory assessment of a patient with positive neurologic findings may include (1) serum electrolytes, complete blood count, and renal, liver, and endocrine studies; (2) cerebrospinal fluid (CSF) examination (see below); (3) neuroimaging studies (Chap. 358); or (4) electrophysiologic studies (Chap. 357). The anatomic localization, mode of onset and course of illness, other medical data, and laboratory findings are then integrated to establish an etiologic diagnosis.

THE NEUROLOGIC HISTORY

Attention to the description of the symptoms as experienced by the patient and substantiated by family members or friends often permits an accurate localization and determination of the probable cause of the complaints even before the neurologic examination is undertaken. Two principles should be followed. First, each complaint should be pursued as far as possible in an effort to delineate where the lesion might be or, more importantly, to formulate a set of questions to be answered by the examination. A patient complains of weakness of the right arm. What are the associated features? Is this weakness for brushing the hair (proximal) or opening a twist-top bottle (distal)? Second, negative associations may also be crucial. A patient with a right hemiparesis without a language deficit likely has a lesion (and likely an etiology) different from that of

a patient with a right hemiparesis and aphasia. Additional features of the history include the following:

1. *Temporal course of the illness*. It is important to ascertain the precise time of appearance and rate of progression of the symptoms experienced by the patient. The rapid onset of a neurologic complaint, occurring within seconds or minutes, usually indicates a cerebrovascular event, a seizure, or rarely migraine. The onset of sensory symptoms located in one extremity that spread over a few seconds to adjacent portions of that extremity and then to the other limb or to the face suggests a seizure. A more gradual onset and less well localized sensory symptoms point to the possibility of a transient ischemic attack (TIA). A similar but slower temporal march of a sensory change occurring with headache, nausea, or visual disturbance suggests migraine. In general, the march of migraine is slower than that of seizure, and a TIA tends to be more generalized in location on the side of the body or extremities. The presence of "positive" sensory symptoms (e.g., tingling) or involuntary motor movements suggests a seizure; in contrast, transient loss of function (negative symptoms) suggests a TIA. A stuttering onset where symptoms appear, stabilize, and then progress over hours or days also suggests cerebrovascular disease; an additional history of transient remission or regression indicates that the process is due to ischemia and not hemorrhage. On occasion, a demyelinating process may also produce new symptoms that evolve rapidly over the course of a few hours. Progressing symptoms associated with the systemic manifestations of fever, stiff neck, and altered level of consciousness raise the possibility of an infectious process. Relapsing and remitting symptoms involving different levels of the neuraxis suggest multiple sclerosis. Slowly progressive symptoms without remissions are characteristic of neurodegenerative disorders.

2. *Subjective descriptions of the complaint*. The same words often mean different things to different patients. "Dizziness" may imply impending syncope, a sense of giddiness, or true spinning vertigo. "Numbness" may mean a complete loss of feeling, a positive sensation of tingling, or paralysis. "Blurred vision" may be used to describe unilateral visual loss, as in transient monocular blindness, or diplopia. It is important to define the contextual meaning of the patient's complaint to understand its true significance.

3. *Corroboration of the history by others*. It is often useful to obtain additional information from family, friends, or observers to corroborate or expand the patient's description. Memory loss, aphasia, loss of insight, drug or alcohol abuse, and other factors may impair the patient's capacity to communicate normally with the examiner or prevent openness about factors that have contributed to the illness. Episodes of loss of consciousness that may be due to syncope or seizures necessitate that details be sought from observers to ascertain the exact circumstances.

4. *Family history*. Many neurologic disorders have an underlying genetic component. The presence of a Mendelian disorder, such as Huntington's disease or Charcot-Marie-Tooth neuropathy, is often obvious if appropriate family data are available. In polygenic disorders such as multiple sclerosis or migraine, a positive family history, when present, may be helpful. It is important to elicit family history about all illnesses, in addition to neurologic and psychiatric disorders. A familial propensity to hypertension or heart disease may be relevant to a patient who presents with a stroke. Many inherited neurologic illnesses are associated with multisystem manifestations that

may provide clues to the correct diagnosis (e.g., the phakomatoses, hepatocerebral disorders, neuro-ophthalmic syndromes).

5. *Medical illnesses*. Many neurologic illnesses occur in the context of systemic disorders. Disorders such as diabetes mellitus, hypertension, and abnormalities of blood lipids predispose to cerebrovascular disease. Marfan's syndrome and related collagen disorders predispose to dissection of the cranial arteries and also to aneurysmal subarachnoid hemorrhage; the latter may also occur with polycystic kidney disease. A recent onset of asthma suggests the possibility of polyarteritis nodosa. Various neurologic disorders occur with dysthyroid states. A solitary mass lesion may be a brain abscess in a patient with valvular heart disease, a primary hemorrhage in a patient with a coagulopathy, a metastasis in a patient with underlying cancer, or a lymphoma or toxoplasmosis in a patient with AIDS. The presence of systemic diseases that are associated with peripheral neuropathy should be explored. Most patients with coma in a hospital setting can be shown to have a metabolic, toxic, or infectious process.

6. *The patient's perception of the disease*. It is frequently helpful to ask patients what they perceive to be wrong. Patients who complain of failing memory are often concerned that they have early symptoms of Alzheimer's disease; more often they are found to suffer from depression. Patients with headaches may fear that a tumor or an impending stroke is a possibility. Patients with sensory symptoms frequently are concerned about the possibility of multiple sclerosis. The patient may seek medical attention because a relative or friend has been diagnosed with a serious neurologic illness.

7. *Drug use and abuse and toxin exposure*. It is essential to inquire about the history of drug use, both prescribed and illicit. Digitalis use may provoke complaints of yellow vision. Excessive vitamin ingestion may lead to disease; for example, vitamin A and pseudotumor cerebri, or pyridoxine and peripheral neuropathy. Aminoglycoside antibiotics may exacerbate symptoms of weakness in patients with disorders of neuromuscular transmission, such as myasthenia gravis. Dizziness may be secondary to ototoxicity caused by aminoglycosides. Many patients are unaware that over-the-counter sleeping pills, cold preparations, and diet pills are actually drugs. Alcohol, the most prevalent neurotoxin, is often not recognized as such by patients. A history of environmental or industrial exposure to neurotoxins may provide an essential clue; consultation with the patient's family or employer may be required.

8. *History of malignancy*. Patients with malignancy may present with nervous system metastases, a paraneoplastic syndrome (Chap. 101), or complications from chemotherapy or radiotherapy.

9. *Formulating an impression of the patient*. Use the opportunity while taking the history to form an impression of the patient. Is there evidence of anxiety, depression, hypochondriasis? Are there any clues to defects of language, memory, inappropriate behavior, or secondary gain? The neurologic assessment begins as soon as the patient walks into the room and the first introduction is made.

THE NEUROLOGIC EXAMINATION

A systematic neurologic examination should encompass a survey of all functions from the cerebrum to the peripheral nerve and muscle, i.e., from the mental status examination to the simplest reflexes. Physicians should acquire skills that come only from the repeated use of the same techniques and instruments on a large number of individuals with and without neurologic disease. Errors and serious omissions are avoided if the examination procedure is orderly and systematic, beginning with mental (cerebral) functions and continuing with cranial nerves; then with motor, reflex, and sensory functions of the arms, trunk, and legs; and finishing with an analysis of posture and gait (Table 356-3).

This detailed examination is undertaken only if there are symptoms of disturbed nervous system functioning. If none are present, it suffices to do an abbreviated examination that includes evaluation only of pupils, ocular movements, optic fundi, facial movements, speech, strength of arm and leg muscles, tendon and plantar reflexes, pain and vibratory sensation in hands and feet, and gait. All this can be completed in 3 to 5 min.

Several additional points about the examination are worth noting. First, in recording observations, it is important to describe what is found rather than to apply a poorly defined medical term (i.e., "patient groans to sternal rub" rather than "obtunded"). Second, if the patient's complaint is brought out by some activity, reproduce the activity in the office. If the complaint is of dizziness when raising the right arm and turning the head to the left, have the patient do it. If pain occurs after walking two blocks, have the patient demonstrate it, and repeat the examination. Finally, the use of tests that are individually tailored to the patient's problem can be of value in assessing changes over time. Tests of walking a 25-ft distance (normal, 5 to 6 s; note assistance, if any), repetitive finger or toe tapping (normal, 20 to 25 taps in 5 s), or handwriting are examples.

The neurologic examination may be normal even in patients with a serious neurologic disease, such as one that causes seizures or syncope. A comatose patient may arrive with no available history; the examination proceeds along the lines described inChap. 24. An inadequate history may be compensated for to some extent by a succession of examinations from which the course of the illness may be plotted.

LUMBAR PUNCTURE

The clinical indications for lumbar puncture (LP) are listed inTable 356-4. In experienced hands, LP is a safe procedure. The patient is asked to lie on his or her side facing away from the examiner. The back is positioned at the edge of the bed or table near the examiner. The patient is asked to "roll up into a ball" -- the neck is gently flexed and the knees drawn up to the abdomen. Proper positioning is essential for success; the examiner should ensure that the shoulders and pelvis are vertically aligned without forward or backward tilt. A pillow is placed under the neck for comfort and a blanket offered for warmth. Because the spinal cord terminates at approximately the L1 vertebral level, the LP is performed below this level; i.e., at or below the L2-L3 interspace. A useful anatomic guidepost is the iliac crest which corresponds to the L3-L4 interspace. The interspace is chosen after gentle palpation to identify the spinous processes at each lumbar level. The skin is cleansed with an antibacterial liquid and alcohol, and the area is draped with sterile cloths. Local anesthetic, typically 1%

lidocaine, is injected into the subcutaneous tissue; a topical anesthetic cream (lidocaine 2.5%/prilocaine 2.5%) applied 90 min before the procedure can eliminate pain associated with injection. Approximately 5 min after the lidocaine injection, the LP needle (typically 22 gauge) is inserted in the midline between two spinous processes and slowly advanced at a slightly cephalic angle aiming at the umbilicus. The bevel of the needle should be maintained in a horizontal position, parallel to the direction of the dural fibers; this minimizes injury to the fibers as the dura is penetrated. In most adults, the needle is advanced 4 to 5 cm (1$1/2$to 2 in.) before the subarachnoid space is reached, and the examiner usually recognizes entry as a sudden release of resistance. Some examiners prefer to remove the stylet periodically as the needle is advanced to look forCSFflow. If the needle cannot be advanced because bone is hit, if the patient experiences sharp radiating pain down one leg, or if the tap is "dry," the needle is removed completely and repositioned.

Once the subarachnoid space is reached, a manometer is attached to the needle and theCSFpressure is measured. The examiner should look for normal oscillations in CSF pressure associated with pulse and respirations. Depending on the clinical indication, fluid is then obtained for studies that include the following: (1) cell count, differential, and presence of microorganisms -- it is often useful to repeat the cell count in the first and last tube; (2) protein, glucose, and other chemical measurements; (3) cytology; (4) bacteriologic cultures and virus isolation; (5) VDRL, cryptococcal antigen, and serologic and genetic tests for other microorganisms as appropriate; (6) immunoelectrophoresis for determination of gamma globulin level (paired serum sample essential), oligoclonal banding, and other special biochemical tests (NH$_3$, pH, CO$_2$, enzymes). Normal values of CSF constituents are shown inAppendix A. Under most conditions, the physician should not be concerned about removing too large a quantity of CSF. A sufficient volume to obtain all the data required is essential. In particular, adequate volumes for cytology, when indicated, should be removed.

Failure to enter the lumbar subarachnoid space after two or three trials can usually be corrected by repositioning the patient in the sitting position and then assisting them to lie on their side. The "dry tap" is more often due to an improperly placed needle than to a pathologic obliteration of subarachnoid space by a compressive lesion of the spinal cord or by chronic adhesive arachnoiditis. A bloody tap due to penetration of a meningeal vessel may be confused with subarachnoid hemorrhage. In these situations a specimen ofCSFshould be centrifuged immediately after it is obtained; clear supernatant CSF after centrifugation supports the diagnosis of a bloody tap, whereas xanthochromic supernatant suggests subarachnoid hemorrhage. In general, bloody CSF due to a meningeal vessel puncture clears gradually in successive tubes, whereas blood due to a subarachnoid hemorrhage does not. In addition to subarachnoid hemorrage, xanthochromic CSF may also be present in patients with liver disease or when the level of CSF protein is elevated [>1.5 to 2.0 g/L (150 to 200 mg/dL)].

There are several absolute or relative contraindications toLP. The procedure should be undertaken with particular care in patients with thrombocytopenia or disorders of blood coagulation because serious hemorrhage into the extradural or intradural space may occur. In these situations, it is prudent whenever possible to transfuse platelets, administer fresh frozen plasma, or reverse therapeutic anticoagulation before the procedure. Patients receiving low-molecular-weight heparin may be at risk of

hemorrhage unless doses are held for 24 h. An LP through areas of cutaneous or soft tissue infection may spread infection to the meninges; thus LP at these sites should be avoided.

In patients with elevated CSF pressure, potentially fatal cerebellar or tentorial herniation may follow LP. This possibility should be considered in all patients with focal neurologic findings, altered mental status, or papilledema. If CSF examination is required in such cases, it is wise to first obtain a neuroimaging scan to exclude a mass lesion. An exception to this rule is suspected meningitis, where immediate CSF examination is indicated. In this situation, the LP may be performed with a fine-bore (24-gauge) needle. If the pressure is >400 mmHg, the minimum required sample of fluid should be obtained, the needle removed, and, according to the suspected clinical disease and the patient's condition, intravenous mannitol administered in a dose of 0.75 to 1.0 mg/kg; unless contraindicated, dexamethasone may also be started in a dose of 4 to 6 mg every 6 h.

After LP, the patient is normally positioned in a comfortable, recumbant position for 1 h before rising. The principal complication of LP is headache, occurring in 10 to 30% of patients, caused by a drop in CSF pressure related to persistent leakage of CSF. Such headaches typically begin 12 to 48 h after the procedure and may last from several days to 2 weeks, rarely longer. These headaches are strikingly positional in character; they are worsened by an upright posture and are relieved by lying flat. *Therapy is discussed in Chap. 15.*

(Bibliography omitted in Palm version)

Back to Table of Contents

357. ELECTROPHYSIOLOGIC STUDIES OF THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS - *Michael J. Aminoff*

ELECTROENCEPHALOGRAPHY

The electrical activity of the brain [the *electroencephalogram* (EEG)] is easily recorded from electrodes placed on the scalp. The potential difference between pairs of electrodes on the scalp (bipolar derivation) or between individual scalp electrodes and a relatively inactive common reference point (referential derivation) is amplified and displayed on paper or the screen of an oscilloscope. The findings depend on the patient's age and level of arousal. The rhythmic activity normally recorded represents the postsynaptic potentials of vertically oriented pyramidal cells of the cerebral cortex and is characterized by its frequency. In normal awake adults lying quietly with the eyes closed, an 8- to 13-Hz alpha rhythm is seen posteriorly in the EEG, intermixed with a variable amount of generalized faster (beta) activity, and it is attenuated when the eyes are opened (Fig. 357-1). During drowsiness, the alpha rhythm is also attenuated; with light sleep, slower activity in the theta (4 to 7 Hz) and delta (<4 Hz) ranges becomes more conspicuous.

The EEG is best recorded from several different electrode arrangements (montages) in turn, and activating procedures are generally undertaken in an attempt to provoke abnormalities. Such procedures commonly include hyperventilation (for 3 or 4 min), photic stimulation, sleep, and the deprivation of sleep on the night prior to the recording.

Electroencephalography is relatively inexpensive and may aid clinical management in several different contexts.

THE EEG AND EPILEPSY

The EEG is most useful in evaluating patients with suspected epilepsy. The presence of *electrographic seizure activity,* i.e., of abnormal, repetitive, rhythmic activity having an abrupt onset and termination, clearly establishes the diagnosis. The absence of such electrocerebral accompaniment does not exclude a seizure disorder, however, because there may be no change in the scalp-recorded EEG during simple or complex partial seizures. With generalized tonic-clonic seizures, however, the EEG is always abnormal during the episode. It is often not possible to obtain an EEG during clinical events that may represent seizures, especially when such events occur unpredictably or infrequently. The development of portable equipment to record the EEG continuously on cassettes for 24 h or longer in ambulatory patients has made it easier to capture the electrocerebral accompaniments of such clinical episodes, and monitoring by this means is sometimes helpful in confirming that seizures are occurring, characterizing the nature of clinically equivocal episodes, and determining the frequency of epileptic events.

The EEG findings may also be helpful in the interictal period by showing certain abnormalities that are strongly supportive of a diagnosis of epilepsy. Such *epileptiform activity* consists of bursts of abnormal discharges containing spikes or sharp waves. The presence of epileptiform activity is not specific for epilepsy, but it has a much greater prevalence in epileptic patients than in normal individuals. When epileptiform

activity is found in the EEG of a patient with episodic behavioral disturbances that clinically might be epileptic in nature, the likelihood that epilepsy is the correct diagnosis is markedly increased.

The EEG findings have also been used in classifying seizure disorders and selecting appropriate anticonvulsant medication for individual patients (Fig. 357-2). The episodic generalized spike-wave activity that occurs during and between seizures in patients with typical absences (petit mal epilepsy) contrasts with the normal findings, focal interictal epileptiform discharges, or ictal patterns found in patients with complex partial seizures. These latter seizures may have no correlates in the scalp-recorded EEG or may be associated with abnormal rhythmic activity of variable frequency, a localized or generalized distribution, and a stereotyped pattern that varies with the patient. Focal or lateralized epileptogenic lesions are important to recognize, especially if surgical treatment is contemplated. Intensive long-term monitoring of clinical behavior and the EEG is required for operative candidates, however, and this generally also involves recording from intracranially placed electrodes (which may be subdural, extradural, or intracerebral in location).

The findings in the routine scalp-recorded EEG may indicate the prognosis of seizure disorders: in general, a normal EEG implies a better prognosis than otherwise, whereas an abnormal background or profuse epileptiform activity suggests a poor outlook. The EEG findings are not helpful in determining which patients with head injuries, stroke, or brain tumors will go on to develop seizures, because in such circumstances epileptiform activity is commonly encountered regardless of whether seizures occur. The EEG findings are sometimes used to determine whether anticonvulsant medication can be discontinued in epileptic patients who have been seizure-free for several years, but the findings provide only a general guide to prognosis: further seizures may occur after withdrawal of anticonvulsant medication despite a normal EEG or, conversely, may not occur despite a continuing EEG abnormality. The decision to discontinue anticonvulsant medication is made on clinical grounds, and the EEG does not have a useful role in this context except for providing guidance when there is clinical ambiguity or the patient requires reassurance about a particular course of action.

The EEG has no role in the management of tonic-clonic status epilepticus except when there is clinical uncertainty whether seizures are continuing in a comatose patient. In patients treated by pentobarbital-induced coma for refractory status epilepticus, the EEG findings are useful in indicating the level of anesthesia and whether seizures are occurring. During status epilepticus, the EEG shows repeated electrographic seizures or continuous spike-wave discharges. In nonconvulsive status epilepticus, a disorder that may not be recognized unless an EEG is performed, the EEG may also show continuous spike-wave activity ("spike-wave stupor") or, less commonly, repetitive electrographic seizures (complex partial status epilepticus).

THE EEG AND COMA

In patients with an altered mental state or some degree of obtundation, the EEG tends to become slower as consciousness is depressed, regardless of the underlying cause (Fig. 357-1). Other findings may also be present and may suggest diagnostic possibilities, as when electrographic seizures are found or there is a focal abnormality

indicating a structural lesion. The EEG generally slows in metabolic encephalopathies, and triphasic waves may be present. The findings do not permit differentiation of the underlying metabolic disturbance but help to exclude other encephalopathic processes by indicating the diffuse extent of cerebral dysfunction. The response of the EEG to external stimulation is helpful prognostically because electrocerebral responsiveness implies a lighter level of coma than a nonreactive EEG. Serial records provide a better guide to prognosis than a single record and supplement the clinical examination in following the course of events. As the depth of coma increases, the EEG becomes nonreactive and may show a burst-suppression pattern, with bursts of mixed-frequency activity separated by intervals of relative cerebral inactivity. In other instances there is a reduction in amplitude of the EEG until eventually electrocerebral activity cannot be detected. Such electrocerebral silence does not necessarily reflect irreversible brain damage, because it may occur in hypothermic patients or with drug overdose. The prognosis of electrocerebral silence, when recorded using an adequate technique, depends upon the clinical context in which it is found. In patients with severe cerebral anoxia, for example, electrocerebral silence in a technically satisfactory record implies that useful cognitive recovery will not occur.

In patients with clinically suspected brain death, an EEG, when recorded using appropriate technical standards, may be confirmatory by showing electrocerebral silence. However, complicating disorders that may produce a similar but reversible EEG appearance (e.g., hypothermia or drug intoxication) must be excluded. The presence of residual EEG activity in suspected brain death fails to confirm the diagnosis but does not exclude it. The EEG is usually normal in patients with locked-in syndrome and helps in distinguishing this disorder from the comatose state with which it is sometimes confused clinically.

THE EEG IN OTHER NEUROLOGIC DISORDERS

In the developed countries, computed tomography (CT) scanning and magnetic resonance imaging (MRI) have taken the place of EEG as a noninvasive means of screening for focal structural abnormalities of the brain, such as tumors, infarcts, or hematomas (Fig. 357-1). Nonetheless, the EEG is still used for this purpose in many parts of the world, although infratentorial or slowly expanding lesions may fail to cause any abnormalities. Focal slow-wave disturbances, a localized loss of electrocerebral activity, or more generalized electrocerebral disturbances are common findings but provide no reliable indication about the nature of the underlying pathology.

In patients with an acute encephalopathy, focal or lateralized periodic slow-wave complexes, sometimes with a sharpened outline, suggest a diagnosis of herpes simplex encephalitis, and periodic lateralized epileptiform discharges (PLEDs) are commonly found with acute hemispheric pathology such as a hematoma, abscess, or rapidly expanding tumor. The EEG findings in dementia are usually nonspecific and do not distinguish between the different causes of cognitive decline except in rare instances when the presence of complexes occurring with a regular repetition rate (so-called periodic complexes) in dementing disorders, for example, supports a diagnosis of Creutzfeldt-Jakob disease (Fig. 357-1) or subacute sclerosing panencephalitis. In most patients with dementias, the EEG is normal or diffusely slowed, and the EEG findings alone cannot indicate whether a patient is demented or distinguish between dementia

and pseudodementia.

EVOKED POTENTIALS

SENSORY EVOKED POTENTIALS

The noninvasive recording of spinal or cerebral potentials elicited by stimulation of specific afferent pathways is an important means of monitoring the functional integrity of these pathways but does not indicate the pathologic basis of lesions involving them. Such evoked potentials (EPs) are so small compared to the background EEG activity that the responses to a number of stimuli have to be recorded and averaged with a computer in order to permit their recognition and definition. The background EEG activity, which has no fixed temporal relationship to the stimulus, is averaged out by this procedure.

Visual evoked potentials (VEPs) are elicited by monocular stimulation with a reversing checkerboard pattern and are recorded from the occipital region in the midline and on either side of the scalp. The component of major clinical importance is the so-called P100 response, a positive peak having a latency of approximately 100 ms. Its presence, latency, and symmetry over the two sides of the scalp are noted. Amplitude may also be measured, but changes in size are much less helpful for the recognition of pathology. VEPs are most useful in detecting dysfunction of the visual pathways anterior to the optic chiasm. In patients with acute severe optic neuritis, the P100 is frequently lost or grossly attenuated; as clinical recovery occurs and visual acuity improves, the P100 is restored but with an increased latency that generally remains abnormally prolonged indefinitely. The VEP findings are therefore helpful in indicating previous or subclinical optic neuritis. They may also be abnormal with ocular abnormalities and with other causes of optic nerve disease, such as ischemia or compression by a tumor. Normal VEPs may be elicited by flash stimuli in patients with cortical blindness.

Brainstem auditory evoked potentials (BAEPs) are elicited by monaural stimulation with repetitive clicks and are recorded between the vertex of the scalp and the mastoid process or earlobe. A series of potentials, designated by roman numerals, occurs in the first 10 ms after the stimulus and represents in part the sequential activation of different structures in the pathway between the auditory nerve (wave I) and the inferior colliculus (wave V) in the midbrain. The presence, latency, and interpeak latency of the first five positive potentials recorded at the vertex are evaluated. The findings are helpful in screening for acoustic neuromas, detecting brainstem pathology, and evaluating comatose patients. The BAEPs are normal in coma due to metabolic/toxic disorders or bihemispheric disease but abnormal in the presence of brainstem pathology.

Somatosensory evoked potentials (SEPs) are recorded over the scalp and spine in response to electrical stimulation of a peripheral (mixed or cutaneous) nerve. The configuration, polarity, and latency of the responses depend on the nerve that is stimulated and on the recording arrangements. SEPs are used to evaluate proximal (otherwise inaccessible) portions of the peripheral nervous system and the integrity of the central somatosensory pathways.

Clinical Utility of Sensory Evoked Potentials EP studies may detect and localize lesions in afferent pathways in the central nervous system (CNS). They have been used

particularly to investigate patients with suspected multiple sclerosis (MS), the diagnosis of which requires the recognition of lesions involving several different regions of the central white matter. In patients with clinical evidence of only one lesion, the electrophysiologic recognition of abnormalities in other sites helps to suggest or support the diagnosis but does not establish it unequivocally. Multimodality EP abnormalities are not specific for MS; they may occur in AIDS, Lyme disease, systemic lupus erythematosus, neurosyphilis, spinocerebellar degenerations, familial spastic paraplegia, and deficiency of vitamin E or B_{12}, among other disorders. The diagnostic utility of the electrophysiologic findings therefore depends upon the circumstances in which they are found. Abnormalities may aid in the localization of lesions to broad areas of the CNS, but attempts at precise localization on electrophysiologic grounds are misleading because the generators of many components of the EP are unknown.

The EP findings are sometimes of prognostic relevance. Bilateral loss of SEP components that are generated in the cerebral cortex implies that cognition may not be regained in posttraumatic or postanoxic coma, and EP studies may also be useful in evaluating patients with suspected brain death. In patients who are comatose for uncertain reasons, preserved BAEPs suggest either a metabolic-toxic etiology or bihemispheric disease. In patients with spinal cord injuries, SEPs have been used to indicate the completeness of the lesion -- the presence or early return of a cortically generated response to stimulation of a nerve below the injured segment of the cord indicates an incomplete lesion and thus a better prognosis for functional recovery than otherwise. In surgery, intraoperative EP monitoring of neural structures placed at risk by the procedure may permit the early recognition of dysfunction and thereby permit a neurologic complication to be averted or minimized.

Visual and auditory acuity may be determined using EP techniques in patients whose age or mental state precludes traditional ophthalmologic or audiologic examinations.

COGNITIVE EVOKED POTENTIALS

Certain EP components depend upon the mental attention of the subject and the setting in which the stimulus occurs, rather than simply on the physical characteristics of the stimulus. Such "event-related" potentials (ERPs) or "endogenous" potentials are related in some manner to the cognitive aspects of distinguishing an infrequently occurring target stimulus from other stimuli occurring more frequently. For clinical purposes, attention has been directed particularly at the so-called P3 component of the ERP, which is also designated the P300 component because of its positive polarity and latency of approximately 300 to 400 ms after onset of an auditory target stimulus. The P3 component is prolonged in latency in many patients with dementia, whereas it is generally normal in patients with depression or other psychiatric disorders that might be mistaken for dementia. ERPs are therefore sometimes helpful in making this distinction when there is clinical uncertainty, although a response of normal latency does not exclude a dementing disorder.

MOTOR EVOKED POTENTIALS

The electrical potentials recorded from muscle or the spinal cord following stimulation of the motor cortex or central motor pathways are referred to as *motor evoked potentials*.

For clinical purposes such responses are recorded most often as the compound muscle action potentials elicited by transcutaneous magnetic stimulation of the motor cortex. A strong but brief magnetic field is produced by passing a current through a coil, and this induces stimulating currents in the subjacent neural tissue. The procedure is painless and apparently safe. Abnormalities have been described in several neurologic disorders with clinical or subclinical involvement of central motor pathways, includingMS and motor neuron disease. In addition to a possible role in the diagnosis of neurologic disorders or in evaluating the extent of pathologic involvement, the technique provides information of prognostic relevance (e.g., in suggesting the likelihood of recovery of motor function after stroke) and is useful as a means of monitoring intraoperatively the functional integrity of central motor tracts.

ELECTROPHYSIOLOGIC STUDIES OF MUSCLE AND NERVE

The motor unit is the basic element subserving motor function. It is defined as an anterior horn cell, its axon and neuromuscular junctions, and all the muscle fibers innervated by the axon. The number of motor units in a muscle ranges from approximately 10 in the extraocular muscles to several thousand in the large muscles of the legs. There is considerable variation in the average number of muscle fibers within the motor units of an individual muscle, i.e., in the innervation ratio of different muscles. Thus the innervation ratio is less than 25 in the human external rectus or platysma muscle and between 1600 and 1700 in the medial head of the gastrocnemius muscle. The muscle fibers of individual motor units are divided into two general types by distinctive contractile properties, histochemical stains, and characteristic responses to fatigue. Within each motor unit, all of the muscle fibers are of the same type.

ELECTROMYOGRAPHY

The pattern of electrical activity in muscle [i.e., the *electromyogram* (EMG)], both at rest and during activity, may be recorded from a needle electrode inserted into the muscle. The nature and pattern of abnormalities relate to disorders at different levels of the motor unit.

Relaxed muscle normally is electrically silent except in the endplate region, but abnormal spontaneous activity (Fig. 357-3) occurs in various neuromuscular disorders, especially those associated with denervation or inflammatory changes in affected muscle. Fibrillation potentials and positive sharp waves (which reflect muscle fiber irritability) and complex repetitive discharges are most often -- but not always -- found in denervated muscle and may also occur after muscle injury and in certain myopathic disorders, especially inflammatory disorders such as polymyositis. After an acute neuropathic lesion they are found earlier in proximal rather than distal muscles and sometimes do not develop distally in the extremities for 4 to 6 weeks; once present, they may persist indefinitely unless reinnervation occurs or the muscle degenerates so completely that no viable tissue remains. Fasciculation potentials (which reflect the spontaneous activity of individual motor units) are characteristic of slowly progressive neuropathic disorders, especially those with degeneration of anterior horn cells (such as amyotrophic lateral sclerosis). Myotonic discharges -- high-frequency discharges of potentials derived from single muscle fibers that wax and wane in amplitude and frequency -- are the signature of myotonic disorders such as myotonic dystrophy or

myotonia congenita but occur occasionally in polymyositis or other, rarer, disorders.

Slight voluntary contraction of a muscle leads to activation of a small number of motor units. The potentials generated by any muscle fibers of these units that are within the pick-up range of the needle electrode will be recorded (Fig. 357-3). The parameters of normal motor unit action potentials depend on the muscle under study and age of the patient, but their duration is normally between 5 and 15 ms, amplitude is between 200 uV and 2 mV, and most are bi- or triphasic. The number of units activated depends on the degree of voluntary activity. An increase in muscle contraction is associated with an increase in the number of motor units that are activated (recruited) and in the frequency with which they discharge. With a full contraction, so many motor units are normally activated that individual motor unit action potentials can no longer be distinguished, and a complete interference pattern is said to have been produced.

The incidence of small, short-duration, polyphasic motor unit action potentials (i.e., having more than four phases) is usually increased in myopathic muscle, and an excessive number of units is activated for a specified degree of voluntary activity. By contrast, the loss of motor units that occurs in neuropathic disorders leads to a reduction in number of units activated during a maximal contraction and an increase in their firing rate, i.e., there is an incomplete or reduced interference pattern; the configuration and dimensions of the potentials may also be abnormal, depending on the duration of the neuropathic process and on whether reinnervation has occurred. The surviving motor units are initially normal in configuration but, as reinnervation occurs, they increase in amplitude and duration and become polyphasic (Fig. 357-3).

Action potentials from the same motor unit sometimes fire with a consistent temporal relationship to each other, so that double, triple, or multiple discharges are recorded, especially in tetany, hemifacial spasm, or myokymia.

Electrical silence characterizes the involuntary, sustained muscle contraction that occurs in phosphorylase deficiency, which is designated a *contracture*.

EMGenables disorders of the motor units to be detected and characterized as either neurogenic or myopathic. In neurogenic disorders, the pattern of affected muscles may localize the lesion to the anterior horn cells or to a specific site as the axons traverse a nerve root, limb plexus, and peripheral nerve to their terminal arborizations. The findings do not enable a specific etiologic diagnosis to be made, however, except in conjunction with the clinical findings and results of other laboratory studies.

The findings may provide a guide to the severity of an acute disorder of a peripheral or cranial nerve (by indicating whether denervation has occurred and the completeness of the lesion), and whether the pathologic process is active or progressive in chronic or degenerative disorders such as amyotrophic lateral sclerosis. Such information is important for prognostic purposes.

Various quantitativeEMGapproaches have been developed. The most common is to determine the mean duration and amplitude of 20 motor unit action potentials using a standardized technique. The technique of macro-EMG provides information about the number and size of muscle fibers in a larger volume of the motor unit territory and has

also been used to estimate the number of motor units in a muscle. Scanning EMG is a computer-based technique that has been used to study the topography of motor unit action potentials and, in particular, the spatial and temporal distribution of activity in individual units. The technique of single-fiber EMG is discussed separately below.

NERVE CONDUCTION STUDIES

Recording of the electrical response of a muscle to stimulation of its motor nerve at two or more points along its course (Fig. 357-4) permits conduction velocity to be determined in the fastest-conducting motor fibers between the points of stimulation. The latency and amplitude of the electrical response of muscle (i.e., of the compound muscle action potential) to stimulation of its motor nerve at a distal site are also compared with values defined in normal subjects. Sensory nerve conduction studies are performed by determining the conduction velocity and amplitude of action potentials in sensory fibers when these fibers are stimulated at one point and the responses are recorded at another point along the course of the nerve. In adults, conduction velocity in the arms is normally between 50 and 70 m/s, and in the legs is between 40 and 60 m/s.

Nerve conduction studies complement theEMGexamination, enabling the presence and extent of peripheral nerve pathology to be determined. They are particularly helpful in determining whether sensory symptoms are arising from pathology proximal or distal to the dorsal root ganglia (in the former instance, peripheral sensory conduction studies will be normal) and whether neuromuscular dysfunction relates to peripheral nerve disease. In patients with a mononeuropathy, they are invaluable as a means of localizing a focal lesion, determining the extent and severity of the underlying pathology, providing a guide to prognosis, and detecting subclinical involvement of other peripheral nerves. They enable a polyneuropathy to be distinguished from a mononeuropathy multiplex when this is not possible clinically, an important distinction because of the etiologic implications. Nerve conduction studies provide a means of following the progression and therapeutic response of peripheral nerve disorders and are being used increasingly for this purpose in clinical trials. They may suggest the underlying pathologic basis in individual cases. Conduction velocity is often markedly slowed, terminal motor latencies are prolonged, and compound motor and sensory nerve action potentials may be dispersed in the demyelinative neuropathies (such as in Guillain-Barre syndrome, chronic inflammatory polyneuropathy, metachromatic leukodystrophy, or certain hereditary neuropathies); conduction block is frequent in acquired varieties of these neuropathies. By contrast, conduction velocity is normal or slowed only mildly, sensory nerve action potentials are small or absent, and there is EMG evidence of denervation in axonal neuropathies such as occur in association with metabolic or toxic disorders.

The utility and complementary role ofEMG and nerve conduction studies are best illustrated by reference to a common clinical problem. Numbness and paresthesia of the little finger, and associated wasting of the intrinsic muscles of the hand may result from a spinal cord lesion, C8/T1 radiculopathy, brachial plexopathy (lower trunk or medial cord), or a lesion of the ulnar nerve. If sensory nerve action potentials can be recorded normally at the wrist following stimulation of the digital fibers in the affected finger, the pathology is probably proximal to the dorsal root ganglia, i.e., there is a radiculopathy or more central lesion; absence of the sensory potentials, by contrast, suggests distal

pathology. EMG examination will indicate whether the pattern of affected muscles conforms to radicular or ulnar nerve territory, or is more extensive (thereby favoring a plexopathy); ulnar motor conduction studies will generally also distinguish between a radiculopathy (normal findings) and ulnar neuropathy (abnormal findings) and will often identify the site of an ulnar nerve lesion -- the nerve is stimulated at several points along its course to determine whether the compound action potential recorded from a distal muscle that it supplies shows a marked alteration in size or area, or a disproportionate change in latency, with stimulation at a particular site. The electrophysiologic findings thus permit a definitive diagnosis to be made and specific treatment instituted in circumstances where there is clinical ambiguity.

F WAVE STUDIES

Stimulation of a motor nerve causes impulses to travel antidromically (i.e., toward the spinal cord) as well as orthodromically (to the nerve terminals). Such antidromic impulses cause a few of the anterior horn cells to discharge, producing a small motor response that occurs considerably later than the direct response elicited by nerve stimulation. The F wave so elicited is sometimes abnormal (absent or delayed) with proximal pathology of the peripheral nervous system, such as a radiculopathy, and may therefore be helpful in detecting abnormalities when conventional nerve conduction studies are normal. In general, however, the clinical utility of F wave studies has been disappointing, except perhaps in Guillain-Barre syndrome, where they are often absent or delayed.

H REFLEX STUDIES

The H reflex is easily recorded only from the soleus muscle (S1) in normal adults. It is elicited by low-intensity stimulation of the tibial nerve and represents a monosynaptic reflex in which spindle (Ia) afferent fibers constitute the afferent arc and alpha motor axons the efferent pathway. The H reflexes are often absent bilaterally in elderly patients or with polyneuropathies and may be lost unilaterally in S1 radiculopathies.

MUSCLE RESPONSE TO REPETITIVE NERVE STIMULATION

The size of the electrical response of a muscle to supramaximal electrical stimulation of its motor nerve relates to the number of muscle fibers that are activated. Neuromuscular transmission can be tested by several different protocols, but the most helpful is to record with surface electrodes the electrical response of a muscle to supramaximal stimulation of its motor nerve by repetitive (2 to 3 Hz) shocks delivered before and at selected intervals after a maximal voluntary contraction.

There is normally little or no change in size of the compound muscle action potential following repetitive stimulation of a motor nerve at 2 to 3 Hz with stimuli delivered at intervals after voluntary contraction of the muscle for about 20 to 30 s, even though preceding activity in the junctional region influences the release of acetylcholine and thus the size of the endplate potentials elicited by a test stimulus. This is because more acetylcholine is normally released than is required to bring the motor endplate potentials to the threshold for generating muscle fiber action potentials. In disorders of neuromuscular transmission this safety factor is reduced. Thus, in myasthenia gravis

repetitive stimulation, particularly at a rate of between 2 and 5 Hz, may lead to a depression of neuromuscular transmission, with a decrement in size of the response recorded from affected muscles. Similarly, immediately after a period of maximal voluntary activity, single or repetitive stimuli of the motor nerve may elicit larger muscle responses than before, indicating that more muscle fibers are responding. This postactivation facilitation of neuromuscular transmission is followed by a longer-lasting period of depression, maximal between 2 and 4 min after the conditioning period and lasting for as long as 10 min or so, during which responses are reduced in size.

Decrementing responses to repetitive stimulation at 2 to 5 Hz are common in myasthenia gravis but may also occur in the congenital myasthenic syndromes. In Lambert-Eaton myasthenic syndrome, in which there is defective release of acetylcholine at the neuromuscular junction, the compound muscle action potential elicited by a single stimulus is generally very small. With repetitive stimulation at rates of up to 10 Hz, the first few responses may decline in size, but subsequent responses increase. If faster rates of stimulation are used (20 to 50 Hz), the increment may be dramatic so that the amplitude of compound muscle action potentials eventually reaches a size that is several times larger than the initial response. In patients with botulism, the response to repetitive stimulation is similar to that in Lambert-Eaton syndrome, although the findings are somewhat more variable and not all muscles are affected.

SINGLE-FIBER ELECTROMYOGRAPHY

The technique is particularly helpful in detecting disorders of neuromuscular transmission. A special needle electrode is placed within a muscle and positioned to record action potentials from two muscle fibers belonging to the same motor unit. The time interval between the two potentials will vary in consecutive discharges, and this is called the *neuromuscular jitter*. The jitter can be quantified as the mean difference between consecutive interpotential intervals and is normally between 10 and 50 us. This value is increased when neuromuscular transmission is disturbed for any reason, and in some instances impulses in individual muscle fibers may fail to occur because of impulse blocking at the neuromuscular junction. Single-fiberEMG is more sensitive than repetitive nerve stimulation or determination of acetylcholine receptor antibody levels in diagnosing myasthenia gravis.

Single-fiberEMG can also be used to determine the mean fiber density of motor units (i.e., mean number of muscle fibers per motor unit within the recording area) and to estimate the number of motor units in a muscle, but this is of less immediate clinical relevance.

BLINK REFLEXES

Electrical or mechanical stimulation of the supraorbital nerve on one side leads to two separate reflex responses of the orbicularis oculi -- an ipsilateral R1 response having a latency of approximately 10 ms and a bilateral R2 response with a latency in the order of 30 ms. The trigeminal and facial nerves constitute the afferent and efferent arcs of the reflex, respectively. Abnormalities of either nerve or intrinsic lesions of the medulla or pons may lead to uni- or bilateral loss of the response, and the findings may therefore be helpful in identifying or localizing such pathology.

(Bibliography omitted in Palm version)

A dramatic increase in the role of imaging in diagnosis of neurologic diseases occurred with the development of computed tomography (CT) in the early 1970s and of magnetic resonance imaging (MRI) in the 1980s. MRI has gradually replaced CT for many indications and has also reduced the indications for invasive neuroimaging techniques, such as myelography and angiography. In general, MRI is more sensitive than CT for the evaluation of most lesions affecting the central nervous system, particularly those in the spinal cord, cranial nerves, and posterior fossa. CT is more sensitive than MRI for visualizing fine osseous detail, such as temporal bone anatomy and fractures. Recent developments, such as helical CT, CT angiography (CTA), MR angiography (MRA), positron emission tomography (PET), Doppler ultrasound, and interventional angiography have continued to advance diagnosis and guide therapy. Conventional angiography is reserved for cases in which small-vessel detail is essential for diagnosis (Table 358-1).

COMPUTED TOMOGRAPHY

Technique The CT image is a computer-generated cross-sectional representation of anatomy created by an analysis of the attenuation of x-ray beams passed through various points around a section of the body. As the x-ray source, collimated to the desired slice thickness, rotates around the patient, sensitive x-ray detectors aligned 180° from the source detect x-rays attenuated by the patient's anatomy. A computer calculates a "back projection" image from the 360° x-ray attenuation profile. Greater x-ray attenuation, as caused by bone, results in areas of high "density," while soft tissue structures, which attenuate x-rays less, are lower in density. The resolution of an image depends on the radiation dose, the collimation (slice thickness), the field of view, and the matrix size of the display. A typical modern CT scanner is capable of obtaining sections 1 to 2, 5, and 10 mm thick at a speed of 1 to 3 s per section; complete studies of the brain can be completed in <2 to 3 min.

Intravenous contrast is often administered prior to or during a CT study to identify vascular structures and to detect defects in the blood-brain barrier (BBB) associated with disorders such as tumors, infarcts, and infections. An intact BBB prevents contrast molecules, which are large, from exiting the intravascular compartment. In the normal central nervous system, only vessels and those structures not having a BBB (e.g., the pituitary gland, choroid plexus, and dura) enhance. The use of contrast agents carries a risk of allergic reaction, increases the dose of radiation when both noncontrast and contrast CT scans are to be obtained, adds expense, and may mask hemorrhage; thus, before contrast is administered, the indication for its use should always be considered carefully.

Helical CT is a new technique in which continuous three-dimensional CT information is obtained. In the helical scan mode, the table moves continuously through the rotating x-ray beam, generating a "helix" of information that can be reformatted into various slice thicknesses. Advantages include shorter scan times, reduced patient and organ motion, and the ability to acquire images during the infusion of intravenous contrast. The contrast images can be used to construct CT angiograms of vascular structures. CTA images require a workstation to threshold and segment CT images for

display ([Fig. 358-1](#)). CTA has proven useful in assessing the carotid bifurcation and intracranial arterial anatomy in selected instances in which a contraindication to[MRA](#)exists. Newer "multidetector" scanners allow multiple sections to be obtained with each revolution of the gantry. These scanners have further decreased the time per examination and permit rapid assessment of vascular anatomy ([Fig. 358-2](#)).

Indications The indications for[CT](#) have decreased since the development of[MRI](#). While MRI gives greater soft tissue contrast and is more sensitive than CT in detecting early brain damage, CT is useful in imaging osseous structures of the spine, skull base, and temporal bones. CT is also more sensitive and specific than MRI for acute subarachnoid hemorrhage. In the spine, CT is useful in evaluating patients with osseous spinal stenosis and spondylosis, but MRI is often preferred in those with neurologic deficits. CT can also be obtained following intrathecal contrast injection to evaluate the intracranial cisterns for cerebrospinal fluid (CSF) fistula, as well as the spinal subarachnoid space.

Complications [CT](#) is safe and reliable. Radiation exposure is between 3 and 5 cGy per examination. The most frequent complications are associated with use of intravenous contrast agents. Two broad categories of contrast media, ionic and nonionic, are in use. Ionic agents are relatively safe and inexpensive but cause a higher incidence of toxicity reactions than nonionic agents.

Nephrotoxicity caused by contrast administration (*contrast nephropathy*) may result from hemodynamic changes, tubular obstruction and cell damage, or immunologic reactions to contrast agents. A rise in serum creatinine of at least 85 umol/L (1 mg/dL) within 48 h of contrast administration is often used as a definition of contrast nephropathy, although other causes of acute renal failure must be excluded. The prognosis is usually favorable, with serum creatinine levels returning to baseline within 1 to 2 weeks. Risk factors for contrast nephropathy include advanced age, preexisting renal disease, diabetes, dehydration, and high contrast dose. Patients with diabetes and those with mild renal failure should be well hydrated prior to the administration of nonionic agents. Nonionic, low-osmolar media produce fewer abnormalities in renal blood flow and less endothelial cell damage than ionic agents (see[Guidelines](#)).

A sensation of heat, *pain*, *nausea*, and *vomiting* are well-known side effects following intravenous administration of ionic contrast media, and they become more important as studies require longer imaging times and repeated contrast injections. Pain and the sensation of heat are probably due to the osmolality of the agent and vasodilation. These side effects are less intense or nonexistent with nonionic contrast media.

Anaphylactoid reactions to intravenous contrast media range from mild hives to bronchospasm to acute anaphylaxis and death. The pathogenesis of these allergic reactions is not fully understood, but it is thought to include the release of mediators such as histamine, antibody-antigen reactions, and complement activation. Severe allergic reactions occur in approximately 0.04% of patients receiving nonionic media, sixfold fewer than with ionic media. Risk factors include a history of prior contrast reaction, allergy (asthma and hay fever), and cardiac disease. In these patients, a noncontrast[CT](#) or[MRI](#)procedure should be considered as an alternative to contrast administration. If contrast is absolutely required, a nonionic agent should be used in conjunction with pretreatment with glucocorticoids and antihistamines ([Table 358-2](#)

and Guidelines). Patients with allergic reactions to iodinated contrast material do not usually react against gadolinium-based magnetic resonance (MR) contrast material, although it would be wise to pretreat in a similar fashion prior to MR contrast administration.

MAGNETIC RESONANCE IMAGING

Technique The phenomenon of magnetic resonance is a complex interaction between protons in biologic tissues, a static and alternating magnetic field (the magnet), and energy in the form of radiofrequency waves of a specific frequency (Rf), introduced by coils placed next to the body part of interest. The energy state of the hydrogen protons is transiently excited. The subsequent return to equilibrium (*relaxation*) of the protons results in a release of Rf energy (the *echo*), which can be measured by the same surface coils that delivered the Rf pulses. The complex Rf signal, or echo, is transformed by Fourier analysis into the information used to form an MR image.

T1 and T2 Relaxation Times The rate of return to equilibrium of perturbed protons is called the *relaxation rate*. The relaxation rate is different for different normal and pathologic tissues. The relaxation rate of a hydrogen proton in a tissue is influenced by surrounding molecular environment and atomic neighbors. Two relaxation rates, the T1 and T2 relaxation times, are measurable. The T1 relaxation rate is the time for 63% of the protons to return to their normal equilibrium state, while the T2 relaxation rate is the time for 63% of the protons to become dephased owing to interactions among adjacent protons. The intensity of the signal, and thus the image contrast, can be modulated by altering certain parameters, such as the interval between Rf pulses (TR) and the time between the Rf pulse and the signal reception (TE). So-called T1-weighted (T1W) images are produced by keeping the TR and TE relatively short. Under these conditions, contrast between structures is based primarily on their T1 relaxation differences. T2-weighted (T2W) images are produced by using longer TR and TE times. Fat and subacute hemorrhage have short T1 relaxation rates and a high signal intensity on T1W images. Watery media, such as CSF and edematous tissue, have long T1 and T2 relaxation rates, a low signal intensity on T1W images, and a high signal intensity on T2W images. As white matter contains more lipid (due to myelin), it contains 10 to 15% less water than gray matter. These two chemical differences account for much of the contrast difference between gray and white matter on MRI (Fig. 358-3). T2W images are more sensitive than T1W images to edema or myelin destruction (Fig. 371-2).

MR images can be generated in sagittal, coronal, axial, and oblique planes without changing the patient's position. Each plane obtained requires a separate sequence lasting 5 to 10 min. Unlike CT, movement of the patient during a sequence will distort *all* the images; therefore, patient cooperation is important. Approximately 5% of the population experience claustrophobia in the MR environment. This can be reduced by mild sedation. Three-dimensional volumetric imaging is also possible with MR, resulting in data that can be reformatted in any plane and manipulated in a real-time fashion to highlight certain disease processes. Fluid-attenuated inversion recovery, or FLAIR, is a pulse sequence that produces T2W images in which the CSF signal is suppressed. FLAIR images are more sensitive than standard spine echo images for cortical lesions and meningeal processes (Fig. 358-2*D*).

Contrast Material The heavy-metal element *gadolinium* forms the basis of all current intravenousMRcontrast agents. Gadolinium is a paramagnetic substance, which means that it reduces the T1 and T2 relaxation times of nearby water protons, resulting in a high signal onT1Wimages. The metal is chelated to an agent such as DTPA, which allows renal excretion without toxicity. Approximately 0.2 mL/kg body weight is administered intravenously (10 to 15 mL for the average-sized adult); the cost is approximately $60 per 20 mL. Gadolinium contrast does not cross a normalBBB, and thus it causes enhancement of brain tissue only at sites of abnormalities in the BBB (Fig. 371-2*D*) and in areas of the brain that normally have no BBB, such as the pituitary gland. Allergic reactions are extremely rare; renal failure does not occur. These agents can be administered safely to children as well as adults.

MAGNETIC RESONANCE ANGIOGRAPHY

Flowing blood exhibits complexMRsignals that range from bright to dark relative to background stationary tissue (Fig. 358-3). Fast-flowing blood, such as arterial blood, shows no signal on routine MR images. Slower flow, as in veins or distal to arterial stenoses, may appear high in signal. It is possible, by varying the MR image parameters, to assess blood flow either qualitatively or quantitatively. This is the basis ofMRA, which capitalizes on the differences in signal between moving blood and stationary tissue on gradient echo images (Fig. 358-4). *Gradient echo images* differ from standard spin echo images in being more sensitive to blood products, calcification, and other susceptibility artifacts. The suppression of background signal achieved in short-flip-angle gradient echo images provides the contrast needed for flowing blood to appear bright in signal on MRA images.

It is important to understand thatMRAprovides a *vascular flow map* rather than the anatomic map given by conventional angiography. Two MRA techniques, time-of-flight (TOF) and phase-contrast, are used. TOF, currently the technique used most frequently, relies on the suppression of nonmoving tissue to provide a background for the high signal intensity of flowing blood. A typical TOF angiography sequence results in a series of contiguous thinMRsections (0.9 mm thick), which can be viewed as a stack to create an angiographic image data set that can be reformatted or viewed in various planes and angles to reveal the vascular relationships (Fig. 358-4). Either arterial (MRA) or venous (MRV) structures may be highlighted.

Phase-contrastMRA has a longer acquisition time thanTOF MRA but reveals the velocity and direction of blood flow in addition to providing anatomic information similar to that of TOF imaging. Through the selection of different imaging parameters, differing blood velocities can be highlighted; selective venous and arterial MRA images thus can be obtained. One advantage of phase-contrast MRA is the excellent suppression of background signal.

MRAhas lower resolution than conventional angiography and therefore cannot detect small-vessel detail, such as is needed in the workup of vasculitis. It is also less sensitive to slow flow and thus may not differentiate occlusive disease from near-occlusive disease. Motion, either by the patient or by anatomic structures, may distort the images, creating artifacts that may be misinterpreted as stenoses or occlusions. These limitations notwithstanding, MRA has proved useful in evaluation of the cervical carotid

artery and larger-caliber intracranial arterial and venous structures. It has also proved useful in the noninvasive detection of intracranial aneurysms (Fig. 361-13) and vascular malformations.

ECHO-PLANAR MR IMAGING

Recent improvements in gradients, software, and high-speed computer processors now permit MR imaging of the brain on the order of milliseconds. With echo-planar MRI (EPI), fast gradients are switched on and off at high speeds to create the information used to form an image. In routine spin echo imaging, images of the brain can be obtained in 5 to 10 min. With EPI, all of the information required for processing an image is accumulated in 50 to 150 ms, and the information for the entire brain is obtained in 5 to 10 s. EPI allows motion-free imaging, as well as perfusion imaging, diffusion imaging (Fig. 358-4*A*), functional MRI, and kinematic motion studies.

EPI techniques are making their way into clinical practice. The hope for these techniques is that they will provide useful functional data in addition to exquisite anatomic images. EPI *perfusion imaging* and *diffusion imaging* are useful in early detection of ischemic injury of the brain, and may be useful in demonstrating "tissue at risk" of further infarction (Fig. 358-4*A*). Diffusion imaging may also be useful in the characterization of white matter tracts. *Functional MRI* of the brain is a technique that localizes regions of activity in the brain following task activation. Tasks alter the balance of oxyhemoglobin and deoxyhemoglobin within specific regions of the activated cortex. Repetitive actions such as finger tapping elicit an increase in the amount of blood flow delivered to a specific region of the brain, resulting in a slight increase in oxyhemoglobin and a 2 to 3% change in signal intensity (Fig. 358-5). Further work will determine whether these techniques are cost-effective or clinically useful, but currently somatosensory and auditory cortex localization are possible.

Complications of MRI and Patient Safety MRI is considered safe for patients, even at very high field strengths. Serious injuries have been caused, however, by the high magnetic fields used. Ferromagnetic (metal) objects are attracted to the magnet and may act as missiles if brought too close to the magnet. Likewise, ferromagnetic aneurysm clips may torque within the magnet, causing hemorrhage and even death. Metallic foreign bodies in the eye have moved and caused hemorrhage, so screening for ocular metallic fragments is indicated in those with a history of ocular metallic foreign bodies. Implanted cardiac pacemakers are a contraindication to MRI owing to the risk of induced arrhythmias. All personnel and patients must be screened and educated thoroughly to prevent such disasters. Table 358-3 lists several of the more common contraindications for MRI.

POSITRON EMISSION TOMOGRAPHY

PET relies on the detection of positrons emitted during the decay of a radionuclide that has been injected into a patient. The most frequently used moiety is 2-[$_{18}$F]fluoro-2-deoxy-D-glucose (FDG), which is an analogue of glucose and is taken up by cells competitively with 2-deoxyglucose. Multiple images of glucose uptake activity are formed after 45 to 60 min. Images reveal differences in regional glucose activity among normal and pathologic brain structures. FDG PET scanning has been used to

assist in differentiating radiation necrosis from active neoplasm following therapy, in localizing temporal lobe epileptic foci, and in detecting metastatic disease and determining cardiac viability. A lower activity of FDG in the parietal lobes has been associated with Alzheimer's disease ([Fig. 362-1](#)).

MYELOGRAPHY

Technique Myelography involves the intrathecal instillation of 8 to 15 mL of water-soluble iodinated contrast medium (180 to 300 mg/mL) into the lumbar or cervical subarachnoid space via a percutaneously placed spinal needle (22 gauge or smaller). Contrast is maneuvered into the area of interest by fluoroscopic guidance and patient rotation. *Conventional myelography* involves a relatively high concentration and volume of contrast material and visualization by x-ray "spot films" and formal "overhead" plain films. The radiation exposure during conventional myelography is 4 to 8 cGy, making it one of the more radiation-intense procedures. The gonads should be shielded if possible, although doing so is sometimes difficult. [CT](#) scanning is often performed after myelography (*CT myelography*), to better demonstrate the spinal cord and roots as filling defects in the opacified subarachnoid space. CT myelography alone, in which CT is performed after the subarachnoid injection of a small amount of relatively dilute contrast material, has replaced conventional myelography for many indications, thereby reducing exposure to radiation and contrast media. CT slices 3 mm thick are routinely obtained through the area of interest.

Indications For diagnosis of diseases of the spinal canal and cord, myelography has been largely replaced by [CT](#), CT myelography, and [MRI](#)([Table 358-1](#)). The remaining indications for conventional plain film myelography include the evaluation of suspected meningeal or arachnoid cysts and the localization of spinal dural arteriovenous fistulas and [CSF](#) fistulas. Conventional myelography and CT myelography provide the most precise information in patients with prior spinal fusion and spinal fixation hardware.

Contraindications Myelography is relatively safe. However, it should be performed with caution in any patient with suspected herniation, elevated intracranial pressure, or a history of allergic reaction to intrathecal contrast media. In patients with a suspected spinal block, only a small amount of contrast medium should be instilled below the level of the block to minimize the risk of deterioration. Lumbar puncture is to be avoided in patients with bleeding disorders, including patients receiving anticoagulant therapy ([Chap. 356](#)).

Complications Complications resulting from myelography are related to the needle puncture and to reactions to intrathecal contrast material.

Vasovagal syncope may occur during lumbar puncture; it is accentuated by the upright position used during lumbar myelography. Adequate hydration before and after myelography will reduce the incidence of this complication.

Headache, *nausea*, and *vomiting* are the most frequent complications of dural puncture and myelography, occurring in up to 38% of patients. These symptoms are thought to result from neurotoxic effects of the contrast agent, persistent leakage of [CSF](#) at the puncture site, or psychological reactions to the procedure. The incidence of headache

has been reduced with the use of smaller-gauge spinal needles and nonionic, water-soluble contrast agents.

Postural headache (post-lumbar puncture headache) is generally due to prolonged leakage of CSF from the puncture site, resulting in CSF hypotension. Intravenous hydration may be helpful, and an autologous epidural blood patch is indicated in patients with persistent headache 48 h after myelography (Chap. 15).

Hearing loss is a rare complication. It may result from a direct toxic effect of the contrast medium or from an alteration of the pressure equilibrium between CSF and perilymph in the inner ear.

Puncture of the spinal cord is a rare but serious complication of cervical (C1-2) and high lumbar puncture. The cervical approach requires proper alignment of the patient and is best performed in the prone position using fluoroscopic guidance. Direct puncture of the spinal cord, laceration of epidural and vertebral venous and arterial structures, and hyperextension of the neck are reported complications. Injection of contrast material into the spinal cord can precipitate acute neurologic decline or subacute hemorrhagic necrosis of the gray matter. The risk of cord puncture is greatest in patients with spinal stenosis or conditions that reduce CSF volume. In these settings, a low-dose lumbar injection followed by thin-section CT is a safer alternative to cervical puncture.

Intrathecal contrast reactions are rare, but aseptic meningitis and encephalopathy may occur. The latter is usually dose-related and associated with contrast entering the intracranial subarachnoid space. *Seizures* occur following myelography in 0.1 to 0.3% of patients. Risk factors include a preexisting seizure disorder and the use of a total iodine dose of >4500 mg. Other reported symptoms include headache, hyperthermia, hallucinations, depression, and anxiety states. These neurotoxic side effects have been reduced by the development of nonionic, water-soluble contrast agents, as well as by head elevation and generous hydration following myelography.

Arachnoiditis, or inflammation of the leptomeninges, has also been ascribed to the use of contrast agents for myelography. Pantopaque, an oil-soluble contrast agent no longer used, was first noted to cause arachnoiditis, especially in cases where myelography resulted in subarachnoid bleeding (i.e., traumatic tap). The incidence of arachnoiditis with new water-soluble, nonionic contrast agents is much lower than with Pantopaque and with ionic, water-soluble agents (metrizamide). Other variables that increase the likelihood of arachnoiditis include trauma, infection, and subarachnoid hemorrhage.

ANGIOGRAPHY

Technique Angiography is essential in the diagnostic evaluation of many patients with vascular pathology. However, it carries the greatest risk of morbidity of all diagnostic imaging procedures, owing to the necessity of inserting a catheter into a blood vessel, directing the catheter to the required location, injecting contrast material to visualize the vessel, and removing the catheter while maintaining hemostasis. Therapeutic transcatheter procedures (see below) have become important options for the treatment of some cerebrovascular diseases. The decision to undertake a diagnostic or therapeutic angiographic procedure requires careful assessment of the goals of the

investigation and its attendant risks.

Patients undergoing angiography should be well hydrated before and after the procedure to better tolerate the contrast agents. Since the femoral route is used most commonly, the femoral artery must be compressed after the procedure to prevent a hematoma from developing. The puncture site and distal pulses should be evaluated carefully after the procedure; complications can include thigh hematoma or distal emboli.

Indications Table 358-1 lists some of the indications for conventional angiography. Over the past 20 years, angiography has been replaced for many indications byCT orMRI. However, it is still used today for evaluating intracranial small-vessel pathology (such as vasculitis), for assessing vascular malformations and aneurysms, and in intravascular therapeutic procedures.

Complications The vast majority of aortic arch, carotid, and vertebral arteriograms are carried out via transfemoral arterial access. A common femoral arterial puncture provides retrograde access via the aorta to the aortic arch and great vessels. The most feared complication of cerebral angiography is stroke. Thrombus can form on or inside the tip of the catheter, and atherosclerotic thrombus or plaque can be dislodged by the catheter or guidewire or by the force of injection and can embolize distally in the cerebral circulation. The duration and extent of the resulting ischemic neurologic deficit depends on the size and length of the embolus, its composition (fresh thrombus is thought to fragment more readily), its location, and the available collateral circulation. Risk factors for ischemic complications include limited experience on the part of the angiographer, atherosclerosis, vasospasm, low cardiac output, decreased oxygen-carrying capacity, advanced age, and possibly migraine. The risk of a neurologic complication varies but is approximately 4% for transient ischemic attack and stroke, 1% for permanent deficit, and<0.1% for death.

Ionic contrast material injected into the cerebral vasculature can be neurotoxic if theBBB is breached, either by an underlying disease or by the injection of hyperosmolar contrast agent. Ionic contrast media are less well tolerated than nonionic media, probably because they can induce changes in cell membrane electrical potentials. Patients with dolichoectasia of the basilar artery can suffer reversible brainstem dysfunction and acute short-term memory loss during angiography owing to the slow percolation of the contrast material and the consequent prolonged exposure of the brain. Rarely, an intracranial aneurysm ruptures during an angiographic contrast injection, causing subarachnoid hemorrhage, perhaps as a result of injection under high pressure.

Spinal Angiography Spinal angiography may be indicated to evaluate vascular malformations and tumors and to identify the artery of Adamkiewicz prior to aortic aneurysm repair. The procedure is lengthy and requires the use of relatively large volumes of contrast; the incidence of serious complications, including paraparesis, subjective visual blurring, and altered speech, is approximately 2%.

Interventional Neuroradiology This rapidly developing field is providing new therapeutic options for patients with difficult neurovascular problems. Available

procedures include detachable coil therapy for aneurysms, particulate or liquid adhesive embolization of arteriovenous malformations, balloon angioplasty and stenting of stenosis or vasospasm, transarterial or transvenous embolization of dural arteriovenous fistulas, balloon occlusion of carotid-cavernous and vertebral fistulas, endovascular treatment of vein-of-Galen malformations, preoperative embolization of tumors, and thrombolysis of acute arterial or venous thrombosis. Many of these disorders place the patient at high risk of cerebral hemorrhage, stroke, or death. The therapeutic risks are comparable to those of neurosurgery rather than routine diagnostic radiographic procedures.

The highest complication rates are found with the therapies designed to treat the highest-risk diseases. In a large series of surgically difficult intracranial aneurysms treated with detachable balloons, Higashida and colleagues reported a 7.4% incidence of stroke and a 9.8% death rate. These figures must be considered in light of the high morbidity and mortality associated with untreated and surgically unapproachable aneurysms (Chap. 361). The advent of the electrolytically detachable coil has reduced these rates and ushered in a new era in the treatment of cerebral aneurysms. It remains to be determined what the role of coils will be relative to surgical options, but in many centers, coiling of aneurysms has become standard therapy for many aneurysms.

(Bibliography omitted in Palm version)

359. MOLECULAR DIAGNOSIS OF NEUROLOGIC DISORDERS - *Joseph B. Martin*, *Frank M. Longo*

Completion of the Human Genome Project, along with evolving strategies for linking disease phenotypes with gene loci, will increase the rate at which genes responsible for neurologic disorders are discovered. The widespread availability of DNA testing has already changed traditional diagnostic approaches and raised novel ethical issues. For example, the discovery of "susceptibility" genes that do not directly cause disease but modify the age of disease onset or rate of disease progression creates complexity in the application of molecular diagnosis, particularly in guiding the use of preventative therapies. In this chapter we review molecular diagnostic approaches relevant to neurologic disease and illustrate how they contribute to patient care.

DNA-BASED DIAGNOSIS OF NEUROLOGIC DISORDERS

Appropriate use of DNA testing in the clinical setting requires that the clinician have a general understanding of the available molecular diagnostic approaches and the limitations in their application and interpretation. For many of the disorders listed inTable 359-1, the identification of specific disease-causing mutations has made direct DNA diagnosis possible. Most disease-causing mutations consist of single base substitutions leading to amino acid substitutions (missense mutations), premature translation stop signals (nonsense mutations), or abnormal RNA transcript splicing. Other mutations result from DNA deletions, DNA duplications, or instability of trinucleotide repeats. The ability to detect a mutation eliminates the need for additional diagnostic studies. For disorders that have been linked to specific gene loci but for which specific mutations have not been identified, DNA diagnosis may be possible by family linkage analysis. Linkage analysis requires that family relationships (such as paternity) are correctly established, that informative markers are available, and that an adequate number of family members are genotyped and clinically evaluated. For many patients, lack of this information makes DNA diagnosis impossible.

Approaches for Detection of DNA Mutations

Direct sequencing of patient DNA These methods generally require amplification of DNA by the polymerase chain reaction (PCR). With most current sequencing methods, only 300 to 400 DNA bases are determined in each sequencing reaction; therefore, sequencing-based strategies are best applied when a limited region contains the majority of potential mutation sites. Direct DNA sequencing allows novel mutations to be detected and decreases the chances of a false-negative result. In some cases, the significance of previously uncharacterized missense mutations will be difficult to interpret. While they may code for harmless amino acid polymorphisms, in other cases they may be the cause of the disease. The segregation of the same mutation with the disease phenotype within a family or the substitution of nonconserved amino acids, especially in critical protein regions, suggests the latter.

Allele-Specific Oligonucleotide Hybridization This technique, in which sequence-specific oligonucleotides differentially recognize DNA segments with normal or mutated sequence, is best suited for detecting known mutations and can be applied to a large number of samples.

Differential restriction endonuclease patterns of[PCR]*-amplified DNA* DNA is digested with restriction enzymes that recognize either normal or mutated DNA sequence; the size pattern of resulting DNA fragments indicates whether the DNA sample contains normal or variant sequence ([Chap. 65]). This method is directed toward detecting known mutations.

Analysis of Unstable Repeats The number of trinucleotide (or other sized) repeats at a specific DNA locus can be counted by amplifying the DNA segment using[PCR] and then applying electrophoresis to determine the repeat number present in the amplified DNA. Large repeat expansions such as occur in myotonic dystrophy often prevent reliable application of PCR and require Southern blot analysis for detection.

Analysis of Single-Stranded Conformation Polymorphisms Gene segments of several hundred bases are[PCR]-amplified and electrophoresed under denaturing conditions. Mutation-induced alterations of DNA structure lead to altered electrophoretic patterns. This approach can detect novel mutations directly. Large numbers of samples, either covering many exons of a large gene or from many patients, can be analyzed using this method.

Detection of DNA deletions by fluorescence in situ hybridization DNA deletions are detected by hybridizing chromosomes with a fluorescent probe corresponding to the gene of interest. Fluorescence in situ hybridization (FISH) can detect deletions smaller than the 2000- to 3000-kb minimum detected by banding techniques.

Detection of DNA deletions or duplications by pulsed-field electrophoresis with Southern blot analysis DNA deletions and duplications can also be detected using an electrophoresis technique (pulsed-field) optimized for separation of large DNA segments. This method is used for the detection of chromosome 17p11.2-12 duplications or deletions in the diagnosis of Charcot-Marie-Tooth disease type 1A (CMT1A) and hereditary neuropathy with liability to pressure palsies (HNPP), respectively.

Cytogenetic Testing Chromosomes isolated from peripheral blood lymphocytes or tissue are stained, allowing identification of insertions, deletions, other chromosome imbalances, and assessment of chromosomal number ([Chap. 66]).

Some disorders such as Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD), neurofibromatosis (type 1 and 2), and familial amyotrophic lateral sclerosis (ALS) can be caused by dozens of different mutations within one gene. Heterogeneity over a wide region of a given gene is common for many genes involved in metabolic diseases. Many methods of DNA analysis focus on recurring point mutations or relatively short segments of DNA. Using such focused DNA analysis, mutations are found in only about two-thirds of DMD, BMD, and neurofibromatosis type 2 patients and in fewer than half of those with neurofibromatosis type 1. Advances in multiplex[PCR] and single-strand chain polymorphism analysis have increased this yield, but the clinician should be aware of the sensitivity of each DNA analysis.

Detection of Protein Abnormalities For some applications, diagnostic methods based

on protein properties or function are more effective and efficient than DNA-based tests. Traditional enzyme activity-based assays continue to be useful for diagnosis of many metabolic diseases, and additional protein functions can be used to detect other disorders. For example, immunostaining or western blot analysis using dystrophin antibodies may reveal decreased protein levels or aberrant distribution of dystrophin in a muscle biopsy of a patient in whom no DNA mutations are detected.

COMPLICATIONS AND LIMITATIONS OF GENETIC TESTING

The limitations of DNA testing must be considered. If the presumed diagnosis is in error or if the phenotype overlaps with other disorders, the failure to detect a given mutation does not rule out other phenotype-causing mutations in the same gene or other genes. Different mutations in the same gene can result in different phenotypes (*allelic heterogeneity*), and mutations in different genes can result in the same phenotype (*nonallelic genetic heterogeneity*). Other phenomena such as phenocopies, incomplete penetrance, age-dependent onset of phenotype, polygenic inheritance, imprinting, mitochondrial inheritance, and dynamic mutations (trinucleotide repeats) may also make interpretation of genetic testing difficult.

Nonallelic Genetic Heterogeneity Nonallelic genetic heterogeneity (also known simply as genetic heterogeneity) exists when individuals or families have similar pathologic and/or clinical syndromes caused by mutations in different genes, as in the multiple demyelinating forms of CMT disease (Table 359-1). For CMT type 1A the locus is 17p11.2, and mutations are present in the *PMP-22* gene encoding the peripheral myelin protein. CMT type 1B, which is clinically similar to CMT 1A but less common, is caused by a mutation on chromosome 1q22 in the *PMZ* gene encoding the P$_o$protein, a component of compact myelin. Familial Alzheimer's disease (AD) is caused by mutations in genes located on chromosome 14 (*presenilin 1*, causing 70 to 80% of early-onset AD), 1 (*presenilin 2*), and 21 (*amyloid precursor protein*). Type I autosomal dominant spinocerebellar ataxias (SCAs) are caused by mutations in at least 10 different genes. The phenotypically similar limb-girdle muscular dystrophies (LGMD) are also caused by mutations in a large number of distinct genes, several of which encode products in distinct protein families. As disease-causing mutations continue to be identified, genetic heterogeneity is emerging as a common theme.

An intriguing basis of genetic heterogeneity consists of mutations in distinct genes encoding proteins that function via direct interaction in common mechanistic pathways. The phenotypically similar X-linked and autosomal dominant Emery-Dreifuss muscular dystrophies are caused by mutations in genes encoding emerin and lamin A/C, respectively. Emerin and lamin A/C interaction is likely to be important for targeting emerin to the nuclear envelope. Tuberous sclerosis types 1 and 2 are caused by mutations in the genes encoding tuberin and hamartin, respectively. Evidence suggests that tuberin binds to hamartin, possibly acting as a chaperon.

Allelic Heterogeneity Different mutations in the same gene (allelic mutations) can cause markedly distinct clinical phenotypes. Mutations in the a1A voltage-gated calcium channel subunit can cause either familial hemiplegic migraine, SCA type 6, or episodic ataxia type 2 (EA-2). Familial Creutzfeldt-Jakob disease, fatal familial insomnia, and Gerstmann-Straussler-Scheinker disease are all caused by allelic mutations in the prion

protein gene (20pter-p12), each of which results in distinct aberrant protein isoforms or alterations in expression. Mutations in the gene encoding the L1 cell adhesion molecule (L1CAM) can cause either isolated hydrocephalus or MASA syndrome (mental retardation, aphasia, shuffling gait, and adducted thumbs). Mutations in the sodium channela subunit (SCN4A) can cause either hyperkalemic periodic paralysis (HYPP) or paramyotonia congenita (PC). In cases of allelic heterogeneity, different mutations cause distinct alterations in protein structure and function, resulting in separate phenotypes.

Phenocopies Patients may have a clinical presentation that resembles the phenotype of a genetic disorder but that has a nongenetic cause. Examples include vascular dementia appearing as familialAD, toxin- or drug-induced chorea mimicking Huntington's disease (HD), and vitamin E deficiency resembling Friedreich's ataxia (FA).

Variable Expressivity Variable expressivity occurs when the severity of a trait resulting from a mutant allele varies from mild to severe. Expression of the disease phenotype can be modified by other factors such as predisposing alleles of other genes, environmental agents, sex, and age. Variation in expression can also occur following somatic variations in trinucleotide repeats, as occurs in myotonic dystrophy (dystrophia myotonica, or DM).

Incomplete Penetrance Penetrance refers to the all-or-none expression of a mutant genotype. If a disease is expressed in <100% of individuals carrying the abnormal allele, it is said to have incomplete penetrance.

Polygenic Inheritance and Complex Traits The majority of diseases listed inTable 359-1 are caused by mutations in single genes. In disorders such as "sporadic"AD, Parkinson's disease, and multiple sclerosis, it is likely that disease onset is determined by concomitant mutations or polymorphisms in large numbers of genes. Genetic testing for susceptibility or diagnosis will require assessment of multigene "panels."

INFLUENCE OF GENETIC BACKGROUND

Machado-Joseph disease (MJD) andSCA type 3 are autosomal dominant ataxias originally described in different ethnic backgrounds with distinct features. MJD occurs in families, often of Portuguese-Azorean origin, and is manifest as hereditary ataxia with dystonia, rigidity, faciolingual fasciculation, and bulging eyes. In French families with a syndrome of progressive ataxia and dysarthria recognizably distinct from that found in Portuguese-Azorean families, the SCA3 gene was mapped to a site on chromosome 14q near the MJD locus. It is now clear that MJD and SCA3 are both caused by expansion of the same tract of CAG repeats in the same gene (*MJD1*) at 14q32.1. Although expansions in *MJD1* are the most common mutations in German SCA patients, the diagnosis of MJD had not previously been considered in this population. The extent to which different genetic background causes phenotypic heterogeneity will require further studies.

SUSCEPTIBILITY GENES

Allelic variations or mutations can cause increased susceptibility to specific diseases.

Detection of such DNA polymorphisms can influence differential diagnosis, as in the genotyping of *APOE* alleles in the diagnosis of AD. Apolipoprotein E (apoE) is a 299-amino-acid protein involved in mobilization and reutilization of lipoprotein cholesterol. ApoE secreted by astrocytes appears to be internalized by neurons via low-density lipoprotein-related receptors where it contributes to neuronal function. The three isoforms of apoE (apoE2, apoE3, and apoE4) are derived from three corresponding alleles of the *APOE* gene located at 19q13.2; the apoE4 allele is overrepresented in sporadic and familial AD and is a significant risk factor for the disease. In contrast, the apoE2 allele is underrepresented and thus may have a "protective" effect (Chap. 362).

The increased incidence of the *APOE4/4* genotype in AD patients has raised the possibility that ascertainment of *APOE* genotype might be useful in the diagnostic assessment of patients with dementia. For example, since AD accounts for some two-thirds of late-onset dementia, the prior probability of an elderly patient with dementia having AD is approximately 0.66. In many populations the probability of a demented patient with the *APOE4/4* genotype having AD increases to >0.90. However, since the relationship between *APOE4* genotype and probability of AD changes with age, gender, and ethnic background, application of population-based probabilities to specific individuals is limited. If a 25-year-old presents with dementia and has the *APOE4/4* genotype, it is very unlikely that this patient has AD. Since individuals with all *APOE* genotypes can have AD, genotypes cannot absolutely rule in or rule out this diagnosis. Even if genotyping increases the odds that a given patient has AD, it does not rule out the possibility that a treatable cause of dementia is present. Current diagnostic studies for demented patients are focused on detecting reversible causes of dementia (Chap. 26) *APOE* genotype results would not change the diagnostic evaluation and should not be ordered on a routine basis. Nevertheless, *APOE* genotyping might eventually be combined with yet-to-be developed diagnostic tests to form a "panel" of data with acceptable sensitivity and specificity for diagnosis (Chap. 362).

The availability of *APOE* genotyping raises questions about the use of predictive testing in asymptomatic individuals. Until preventive therapy is available, many clinicians would consider such predictive testing unethical. Moreover, useful predictions of age of onset based solely on *APOE* genotyping are not possible. For the 2% of the population with the high-risk *E4/E4* genotype, the period of risk extends from the fifties to beyond the nineties.

APPROACHES TO GENETIC TESTING

One indication for DNA analysis is to confirm the diagnosis of a specific disease already suggested by clinical assessment. DNA testing can also be used to narrow the differential diagnosis in cases with multiple diagnostic possibilities. DNA testing for HD allows patients to avoid neuroimaging and other diagnostic studies. When a patient presents with a well-established and relatively specific clinical phenotype, such as that of HD, initial genetic testing can focus on a single gene. In other disorders, such as the SCAs, the high degree of phenotypic overlap calls for concomitant testing of a panel of genes (*SCA1*, *SCA2*, *SCA3*, *SCA6*, and *SCA7*). Another application of genetic analysis is presymptomatic testing in members of families known or suspected to have a specific disorder. In these cases the most common reasons for testing are life

management issues, reproductive planning decisions, and eliminating the stress of unknown carrier status. Development of new therapies that delay onset or progression of neurodegenerative diseases will provide additional indications for presymptomatic testing.

Genetic testing should be conducted only in the context of comprehensive genetic counseling, in which the implications of potential test results are fully explained and adequate support services are available. Clinicians ordering genetic studies should be familiar with issues regarding informed consent, suicide risk, ongoing patient support, insurance, employment discrimination, testing of minors, and testing of fetal tissue.

A directory of diseases for which DNA diagnostic testing is available, along with a listing of testing sites, on the http://www.genetests.org, web site. This site was developed at the University of Washington, Seattle, and is supported by the U.S. National Library of Medicine and Maternal and Child Health Bureau.

CLINICAL AND GENETIC CLASSIFICATION OF GENE DISORDERS

Neurogenetic disorders have traditionally been classified and subtyped on the basis of clinical and pathophysiologic concepts. Their complexity, phenotypic variability, and overlapping features limit the resolution of phenotype-based classification and confound nosology. Identification of tightly linked disease markers and discovery of disease-causing mutations have provided a basis for refining such classifications. For example, the clinical distinction between neurofibromatosis type 1 and type 2 has been upheld by the discovery that they are caused by mutations in different genes, encoding the GTPase-activating protein and the merlin (schwannomin) cytoskeletal protein, respectively. In contrast, the finding that DMD and BMD are caused by mutations in the same gene points to shared pathophysiologic mechanisms and blurs the distinctions between these disorders. Mutations in different genes can lead to overlapping clinical syndromes as in the inherited ataxias. In other instances, phenotypically dissimilar disorders are caused by mutations in the same gene, as described above for the a1A voltage-gated calcium channel subunit gene.

Neurogenetic disorders with known chromosomal gene localization are organized primarily by clinical phenotype in Table 359-1. As in any clinical classification, there is frequent overlap in specific phenotypic features. Reference numbers from the Online Mendelian Inheritance in Man (OMIM) database (described below) are included to facilitate access to continuously updated disease information.

Different modes of inheritance occur in each of these categories. Neurologic genetic disorders inherited in Mendelian autosomal dominant mutations include HD, familial AD, ALS, DM, CMT, familial HYPP, SCA, and tuberous sclerosis. Autosomal recessive disorders include FA, Wilson's disease, and ataxia telangiectasia. X-linked recessive traits include DMD, spinobulbar muscular atrophy (Kennedy syndrome), and fragile X syndrome. Non-Mendelian patterns of transmission such as maternal inheritance can result from mitochondrial mutations (Chap. 383) and unstable trinucleotide repeats (see below).

The types of mutations causing neurologic genetic disorders include gene deletions (the

most common finding in DMD), insertions (e.g., Fukuyama-type congenital muscular dystrophy), duplications (e.g., CMT1A), translocations that interrupt the gene (neurofibromatosis type 1), and point mutations (e.g., in the superoxide dismutase gene in ALS). Point mutations, either missense or nonsense, are considered "static" mutations because they generally remain stable during meiosis and provide the basis for classic Mendelian inheritance. Unstable trinucleotide repeats cause "dynamic" mutations and account for the clinical phenomenon of anticipation.

GENETICALLY INDUCED MECHANISMS OF CELL DEATH

Three general mechanisms of cell death in genetic disorders have been proposed: loss of function, dominant-negative effects, and gain of function.

In *loss-of-function disorders*, the mutation causes a deficiency in an enzyme or protein resulting in cellular dysfunction. The best defined examples are the lysosomal storage disorders in which enzymatic deficiencies in complex lipid metabolism lead to accumulation of normal or abnormal cellular constituents. The mode of inheritance in these disorders is most often autosomal recessive, but it can also be X-linked or the combined result of an inherited germline mutation and an acquired somatic mutation ("second hit") that knocks out both alleles (such as the loss of a growth-suppressor gene in tumors such as retinoblastoma). It is less common for loss-of-function disorders to result from autosomal dominant mutations.

In the case of a *dominant-negative effect*, the abnormal mutation competes with or abolishes the normal allelic function at either the DNA, RNA, or protein level. A dominant-negative mechanism in DM has been suggested by observations that RNA transcripts with expanded CTG repeats precipitate normal RNA transcripts. In myotonia congenita, abnormal protein isoforms combined with normal isoforms of the CLC-1 chloride channel disrupt overall homomultimeric channel function.

In *gain-of-function effects*, the abnormal cellular function exerted by the mutation at one allele in some way renders the cell susceptible to toxic effects, whether or not the normal allele is expressed.

True dominant disorders such as HD and SCA1, in which the heterozygote genotype elicits the full disease phenotype, could be the result of (1) dominant-negative effects, in which proteins with expanded polyglutamine tracts would form oligomers with normal protein isoforms and thereby interfere with their function; or (2) toxic gain-of-function effects.

DISORDERS ASSOCIATED WITH TRINUCLEOTIDE REPEATS

An important group of neurologic disorders is caused by abnormal expansions of trinucleotide repeats (Table 359-2). A useful way of organizing repeat diseases and understanding their mechanisms is based on the location of the repeat expansions within the gene. Expansions can occur in the 5¢ untranslated region (UTR), within the open reading frame (translated portion of the gene), within the 3¢ UTR, or within introns.

The first category of trinucleotide repeat disorders in which repeats are located in the

5¢UTR includes fragile X syndrome and SCA12. Expansions in this region lead to impaired transcription with subsequent loss of protein expression.

The second category of trinucleotide diseases consists of neurodegenerative disorders in which expansion of a CAG repeat in the open reading frame encodes an aberrant protein with an expanded polyglutamine tract. The stretches of CAG repeats, which vary between 5 and 37 in the normal alleles of each gene, are increased two- to fourfold in the mutation. There is a striking correlation between larger numbers of repeats and both increased severity of the neurologic disorder and earlier age of onset. HD patients homozygous for the disease allele have phenotypes similar to heterozygous patients.

These observations suggest a model in which a gain of function is toxic to neurons. One possibility is that expanded polyglutamine tracts provide a substrate for aberrant protein-protein interactions. Such interactions might lead to a loss of function of a critical protein or toxic accumulations of protein aggregates. Several lines of evidence support such a protein-based hypothesis. Open reading frame CAG repeats are indeed translated into protein. Transgenic mice expressing a human *SCA1* gene with an expanded CAG repeat develop the characteristic phenotype only when the transgene is expressed. One study has suggested that polyglutamine tracts of the proteins that cause HD and dentatorubral-pallidoluysian atrophy (DRPLA) interact with glyceraldehyde-3-phosphate dehydrogenase and thereby might have a deleterious effect on neuronal energy metabolism.

Gain-of-function models of polyglutamine tract diseases must also reconcile the observations that each neurodegenerative disease affects only regionally specific populations of neurons, yet proteins associated with these disorders are widely expressed. One possibility is that each of these polyglutamine tract proteins interacts with yet-to-be discovered proteins that are indeed cell-type specific. The huntingtin-associated protein (HAP1) is one such candidate. HAP1 is selectively expressed in brain tissue and demonstrates enhanced association with forms of huntingtin protein with increased lengths of glutamine repeats. Another potential mechanism of cell-type specificity is that somatic instability of CAG repeats leads to greater expansions in specific cell populations. However, the relatively small variations of three to five in the number of triplet repeats in the HD gene in different regions of the brain makes this explanation less likely.

In the third category of trinucleotide repeat disorders, repeat expansion occurs in the 3¢UTR. In DM, a GTC (CAG in the antisense) repeat in the 3¢ UTR of the DM kinase gene expands manyfold, from 5 to 40 repeats in normal alleles to up to 2700 in severe cases. The repeat expansion is variable in different tissues, indicating that errors in DNA replication can occur during meiosis and during somatic cell mitosis. Since DNA sequence motifs in the 3¢ UTR of RNA transcripts regulate transcript stability and processing, expansions in this region might affect transcript levels and alter DM kinase protein levels. Quantitative analysis of messenger RNA in muscle biopsies has demonstrated marked disease-specific decreases in DM kinase mRNA in adult-onset DM patients. Levels of normal as well as mutant DM transcripts were decreased, suggesting a novel mechanism of a dominant-negative mutation occurring at the RNA level.

A second potential mechanism in DM is that expansion of the GTC repeat could inhibit expression of the adjacent *DMR-N9* (telomeric) and *DMAHP* (centromeric) genes. Disruption of adjacent chromatin structure by repeat expansion is one mechanism by which expression of adjacent genes could be inhibited. Alternatively, repeat expansions at one locus might affect expression of more than one gene by a *cis*-acting effect, and expansion-containing transcripts may alter levels of transcripts derived from a separate allele by a *trans*-acting effect.

A fourth category of trinucleotide repeat disease occurs with repeat expansion in an intron. FA is caused by the expansion of a GAA triplet in intron 1 of the *X25* gene. Repeat expansions with associated alterations in DNA structure are likely to impair *X25* transcription. Consistent with an autosomal recessive pattern of inheritance and a loss-of-function disease mechanism, the majority of FA patients tested to date demonstrate homozygosity for expanded alleles, while a smaller number are heterozygous with a combination of one expanded allele and point mutations in the other allele. FA is not associated with anticipation (see below) and manifests more often during adolescence rather than middle age, distinguishing it from other trinucleotide repeat diseases.

The discovery of triplet repeats has given molecular precision to old concepts such as *anticipation* (earlier onset of the disease in successive generations, which is associated with further expansion of the abnormal repeats in more severely affected individuals) and has helped to account for variations in gene expression. Variations in trinucleotide repeats in HD, and particularly in DM have given a molecular explanation for *variable expression*, where variations in repeats occurring among individual members of the family can lead to earlier onset or more severe symptoms and signs, as occurs in juvenile HD. Studies suggesting that other neurologic and psychiatric disorders involve anticipation raise the possibility that additional trinucleotide repeat diseases will be discovered.

ONLINE MENDELIAN INHERITANCE IN MAN

The OMIM catalogue contains a frequently updated listing of all known genetic traits. For each disease it includes information on clinical manifestations, mapping studies and identity (if available) of the relevant gene, and status of genetic testing. OMIM is administered by the National Center for Biotechnologic Information and is on the Internet at

www3.ncbi.nlm.nih.gov/omim/.

(Bibliography omitted in Palm version)

SECTION 2 - DISEASES OF THE CENTRAL NERVOUS SYSTEM

360. SEIZURES AND EPILEPSY - *Daniel H. Lowenstein*

A *seizure* (from the Latin *sacire*, "to take possession of") is a paroxysmal event due to abnormal, excessive, hypersynchronous discharges from an aggregate of central nervous system (CNS) neurons. Depending on the distribution of discharges, this abnormal CNS activity can have various manifestations, ranging from dramatic convulsive activity to experiential phenomena not readily discernible by an observer. Although a variety of factors influence the incidence and prevalence of seizures, approximately 5 to 10% of the population will have at least one seizure during their lifetime, with the highest incidence occurring in early childhood and late adulthood. Because seizures are common, this clinical problem is encountered frequently during medical practice in a variety of settings.

The meaning of the term seizure needs to be carefully distinguished from that of epilepsy. *Epilepsy* describes a condition in which a person has *recurrent* seizures due to a chronic, underlying process. This definition implies that a person with a single seizure, or recurrent seizures due to correctable or avoidable circumstances, does not necessarily have epilepsy. Epilepsy refers to a clinical phenomenon rather than a single disease entity, since there are many forms and causes of epilepsy. However, among the many causes of epilepsy there are various *epilepsy syndromes* in which the clinical and pathologic characteristics are distinctive and suggest a specific underlying etiology.

Using the definition of epilepsy as two or more unprovoked seizures, the incidence of epilepsy is approximately 0.3 to 0.5% in different populations throughout the world, and the prevalence of epilepsy has been estimated at 5 to 10 persons per 1000.

CLASSIFICATION OF SEIZURES

An essential step in the evaluation and management of a patient with a seizure is to determine the type of seizure that has occurred. The importance of this cannot be overemphasized -- classifying the seizure is essential for focusing the diagnostic approach on particular etiologies, selecting the appropriate therapy, and providing potentially vital information regarding prognosis. In 1981, the International League Against Epilepsy (ILAE) published a modified version of the International Classification of Epileptic Seizures that has continued to be a useful classification system (Table 360-1). This system is based on the clinical features of seizures and associated electroencephalographic findings. Other potentially distinctive features such as etiology or cellular substrate are not considered in this classification system, although this will undoubtedly change in the future as more is learned about the pathophysiologic mechanisms that underlie specific seizure types.

The main characteristic that distinguishes the different categories of seizures is whether the seizure activity is partial (synonymous with focal) or generalized. *Partial seizures* are those in which the seizure activity is restricted to discrete areas of the cerebral cortex. *Generalized seizures* involve diffuse regions of the brain simultaneously in a bilaterally symmetric fashion. Partial seizures are often associated with structural abnormalities of the brain. In contrast, generalized seizures may result from cellular, biochemical, or

structural abnormalities that have a more widespread distribution.

PARTIAL SEIZURES

Partial seizures occur within discrete regions of the brain. If consciousness is fully preserved during the seizure, the clinical manifestations are considered relatively simple and the seizure is termed a *simple partial seizure*. If consciousness is impaired, the symptomatology is more complex and the seizure is termed a *complex partial seizure*. An important additional subgroup comprises those seizures that begin as partial seizures and then spread diffusely throughout the cortex, i.e., *partial seizures with secondary generalization*.

Simple Partial Seizures Simple partial seizures cause motor, sensory, autonomic, or psychic symptoms without an obvious alteration in consciousness. For example, a patient having a partial motor seizure arising from the right primary motor cortex in the vicinity controlling hand movement will note the onset of involuntary movements of the contralateral, left hand. These movements are typically clonic (i.e., repetitive, flexion/extension movements) at a frequency of approximately 2 to 3 Hz; pure tonic posturing may be seen as well. Since the cortical region controlling hand movement is immediately adjacent to the region for facial expression, the seizure may also cause abnormal movements of the face synchronous with the movements of the hand. The electroencephalogram (EEG) recorded with scalp electrodes during the seizure (i.e., an ictal EEG) may show abnormal discharges in a very limited region over the appropriate area of cerebral cortex if the seizure focus involves the cerebral convexity (Chap. 357). Seizure activity occurring within deeper brain structures is often not recorded by the standard EEG, however, and may require intracranial electrodes for its detection.

Three additional features of partial motor seizures are worth noting. First, in some patients the abnormal motor movements may begin in a very restricted region such as the fingers and gradually progress (over seconds to minutes) to include a larger portion of the extremity. This phenomenon was originally described by Hughlings Jackson and is known as a "Jacksonian march," representing the spread of seizure activity over a progressively larger region of motor cortex. Second, patients may experience a localized paresis (Todd's paralysis) for minutes to many hours in the involved region following the seizure. Third, in rare instances the seizure may continue for hours or days. This condition, termed *epilepsia partialis continua*, is often quite refractory to medical therapy.

Other forms of simple partial seizures include those that cause changes in somatic sensation (e.g., paresthesias), vision (flashing lights or formed hallucinations), equilibrium (sensation of falling or vertigo), or autonomic function (flushing, sweating, piloerection). Simple partial seizures arising from the temporal or frontal cortex may also cause alterations in hearing, olfaction, or higher cortical function (psychic symptoms). This includes the sensation of unusual, intense odors (e.g., burning rubber or kerosene) or sounds (crude or highly complex sounds), or an epigastric sensation that rises from the stomach or chest to the head. Some patients describe odd, internal feelings such as fear, a sense of impending change, detachment, depersonalization, deja vu, or illusions that objects are growing smaller (micropsia) or larger (macropsia). When such symptoms precede a complex partial or secondarily generalized seizure, these simple

partial seizures serve as a warning, or *aura*.

Complex Partial Seizures Complex partial seizures are characterized by focal seizure activity accompanied by a transient impairment of the patient's ability to maintain normal contact with the environment. Operationally this means that the patient is unable to respond appropriately to visual or verbal commands during the seizure and has impaired recollection or awareness of the ictal phase. The seizures frequently begin with an aura (i.e., a simple partial seizure) that is stereotypic for the patient. The start of the ictal phase is often a sudden behavioral arrest or motionless stare, and this marks the onset of the event for which the patient will be amnestic. The behavioral arrest is usually accompanied by automatisms, which are involuntary, automatic behaviors that have a wide range of manifestations. Automatisms may consist of very basic behaviors such as chewing, lip smacking, swallowing, or "picking" movements of the hands, or more elaborate behaviors such as a display of emotion or running. The patient is typically confused following the seizure, and the transition to full recovery of consciousness may range from seconds up to an hour. Careful examination of the patient immediately following the seizure may show an anterograde amnesia or, in cases involving the dominant hemisphere, a postictal aphasia.

The routine, interictal (i.e., between seizures) EEG in patients with complex partial seizures is often normal, or may show brief discharges termed *epileptiform spikes*, or *sharp waves*. Since complex partial seizures can arise from the medial temporal lobe or inferior frontal lobe, i.e., regions distant from the scalp, the EEG recorded during the seizure may be nonlocalizing. However, the seizure focus is often detected using special electrodes such as sphenoidal or surgically placed intracranial electrodes.

The range of potential clinical behaviors linked to complex partial seizures is so broad that extreme caution is advised before concluding that stereotypic episodes of bizarre or atypical behavior are not due to seizure activity. In such cases it is imperative to consider more detailed EEG studies to determine whether the behaviors are caused by a seizure disorder.

Partial Seizures with Secondary Generalization Partial seizures can spread to involve both cerebral hemispheres and produce a generalized seizure, usually of the tonic-clonic variety (discussed below). Secondary generalization is observed frequently following simple partial seizures, especially those with a focus in the frontal lobe, but may also be associated with partial seizures occurring elsewhere in the brain. A partial seizure with secondary generalization is often difficult to distinguish from a primarily generalized tonic-clonic seizure, since bystanders tend to emphasize the more dramatic, generalized convulsive phase of the seizure and overlook the more subtle, focal symptoms present at onset. In some cases, the focal onset of the seizure becomes apparent only when a careful history identifies a preceding aura (i.e., simple partial seizure). Often, however, the focal onset is not clinically evident and may be established only through careful EEG analysis. Nonetheless, distinguishing between these two entities is extremely important, as there may be substantial differences in the evaluation and treatment of partial versus generalized seizure disorders.

GENERALIZED SEIZURES

By definition, generalized seizures arise from both cerebral hemispheres simultaneously. However, it is currently impossible to exclude entirely the existence of a focal region of abnormal activity that initiates the seizure prior to rapid secondary generalization. For this reason, generalized seizures may be practically defined as bilateral clinical and electrographic events without any detectable focal onset. Fortunately, a number of the subtypes of generalized seizures have distinctive features that facilitate clinical diagnosis.

Absence Seizures (Petit Mal) Absence seizures are characterized by sudden, brief lapses of consciousness without loss of postural control. The seizure typically lasts for only seconds, consciousness returns as suddenly as it was lost, and there is no postictal confusion. Although the brief loss of consciousness may be clinically inapparent or the sole manifestation of the seizure discharge, absence seizures are usually accompanied by subtle, bilateral motor signs such as rapid blinking of the eyelids, chewing movements, or small-amplitude, clonic movements of the hands.

Absence seizures usually begin in childhood (ages 4 to 8) or early adolescence and are the main seizure type in 15 to 20% of children with epilepsy. The seizures can occur hundreds of times per day, but the child may be unaware of or unable to convey their existence. This can lead to a situation in which the patient is constantly struggling to piece together experiences that have been interrupted by the seizures. Since the clinical signs of the seizures are subtle, especially to new parents, it is not surprising that the first clue to absence epilepsy is often unexplained "daydreaming" and a decline in school performance recognized by a teacher.

The electrophysiologic hallmark of typical absence seizures is a generalized, symmetric, 3-Hz spike-and-wave discharge that begins and ends suddenly on a normalEEGbackground (Chap. 357). Periods of spike-and-wave discharges lasting more than a few seconds usually correlate with the clinical signs, but the EEG often shows many more periods of abnormal cortical activity than were suspected clinically. Hyperventilation tends to provoke these electrographic discharges and even the seizures themselves and is routinely used when recording the EEG.

Typical absence seizures are often associated with generalized, tonic-clonic seizures, but patients usually have no other neurologic problems and respond well to treatment with specific anticonvulsants. Although estimates vary, approximately 60 to 70% of such patients will have a spontaneous remission during adolescence.

Atypical Absence Seizures Atypical absence seizures have features that deviate from both the clinical andEEGfeatures of typical absence seizures. For example, the lapse of consciousness is usually of longer duration and less abrupt in onset and cessation, and the seizure is accompanied by more obvious motor signs that may include focal or lateralizing features. The EEG shows a generalized, slow spike-and-wave pattern with a frequency of £2.5/s, as well as other abnormal activity. Atypical absence seizures are usually associated with diffuse or multifocal structural abnormalities of the brain and therefore may accompany other signs of neurologic dysfunction such as mental retardation. Furthermore, the seizures are less responsive to anticonvulsants compared to typical absence seizures.

Generalized, Tonic-Clonic Seizures (Grand Mal) Primarily generalized, tonic-clonic seizures are the main seizure type in approximately 10% of all persons with epilepsy. They are also the most common seizure type resulting from metabolic derangements and are therefore frequently encountered in many different clinical settings. The seizure usually begins abruptly without warning, although some patients describe vague premonitory symptoms in the hours leading up to the seizure. This prodrome should be distinguished from the stereotypic auras associated with focal seizures that secondarily generalize. The initial phase of the seizure is usually tonic contraction of muscles throughout the body, accounting for a number of the classic features of the event. Tonic contraction of the muscles of expiration and the larynx at the onset will produce a loud moan or cry. Respirations are impaired, secretions pool in the oropharynx, and the patient becomes cyanotic. Contraction of the jaw muscles may cause biting of the tongue. A marked enhancement of sympathetic tone leads to increases in heart rate, blood pressure, and pupillary size. After 10 to 20 s, the tonic phase of the seizure typically evolves into the clonic phase, produced by the superimposition of periods of muscle relaxation on the tonic muscle contraction. The periods of relaxation progressively increase until the end of the ictal phase, which usually lasts no more than 1 min. The postictal phase is characterized by unresponsiveness, muscular flaccidity, and excessive salivation that can cause stridorous breathing and partial airway obstruction. Bladder or bowel incontinence may occur at this point as well. Patients gradually regain consciousness over minutes to hours, and during this transition there is typically a period of postictal confusion. Patients will subsequently complain of headache, fatigue, and muscle ache that can last for many hours. The duration of impaired consciousness in the postictal phase can be extremely long, i.e., many hours, in patients with prolonged seizures or underlying CNS diseases such as alcoholic cerebral atrophy.

The EEG during the tonic phase of the seizure shows a progressive increase in generalized low-voltage fast activity, followed by generalized high-amplitude, polyspike discharges. In the clonic phase, the high-amplitude activity is typically interrupted by slow waves to create a spike-and-wave pattern. The postictal EEG shows diffuse slowing that gradually recovers as the patient awakens.

There are many variants of the generalized tonic-clonic seizure, including pure tonic and pure clonic seizures. Brief tonic seizures lasting only a few seconds are especially noteworthy since they are usually associated with known epileptic syndromes having mixed seizure phenotypes, such as the Lennox-Gastaut syndrome (discussed below).

Atonic Seizures Atonic seizures are characterized by sudden loss of postural muscle tone lasting 1 to 2 s. Consciousness is briefly impaired, but there is usually no postictal confusion. A very brief seizure may cause only a quick head drop or nodding movement, while a longer seizure will cause the patient to collapse. This can be quite dramatic and extremely dangerous, since there is a substantial risk of direct head injury with the fall. The EEG shows brief, generalized spike-and-wave discharges followed immediately by diffuse slow waves that correlate with the loss of muscle tone. Similar to pure tonic seizures, atonic seizures are usually seen in association with known epileptic syndromes.

Myoclonic Seizures Myoclonus is a sudden and brief muscle contraction that may

involve one part of the body or the entire body. A normal, common physiologic form of myoclonus is the sudden jerking movement observed while falling asleep. Pathologic myoclonus is most commonly seen in association with metabolic disorders, degenerativeCNSdiseases, or anoxic brain injury (Chap. 22). Although the distinction from other forms of myoclonus is imprecise, myoclonic seizures are considered to be true epileptic events since they are caused by cortical (versus subcortical or spinal) dysfunction. TheEEGshows bilaterally synchronous spike-and-wave discharges. Myoclonic seizures usually coexist with other forms of generalized seizure disorders but are the predominant feature of juvenile myoclonic epilepsy (discussed below).

UNCLASSIFIED SEIZURES

Not all seizure types can be classified as partial or generalized. This appears to be especially true of seizures that occur in neonates and infants. The distinctive phenotypes of seizures at these early ages likely result, in part, from differences in neuronal function and connectivity in the immature versus matureCNS.

EPILEPSY SYNDROMES

In addition to recognizing the patterns of different types of seizures, it is also useful to be familiar with some of the more common epilepsy syndromes, since this often helps in the determination of therapy and prognosis. Epilepsy syndromes are disorders in which epilepsy is a predominant feature, and there is sufficient evidence (e.g., through clinical,EEG, radiologic, or genetic observations) to suggest a common underlying mechanism. Three important epilepsy syndromes are listed below; additional examples with a known genetic basis are shown in Table 360-2.

JUVENILE MYOCLONIC EPILEPSY

Juvenile myoclonic epilepsy (JME) is a generalized seizure disorder of unknown cause that appears in early adolescence and is usually characterized by bilateral myoclonic jerks that may be single or repetitive. The myoclonic seizures are most frequent in the morning after awakening and can be provoked by sleep deprivation. Consciousness is preserved unless the myoclonus is especially severe. Many patients also experience generalized tonic-clonic seizures, and up to one-third have absence seizures. The condition is otherwise benign, and although complete remission is uncommon, the seizures respond well to appropriate anticonvulsant medication. There is often a family history of epilepsy, and genetic linkage studies suggest a polygenic cause.

LENNOX-GASTAUT SYNDROME

Lennox-Gastaut syndrome occurs in children and is defined by the following triad: (1) multiple seizure types (usually including generalized tonic-clonic, atonic, and atypical absence seizures); (2) anEEGshowing slow (<3 Hz) spike-and-wave discharges and a variety of other abnormalities; and (3) impaired cognitive function in most but not all cases. Lennox-Gastaut syndrome is associated withCNSdisease or dysfunction from a variety of causes, including developmental abnormalities, perinatal hypoxia/ischemia, trauma, infection, and other acquired lesions. The multifactorial nature of this syndrome suggests that it is a nonspecific response of the brain to diffuse neural injury.

Unfortunately, many patients have a poor prognosis due to the underlying CNS disease and the physical and psychosocial consequences of severe, poorly controlled epilepsy.

MESIAL TEMPORAL LOBE EPILEPSY SYNDROME

Mesial temporal lobe epilepsy (MTLE) is the most common syndrome associated with complex partial seizures and is an example of a symptomatic, partial epilepsy. Distinctive clinical, electroencephalographic, and pathologic features define this syndrome (Table 360-3). High-resolution magnetic resonance imaging (MRI) can detect the characteristic hippocampal sclerosis that appears to be an essential element in the pathophysiology of MTLE for many patients (Fig. 360-1). Recognition of this syndrome is especially important because it tends to be refractory to treatment with anticonvulsants but responds extremely well to surgical intervention. Major advances in the understanding of basic mechanisms of epilepsy have come through studies of experimental models of MTLE, discussed below.

THE CAUSES OF SEIZURES AND EPILEPSY

Seizures are a result of a shift in the normal balance of excitation and inhibition within theCNS. Given the numerous properties that control neuronal excitability, it is not surprising that there are many different ways to perturb this normal balance, and therefore many different causes of both seizures and epilepsy. Our understanding of the basic mechanisms involved remains very limited, and consequently there is not a rigorous, mechanistic-based framework for organizing all the etiologies. Conceptually, however, three important clinical observations emphasize how a variety of factors determine why certain conditions may cause seizures or epilepsy in a given patient.

1. *The normal brain is capable of having a seizure under the appropriate circumstances, and there are differences between individuals in the susceptibility or threshold for seizures*. For example, seizures may be induced by high fevers in children who are otherwise normal and who never develop other neurologic problems, including epilepsy. However, febrile seizures occur only in a relatively small proportion of children. This implies there are various underlying, *endogenous factors* that influence the threshold for having a seizure. Some of these factors are clearly genetic, as it has been shown that a family history of epilepsy will influence the likelihood of seizures occurring in otherwise normal individuals. Normal development also plays an important role, since the brain appears to have different seizure thresholds at different maturational stages.

2. *There are a variety of conditions that have an extremely high likelihood of resulting in a chronic seizure disorder*. One of the best examples of this is severe, penetrating head trauma, which is associated with up to a 50% risk of subsequent epilepsy. The high propensity for severe traumatic brain injury to lead to epilepsy suggests that the injury results in a long-lasting, pathologic change in theCNS that transforms a presumably normal neural network into one that is abnormally hyperexcitable. This process is known as *epileptogenesis*, and the specific changes that result in a lowered seizure threshold can be considered *epileptogenic factors*. Other processes associated with epileptogenesis include stroke, infections, and abnormalities of CNS development. Likewise, the genetic abnormalities associated with epilepsy likely involve processes that trigger the appearance of specific sets of epileptogenic factors.

3. *Seizures are episodic.* Patients with epilepsy have seizures intermittently and, depending on the underlying cause, many patients are completely normal for months or even years between seizures. This implies there are important provocative or *precipitating factors* that induce seizures in patients with epilepsy. Similarly, precipitating factors are responsible for causing the single seizure in someone without epilepsy. Precipitants include those due to intrinsic physiologic processes, such as psychological or physical stress, sleep deprivation, or hormonal changes associated with the menstrual cycle. They also include exogenous factors such as exposure to toxic substances and certain medications.

These observations emphasize the concept that the many causes of seizures and epilepsy result from a dynamic interplay between endogenous factors, epileptogenic factors, and precipitating factors. The potential role of each needs to be carefully considered when determining the appropriate management of a patient with seizures. For example, the identification of predisposing factors (e.g., family history of epilepsy) in a patient with febrile seizures may increase the necessity for closer follow-up and a more aggressive diagnostic evaluation. Finding an epileptogenic lesion may help in the estimation of seizure recurrence and duration of therapy. Finally, removal or modification of a precipitating factor may be an effective and safer method for preventing further seizures than the prophylactic use of anticonvulsant drugs.

CAUSES ACCORDING TO AGE

In practice, it is useful to consider the etiologies of seizures based on the age of the patient, as age is one of the most important factors determining both the incidence and likely causes of seizures or epilepsy (Table 360-4). During the *neonatal period and early infancy*, potential causes include hypoxic-ischemic encephalopathy, trauma,CNSinfection, congenital CNS abnormalities, and metabolic disorders. Babies born to mothers using neurotoxic drugs such as cocaine, heroin, or ethanol are susceptible to drug-withdrawal seizures in the first few days after delivery. Hypoglycemia and hypocalcemia, which can occur as secondary complications of perinatal injury, are also causes of seizures early after delivery. Seizures due to inborn errors of metabolism usually present once regular feeding begins, typically 2 to 3 days after birth. Pyridoxine (vitamin B$_6$) deficiency, an important cause of neonatal seizures, can be effectively treated with pyridoxine replacement. The idiopathic or inherited forms of benign neonatal convulsions are also seen during this time period.

The most common seizures arising in *late infancy and early childhood* are febrile seizures, which are seizures associated with fevers but without evidence ofCNSinfection or other defined causes. The overall prevalence is 3 to 5% and even higher in some parts of the world, such as Asia. Patients often have a family history of febrile seizures or epilepsy. Febrile seizures usually occur between 3 months and 5 years of age and have a peak incidence between 18 and 24 months. The typical scenario is a child who has a generalized, tonic-clonic seizure during a febrile illness in the setting of a common childhood infection such as otitis media, respiratory infection, or gastroenteritis. The seizure is likely to occur during the rising phase of the temperature curve (i.e., during the first day) rather than well into the course of the illness. A *simple* febrile seizure is a single, isolated event, brief, and symmetric in appearance. *Complex* febrile seizures

have repeated seizure activity, last>15 min, or have focal features. Approximately one-third of patients with febrile seizures will have a recurrence, but<10% have three or more episodes. Recurrences are much more likely when the febrile seizure occurs in the first year of life. Simple febrile seizures are not associated with an increase in the risk of developing epilepsy, while complex febrile seizures have a risk of 2 to 5%; other risk factors include the presence of preexisting neurologic deficits and a family history of nonfebrile seizures.

Childhood marks the age at which many of the well-defined epilepsy syndromes present. Some children who are otherwise normal develop idiopathic, generalized tonic-clonic seizures without other features that fit into specific syndromes. Temporal lobe epilepsy usually presents in childhood and may be related to mesial temporal lobe sclerosis (as part of the MTLE syndrome) or other focal abnormalities such as cortical dysgenesis. Other types of partial seizures, including those with secondary generalization, may be the relatively late manifestation of a developmental disorder, an acquired lesion such as head trauma, CNS infection (especially viral encephalitis), or very rarely a CNS tumor.

The period of *adolescence and early adulthood* is one of transition during which the idiopathic or genetically based epilepsy syndromes, including JME and juvenile absence epilepsy, become less common, while epilepsies secondary to acquired CNS lesions begin to predominate. Seizures that begin in patients in this age range may be associated with head trauma, CNS infections (including parasitic infections such as cysticercosis), brain tumors, congenital CNS abnormalities, illicit drug use, or alcohol withdrawal.

Head trauma is a common cause of epilepsy in adolescents and adults. The head injury can be caused by a variety of mechanisms, and the likelihood of developing epilepsy is strongly correlated with the severity of the injury. A patient with a penetrating head wound, depressed skull fracture, intracranial hemorrhage, or prolonged posttraumatic coma or amnesia has a 40 to 50% risk of developing epilepsy, while a patient with a closed head injury and cerebral contusion has a 5 to 25% risk. Recurrent seizures usually develop within 1 year after head trauma, although intervals of 10 years or longer are well known. In controlled studies, mild head injury, defined as a concussion with amnesia or loss of consciousness of <30 min, was not found to be associated with an increased likelihood of epilepsy. Nonetheless, most epileptologists know of patients who have partial seizures within hours or days of a mild head injury and subsequently develop chronic seizures of the same type; such cases may represent rare examples of chronic epilepsy resulting from mild head injury.

The causes of seizures in *older adults* include cerebrovascular disease, trauma (including subdural hematoma), CNS tumors, and degenerative diseases. Cerebrovascular disease may account for approximately 50% of new cases of epilepsy in patients older than 65. Acute seizures (i.e., occurring at the time of the stroke) are seen more often with embolic rather than hemorrhagic or thrombotic stroke. Chronic seizures typically appear months to years after the initial event and are associated with all forms of stroke.

Metabolic disturbances such as electrolyte imbalance, hypo- or hyperglycemia, renal

failure, and hepatic failure may cause seizures at any age. Similarly, endocrine disorders, hematologic disorders, vasculitides, and many other systemic diseases may cause seizures over a broad age range. A wide variety of medications and abused substances are known to precipitate seizures as well (Table 360-5).

BASIC MECHANISMS

MECHANISMS OF SEIZURE INITIATION AND PROPAGATION

Partial seizure activity can begin in a very discrete region of cortex and then spread to neighboring regions, i.e., there is a *seizure initiation* phase and a *seizure propagation* phase. Studies of experimental models of these phases suggest that the initiation phase is characterized by two concurrent events in an aggregate of neurons: (1) high-frequency bursts of action potentials, and (2) hypersynchronization. The bursting activity is caused by a relatively long-lasting depolarization of the neuronal membrane due to influx of extracellular calcium (Ca_{2+}), which leads to the opening of voltage-dependent sodium (Na_+) channels, influx of Na_+, and generation of repetitive action potentials. This is followed by a hyperpolarizing afterpotential mediated byg-aminobutyric acid (GABA) receptors or potassium (K_+) channels, depending on the cell type. The synchronized bursts from a sufficient number of neurons result in a so-called spike discharge on theEEG.

Normally, the spread of bursting activity is prevented by intact hyperpolarization and a region of surrounding inhibition created by inhibitory neurons. With sufficient activation there is a recruitment of surrounding neurons via a number of mechanisms. Repetitive discharges lead to the following: (1) an increase in extracellular K_+, which blunts the extent of hyperpolarization and depolarizes neighboring neurons; (2) accumulation of Ca_{2+} in presynaptic terminals, leading to enhanced neurotransmitter release; and (3) depolarization-induced activation of the *N*-methyl-D-aspartate (NMDA) subtype of the excitatory amino acid receptor, which causes more Ca_{2+}influx and neuronal activation. The recruitment of a sufficient number of neurons leads to a loss of the surrounding inhibition and propagation of seizure activity into contiguous areas via local cortical connections, and to more distant areas via long commissural pathways such as the corpus callosum.

Many factors control neuronal excitability, and thus there are many potential mechanisms for altering a neuron's propensity to have bursting activity. Examples of mechanisms *intrinsic* to the neuron include changes in the conductance of ion channels, response characteristics of membrane receptors, cytoplasmic buffering, second-messenger systems, and protein expression as determined by gene transcription, translation, and posttranslational modification. Mechanisms *extrinsic* to the neuron include changes in the amount or type of neurotransmitters present at the synapse, modulation of receptors by extracellular ions and other molecules, and temporal and spatial properties of both synaptic and nonsynaptic input. Nonneural cells, such as astrocytes and oligodendrocytes, have an important role in many of these mechanisms as well.

Certain known causes of seizures are explained by these mechanisms. For example, accidental ingestion of domoic acid, which is an analogue of glutamate (the principal

excitatory neurotransmitter in the brain), causes profound seizures via direct activation of excitatory amino acid receptors throughout the CNS. Penicillin, which can lower the seizure threshold in humans and is a potent convulsant in experimental models, reduces inhibition by antagonizing the effects of GABA at its receptor. The basic mechanisms of other precipitating factors of seizures, such as sleep deprivation, fever, alcohol withdrawal, hypoxia, and infection, are not as well understood but presumably involve analogous perturbations in neuronal excitability. Similarly, the endogenous factors that determine an individual's seizure threshold may relate to these properties as well.

Knowledge of the mechanisms responsible for the initiation and propagation of most generalized seizures (including tonic-clonic, myoclonic, and atonic types) remains rudimentary and reflects the limited understanding of the connectivity of the brain at a systems level. Much more is understood about the origin of generalized spike-and-wave discharges in absence seizures. These appear to be related to oscillatory rhythms that are normally generated during sleep by circuits connecting the thalamus and cortex. This oscillatory behavior involves an interaction between $GABA_B$ receptors, T-type Ca_{2+} channels, and K_+ channels located within the thalamus. Pharmacologic studies indicate that modulation of these receptors and channels can induce absence seizures, and there is speculation that the genetic forms of absence epilepsy may be associated with mutations of components of this system.

MECHANISMS OF EPILEPTOGENESIS

Epileptogenesis refers to the transformation of a normal neuronal network into one that is chronically hyperexcitable. For example, there is often a delay of months to years between an initial CNS injury such as trauma, stroke, or infection and the first seizure. The injury appears to initiate a process that gradually lowers the seizure threshold in the affected region until a spontaneous seizure occurs. In many genetic and idiopathic forms of epilepsy, epileptogenesis is presumably determined by developmentally regulated events.

Pathologic studies of the hippocampus from patients with temporal lobe epilepsy have led to the suggestion that some forms of epileptogenesis are related to *structural changes in neuronal networks*. For example, many patients with MTLE syndrome have a highly selective loss of neurons that has been proposed to contribute to inhibition of the main excitatory neurons within the dentate gyrus. There is also evidence that, in response to the loss of neurons, there is reorganization or "sprouting" of surviving neurons in a way that affects the excitability of the network. Some of these changes can be seen in experimental models of prolonged electrical seizures or traumatic brain injury. Thus, an initial injury such as head injury may lead to a very focal, confined region of structural change that causes local hyperexcitability. The local hyperexcitability leads to further structural changes that evolve over time until the focal lesion produces clinically evident seizures. Similar models have also provided strong evidence for long-term alterations in *intrinsic, biochemical properties of cells* within the network, such as chronic changes in glutamate receptor function.

GENETIC CAUSES OF EPILEPSY

The most important recent progress in epilepsy research has been the identification of genetic mutations associated with a variety of epilepsy syndromes.Table 360-2 describes some of these in further detail. Although all of the mutations identified to date cause rare forms of epilepsy, they have led to extremely important conceptual advances. For example, it appears that many of the inherited, idiopathic epilepsies (i.e., the relatively "pure" forms of epilepsy in which seizures are the phenotypic abnormality and brain structure and function are otherwise normal) are due to mutations affecting ion channel function. These syndromes are therefore part of the larger group of "channelopathies" causing paroxysmal disorders such as cardiac arrhythmias, episodic ataxia, periodic weakness, and familial hemiplegic migraine (Chap. 15). In contrast, gene mutations observed in symptomatic epilepsies (i.e., disorders in which other neurologic abnormalities, such as cognitive impairment, coexist with seizures) are proving to be associated with pathways influencingCNSdevelopment or neuronal homeostasis. A current challenge is to identify the multiple susceptibility genes that underlie the more common forms of idiopathic epilepsies.

MECHANISMS OF ACTION OF ANTIEPILEPTIC DRUGS

Currently available antiepileptic drugs appear to act primarily by blocking the initiation or spread of seizures. This occurs through a variety of mechanisms, and in most cases the drugs have pleiotropic effects. The mechanisms include inhibition of Na_+-dependent action potentials in a frequency-dependent manner (e.g., phenytoin, carbamazepine, topiramate, zonisamide), inhibition of voltage-gated Ca_{2+}channels (phenytoin), decrease of glutamate release (lamotrigine), potentiation of GABA receptor function (benzodiazepines and barbiturates), and increase in the availability ofGABA(valproic acid, gabapentin, tiagabine). The two most effective drugs for absence seizures, ethosuximide and valproic acid, probably act by inhibiting T-type Ca_{2+}channels in thalamic neurons.

In contrast to the relatively large number of antiepileptic drugs that can attenuate seizure activity, there are currently no drugs known to prevent the formation of a seizure focus followingCNSinjury in humans. The eventual development of such "antiepileptogenic" drugs will provide an important means of preventing the emergence of epilepsy following injuries such as head trauma, stroke, and CNS infection.

EVALUATION OF THE PATIENT WITH A SEIZURE

When a patient presents shortly after a seizure, the first priorities are attention to vital signs, respiratory and cardiovascular support, and treatment of seizures if they resume (see "Treatment"). Life-threatening conditions such asCNSinfection, metabolic derangement or drug toxicity must be recognized and managed appropriately.

When the patient is not acutely ill, the evaluation will initially focus on whether or not there is a history of earlier seizures (Fig. 360-2). If this is the patient's first seizure, then the emphasis will be to (1) establish whether the reported episode was a seizure rather than another paroxysmal event, (2) determine the cause of the seizure by identifying risk factors and precipitating events, and (3) decide whether anticonvulsant therapy is required in addition to treatment for any underlying illness.

In the patient with prior seizures or a known history of epilepsy, the evaluation is directed toward (1) identification of the underlying cause and precipitating factors, and (2) determination of the adequacy of the patient's current therapy.

HISTORY AND EXAMINATION

The history should first determine whether the event was truly a seizure. It is essential to take the time to gather an in-depth history, for *in many cases the diagnosis of a seizure is based solely on clinical grounds -- the examination and laboratory studies are often normal*. Keeping in mind the characteristics of different seizure types, questions need to focus precisely on the symptoms before, during, and after the episode in order to discriminate a seizure from other paroxysmal events (see "Differential Diagnosis of Seizures"). Seizures frequently occur out-of-hospital, and the patient may be unaware of the ictal and immediate postictal phases; thus witnesses to the event should be interviewed carefully.

The history should also focus on risk factors and predisposing events. Clues for a predisposition to seizures include a history of febrile seizures, earlier auras or brief seizures not recognized as such, and a family history of seizures. Epileptogenic factors such as prior head trauma, stroke, tumor, or vascular malformation should be identified. In children, a careful assessment of developmental milestones may provide evidence for underlyingCNSdisease. Precipitating factors such as sleep deprivation, systemic diseases, electrolyte or metabolic derangements, acute infection, drugs that lower the seizure threshold (Table 360-5), or alcohol or illicit drug use should also be identified.

The general physical examination includes a search for signs of infection or systemic illness. Careful examination of the skin may reveal signs of neurocutaneous disorders, such as tuberous sclerosis or neurofibromatosis, or chronic liver or renal disease. A finding of organomegaly may indicate a metabolic storage disease, and limb asymmetry may provide a clue for brain injury early in development. Signs of head trauma and use of alcohol or illicit drugs should be sought. Auscultation of the heart and carotid arteries may identify an abnormality that predisposes to cerebrovascular disease.

All patients require a complete neurologic examination, with particular emphasis on eliciting signs of cerebral hemispheric disease (Chap. 356). Careful assessment of mental status (including memory, language function, and abstract thinking) may suggest lesions in the anterior frontal, parietal, or temporal lobes. Testing of visual fields will help screen for lesions in the optic pathways and occipital lobes. Screening tests of motor function such as pronator drift, deep tendon reflexes, gait, and coordination may suggest lesions in motor (frontal) cortex, and cortical sensory testing (e.g., double simultaneous stimulation) may detect lesions in the parietal cortex.

LABORATORY STUDIES

Routine blood studies are indicated to identify the more common metabolic causes of seizures, such as abnormalities in electrolytes, glucose, calcium, or magnesium, and hepatic or renal disease. A screen for toxins in blood and urine should also be obtained from all patients in the appropriate risk groups, especially when no clear precipitating factor has been identified. A lumbar puncture is indicated if there is any suspicion of

meningitis or encephalitis and is mandatory in all patients infected with HIV, even in the absence of symptoms or signs suggesting infection.

All patients who have a possible seizure disorder should be evaluated with an EEG (Chap. 357) as soon as possible. The EEG may help to establish the diagnosis of epilepsy, classify the seizure type, and provide evidence for the existence of a particular epilepsy syndrome. If the patient is having frequent seizures, such as a child with absence epilepsy, the EEG may confirm the presence of seizures and help to identify the seizure type. In patients with infrequent seizures, the EEG may reveal potentially abnormal interictal activity that, when combined with clinical or radiologic data, aids in establishing the diagnosis. However, the existence of epileptiform patterns such as spikes or sharp waves are not diagnostic in themselves, since similar patterns can be seen in 1 to 2% of normal individuals. Ideally, the EEG should be performed after sleep deprivation to increase the potential diagnostic yield of the study.

Almost all patients with new-onset seizures should have a brain imaging study to determine whether there is an underlying structural abnormality that is responsible. The main exception to this rule is children who have an unambiguous history and examination suggestive of a benign, generalized seizure disorder such as absence epilepsy. MRI has been shown to be superior to computed tomography (CT) in scanning for the detection of cerebral lesions associated with epilepsy. In some cases MRI will identify lesions such as tumors, vascular malformations, or other pathologies that need immediate therapy. The use of newer MRI methods, such as fluid-attenuated inversion recovery (FLAIR), has increased the sensitivity for detection of abnormalities of cortical architecture, including hippocampal atrophy associated with mesial temporal sclerosis, and abnormalities of cortical neuronal migration. In such cases the findings may not lead to immediate therapy, but they do provide an explanation for the patient's seizures and point to the need for chronic anticonvulsant therapy or possible surgical resection.

In the patient with suspected CNS infection or mass lesions, CT scanning should be performed emergently when MRI is not immediately available. Otherwise, it is usually appropriate to obtain an MRI study within a few days of the initial evaluation. Functional imaging procedures such as positron emission tomography (PET) and single photon emission computed tomography (SPECT) are also used to evaluate certain patients with medically refractory seizures (discussed below).

DIFFERENTIAL DIAGNOSIS OF SEIZURES

The various disorders that may mimic seizures are listed in Table 360-6. In most cases seizures can be distinguished from these other conditions by meticulous attention to the history and relevant laboratory studies. On occasion, additional studies, such as video-EEG monitoring, sleep studies, tilt table analysis, or cardiac electrophysiology, may be required to reach a correct diagnosis. Two of the more common syndromes in the differential diagnosis are detailed below.

Syncope The diagnostic dilemma encountered most frequently is the distinction between a generalized seizure and syncope. Observations by the patient and bystanders that can help discriminate between the two are listed in Table 360-7. Characteristics of a seizure include the presence of an aura, cyanosis,

unconsciousness, motor manifestations lasting more than 30 s, postictal disorientation, muscle soreness, and sleepiness. In contrast, a syncopal episode is more likely if the event was provoked by acute pain or anxiety or occurred immediately after arising from the lying or sitting position. Patients with syncope often describe a stereotyped transition from consciousness to unconsciousness that includes tiredness, sweating, nausea, and tunneling of vision, and they experience a relatively brief loss of consciousness. Headache or incontinence usually suggests a seizure but may on occasion also occur with syncope. A brief period (i.e., 1 to 10 s) of convulsive motor activity is frequently seen immediately at the onset of a syncopal episode, especially if the patient remains in an upright posture after fainting (e.g., in a dentist's chair) and therefore has a sustained decrease in cerebral perfusion. Rarely, a syncopal episode can induce a full tonic-clonic seizure. In such cases the evaluation must focus on both the cause of the syncopal event as well as the possibility that the patient has a propensity for recurrent seizures.

Psychogenic Seizures Psychogenic seizures are nonepileptic behaviors that resemble seizures. The behavior is often part of a conversion reaction precipitated by underlying psychological distress. Certain behaviors, such as side-to-side turning of the head, asymmetric and large-amplitude shaking movements of the limbs, twitching of all four extremities without loss of consciousness, pelvic thrusting, and crying or talking during the event, are more commonly associated with psychogenic rather than epileptic seizures. However, the distinction is sometimes difficult on clinical grounds alone, and there are many examples of diagnostic errors made by experienced epileptologists. This is especially true for psychogenic seizures that resemble complex partial seizures, since the behavioral manifestations of complex partial seizures (especially of frontal lobe origin) can be extremely unusual, and in both cases the routine surfaceEEG may be normal. Video-EEG monitoring is often useful when the clinical observations are nondiagnostic. Generalized tonic-clonic seizures always produce marked EEG abnormalities during and after the seizure. For suspected complex partial seizures of temporal lobe origin, the use of additional electrodes beyond the standard scalp locations (e.g., sphenoidal electrodes) may be required to localize a seizure focus. Measurement of serum prolactin levels may also help to discriminate between organic and psychogenic seizures, since most generalized seizures and many complex partial seizures are accompanied by rises in serum prolactin (during the immediate 30-min postictal period), whereas psychogenic seizures are not. It is important to note that the diagnosis of psychogenic seizures does not exclude a concurrent diagnosis of epilepsy, since the two often coexist.

TREATMENT

Therapy for a patient with a seizure disorder is almost always multimodal and includes treatment of underlying conditions that cause or contribute to the seizures, avoidance of precipitating factors, suppression of recurrent seizures by prophylactic therapy with antiepileptic medications or surgery, and addressing a variety of psychological and social issues. Treatment plans must be individualized, given the many different types and causes of seizures as well as the differences in efficacy and toxicity of antiepileptic medications for each patient. In almost all cases a neurologist with experience in the treatment of epilepsy should design and oversee implementation of the treatment strategy. Furthermore, patients with refractory epilepsy or those who require polypharmacy with antiepileptic drugs should remain under the regular care of a

neurologist.

Treatment of Underlying Conditions If the sole cause of a seizure is a metabolic disturbance such as an abnormality of serum electrolytes or glucose, then treatment is aimed at reversing the metabolic problem and preventing its recurrence. Therapy with antiepileptic drugs is usually unnecessary unless the metabolic disorder cannot be corrected promptly and the patient is at risk of having further seizures. If the apparent cause of a seizure was a medication (e.g., theophylline) or illicit drug use (e.g., cocaine), then appropriate therapy is avoidance of the drug and there is usually no need for antiepileptic medications unless subsequent seizures occur in the absence of these precipitants.

Seizures caused by a structural CNS lesion such as a brain tumor, vascular malformation, or brain abscess may not recur after appropriate treatment of the underlying lesion. However, despite removal of the structural lesion, there is a risk that the seizure focus will remain in the surrounding tissue or develop de novo as a result of gliosis and other processes induced by surgery, radiation, or other therapies. Most patients are therefore maintained on an antiepileptic medication for at least 1 year, and an attempt is made to withdraw medications only if the patient has been completely seizure-free. If the seizures are refractory to medication, the patient may benefit from surgical removal of the epileptic brain region (see "Surgical Treatment of Refractory Epilepsy").

Avoidance of Precipitating Factors Unfortunately, little is known about the specific factors that determine precisely when a seizure will occur in a patient with epilepsy. Some patients can identify particular situations that appear to lower their seizure threshold; these situations should be avoided. For example, a patient who has seizures in the setting of sleep deprivation should obviously be advised to maintain a normal sleep schedule. Many patients note an association between alcohol intake and seizures, and they should be encouraged to modify their drinking habits accordingly. There are also relatively rare cases of patients with seizures that are induced by highly specific stimuli such as a video game monitor, music, or an individual's voice ("reflex epilepsy"). If there is an association between stress and seizures, stress reduction techniques such as physical exercise, meditation, or counseling may be helpful.

Antiepileptic Drug Therapy Antiepileptic drug therapy is the mainstay of treatment for most patients with epilepsy. The overall goal is to completely prevent seizures without causing any untoward side effects, preferably with a single medication and a dosing schedule that is easy for the patient to follow. Seizure classification is an important element in designing the treatment plan, since some antiepileptic drugs have different activities against various seizure types. However, there is considerable overlap between many antiepileptic drugs, such that the choice of therapy is often determined more by specific needs of the patient, especially the patient's assessment of side effects.

When to Initiate Antiepileptic Drug Therapy Antiepileptic drug therapy should be started in any patient with recurrent seizures of unknown etiology or a known cause that cannot be reversed. Whether to initiate therapy in a patient with a single seizure is controversial. Patients with a single seizure due to an identified lesion such as a CNS tumor, infection, or trauma, in which there is strong evidence that the lesion is

epileptogenic, should be treated. The risk of seizure recurrence in a patient with an apparently unprovoked or idiopathic seizure is uncertain, with estimates ranging from 31 to 71% in the first 12 months after the initial seizure. This uncertainty arises from differences in the underlying seizure types and etiologies in various published epidemiologic studies. Generally accepted risk factors associated with recurrent seizures include the following: (1) an abnormal neurologic examination, (2) seizures presenting as status epilepticus, (3) postictal Todd's paralysis, (4) a strong family history of seizures, or (5) an abnormal EEG. Most patients with one or more of these risk factors should be treated. Issues such as employment or driving may influence the decision whether or not to start medications as well. For example, a patient with a single, idiopathic seizure and whose job depends on driving may prefer taking antiepileptic drugs rather than risking a seizure recurrence and the potential loss of driving privileges.

Selection of Antiepileptic Drugs The choices of antiepileptic drugs in the United States for different seizure types are shown in Table 360-8, and the main pharmacologic characteristics of commonly used drugs are listed in Table 360-9. Older medications such as phenytoin, valproic acid, carbamazepine, and ethosuximide are generally used as first-line therapy for most seizure disorders since, overall, they are as effective as recently marketed drugs and significantly less expensive. Of the new drugs that have become available in the United States in the past decade, most are currently being used as add-on or alternative therapy.

In addition to efficacy, other factors influencing the specific choice of an initial medication for a patient include the relative convenience of dosing schedule (e.g., once daily versus three or four times daily) and potential side effects. Almost all of the commonly used antiepileptic drugs can cause similar, dose-related side effects such as sedation, ataxia, and diplopia. Close follow-up is required to ensure these are promptly recognized and reversed. Most of the drugs may also cause idiosyncratic toxicity such as rash, bone marrow suppression, or hepatotoxicity. Although rare, these side effects need to be carefully considered during drug selection, and patients require laboratory tests (e.g., complete blood count and liver function tests) prior to the institution of a drug (to establish baseline values) and during initial dosing and titration of the agent.

ANTIEPILEPTIC DRUG SELECTION FOR PARTIAL SEIZURES Carbamazepine or phenytoin is currently the initial drug of choice for the treatment of partial seizures, including those that secondarily generalize. Overall they have very similar efficacy, but differences in pharmacokinetics and toxicity are the main determinants for use in a given patient. Phenytoin has a relatively long half-life and offers the advantage of once or twice daily dosing compared to two or three times daily dosing for carbamazepine (although a more expensive, extended-release form of carbamazepine is now available). An advantage of carbamazepine is that its metabolism follows first-order pharmacokinetics, and the relationship between drug dose, serum levels, and toxicity is linear. By contrast, phenytoin shows properties of saturation kinetics, such that small increases in phenytoin doses above a standard maintenance dose can precipitate marked side effects. This is one of the main causes of acute phenytoin toxicity. Long-term use of phenytoin is associated with untoward cosmetic effects (e.g., hirsutism, coarsening of facial features, and gingival hypertrophy), so it is often avoided in young patients who are likely to require the drug for many years. Carbamazepine can cause leukopenia, aplastic anemia, or hepatotoxicity and would therefore be

contraindicated in patients with predispositions to these problems.

Valproic acid is an effective alternative for some patients with partial seizures, especially when the seizures secondarily generalize. Gastrointestinal side effects are fewer when using the valproate semisodium formulation (Depakote). Valproic acid also rarely causes reversible bone marrow suppression and hepatotoxicity, and laboratory testing is required to monitor toxicity. This drug should generally be avoided in patients with preexisting bone marrow or liver disease. Irreversible, fatal hepatic failure appearing as an idiosyncratic rather than dose-related side effect is a relatively rare complication; its risk is highest in children<2 years old, especially those taking other antiepileptic drugs or with inborn errors of metabolism. Valproic acid therapy should therefore only be used in infants and young children when the benefits clearly exceed this risk.

Lamotrigine, gabapentin, topiramate, tiagabine, and phenobarbital are additional drugs currently used for the treatment of partial seizures with or without secondary generalization. Lamotrigine appears to have an overall efficacy profile similar to the more standard drugs and is now being used as monotherapy. All patients, particularly children, need to be monitored closely for a lamotrigine-induced rash during the initiation of therapy. Also, lamotrigine must be started very slowly when used as add-on therapy with valproic acid, since valproic acid can inhibit its metabolism, thereby substantially prolonging its half-life. Gabapentin is unique in not having any significant drug interactions. This makes it potentially useful as add-on therapy, especially in patients who are particularly susceptible to side effects of other medications. Until recently, phenobarbital and other barbiturate compounds were commonly used as first-line therapy for many forms of epilepsy. However, the barbiturates frequently cause sedation in adults, hyperactivity in children, and other more subtle cognitive changes; thus, their use should be limited to situations in which no other suitable treatment alternatives exist.

ANTIEPILEPTIC DRUG SELECTION FOR GENERALIZED SEIZURES Valproic acid is currently considered the best initial choice for the treatment of primarily generalized, tonic-clonic seizures and lamotrigine, followed by carbamazepine and phenytoin, are suitable alternatives. Valproic acid is also particularly effective in absence, myoclonic, and atonic seizures and is therefore the drug of choice in patients with generalized epilepsy syndromes having mixed seizure types. Ethosuximide remains the preferred drug for the treatment of uncomplicated absence seizures, but it is not effective against tonic-clonic or partial seizures. Ethosuximide rarely causes bone marrow suppression, so that periodic monitoring of blood cell counts is required. Although approved for use in partial seizure disorders, lamotrigine appears to be effective in epilepsy syndromes with mixed, generalized seizure types such as JME and Lennox-Gastaut syndrome.

Initiation and Monitoring of Therapy Because the response to any antiepileptic drug is unpredictable, patients should be carefully educated about the approach to therapy. Patients need to understand that the goal is to prevent seizures and minimize the side effects of therapy; determination of the optimal dose is often a matter of trial and error. This process may take months or longer if the baseline seizure frequency is low. Most anticonvulsant drugs need to be introduced relatively slowly to minimize side effects, and patients should expect that minor side effects such as mild sedation, slight changes in cognition, or imbalance will typically resolve within a few days. Starting doses are

usually the lowest value listed under the dosage column in<u>Table 360-9</u>. Subsequent increases should be made only after achieving a steady state with the previous dose (i.e., after an interval of five or more half-lives).

Monitoring of serum antiepileptic drug levels can be very useful for establishing the initial dosing schedule. However, the published therapeutic ranges of serum drug concentrations are only an approximate guide for determining the proper dose for a given patient. The key determinants are the clinical measures of seizure frequency and presence of side effects, not the laboratory values. Conventional assays of serum drug levels measure the total drug (i.e., both free and protein-bound), yet it is the concentration of free drug that reflects extracellular levels in the brain and correlates best with efficacy. Thus, patients with decreased levels of serum proteins (e.g., decreased serum albumin due to impaired liver or renal function) may have an increased ratio of free to bound drug, yet the concentration of free drug may be adequate for seizure control. These patients may have a "subtherapeutic" drug level, but the dose should be changed only if seizures remain uncontrolled, not just to achieve a "therapeutic" level. It is also useful to monitor free drug levels in such patients. In practice, other than during the initiation or modification of therapy, monitoring of antiepileptic drug levels is most useful for documenting compliance.

If seizures continue despite gradual increases to the maximum tolerated dose and documented compliance, then it becomes necessary to switch to another antiepileptic drug. This is usually done by maintaining the patient on the first drug while a second drug is added. The dose of the second drug should be adjusted to decrease seizure frequency without causing toxicity. Once this is achieved, the first drug can be gradually withdrawn (usually over weeks unless there is significant toxicity). The dose of the second drug is then further optimized based on seizure response and side effects.

When to Discontinue Therapy Overall, about 70% of children and 60% of adults who have their seizures completely controlled with antiepileptic drugs can eventually discontinue therapy. Clinical studies suggest that the following patient profile yields the greatest chance of remaining seizure-free after drug withdrawal: (1) complete medical control of seizures for 1 to 5 years; (2) single seizure type, either partial or generalized; (3) normal neurologic examination, including intelligence; and (4) normal<u>EEG</u>. The appropriate seizure-free interval is unknown and undoubtedly varies for different forms of epilepsy. However, it seems reasonable to attempt withdrawal of therapy after 2 years in a patient who meets all of the above criteria, is motivated to discontinue the medication, and clearly understands the potential risks and benefits. In most cases it is preferable to reduce the dose of the drug gradually over 2 to 3 months. Most recurrences occur in the first 3 months after discontinuing therapy, and patients should be advised to avoid potentially dangerous situations such as driving or swimming during this period.

Treatment of Refractory Epilepsy Approximately one-third of patients with epilepsy do not respond to treatment with a single antiepileptic drug, and it becomes necessary to try a combination of drugs to control seizures. Patients who have focal epilepsy related to an underlying structural lesion or those with multiple seizure types and developmental delay are particularly likely to require multiple drugs. There are currently no clear guidelines for rational polypharmacy, but in most cases the initial combination therapy

combines first-line drugs, i.e., carbamazepine, phenytoin, valproic acid, and lamotrigine. If these drugs are unsuccessful, then the addition of a newer drug such as topiramate or gabapentin is indicated. Patients with myoclonic seizures resistant to valproic acid may benefit from the addition of clonazepam, and those with absence seizures may respond to a combination of valproic acid and ethosuximide. The same principles concerning the monitoring of therapeutic response, toxicity, and serum levels for monotherapy apply to polypharmacy, and potential drug interactions need to be recognized. If there is no improvement, a third drug can be added while the first two are maintained. If there is a response, the least effective of the first two drugs should be gradually withdrawn.

Surgical Treatment of Refractory Epilepsy Approximately 20% of patients with epilepsy are resistant to medical therapy despite efforts to find an effective combination of antiepileptic drugs. For some, surgery can be extremely effective in substantially reducing seizure frequency and even providing complete seizure control. Understanding the potential value of surgery is especially important when, at the time of diagnosis, a patient has an epilepsy syndrome that is considered likely to be drug-resistant. Rather than submitting the patient to years of unsuccessful medical therapy and the associated psychosocial trauma of ongoing seizures, the patient should have an efficient but relatively brief attempt at medical therapy and then be referred for surgical evaluation.

The most common surgical procedure for patients with temporal lobe epilepsy involves resection of the anteromedial temporal lobe (temporal lobectomy) or a more limited removal of the underlying hippocampus and amygdala. Focal seizures arising from extratemporal regions may be suppressed by a focal neocortical resection or precise removal of an identified lesion (lesionectomy). When the cortical region cannot be removed, multiple subpial transection, which disrupts intracortical connections, is sometimes used to prevent seizure spread. Hemispherectomy or multilobar resection is useful for some patients with severe seizures due to hemispheric abnormalities such as hemimegaloencephaly or other dysplastic abnormalities, and corpus callosotomy has been shown to be effective for disabling tonic or atonic seizures, usually when they are part of a mixed-seizure syndrome (e.g., Lennox-Gastaut syndrome).

Presurgical evaluation is designed to identify the functional and structural basis of the patient's seizure disorder. Inpatient video-EEGmonitoring is used to define the anatomic location of the seizure focus and to correlate the abnormal electrophysiologic activity with behavioral manifestations of the seizure. Routine scalp or scalp-sphenoidal recordings are usually sufficient for localization, and advances in neuroimaging have made the use of invasive electrophysiologic monitoring such as implanted depth electrodes or subdural electrodes much less common. A high-resolutionMRI scan is routinely used to identify structural lesions. Functional imaging studies such asSPECTand PETare adjunctive tests that may help verify the localization of an apparent epileptogenic region with an anatomic abnormality. Once the presumed location of the seizure onset is identified, additional studies, including neuropsychological testing and the intracarotid amobarbital test (Wada test) may be used to assess language and memory localization and to determine the possible functional consequences of surgical removal of the epileptogenic region. In some cases, the exact extent of the resection to be undertaken is determined by performing cortical mapping at the time of the surgical procedure. This involves electrophysiologic recordings and cortical stimulation in the awake patient to identify the extent of

epileptiform disturbances and the function of cortical regions in question.

Advances in presurgical evaluation and microsurgical techniques have led to a steady increase in the success of epilepsy surgery. Clinically significant complications of surgery are<5%, and the use of functional mapping procedures has markedly reduced the neurologic sequelae due to removal or sectioning of brain tissue. For example, about 70% of patients treated with temporal lobectomy will become seizure-free, and another 15 to 25% will have at least a 90% reduction in seizure frequency. Marked improvement is also usually seen in patients treated with hemispherectomy for catastrophic seizure disorders due to large hemispheric abnormalities. Postoperatively, patients generally need to remain on antiepileptic drug therapy, but the marked reduction of seizures following surgery can have a very beneficial effect on their quality of life.

Vagus Nerve Stimulation (VNS) VNS is a new treatment option for patients with medically refractory epilepsy who are not candidates for resective brain surgery. The procedure involves placement of a bipolar electrode on the midcervical portion of the left vagus nerve. The electrode is connected to a small, subcutaneous generator located in the infraclavicular region, and the generator is programmed to deliver intermittent electrical pulses to the vagus nerve. The precise mechanism of action of VNS is unknown, although experimental studies have shown that stimulation of vagal nuclei leads to widespread activation of cortical and subcortical pathways and an associated increased seizure threshold. In practice, the efficacy of VNS appears to be no greater than recently introduced anticonvulsant medications. Adverse effects of the surgery are rare, and stimulation-induced side effects, including transient hoarseness, cough, and dyspnea, are usually mild and well tolerated.

STATUS EPILEPTICUS

Status epilepticus refers to continuous seizures or repetitive, discrete seizures with impaired consciousness in the interictal period. The duration of seizure activity sufficient to meet the definition of status epilepticus has traditionally been specified as 15 to 30 min. However, a more practical definition is to consider status epilepticus as a situation in which the duration of seizures prompts the acute use of anticonvulsant therapy, typically when seizures last beyond 5 min.

Status epilepticus is an emergency and must be treated immediately, since cardiorespiratory dysfunction, hyperthermia, and metabolic derangements can develop as a consequence of prolonged seizures, and these can lead to irreversible neuronal injury. Furthermore, CNS injury can occur even when the patient is paralyzed with neuromuscular blockade but continues to have electrographic seizures. The most common causes of status epilepticus are anticonvulsant withdrawal or noncompliance, metabolic disturbances, drug toxicity, CNS infection, CNS tumors, refractory epilepsy, and head trauma.

Generalized status epilepticus is obvious when the patient is having overt convulsions. However, after 30 to 45 min of uninterrupted seizures, the signs may become increasingly subtle. Patients may have mild clonic movements of only the fingers, or fine, rapid movements of the eyes. There may be paroxysmal episodes of tachycardia,

hypertension, and pupillary dilation. In such cases, the EEG may be the only method of establishing the diagnosis. Thus, if the patient stops having overt seizures, yet remains comatose, an EEG should be performed to rule out ongoing status epilepticus.

The first step in the management of a patient in status epilepticus is to attend to any acute cardiorespiratory problems or hyperthermia, perform a brief medical and neurologic examination, establish venous access, and send samples for laboratory studies to identify metabolic abnormalities. Anticonvulsant therapy should then begin without delay; a treatment approach is shown in Fig. 360-3.

BEYOND SEIZURES: OTHER MANAGEMENT ISSUES

Interictal Behavior The adverse effects of epilepsy often go beyond the occurrence of clinical seizures, and the extent of these effects depends largely upon the etiology of the seizure disorder, the degree to which the seizures are controlled, and the presence of side effects from antiepileptic therapy. Many patients with epilepsy are completely normal between seizures and able to live highly successful and productive lives. In contrast, patients with seizures secondary to developmental abnormalities or acquired brain injury may have impaired cognitive function and other neurologic deficits. Frequent interictal EEG abnormalities have been shown to be associated with subtle dysfunction of memory and attention. Patients with many seizures, especially those emanating from the temporal lobe, often note an impairment of short-term memory that may progress over time.

Patients with epilepsy are at risk of developing a variety of psychiatric problems including depression, anxiety, and psychosis. This risk varies considerably depending on many factors, including the etiology, frequency, and severity of seizures and the patient's age and previous history. Depression occurs in approximately 20% of patients, and the incidence of suicide is higher in epileptic patients than in the general population. Depression should be treated through counseling or medication. The selective serotonin reuptake inhibitors typically have no effect on seizures, while the tricyclic antidepressants may lower the seizure threshold. Anxiety can appear as a manifestation of a seizure, and anxious or psychotic behavior can sometimes be observed as part of a postictal delirium. Interictal psychosis is a rare phenomenon that typically occurs after a period of increased seizure frequency. There is usually a brief lucid interval lasting up to a week, followed by days to weeks of agitated, psychotic behavior. The psychosis will usually resolve spontaneously but may require treatment with antipsychotic or anxiolytic medications.

There is ongoing controversy as to whether some patients with epilepsy (especially temporal lobe epilepsy) have a stereotypical "interictal personality." The predominant view is that the unusual or abnormal personality traits observed in such patients are, in most cases, not due to epilepsy but result from an underlying structural brain lesion, the effects of antiepileptic drugs, or psychosocial factors.

Mortality of Epilepsy Patients with epilepsy have an increased risk of death that is roughly two to three times greater than what would be expected in a matched population without epilepsy. Most of the increased mortality is due to the underlying etiology of epilepsy, i.e., more widespread neurologic or systemic diseases in children and tumors

or strokes in older adults. However, a small number of patients die from a syndrome known as *sudden unexpected death in epileptic patients* (SUDEP), which usually affects young people with convulsive seizures and tends to occur at night. The cause(s) remain unknown, although the leading theories propose brainstem-mediated effects of seizures on cardiac rhythms or pulmonary function.

Psychosocial Issues There continues to be a cultural stigma about epilepsy, although it is slowly declining in societies with effective health education programs. Because of this stigma, many patients with epilepsy harbor fears, such as the fear of becoming mentally retarded or dying during a seizure. These issues need to be carefully addressed by educating the patient about epilepsy and by ensuring that family members, teachers, fellow employees, and other associates are equally well informed. The Epilepsy Foundation of America (1-800-EFA-1000) is a patient advocacy organization and a useful source of educational material.

Employment and Driving Many patients with epilepsy face difficulty in obtaining or maintaining employment, even when their seizures are well controlled. Federal and state legislation is designed to prevent employers from discriminating against patients with epilepsy, and patients should be encouraged to understand and claim their legal rights. Patients in these circumstances also benefit greatly from the assistance of health providers who act as strong patient advocates.

Loss of driving privileges is one of the most disruptive social consequences of epilepsy. Physicians should be very clear about local regulations concerning driving and epilepsy, since the laws vary considerably among states and countries. In all cases, it is the physician's responsibility to warn patients of the danger imposed on themselves and others while driving if their seizures are uncontrolled (unless the seizures are not associated with impairment of consciousness or motor control). In general, most states allow patients to drive after a seizure-free interval (on or off medications) between 3 months and 2 years.

SPECIAL ISSUES RELATED TO WOMEN AND EPILEPSY

Catamenial Epilepsy Some women experience a marked increase in seizure frequency around the time of menses. This is thought to reflect either the effects of estrogen and progesterone on neuronal excitability or changes in antiepileptic drug levels due to altered protein binding. Acetazolamide (250 to 500 mg/d) may be effective as adjunctive therapy in some cases when started 7 to 10 days prior to the onset of menses and continued until bleeding stops. Some patients may benefit from increases in antiepileptic drug dosages during this time or from control of the menstrual cycle through the use of oral contraceptives.

Pregnancy Most women with epilepsy who become pregnant will have an uncomplicated gestation and deliver a normal baby. However, epilepsy poses some important risks to a pregnancy. Seizure frequency during pregnancy will remain unchanged in approximately 50% of women, increase in 30%, and decrease in 20%. Changes in seizure frequency are attributed to endocrine effects on the CNS, variations in antiepileptic drug pharmacokinetics (such as acceleration of hepatic drug metabolism or effects on plasma protein binding), and changes in medication compliance. It is

therefore useful to see patients at more frequent intervals during pregnancy and monitor serum antiepileptic drug levels. Measurement of the unbound drug concentrations may be useful if there is an increase in seizure frequency or worsening of side effects of antiepileptic drugs.

The overall incidence of fetal abnormalities in children born to mothers with epilepsy is 5 to 6%, compared to 2 to 3% in healthy women. Part of the higher incidence is due to teratogenic effects of antiepileptic drugs, and the risk increases with the number of medications used (e.g., 10% risk of malformations with three drugs). A syndrome comprising facial dysmorphism, cleft lip, cleft palate, cardiac defects, digital hypoplasia, and nail dysplasia was originally ascribed to phenytoin therapy, but it is now known to occur with other first-line antiepileptic drugs (i.e., valproic acid and carbamazepine) as well. Also, valproic acid and carbamazepine are associated with a 1 to 2% incidence of neural tube defects compared with a baseline of 0.5 to 1%. Little is currently known about the safety of newer drugs.

Since the potential harm of uncontrolled seizures on the mother and fetus is considered greater than the teratogenic effects of antiepileptic drugs, it is currently recommended that pregnant women be maintained on effective drug therapy. When possible, it seems prudent to have the patient on monotherapy at the lowest effective dose, especially during the first trimester. Patients should also take folate (1 to 4 mg/d), since the antifolate effects of anticonvulsants are thought to play a role in the development of neural tube defects, although the benefits of this treatment remain unproved in this setting.

Enzyme-inducing drugs such as phenytoin, phenobarbital, and primidone cause a transient and reversible deficiency of vitamin K-dependent clotting factors in approximately 50% of newborn infants. Although neonatal hemorrhage is uncommon, the mother should be treated with oral vitamin K (20 mg daily) in the last 2 weeks of pregnancy, and the infant should receive an intramuscular injection of vitamin K (1 mg) at birth.

Contraception Special care should be taken when prescribing antiepileptic medications for women who are taking oral contraceptive agents. Drugs such as carbamazepine, phenytoin, phenobarbital, and topiramate can significantly antagonize the effects of oral contraceptives via enzyme induction and other mechanisms. Patients should be advised to consider alternative forms of contraception, or their contraceptive medications should be modified to offset the effects of the antiepileptic medications.

Breast Feeding Antiepileptic medications are excreted into breast milk to a variable degree. The ratio of drug concentration in breast milk relative to serum is approximately 80% for ethosuximide, 40 to 60% for phenobarbital, 40% for carbamazepine, 15% for phenytoin, and 5% for valproic acid. Given the overall benefits of breast feeding and the lack of evidence for long-term harm to the infant by being exposed to antiepileptic drugs, mothers with epilepsy should be encouraged to breast feed. This should be reconsidered, however, if there is any evidence of drug effects on the infant, such as lethargy or poor feeding.

(Bibliography omitted in Palm version)

361. CEREBROVASCULAR DISEASES - *Wade S. Smith, Stephen L. Hauser, J. Donald Easton*

Cerebrovascular diseases occur predominately in the middle and late years of life. They cause approximately 200,000 deaths in the United States each year, as well as considerable neurologic disability. The incidence of stroke increases with age; thus the disability affects many people in their "golden years," a segment of the population that is growing rapidly in Western countries. Categories of cerebrovascular diseases include ischemia-infarction and intracranial hemorrhage (Table 361-1). Many of the arterial and cardiac disorders underlying these diseases are preventable; the morbidity and mortality from cerebrovascular diseases has been diminishing in recent years, apparently because of better recognition and treatment of hypertension.

Most cerebrovascular diseases are manifest by the abrupt onset of a focal neurologic deficit. The deficit may remain fixed or may rapidly improve or progressively worsen. It is this abrupt onset of a nonconvulsive and focal neurologic deficit that defines a *stroke*, or cerebrovascular accident (CVA).

Cerebral ischemia is caused by a reduction in blood flow that lasts for several seconds to a few minutes. Neurologic symptoms are manifest within 10 s because neurons lack glycogen and suffer rapid energy failure. When blood flow is rapidly restored brain tissue can recover fully, and the patient's symptoms are only transient: a transient ischemic attack (TIA) is said to have occurred. Typically the neurologic signs and symptoms of TIA last for 5 to 15 min but, by definition, must last<24 h. If the cessation of flow lasts for more than a few minutes, *infarction* or death of brain tissue results. Stroke has occurred if the neurologic signs and symptoms last for>24 h. A *generalized* reduction in cerebral blood flow due to systemic hypotension (e.g., cardiac arrhythmia, myocardial infarction, or hemorrhagic shock) usually produces syncope (Chap. 21). If low cerebral blood flow is maintained for a longer duration, then infarction in the border zones between the major cerebral artery distributions or widespread brain necrosis develops. This process is termed *global hypoxia-ischemia* and a patient with cognitive sequelae is said to have *hypoxic-ischemic encephalopathy* (Chap. 376). *Focal* ischemia or infarction, on the other hand, is usually caused by thrombosis of the cerebral vessels themselves or by emboli from a proximal arterial source or the heart. A comprehensive list of causes of ischemia-infarction is shown inTable 361-2.

Cerebral hemorrhage produces neurologic symptoms by producing a mass effect on neural structures or from the toxic effects of blood itself; debate exists as to how much injury occurs from tamponade of surrounding blood vessels. As with ischemia-infarction, the causes are numerous (Tables 361-1 and361-7).

When faced with an acute stroke, the clinician must rapidly differentiate between ischemia-infarction and hemorrhage, because the method of emergency treatment depends on cause. The clinician should focus on two goals: (1) to prevent or reverse acute brain injury, and (2) to prevent future neurologic injury. The first goal involves identifying those patients who may benefit from thrombolysis and attending to acute medical issues of airway, blood pressure, and concomitant organ failure; the second goal is achieved once the mechanism of stroke is elucidated and the proper secondary prevention strategy prescribed.

ISCHEMIC STROKE

MECHANISMS AND DEFINITIONS

Several pathophysiologic processes may produce cerebral ischemia and infarction. A common form is atherosclerotic damage to the aortic arch, carotid bifurcation, or intracranial vessels that produces local thrombosis and distal embolism of the clot. The released clot travels until it occludes a distal vessel and prevents distal cerebral blood flow. Such strokes are called *atherothromboembloic strokes*, or simply *embolic strokes*, and are a subset of *artery-artery embolic strokes*. Stroke produced by thrombosis of large (~0.5 to 3 mm) intracranial vessels in situ from atherosclerosis is termed *atherothrombotic stroke*. Unlike coronary arteries in which vascular occlusion may be sudden and complete, sudden thrombosis of intracranial vessels occurs less frequently. It may be more likely that atherosclerosis of an intracranial vessel will produce stroke by distal embolism rather than by occlusion. Stenosis of an extra- or intracranial vessel may produce a *low-flow stroke* or TIA if cardiac output or systemic blood pressure is reduced below some threshold. This mechanism was thought to be the cause of stroke from carotid atherosclerosis, but it is now clear that carotid disease produces stroke primarily by an embolic mechanism. True flow-related TIAs and stroke are rare, but it is important to identify them since they will respond to revascularization procedures rather than standard antithrombotic treatment. Thrombotic occlusion of smaller intracranial vessels (~30 to 100 um), in contrast to larger vessel thrombosis, is a frequent cause of stroke. These end-arteries typically supply a small volume of brain tissue, and their occlusion may result in a *lacunar syndrome*, of which there are >30 types. A patient who has a stroke from this event is said to have *lacunar stroke*. The underlying pathology of this form of stroke is usually lipohyalinosis or microatheromata with thrombosis of the vascular lumen.

Clinically, thrombotic strokes are more gradual in onset or may stutter. It is common for a person to experience several TIAs of the lacunar type prior to eventual stroke. *Crescendo TIAs* -- the occurrence of increasing number and frequency of TIAs -- have a particularly high likelihood of evolving to stroke. *Stroke in progression* is said to be present if a patient suffers progressive neurologic deficits over a few hours or days that are not accounted for by cerebral edema. This may happen as a small vessel slowly thromboses or, more ominously, as a larger intracranial vessel such as the basilar artery progressively thromboses, producing an ever enlarging region of cerebral ischemia. Heparin and thrombolytic treatment may arrest progression, but it has been difficult to demonstrate in clinical trials whether or not such treatments improve outcome.

Embolism from a cardiac source is most commonly from red atrial thrombi but can arise from numerous sources. In most cases the clinician does not observe clot within the heart and makes the diagnosis by associating a known cardiac cause (e.g., atrial fibrillation, recent myocardial infarction) with a sudden large-vessel occlusion in the brain. Such strokes are called *cardioembolic strokes*. Some patients, however, may develop sudden occlusion of a large intracranial vessel, and despite extensive evaluation, no cause is apparent. These strokes are called *cryptogenic strokes*.

Clinically, embolic events usually produce a sudden onset of neurologic dysfunction that

is maximum at onset. The extent of neuronal ischemia is determined by the location of the occlusion and the degree to which collateral flow is offered to the ischemic tissue bed, the blood pressure and body temperature, and other factors. Embolic strokes have a higher risk of transforming into hemorrhagic stroke in which petechial bleeding or frank hemorrhage occurs into the infarcted tissue hours or days following the initial embolic occlusion. This natural history risk of spontaneous hemorrhage must be taken into account in acute stroke trials testing the safety of thrombolytic treatment.

RISK FACTORS FOR ISCHEMIC STROKE

Older age, family history of thrombotic stroke, diabetes mellitus, hypertension, tobacco smoking, elevated blood cholesterol levels, and other factors are risk factors for atherosclerosis and hence either proven or probable risk factors for ischemic stroke. Risk of second stroke is strongly influenced by prior stroke or TIA (Table 361-3). Many cardiac conditions predispose to stroke, including atrial fibrillation and recent myocardial infarction. Oral contraceptives may increase stroke risk slightly, and certain inherited and acquired hypercoagulable states predispose to stroke. Identification of modifiable risk factors and prophylactic interventions to lower risk is probably the best treatment for stroke overall, as the total number of strokes could be reduced substantially by these means. (See below for recommendations for risk factor modification.)

ACUTE STROKE

Clinical Encounter Patients with acute stroke often do not seek medical assistance on their own, perhaps because it is rarely painful but also because they may lose the appreciation that something is wrong with them (*anosognosia*). It is often a family member or a bystander who calls for help, and many gain entry into the medical system through emergency medical services, such as the 911 system in the United States. Use of such a system allows rapid evaluation of patients for consideration for time-sensitive treatments such as thrombolysis. Patients at risk for stroke should be counseled to call emergency medical services if they experience the sudden onset of any of the following: loss of sensory and/or motor function on one half of the body; change in vision, gait, or ability to speak or understand; or a sudden, severe headache.

The differential diagnosis of neurologic symptoms of sudden onset includes stroke (ischemic or hemorrhagic), TIA, seizure with postictal Todd's paralysis, intracranial tumor, migraine, and metabolic encephalopathy (Table 361-4). An adequate history from an observer that no convulsive activity occurred at the onset reasonably excludes seizure. Tumors may present with acute neurologic symptoms due to hemorrhage, seizure, or hydrocephalus. Surprisingly, migraine can mimic cerebral ischemia, even in patients without a significant migraine history. When it develops without head pain (*acephalgic migraine*), the diagnosis may remain elusive. Elderly patients without any prior history of complicated migraine may develop acephalgic migraine after age 65. The sensory disturbance is often prominent, and the sensory deficit, as well as any motor deficits, tends to migrate slowly across a limb over minutes. The diagnosis of migraine becomes more likely as the cortical disturbance begins to cross vascular boundaries. At times it may be difficult to make the diagnosis until multiple episodes have occurred leaving behind no residual stroke or brain imaging abnormality. Classically, metabolic encephalopathies produce a fluctuating mental status without

focal neurologic findings. However, in the setting of prior stroke or brain injury, a patient with fever or sepsis will manifest hemiparesis, which clears rapidly when the infection is remedied. The metabolic process serves to "unmask" a prior deficit.

STROKE SYNDROMES

A careful history and neurologic examination can often localize the region of brain dysfunction; if this region corresponds to a particular arterial distribution, the possible causes responsible for the syndrome can be narrowed. This is of particular importance when the patient presents with a TIA and a normal examination. For example, if a patient develops language loss and a right homonymous hemianopia, a search for causes of left middle cerebral emboli should be performed. A finding of an isolated stenosis of the right internal carotid artery in that patient suggests an asymptomatic carotid stenosis, which carries a significantly lower risk than symptomatic stenosis (i.e., stenosis of the left internal carotid artery). The following sections describe the clinical findings of arterial ischemia associated with cerebral vascular territories depicted in Figs. 361-1, 361-2, 361-3, 361-4, 361-5, 361-6, 361-7, 368-8, and 361-9. Stroke syndromes are divided into: (1) large vessel stroke within the anterior circulation, (2) large vessel stroke within the posterior circulation, and (3) small vessel disease of either vascular bed.

Large Vessel Stroke within the Anterior Circulation

Pathophysiology The internal carotid artery and its branches comprise the anterior circulation of the brain. These vessels can be occluded by intrinsic disease of the vessel (e.g., atherosclerosis or dissection) or by embolic occlusion from a proximal source. The causes of occlusion are enumerated here, and the clinical manifestations are listed in the next section.

EXTRACRANIAL INTERNAL CAROTID ARTERY The origin of the internal carotid artery is probably the most common site of atherosclerosis that leads to TIA or stroke. Atherosclerosis is usually most severe in the first 2 cm and arises from the posterior wall, often extending downward into the common carotid artery. Atherosclerosis at this site is often manifested by a TIA or minor stroke, presumably caused by embolism or, less frequently, low flow.

Dissection of the carotid artery produces cerebral ischemia by distal embolization and/or low flow to the anterior circulation. When low flow is the mechanism, there is presumably inadequate collateral flow through the circle of Willis. Fibromuscular dysplasia of the carotids may produce distal emboli or dissection.

Rarely a large embolus will lodge in the common or internal carotid artery. Emboli of a size sufficient to block the internal carotid most often originate from pulmonary veins or extensive atrial or myocardial thrombi. *Takayasu's arteritis* (Chap. 317) is the most common form of vasculitis that affects the carotid artery.

INTRACRANIAL INTERNAL CAROTID ARTERY Atheromatous disease at the petrous inlet, the siphon (S-shaped portion of the internal carotid artery in the cavernous sinus), or the proximal segment of the middle or anterior cerebral arteries may produce distal embolization. These intracranial sites predominate in African Americans, Hispanics, and

Asians. *Moyamoya syndrome* results from progressive stenosis and occlusion of the distal internal carotid artery and/or proximal middle and anterior cerebral arteries. It is idiopathic in children and acquired secondary to atherosclerosis in adults. Ischemia is produced by breakdown in lenticulostriate collaterals that form to reconstitute flow in the middle cerebral artery (MCA) or by progressive sclerosis of cortical vessels. Capsular hemorrhage may occur from rupture of the enlarged lenticulostriate vessels.

MIDDLE CEREBRAL ARTERY In contrast to the internal carotid artery, occlusion of the proximal MCA or one of its major branches is most often due to an embolus (artery-to-artery, cardiac, or of unknown source) rather than intracranial atherothrombosis. Atherosclerosis of the proximal MCA may cause distal emboli to the middle cerebral territory or, less commonly, may produce low-flow TIAs. Collateral formation via leptomeningeal vessels often prevents MCA stenosis from becoming symptomatic.

ANTERIOR CEREBRAL ARTERY Atheromatous deposits in the proximal segment of the anterior cerebral artery rarely cause symptoms because the effects of occlusion are usually circumvented by collateral circulation through the anterior communicating artery. If the anterior communicating artery is congenitally atretic or the atheromatous lesion occurs distally in the anterior cerebral artery, TIAs and stroke may occur. The anterior cerebral artery is rarely the recipient of emboli.

Clinical Manifestations

MIDDLE CEREBRAL ARTERY The cortical branches of the MCA supply the lateral surface of the hemisphere except for (1) the frontal pole and a strip along the superomedial border of the frontal and parietal lobes supplied by the anterior cerebral artery and (2) the lower temporal and occipital pole convolutions supplied by the posterior cerebral artery (Figs. 361-2,361-4, and 361-5).

The proximal MCA (M1 segment) gives rise to penetrating branches (termed *lenticulostriate arteries*) that supply the putamen, outer globus pallidus, posterior limb of the internal capsule above the plane of the upper border of the globus pallidus, the adjacent corona radiata, and the body and upper and lateral head of the caudate nucleus. In the sylvian fissure, the middle cerebral artery in most patients divides into *superior* and *inferior* divisions (M2 branches). Branches of the inferior division supply the inferior parietal and temporal cortex, and those from the superior division supply the frontal and superior parietal cortex (Fig. 361-3). There is considerable variability in the parietal lobe supply between the two divisions, with about two-thirds of individuals having an inferior division that supplies regions above the angular gyrus.

If the entire MCA is occluded at its origin (blocking both its penetrating and cortical branches) and the distal collaterals are limited, the clinical findings are contralateral hemiplegia, hemianesthesia, homonymous hemianopia, and a day or two of gaze preference to the ipsilateral side. When the dominant hemisphere is involved, global aphasia is present also, and when the nondominant hemisphere is affected, anosognosia, constructional apraxia, and neglect are found (Fig. 361-3). Dysarthria may also occur.