Sphingolipids differ from lecithin and cephalin. They are phosphate esters of sphingosine bound to choline or ethanolamine and primarily are found in brain tissue (eg, sphingomyelin, galactolipin). The *ratio of lecithin to sphingomyelin* (L/S) in amniotic fluid or resuscitated amniotic fluid from the oral cavity of the newborn is an accurate assessment of fetal maturity and the respiratory-distress syndrome. Changes in phospholipid biosynthesis during gestation reflect the aging of the fetal lung, as the L/S ratio normally increases.

Tay-Sachs disease is a lipid-storage disease in which the central nervous system degenerates because of the progressive intraneuronal accumulation of excess amounts of the sphingolipid ganglioside GM_2. The accumulation of GM_2 in Tay-Sachs disease has been shown to be caused by a lack of the enzyme hexosaminidase A. Therefore, the measurement of serum, WBC or amniotic fluid *hexosaminidase A* is important in evaluating carriers and in diagnosing Tay-Sachs disease in the fetus.

Both hexosaminidase A (heat-labile) and hexosaminidase B (heat-stable) can catalyze the conversion of 4-methylumbelliferyl-*N*-acetylgalactosamine (a synthetic substrate) to *N*-acetylgalactosamine and 4-methylumbelliferone. The cleavage product, 4-methylumbelliferone, fluoresces under ultraviolet radiation and the intensity of the fluorescence is a measure of the activity of the enzyme. In noncarriers, 50 to 75% of the total hexosaminidase activity is heat-labile (hexosaminidase A), and in carriers 20 to 45% of the total hexosaminidase activity is heat-labile.

The blood fatty acids occur in esterified (EFA) and nonesterified (NEFA) forms. *Triglyceride* determinations are of value in differentiating the hyperlipidemic states, ie, essential (diet-induced) hypertriglyceridemia from familial hypocholesterolemia with or without triglyceridemia. After the preliminary separation from phospholipids, triglycerides most often are determined in terms of their glycerol moiety. The glycerol released by saponification is oxidized to formaldehyde and the latter determined by fluorimetric or colorimetric procedures. Triglycerides also can be determined by coupling the glycerol liberated from lipase/α-chymotrypsin treatment of serum with a glycerol kinase-pyruvate kinase-LDH system and spectrometric estimation of NADH. Normal triglyceride levels are 110 to 140 mg/100 mL. An increase in triglycerides will produce a milky appearance in serum (lipemic). EFA analyses are based also on the reaction of alkaline hydroxylamine with esters of fatty acids to form hydroxamic acids which produce a red color with ferric chloride.

Gas chromatographic procedures have been used to quantitate the various *fatty acids;* ie, palmitic, stearic, oleic, linoleic and linolenic acids. Mono-, di- and triglycerides also can be separated into classes and quantitated by column or thin-layer chromatography, and infrared spectrometry. The total fatty acids of plasma range from 200 to 450 mg/100 mL in the fasting state; they are derived from glycerides, cholesterol esters and phospholipids.

All the lipids in plasma circulate in combination with protein. The free fatty acids are bound to albumin and the lipids aggregate with other proteins to form *lipoproteins*. Electrophoresis and ultracentrifugation are the principal methods used to separate and identify lipoprotein families. *Chylomicrons* (S_f >400), *pre-β-lipoproteins* (S_f 20–400), *β-lipoproteins* (S_f 0–20) and *α-lipoproteins* are the four major classes in order of increasing density and migration on cellulose acetate electrophoresis. Chylomicrons are representative primarily of dietary or exogenous triglycerides, pre-β-lipoproteins of endogenous glycerides, β-lipoproteins of cholesterol and its esters and α-lipoproteins of cholesterol and phospholipids. Abnormal lipoproteins that may appear in plasma include floating β-lipoproteins, lipoprotein X and complexes of normal lipoproteins with IgA and IgG myeloma proteins (autoimmune hyperlipoproteinemia). Age, sex, diet, fasting, posture changes and trauma can alter the lipid profile.

The lipoprotein classes usually are separated by paper, agarose or cellulose acetate electrophoresis. The strips are stained with fat-soluble dyes (Sudan Black or Oil Red O) and quantitated by densiometric scanning. Primary hyperlipoproteinemias are classified into normal and five abnormal types based on cholesterol and triglyceride levels and lipoprotein analysis. Hyperchylomicronemia (Type I), hyper-β-lipoproteinemia (Type II), broad β-band (Type III), hyper-pre-β-lipoproteinemia (Type IV) and hyper-pre-β-lipoproteinemia and chylomicronemia (Type V) are the major classes. Carbohydrate and fat-tolerance studies, post-heparin lipase activity and clinical symptomatology also are integrated into the diagnosis of the various subclasses. The presence or predisposition to coronary artery disease and other disease states is associated with the various types.[17]

Steroids and Other Hormones—The steroids possess a common structure, the perhydrocyclopentanophenanthrene nucleus, and include cholesterol, bile acids, androgens and the adrenocortical, adrenomedullary, estrogenic and progestational hormones.

Androsterone, dehydroepiandrosterone, etiocholan-3α-ol-17-one, 11-ketoandrosterone, 11-ketoetiocholanolone, 11β-hydroxyandrosterone and 11β-hydroxyetiocholanolone are the principal *urinary 17-ketosteroids* (*17KS*). These androgenic hormones are derived from the adrenal and, in males, testicular function. The principal urinary steroid metabolites in this group of androgens are found both in the free form, and as conjugates of glucuronides, sulfates or acetates. Their determination in urine involves the acid hydrolysis of the conjugates, extraction with organic solvent, reaction with alkaline *m*-dinitrobenzene (Zimmerman reaction) and colorimetric estimation of the chromogen. The individual 17KS can be separated by TLC prior to analysis to obtain further information on the individual steroids. The normal adult urine values are: male, 9 to 24 mg/day; female, 5 to 17 mg/day. Decreased excretion is seen in hyponactive disease of the pituitary, gonads and adrenals. Increased excretion is seen in hyperplasia, cancer or tumors of the adrenals.

Testosterone is the most potent androgen in blood. The measurement of urinary or serum testosterone is useful in distinguishing normal and hypogonadal males and in treating hirsutism in the female. This hormone is determined by gas chromatography, competitive protein-binding, isotope dilution or RIA procedures. Normal serum testosterone is 0.2 to 1.1 μg/100 mL in the male and <0.1 μg/100 mL in the female.

The natural *estrogenic hormones* are estradiol, estrone and estriol, produced in the gonads, adrenals and placenta. The relative amounts of the three estrogens rise and fall concomitantly during the menstrual cycle. Maternal, urinary total-estrogen excretion, especially estriol, is an indirect index of the integrity and viability of the fetoplacental unit. Analysis involves acid or glucuronidase-arylsulfatase hydrolysis of the conjugates, removal of urinary glucose if present, extraction and colorimetric or fluorimetric analysis. In the determination, after acid hydrolysis and ether extraction of the urine, the estrogens are methylated with dimethyl sulfate and chromatographically separated prior to reaction with phenolsulfuric acid to yield a red chromogen for colorimetric analysis. The normal estrogen output is 4 to 60 μg/24 hr in the female and up to 25 μg in the male. Estrogen deficiency can be related to ovarian failure and pituitary deficiency.

Progesterone is a progestational hormone which is secreted by the corpus luteum of the ovary and also by the adrenal

cortex. Serum progesterone determination is of value in the detection of ovulation and is a measure of the secretory activity of the placenta during pregnancy. Progesterone is determined in serum by RIA, double-isotope derivatization, gas-liquid chromatography or competitive protein-binding techniques. Normal, menstrual-cycle serum progesterone levels vary between 0 and 1.6 μg/100 mL.

Pregnanediol is the principal metabolite of progesterone. The urinary determination of pregnanediol excretion is an indirect index of progesterone levels but is subject to variation due to individual differences in hepatic metabolism of this hormone and is not representative of total endogenous progesterone production.

Adrenal cortex steroids include glucocorticoids, androgens, estrogens, progesterone and mineralocorticoids. Glucocorticoids can be determined as plasma cortisol (plasma 17-OH corticosteroids), urinary-free (unconjugated cortisol) or total-urinary *17-OH corticosteroids*. The latter are determined in urine as *17-ketogenic steroids* (17KGS). The 17KS in urine are reduced with borohydride to alcohols; the 17-OH steroids are oxidized with sodium bismuthate or periodate to 17KS and quantitated by the alkaline dinitrobenzene method. The 17-OH steroids can be quantitated directly by the phenylhydrazine-sulfuric acid reaction after hydrolysis of glucuronide conjugates and chromatographic purification. The 17-OH steroid analysis only determines compounds with the dihydroxyacetone side chain, such as tetrahydrocortisol or tetrahydrocortisone; the 17KGS analysis includes the 17-OH-corticosteroids with the dihydroxyacetone side chain and the pregnanetriol type of compound. Normal 17KGS daily urinary excretion is 5 to 23 mg in the male and 3 to 15 mg in the female. They are reduced significantly in myxedema and adrenal or anterior pituitary insufficiency. Plasma cortisol usually is measured by fluorimetric or gas-chromatographic procedures.

Aldosterone is the most active member of the mineralocorticoid group. The determination of urinary aldosterone is of value in differentiating benign essential hypertension from primary aldosteronism (Conn's syndrome), which is caused by an adrenal adenoma and is accompanied by hypertension. A double-isotope derivatization technique is used. Urinary aldosterone is acetylated with ^{3}H-acetic anhydride; aldosterone-^{14}C-diacetate standard is added early in the procedure. The $^{3}H/^{14}C$ specific activity of the final product is measured after chromatographic purification and is a direct measurement of aldosterone. The normal aldosterone levels of about 10 μg/day are elevated in Conn's disease and usually are associated with low serum potassium, sodium retention and low-concentration alkaline urine.

The anterior pituitary secretes three substances (*gonadotropins*) which regulate gonadal activity: *follicle-stimulating hormone* (*FSH*), *luteinizing hormone* or *interstitial cell hormone* (*ICSH, LH*) and *luteotropin* (*LTH*). The gonadotropins are glycoproteins. Bioassay methods can be used to determine gonadotrophic activity. After fractionation and isolation the urine extract is assayed in test animals as to the follicular growth of the ovaries in hypophysectomized animals or increase in testicular, ovarian or uterine weight in various animal models. RIA techniques have been developed for these gonadotropins and represent the most sensitive and precise measurement method.

Analysis of serum or urinary *placental lactogen* (*HPL*) and *chorionic gonadotropin* (*HCG*), a placental-derived protein hormone, is useful in the diagnosis of threatened abortion, hydatiform mole and choriocarcinoma. HCG, pregnanediol and progesterone as well as total and fractionated estrogens are useful in testing for pregnancy. HCG and HPL readily are measured by RIA and low values are seen in threatened abortion and intrauterine fetal death.

The increase in HCG in the serum or urine of the pregnant female is the basis of a routine *pregnancy test*. Test components consist of an antigen in the form of HCG latex particles and an HCG antiserum. When antiserum is mixed with urine containing a detectable level of HCG, it is neutralized and no agglutination of latex-antigen particles occur (*agglutination inhibition test*). The commercial application of the HCG assay gives laboratories a rapid, accurate pregnancy test by taking advantage of monoclonal antibody specificity and sensitivity. A monoclonal slide procedure on urine, Duoclon (*Organon Diagnostics*), uses two different monoclonal antibodies, one against HCG and one against the HCG_{β} subunit for maximum specificity. Agglutination indicates a positive test with a sensitivity level of 500 mIU HCG/mL, detecting pregnancy a few days after conception.

Human growth hormone and *insulin* are proteins which are of diagnostic value in growth-rate studies and diabetes. They are best quantitated by RIA.

Epinephrine and *norepinephrine* are biologically active catecholamines derived from the adrenal medulla and sympathetic nerve endings. Catecholamines are measured in the blood and urine after fractionation on alumina or ion-exchange columns, oxidation at pH 3.5 or 6.0 and subsequent fluorimetric analysis. Urine catecholamines are increased to >350 μg/24 hr in adrenal medullary tissue tumors (pheochromocytoma). The normal plasma level is 2.1 to 6.5 μg/L with about 80% as norepinephrine.

Vanillylmandelic acid (*VMA*) is the urine metabolite of these two catecholamines. Its quantity in urine reflects the endogenous secretion of catecholamines. VMA can be determined colorimetrically, after extraction of the urine with ethyl acetate and diazotization with *p*-nitroaniline and ethanolamine in the presence of carbonate ion. VMA also can be measured spectrometrically following periodate oxidation to vanillin and solvent extraction. The normal output is 0 to 12 mg/24 hr.

Homovanillic acid (*HVA*) is not a metabolite of epinephrine or norepinephrine, but is produced from a common precursor, dopamine. Elevated HVA excretion is diagnostic in cases of neuroblastoma.

The biosynthesis of *serotonin* (5-hydroxytryptamine) and urinary excretion of its metabolite, 5-hydroxyindoleacetic acid (5-HIAA), are increased in argentaffine tumors. These have a very large capacity to metabolize tryptophane stores to serotonin. Urinary 5-HIAA increases from 1 to 7 mg/24 hr to as much as 1 g/24 hr in this type of tumor.

Bilirubin, a tetrapyrrole which is derived from senescent red-cell degradation, normally occurs in low concentration in the blood. In bile, it is present as the water-soluble conjugated acyldiglucuronide. In blood, bilirubin is tightly bound to plasma albumin. The reduction of bilirubin in the intestine yields urobilinogen which is, in turn, oxidized to a brown pigment—urobilin.

Serum bilirubin is determined by coupling with diazotized sulfanilic acid to form azobilirubin for colorimetric analysis. The *direct* or *conjugated bilirubin* test is performed in aqueous media; the *indirect* or *free bilirubin* analysis is performed in methanol or caffeine-sodium benzoate solution. Normal values in serum are: direct, 0 to 0.3 mg/100 mL; total, 0 to 1.5 mg/100 mL.

Clinical jaundice is a yellowing of the tissues associated with hyperbilirubinemia; in hemolytic disease of the newborn due to Rh and ABO incompatibilities, indirect serum bilirubin is elevated, whereas acute hepatitis results in increases in the direct type.

Electrolytes—The normal plasma electrolyte level is 154 mEq/L of cations and 154 mEq/L of anions. The osmotic effects of chloride, bicarbonate, sodium and potassium are important in the maintenance of normal muscle contraction and water distribution between cells, plasma and interstitial fluid.

Flame photometry, atomic-absorption spectrometry, neutron-activation analysis, X-ray fluorescence, ion-specific electrodes and colorimetric techniques are used in the identification and determination of cations or anions in biological fluids. Advances in technology have developed multiphase systems capable of measuring not only sodium and potassium but also chloride, carbon dioxide and calcium simultaneously.

Sodium and *potassium* serum concentrations are readily measured by flame photometry or highly sensitive and specific atomic-absorption spectrometry. The latter technique is similar to emission-flame photometry, except that it measures energy as it is absorbed by atoms rather than as it is emitted by atoms. Both techniques are based on the characteristic absorption or emission wavelengths of the cations. Ion-specific glass electrodes also are used for Na^+ and K^+ determinations, eliminating the use of a flame or combustible gas and can be performed on whole blood, plasma or serum.

Chloride levels in serum or urine are determined by titration with acid mercuric nitrate solution in the presence of s-diphenylcarbazone indicator. They also may be determined potentiometrically with a silver-silver chloride pH electrode assembly. The normal serum values are 135 to 155 mEq Na/L, 3.9 to 5.6 mEq K/L and 95 to 106 mEq Cl/L; urine levels are 150 to 197 mEq Na/day, 20 to 64 mEq K/day and 180 to 270 mEq Cl/day.

Serum sodium, potassium, chloride and *bicarbonate* determinations are useful indicators in adrenal cortical insufficiency, renal and cardiac failure, anuria, dehydration, alimentary tract diseases associated with diarrhea and vomiting and increased renal electrolyte excretion (diuretic therapy).

The determination of excess *chloride* (>50 mEq/L) in the perspiration of patients with pancreatic *cystic fibrosis* is an accurate diagnostic tool. Perspiration is stimulated by placing the patient's hand in a plastic bag for 15 to 20 min or, preferably, by an iontophoresis technique in which pilocarpine nitrate ions are transported through small areas of the skin to produce local perspiration. The chloride content may be quantitated with silver nitrate–potassium chromate-impregnated papers or with ion-selective electrodes.

Bicarbonate, phosphates, sodium, potassium and chloride concentrations are related to maintenance of acid-base balance in the body. The pH of the blood reflects the state of the acid-base balance and is related mathematically to HCO_3^- concentration and partial pressure of CO_2 (pCO_2) in blood by the Henderson-Hasselbach equation.

$$pH = 6.1 + \log \frac{[HCO_3^-]}{[H_2CO_3]} \qquad (2)$$

Blood pH, as measured electrometrically, has a normal range of 7.36 to 7.40 for venous samples and 7.38 to 7.42 for arterial samples. The *pCO_2 level* in blood is determined by measuring the pH of the blood at three different pCO_2 concentrations—one native to the blood and the other two obtained by equilibration with gas mixtures of known pCO_2. Blood bicarbonate levels also may be determined by measuring the amount of acid neutralized by plasma or serum and pCO_2 calculated by Eq 2. The relationship between pCO_2 and carbonic acid concentration is

$$[H_2CO_3] \approx 0.03 \times pCO_2$$
$$\text{mM per L} \qquad \text{torr} \qquad (3)$$

The role of oxygen and hemoglobin in respiration has been discussed previously. Measurements of blood pH and CO_2 content are used in differentiating respiratory acidosis (low pH, high CO_2) from metabolic acidosis (low pH, low CO_2).

Blood oxygen (pO_2) and *percent oxygen saturation* are measured by a polarographic method; the blood sample is placed in a chamber and separated from a combined platinum and silver-silver chloride electrode by a polypropylene membrane. By diffusion through the membrane, equilibrium is established between the pO_2 of the blood and a film of solution in contact with the electrode. A current, which is proportional to blood pO_2, is generated after the application of a polarizing voltage.

Calcium and phosphorus are important minerals in the processes of bone calcification, nerve irritability, muscle contraction and blood coagulation. Calcium is present in plasma as an ultrafilterable (ionic and nonionic) form and a protein-bound fraction. Blood phosphorus consists of inorganic phosphorus, organic phosphate ester (G6P, ATP) and phospholipids.

Serum and urine calcium levels are determined routinely by titration with EDTA or EGTA using a fluorescent calcein or calcichrome indicator. Other methods are based on the colorimetric analysis of calcium-methylthymol blue complex in the presence of 8-quinolinol to prevent interference by magnesium. Bis-(*o*-hydroxyphenylimino)ethane forms a colored complex with calcium and, in the presence of polyvinylpyrrolidone to inhibit phosphate interference, is a sensitive and specific method for calcium. Calcium is determined best by atomic-absorption spectrometry. As with all cations, calcium can be determined by emission- or absorption-flame photometry or ion-selective electrodes.

Inorganic phosphorus levels are determined by reaction with acid molybdate reagent to form phosphomolybdic acid which, in turn, is reduced with aminonaphtholsulfonic acid or *p*-dimethylaminophenol sulfate to give a blue complex which is estimated colorimetrically. Normal serum levels are 2.5 to 4.5 mg P/100 mL and 9 to 11 mg Ca/100 mL.

Calcium levels are decreased and phosphorus increased in hypoparathyroidism; an opposite effect is seen in hyperactivity of this gland. In rickets and osteomalacia, the concentrations of both elements are decreased. In establishing primary hyperparathyroidism and other causes of hypercalcemia, daily measurements for ionized calcium (Ca^{2+}) are replacing total Ca measurements using ISE technology.

Copper, magnesium, zinc and iron are trace elements in blood. They are quantitated readily by flame photometric, colorimetric or atomic-absorption techniques.

Organ Function Tests—The analyses of various blood or urine constituents, determination of metabolic excretion rates of exogenous compounds or endogenous metabolites and effect of exogenous stimuli on these parameters are used for evaluation of *in situ* activity and function of various organs. Organ function studies are performed in diseases associated with the liver, kidney, parathyroid, thyroid and pituitary gland, gastrointestinal tract, pancreas, adrenals and gonads. The principles and significance of the analysis used in such evaluations have been described also in other sections of this chapter.

Tests for *hepatic function* are based on bilirubin metabolism and excretion, carbohydrate metabolism (galactose tolerance test), plasma-protein changes (cephalin flocculation test and A/G ratio), abnormal fat metabolism, detoxification mechanisms (hippuric acid synthesis), excretion of injected substances [BSP], prothrombin formation and previously discussed enzyme levels.

Diseases of the liver are due to cellular alterations (hepatocellular) or obstructions to the flow of bile (obstructive jaundice). Hepatocellular liver disease can be chronic (postnecrotic cirrhosis, carcinoma) or acute (viral hepatitis, alcoholism, toxin- and chemical-induced).

The *cephalin flocculation test* is based on the flocculation of cephalin-emulsified cholesterol by γ-globulin. In normal serum an albumin-like protein will inhibit this reaction; in hepatic diseases, which produce abnormal γ-globulin or reduced albumin levels, the flocculation will occur.

The *detoxification mechanisms of the liver* can be evalu-

ated by intravenous administration of sodium benzoate and estimation of the benzoic acid metabolite, hippuric acid, in the urine. In hepatoparenchymal disease, a reduced capacity of the liver to form hippuric acid by conjugation of glycine and benzoic acid is observed.

The ability of the liver to excrete an injected dye is determined in the *BSP test;* the serum is analyzed for dye concentration at a suitable time interval after IV administration of 2 to 5 mg BSP/kg. Radioiodinated (^{131}I) Rose Bengal Sodium dye also has been used in dye-excretion studies with isotopic estimation of urine dye levels.

Kidney function tests are based on the determination of blood nonprotein nitrogen (urea, uric acid and creatinine), electrolytes, blood acid-base balance, routine urinalysis and the clearance of administered compounds in the urine. Most *clearance studies* are performed with substances that are not resorbed or secreted by the renal tubules: inulin, mannitol, sodium *p*-aminohippurate or ^{125}I-iothalamate sodium (sodium 5-acetamido-2,4,6-triiodo-*N*-methylisophthalamate). These are administered intravenously and the rate of urine clearance and glomerular filtration is estimated by analysis of the urine. The excretory capacity of the renal tubular epithelium can be determined by measuring the clearance rate of PSP. The dye is injected IV and the rate of its clearance in urine is determined. PSP is bound loosely to serum albumin and is removed rapidly from the blood by the renal tubules.

Sodium iodohippurate-(^{125}I), which is extracted almost completely from the blood on a single passage through the kidney, also has been used in renal function studies; a *renogram* or isotopic scan of both kidneys is performed. The test provides data on renal tubular secretion, renal vascular competence and renal evacuation and is primarily useful as a comparison of individual kidney function. It is important to note that 50% of kidney function can be compromised without any significant change in the routine renal function parameters.

Thyroid function tests usually measure the circulating levels of the thyroid hormones, and not the end-organ effect. The thyroid gland converts inorganic iodide to *thyroxine* (T_4) and *triiodothyronine* (T_3). T_3 and T_4 are stored in the colloid part of the gland as part of the thyroglobulin molecule. Hypothalamic *thyrotropin-releasing hormone* (TRH) mediates the release of the pituitary thyrotropin (*thyroid-stimulating hormone, TSH*). Excess levels of circulating T_4 depress, and low levels of T_4 increase, TSH release. TSH stimulates the proteolytic degradation of thyroglobulin to release T_4 and T_3, and increases organification of iodine. T_4 accounts for 90% of secreted thyroid hormones and exists in blood bound to *thyroxine-binding globulin* (*TBG*) or *thyroxine-binding prealbumin* (*TBPA*) or to albumin. T_3 is not protein-bound and has 5 to 10 times the biological potency of T_4 on a weight basis. Therefore, T_4 represents the major part of protein-bound iodine (PBI). The level of *free thyroxine* (FT_4), the active fraction in blood, is regulated by T_4 and T_3 release and the levels of binding proteins in blood and tissues.

The uptake of orally administered Na ^{131}I preparations by the thyroid gland can be estimated by isotopic scanning of the gland 24 hours after ^{131}I administration and is an index of glandular function (hyperactive, >50% uptake; hypoactive, <15%).

PBI determinations are based on the precipitation of protein-bound thyroxine, removal of inorganic iodine by basic- or anion-exchange chromatography, alkaline incineration to convert thyroxine to inorganic iodide and, finally, quantitation of iodide by reaction with arsenous acid and ceric ammonium sulfate. PBI is a good estimate of total circulating hormonal iodine. The normal range is 4 to 8 μg/100 mL serum.

T_4 can be determined by column chromatography in which it is separated and isolated by ion-exchange chromatography, and then analyzed colorimetrically. Nonisotope thyroid assays have been developed using fluorescence polarization methods for T_4 and free-thyroxin index. In the competitive protein-binding assay for T_4, serum T_4 competes with ^{125}I-T_4 for binding sites on a known amount of TBG. The ratio of bound to free ^{125}I is determined by adsorption of ^{125}I-T_4 not bound to TBG on an anion-exchange resin embedded in a polyurethane sponge or a porous dextran gel, and is a direct index of T_4 levels. The presence of mercurials, inorganic iodide or iodinated radiographic compounds in serum interferes with the T_4 column and PBI procedures. The competitive-binding procedure is affected by the presence of highly protein-bound drugs or changes in TBG levels in serum. The normal range of serum T_4 is 2.9 to 6.4 μg/100 mL by column and 3.0 to 7.0 μg/100 mL by binding assay. T_4 and PBI are increased in hyperthyroidism and the early stages of hepatitis. T_4 and PBI are decreased in hypothyroidism and nephrosis.

FT_4 also is determined in a competitive protein-binding assay in which ^{125}IT_4 and serum are incubated, and then dialysed to determine the percent dialyzable ^{125}I-T_4. FT_4 analysis is used in suspected abnormalities in protein-binding globulins. T_4 binding capacity of serum TBG, albumin and prealbumin can be determined after electrophoretic separation of these proteins.

T_3 analysis is determined by the resin-uptake test. The uptake of ^{125}I-T_3 by a resin is determined in the presence of the test serum. In hyperthyroidism, the primary TBG-binding sites are saturated and ^{125}I-T_3 is taken up by the resin. The resin uptake is decreased in hypothyroidism, and most of ^{125}I-T_3 is bound to TBG in serum. A *free thyroxine index* can be obtained by multiplying T_3 (resin) × T_4 (competitive binding) × 0.01. This product deviates from normal in the same direction as T_3 and T_4 in hyper- and hypothyroidism. This product is stable during euthyroidism in spite of changes in binding proteins; eg, a euthyroid patient on phenytoin therapy will show a decreased TBG and T_4 and increased T_3, but (T_4 × T_3) is normal. The indication of hyper- or hypothyroidism in the presence of abnormal amounts of TBG is observed in the (T_3 × T_3) product.

The determination of *TSH* by RIA appears to be the most useful test in discriminating patients with primary hyperthyroidism from the euthyroidism or hypothyroidism secondary to pituitary disease. Serum TSH is increased in the primary disease state.

The *PBI conversion ratio* is an estimate of the rate of conversion of inorganic iodide to PBI. Radioiodide-(^{131}I) is administered to the subject; after 24 hr, a sample of blood is obtained and the ^{131}I to PB^{131}I is estimated by radiochromatographic procedures with ion-exchange resins (normal conversion, 13 to 42%).

Adrenocortical function is evaluated by estimation of serum or urinary 17-ketosteroids (17-KS) and 17-hydroxycorticosteroids (17-OH-CS) (androgen and corticosteroid metabolism), serum electrolytes (aldosterone metabolism) and blood adrenocorticotrophic hormone (ACTH) levels in the basal state, after stimulation with IM or IV ACTH, or after adrenal inhibition with dexamethasone. In the normal individual, ACTH will increase plasma cortisol and urine 17-OH-CS, and dexamethasone will suppress plasma cortisol. Metapirone, an inhibitor of 11β-hydroxylase, will cause selective secretion of compound S (11-deoxycortisol) by the adrenals in place of cortisol. Compound S will not inhibit the adrenal-pituitary feedback mechanism, the pituitary will secrete more ACTH and the adrenal will secrete more compound S. The determination of urinary 17-OH-CS or tetrahydro-compound S (THS) following metapirone administration is a good index of the functional integrity of the pituitary-adrenal axis; patients with virilizing adrenal hy-

Table IV—Reference Values[a]

Electrolytes		
Calcium	9.0–10.6 mg/dL	2.25–2.65 mmol/L
Chloride		98–109 mmol/L
CO_2 content		23–30 mmol/L
Magnesium	1.2–2.4 mEq/L	0.6–1.2 mmol/L
Phosphorus	2.5–5.0 mg/dL	0.81–1.62 mmol/L
Potassium		3.7–5.3 mmol/L
Sodium		136–146 mmol/L
Metabolites		
Bilirubin	0.1–1.2 mg/dL	1.7–20.5 μmol/L
Cholesterol	150–250 mg/dL	3.9–6.5 mmol/L
Creatinine	0.7–1.5 mg/dL (adults)	62–123 μmol/L
Glucose	60–95 mg/dL	3.33–5.28 mmol/L
Iron	50–165 μg/dL	9.0–29.5 μmol/L
Triglycerides	20–180 mg/dL	0.22–1.98 mmol/L
Urea nitrogen (BUN)	8–26 mg/dL	2.9–9.3 mmol/L
Uric acid	2.5–7.0 mg/dL	0.15–0.41 mmol/L
Proteins and enzymes		
Alanine aminotransferase	(ALT, SGPT)	5–40 U/L at 37°
Albumin	3.5–5.0 g/dL	35–50 g/L
Alkaline phosphatase	35–120 U/L at 37° (adults)	50–400 U/L at 37° (children)
Amylase	60–180 Somogyi Units	110–330 U/L
Aspartate aminotransferase	(AST, SGOT)	8–40 U/L at 37°
Carcinoembryonic antigen (CEA)	<2.5 ng/mL	<2.5 μg/L
Creatine kinase (CK)		10–180 U/L at 37°
Glutamyl transferase (GGT)		5–40 U/L at 37°
Lactate dehydrogenase (LDH)	60–220 U/L at 37°	(lactate → pyruvate)
Total protein	6.0–8.0 g/dL	60–80 g/L
Hormones		
Cortisol in plasma	7–20 μg/dL (at 8:00 AM) (200–550 mmol/L)	3–13 μg/dL (at 4:00 PM) (80–360 mmol/L)
Cortisol (free) in urine	20–90 μg/24 hr	55–248 nmol/24 hr
Follicle-stimulating hormone (FSH)	Adult males	Adult Females
	2–15 mIU/mL	Follicular phase 3–15 mIU/mL Ovulatory spike 10–50 mIU/mL Luteal Phase 3–15 mIU/mL Postmenopause 30–200 mIU/mL
17-Hydroxycorticosteroids in urine	3–10 mg/24 hr	
17-Ketosteroids in urine	5–15 mg/24 hr	(adult females)
	8–20 mg/24 hr	(adult males)
	0.1–3.0 mg/24 hr	(prepubertal children)
Luteinizing hormone (LH)	Adult males	Adult females
	5–25 mIU/mL	Follicular phase 5–30 mIU/mL Ovulatory spike 50–150 mIU/mL Luteal phase 5–40 mIU/mL Postmenopause 30–200 mIU/mL
Metanephrine in urine	<1.3 mg/24 hr	
Prolactin	1–20 ng/mL (males) (1–20 μg/L)	1–25 ng/mL (females) (1–25 μg/L)
Thyroxine (Ty)	5.5–12.5 g/dL (adults) (72–163 nmol/L)	7.8–16.0 μg/dL (newborns) (101–208 nmol/L)
Vanillylmandelic acid (VMA) in urine	<6.8 mg/24 hr	

[a] Serum specimens unless otherwise indicated.[18]

perplasia excrete excessive THS due to a 11β-hydroxylase defect.

Common, chemistry, reference values are listed in Table IV.[18]

Automated Analysis—The automation of analytical techniques used in blood and urine chemistry, hematology, blood typing and immunology has increased the productivity and accuracy of the clinical laboratory.[19] *Computerization* of the automated analytical system also has increased the rapidity of reporting test results, reduced clerical error and provided a unified and updated report of the laboratory tests for each patient.

In the SMA-12 (or SMA-20) Autoanalyzer (*Technicon*), a continuously operating, multiple-channel proportioning pump moves the samples, diluents and reagent streams. Air bubbles segment the flowing streams of samples and reagents, which then may flow through dialyzers to remove interfering substances, move them into chambers preset at desired temperatures and, finally, into detection devices (colorimeters, fluorometers, flame photometers, spectrophotometers). A serum standard is run simultaneously with the samples. The results can be read directly from a recorder or can be coupled into a digital computer output. Sequential, multiple analyses in the SMA-12 are accomplished by distributing the sample to 12 different analytical streams, so that all 12 analyses are in progress at the samo time. The *SMA-12 profile* usually determines calcium, inorganic phosphorus, glucose, BUN, uric acid, cholesterol, total protein, albumin, total bilirubin, alkaline phosphatase, LDH and AST. The Mark X (*Hycel*), Ektachom 400 (*Kodak*), ACA (*Dupont*) and DSA-560 (*Beckman*) also are used in automated clinical-laboratory techniques.

Technicon recently developed the "capsule chemistry" analysis on the Chem 1 analyzer. Microaliquots of the sample (1 μL) and reagents (14 μL) are encapsulated within an inert fluorocarbon liquid. The resulting "test capsule" is introduced into a single, analytical flow path (composed of a solid fluorocarbon liquid, Teflon) where the sample is incubated, mixed, reacted and measured as a moving series of individual tests. The reactions are monitored at in-line detector stations for colorimetric and nephelometric measurements. On each sample 35 chemistries can be run sequentially.

The rapid growth of more-sophisticated chemistry analyzers increases the capacity of any clinical laboratory and is associated with small-specimen requirements incorporating batch analysis, profiles and stat capabilities. In addition to routine chemistry testing, the systems test for enzymes, immunoassay, therapeutic-drug tests, coagulation (fibrinogen, antithrombin III, plasminogen) and electrolytes. Techniques eliminating liquid requirements of other reagent systems are available from *Kodak* and *Ames* using dry reagents,

which are impregnated in pads on a strip or slide and read by a reflectance photometer.

Automated hematology and simultaneous determination of RBC, WBC, hemoglobin and hematocrit, MCV, MCH and MCHC can be performed on the SMA-7A (*Technicon*) Analyzer. The automated Technicon Hemalog system will provide data of SMA-7A and CCV (conductivity cell volume), prothrombin time, partial thromboplastin time and platelet count. *Automated leukocyte differential* was discussed previously under *Hematology*.

Urine

The formation of urine and its excretion are critical physiological activities of the body which provide a mechanism for the maintenance of a constant internal environment for all cells, tissues and organs. This internal ecology of the body is well-recognized and known as homeostasis. Inasmuch as the urine reflects what is occurring within the body, it offers a fluid which is an important source of information that is most useful as an aid in the definition of the states of health and disease. More specifically, the kidney, by means of urine formation

1. Regulates the body water.
2. Excretes metabolic waste products, many of which are of a nitrogenous nature.
3. Excretes toxic substances of both endogenous and exogenous origin.
4. Regulates the electrolyte equilibrium of the body by either excreting or retaining each specific ion.
5. Maintains the delicate balance of pH within the body by excretion of excess acid or excess base.
6. Provides an important route for the elimination of pharmaceutical agents and their breakdown products from the body.

Normal urine contains several thousand compounds most of which occur in minute quantities. Table V identifies some of the constituents of normal urine which are of particular significance.

Urine is studied quite widely as a means of identifying abnormalities associated with disease. The importance of such study is emphasized by the fact that the number of tests carried out on urine far exceeds those made on all other body fluids combined. Urine not only is important in providing information relating to kidney disease, but it may provide information relative to many other body activities. Information from urine studies is of diagnostic value in functional diseases of the kidney, liver, pancreas, blood, bone, muscle and the urinary, gastrointestinal and cardiovascular systems. Urine studies provide vital clinical information on electrolyte and water balance, acid-base equilibrium, intermediary metabolism, inborn errors of metabolism, drug abuse, intoxication, pregnancy and hormone balance. Most of these parameters have been discussed earlier and this section will be devoted to routine urinalysis.

Table V—Normal Constituents of Urine

Constituent	g/day	Constituent	g/day
Water	1400	Amino acids	2.1
Total solids	60	Purine bases	0.01
Urea	30	Phenols	0.03
Uric acid	0.4	Proteins (total)	0.025
Hippuric acid	0.9	Chloride (as NaCl)	12
Creatinine	1.2	Sodium	5
Indican	0.01	Potassium	2
Citric acid	0.8	Calcium	0.2
Lactic acid	0.2	Magnesium	0.15
Oxalic acid	0.03	Sulfur (total)	1.0
Nicotinic acid	0.00025	Phosphate (as P)	1.1
Allantoin	0.04	Ammonia	0.7

It is important to recognize that urine test information, like all other laboratory data, helps provide a picture of the whole body, but any single result requires interpretation to be most meaningful. It also should be recognized that negative results can be essentially as useful as positive results in a great many instances. The ready availability of urine is an advantage that makes it practical as a material for monitoring the course of the treatment of disease as well as for its recognition and definition.

Most urine examinations include observations with regard to the majority of the following—color, odor, turbidity, pH, protein, glucose (or reducing substances), ketone bodies (acetone), occult blood, bilirubin, urobilinogen, bacteria (culture or chemical tests), specific gravity and microscopic examination of sediment, including erythrocytes, leukocytes, casts, epithelial cells, crystals, bacteria, parasites and exfoliative cytology. A *"routine" urinalysis* varies in different institutions but ordinarily involves the inclusion of the majority of the above tests.

Urine for laboratory study should be collected in clean containers—preferably into a disposable unit (polystyrene tube) with a capacity of 15 mL which can be used for collecting, transporting, centrifuging and testing. Refrigeration is desirable for any specimen which is not tested within 1 to 2 hours.

If urine is to be transported through the mails or is to be held for a significant time at room temperature, it is desirable to add a urine preservative (formalin, methenamine, thymol, toluene) which will interfere with microbial growth in the specimen. Several proprietary urine preservative tablets are available. If urine is allowed to stand at room temperature, bacteria will grow in the specimen and cause degradation of many constituents. Frequently, the bacteria decompose urea into ammonium carbonate with a resulting increase in the alkalinity of the specimen. Formed elements, particularly casts and red blood cells, disintegrate in alkaline solution.

The majority of urine tests are done on random specimens but, in certain instances, it is necessary to have a 24-hr specimen for certain specialized analyses. For urine-sugar testing in diabetes detection, it is desirable to use a postprandial urine specimen (ie, after a meal). For protein tests, as well as chemical or culture tests for bacteriuria, the first morning specimen is preferred. Most laboratories use commercially available, standardized, reagent-impregnated strips ("dipstrips") or tablets (*Ames*) for routine urinalysis.

Instrumentation in Urinalysis—Automated urine-testing systems, semiautomated reagent-strip readers and a system which performs the complete urinalysis procedure have been developed. The strip reader is a reflectance photometer which measures urine pH, protein, glucose, ketones, blood, bilirubin, nitrate and urobilinogen. The IRIS AIM (*International Remote Imaging Systems*) measures urine specific gravity by refractometry, urine sediment by staining and classifies analytes, controlled fluid dynamics, video microscopy with an image processor, a chemistry system to read a standard dipstick by reflectance photometry, and color and appearance. These systems achieve standard results for routine urinalysis and increase accuracy and precision.

Volume—The normal volume of urine excreted during a 24-hr period is usually in the range of 1000 to 1500 mL. It is possible for a healthy person to modify the volume either by severe fluid restriction or by ingestion of excessive quantities of fluid. In certain disorders there is a change in urine volume. Urine-volume increases are identified as polyuria and are encountered in diabetes mellitus, diabetes insipidus and in certain stages of chronic renal disease. Urine volume is increased during diuretic therapy and with the ingestion or injection of large volumes of fluid. A decrease in urine

volume usually occurs in dehydration, water restriction and in acute or terminal renal disease. Extensive water loss from severe diarrhea or vomiting causes oliguria or decreased urine volume. Acute renal failure precipitated by shock, poisons or transfusion reaction may result in a complete absence of urine excretion or anuria. In the majority of instances urine study does not require volume measurements, but these are quite critical in severely ill persons where oliguria or anuria is present.

Specific Gravity-Osmolality—The urine density or specific gravity is related to the amount of solids excreted in a given volume of urine. In the majority of instances in healthy persons the specific gravity varies between 1.010 and 1.030 and is related to dietary habits of fluid and food ingestion and, secondarily, to the loss of fluid by other routes such as extensive sweating. The measurement of urine density or specific gravity is a part of "routine urinalysis," and as such provides information with regard to water and solids turnover in the body. The specific-gravity information alone is not nearly so important as it may be in conjunction with other observations. Thus, if dehydration is suspected, a specific gravity in the midrange of 1.015 would cast a doubt about dehydration unless there was a concurrent renal dysfunction.

The kidney possesses a remarkable ability to either form a concentrated urine or a very dilute urine ranging from a specific gravity of 1.001 to 1.032. This concentrating or diluting capacity is diminished in cases where there is a loss of renal function. In fact, one of the sensitive tests for measuring renal function involves the so-called dilution-concentration tests where fluid is administered or withheld, and the specific gravity of the urine is measured. With a serious loss of renal function, the kidney cannot excrete a urine in excess of 1.020 even with marked fluid restriction. In advanced renal disease the specific gravity of the urine may become "fixed" or constant in the range of 1.010 to 1.012 with all urine being of this specific gravity regardless of whether there is overhydration or dehydration.

Specific gravity is measured readily with a special hydrometer, called a urinometer. There is a correlation between the density of urine and its refractive index, and a special refractometer has been designed which gives readings in specific-gravity units on a single drop of urine.

Certain abnormal constituents of urine, such as glucose or protein, when present in high concentrations, will cause significant increases in specific gravity. Certain X-ray contrast media, when excreted in the urine, also will cause marked increases in specific gravity.

Urine specific gravity is only an indirect index of solute concentration, ie, 1 mole of urea will produce a lower specific gravity than 1 mole of glucose. Osmolality is a direct measure of the molal concentration of solutes in solution regardless of their molecular weight, ie, 1 mole of NaCl dissociates into 1 mole of chloride ion and 1 mole of sodium ion. Osmolality is determined in a direct-reading osmometer by comparing the freezing point of urine with that of a standard sodium chloride solution.

The kidneys normally excrete 800 to 1400 mOsm/kg (an osmol is that weight of any substance when dissolved in water depresses the freezing point 1.86°) of solutes per day. Man concentrates urine and eliminates the daily solute load at a maximum volume of 1200 mOsm/kg water. Urine osmolality is an inverse function of urine volume in the normal catabolic state. Urine volume is regulated by the antidiurotic hormone (ADH) and sodium excretion by the hormone aldosterone. Increased osmolality of body fluids stimulates, and increased dilution inhibits, the release of ADH. The major determinant of body-fluid osmolality is sodium. Sodium conservation is mediated through the renin-angiotensin-aldosterone axis. Determinations of plasma and urine sodium, and osmolality and urinary volume, are of diagnostic value in Addison's disease, vasomotor nephropathy (acute tubular necrosis), inapparent volume depletion, incomplete urinary tract obstruction and hepatorenal disease.

pH—Freshly voided urine usually has a slightly acid pH. The normal range is 5 to 8 and, essentially, this is also the abnormal pH range. The kidneys, by reason of excreting a urine of variable pH, provide a regulatory mechanism for the body to get rid of excess acid or alkaline waste products. Since the normal pH range and the abnormal pH range are comparable, the measurement of pH alone provides minimal information, but when used in conjunction with other information, it is a very useful urinary parameter. In conditions of acidosis, the urine is quite acid; in conditions of alkalosis, the urine pH is above 7. When metabolic or respiratory acidosis is suspected, an alkaline-urine pH result almost eliminates the possibility of acidosis. Conversely, if respiratory or metabolic alkalosis is suspected, the excretion of an acid urine indicates that alkalosis is likely not present.

Dip-and-read tests are used widely for pH testing, but pH-meter measurements are used less commonly. In certain situations involving kidney stone susceptibility, it is quite important to maintain a narrow range of urinary pH. For example, in cystinuria an alkaline pH is maintained to keep the cystine solubilized and to avoid as much as possible the crystallization of cystine into renal calculi. The maintenance of urinary pH is also important for optimum results in certain types of drug therapy.

Color—Urine normally has a yellow color, mostly due to urochrome; the color varies from pale straw to dark amber. Darker specimens usually have a high specific gravity. Occasionally, either normal or abnormal urine may show a color different from yellow. Bilirubin may cause fresh urine to be dark in color. In addition, urine which is allowed to stand darkens because of the oxidation of urobilinogen to urobilin. Red, reddish-brown or "smoky" urine usually is due to the presence of hemoglobin (hemoglobinuria), myoglobin (myoglobinuria) or red blood cells (hematuria). Porphyria is an uncommon cause of red coloration. Black urine can be caused by melanin, which may occur in the urine of patients with far-advanced malignant melanoma. An inborn error of metabolism, alkaptonuria, is characterized by the urinary excretion of homogentisic acid, which causes the urine to turn dark brown or black on standing. Many of the unusual colors occasionally found in urine are derived from exogenous sources, including both foods and drugs. Among these are the red color caused by beets, particularly in infants, the golden-yellow or orange-red color of metabolites of pyridium-like drugs or azo drugs and the green or blue color from methylene blue.

Odor—Normal, freshly voided urine has a faint aromatic and characteristic odor, which is more intense in concentrated specimens. If the urine is allowed to stand, the odor becomes strongly ammoniacal and unpleasant because of bacterial destruction of urea. Freshly voided urine having a foul odor indicates severe infection. A sweet, fruity odor may be due to ketones.

Appearance—Freshly voided urine is usually clear. On standing, a precipitate may form which usually consists of amorphous urates if the urine is acid or calcium and magnesium phosphates if the urine is alkaline. The formation of a precipitate is more likely to occur if the urine is refrigerated. Most specimens will become clear again if they are warmed gently to room temperature. Large quantities of mucus, cells, leukocytes or bacteria may cause cloudiness. Protein usually does not cause cloudiness.

Protein—A small amount of protein is present in the urine obtained from healthy subjects although the quantity is not sufficient to give a positive reaction with the tests

commonly used for the recognition of protein in urine. The majority of the 25 to 50 mg of protein that is excreted daily is microprotein (low-molecular-weight polypeptide), with properties quite different than those of albumin and globulin, which are the principal proteins of the blood serum. Albumin and globulins do occur in the normal urine in minute concentrations.

Plasma proteins, hemoglobin, abnormal Bence-Jones protein and proteins (nucleo-, phospho- and glyco-proteins) derived from leukocytes and mucus may be present in urine in nephritis, nephrosis, lesions of the urinary tract, GI dehydration and renal congestion. Abnormal amounts of protein in the urine may be recognized by either precipitation or colorimetric tests. The precipitation depends on the heat coagulation of the protein or on the chemical precipitation of the protein. The most popular of the heat-precipitation tests is the heat-and-acetic acid test in which a tube of urine is heated to boiling after the addition of a drop or two of acetic acid. Sulfosalicylic acid is employed commonly in chemical precipitation tests and, in this test, equal quantities of 3% sulfosalicylic acid and urine are mixed in a test tube and the mixture examined for turbidity indicative of precipitated protein.

Colorimetric tests for proteins involve *dip-and-read* type of systems and are based on the *protein error* of indicators. Certain indicators have a point of color change which is different in the presence of protein compared to the same system in the absence of protein. Thus, by buffering the indicator tetrabromophenol blue on this dip-strip at a specific pH, it is possible to have a yellow color in the absence of protein and a green or blue color in the presence of protein. This test, Albustix (*Ames*), not only indicates the presence or absence of protein in the urine but also can be made to indicate the approximate amount of protein. Strongly alkaline or fermented urines will give false-positive results. The sensitivity of the colorimetric method is such that quantities of 10 to 20 mg of albumin per 100 mL of urine are recognized with confidence.

A positive test for protein in the urine may have any one of several meanings, and it is only when this information is related to other observations that it has optimum value. Proteinuria may be benign and appear following strenuous exercise or simply as a result of standing (orthostatic proteinuria). Protein frequently occurs in the urine during pregnancy and in some instances this is benign, but in other cases it indicates renal complications. Transient proteinuria may occur following severe infections, high fever, exposure to cold and in congestive heart failure. Proteinuria may be an early and sensitive indicator of renal disease and may indicate an abnormality prior to other signs and symptoms of renal impairment in the glomerulus or tubules. In the majority of instances there is not a correlation between the amount of protein in the urine and the severity of the renal disease.

Patients with severe nephrosis may lose up to 25 g of protein per day. Such a marked loss of protein causes a decrease in plasma protein concentration with an accompanying edema. In both chronic and acute glomerulonephritis there is protein in the urine. Tumors of the kidney and renal infection usually will have an accompanying proteinuria. Bence-Jones protein is a unique protein which occurs in the urine of about 50% of patients with multiple myeloma. It has the unusual property of precipitating between 50 and 60° and dissolving at higher temperatures.

Glucose (Reducing Substances)—Glucose normally occurs in urine in such low concentration that it escapes detection by the usual testing methods. The urine of untreated or poorly controlled diabetic patients characteristically contains easily detectable amounts of glucose. A positive test for glucose in urine usually suggests hyperglycemia and the diagnosis of diabetes mellitus; further studies, such as the glucose tolerance test to confirm the diagnosis, are indicated. Glycosuria also may occur when the renal tubules fail to reabsorb glucose normally, and glucose appears in the urine despite normal blood glucose levels, in contrast to true diabetes.

Glucose is the sugar almost always found in urine; however, lactose, galactose, levulose, sucrose and pentoses may be encountered. These other sugars are identified by paper chromatography, selective fermentation, polarimetry, special chemical tests or the formation of their osazones. Other reducing substances occur in urine and may cause falsely positive reducing reactions for glucose. Examples are ascorbic acid, glucuronides, many drugs, homogentisic acid and the preservatives formalin and chloroform.

The traditional test for glucose in urine (Benedict's test) relies on the reduction of cupric ions in alkaline solution to reddish-orange insoluble cuprous oxide. The copper is reduced totally by large amounts of glucose and results in a brick-red sediment with no remaining blue color. Lesser concentrations form green- to rust-colored solutions with some red sediment. A modification of this test, Clinitest (*Ames*), is available in tablet form. The tablet contains copper sulfate, anhydrous sodium hydroxide, citric acid and sodium carbonate. When added to dilute urine, the tablet dissolves and generates enough heat and effervescence to yield results comparable with the Benedict test.

A specific but extremely simple enzyme test for glucose is available—Tes-Tape (*Lilly*), Clinistix (*Ames*) and Multistix (*Ames*). Reagent strips are impregnated with glucose oxidase, peroxidase and orthotolidine. When dipped into a solution of glucose, oxidation occurs and hydrogen peroxide is formed which oxidizes orthotolidine to a blue color. This test is more sensitive than Clinitest, but is not as reliable for estimating the concentration of glucose. The enzymatic test is specific and thus useful in determining whether or not a reducing substance is glucose. Diastix (*Ames*) is a specific urine glucose test using glucose oxidase, which also indicates the quantity of glucose present.

Ketone Bodies—The ketone bodies acetone, acetoacetic acid and beta-hydroxybutyric acid are present in the urine when fats are metabolized incompletely. Ketonuria is seen most commonly in poorly controlled diabetes and indicates ketonemia and diabetic acidosis. Other causes for ketonuria are starvation, fever, protracted vomiting and Von Gierke's disease. Ketonuria also occurs following anesthesia. Acetoacetic acid and acetone produce a distinctive purple color when treated with a mixture of sodium nitroprusside, ammonium sulfate and concentrated ammonium hydroxide. A similar reagent is available in tablet form (Acetest, *Ames*). A drop of urine is placed on the tablet; if ketones are present, a lavender to deep-purple color develops in 30 sec. The color intensity indicates the concentration of ketones. The reagent strip Ketostix (*Ames*), used as a dip-and-read test on urine or serum, contains the same reagents, which are available on Multistix (*Ames*) and other multiple reagents as well. These tests will detect 5 to 10 mg of acetoacetic acid per 100 mL of urine.

Phenylpyruvic Acid—Phenylketonuria (or PKU) is an inborn error of metabolism in which the normal conversion of phenylalanine to tyrosine in the body does not occur and there is a buildup of phenylalanine concentration in the blood. This metabolic disorder causes mental retardation. A portion of the phenylalanine is excreted by the kidneys into the urine and in the process is converted to phenylpyruvic acid (or phenylketone). If this genetic disorder is discovered soon after birth, it is possible to place the infant on a diet very low in phenylalanine-containing proteins and thus minimize the phenylalanine buildup in the body, averting

the serious mental retardation which ordinarily is seen in the untreated PKU patient.

Recognition of PKU can be made by the use of a test for phenylpyruvic acid using a dip-and-read reagent composition containing ferric ions. This test, Phenistix (*Ames*), can be used on urine from all newborn babies. A positive reaction gives a green color, whereas a normal infant's urine gives a pale-ivory or yellow color to the strip. PKU also can be recognized by employing a chemical or microbiological test for elevated phenylalanine in serum, as discussed under *Amino Acids*.

Bilirubin—Bilirubin is found in the urine of patients with hepatitis or obstructive jaundice but not in patients with hemolytic jaundice. Tests for bilirubin and urobilinogen combine to give excellent information in the differential diagnosis of jaundice. Tests for bilirubin are of two kinds; oxidation tests form a green color of biliverdin from bilirubin usually using ferric chloride as the oxidative reagent, and diazotization tests form colored compounds when bilirubin reacts with diazonium salts in a strongly acid medium. Most oxidation tests adsorb the bilirubin onto barium sulfate or similar material before the addition of Fouchet's reagent. The tablet test Ictotest (*Ames*) is the most sensitive diazo test and it uses an absorption mat to concentrate the bilirubin from 5 drops of urine. A reagent tablet is added to the moist spot on the mat and 2 drops of water are added to dissolve the effervescent reagent and wash some of it off the tablet onto the mat where the reaction takes place. A blue or purple color on the mat around the tablet in 30 sec indicates the presence of bilirubin. In addition, a dip-and-read test composition also based on the diazo reaction has been incorporated into the multiple urinalysis reagent strips, Bili-Labstix and Multistix (*Ames*). It is less sensitive than the tablet test, but its convenience allows it to be used in routine urinalysis quite readily. An incidence of approximately 0.1% positives on health-screening population groups, 0.2% on clinic patients and 0.9% on hospitalized patients has been reported.

Urobilinogen—Bilirubin in the bile is reduced to urobilinogen by bacteria in the lower intestine. A portion of the urobilinogen is reabsorbed from the intestine into the blood. A portion of this urobilinogen is excreted into the urine by the kidney and the balance is re-excreted via the bile into the intestine. Although the quantity of urobilinogen in the urine is quite small, it is an important indicator of liver function and red-blood-cell catabolism.

If there is an obstruction to bile flow such as in obstructive jaundice, the amount of urobilinogen formed and reabsorbed into the blood and excreted in the urine is decreased. With impairment of liver function, the excretion of urobilinogen in the bile is decreased, the blood concentration increases and there is a corresponding increase in urinary urobilinogen excretion. Actually, the increase in urinary urobilinogen is one of the most sensitive tests for impaired liver function and this test may indicate an abnormality when all other tests of liver function remain unchanged from normal.

In hemolytic diseases where there is an increased rate of hemoglobin breakdown, the amount of bilirubin formation is increased with a corresponding increase in urobilinogen formation and excretion in the urine. The concentration of urobilinogen in urine can be established by the use of a dip-and-read test which uses the interaction of urobilinogen and *p*-dimethylaminobenzaldehyde (Urobilistix, *Ames*).

Hematuria, Hemoglobinuria and Myoglobinuria—Hematuria refers to a condition in which intact red blood cells appear in the urine. This condition is indicative of a specific defect in the microscopic functional unit (the nephron) of the kidney or it may be indicative of bleeding in the kidney, the ureter, the bladder or the urethra. In the female there may be variable numbers of red blood cells in the urine during menstruation.

Hemoglobinuria is a condition in which free hemoglobin is present in the urine without red blood cells. This may be caused by intravascular hemolysis as a result of a transfusion reaction or by poisoning or toxins. The free hemoglobin in the plasma is excreted by the kidney into the urine. In some situations actual total hemolysis of the red cells occurs after they have entered the urine. This occurs particularly with alkaline urines.

Myoglobin is the red respiratory pigment of muscle. This pigment is quite comparable to hemoglobin in its composition and chemical reactions. Myoglobin may be liberated from muscle cells in certain types of injury and, in such cases, will circulate in the plasma and be excreted in the urine. There are also certain genetic muscle disorders in which myoglobin is lost from the muscles and appears in the plasma and subsequently in the urine.

Chemical tests for red cells, free hemoglobin and myoglobin are based on the peroxidase-like activity of hemoglobin or myoglobin. When a chromogen mixture such as orthotolidine and peroxide is exposed to this peroxidase activity, it will interact rapidly to generate an intense blue color. A dip-and-read solid state system is available which is called Hemastix (*Ames*). This specific composition uses cumene hydroperoxide as the peroxide. The same dip-and-read test for occult blood is incorporated as a component part of multiple, urine dip-and-read tests, eg, Multistix (*Ames*).

Microscopic Examination—Ordinarily, urine contains a number of formed elements or solid structures of microscopic dimensions. These are studied readily by centrifuging 10 to 15 mL of urine, pouring off the supernatant and resuspending the sediment in the drop or so of urine which remains in the tube. This suspension of sediment is placed on a microscope slide and viewed with low-power magnification. Specific structures can be studied with higher magnification. The urinary sediments can be classified into unorganized (chemical substances) and organized (cells and casts) constituents.

In an alkaline urine, amorphous or crystalline ammonium-magnesium phosphates, calcium carbonate or oxalate crystals and ammonium urate may occur normally. Amorphous or crystalline urates, uric acid and calcium oxalates normally are seen in acid urines. The presence of tyrosine, leucine or cystine crystals is associated with various diseases. Chemical crystals are identified by solubility in acid and/or alkali, colorimetric reactions and crystalline structure.

The urine sediment ordinarily contains residues of epithelial cells, crystals and an occasional red or white blood cell. Increased numbers of erythrocytes are seen where there is bleeding into the urinary tract. If the red cells are formed into a red-cell cast, it is suggestive that bleeding has occurred at the glomerular level. An increased number of leukocytes is suggestive of infection and inflammation of the kidney. Casts are microscopic concretions which have the form of a tubule; they have a matrix of precipitated protein and, depending on their appearance, may be identified as hyaline, granular, waxy or red-cell casts. Renal-failure casts are larger and are associated with severe necrosis of the kidney.

Numerous crystals, mucus fibers, bacteria, yeast cells, spermatozoa and parasites (*Trichomonas vaginalis*) may be indentified in the urine sediment. The majority of these crystals do not have any unusual significance but in certain disorders may be indicative of crystal deposits in kidney tissue or predisposition to formation of calculi.

Tissue cells can be recognized in urine sediment. This provides an excellent means of detection and diagnosis of cancer of the lower urinary tract when the sediment is fixed in alcohol and stained by the Papanicolaou procedure. Ex-

foliative cytology of urine may be applied as a routine to all urology patients. In one large clinic the number of positive cases found among urology patients was almost 5%, which is a much higher return of positive results than is obtained with routine staining of cervical smears.

Bacteria—Freshly voided specimens of urine ordinarily contain a few microorganisms, which primarily represent bacteria picked up from the external genitalia. There are fewer contaminating organisms in a *clean-catch* specimen, which involves extensive washing of the external genitalia prior to collection of the specimen. A specimen collected at the midpoint of urination or a "midstream" specimen ordinarily has more organisms than a clean-catch specimen, but fewer than a so-called random specimen. When there is an infection of the kidney or urinary tract, the number of organisms in the urine is increased markedly. Ordinarily, if the urine contains 100,000 or more organisms per mL, the result strongly suggests the presence of an active infection. Infection of the urinary tract with accompanying bacteriuria is relatively common in young girls and women. Quite often the condition is asymptomatic and is recognized only as a result of a study of the urine. If bacteriuria is not treated, it may lead to serious renal injury.

If there is a very large number of bacteria in the urine, the specimen actually may be turbid. This can be recognized by gross visual inspection of the urine. Bacteriuria also can be recognized by microscopic examination of the urine sediment particularly if there is a large number of organisms present. The most widely employed procedure for recognizing bacteria involves plating a specimen of diluted urine on a culture plate, incubating it and counting the number of colonies. A more convenient approach to this same measurement involves the use of a microscope slide which is coated with nutrient agar. Such a slide, when dipped in a urine specimen and then incubated, will indicate the presence or absence of bacteriuria and also the approximate count.

Methods to determine the presence of significant numbers of bacteria in urine samples are available on various automated systems.[20] The Bac-T-Screen (*Marion*) system is a dispensing and filtering system used with a straining process to detect the presence of bacteria on special filter cards by noting the color change on the card. Analysis on the Abbott MS-2 performed by photometric monitoring of bacterial growth changes the light transmitted in a broth culture over a period of time. A decrease in the light transmission due to turbidity or color identifies a positive specimen. The Lumac Biocounter M2010 measures bacterial adenosine triphosphate (ATP) in urine by the bioluminescence produced in a luciferin-luciferase system. Once these rapid techniques are performed to determine which specimens have increased bacteria, further identification and sensitivity testing are performed. Chemical tests for the metabolic activity of bacteria have been used in studying bacteriuria. The most popular chemical test is that for nitrite. Ordinarily, all urine specimens contain nitrate, but do not contain nitrite. If *E coli*, or certain other organisms, are present in sufficient numbers, they will reduce the nitrate to nitrite.

Calculi—Knowledge of the composition of renal and bladder calculi ("stones") is essential in planning the therapeutic regimen for such diseases. Mixed calcium phosphate and oxalate stones usually occur over the entire urine pH range. Uric acid, cystine and calcium hydrogen phosphate calculi generally are associated with acid urines, while magnesium ammonium phosphate calculi usually occur in alkaline urine. Hyperexcretion of one of the calculi components, pH, renal blockage and the presence of foreign objects in the urinary tract are the most probable causal factors in the formation of renal calculi. Calcium oxalate stones are the most common type. The chemical content of the stones is established by routine qualitative analysis for calcium, magnesium, ammonium, phosphate, carbonate, oxalate, uric acid and cystine. Subsequent confirmation by optical crystallography, X-ray diffraction and infrared spectroscopy is also used in the characterization of the physical properties of the calculi.

Feces

Normal feces consists of undigested food remnants, products of digestion, bacteria and secretions of the gastrointestinal tract. *Macroscopic, chemical* and *microscopic* determinations are performed routinely. The normal quantity of feces is about 200 g/day. The brown color is due to the reduction of bilirubin to urobilinogen and then to uribilin (stercobilin); bilirubin is not normally present in feces, but porphyrins and biliverdin (a component of meconium) are excreted during the first days of life. Bilirubin can be detected by tests previously described for bile pigments.

Color changes in the stool can be the result of dietary intake or diagnostic for biliary obstruction and gastrointestinal bleeding.[21] Patients with steatorrhea and malabsorption may show a yellow bulky stool containing fat and gas. The feces is clay colored when bile is prevented from entering the gut. A red or black stool can occur when excessive doses of anticoagulants, phenylbutazone or salicylates are taken, producing bleeding in the gastrointestinal tract. Substances which interfere with the coloration of the stool include antacids (whitish or speckling), bismuth salts (black), iron salts (black), pyridium (orange), senna (yellow to brown) and tetracyclines (red).

Fecal urobilinogen can be determined colorimetrically by reduction of urobilin to urobilinogen with alkaline ferrous sulfate, and then reaction with acidified *p*-dimethylaminobenzaldehyde (Ehrlich's reagent). It is increased from a normal range of 40 to 280 mg a day to 400 to 1400 mg in hemolytic jaundice (dark brown stool), and is decreased in obstructive jaundice (clay-colored stool).

Porphyrins and porphyrinogens do not arise from hemoglobin catabolism, such as bilirubin, but are by-products of the synthesis of heme. Increases in fecal and urinary elimination of coproporphyrin, uroporphyrin and protoporphyrin are valuable diagnostic aids in distinguishing the various hepatic and erythropoietic porphyrias. Fecal coproporphyrins (CP) and coproporphyrinogens (CPP) are determined after extraction, conversion of CPP to CP by iodine and triple-point spectrophotometric estimation at 380, 401 and 430 nm to correct for interfering substances (also see *Urinalysis*).

Fecal occult blood is detected readily by the *o*-tolidine, benzidine, guaiac or diphenylamine tests; this is valid only if the patient has been on a meat-free diet for 3 days. Guaiac and diphenylamine are preferred due to the carcinogenic potential of the other two chemicals.

The Hemoccult test kit (*SmithKline Diagnostics*) uses an impregnated guaiac paper slide for detecting occult blood, which is a useful screening test for colon cancer. Two slides are prepared each day for 3 days from different parts of the same stool while the patient is on a meat-free high-bulk diet. Interfering substances include aspirin, indomethacin and corticosteroids, because they can produce bleeding, and Vitamin C, which interferes with the oxidation reaction of the test. If bleeding occurs high in the GI tract, the blood is digested and converted to acid hematin; 50 mL of blood in the feces will cause melena (black stool). Bleeding from the lower GI tract is apparent from red streaking of stools. The use of ^{51}Cr-tagged erythrocytes has been used to quantitate and locate the source of gastrointestinal bleeding. The subject's red cells are mixed with an isotonic ^{51}Cr solution and then reinjected intravenously. If bleeding occurs, the ^{51}Cr-

isotope content of the feces will be increased. Location of the hemorrhagic area also can be approximated by an isotopic scan of the abdominal area.

The presence of excessive quantities of *mucus* is usually indicative of dysentery, colitis or other inflammatory processes in the intestinal mucosa. Strongly alkaline or acidic reaction in the feces is indicative of excessive quantities of protein or carbohydrate in the diet, respectively.

Quantitative determination of *fecal nitrogen* is useful in analysis of pancreatic function. In pancreatic disease, increases in fecal nitrogen will occur as a result of decreased secretion of pancreatic proteolytic enzymes. The normal individual will excrete 4 to 13% of ingested nitrogen in the feces; in chronic pancreatitis, 9 to 30%. Fecal nitrogen can be determined by the Kjeldahl digestion procedure.

Fecal fat is present in the form of triglycerides of fatty acids (neutral fat), free fatty acids (FFA) and soaps. Fat determinations are based on the solubility of neutral fat and FFA in ether; the soaps are insoluble in ether and have to be acid-hydrolyzed to their respective FFA prior to extraction. Neutral fat will liberate FFA only on alkaline hydrolysis. The FFA, isolated from the above fractionations, are then determined by titrimetric, colorimetric or gas-chromatographic procedures.

Determinations of blood, urine and fecal ^{125}I after oral administration of an iodinated glyceryl trioleate or ^{125}I-oleic acid preparation is an index of *pancreatic, biliary and intestinal absorptive function* and correlates with *fecal fat excretion*. The bile must emulsify the ^{125}I-triglyceride prior to enzymatic hydrolysis by pancreatic lipase to yield FFA-^{125}I, which subsequently is absorbed and metabolized. An increased amount of ^{125}I in the feces is associated with pancreatic diseases (cystic fibrosis with achylia), obstructive jaundice, malabsorption disease (sprue, celiac disease) and steatorrhea. The latter entity can be differentiated as to a pancreatic lipase or intestinal absorptive defect. In the "absorptive" disease, increased excretion of ^{125}I is seen after administration of ^{125}I-triolein or oleic acid. In the pancreatic defect, adequate absorption of ^{125}I oleic acid occurs but fecal ^{125}I is increased after the triolein meal.

A *microscopic examination* of emulsified feces includes analysis for the presence of crystals, food residues, body cells, bacteria and parasites. Crystals of triple phosphate, calcium oxalate, fat and cholesterol, starch granules, vegetable fibers and neutral fat globules are normally present. Octahedral needle-shaped crystals (Charcot-Leyden crystals) are present in parasitic infestation and mucous colitis. Excessive quantities of fat or starch are seen in malabsorption disease.

Adult, larval or ova phases of parasites may be encountered in the feces. The most common parasitic infestations are caused by *cestodes* (tapeworms), *trematodes* (flukes), *nematodes* (roundworms) and *protozoa* (amoeba) (see *Microbiology*).

Toxicology

The determination of drug or chemical concentrations in biological fluids is an important aspect in diagnosing and treating the toxic syndrome induced by various agents in acute or chronic drug-abuse situations or in chemical poisoning.

Barbiturates, glutethimide, methaqualone, chlordiazepoxide, diazepam, diphenhydramine, ethchlorvynol, morphine, phenothiazines and salicylates are encountered in drug-abuse situations. Preliminary screening of serum or urine samples for drug substances is accomplished by TLC procedures. The analysis of serum or urine levels of intact drug or its metabolites usually is performed by extraction of the sample with an organic solvent, separation by gas-liquid (GLC), or high-performance liquid (HPLC) chromatography, and quantitation by spectrophotometric, fluorometric or electrochemical techniques. The interpretation of the serum-concentration data in relation to clinical significance and toxicology must not be limited to numbers.

In acute drug overdosage the time of drug ingestion, time of blood or urine sampling and severity of clinical symptoms or time of death must be interpreted in reference to data on the absorption, tissue distribution, metabolism and elimination of the drug and its metabolites. The specificity of the chemical assay as to interference from other drugs or metabolites of the parent drug must be considered. The combined techniques of GLC or HPLC and mass spectrometry confirms the identity of specific drugs in biological matrices. The extent of absorption of many drug substances is not related directly to the dose when large amounts of a drug are ingested, in comparison to the therapeutic dose.

The tissue-distribution and metabolic rates can be affected by large drug overdoses in which renal or hepatic failure is encountered. The plasma-elimination rate also can be affected, and it is important to recognize the change in elimination kinetics and to be aware of the nature of plasma elimination as defined by a mono-, bi- or polyexponential elimination curve. The drug overdose usually involves several drug substances and the chemical, metabolic and pharmacological aspects of drug interaction must be considered.

The methodology for the analysis of drugs in biological fluids or tissues can be found in the books listed in the *Bibliography*. Analysis for serum *barbiturate* levels will be described in this section as a specific example of the analytical methodology.

Serum is extracted at pH 6.5 with chloroform; the chloroform extract is washed with pH 7.0 phosphate buffer and extracted with 0.45*N* NaOH. The UV spectrum of the alkaline aqueous layer is determined at pH 13 and 10.5. The UV spectra are characteristic and distinguish barbiturates, *N*-methylbarbituric acids and thiobarbiturates. The barbiturates also can be detected by acidifying the alkaline layer, extracting with chloroform and spotting this organic extract on a silica-gel TLC plate. Sequential spraying of the plate with $KMnO_4$, $HgSO_4$ and diphenylcarbazone will show R_f values and color reactions typical of the various barbiturates. Blood barbiturates can be determined more accurately by a GLC procedure in which the retention times are used to identify the specific barbiturates. The degree of severity of clinical symptoms has been correlated with blood barbiturate levels. Comatose, areflexic signs are observed at 5.0 mg% amobarbital, 2.0 mg% pentobarbital, 8.0 mg% phenobarbital and 1.5 mg% secobarbital.

Opiates, amphetamines, barbiturates and methadone can be detected rapidly by *"homogenous" enzyme assay*.[22] In this procedure, the addition of drug antibodies to a conjugate of drug and lysozyme results in the inhibition of lysozyme activity. The addition of free drug to this reaction mixture increases the enzyme activity in proportion to the amount of free drug added. The sensitivity of this type of assay is 0.1 μg/mL of amphetamine and barbiturates, 0.5 μg/mL of methadone, 0.3 μg/mL of opiates and 1.0 μg/mL of benzoylecgonine, a cocaine metabolite. This assay is applicable to large drug-screening programs.

Electron-spin-labeling techniques also can be employed on large-scale drug-screening programs. In this procedure known amounts of drug antibodies are mixed with drug labeled with a stable nitroxide radical (spin-label) and with the specimen to be analyzed. Due to the competition for antibody between spin-labeled drug and drug in the specimen, the spin-labeled drug becomes detached from the antibody and can be detected by electron-spin resonance spectroscopy. This procedure is 1000 times more sensitive than TLC.

Blood-alcohol levels may be determined by aeration, distillation, gas chromatography or specific enzymatic analysis with alcohol dehydrogenase. In the chemical techniques the blood sample is either oxidized or distilled into a dichromate-sulfuric acid mixture; the excess dichromate is then determined by titration with potassium iodide or methyl orange-ferrous sulfate solutions or by colorimetric analysis. The gas-chromatographic and enzyme procedures are specific for ethanol, whereas the chemical techniques are influenced by other volatile or oxidizable substances in the blood. The enzymatic method is based on the reaction of ethanol and NAD in the presence of alcohol dehydrogenase to form acetaldehyde and NADH; the acetaldehyde is removed with semicarbazide and the NADH formed in the reaction is estimated spectrophotometrically at 340 nm. Ethanol levels of >0.10% are indicative of intoxication and apparent psychomotor disturbance. Levels of 0.40 to 0.50% are associated with medullary and diencephalic disturbances such as tremors, coma, respiratory depression, peripheral collapse and death.

Specific analysis of heavy metals is best performed by atomic-absorption spectroscopy. Analyses for arsenic, beryllium, bismuth, copper, iron, lead, lithium, mercury, nickel, thallium and zinc are encountered frequently in the toxicology laboratory. *Blood lead* is determined by forming a lead-dithiocarbamate chelate in the presence of ammonium pyrrolidinedithiocarbamate and extracting the chelate into methyl isobutyl ketone for subsequent atomic-absorption analysis. A lead concentration of >60 μg/mL in children usually reflects significant absorption and accumulation of lead and is interpreted as an indicator of lead toxicity (plumbism).

Increased lead exposure will result in a decrease in delta-*aminolevulinic acid* (*ALA*) conversion to porphobilinogen by ALA-dehydrase in heme synthesis. ALA blood levels will increase to the point that ALA is excreted in the urine. Determination of urinary ALA is performed by removing urine porphobilinogen and urea by ion-exchange chromatography, reacting ALA with *p*-dimethylaminobenzaldehyde and determining the chromogen colorimetrically. Urinary ALA levels >2.5 mg/100 mL are unacceptable in children and industrial lead workers. Urinary ALA levels are not as sensitive an indicator of lead toxicity as blood lead, but they can be used to monitor prophylactic treatment procedures.

Cholinesterase determinations are of value in the diagnosis of suspected cases of organophosphate or carbamate pesticide poisoning. Two types of cholinesterase are found in tissues. True cholinesterase is found in RBC and nerve tissue and exhibits a specificity for acetylcholine substrate. Pseudocholinesterase is found in plasma and has a greater affinity for hydrolyzing butyrylcholine and other esters. The organophosphate and carbamate insecticides inhibit both enzymes. The activity of the plasma enzyme is inhibited more rapidly than the RBC cholinesterase, and recovers more rapidly due to synthesis of new enzyme by the liver. The recovery of the erythrocyte enzyme is slow and is governed by red-cell turnover rate. Cholinesterase activity usually is determined by measuring changes in pH after the incubation of plasma or RBC lysates with acetylcholine. The normal range of this enzyme is 4.5 to 10.9 (plasma), 3.4 to 5.7 (whole blood) and 6 to 10.5 (RBC) units/mL.

Gastric Analysis

The chief constituents of gastric juice are hydrochloric acid, gastric proteases (pepsin and gastricsin), hematopoietic factor (intrinsic factor and vitamin B_{12} binders), gastric hormones and mucosubstances (aminopolysaccharides, mucopolyuronides, mucoids and mucoproteins). Tests for *gastric function*[23] usually are performed on gastric juice samples collected by direct intubation into the stomach. The fasting content (normal, <100 mL) of the stomach is removed and gastric secretion is collected in the basal state, or after stimulation by the oral administration of caffeine-benzoate or alcohol, or parenteral administration of histamine, insulin or the hormone pentagastrin. Samples are collected by continuous aspiration and analyzed for acidity and gastric protease activity at various time intervals. The extent of recovery of total juice can be estimated by oral, nonabsorbable indicators (polyethylene glycol-^{14}C, phenol red and ^{125}I-HSA) instilled into the stomach prior to the aspiration. The recovery and specific concentration of these indicators in gastric juice is an index of gastric secretory volume, completeness of collection and gastric emptying rate.

Gastric juice is a heterogeneous mixture of clear juice and flocculent, clear mucus. The *color* of the juice should be noted as to the appearance of blood, bile and excessive quantities of mucus. The *acidity* can be determined by a simple pH measurement and conversion to mEq H^+ or by titration of centrifuged gastric juice to pH 3.5, 4.5 and 7.4, the respective end-points for free acid (HCl), protease activity and physiological neutrality. The *basal acid output* is about 1 mEq/hr in normal subjects and 2 to 4 mEq/hr in duodenal ulcer patients. The *peak acid output* (*PAO*) after histamine stimulation is 10 to 20 mEq/hr in normals and 40 to 50 mEq/hr in duodenal ulcer; PAO following pentagastric stimulation is similar to histamine. Gastric acid secretion is decreased in atrophic gastritis, gastric carcinoma and certain types of gastric ulcer. Hypersecretion is seen in duodenal ulcer, Zollinger-Ellison (ZE) syndrome and hyperparathyroidism.

In situ measurements of pH may be made with a *Heidelberg capsule apparatus*. In this technique the subject swallows a small pH-sensitive capsule (transmitter); radiowaves are transmitted from the capsule to a sensing device (receiver), and the signals are recorded as a function of pH. The normal pH of the stomach is 1.2 to 1.8.

Tubeless gastric acidity analysis is performed by oral administration of Diagnex Blue (*Squibb*), a carbacrylic ion-exchange resin reacted with azure blue dye. The hydrogen ions in the gastric juice exchange with the dye on the resin; the dye is absorbed and then excreted in the urine. The dye concentration in the urine is a function of gastric acidity. The normal value is >0.6 mg of dye in the urine 2 hours after administration.

The principal gastric proteases are *pepsin and gastricsin*; pepsinogen is a precursor which is converted to active pepsin by free HCl and by an autocatalytic process. *Total gastric protease activity* is determined on hemoglobin or radioiodinated human serum albumin (RISA) substrates at pH 1.8 to 3.1 (RISA-^{125}I); protease activity on hemoglobin will liberate tyrosine which can be estimated spectrophotometrically at 280 nm; with RISA, liberated tyrosine-^{125}I, as estimated by isotopic procedures, is an index of proteolytic activity.

Pepsin activity can be distinguished from the total protease activity by estimation of the 3,5-diiodotyrosine liberated from *N*-acetyl-L-phenylalanyl-3,5-diiodotyrosine substrate at pH 2.1. Pepsin will react on this substrate; gastricsin will not. Normal gastric juice protease activity ranges from 200 to 1200 μg total protease activity/mL and 50 to 300 μg pepsin/mL. The presence of bile, blood, saliva or excess mucus in the sample will decrease both acidity and gastric protease activity.

Gastrin, cholecystokinin, secretin and pancreozymin are gastrointestinal hormones.[24] The role of gastrin and its interaction with other gastrointestinal hormones in the etiology and proliferation of ulcer disease is of recent interest. Accurate RIA techniques have been developed for gastrin and secretin-6-tyrosine due to the availability of a pure synthetic polypeptide. Biological assays based on the effect of

these substances on gastric, pancreatic and biliary secretion also have been used.

Gastrin is found in various species in two forms, G-I and G-II. The only difference is in sulfation of the 12-tyrosyl residue in G-II of the heptadecapeptide amides. Gastrin is found primarily in the gastrin-producing cells (G-cells) of the antral mucosa. The C-terminal tetrapeptide represents the biologically active part of the molecule. Gastrin infusion will stimulate secretion of gastric acid, pepsin and intrinsic factor. It has a slight secretin-like effect and a powerful pancreozymin-like effect on pancreatic secretion. Gastrin also stimulates bile flow. The instillation of HCl into the stomach will inhibit gastrin release; protein and meal stimulation will increase serum gastrin.

The RIA of serum gastrin is of diagnostic value in the ZE syndrome, pernicious anemia and duodenal ulcer. Basal serum gastrin levels in the normal individual are 20 to 30 μg/mL and increase about 2-fold after a protein meal stimulus. Basal serum gastrin levels in duodenal ulcer are normal or slightly elevated, but increase 4- to 5-fold after a protein-meal stimulus. Basal serum gastrin levels are elevated in ZE to 500 to 4000 pg/mL due to the presence of a gastrin-producing tumor. The ZE patient is uniquely sensitive to IV calcium stimulation which will increase both gastric acid secretion and serum gastrin in this syndrome. Basal serum gastrin levels also are elevated in gastric hyposecretion as seen in pernicious anemia and Type A gastritis, and in chronic renal failure due to the decreased metabolic turnover of gastrin in the kidney.

The RIA of serum gastrin is based on the competition of gastrin in test sample with ^{125}I-gastrin for gastrin antibody binding sites. The antibodies used in this procedure are usually cospecific for G-I and G-II. However, they detect all forms of circulating gastrin, ie, Big-Big Gastrin (G-39), Big Gastrin (mol wt 7000; G-33), gastrin heptadecapeptide (G-17, mol wt 2200), G-13 and G-8 (mini-gastrin). The Big components can be converted to gastrin by trypsin hydrolysis. The significance of changes in the ratio of the circulating gastrins is not known, but it has been suggested that G-39 and G-33 predominate in the basal state and cleave to G-17, which is the major serum form after a protein meal.

Other Body Fluids

Physical, chemical and microscopic examination of cerebrospinal fluid, synovial fluid, human milk, transudates and exudates also are performed by the clinical laboratory. The principles of the various determinations are similar to those described for blood and urine.

Microbiology

Clinical medical microbiology is a science which is concerned with the isolation and identification of disease-producing microorganisms, ie, bacteria, fungi (including yeast), viruses, rickettsia and parasites. The techniques employed in the isolation and identification of the suspect organism(s) involve the propagation on suitable primary culture media, selective isolation on special culture media, use of suitable living host material (mouse, embryonated egg, tissue culture, etc), determination of morphological and, where applicable, staining characteristics of the organism and confirmation by biochemical and/or immunochemical analysis. Suitable animal inoculation, where applicable, may be employed to determine pathogenicity. Site, timing, technique (aseptic), instrumentation, and transportation of clinical specimens (blood, urine, feces, cerebrospinal fluid, etc) are prime variables involved in the final differentiation and confirmation process.

Rapid manual enzymatic and immunological test kits have been introduced to identify pathogens for cerebrospinal fluid analysis. The latex-agglutination test coats a specific antibody onto latex particles and when an antigen is present, the latex particles are visible.[26] In the coagglutination test, the specific antibody is bound to protein A on the surface of a staphylococcal cell and the presence of antigen produces agglutination.[25]

Staphylococcus aureus (*Micrococcus pyogenes* var *aureus*) is a Gram-positive coccus frequently found on normal human skin and mucous membranes and frequently associated with abscesses, septicemia, endocarditis and osteomyelitis. Some strains elaborate an exotoxin capable of causing food poisoning. The primary isolation is on blood agar and in thioglycollate broth. With feces and other heavily contaminated specimens, phenylethyl alcohol agar and/or mannitol-salt agar should be inoculated to suppress growth of other bacteria. The identification of pathogenic staphylococci is based on colonial (pigmentation) and microscopic morphology (grape-like clusters), positive catalase production, positive coagulase production (staphylocoagulase-plasma clotting factor) and positive mannitol fermentation.

Streptococcus pyogenes is another Gram-positive coccus frequently associated with tonsillitis or pharyngitis, erysipelas, pyoderma and endocarditis. Neopeptone agar containing 5% defibrinated sheep blood is preferred for primary isolation and to demonstrate characteristic hemolysin production by observing a zone of clear (beta) hemolysis around the colonies on blood agar. Streptococcal groups are identified by precipitin tests with group-specific antisera for A, B, C, D, F and G. Streptex (*Welleome Diagnostics*) uses a latex agglutination system for identifying the Lancefield group of streptococci. Other groups usually are not associated with human clinical materials.

Legionella pneumophilia identification includes specimen cultures on lung tissue or sterile body fluids (eg, pleural fluid or pericardial fluid). Direct fluorescent antibody method is a test for *L pneumophilia*. Organisms are best seen in the acute stage of the disease. Since the antiserum is species-specific, polyvalent antisera are necessary for identification.

Neisseria gonorrheae is a Gram-negative diplococcus associated with the venereal disease gonorrhea. The identification is based on the primary isolation of the gonococcus from urethral exudates on chocolate agar or Thayer-Martin (TM) medium. The microscopic observation of Gram-negative intracellular diplococci resembling the gonococcus constitutes a presumptively positive diagnosis of gonorrhea. Confirmation of the oxidase enzyme activity of the gonococci is performed by a reaction with *p*-dimethylaminoaniline which turns oxidase-positive colonies black. A positive oxidase test by Gram-negative diplococci isolated on TM medium constitutes a presumptively positive test for *N gonorrheae*. Final identification rests on typical sugar fermentation or specific (fluorescent antibody) staining.

Neisseria meningitidis is the primary cause of bacterial meningitis and septicemia. The primary isolation is based on culturing of a specimen (blood, spinal fluid or nasopharyngeal secretions) on a Mueller-Hinton medium or chocolate agar containing a vancomycin-colistimethate-nystatin antibiotic mixture. The confirmation of the isolate by biochemical reactions (positive oxidase, positive catalase, etc) and serological agglutination with group-specific (A, B and C) antiserum is used in the differentiation. Young cultures of groups A and C may show capsular swelling (Quellung reaction) in the presence of a specific antiserum.

The enteric bacilli (*Enterobacteriaceae*) are Gram-negative, nonsporulating rods associated with dysentery (*Shigella sp*), typhoid fever (*Salmonella typhi*), urinary tract and tissue infections (*Escherichia coli*, *Proteus sp* and *Pseudomonas sp*), and pulmonary infections (*Klebsiella sp*). The primary isolation of enteric bacilli is on selective and differential infusion agar such as MacConkey and eosin-methyl-

ene blue (EMB), and enrichment media such as selenite broth and tetrathionate broth. The primary isolation of *Salmonella sp.* is on Leifson's deoxycholate citrate agar (LDC) or *Salmonella-shigella* agar (SS); if *Salmonella typhi* is suspected, brilliant green agar (BG) and bismuth sulfite agar (BS) may be used and would constitute a presumptively positive diagnosis of *S typhi*.

The confirmation and identification of enteric bacilli may be performed by serological tests and biochemical reactions: H_2S production (triple-sugar iron agar), indole production, acetylmethylcarbinol production, citrate utilization, urease, lysine and arginine decarboxylase and phenylalanine deaminase activity. Enterotube (*Roche Diagnostics*) employs conventional media to perform 11 standard biochemical tests which can be inoculated simultaneously in one compartmented tube, with a single bacterial colony. The serological identification of *Salmonella* and *Shigella sp* is based on the agglutination of antigens that fall into three categories: "K" capsular (*Klebsiella sp* and *Shigella sp*), "O" (*Salmonella sp, Arizona sp, E coli, Shigella sp*, etc) and "H" flagellar (*Salmonella sp*).

Other Gram-negative rods of medical importance are the hemophilic bacilli (*Bordetella pertussis*, whooping cough and *Hemophilus influenzae*, bacterial meningitis), the hemorrhagic bacilli (*Pasteurella pestis*, bubonic plague, and *P tularensis*, tularemia) and pyrogenic bacillus (*Brucella melitensis*, undulant fever).

Spore-forming Gram-positive rods of medical importance belong to the genus *Clostridium*, which are associated with tetanus (*Cl tetani*), gas gangrene (*Cl perfringens* or *welchii*) and botulism (*Cl botulinum*). The isolation of these organisms requires anaerobic conditions. Once the strain to be identified is obtained in pure culture by single-colony selection, its morphological characteristics are noted; the strain then is grown in a variety of definitive media to determine catalase activity, hydrogen peroxide decomposition and fermentation or hydrolysis of carbohydrates and organic acids. The analysis of fermentation products (gas chromatography) also is used for the identification of pathogenic anaerobic *Clostridia*. The major clostridial exotoxin type can be determined by typing with specific antitoxin sera. A Gram-positive, aerobic, spore-former of medical importance is *Bacillus anthracis*, responsible for anthrax, a disease of animals transmissible to man.

The mycobacteria are acid-fast bacilli associated with tuberculosis in man (*Mycobacterium tuberculosis*), in cattle (*Mycobacterium bovis*) and leprosy (*Mycobacterium leprae*). Tubercle bacilli in man are isolated from sputum cultured on a tubed or bottled egg medium (Lowenstein-Jensen) following enzymatic digestion and concentration of the specimens. A provisional diagnosis of tuberculosis usually is made by demonstrating acid-fast bacilli microscopically, X-ray diagnosis and a positive tuberculin skin test.

Other weakly and partially acid-fast bacilli of medical importance are members of the *Actinomycetales, Nocardia asteroides* and *Nocardia brasiliensis*, which are responsible for severe pulmonary infections and cutaneous and subcutaneous abscesses.

Bacteriophages (phages) are a special group of viruses that are hosted by bacteria. Any given phage is highly host-specific and when in contact, lysis of the host occurs (phage-typing). They are used primarily as epidemiological tools in subtyping strains of *E coli*, staphylococci or *Salmonella sp* that are presumed to be related epidemiologically. Phages also furnish ideal material for studying host-parasite relationships and virus multiplication.

The medically important fungal diseases include the superficial mycoses, ie, fungal invasion is restricted to the outermost layers of the skin or to the hair shafts (*Microsporum audouini*, ringworm of the scalp, *Trichophyton sp*, athlete's foot and *Epidermophyton floccosum*, *Tinea pedis*) and the systemic pathogenic fungi (*Blastomyces dermatitidis, Coccidioides immitis, Histoplasma capsulatum, Candida albicans*). The diagnosis of the causative agent is based on the isolation of organisms on Sabouraud's dextrose agar or trypticase soy agar with or without cycloheximide and chloramphenicol to suppress the growth of saprophytic fungi and bacteria, macroscopic examination of morphological characteristics and microscopic examination using KOH or lactophenol cotton-blue stain. Biochemical reactions usually are limited to *Candida sp*. Immunological reactions include skin tests, where applicable, agglutination tests, such as latex particle agglutination for histoplasmosis and tube precipitin and complement-fixation tests.

An *antimicrobial susceptibility test* is a determination of the least amount of an antimicrobial chemotherapeutic agent that will inhibit the growth of a microorganism *in vitro*, using a tube-dilution method, agar cup or disk-diffusion method. The test may function as an aid in the selection of a chemotherapeutic agent by the physician. Also, the concentration of antimicrobial agents in body fluids may be determined by biological assay with an organism of known susceptibility for the specific agent.

The laboratory diagnosis of *viral infections* is based upon (1) examination of the infected tissues for pathognomonic changes or for the presence of viral material; (2) isolation and identification of the viral agent; (3) demonstration of a significant increase in antibody titer to a given virus during the course of the illness; (4) detection of viral antigens in lesions—using fluorescein-labeled antibodies and (5) electron microscopic examination of vesicular fluids or tissue extracts. Blood is used for serological tests but seldom for virus isolation. Acute and convalescent-phase blood specimens must be examined in parallel to determine whether or not antibodies have appeared or increased in titer during the course of the disease. Some examples of human viral infections are respiratory infections (Adenovirus group); diseases of the nervous system, ie, polio and coxsackie viruses of the picornavirus group; smallpox (poxvirus group); measles (paramyxovirus group); chicken pox (herpesvirus group) and influenza (myxovirus group).

Members of *Mycoplasmataceae* pleuropneumonia-like organisms (PPLO) are of a range of size similar to the larger viruses. They are highly pleomorphic because they lack a rigid cell wall, they can reproduce in cell-free media and they do not revert to or from bacterial parental forms as the L-forms. Specimens (sputum, bronchial secretions, urinary sediment, etc) for the primary isolation of mycoplasmas (*M pneumoniae, M hominis*, etc) should be cultured on agar media containing peptone, serum, ascitic fluid, whole blood or egg yolk. The species identification may be by growth inhibition on agar medium containing type-specific rabbit antisera. Antigenic variants or subspecies may be detected by immunodiffusion. Various PPLO are pathogenic, parasitic or saprophytic. Mycoplasmas have a predilection for mucous membranes and are associated with primary atypical pneumonia and bronchitis.

Clinical parasitology is a science which is concerned with the parasitic protozoa (amoeba), the helminths (cestodes, tapeworms; trematodes, flukes; nematodes, roundworms) and the arthropods. The identification of protozoan ova is based on detailed microscopic morphology (nuclei, etc) using wet mounts (saline or iodine) or stained preparations (iron hematoxylin, etc) obtained from fecal specimens (fresh or preserved with polyvinyl alcohol), which are concentrated by sedimentation, centrifugation or flotation techniques. Trophozoite and/or cystic stages may be detected in fecal specimens associated with intestinal protozoa as in amoebic dysentery caused by *Entamoeba histolytica*.

The commonly encountered helminths are *Necator americanus* (hookworm), *Trichuris trichiura* (whipworm) and *Enterobius vermicularis* (pinworm); they are identified by

characteristic ova. Characterization of tapeworm segments (proglottids) or head (scolex) in a fecal specimen will differentiate *Taenia saginata* (beef tapeworm) from *Taenia solium* (pork tapeworm). Eggs of *T solium* and *T saginata* cannot be differentiated on a morphological basis.

Adult flukes oviposit a characteristic egg which may reach the urine, sputum or feces. *Schistosoma japonicum* eggs have a small, indistinct spine; *S mansoni*, a distinct, large, lateral spine; and *S haematobium*, a distinct terminal spine.

Arthropoda constitute the largest of the animal phyla which are characterized by a segmented body with the segments usually grouped in two or three distinct body regions, by a chitinous exoskeleton, several pairs of jointed appendages and characteristic internal organs. Most arthropods can be preserved in 70% alcohol. They are of medical importance since they can infest man and cause mechanical trauma or produce hypersensitivity from repeated exposure (*Cimex lectularius*, the bedbug) or by toxin injection (*Latrodectus mactans*, the black widow spider), by skin invasion (*Sarcoptes scabiei*, the itch mite) and by transmitting disease (*Anopheles* mosquitoes, malaria), and *Yersinia pestis* in fleas (plague).

The serodiagnosis of parasitic diseases includes the following immunodiagnostic tests: complement-fixation (trichinosis), precipitin test (schistosomiasis), bentonite flocculation (ascariasis), hemagglutination (echinococcosis), latex agglutination (trichinosis), cholesterol flocculation (schistosomiasis), fluorescent antibody (malaria) and methylene blue dye test (toxoplasmosis).

Immunochemistry

Clinical immunopathology[26] includes *general immunology* (immunofluorescence, immunodiffusion, immunoelectrophoresis and agglutination tests), *radioimmunoassay* (RIA-hormones, vitamins, drugs, immunoglobulins), *tissue typing* (histocompatibility tests in organ transplants), *cellular immunology*, *cancer immunology* and *immunohematology*. Examples of each of these disciplines are discussed in this section and other parts of this chapter.

The ELISA, *enzyme-linked immunosorbent assay*, detects antibodies by an indirect technique using enzyme-linked antibodies to label antigenic substances in tissue or body fluid. The antigen is attached to a solid matrix and reacts with a specimen that may contain a complimentary antibody. The antihuman globulin, which is conjugated with the enzyme, is added and the antigen reacts with the bound antibody of the patient. By adding the substrate molecule the enzyme is detected. This system has been used to identify antibodies to viruses, parasites, bacterial products and quantitation of some drugs.

Antibody response is a complex process involving the lymphoid cell system response to foreign stimulus or antigen. Hematopoietic cells in the fetal yolk sac, liver or marrow develop into lymphoid stem cells which, in turn, differentiate into T-lymphocytes of thymic origin and B-lymphocytes of bone-marrow origin. The T-cells further differentiate into lymphoblasts which are responsible for *cell-mediated cellular immunity* (graft vs host reaction, tissue transplant rejection, tuberculin skin testing, *delayed-type hypersensitivity*). B-cells differentiate into plasma cells which are responsible for humoral immunity which is mediated by circulating serum immunoglobulins (*immediate-type hypersensitivity*).

Macrophages can cooperate in presentation of antigen to the T- or B-lymphoblasts. Cooperation between T- and B-cells, immunological memory, development of immune tolerance to antigens and genetic control of the immune response are integral properties of the immune system and are related to development of immune deficiency and autoimmune disease.

The identification and determination of *immunoglobulins* (IgG, IgM, IgA) by radial immunodiffusion and immunoelectrophoresis have been discussed under *Proteins*. *IgM* (γM) is the earliest antibody found in the primary immune response and falls rapidly after the onset of IgG antibody synthesis. *IgG* (γG) is the major class of antibody in both the primary and secondary immune response. IgG can cross the placenta to provide the early forms of antibody protection for the newborn. IgG and IgM can participate in the complement fixation reaction. *IgA* (γA) is found predominantly in saliva and secretions of the gastrointestinal and respiratory tracts. In contrast to IgM and IgG, only a small portion of total IgA is found in blood. IgA functions in protection against pathogens that enter the host through the respiratory or gastrointestinal tract. *IgD* (γD) is found in trace quantities in sera and its function is unknown. *IgE* (γE) is probably the most important antibody in acute hypersensitivity or allergic reactions. Reaction of mast cell- or basophil-bound IgE with antigen initiates the release of histamine, slow-reacting substance (SRS), serotonin and bradykinin and the subsequent allergic response. IgE is best quantitated by RIA. Mean serum levels (mg%) in healthy adults are IgG 1200 ± 500, IgA 210 ± 140, IgM 140 ± 70, IgD 3 and IgE <0.1.

Heterophile antibodies are agglutinins which are capable of reacting with antigens that are entirely unrelated to those which stimulate their production. These antibodies, which occur in the serum of patients with infectious mononucleosis or serum sickness, will agglutinate formalized horse erythrocytes. In order to distinguish the specific *heterophile agglutinins of infectious mononucleosis*, the serum sample is mixed with guinea-pig kidney tissue or beef erythrocyte stromata; the infectious mononucleosis antibody will be absorbed and inactivated by the beef cells but not by the kidney tissue, and subsequent agglutination of horse erythrocytes will occur only in the kidney-tissue system. This test is used to detect infectious mononucleosis even prior to clinical symptoms. The heterophile titer has no relation to the course or severity of the disease.

Two protein constituents of human plasma, *rheumatoid factor* (RF) and *C-reactive protein* (CRP) are of value in the differential diagnosis of rheumatoid diseases. CRP is a protein present in the serum of patients in the acute stages of bacterial and viral infections, collagen diseases and other inflammatory processes. The presence of this antigen in serum is detected by agglutination of polystyrene latex particles sensitized with specific CRP antibody globulin. In the management of rheumatic fever, decreases in CRP blood levels are used to measure the effectiveness of therapy.

Rheumatoid arthritis is characterized by the presence of a reactive group of macroglobulins known as RF in blood and synovial fluid. RF is a protein of the IgM globulin fraction and is regarded as an autoantibody against antigenic determinants of IgG. Analysis of RF is based on agglutination procedures employing polystyrene latex particles coated with a layer of adsorbed human gamma globulin. The RF-antibody reaction causes a visible agglutination of the inert latex particles. CRP is not elevated in rheumatoid arthritis.

β-Hemolytic streptococci, the causative agent in rheumatic fever, produce streptolysin O and S, streptokinase, hyaluronidase, desoxyribonuclease and NADase in the body. The growth of streptococci in tissue with elaboration of these proteins serves as the antigenic stimulus to evoke the production of specific antibodies (eg, *antistreptolysin-O, ASO*). The quantitation of the antibody titer to these enzymes is an index of the strength of the antigenic stimulus and the extent of the streptococcal infection. These antibodies can be detected by latex agglutination (ASO) or tests

dependent on the inhibition of enzyme action by the antibody (anti-hyaluronidase inhibition of hyaluronic acid depolymerization by hyaluronidase).

The laboratory diagnosis of *syphilis* (treponemal disease) and the evaluation of a chemotherapeutic approach is based on serological tests. Demonstration of an antibody-like substance, *reagin*, or of true antitreponemal antibody in the serum of infected individuals is accomplished by complement fixation or flocculation tests for reagin, or immunofluorescent techniques for treponemal antibody.

In the *complement fixation* tests (Kolmer CF), reagin reacts with a complex phosphatidic acid antigen (cardiolipin) and complement; the complement is bound and will not lyse hemolysin-sensitized red cells which were added in the second phase of the test. In normal serum the reagin-cardiolipin complex is not formed and the complement is free to react with hemolysin and lyse the erythrocytes.

Flocculation tests for determining of syphilis use a cardiolipin-lecithin-cholesterol antigen which clumps in the presence of serum reagin occurring in nontreponemal diseases and syphilis (*Venereal Disease Research Laboratory—VDRL Test; rapid plasma reagin—RPR test*).

Treponemal antibody can be detected also by the reaction of the patient's serum with treponemal antigen and subsequent confirmation with fluorescein-labeled antihuman globulin as an indicator of primary antigen-antibody reaction (*fluorescent treponemal antibody-FTA test*). The patient's serum can be treated with an extract of treponemes prior to the FTA test to remove interfering antibodies and eliminate biological false-positives (FTA-Abs Test). False-positives occur in related treponematosis such as yaws, pinta and bejel. Increased reagin titers also occur in malaria, leprosy, infectious mononucleosis, chronic rheumatoid arthritis or systemic lupus erythematosus and in patients on hydralazine therapy.

Febrile antibodies are present in the serum of patients with certain bacterial or rickettsial infections (spotted, typhus or Q fever). In typhus the patient's serum contains a febrile antibody which will agglutinate a suspension of *Proteus OX-19* bacteria (Weil-Felix Reaction). *Salmonella* O-H, *Pasteurella tularensis* and *Brucella abortus* antigens are used in febrile antibody tests for diagnosis of typhoid or paratyphoid fever, tularemia and brucellosis, respectively.

Toxoplasmosis is a major cause of birth defects. An expectant mother may become infected with oocysts in uncooked meat, or from cat fur, and infect the fetus transplacentally. Toxoplasmosis testing is based on detecting serum antibody by a hemagglutination procedure. Red cells sensitized by exposure to toxoplasmosis antigen are agglutinated by the specific antibody.

Radioimmunoassay (RIA)[6,27] has been mentioned in various sections of this chapter as an analytical tool in the measurement of hormones, immunoglobulins, drugs and steroids. The basic principle of RIA is

$$Ag^* + Ag + Ab \rightleftarrows Ag^*Ab + AgAb + Ag^* + Ag \qquad (4)$$

RIA is not to be confused with the *specific reactor assay* using labeled antigen and nonantibody protein receptors which is used for vitamin B_{12}, T^4, T^3 and cortisol assays.

All procedures are based on the observation that radiolabeled antigens (Ag^*) compete with nonlabeled antigen (Ag) for binding sites on specific antibody (Ab) in the formation of antigen-antibody complexes (Ag^*Ab, AgAb). When increasing amounts of Ag are added to the assay, the binding sites of Ab are saturated progressively and the antibody can bind less Ag^*. Therefore, the ratio of bound to free Ag^* (B/F) or % Ag^* bound is a direct index of the concentration of Ag in the assay.

The requirements for RIA are (1) preparation and characterization of Ag (2) radiolabeling of Ag, (3) preparation of specific Ab and (4) development of the assay system and methods to separate free (Ag, Ag^*) from antibody bound (AgAb, Ag^*Ab) antigen.

Antigens can be prepared from natural tissue sources or preferably synthesized. ^3H, ^{14}C or ^{125}I-labeled antigens are used routinely in the assay. The biological and immunochemical activity of the antigen must not be altered in the tagging procedure, and the specific activity of Ag^* must be extremely high so that tracer quantities can be used in the assay. Tritium labeling and iodination (^{125}I) produce the highest specific activity, but also increase susceptibility of Ag^* to internal degradation and self-radiolysis, in contrast to ^{14}C. In many instances, the original antigen cannot be iodinated, but can be altered chemically in such a way as to retain full antigenic cross-reactivity in RIA; eg, cyclic AMP has no tyrosyl or histidyl residue for iodination; ^{125}I-succinylcyclic AMP-tyrosine methyl ester retains full cross-reactivity with antibodies to cyclic AMP and is used in the assay.

Hormones, steroids and drug substances are *haptens*. They do not produce the antibody response when injected by themselves, but will produce antibodies specific for the hapten when injected as a hapten-protein carrier conjugate. Gastrin (hapten) is coupled to albumin (protein-carrier) by treatment with carbodiimides (CCD), which couple functional carboxyl, amino, alcohol, phosphate or thiol groups. Morphine must be converted to the 3-*O*-carboxymethyl derivative prior to CCD coupling with albumin to provide a functional coupling group in the hapten. The hapten-conjugate usually is emulsified in a mineral oil preparation of killed *Mycobacterium* (Complete Freunds Adjuvant) and injected intradermally in rabbits or guinea pigs on several occasions. The serum antibody must have both high specificity and affinity for the antigens.

The *assay system* contains Ag^*, sample-containing endogenous Ag or a standard Ag and antibody, at specified pH (6.5 to 8.5). After incubation at 5 to 37° for anywhere from 1 hour to several days, free and antibody-bound antigen must be separated. This is accomplished by *double-antibody technique, solid-phase RIA, resin techniques* or *salt or solvent precipitation*. In the double-antibody technique, antiglobulin (Ab′) serum is added to the assay system after incubation. $Ab\text{-}Ag^*$ and Ab-Ag complexes are antibody-globulin antigen complexes. The antiglobulin will react to form insoluble $Ab'\text{-}Ab\text{-}Ag^*$ and Ab′-Ab-Ag complexes, which can be removed by centrifugation. The free Ag^*, Ag is in the supernate.

The solid phase RIA is performed by coating tubes with Ab; Ag and Ag^* react, compete and bind with Ab on the wall of tube. Unreacted Ag and Ag^* is separated by decanting and rinsing the tube. Ab also can be bound covalently with isothiocyanate to dextran gel particles. Ag and Ag^* will compete and bind with Ab on particles. Bound antigen then can be separated from free antigen by centrifugation.

RIA has been applied to analysis of hormones (ACTH, angiotensin I and II, gastrin, HCG, FSH, GH, glucagon, HLH, HPL, insulin, thyroxine), steroid hormones (aldosterone, androstenedione, glucocorticoids, testosterone, estrones, progesterone), drug substances (digoxin, digitoxin, amphetamines, barbiturates, morphine, LSD, ouabain), endogenous substances (cyclic AMP, cyclic GMP, prostaglandins, immunoglobulins, hepatitis antigen, carcinoembryonic antigen—CEA). Examples of the specific assays are discussed in other sections.

CEA and AFP (α-1-fetoprotein) are proteins found in fetal tissue. CEA analysis was first proposed as a specific test for the early detection of bowel cancer. Although the test does not have absolute specificity for this disease, it may prove of value as a diagnostic aid and therapy monitor. CEA can be detected by RIA. Serum levels >2.5 ng CEA/mL are found in 60 to 70% of patients with adenocarci-

noma of the colon; positive levels also are found in lower percentages in carcinomas of the pancreas, stomach, liver, breast, endometrium, ovary, kidney and bronchus, as well as in other conditions such as gastrointestinal polyps, colitis, diverticulitis and cirrhosis. CEA appears to be associated primarily with tumors of entodermally derived epithelial tissue. The similarity between CEA and cell-surface glycoproteins and sialic acids has stimulated considerable research interest in a new approach to cancer chemotherapy.

The study of *tissue-transplantation antigens* is an important factor in studies on tissue and organ transplants. ABO blood group antigens are involved in survival of skin and renal grafts. Because of the presence of natural occurring anti- A and B, avoidance of ABO incompatibility is important in clinical grafting. The *HL-A antigens* are found on tissue and on the white cells. There is one major histocompatibility locus, comprising a number of alleles or linked genes, on a single chromosome segment. Each allele controls four to five groups of major transplantation antigens. These HL-A isoantigens affect the survival of allogenic tissue grafts and organ transplants. HL-A antigens can be typed by a leukoagglutination method in which the patient's or donor's white cells are reacted with specific HL-A antisera. HL-A typing also can be performed by a cytotoxicity test in which lymphocytes are mixed with antisera and complement. The antibody can destroy the lymphocytes if a corresponding antigen is present on the cell surface.

References

1. Mitruka BM, Rawnsley AM: *Clinical and Hematological Reference Values in Normal Experimental Animals and Normal Humans*, Yearbook Medical Pub, Chicago, 1981.
2. Christensen RL, Triplett DA: *Lab Med, 13(11)*: 666-672, 1982.
3. Bollinger P, Brailas CD, Drewinko B: Evaluation of whole-blood platelet analyzers. *Ibid 14*: 492, 1983.
4. *Central File for Rare Donors*, Am Assoc Blood Banks, Milwaukee.
5. *ABO and Rh Systems*, Ortho Diagnostics, Raritan NJ 1969.
6. Berson S, Yalow R: *Gastroenterol 62*: 1061, 1972.
7. *Fed Reg 37FR17419*, Aug 26, 1972.
8. Broughton PMG, Dawson JB: *Advan Clin Chem 15*: 288, 1972.
9. Mears T, Young D: *Am J Clin Pathol 50*: 411, 1968.
10. Meinke W: *Anal Chem 43*: 28A, 1971.
11. Dybaker R: *Std Methods Clin Chem 6*: 223, 1970.
12. Flokstra JH, Soda JA: *Am J Med Sci 254*: 429, 1967.
13. Young DS *et al*: *Clin Chem 21*: 1D, 1975.
14. Constantino NV, Kabat HF: *Am J Hosp Pharm 30*: 24, 1973.
15. Peterson CM: *Diagn Med*: 78, Jul/Aug 1980.
16. *Radial Immunodiffusion and Immunoelectrophoreses for Qualitation and Quantitation of Immunoglobulins* (DHEW Publ HSM-72-8102), USD-HEW, Washington DC, 1972.
17. *Bull WHO 43*: 891, 1970.
18. Statland BE: *Clinical Decision Levels For Lab Tests*, Med Econ Co Inc., Oradell NJ, 1983.
19. White W *et al*: *Practical Automation for the Clinical Laboratory*, Mosby, St Louis, 1972.
20. Szilagyi G, Aning V, Karmen A: *J Clin Lab Automation 3*: 117, 1983.
21. Bradley GM: Fecal analysis. *Diagn Med* 63 Mar/Apr 1980.
22. Rubenstein K *et al*: *Biochem Biophys Res Comm 47*: 846, 1972.
23. Baron J: *Scand J Gastroenterol 5*: 9, 1970.
24. Jorpes J, Mott V: *Secretin, CCK, Pancreozymin and Gastrin*, Springer Verlag, New York, 1973.
25. Kuhn PJ: *Med Lab Observer* 108, Sept 1983.
26. Feldman M, Nossal GJ: *Quart Rev Biol 47*: 269, 1972.
27. Skelley DS, *et al*: *Clin Chem 19*: 146, 1973.

Bibliography

Wintrobe M: *Clinical Hematology*, 6th ed, Lea & Febiger, Philadelphia, 1967.

Lynch MJ: *Medical Laboratory Technology*, 2nd ed, Saunders, Philadelphia, 1969.

Faulkner W, King J: *Manual Clinical Laboratory Procedures*, Chemical Rubber Co, Cleveland, 1970.

Faulkner W *et al*: *Handbook Clinical Laboratory Data*, Chemical Rubber Co, Cleveland, 1980.

Roth K, Saunders A: *Evaluation of Methods for White Cell Identification and Counting-Advances in Automated Analysis*, Technicon Intl Congr, 1970.

Dacie J, Lewis S: *Practical Hematology*, 3rd ed, Churchill, London, 1963.

Frankel S, Reitman S, eds: *Clinical Laboratory Methods and Diagnosis*, 6th ed, Mosby, St. Louis, 1963.

Wintrobe MM: *Laboratory Medicine-Hematology*, 2nd ed, Mosby, St. Louis, 1962.

Manual of Blood Coagulation Technics, 2nd ed, Warner-Chilcott, Morris Plans NJ, 1966.

Detection of Fibrinogen Degradation Products, Wellcome Res Labs, England, 1973.

A Manual of Methods for the Coagulation Laboratory, BD & Co, Rutherford NJ, 1965.

Technical Methods and Procedures of the American Association of Blood Banks, Am Assoc Blood Banks, Chicago, 1962.

Standards for Blood Transfusion Service, 4th ed, Am Assoc Blood Banks, Chicago, 1963.

Griffiths JJ, Elliott J: *Blood Bank Procedures*, Dade Reagents, Miami, 1967.

Chromatography in Mass Screening for Disorders of Amino Acid Metabolism, Hyland, Los Angeles, 1966.

Rosalki S, Wilkinson J: *Diagnostic Enzymology*, Dade Reagents, Miami, 1966.

Wilkinson J: *Introduction to Diagnostic Enzymology*, Edward Arnold, Ltd. London, 1962.

Davidsohn I, Henry J: *Todd-Sanford Clinical Diagnosis by Laboratory Methods*, 15th ed, Saunders, Philadelphia, 1974.

Peron FG, Caldwell BV: *Immunologic Methods in Steroid Determination*, Appleton-Century, New York, 1970.

Winsten S, Dalal F: *Clinical Laboratory Procedures for Nonroutine Problems*, Chemical Rubber Co, Cleveland, 1972.

Specialized Diagnostic Laboratory Tests, Bioscience Labs, Van Nuys CA, 1971.

Kark RM *et al*: *A Primer of Urinalysis*, 2nd ed, Harper & Row, New York, 1963.

Sunderman FW, Sunderman FW, Jr: *Laboratory Diagnosis of Renal Diseases*, Warren H Green, Inc, St Louis, 1970.

Faust E, Russell P: *Clinical Parasitology*, 7th ed, Lea & Febiger, Philadelphia, 1964.

Sunshine I: *Manual of Analytical Toxicology*, Chemical Rubber Co., Cleveland, 1972.

Clarke, E: *Isolation and Identification of Drugs*, vols 1 and 2, Pharmaceutical Press, London, 1969 and 1975.

Blair JE *et al*: *Manual of Clinical Microbiology*, Williams & Wilkins, Baltimore, 1970.

Edwards PR, Ewing WH: *Identification of Enterobacteriaceae*, 3rd ed, Burgess, Minneapolis, 1972.

Holdeman LV, Moore WEC: *Anaerobe Laboratory Manual*, 2nd ed, VPI Anaerobe Lab, Blacksburg VA, 1973.

Conant NF *et al*: *Manual of Clinical Mycology*, 3rd ed, Saunders, Philadelphia, 1971.

Buch F, Good R: *Clinical Immunobiology*, Academic, New York, 1972.

Clinical RIA. *Lab Manag*: May 1973.

Manual of tissue typing techniques. *Natl Inst All Infect Dis Bull*: 1972.

Directory of Rare Analysis. *Clin Chem 23*: 323, 1977.

Doucet LD, *Medical Technology Review*, Lippincott, Philadelphia, 1981.

Hansten PD, *Drug Interactions*, 3rd ed, Lea & Febiger, Philadelphia, 1975.

Kaplan A, Szabo LL, *Clinical Chemistry: Interpretation and Techniques*, 2nd ed, Lea & Febiger, Philadelphia, 1983.

Miller SE, Weller JM, *Textbook of Clinical Pathology*, 8th ed, Williams & Wilkins, Baltimore, 1971.

Peacock J, Tomar R, *Manual of Laboratory Immunology*, Lea & Febiger, Philadelphia, 1980.

Pertinent Reference Journals

Advan Clin Chem	*J Clin Lab Automation*
Am J Clin Pathol	*Diagn Med*
Am Clin Prod Rev	*J Lab Clin Med*
Am J Hosp Pharm	*Lab Med*
Am J Med Technol	*Lab Notes Med Diag*
Anal Chem	*Med Lab Obs*
BioTechniques	*Med Lab Tech*
Clin Chem	*Scand J Clin Lab Invest*
Clin Chim Acta	*Std Methods Clin Chem*

CHAPTER 35

Drug Absorption, Action and Disposition

Stewart C Harvey, PhD
Professor of Pharmacology
School of Medicine, University of Utah
Salt Lake City, UT 84132

Although drugs differ widely in their pharmacodynamic effects and clinical application, in penetrance, absorption and usual route of administration, in distribution among the body tissues and in disposition and mode of termination of action, there are certain general principles that help explain these differences. These principles have both pharmaceutic and therapeutic implications. They facilitate an understanding of both the features that are common to a class of drugs and the differentia among the members of that class.

In order for a drug to act it must be absorbed, transported to the appropriate tissue or organ, penetrate to the responding subcellular structure and elicit a response or alter ongoing processes. The drug may be distributed simultaneously or sequentially to a number of tissues, bound or stored, metabolized to inactive or active products or excreted. The history of a drug in the body is summarized in Fig 35-1. Each of the processes or events depicted relates importantly to therapeutic and toxic effects of a drug and to the mode of administration, and drug design must take each into account. Since the effect elicited by a drug is its *raison d'etre*, *drug action* and *effect* will be discussed first in the text that follows, even though they are preceded by other events.

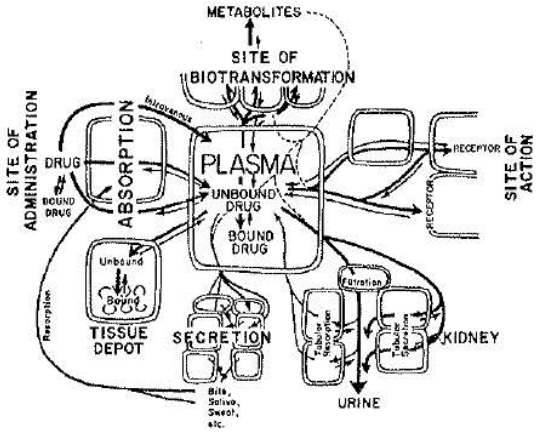

Fig 35-1. The absorption, distribution, action and elimination of a drug (arrows represent drug movement). Intravenous administration is the only process whereby a drug may enter a compartment without passing through a biological membrane. Note that drugs excreted in bile and saliva may be resorbed.

Drug Action and Effect

The word *drug* imposes an action–effect context within which the properties of a substance are described. The description of necessity must include the pertinent properties of the recipient of the drug. Thus, when a drug is defined as an analgesic, it is implied that the recipient reacts in a certain way, called pain,* to a noxious stimulus. Both because the pertinent properties are locked into the complex and somewhat imprecise biological context and because the types of possible response are many, descriptions of the properties of drugs tend to emphasize the qualitative features of the effects they elicit. Thus, a drug may be described as having analgesic, vasodepressor, convulsant, antibacterial, etc, properties. The specific effect (or use) categories into which the many drugs may be placed are the subject of Chapters 38 through 65 and will not be elaborated upon in this chapter. However, the description of a drug does not end with the enumeration of the responses it may elicit. There are certain intrinsic properties of the drug–recipient system that can be described in quantitative terms and which are essential to the full description of the drug and to the validation of the drug for specific uses. Under *Definitions and Concepts*, below, certain general terms are defined in qualitative language; under *Dose–Effect Relationships* the foundation is laid for an appreciation of some of the quantitative aspects of pharmacodynamics.

Definitions and Concepts

In the field of pharmacology, the vocabulary that is unique to the discipline is relatively small, and the general vocabulary is that of the biological sciences and chemistry. Nevertheless, there are a few definitions that are important to the proper understanding of pharmacology. It is necessary to differentiate among action, effect, selectivity, dose, potency and efficacy.

Action vs Effect—The *effect* of a drug is an *alteration of function* of the structure or process upon which the drug acts. It is common to use the term action as a synonym for effect. However, action precedes effect. *Action* is the *alteration of condition* that brings about the effect.

The final effect of a drug may be far removed from its site of action. For example, the diuresis subsequent to the ingestion of ethanol does not result from an action on the kidney but instead from a depression of activity in the supraopticohypophyseal region of the hypothalamus, which regulates the release of antidiuretic hormone from the posterior pituitary gland. The alteration of supraopticohypo-

* Sophisticated studies indicate that pain is not simply the *perception* of a certain kind of stimulus but, rather, a *reaction* to the perception of a variety of kinds of stimuli or stimulus patterns.

physeal function is, of course, also an effect of the drug, as is each subsequent change in the chain of events leading to diuresis. The action of ethanol was exerted only at the initial step, each subsequent effect being then the action to a following step.

Multiple Effects—No known drug is capable of exerting a single effect, although a number are known that appear to have a single mechanism of action. Multiple effects may derive from a single mechanism of action. For example, the inhibition of acetylcholinesterase by physostigmine will elicit an effect at every site where acetylcholine is produced, is potentially active, and is hydrolyzed by cholinesterase. Thus, physostigmine elicits a constellation of effects.

A drug also can cause multiple effects at several different sites by a single action at only one site, providing that the function initially altered at the site of action ramifies to control other functions at distant sites. Thus, a drug that suppresses steroid synthesis in the liver may not only lower serum cholesterol, impair nerve myelination and function and alter the condition of the skin (as a consequence of cholesterol deficiency) but also may affect digestive functions (because of a deficiency in bile acids) and alter adrenocortical and sexual hormonal balance.

Although a single action can give rise to multiple effects, most drugs exert multiple actions. The various actions may be related, as, for example, the sympathomimetic effects of metaraminol that accrue to its structural similarity to norepinephrine and its ability partially to suppress sympathetic responses because it occupies the catecholamine storage pools in lieu of norepinephrine; or the actions may be unrelated, as with the actions of morphine to interfere with the release of acetylcholine from certain autonomic nerves, block some actions of 5-hydroxytryptamine (serotonin) and release histamine. Many drugs bring about immunologic (allergic or hypersensitivity) responses that bear no relation to the other pharmacodynamic actions of the drug.

Selectivity—Despite the potential most drugs have for eliciting multiple effects, one effect is generally more readily elicitable than another. This differential responsiveness is called *selectivity*. It usually is considered to be a property of the drug, but it is also a property of the constitution and biodynamics of the recipient subject or patient.

Selectivity may come about in several ways. The subcellular structure (receptor) with which a drug combines to initiate one response may have a higher affinity for the drug than that for some other action. Atropine, for example, has a much higher affinity for muscarinic receptors (page 889) that subserve the function of sweating than it does for the nicotinic receptors (page 889) that subserve voluntary neuromuscular transmission, so that suppression of sweating can be achieved with only a tiny fraction of the dose necessary to cause paralysis of the skeletal muscles. A drug may be distributed unevenly, so that it reaches a higher concentration at one site than throughout the tissues generally; chloroquine is much more effective against hepatic than intestinal (colonic) amebiasis because it reaches a much higher concentration in the liver than in the wall of the colon. An affected function may be much more critical to or have less reserve in one organ than in another, so that a drug will be predisposed to elicit an effect at the more critical site. Some inhibitors of dopa decarboxylase (which is also 5-hydroxytryptophan decarboxylase) depress the synthesis of histamine more than that of either norepinephrine or 5-hydroxytryptamine (serotonin), even though histidine decarboxylase is less sensitive to the drug, simply because histidine decarboxylase is the only step and, hence, is rate-limiting in the biosynthesis of histamine. Dopa decarboxylase is not rate-limiting in the synthesis of either norepinephrine or 5-hydroxytryptamine until the enzyme is nearly completely inhibited. Another example of the determination of selectivity by the critical balance of the affected function is that of the mercurial diuretic drugs. An inhibition of only 1% in the tubular resorption of glomerular filtrate usually will double urine flow, since 99% of the glomerular filtrate is normally resorbed. Aside from the question of the possible concentration of diuretics in the urine, a drug-induced reduction of 1% in sulfhydryl enzyme activity in tissues other than the kidney usually is not accompanied by an observable change in function. Selectivity also can be determined by the pattern of distribution of destructive or activating enzymes among the tissues and by other factors.

Dose—Even the uninitiated person knows that the *dose* of a drug is the amount administered. However, the appropriate dose of a drug is not some unvarying quantity, a fact sometimes overlooked by pharmacists, official committees and physicians. The practice of pharmacy is entrapped in a system of fixed-dose formulations, so that fine adjustments in dosage are often difficult to achieve. Fortunately, there is usually a rather wide latitude allowable in dosages. It is obvious that the size of the recipient individual should have a bearing upon the dose, and the physician may elect to administer the drug on a body-weight or surface-area basis rather than as a fixed dose. Usually, however, a fixed dose is given to all adults, unless the adult is exceptionally large or small. The dose for infants and children often is determined by one of several formulas which take into account age or weight, depending on the age group of the child and the type of action exerted by the drug. Infants, relatively, are more sensitive to many drugs, often because enzyme systems which destroy the drugs may not be developed fully in the infant.

The nutritional condition of the patient, the mental outlook, the presence of pain or discomfort, the severity of the condition being treated, the presence of secondary disease or pathology, genetic and many other factors affect the dose of a drug necessary to achieve a given therapeutic response or to cause an untoward effect (Chapter 67). Even two apparently well-matched normal persons may require widely different doses for the same intensity of effect. Furthermore, a drug is not always employed for the same effect and, hence, not in the same dose. For example, the dose of a progestin necessary for an oral contraceptive effect is considerably different from that necessary to prevent spontaneous abortion, and a dose of an estrogen for the treatment of the menopause is much too small for the treatment of prostatic carcinoma.

From the above it is evident that the wise physician knows that *the dose of a drug is "enough"* (ie, no rigid quantity but rather that which is necessary and can be tolerated) and individualizes the regimen accordingly. The wise pharmacist also will appreciate this dictum and recognize that official or manufacturer's recommended doses are sometimes quite narrowly defined and may be very wide of the mark. They should serve only as a useful guide rather than as an imperative.

Potency and Efficacy—The *potency* of a drug is the reciprocal of dose. Thus, it will have the units of persons/unit weight of drug or body weight/unit weight of drug, etc. Potency generally has little utility other than to provide a means of comparing the relative activities of drugs in a series, in which case *relative potency*, relative to some prototype member of the series, is a parameter commonly used among pharmacologists and in the pharmaceutical industry.

Whether a given drug is more potent than another has little bearing on its clinical usefulness, provided that the potency is not so low that the size of the dose is physically unmanageable or the cost of treatment is higher than with an equivalent drug. If a drug is less potent but more selective, it is the one to be preferred. Promotional arguments in favor of a more potent drug thus are irrelevant to the impor-

tant considerations that should govern the choice of a drug. However, it sometimes occurs that drugs of the same class differ in the maximum intensity of effect; that is, some drugs of the class may be less efficacious than others, irrespective of how large a dose is used.

Efficacy connotes the property of a drug to achieve the desired response, and *maximum efficacy* denotes the maximum achievable effect. Even huge doses of codeine often cannot achieve the relief from severe pain that relatively small doses of morphine can; thus, codeine is said to have a lower maximum efficacy than morphine. Efficacy is one of the primary determinants of the choice of a drug.

Dose–Effect Relationships

The importance of knowing how changes in the intensity of response to a drug vary with the dose is virtually self-evident. Both the physician, who prescribes or administers a drug, and the manufacturer, who must package the drug in appropriate dose sizes, must translate such knowledge into everyday practice. Theoretical or molecular pharmacologists also study such relationships in inquiries into mechanism of action and receptor theory (see page 702). It is necessary to define two types of relationships: (1) dose–intensity relationship—ie, the manner in which the intensity of effect in the individual recipient relates to dose—and (2) dose–frequency relationship—ie, the manner in which the number of responders among a population of recipients relates to dose.

Dose–Intensity of Effect Relationships—Whether the intensity of effect is determined *in vivo* (eg, the blood-pressure response to epinephrine in the human patient) or *in vitro* (eg, the response of the isolated guinea pig ileum to histamine), the dose–intensity of effect (often called dose-effect) curve usually has a characteristic shape, namely a curve that closely resembles one quadrant of a rectangular hyperbola.

In the dose–intensity curve depicted in Fig 35-2, the curve appears to intercept the x axis at 0 only because the lower doses are quite small on the scale of the abscissa, the smallest dose being 1.5×10^{-3} µg. Actually, the x intercept has a positive value, since a finite dose of drug is required to bring about a response, this lowest effective dose being known as the *threshold dose*. Statistics and chemical kinetics predict

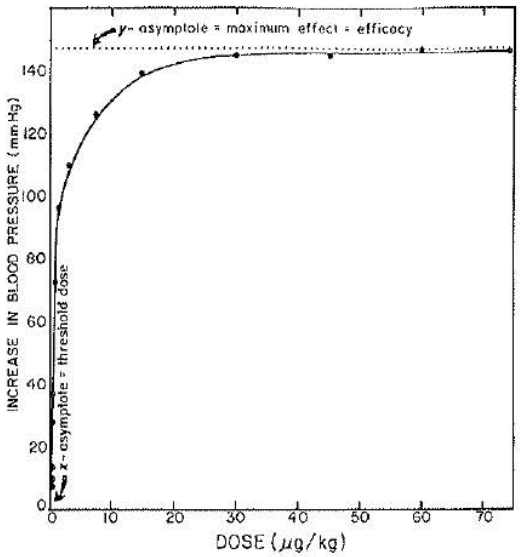

Fig 35-2. The relationship of the intensity of the blood-pressure response of the cat to the intravenous dose of norepinephrine.

that the curve should approach the y axis asymptotically. However, if the intensity of the measured variable does not start from zero, the curve possibly may have a positive y intercept (or negative x intercept), especially if the ongoing basal activity before the drug is given is closely related to that induced by the drug.

In practice, instead of an asymptote to the y axis, dose–intensity curves nearly always show an upward concave foot at the origin of the curve, so that the curve has a lopsided sigmoid shape. At high doses, the curve approaches an asymptote which is parallel to the x axis, and the value of the asymptote establishes the maximum possible response to the drug, or *maximum efficacy*. However, experimental data in the regions of the asymptotes generally are too erratic to permit an exact definition of the curve at the very low and very high doses. The example shown represents an unusually good set of data.

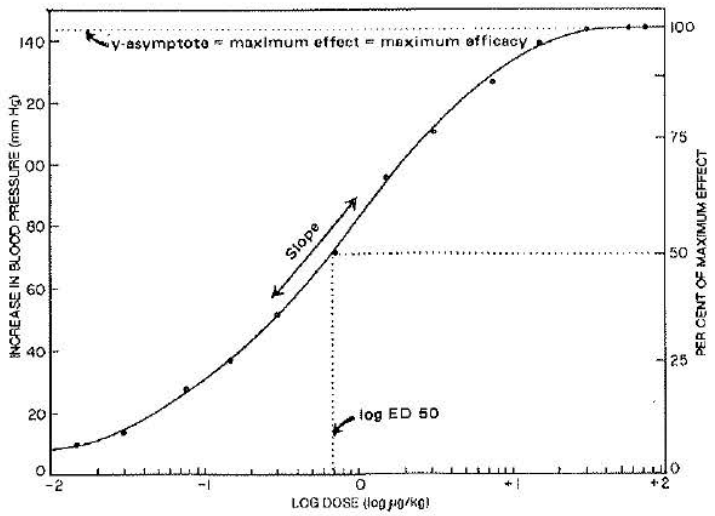

Fig 35-3. The relationship of the intensity of the blood-pressure response of the cat to the log of the intravenous dose of norepinephrine.

Because the dose range may be 100- or 1000-fold from the lowest to the highest dose, it has become the practice to plot dose–intensity curves on a logarithmic scale of abscissa; ie, to plot the log of dose versus the intensity of effect. Figure 35-3 is such a semilogarithmic plot of the same data as in Fig 35-2. In the figure the intensity of effect is plotted both in absolute units (at the left) or in relative units, as percent (at the right).

Although no new information is created by a semilogarithmic representation, the curve is stretched in such a way as to facilitate the inspection of the data; the comparison of results from multiple observations and the testing of different drugs also is rendered easier. In the example shown, the curve is essentially what is called a *sigmoid curve* and is nearly symmetrical about the point which represents an intensity equal to 50% of the maximal effect, ie, about the midpoint. The symmetry follows from the rectangular hyperbolic character of the previous Cartesian plot (Fig 35-2). The semilogarithmic plot reveals better the dose–effect relationships in the low-dose range, which are lost in the steep slope of the Cartesian plot. Furthermore, the data about the midpoint are almost a straight line; the nearly linear portion covers approximately 50% of the curve. The slope of the linear portion of the curve or, more correctly, the slope at the point of inflection, has theoretical significance (see *Drug Receptors and Receptor Theory*, page 702).

The upper portion of the curve approaches an asymptote, which is the same as that in the Cartesian plot. If the response system is completely at rest before the drug is administered, the lower portion of the curve should be asymptotic to the x axis. Both asymptotes and the symmetry derive from the law of mass action (see page 703).

Dose–intensity curves often deviate from the ideal configuration illustrated and discussed above. Usually, the deviate curve remains sigmoid but not extended symmetrically about the midpoint of the *linear* segment. Occasionally, other shapes occur, sometimes quite bizarre ones. Deviations may derive from multiple actions that converge upon the same final effector system, from varying degrees of metabolic alteration of the drug at different doses, from modulation of the response by feedback systems, from nonlinearity in the relationship between action and effect or from other causes.

It is frequently necessary to identify the dose which elicits a given intensity of effect. The intensity of effect that is generally designated is the 50% of maximum intensity. The corresponding dose is called the *50% effective dose*, or *individual ED50* (see Fig 35-3). The use of the adjective, *individual*, distinguishes the ED50 based upon the intensity of effect from the median effective dose, also abbreviated ED50, determined from frequency of response data in a population (see *Dose–Frequency Relationships*, this page).

Drugs that elicit the same quality of effect may be compared graphically. In Fig 35-4, five hypothetical drugs are compared. Drugs *A*, *B*, *C* and *E* all can achieve the same maximum effect, which suggests that the same effector system may be common to all. *D* possibly may be working through the same effector system, but there are no *a priori* reasons to think this is so. Only *A* and *B* have parallel curves and common slopes. Common slopes are consistent with, but in no way prove, the idea that *A* and *B* not only act through the same effector system but also by the same mechanism. Although drug–receptor theory (see *Drug Receptors and Receptor Theory*, page 702) requires that the curves of identical mechanism have equal slopes, examples of exceptions are known. Furthermore, mass-law statistics require that all simple drug-receptor interactions generate the same slope; only when slopes depart from this universal slope in accordance with distinctive characteristics of the response system do they provide evidence of specific mechanisms.

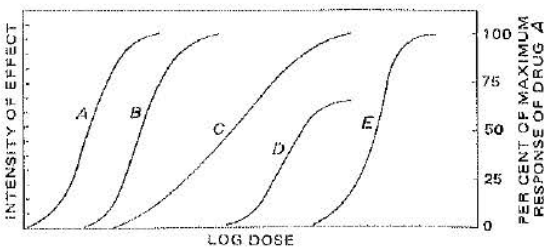

Fig 35-4. Log dose–intensity of effect curves of five different hypothetical drugs (see text for explanation).

The relative potency of any drug may be obtained by dividing the ED50 of the standard, or prototype, drug by that of the drug in question. Any level of effect other than 50% may be used, but it should be recognized that when the slopes are not parallel the relative potency depends upon the intensity of effect chosen. Thus, the potency of *A* relative to *C* (in Fig 35-4) calculated from the ED50 will be smaller than that calculated from the ED25.

The low maximum intensity inducible by *D* poses even more complications in the determination of relative potency than do the unequal slopes of the other drugs. If its dose–intensity curve is plotted in terms of percent of its own maximum effect, its relative inefficacy is obscured and the limitations of relative potency at the ED50 level will not be evident. This dilemma simply underscores the fact that drugs can be compared better from their entire dose–intensity curves than from a single derived number like ED50 or relative potency.

Drugs that elicit multiple effects will generate a dose–intensity curve for each effect. Even though the various effects may be qualitatively different, the several curves may be plotted together on a common scale of abscissa, and the intensity may be expressed in terms of percent of maximum effect; thus, all curves can share a common scale of ordinates in addition to common abscissa. Separate scales of ordinates could be employed, but this would make it harder to compare data.

The selectivity of a drug can be determined by noting what percent of maximum of one effect can be achieved before a second effect occurs. As with relative potency, selectivity may be expressed in terms of the ratio between the ED50 for one effect to that for another effect, or a ratio at some other intensity of effect. Similarly to relative potency, difficulties follow from nonparallelism. In such instances, selectivity expressed in dose ratios varies from one intensity level to another.

When the dose–intensity curves for a number of subjects are compared, it is found that they vary considerably from individual to individual in many respects; eg, threshold dose, midpoint, maximum intensity and sometimes even slope. By averaging the intensities of the effect at each dose, an average dose–intensity curve can be constructed.

Average dose–intensity curves enjoy a limited application in comparing drugs. A single line expressing an average response has little value in predicting individual responses unless it is accompanied by some expression of the range of the effect at the various doses. This may be done by indicating the standard error of the response at each dose. Occasionally, a simple scatter diagram is plotted in lieu of an average curve and statistical parameters (see Fig 10-21). An average dose–intensity curve also may be constructed from a population in which different individuals receive different doses; if sufficiently large populations are employed, the average curves determined by the two methods will approximate each other.

It is obvious that the determination of such average curves from a population sufficiently large to be statistically meaningful requires a great deal of work. Retrospective clinical data occasionally are treated in this way, but prospective studies infrequently are designed in advance to yield average curves. The usual practice in comparing drugs is to employ a quantal (all-or-none) end-point and plot the frequency or cumulative frequency of response over the dose range, as discussed below.

Dose–Frequency of Response Relationships—When an end-point is truly all-or-none, such as death, it is an easy matter to plot the number of responding individuals (eg, dead subjects) at each dose of drug or intoxicant. Many other responses that vary in intensity can be treated as all-or-none if simply the presence or absence of a response (eg, cough or no cough, convulsion or no convulsion) is recorded, without regard to the intensity of the response when it occurs.

When the response changes from the basal or control state in a less abrupt manner (eg, tachycardia, miosis, rate of gastric secretion) it may be necessary to designate arbitrarily some particular intensity of effect as the end-point. If the end-point is taken as an increase in heart rate of 20 beats/min, all individuals whose tachycardia is less than 20/min would be recorded as nonresponders, while all those with 20 or above would be recorded as responders. When the percent of responders in the population is plotted against the dose, a characteristic dose–response curve, more properly called a *dose–cumulative frequency* or *dose–percent* curve, is generated. Such a curve is, in fact, a cumulative frequency–distribution curve, the percent of responders at a given dose being the frequency of response.

Dose–cumulative frequency curves are generally of the same geometric shape as dose–intensity curves (namely, sigmoid) when frequency is plotted against log dose (see Fig 35-5). The tendency of the cumulated frequency of response (ie, percent) to be linearly proportional to the log of the dose in the middle of the dose range is called the *Weber–Fechner law*, although it is not invariable, as a true natural law should be. In many instances, the cumulative frequency is simply proportional to dose rather than log dose. The Weber–Fechner law applies to either dose–intensity or dose–cumulative frequency data. The similarity between dose–frequency and dose–intensity curves may be more than fortuitous, since the intensity of response will usually have an approximately linear relationship to the percent of responding *units* (smooth muscle cells, nerve fibers, etc) and, hence, is also a type of cumulative frequency of response. These are the same kind of statistics that govern the law of mass action.

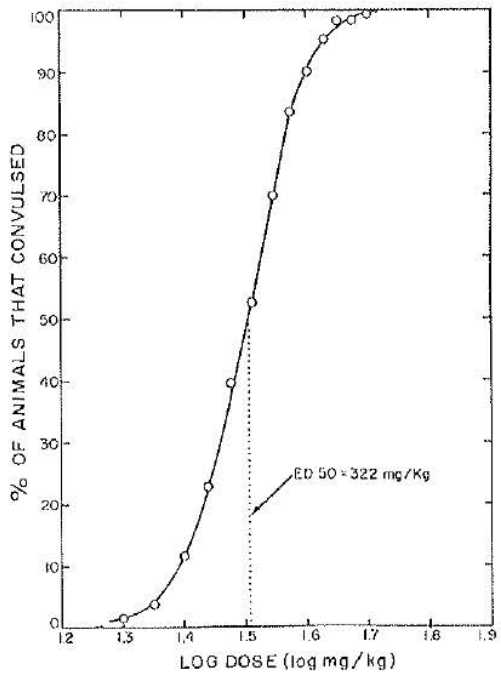

Fig 35-5. The relationship of the number of responders in a population of mice to the dose of pentylenetetrazol (courtesy, Drs DG McQuarry and EG Fingl, University of Utah).

If only the increase in the number of responders with each new dose is plotted, instead of the cumulative percent of responders, a bell-shaped curve is obtained. This curve is the first derivative of the dose–cumulative frequency curve and is a *frequency–distribution* curve (see Chapter 10). The distribution will be symmetrical—ie, *normal* or Gaussian (see Fig 10-5)—only if the dose–cumulative frequency curve is symmetrically hyperbolic. Because most dose–cumulative frequency curves are more nearly symmetrical when plotted semilogarithmically (ie, as log dose), dose–cumulative frequency curves are usually *log-normal*.

Since the dose–intensity and dose–cumulative frequency curves are basically similar in shape, it follows that the curves have similar defining characteristics, such as ED50, maximum effect (maximum efficacy) and slope. In dose–cumulative frequency data, the ED50 (*median effective dose*) is the dose to which 50% of the population responds (see Fig 35-5). If the frequency distribution is normal, the ED50 is both the arithmetic mean and median dose and is represented by the midpoint on the curve; if the distribution is log-normal, the ED50 is the median dose but not the arithmetic mean dose. The efficacy is the cumulative frequency summed over all doses; it is usually, but not always, 100%. The slope is characteristic of both the drug and test population. Even two drugs of identical mechanism may give rise to different slopes in dose–percent curves, whereas in dose–intensity curves the slopes are the same.

Statistical parameters (such as standard deviation), in addition to ED50, maximum cumulative frequency (efficacy) and slope, characterize dose-cumulative frequency relationships (see Chapter 10).

There are several formulations for dose–cumulative frequency curves, some of which are employed only to define the linear segment of a curve and to determine the statistical parameters of this segment. For the statistical treatment of dose–frequency data, see Chapter 10. One simple mathematical expression of the entire log-symmetrical sigmoid curve is

$$\log \text{dose} = K + f \log\left(\frac{\%\ \text{response}}{100\% - \text{response}}\right) \qquad (1)$$

where percent response may be either the percent of maximum intensity or the percent of a population responding. The equation is thus basically the same for both log normal dose–intensity and log normal dose–percent relationships. K is a constant that is characteristic of the midpoint of the curve, or ED50, and $1/f$ is characteristically related to the slope of the linear segment, which, in turn is closely related to the standard deviation of the derivative log-normal frequency-distribution curve.

The comparison of dose–percent relationships among drugs is subject to the pitfalls indicated for dose–intensity comparisons (see page 699), namely, that when the slopes of the curves are not the same (ie, the dose–percent curves are not parallel), it is necessary to state at which level of response a potency ratio is calculated. As with dose–intensity

data, potencies generally are calculated from the ED50, but potency ratios may be calculated for any arbitrary percent response. The expression of selectivity is, likewise, subject to similar qualifications, inasmuch as the dose–percent curves for the several effects are usually nonparallel.

The term *therapeutic index* is used to designate a quantitative statement of the selectivity of a drug when a therapeutic and an untoward effect are being compared. If the untoward effect is designated as *T* (for toxic) and the therapeutic effect as *E*, the therapeutic index may be defined as TD50/ED50 or a similar ratio at some other arbitrary levels of response. The TD and the ED are not required to express the same percent of response; some clinicians use the ratio TD1/ED99 or TD5/ED95, based on the rationale that if the untoward effect is serious, it is important to use a most-severe therapeutic index in passing judgment upon the drug. Unfortunately, therapeutic indices are known in man for only a few drugs.

There will be a different therapeutic index for each untoward effect that a drug may elicit, and, if there is more than one therapeutic effect, a family of therapeutic indices for each therapeutic effect. However, in clinical practice, it is customary to distinguish among the various toxicities by indicating the percent incidence of a given side effect.

Variations in Response and Responsiveness—From the above discussion of dose–frequency relationships and Chapter 10, it is obvious that in a normal population of persons there may be quite a large difference in the dose required to elicit a given response in the least-responsive member of the population and that to elicit the response in the most-responsive member. The difference ordinarily will be a function of the slope of the dose–percent curve, or, in statistical terms, of the standard deviation. If the standard deviation is large, the extremes of responsiveness of responders are likewise large.

In a normal population 95.46% of the population responds to doses within two standard deviations from the ED50 and 99.73% within three standard deviations. In log-normal populations the same distribution applies when standard deviation is expressed as log dose.

In the population represented in Fig 35-5, 2.25% of the population (two standard deviations from the median) would require a dose more than 1.4 times the ED50; an equally small percent would respond to 0.7 the ED50. The physician who is unfamiliar with statistics is apt to consider the 2.25% at either extreme as abnormal reactors. The statistician will argue that these 4.5% are within the normal population and only those who respond well outside of the normal population, at least three standard deviations from the median, deserve to be called abnormal.

Irrespective of whether the criteria of abnormality that the physician or the statistician obtain, the term *hyporeactive* applies to those individuals who require abnormally high doses and *hyperreactive* to those who require abnormally low doses. The terms *hyporesponsive* and *hyperresponsive* also may be used. It is incorrect to use the terms hyposensitive and hypersensitive in this context; *hypersensitivity* denotes an allergic response to a drug and should not be used to refer to hyperreactivity. The term *supersensitivity* correctly applies to hyperreactivity that results from denervation of the effector organ; it is often more definitively called denervation supersensitivity. Sometimes hyporeactivity is the result of an immunochemical deactivation of the drug, or *immunity*. Hyporeactivity should be distinguished from an increased dose requirement that results from a severe pathological condition. Severe pain requires large doses of analgesics, but the patient is not a hyporeactor; what has changed is the baseline from which the endpoint quantum is measured. The responsiveness of a patient to certain drugs sometimes may be determined by the history of previous exposure to appropriate drugs.

Tolerance is a diminution in responsiveness as use of the drug continues. The consequence of tolerance is an increase in the dose requirement. It may be due to an increase in the rate of elimination of drug (as discussed elsewhere in this chapter), to reflex or other compensatory homeostatic adjustments, to a decrease in the number of receptors or in the number of enzyme molecules or other coupling proteins in the effector sequence, to exhaustion of the effector system or depletion of mediators, to the development of immunity or to other mechanisms. Tolerance may be gradual, requiring many doses and days to months to develop, or acute, requiring only the first or a few doses and only minutes to hours to develop. Acute tolerance is called *tachyphylaxis*.

Drug resistance is the decrease in responsiveness of microorganisms, neoplasms or pests to chemotherapeutic agents, antineoplastics or pesticides, respectively. It is not tolerance in the sense that the sensitivity of the individual microorganism or cancer cell decreases; rather, it is the survival of normally unresponsive cells which then pass the genetic factors of resistance on to their progeny.

Patients who fail to respond to a drug are called *refractory*. Refractoriness may result from tolerance or resistance, but it also may result from the progression of pathological states that negate the response or render the response incapable of surmounting an overwhelming pathology. Rarely, it may result from a poorly developed receptor or response system.

Sometimes a drug evokes an unusual response that is *qualitatively* different from the expected response. Such an unexpected response is called a *meta-reaction*. A not uncommon meta-reaction is a central nervous stimulant rather than depressant effect of phenobarbital, especially in women. Pain and certain pathological states sometimes favor *meta*-reactivity. Responses that are different in infants or the aged than in young and middle-aged people are not *meta*-reactions if the response is usual in the age group. The term *idiosyncrasy* also denotes *meta*-reactivity, but the word has been so abused that it is recommended that it be dropped. Although hypersensitivity may cause unusual effects, it is not included in *meta*-reactivity.

Drug Receptors and Receptor Theory

Most drugs act by combining with some key substance in the biological milieu that has an important regulatory function in the target organ or tissue. This biological partner of the drug goes by the name of *receptive substance* or *drug receptor*. The receptive substance is considered mostly to be a cellular constituent, although in a few instances it may be extracellular, as the cholinesterases are, in part. The receptive substance is thought of as having a special chemical affinity and structural requirements for the drug. Drugs such as emollients, which have a physical rather than chemical basis for their action, obviously do not act upon receptors. Drugs such as demulcents and astringents, which act in a nonselective or nonspecific chemical way, also are not considered to act upon receptors, since the candidate receptors have neither sharp chemical nor biological definition. Even antacids, which react with the extremely well-defined hydronium ion, cannot be said to have a receptor, since the reactive proton has no permanent biological residence.

Because of early preoccupation with physical theories of action and the classical and illogical dichotomy of chemical

and physical molecular interaction, there is a reluctance to admit receptors for drugs such as local anesthetics, general anesthetics, certain electrolytes, etc, which generally are not accepted to combine selectively with distinct cellular or organelle membrane constituents. The word receptor often is used inconsistently and intuitively. However, the term is a legitimate symbol for that biological structure with which a drug interacts to initiate a response. Ignorance of the identities of many receptors does not detract from, but rather increases, the importance of the term and general concept.

Once a receptor is identified, it frequently is no longer thought of as a receptor, although such identification may afford the basis of profound advances in receptor theory. Since the effects of anticholinesterases are derived only indirectly from inhibition of cholinesterase and no drugs are known that stimulate the enzyme, it may be argued that it is not a receptor. Nevertheless, a number of drugs ultimately act indirectly through the inhibition of such modulator enzymes and it is important for the theoretician to develop models based upon such indirect interrelations.

Enzymes, of course, readily suggest themselves as candidates for receptors. However, there is more to cellular function than enzymes. Receptors may be membrane or intracellular constituents that govern: the spatial orientation of enzymes, gene expression, compartmentalization of the cytoplasm, contractile or compliant properties of subcellular structures or permeability and electrical properties of membranes. For nearly every cellular constituent there can be imagined a possible way for a drug to affect its function; therefore, few cellular constituents can be dismissed *a priori* as possible receptors. All the receptors for neurotransmitters and autonomic agonists are membrane proteins with agonist-binding groups projecting into the extracellular space. The transducing apparatus, whereby an occupied receptor elicits a response, is called a *coupling system*. Excitatory neurotransmitters in the central nervous system, and nicotinic receptors elsewhere, are coupled to ion channels which, when opened, permit the rapid ingress, especially of sodium ions. GABA (γ-amino-butyric acid) and glycine are coupled to inhibitory chloride channels. Benzodiazepine receptors are coupled to the GABA-receptor. Beta-adrenergic receptors and a number of receptors for polypeptide hormones interact with a stimulatory GDP/GTP-binding protein (G-protein) which can activate the enzyme adenylate cyclase. The cyclase then produces 3′,5′-cyclic AMP (cAMP) which, in turn, activates protein kinases. Other receptors interact with inhibitory G-proteins. Some receptors couple to guanylate cyclase.

Alpha-adrenergic, some muscarinic and various other receptors couple to the membrane enzyme, phospholipase-C, which cleaves inositol phosphates from phosphoinositides. The cleavage product, 1,4,5-inositol triphosphate (IP_3), then causes an increase in intracellular calciun, whereas the product, diacylglycerol (DAG), activates kinase-C. There are a number of other less ubiquitous coupling systems. Substances such as cAMP, cGMP, IP_3 and DAG are called *second messengers*.

It has been found that there may be several different receptors for a given agonist. Differences may he shown not only in the types of coupling systems and effects hut also by differential binding of agonists and antagonists, desensitization kinetics, physical and chemical properties, genes and amino acid sequences. The differentiation among receptor subtypes is called *receptor classification*. Receptor subtypes are designated by Greek or Arabic alphabetical prefixes and/or numerical subscripts. There are at least two each of beta-adrenergic, histaminergic, serotoninergic, GABAergic and benzodiazepine receptors, probably three of muscarinic and alpha-adrenergic and five of opioid receptor subtypes.

Occupation and Other Theories

Drug-receptor interactions are governed by the law of mass action, a concept initiated by Langley in 1878. However, most chemical applications of mass law are concerned with the rate at which reagents disappear or products are formed, whereas receptor theory usually concerns itself with the fraction of the receptors combined with a drug, similar to theories of adsorption. The usual concept is that only when the receptor actually is occupied by the drug is its function transformed in such a way as to elicit a response. This concept has become known as the *occupation theory*. The earliest clear statement of its assumptions and formulations is often credited to Clark in 1926, but both Langley and Hill made important contributions to the theory in the first two decades of this century.

In all receptor theories, the terms agonist, partial agonist and antagonist are employed. An *agonist* is a drug that combines with a receptor to initiate a response.

In the classical occupation theory, two attributes of the drug are required: (1) *affinity*, a measure of the equilibrium constant of the drug-receptor interaction, and (2) *intrinsic activity*, or *intrinsic efficacy* (not to be confused with efficacy as intensity of effect), a measure of the ability of the drug to induce a positive change in the function of the receptor.

A *partial agonist* is a drug that can elicit some but not a maximal effect and which antagonizes an agonist. In the occupation theory it would be a drug with a favorable affinity but a low intrinsic activity.

A *competitive antagonist* is a drug that occupies a significant proportion of the receptors and thereby preempts them from reacting maximally with an agonist. In the occupation theory the prerequisite property is affinity without intrinsic activity.

A *noncompetitive antagonist* may react with the receptor in such a way as not to prevent agonist–receptor combination but to prevent the combination from initiating a response, or it may act to inhibit some subsequent event in the chain of action–effect–action–effect that leads to the final overt response.

The mathematical formulation of the receptor theories derives directly from the law of mass action and chemical kinetics. Certain assumptions are required to simplify calculations. The key assumption is that the intensity of effect is a direct linear function of the proportion of receptors occupied. The correctness of this assumption is most improbable on the basis of theoretical considerations, but empirically it appears to be a close enough approximation to he useful. A second assumption upon which formulations are based is that the drug-receptor interaction is at equilibrium. Another common assumption is that the number of molecules of receptor is negligibly small compared to that of the drug. This assumption is undoubtedly true in most instances, and departures from this situation greatly complicate the mathematical expression of drug-receptor interactions.

The first clearly stated mathematical formulation of drug–receptor kinetics was that of Clark.[1] In his equation,

$$Kx^n = \frac{y}{100 - y} \tag{2}$$

where K is the affinity constant, x is the concentration of drug, n is the molecularity of the reaction, and y is the percent of maximum response. Clark assumed that y was a linear function of the percent of receptors occupied by the drug, so that y could also symbolize the percent of receptors occupied. When the equation is rearranged to solve for y,

$$y = \frac{100Kx^n}{1 + Kx^n} \tag{3}$$

A Cartesian plot of this equation is identical in form to that shown in Fig 35-2. When y is plotted against $\log x$ instead of x, the usual sigmoid curve is obtained. Thus, it may be seen that the dose–intensity curve derives from mass action equilibrium kinetics, which in turn derive from the statistical nature of molecular interaction. The fact that dose–intensity and dose–percent curves have the same shape shows that they involve similar statistics.

If Eq 2 is put into log form

$$\log K + n \log x = \log \frac{y}{100 - y} \qquad (4)$$

a plot of $\log y/100 - y$ against $\log x$ then will yield a straight line with a slope of n; n is theoretically the number of molecules of drug which react with each molecule of receptor. At present, there are no known examples in which more than one molecule of agonist combines with a single receptor, hence, n should be equal to 1, universally. Nevertheless, n often deviates from 1; deviations occur because of cooperative interactions among receptors (*cooperativity*), *spare receptors* (see below), amplifications in the response system (*cascades*), receptor coupling to more than one sequence (eg, to both adenylate cyclase and calcium channels) and other reasons. In these departures from $n = 1$, the slope becomes a characteristic of the mechanism of action and response system.

The probability that a molecule of drug will react with a receptor is a function of the concentration of both drug and receptor. The concentration of receptor molecules cannot be manipulated like the concentration of a drug. But, as each molecule of drug combines with a receptor, the population of free receptors is diminished accordingly. If the drug is a competitive antagonist, it will diminish the probability of an agonist–receptor combination in direct proportion to the percent of receptor molecules preempted by the antagonist. Consequently, the intensity of effect will be diminished. However, the probability of agonist–receptor interaction can be increased by increasing the concentration of agonist, and the intensity of effect can be restored by appropriately larger doses of agonist. Addition of more antagonist will again diminish the response, which can, again, be overcome or *surmounted* by more agonist.

Clark showed empirically, and by theory, that as long as the ratio of antagonist to agonist was constant, the concentration of the competitive drugs could be varied over an enormous range without changing the magnitude of the response (see Fig 35-6). Since the presence of competitive antagonist only diminishes the probability of agonist–receptor combination at a given concentration of agonist and does not alter the molecularity of the reaction, it also follows that the effect of the competitive antagonist is to shift the dose–intensity curve to the right in proportion to the amount of antagonist present; neither shape nor slope of the curve is changed (see Fig 35-7). Both Figs 35-6 and 35-7 are from Clark's original paper on competitive antagonism.[1]

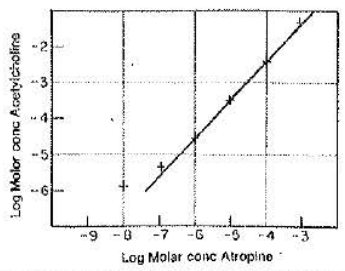

Fig 35-6. Direct proportionality of the dose of agonist (acetylcholine) to the dose of antagonist (atropine) necessary to cause a constant degree of inhibition (50%) of the response of the frog heart (courtesy, adaptation, Clark[1]).

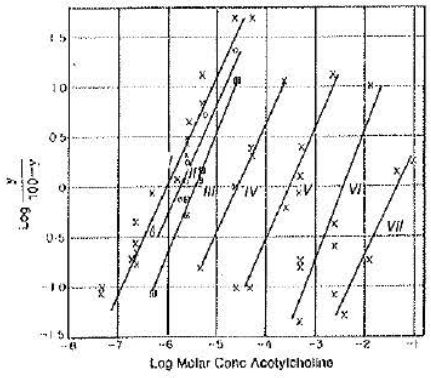

Fig 35-7. Effect of an antagonist to shift the log dose–intensity curve to the right without altering the slope. The effector is the isolated heart. *I:* no atropine; *II:* atropine, $10^{-9}M$; *III:* $10^{-7}M$; *IV:* $10^{-6}M$; *V:* $10^{-5}M$; *VI:* $10^{-4}M$; *VII:* $10^{-3}M$. Y: % of maximum intensity of response; the function $\log y/(100\text{-}y)$ converts the log dose–intensity relationship to a straight line (courtesy, adaptation, Clark[1]).

Many refinements of the Clark formula have been made, but they will not be treated here; details and citations of relevant literature can be found among various works on receptors cited in the Bibliography. Several refinements are introduced to facilitate studies of competitive inhibition. The introduction of the concepts of intrinsic activity[2] and efficacy[3] required appropriate changes in mathematical treatment.

Another important concept has been added to the occupation theory, namely the concept of *spare receptors*. Clark assumed the maximal response to occur only when the receptors were completely occupied, which does not account for the possibility that the maximum response might be limited by some step in the action–effect sequence subsequent to receptor occupation. Work with isotopically labeled agonists and antagonists and with dose–effect kinetics has shown that the maximal effect sometimes is achieved when only a small fraction of the receptors are yet occupied. The mathematical treatment of this phenomenon has enabled theorists to explain several puzzling observations that previously appeared to contradict occupation theory.

The classical occupation theory fails to explain several phenomena satisfactorily, and it is unable to generate a realistic model of intrinsic activity and partial agonism. A rate theory, in which the intensity of response is proportional to the rate of drug–receptor interaction instead of occupation, was proposed to explain some of the phenomena that occupation theory could not, but the rate theory was unable to provide a realistic mechanistic model of response generation, and it had other serious limitations as well.

The phenomena that neither the classical occupation nor rate theory could explain can be explained by various theories in which the receptor can exist in at least two conformational states, one of which is the active one; the drug can react with one or more conformers. In a *two-state* model[4]

$$R \rightleftharpoons R^*$$

where R is the inactive and R^* is the active conformer. The agonist combines mainly with R^*, the partial agonist can combine with both R and R^* and the antagonist can combine with R, the equilibrium being shifted according to the extent of occupation of R and R^*. Other variations of occupation theory treat the receptor as an aggregate of subunits which interact cooperatively.[5]

The Nature of Receptor Groups and Models of Receptors

A *receptor group* is that portion of the receptor molecule with which an agonist acts and which is vital to the function of the receptor. Studies of receptor group composition and configuration are too complex for the purposes of this text; consequently, only a brief sketch will be made here to orient the reader to the nature of the approach.

From the chemical configuration and reactivity of agonists and antagonists, certain deductions can be made about the structure of a receptor group. For example, all highly active agonists of muscarinic receptors are cations at physiological pH. This suggests that the receptor group contains an anionic group and that the force of attraction is electrostatic, at least in part, which agrees with thermodynamic data. That van der Waals forces (especially Heitler-London fluctuation forces) may also make an important contribution to binding is suggested by the requirement for *N*-methyl groups and by the low but definite activity of the nonionizable quaternary carbon analog of acetylcholine, 3,3-dimethylbutyl acetate. This establishes a requirement for an auxiliary structure close to the anionic site. Studies of the contribution to activity of ester and carbonyl oxygen among analogs of acetylcholine, intramolecular distances and the stereospecificity of various isomers and conformers have indicated a partial cationic (proton donor) site between 2.5 and 4 Å and a region of high electronic density (electron donor) between 5 and 7 Å from the anionic site. This is similar to the way in which the active site of acetylcholinesterase was mapped (see page 427, and Figs 25-44, 45 and 46).

The structure–activity relationships among competitive inhibitors also must be consistent with any model of a receptor. However, binding sites additional to the receptor group can be involved, and results are frequently more difficult to interpret than those with agonists. Nevertheless, studies with antagonists have made a substantial contribution to receptor group analysis. There is considerable interest in antagonists that combine irreversibly with the receptor, since such drugs offer a way of marking (affinity labeling) the receptor for isolation and for identification of the receptor group.

Since receptors for autonomic agonists are embedded in the cell membrane, they have been difficult to isolate without inactivation. Several laboratories have succeeded in isolating proteins, the chemical properties of which are consistent with those expected of various receptors. Receptors for steroid hormones have been easier to isolate, and some have been characterized relatively well. Further details of drug-receptor interactions and the nature of receptors can be found in the works on receptors and molecular pharmacology.

Up- and Down-Regulation—In many receptor-effector systems, if there is a paucity of agonist, the system will respond by increasing the responsiveness, number of receptors on the effector membrane or number of coupling proteins or enzymes in the effector system. This is known as *up-regulation*. In adrenergic systems, sympathetic denervation has been shown to increase the number of post-synaptic β-adrenoreceptors at some junctions and the availability of nucleotide-binding protein units and/or adenylate cyclase molecules at others. Hyperthyroid activity also increases the number of β-adrenoreceptors in heart muscle, which explains the excessive heart rate. Denervation of skeletal muscle causes a great multiplication of what is normally a minor type of nicotinic receptor, and the new receptors spread across the entire myocyte membrane. Prolonged blockade of receptors by antagonists also may cause up-regulation. The abrupt discontinuation of treatment, such that drug levels fall faster than re-regulation, may be followed by excessive activity, eg, in pernicious tachycardia and angina pectoris from abrupt withdrawal of propranolol.

Excessive agonism will lead to a decrease in the number of receptors or in stimulus-response coupling. This is one cause of tachyphylaxis or tolerance, such as occurs to the bronchodilator effects of β-adrenoreceptor agonists. Abrupt withdrawal may result in poor residual function or in rebound effects, depending upon the type of effect caused by the agonist. Excessive agonism also may cause desensitization by agonist-induced changes in receptor conformation to inactive, slowly reconformable states.

Mechanism of Drug Action

Any metabolic or physiological function provides a potential mechanism of action of a drug. The term *mechanism of action* has been employed in a number of ways. In the past it was often the habit to confuse the site, or locus of action, with the mechanism of action. For example, the mechanism of the hypotensive action of tetraethylammonium ion originally was described as that of ganglionic blockade, which did nothing more than identify the anatomical structure upon which the drug acted. In a general sense, this was a partial elucidation of the mechanism of action, if mechanism is used in the mechanical sense of the entire linkage between the input and output of a machine. However, there has been a gradual narrowing of the definition of mechanism of action to be restricted to only the first event in the action–effect sequence, that is, only to the alteration of receptor function by the drug. In this sense, the mechanism of action of tetraethylammonium is defined more appropriately as that of competition with acetylcholine for nicotine cholinergic receptors on the postsynaptic ganglion cell membrane, even though the alteration in receptor function is not defined. The ultimate mechanism of action is known for only a few drugs.

It is customary to speak of a drug as a stimulant or a depressant, of the action as being excitatory or inhibitory, etc. Such terms describe only the effect and not the action, and they have no bearing upon whether the drug augments receptor function or diminishes it. In biological systems, positive and negative modulation and feedback occur at every level, the organ as well as the subcellular. Thus, an agonist to a negative modulator may be able to bring about the same effect as an antagonist to a positive modulator. It is possible for an antagonist or inhibitor to elicit an excitatory effect. An example is the convulsant action of strychnine, which results from its antagonism of glycine, an important mediator of postsynaptic inhibition in the central nervous system. Conversely, it is possible for an agonist to elicit an inhibitory effect. An example is the reflex bradycardia that results from the stimulant action of veratrum alkaloids on chemoreceptors in the left ventricle.

Because of the central role *enzymes play* in cellular function, it is not surprising that thoughts about the mechanism of action of drugs has focused largely upon enzymes. Agonist drugs conceivably could serve as substrates, cofactors or activators. At the present time, no drug is known definitely to exert its action as a substrate or as a cofactor, exclusive of vitamins and known nutrients. However, at least three classes of drugs are known and several are suspected to work through the activation of enzymes.

The most notable example of enzyme activation is that of epinephrine and similar β-adrenoreceptor agonists, which activate adenyl cyclase to increase the production of 3′,5′-cyclic adenylic acid (cyclic AMP; cAMP). The metabolic and cardiac effects of catecholamines are attributable, in part, to the increment in cAMP. One modulator of adenyl cyclase is the β-adrenergic receptor. The β-adrenoreceptor is coupled to adenylate cyclase through a regulatory protein that binds GDP and GTP (G-protein). When GDP is present, the agonist-receptor complex is associated with the regulatory protein. GTP causes transfer of the regulatory protein to adenylate cyclase and dissociation of the β-adrenoreceptor. Glucagon also owes its hyperglycemic action to activation of hepatic adenylate cyclase. A number of other agonists also activate adenylate cyclase. There is, thus, the interesting phenomenon of one enzyme, adenyl cyclase, being activated by numerous chemically unrelated drugs. Since β-adrenergic-blocking agents do not antagonize glucagon, it is obvious that glucagon works upon a different receptor than does epinephrine.

Thus, cAMP activates protein kinases that increase the activity of phosphorylase, actomyosin, the sequestration of calcium by the sarcoplasmic reticulum and calcium channels. Therefore, a brief activation of the β-adrenoreceptor sets in motion a cascade of events that greatly amplify the signal. Kinases also participate in down-regulation and desensitization.

Other important enzymes coupled to receptors are guanylate cyclase and phospholipases A and C, which are involved with membrane fluidity and calcium channels, respectively.

Many drugs are inhibitors of enzymes. When the drug is a *competitive inhibitor* of a natural endogenous substrate of the enzyme, it is called an *antimetabolite* (see also page 431). Examples of antimetabolites are sulfonamides, which compete with para-aminobenzoic acid and, thus, interfere with its incorporation into dihydrofolic acid and methotrexate, which competes with folic acid for dihydrofolate reductase and, thus, interferes with the formation of folinic acid. It might seem that anticholinesterases are also antimetabolites, although they are never placed into that classification. The reason is that the products of cholinesterase–acetylcholine interaction do not subserve important metabolic functions, as do folic and folinic acids, so that the organism is not deprived of an important metabolite by the action of the cholinesterase inhibitors.

Some drugs are competitive inhibitors of enzyme systems whose natural function appears not to produce useful metabolites but to rid the body of foreign substances. Inhibitors of the hepatic microsomes and probenecid fall into this category; the hepatic microsomes do perform a few biotransformations on endogenous substrates, but the renal tubular anion transport system does not appear to be required to eliminate any important endogenous substances.

Since neither the hepatic microsomes nor the tubular anion transport system seems to be involved in response systems, inhibitors of these enzyme systems are antagonists without corresponding agonists. Indeed, even natural endogenous substrates of enzymes are rarely considered to be agonists.

Noncompetitive enzyme inhibitors among drugs also are known. Examples are cyanide, fluoride, disulfiram and cardiac glycosides. When enzyme inhibition brings about a positive response—eg, the cholinergic effects of the anticholinesterases or the effects of diazoxide consequent to inhibition of phosphodiesterase—the drug appears to be an agonist. Yet, there can be no competitive antagonist to such an inhibitor, since the competitor to the drug is more substrate, to which the effect of the drug is actually attributable.

Acetylcholine increases the permeability of the subsynaptic membrane to cations and the heart muscle membrane to potassium. The mechanism is thought generally to involve a change in conformation of a protein constituent of the potassium channel, so that pore size or permeability constant is affected. The muscarinic receptor is coupled to the potassium channel through a G-protein. Other autonomic agonists also are known to alter the permeability to ions, in part through activation of adenyl cyclase, guanyl cyclase, phospholipase-c or other enzymes. Many drugs and toxins act through *alterations in the structural and physical properties of membranes*. To the extent that some of such substances may disperse themselves generally throughout the lipid phase of the membrane rather than to combine with special chemical entities, no definite receptors for such drugs can be said to exist.

The mechanism of action of certain drugs, especially autonomic drugs, often is stated to be *mimicry* of a natural neurohumor or hormone. Thus, methacholine mimics acetylcholine as an agonist. This does not define the mechanism of action, unless the mechanism of action of the natural substance is known.

Mimicry usually occurs because of a structural similarity between the natural substance and the mimetic drug. Mimicry in agonist functions is easy to demonstrate, but the site of action may not always be mimicry of the natural agonist at its receptor but rather at an allosteric site on a receptor or at its storage site to *release* the natural agonist.

Examples of mimetics that act by release of the natural mediator are indirectly acting sympathomimetics such as *d*-amphetamine, mephentermine, ephedrine (in part), tyramine and others, which are now known to act by displacing norepinephrine from storage sites within the adrenergic neuron. Many of such indirectly acting sympathomimetics lack a direct action on the adrenergic receptor, although some, like ephedrine, act both upon the receptor and the storage complex. Another mimetic by a release mechanism is carbachol, which promotes the presynaptic discharge of acetylcholine.

In these examples, there is a close structural similarity between the mimetic and the released mediator. In the case of many releasers of histamine (such as tubocurarine, polymyxin or morphine), no close chemical relationship exists between the releaser and the released. In such instances, release has been explained by activation of receptors on the mast-cell membrane which promote exocytosis of the histamine-containing granules, by an influx of calcium and activation of microtubules, all of which may be involved in moving the granules out of the mast cell.

Structural similarity also may aid mimicry by promoting chemical combination with an enzyme of destruction or some other means of disposition. For example, metaraminol, amphetamine, etc *inhibit membrane transport* into the neuron and, hence, inhibit the neuronal recapture of released norepinephrine. Consequently, the extraneuronal concentration of norepinephrine in the nearby region of the receptors does not drop as rapidly as in the absence of the mimetic, and the action of the mediator is sustained.

Some inhibitors of the enzymes of the destruction of mediators are structurally similar enough to the mediator to have some agonist action. This is true of neostigmine, which has a direct stimulant action on nicotinic receptors in addition to its anticholinesterase action. In contrast, the anticholinesterase, physostigmine, has some antagonist actions on cholinergic receptors and also an effect to interfere with acetylcholine synthesis.

The above multiple actions come about because all the structures that interact with a small molecule mediator (the receptor, synthesizing enzyme, destructive enzyme, storage molecule, membrane transport carrier) must have some common structural features and affinities. A drug that re-

acts with one of these molecules has a distinct probability of interacting with another.

The recognition of the critical role of *ions* in the function of membranes, the excitability of cells and the activity of many enzymes has generated a renewed interest in ions in the mechanism of action of certain drugs. The inorganic ions, some of which are used as drugs, lend themselves automatically to a discussion of ionic mechanisms. The repair of electrolyte deficiencies by replacement therapy warrants no further comment here. Some nonphysiological ions act as imperfect impersonators of physiological ions; lithium partly substitutes for sodium, bromide for chloride and thiocyanate for iodide, and each may owe its pharmacological action, in part, to a sluggish mobility through membrane channels, through which their sister ions normally pass readily when traffic is not impeded by "slowly moving vehicles." Iodide has an effect to increase the penetrance of drugs into caseous and necrotic areas, to aid in the resolution of gummatous lesions, to reduce the viscosity of mucous secretions and other odd effects; it is thought to do so by increasing the hydration of collagen and mucoproteins by a poorly understood mechanism. The transition elements and heavy metals have in common the ability to form complexes with a variety of physiologically active substances, particularly the active centers of many enzymes. *Chelation* and other types of *complexation* are the mechanisms of action of several drugs used to treat heavy-metal intoxication, diseases that involve abnormal body burdens or plasma levels of heavy metals and hypercalcemia. Chelates and chelation are discussed in more detail in Chapter 14.

There is much interest in the effects of drugs on ion movements. Cardiac glycosides are known to inhibit an ATPase involved in the membrane transport of sodium and several other substances, which indirectly causes an increase in intracellular calcium content. In part, the mechanisms of action of local anesthetics, quinidine and various other drugs also are speculated to involve calcium movements. In the past decade there has appeared a whole new class of drugs, the calcium channel blockers.

Concomitant with the development of molecular biology was the appreciation that drugs act through *nuclear* and *extranuclear genetic mechanisms*. Nitrogen mustards have long been known to interfere with the replication of DNA. Streptomycin, kanamycin, neomycin and gentamicin cause misreading by the ribosomes of the code incorporated into messenger RNA; tetracyclines, erythromycin and chloramphenicol inhibit the synthesis of protein at the ribosomes; and chloroquine, novobiocin and colchicine inhibit DNA polymerase. Other drugs induce the production of enzymes; aldosterone appears to act by inducing the synthesis of the enzyme, membrane ATPase, necessary to sodium transport. In general, steroid hormones combine with a cytosolic receptor, the complex of which is processed and translocated to the chromatin, where gene expression is altered. Many drugs induce one or more of the hepatic and extrahepatic cytochrome P-450 enzymes.

A number of drugs have simple mechanisms that do not involve an action at the cellular level. Examples are bulk and saline cathartics, osmotic diuretics and cholestyramine. Although such drugs usually do not generate much excitement among pharmacologists, they do serve as a reminder of the many avenues through which a mechanism of action may be expressed. Throughout the various chapters of Part 6, specific mechanisms of action may be mentioned.

Absorption, Distribution and Excretion

No matter by which route a drug is administered it must pass through several to many biological membranes during the processes of absorption, distribution, biotransformation and elimination. Since membranes are traversed in all of these events, the subject of this section will begin with a brief description of biological membranes and membrane processes and the relationship of the physiochemical properties of a drug molecule to penetrance and transport.

Structure and Properties of Membranes

The concept that a membrane surrounds each cell arose shortly after the cellular nature of tissue was discovered. The biological and physiochemical properties of cells seemed in accord with this view. In the past, from time to time, the actual existence of the membrane has been questioned by brilliant men, and ingenious explanations have been advanced to explain cellular integrity and the osmotic and electrophysiological properties of cells. Microchemical, x-ray diffraction, electron microscopic, nuclear magnetic resonance, electron spin resonance and other investigations have proved both the existence and nature of the plasma, mitochondrial, nuclear and other cell membranes. The description of the plasma membrane that follows is much oversimplified, but it will suffice to provide a background for an understanding of penetrance into and through membranes.

Structure and Composition—The cell membrane has been described as a "mayonnaise sandwich," in which a bimolecular layer of lipid material is entrained between two parallel monomolecular layers of protein. However, the protein does not make continuous layers, like the bread in a sandwich, but rather is sporadically scattered over the surfaces, like icebergs; ie, much of the protein is below the surface. In Fig 35-8 the lipid layers are represented as a somewhat orderly, closely packed lamellar array of phospholipid molecules associated tail-to-tail, each "tail" being an alkyl chain or steroid group and the "heads" being polar groups, including the glycerate moieties, with their polar ether and carbonyl oxygens and phosphate with attached polar groups. In reality, the lamellar portion is probably not

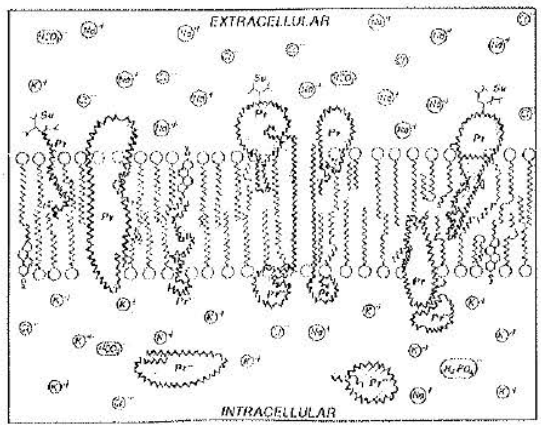

Fig 35-8. Simplified cross section of a cell membrane (components are not to scale). The lipid interior of the lamellar portion of the membrane consists of various phospholipids, fatty acids, cholesterol and other steroids. Ions are indicated in order to illustrate differences in size relative to the channel. *Pr:* protein; *Su:* sugar.

so orderly, since its composition is quite complex. Chains of fatty acids of different degrees of saturation and cholesterol cannot array themselves in simple parallel arrangements. Furthermore, the polar heads will assume a number of orientations depending upon the substances and groups involved. Moreover, the lamellar portion is penetrated by large globular proteins, the interior of which, like the lipid layers, has a high hydrophobicity, and some fibrous proteins.

The plasma membrane appears to be asymmetrical. The lipid composition varies from cell type to cell type and perhaps from site to site on the same membrane. There are, for example, differences between the membrane of the endoplasmic reticulum and the plasma membrane, even though the membranes are coextensive. Where membranes are double, the inner and outer layers may differ considerably; the inner and outer membranes of mitochondria have been shown to have strikingly different compositions and properties. Some authorities have expressed doubt as to the existence of the protein layers in biological membranes, although the evidence is preponderantly in favor of at least an outer glycoprotein coat. Sugar moieties also are attached to the outer proteins; these sugar moieties are important to cellular and immunological recognition and adhesion and have other functions as well.

The cell membrane appears to be perforated by water-filled pores of various sizes, varying from about 4 to 10 Å, the majority of which are about 7 Å. Probably all major ion channels are through the large globular proteins that traverse the membrane. Through these pores pass inorganic ions and small organic molecules. Since sodium ions are more hydrated than potassium and chloride ions, they are larger and do not pass as freely through the pores as potassium and chloride. The vascular endothelium appears to have pores at least as large as 40 Å, but these seem to be interstitial passages rather than transmembrane pores. Lipid molecules small enough to pass through the pores may do so, but they have a higher probability of entering into the lipid layer, from where they will equilibrate chemically with the interior of the cell. From work on monolayers, some researchers contend that it is not necessary to postulate pores to explain the permeability to water and small water-soluble molecules.

Stratum Corneum—Although the stratum corneum is not a membrane in the same sense as a cell membrane, it offers a barrier to diffusion, which is of significance in the topical application of drugs. The stratum corneum consists of several layers of dead keratinized cutaneous epithelial cells enmeshed in a matrix of keratin fibers and bound together with cementing desmosomes and penetrating tonofibrils of keratin. Varying amounts of lipids and fatty acids from dying cells, sebum and sweat are contained among the dead squamous cells. Immediately beneath the layer of dead cells and above the viable epidermal epithelial cells is a layer of keratohyaline granules and various water-soluble substances, such as alpha-amino acids, purines, monosaccharides and urea.

Both the upper and lower layers of the stratum corneum are involved in the cutaneous barrier to penetration. The barrier to penetration from the surface is in the upper layers for water-soluble substances and the lower layers for lipid-soluble substances, and the barrier to the outward movement of water is in the lowest layer.

Membrane Potentials—Across the cell membrane there exists an electrical potential, always negative on the inside and positive on the outside. If a cell did not have special-membrane electrolyte-transport processes, its membrane potential would be mainly the result of the Donnan equilibrium (see Chapter 14) consequent to the semipermeability of the membrane. Such potentials generally lie between 2 and 5 mv.

A cell with a membrane across which diffusible electrolyte distribution is purely passive would be expected to have a high internal concentration of sodium, such as is true for the erythrocytes of some species. However, the interior of most cells is high in potassium and low in sodium, as depicted in Fig 35-8. This unequal distribution of cations attests to special electrolyte-transport processes and to differential permeabilities of diffusible ions, so that the membrane potential is higher than that which would result from a purely passive Donnan distribution. In nerve tissue or skeletal and cardiac muscle, the membrane potential ranges upwards to about 90 mv. The electrical gradient is on the order of 50,000 v/cm, because of the extreme thinness of the membrane. Obviously, such an intense potential gradient will influence strongly the transmembrane passages of charged drug molecules.

Diffusion and Transport

Transport is the movement of a drug from one place to another within the body. The drug may diffuse freely in uncombined form with a kinetic energy appropriate to its thermal environment, or it may move in combination with extracellular or cellular constituents, sometimes in connection with energy-yielding processes that allow the molecule or complex to overcome barriers to simple diffusion.

Simple Nonionic Diffusion and Passive Transport—Molecules in solution move in a purely random fashion, provided they are not charged and moving in an electrical gradient. Such random movement is called *diffusion;* if the molecule is uncharged, it is called *nonionic diffusion.*

In a population of drug molecules, the probability that during unit time any drug molecule will move across a boundary is directly proportional to the number of molecules adjoining that boundary and, therefore, to the drug concentration. Except at dilutions so extreme that only a few molecules are present, the actual rate of movement (molecules/unit time) is directly proportional to the probability and, therefore, to the concentration. Once molecules have passed through the boundary to the opposite side, their random motion may cause some to return and others to continue to move further away from the boundary. The rate of return is likewise proportional to the concentration on the opposite side of the boundary. It follows that, although molecules are moving in both directions, there will he a net movement from the region of higher to that of lower concentration, and the net transfer will be proportional to the concentration differential. If the boundary is a membrane, which has both substance and dimension, the rate of movement is also directly proportional to the permeability and inversely proportional to the thickness. These factors combine into Fick's Law of Diffusion,

$$\frac{dQ}{dt} = \frac{\overline{D}A(C_1 - C_2)}{x} \qquad (5)$$

where Q is the net quantity of drug transferred across the membrane, t is time, C_1 is the concentration on one side and C_2 on the other, x is the thickness of the membrane, A is the area and $\overline{D}$ is the diffusion coefficient, related to permeability. The equation is more nearly correct if chemical activities are used instead of concentrations. Since a biological membrane is patchy, with pores of different sizes and probably with varying thickness and composition, both $\overline{D}$ and x probably vary from spot to spot. Nevertheless, some mean values can be assumed.

It is customary to combine the membrane factors into a single constant, called a permeability constant or coefficient, P, so that $P = \overline{D}/x$, A in Eq 5 having unit value. The rate of net transport (diffusion) across the membrane then becomes

$$\frac{dQ}{dt} = P(C_1 - C_2) \quad (6)$$

As diffusion continues, C_1 approaches C_2, and the net rate, dQ/dt, approaches zero in exponential fashion characteristic of a first-order process. Equilibrium is defined as that state in which $C_1 = C_2$. The equilibrium is, of course, dynamic, with equal numbers of molecules being transported in each direction during unit time. If water is also moving through the membrane, it may either facilitate the movement of drug or impede it, according to the relative directions of movement of water and drug; this effect of water movement is called *solvent drag*.

Ionic or Electrochemical Diffusion—If a drug is ionized, the transport properties are modified. The probability of penetrating the membrane is still a function of concentration, but it is also a function of the potential difference or electrical gradient across the membrane. A cationic drug molecule will be repelled from the positive charge on the outside of the membrane, and only those molecules with a high kinetic energy will pass through the ion barrier. If the cation is polyvalent, it may not penetrate at all.

Once inside the membrane, a cation simultaneously will be attracted to the negative charge on the intracellular surface of the membrane and repelled by the outer surface; it is said to be moving along the *electrical gradient*. If it also is moving from a higher towards a lower concentration, it is said to be moving along its *electrochemical gradient*, which is the sum of the influences of the electrical field and the concentration differential across the membrane.

Once inside the cell, cations will tend to be kept inside by the attractive negative charge on the interior of the cell, and the intracellular concentration of drug will increase until, by sheer numbers of accumulated drug particles, the outward diffusion or mass escape rate equals the inward transport rate, and electrochemical equilibrium is said to have occurred. At electrochemical equilibrium at body temperature (37°C), ionized drug molecules will be distributed according to the Nernst equation,

$$\pm \log \frac{C_o}{C_i} = \frac{ZE}{61} \quad (7)$$

where C_o is the molar extracellular and C_i the intracellular concentration, Z is the number of charges per molecule and E is the membrane potential in millivolts. Log C_o/C_i is positive when the molecule is negatively charged and negative when the molecule is positively charged.

Facilitated Diffusion—Sometimes a substance moves more rapidly through a biological membrane than can be accounted for by the process of simple diffusion. This accelerated movement is termed *facilitated diffusion*. It is thought to be due to the presence of a special molecule within the membrane, called a *carrier*, with which the transported substance combines. There is considered to be a greater permeability to the carrier–drug complex than to the drug alone, so that the transport rate is enhanced. After the complex traverses the membrane, it dissociates. The carrier must either return to the original side of the membrane to be reused or constantly be produced on one side and eliminated on the other in order for the carrier process to be continuous. Many characteristics of facilitated diffusion, formerly attributed to ion carriers, can be explained by ion exchange. Although facilitated diffusion resembles active transport, below, in its dependence upon a continuous source of energy, it differs in that facilitated diffusion will only transport a molecule along its electrochemical gradient.

Active Transport—Active transport may be defined as energy-dependent movement of a substance through a biological membrane against an electrochemical gradient. It is characterized by the following:

1. The substance is transported from a region of lower to one of higher electrochemical activity.
2. Metabolic poisons interfere with transport.
3. The transport rate approaches an asymptote (ie, saturates) as concentration increases.
4. The transport system usually shows a requirement for specific chemical structures.
5. Closely related chemicals are competitive for the transport system.

Many drugs are secreted from the renal tubules into urine, from liver cells into bile or from the cerebrospinal fluid into blood by active transport, but the role of active transport of drugs in the distribution into most body compartments and tissues is less well known. Active transport is required for the penetrance of a number of sympathomimetics into neural tissue and for the movement of several anticancer drugs across cell membranes.

Pinocytosis and Exocytosis—Many, perhaps all, cells are capable of a type of phagocytosis called *pinocytosis*. The cell membrane has been observed to invaginate into a saccular structure containing extracellular materials and then pinch off the saccule at the membrane, so that the saccule remains as a vesicle or vacuole within the interior of the cell. Since metabolic activity is required and since an extracellular substance may be transported against an electrochemical gradient, pinocytosis shows some of the same characteristics as active transport. However, pinocytosis is relatively slow and inefficient compared to most active transport, except in gastrointestinal absorption, in which pinocytosis is of considerable importance.

It is not known to what extent pinocytosis contributes to the transport of most drugs, but many macromolecules and even larger particles can be absorbed by the gut. Pinocytosis probably explains the oral efficacy of the Sabin polio vaccine. Some drugs themselves affect pinocytosis; eg, adrenal glucocorticoids markedly inhibit the process in macrophages and other cells involved in inflammation.

Exocytosis is more or less the reverse of pinocytosis. Granules, vacuoles or other organelles within the cell move to the cell membrane, fuse with it and extrude their contents into the interstitial space.

Physicochemical Factors in Penetrance

Drugs and other substances may traverse the membrane primarily either through the pores or by dissociation into the membrane lipids and subsequent diffusion from the membrane into the cytosol or other fluid on the far side of the membrane. The physicochemical prerequisites are different according to which route is taken. To pass through the pores, the "diameter" of the molecule must be smaller than the pore, but the molecule can be longer than the pore diameter. The probability that a long, thin molecule will be oriented properly is low, unless there is also bulk flow and the transmembrane passage of large molecules is slow.

Water-soluble molecules with low lipid solubility usually are thought to pass through the membrane mainly via the pores and, to a small extent, by pinocytosis, but recent work with lipid monolayers suggests that small water-soluble molecules also may be able to pass readily through the lipid, and the necessity of postulating the existence of pores has been questioned. Nevertheless, experimental data on penetrance overwhelmingly favor the concept of passage of water-soluble lipid-insoluble substances through pores. If there is a membrane carrier or active transport system, a low solubility of the drug in membrane lipids is no impediment to penetration, since the drug-carrier complex is assumed to have an appropriate solubility and energy from an active transport system enables the drug to penetrate the energy barrier "imposed by the lipids." Actually, the lipids are not an important energy barrier; rather, the barrier is the force

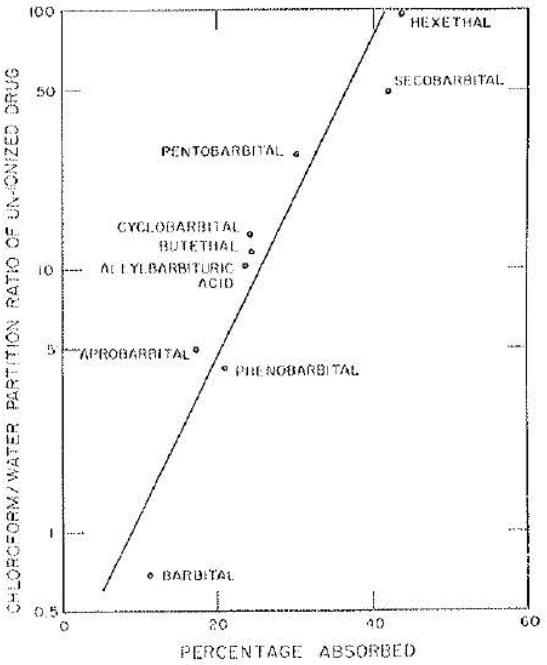

Fig 35-9. The relationship of absorption of the un-ionized forms of drugs from the colon of the rat to the chloroform:water partition coefficient (courtesy, Schanker[6]).

of attraction of the solvent water for its dipolar to polar solute, so that it is difficult for the solute to leave the water and enter the lipid.

Drugs with a high solubility in the membrane lipids, of course, pass easily through the membrane. Even when their dimensions are small enough to permit passage through pores, lipid-soluble drugs primarily pass through the membrane lipids, not only because chemical partition favors the lipid phase but also because the surface area occupied by pores is only a small fraction of the total membrane area.

Lipid Solubility and Partition Coefficients—As early as 1902, Overton investigated the importance of lipid solubility to the penetrance and absorption of drugs. Eventually, it was recognized that more important than lipid solubility was the lipid–water distribution coefficient; ie, a high lipid solubility does not favor penetrance unless the water solubility is low enough so that the drug is not entrained in the aqueous phase.

In Fig 35-9 is illustrated the relationship between the chloroform–water partition coefficient and the colonic absorption of barbiturates. Chloroform probably is not the optimal lipid solvent for such a study, and natural lipids from nerve or other tissues have been shown to be superior in the few instances in which they have been employed. Nevertheless, the correlation shown in the figure is a convincing one.

When the water solubility of a substance is so low that a significant concentration in water or extracellular fluid cannot be achieved, absorption may be negligible in spite of a favorable partition coefficient. Hence, mineral oil, petrolatum, etc are virtually unabsorbed. The optimal partition coefficient for permeation of the skin appears to be lower than that for the permeation of the cell membrane, perhaps being as low as one.

Dipolarity, Polarity and Nonionic Diffusion—The partition coefficient of a drug depends upon the polarity and the size of the molecule. Drugs with a high dipole moment, even though un-ionized, have a low lipid solubility and, hence, poor penetrance. An example of a highly dipolar substance with a low partition coefficient, which does not penetrate into cells, is sulfisoxazole. Sulfadiazine is somewhat less dipolar, has a chloroform–water partition coefficient ten times that of sulfisoxazole and readily penetrates cells. Ionization not only diminishes lipid solubility greatly but also may impede passage through charged membranes (see *Ionic Diffusion*, page 709).

It is often stated that ionized molecules do not penetrate membranes, except for ions of small diameter. This is not necessarily true, because of the presence of membrane carriers for some ions, which effectively may shield or neutralize the charge (ion-pair formation). The renal tubular transport systems, which transport such obligate ions as tetraethylammonium, probably form ion-pairs. Furthermore, if an ionized molecule has a large nonpolar moiety such that an appreciable lipid solubility is imparted to the molecule in spite of the charge, the drug may penetrate, although usually at a slow rate. For example, various morphinan derivatives are absorbed passively from the stomach even though they

Table I—Rates of Entry of Drugs in CSF and the Degrees of Ionization of Drugs at pH 7.4[7]

Drug	% binding to plasma protein	pK_a[a]	% un-ionized at pH 7.4	Permeability constant (P^2 min^{-1}) ± S.E.
		Drugs mainly ionized at pH 7.4		
5-Sulfosalicylic acid	22	(strong)	0	<0.0001
N-Methylnicotinamide	<10	(strong)	0	0.0005 ± 0.00006
5-Nitrosalicylic acid	42	2.3	0.001	0.001 ± 0.0001
Salicylic acid	40	3.0	0.004	0.006 ± 0.0004
Mecamylamine	20	11.2	0.016	0.021 ± 0.0016
Quinine	76	8.4	9.09	0.078 ± 0.0061
		Drugs mainly un-ionized at pH 7.4		
Barbital	<2	7.5	55.7	0.026 ± 0.0022
Thiopental	75	7.6	61.3	0.50 ± 0.051
Pentobarbital	40	8.1	83.4	0.17 ± 0.014
Aminopyrine	20	5.0	99.6	0.25 ± 0.020
Aniline	15	4.6	99.8	0.40 ± 0.042
Sulfaguanidine	6	>10.0[b]	>99.8	0.003 ± 0.0002
Antipyrine	8	1.4	>99.9	0.12 ± 0.013
N-Acetyl-4-aminoantipyrine	<3	0.5	>99.9	0.012 ± 0.0010

[a] The dissociation constant of both acids and bases is expressed as a pK_a—a negative logarithm of the acidic dissociation constant.
[b] Sulfaguanidine has a very weakly acidic group (pK_a > 10) and two very weakly basic groups (pK_a 2.75 and 0.5). Consequently, the compound is almost completely undissociated at pH 7.4.

are ionized completely at the pH of gastric fluid. Nevertheless, when a drug is a weak acid or base, the un-ionized form, with a favorable partition coefficient, passes through a biological membrane so much more readily than the ionized form that, for all practical purposes, only the un-ionized form is said to pass through the membrane. This has become known as the *principle of nonionic diffusion*.

This principle is the reason that only the concentrations of the un-ionized form of the barbiturates are plotted in Fig 35-9.

For the purpose of further illustrating the principle, Table I is provided. In the table, the permeability constants for penetrance into the cerebral spinal fluid of rats are higher for un-ionized drugs than for ionized ones. The apparent exceptions—barbital, sulfaguanidine and acetylaminoantipyrine—may be explained by the dipolarity of the un-ionized molecules. With barbital, the two lipophilic ethyl groups are too small to compensate for the considerable dipolarity of the un-ionized barbituric acid ring; also it may be seen that barbital is appreciably ionized, which contributes to the relatively small permeability constant. Sulfaguanidine and acetylaminoantipyrine are both very polar molecules. Mecamylamine also might be considered an exception, since it shows a modest permeability even though strongly ionized; there is no dipolarity in mecamylamine except in the amino group.

Absorption of Drugs

Absorption is the process of movement of a drug from the site of application into the extracellular compartment of the body. Inasmuch as there is a great similarity among the various membranes that a drug may pass through in order to gain access to the extracellular fluid, it might be expected that the particular site of application (or *route*) would make little difference to the successful absorption of the drug. In actual fact, it makes a great deal of difference; many factors, other than the structure and composition of the membrane, determine the ease with which a drug is absorbed. These factors are discussed in the following sections, along with an account of the ways that drug formulations may be manipulated to alter the ability of a drug to be absorbed readily.

Routes of Administration

Drugs may be administered by many different routes. The various routes include oral, rectal, sublingual or buccal, parenteral, inhalation and topical. The choice of a route depends upon both convenience and necessity.

Oral Route—This is obviously the most convenient route for access to the systemic circulation, providing that various factors do not militate against this route. Oral administration does not always give rise to sufficiently high plasma concentrations to be effective; some drugs are absorbed unpredictably or erratically; patients occasionally have an absorption malfunction. Drugs may not be given by mouth to patients with gastrointestinal intolerance, or who are in preparation for anesthesia or who have had gastrointestinal surgery. Oral administration also is precluded in coma.

Rectal Route—Drugs that ordinarily are administered by the oral route usually can be administered by injection or by the alternative *lower enteral* route, through the anal portal into the rectum or lower intestine. With regard to the latter, *rectal suppositories* or *retention enemas* formerly were used quite frequently, but their popularity has abated somewhat, owing to improvements in parenteral preparations. Nevertheless, they continue to be valid and, sometimes, very important ways of administering a drug, especially in pediatrics and geriatrics. In Fig 35-10 the availability of a drug by retention enema may be compared with that by the intravenous and oral route and rectal suppository administration. It is apparent that the retention enema may be a very satisfactory means of administration but that rectal suppositories may be inadequate where rapid absorption and high plasma levels are required. The illustration is not intended to lead the reader to the conclusion that a retention enema always will give more prompt and higher blood levels than the oral route, for converse findings for the same drug have been reported,[21] but, rather, to show that the retention enema may offer a useful substitute for the oral route.

Sublingual or Buccal Route—Even though an adequate plasma concentration eventually may be achievable by the oral route, it may rise much too slowly for use in some situations where a rapid response is desired. In such situations parenteral therapy usually is indicated. However, the patients with angina pectoris may get quite prompt relief from an acute attack by the *sublingual* or *buccal* administration of nitroglycerin, so that parenteral administration may be avoided. When only small amounts of drugs are required to gain access to the blood, the buccal route may be very satisfactory, providing the physicochemical prerequisites for absorption by this route are present in the drug and dosage form. Only a few drugs may be given successfully by this route.

Parenteral Routes—These routes, by definition, include any route other than the oral-gastrointestinal (enteral) tract, but in common medical usage the term excludes topical administration and includes only various hypodermic routes. Parenteral administration includes the intravenous, intramuscular and subcutaneous routes. Parenteral routes may be employed whenever enteral routes are contraindicated (see above) or inadequate.

The *intravenous* route may be preferred on occasion, even when a drug may be well-absorbed by the oral route. There is no delay imposed by absorption before the administered drug reaches the circulation, and blood levels rise virtually

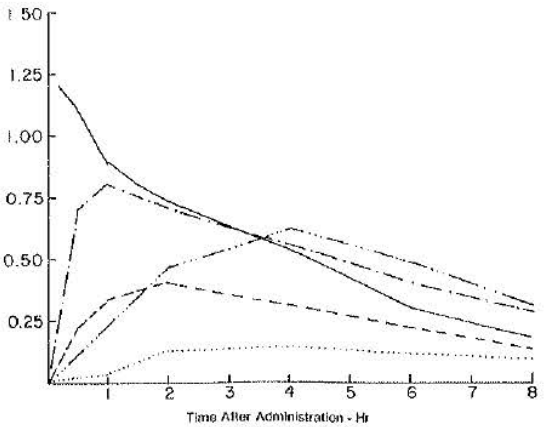

Fig 35-10. Blood concentration in mg/100 mL of theophylline (ordinates) following administration to humans of aminophylline in the amounts and by the routes indicated. Doses: per 70 kg. Theophylline-ethylenediamine by various routes: —— *intravenous, 0.5 g;* – • – • – *retention enema, 0.5 g;* — ··· — ··· — *oral tablets-Pl., 0.5 g;* - - - *oral tablets-Pl., 0.3 g;* ···· *rectal suppository, 0.5 g* (courtesy, Truitt, *et al*,[20] adapted).

as rapidly as the time necessary to empty the syringe or infusion bottle. Consequently, the intravenous route is the preferred route when an emergency calls for an immediate response.

In addition to the rapid rise in plasma concentration of drug, another advantage of intravenous administration is the greater predictability of the peak plasma concentration, which, with some drugs, can be calculated with a fair degree of precision. Smaller doses generally are required by the intravenous than by other routes, but this usually affords no advantage, inasmuch as the sterile injectable dosage form costs more than enteric preparations and the requirements for medical or paramedical supervision of administration also may add to the cost and inconvenience.

Because of the rapidity with which drug enters the circulation, dangerous side effects to the drug may occur which are often not extant by other routes. The principal untoward effect is a depression of cardiovascular function, which is often called *drug shock*. Consequently, some drugs must be given quite slowly to avoid vasculotoxic concentrations of drug in the plasma. Acute, serious, allergic responses also are more likely to occur by the intravenous route than by other routes.

Many drugs are too irritant to be given by the oral, intramuscular or subcutaneous route and must, of necessity, be given intravenously. However, such drugs also may cause damage to the veins (phlebitis) or, if extravasated, cause necrosis (slough) around the injection site. Consequently, such irritant drugs may be diluted in isotonic solutions of saline, dextrose or other media and given by slow infusion, providing that the slower rate of delivery does not negate the purpose of the administration in emergency situations.

Absorption by the *intramuscular* route is relatively fast and this parenteral route may be used where an immediate effect is not required but a prompt effect is desirable. Intramuscular deposition also may be made of certain repository preparations, rapid absorption not being desired. Absorption from an intramuscular depot is more predictable and uniform than from a subcutaneous site.

Irritation around the injection site is a frequent accompaniment of intramuscular injection, depending upon the drug and other ingredients. Because of the dangers of accidental intravenous injection, medical supervision generally is required. Sterilization is necessary.

In *subcutaneous* administration the drug is injected into the alveolar connective tissue just below the skin. Absorption is slower than by the intramuscular route but, nevertheless, may be prompt with many drugs. Often, however, absorption by this route may be no faster than by the oral route. Therefore, when a fairly prompt response is desired with some drugs, the subcutaneous route may not offer much advantage over the oral route, unless for some reason the drug cannot be given orally.

The slower rate of absorption by the subcutaneous route is usually the reason why the route is chosen, and the drugs given by this route are usually those in which it is desired to spread the action out over a number of hours, in order to avoid either too intense a response, too short a response or frequent injections. Examples of drugs given by this route are insulin and sodium heparin, neither of which is absorbed orally and both of which should be absorbed slowly over many hours. In the treatment of asthma, epinephrine usually is given subcutaneously to avoid the dangers of rapid absorption and consequent dangerous cardiovascular effects. Many repository preparations, including tablets or pellets, are given subcutaneously. As with other parenteral routes, irritation may occur. Sterile preparations also are required. However, medical supervision is not required always and self-administration by this route is customary with certain drugs, such as insulin.

Intradermal injection, in which the drug is injected into, rather than below the dermis, is rarely employed, except in certain diagnostic and test procedures, such as screening for allergic or local irritant responses.

Occasionally, even by the intravenous route, it is not possible, practical or safe to achieve plasma concentrations high enough so that an adequate amount of drug penetrates into special compartments, such as the cerebrospinal fluid, or various cavities, such as the pleural cavity. The brain is especially difficult to penetrate with water-soluble drugs. The name *blood-brain barrier* is applied to the impediment to penetration. When drugs do penetrate, the choroid plexus often secretes them back into the blood very rapidly, so that adequate levels of drugs in the cerebrospinal fluid may be difficult to achieve. Consequently, *intrathecal** or *intraventricular* administration may be indicated.

Body cavities such as the pleural cavity normally are wetted by a small amount of effusate which is in diffusion equilibrium with the blood and, hence, is accessible to drugs. However, infections and inflammations may cause the cavity to fill with serofibrinous exudate which is too large to be in rapid diffusion equilibrium with the blood. *Intracavitary* administration, thus, may be required. It is extremely important that sterile and nonirritating preparations be used for intrathecal or intracavitary administration.

Inhalation Route—Inhalation may be employed for delivering gaseous or volatile substances into the systemic circulation, as with most general anesthetics. Absorption is virtually as rapid as the drug can be delivered into the alveoli of the lungs, since the alveolar and vascular epithelial membranes are quite permeable, blood flow is abundant and there is a very large surface for absorption.

Aerosols of nonvolatile substances also may be administered by inhalation, but the route is used infrequently for delivery into the systemic circulation because of various factors which contribute to erratic or difficult-to-achieve blood levels. Whether or not an aerosol reaches and is retained in pulmonary alveoli depends critically upon particle size. Particles greater than 1 μm in diameter tend to settle in the bronchioles and bronchi, whereas particles less than 0.5 μm fail to settle and mainly are exhaled. Aerosols are employed mostly when the purpose of administration is an action of the drug upon the respiratory tract itself. An example of a drug commonly given as an aerosol is isoproterenol, which is employed to relax the bronchioles during an asthma attack.

Topical Route—Topical administration is employed to deliver a drug at, or immediately beneath, the point of application. Although occasionally enough drug is absorbed into the systemic circulation to cause systemic effects, absorption is too erratic for the topical route to be used routinely for systemic therapy. However, various transdermal preparations of nitroglycerin and clonidine are employed quite successfully for systemic use. Some work with aprotic solvent vehicles such as dimethyl sulfoxide (DMSO) also has generated interest in topical administration for systemic effects. A large number of topical medicaments are applied to the skin, although topical drugs are also applied to the eye, nose and throat, ear, vagina, etc.

In man, percutaneous absorption probably occurs mainly from the surface. Absorption through the hair follicles occurs, but the follicles in man occupy too small a portion of the total integument to be of primary importance. Absorption through sweat and sebaceous glands generally appears to be minor. When the medicament is rubbed on vigorously,

* Intrathecal administration denotes administration into the cerebrospinal fluid at any level of the cerebrospinal axis, including injection into the cerebral ventricles, which is the most common mode of intrathecal administration.

the amount of the preparation that is forced into the hair follicles and glands is increased. Rubbing also forces some material through the stratum corneum without molecular dispersion and diffusion through the barrier. Rather large particles of substances such as sulfur have been demonstrated to pass intact through the stratum corneum. When the skin is diseased or abraded, the cutaneous barrier may be disrupted or defective, so that percutaneous absorption may be increased. Since much of a drug that is absorbed through the epidermis diffuses into the circulation without reaching a high concentration in some portions of the dermis, systemic administration may be preferred in lieu of, or in addition to, topical administration.

Factors That Affect Absorption

In addition to the physicochemical properties of drug molecules and biological membranes, various factors affect the rate of absorption and determine, in part, the choice of route of administration.

Concentration—It is self-evident that the concentration, or, more exactly, the thermodynamic activity, of a drug in a drug preparation will have an important bearing upon the rate of absorption, since the rate of diffusion of a drug away from the site of administration is directly proportional to the concentration. Thus, a 2% solution of lidocaine will induce local anesthesia more rapidly than a 0.2% solution. However, drugs administered in solid form are not absorbed necessarily at the maximal rate (see *Physical State of Formulation and Dissolution Rate*, below).

After oral administration the concentration of drugs in the gut is a function of the dose, but the relationship is not necessarily linear. Drugs with a low aqueous solubility (eg, digitoxin) quickly saturate the gastrointestinal fluids, so that the rate of absorption tends to reach a limit as the dose is increased. The peptizing and solubilizing effects of bile and other constituents of the gastrointestinal contents assist in increasing the rate of absorption but are in themselves somewhat erratic. Furthermore, many drugs affect the rates of gastric, biliary and small intestinal secretion, which causes further deviations from a linear relationship between concentration and dose.

Drugs that are administered subcutaneously or intramuscularly also may not show always a direct linear relationship between the rate of absorption and the concentration of drug in the applied solution, because osmotic effects may cause dilution or concentration of the drug, if the movement of water or electrolytes is different from that of the drug. Whenever possible, drugs for hypodermic injection are prepared as isotonic solutions. Some drugs affect the local blood flow and capillary permeability, so that at the site of injection there may be a complex relationship of concentration achieved to the concentration administered.

Physical State of Formulation and Dissolution Rate—The rate of absorption of a drug may be affected greatly by the rate at which the drug is made available to the biological fluid at the site of administration. The intrinsic physicochemical properties, such as solubility and the thermodynamics of dissolution, are only some of the factors which affect the rate of dissolution of a drug from a solid form. Other factors include not only the unavoidable interactions among the various ingredients in a given formulation but also deliberate interventions to facilitate dispersion (eg, comminution, Chapter 75 and dissolution, Chapter 31) or retard it (eg, coatings, Chapter 90 and slow-release formulations, Chapter 91). There are also factors that affect the rate of delivery from liquid forms. For example, a drug in a highly viscous vehicle is absorbed more slowly from the vehicle than a drug in a vehicle of low viscosity; in oil-in-water emulsions, the rate depends upon the partition coefficient. These manipulations are the subject of biopharmaceutics (see Chapter 91).

Area of Absorbing Surface—The area of absorbing surface is an important determinant of the rate of absorption. To the extent that the therapist must work with the absorbing surfaces available in the body, the absorbing surface is not subject to manipulation. However, the extent to which the existing surfaces may be used is subject to variation. In those rare instances in which percutaneous absorption is intended for systemic administration, the entire skin surface is available.

Subsequent to subcutaneous or intramuscular injections, the site of application may be massaged in order to spread the injected fluid from a compact mass to a well-dispersed deposit. Alternatively, the dose may be divided into multiple small injections, although this recourse is generally undesirable.

The different areas for absorption afforded by the various routes account, in part, for differences in the rates of absorption by those routes. The large alveolar surface of the lungs allows for extremely rapid absorption of gases, vapors and properly aerosolized solutions; with some drugs the rate of absorption may be nearly as fast as intravenous injection. In the gut the small intestine is the site of the fastest, and hence most, absorption because of the small lumen and highly developed villi and microvilli; the stomach has a relatively small surface area, so that even most weak acids are absorbed predominately in the small intestine despite a pH partition factor that should favor absorption from the stomach (see *The pH Partition Principle*, page 716).

Vascularity and Blood Flow—Although the thermal velocity of a freely diffusible average drug molecule is on the order of meters per second, in solution the rate at which it will diffuse away from a reference point will be much slower. Collisions with water and/or other molecules, which cause a random motion, and the forces of attraction between the drug and water or other molecules slow the net mean velocity.

The time taken to traverse a given distance is a function of the square of the distance; on the average it would take about 0.01 sec for a net outward movement of 1 μm, 1 sec for 10 μm, 100 sec for 100 μm, etc. In a highly vascular tissue, such as skeletal muscle, in which there may be more than 1000 capillaries/sq mm of cross section, a drug molecule would not have to travel more than a few microns, hence, less than a second on the average, to reach a capillary from a point of extravascular injection.

Once the drug reaches the blood, diffusion is not important to transport and the rate of blood flow determines the movement. The velocity of blood flow in a capillary is about 1 mm/sec, which is 100 times faster than the mean net velocity of drug molecules 1 mm away from their injection site. The velocity of blood flow is even faster in the larger vessels. Less than a minute is required to distribute drug molecules from the capillaries at the injection site to the rest of the body.

From the above discussion it follows that absorption is most rapid in the vascular tissues. Drugs are absorbed more rapidly from intramuscular sites than from less vascular subcutaneous sites, etc. Despite the small absorbing surface for buccal or sublingual absorption, the high vascularity of the buccal, gingival and sublingual surfaces favors an unexpectedly high rate of absorption. Because of hyperemia, absorption will be faster from inflamed than from normal areas, unless the presence of edema lengthens the mean distance between capillaries and, thus, negates the effects of hypermia on absorption.

Vasoconstriction may have a profound effect upon the rate of absorption. When a local effect of a drug is desired,

as in local anesthesia, absorption away from the infiltered site may be impeded greatly by vasoconstrictors included in the preparation. Unwanted vasoconstriction sometimes may cause serious problems. For example, in World War II many wounded soldiers were given subcutaneous morphine without evident effect. As a result, injections were sometimes repeated more than once. When the patient was removed to the field hospital, toxic effects would occur suddenly. The explanation is that cold-induced vasoconstriction occurred in the field; when the patient was warmed in the hospital, vasodilation would result and the victim would be flooded with drug. Shock also contributes to the effect, since during shock the blood flow is diminished and there also may be a superimposed vasoconstriction; repair of the shock condition then facilitates absorption.

Extravascularly injected molecules too large to pass through the capillary endothelium will, of necessity, enter the systemic circulation through the lymph. Thus, the lymph flow may be important to the absorption of a few drugs.

Movement—A number of factors combine so that movement at the site of injection increases the rate of absorption. In the intestine, segmental movements and peristalsis aid in dividing and dispersing the drug mass. The continual mixing of the chyme helps keep the concentration maximal at the mucosal surface. The pressures developed during segmentation and peristalsis also may favor a small amount of filtration. Movement at the site of hypodermic injection also favors absorption, since it tends to force the injected material through the tissue, increasing the surface area of drug mass and decreasing the mean distance to the capillaries. Movement also increases the flow of blood and lymph. The selection of a site for intramuscular injection may be determined by the amount of expected movement, according to whether the preparation is intended as a fast-acting or a repository preparation.

Gastric Motility and Emptying—The motility of the stomach is more important to the rate at which an orally administered drug is passed on to the small intestine than it is to the rate of absorption from the stomach itself, since for various reasons noted above, absorption from the stomach is usually of minor importance.

The average emptying time of the unloaded stomach is about 40 min and the half-time is around 10 min, though it varies according to its contents, reflex and psychological factors, as well as the action of certain autonomic drugs or disease. The effect of food to delay absorption is due, in part, to its action to prolong emptying time. The emptying time causes a delay in the absorption of drug, which may be unfavorable or favorable according to what is desired. In the case of therapy with antacids, gastric emptying is a nuisance, since it removes the antacid from the stomach where it is needed.

Solubility and Binding—The dissolution of drugs of low solubility is generally a slow process. Indeed, low solubility is the result of a low rate of departure of drug molecules from the undispersed phase. Furthermore, since the concentration around the drug mass is low, the concentration gradient from the site of deposition to the plasma is small and the rate of diffusion is low, accordingly.

When it is desired that a drug have a prolonged action but not a high plasma concentration, a derivative of low solubility is often sought. The "insoluble" estolates and other esters of several steroids have durations of action of weeks because of the slow rates of absorption from the sites of injection. Insoluble salts or complexes of acidic or basic drugs also are employed as repository preparations; for example, the procaine salt of penicillin G has a low solubility and is used in a slow release form of the antibiotic.

The solubility of certain macromolecules is dependent critically on the ionization of substituent groups. When they are amphiprotic, they are least soluble at their isoelectric pH. Insulin is normally soluble at the pH of the extracellular fluid, but by combining insulin with the right proportion of a basic protein, such as protamine, the isoelectric pH can be made to be approximately 7.4 and the complex can be used as a low-solubility prolonged-action drug. For more details, see Chapter 91.

Some drugs may bind with natural substances at or near the site of application. The strongly ionized mucopolysaccharides in connective tissue, ground substance and mucous secretions of the gut are retardants to the absorption of a number of drugs, especially large cationic or polycationic molecules. In the gut, the binding is the least at low pH, which should favor absorption of large cations from the stomach; however, absorption from the stomach is slow (see above), so that the absorption of large cations occurs mainly in the upper duodenum where the pH is still relatively low. Pharmacologically inactive quaternary ammonium compounds sometimes are included in an oral preparation of a quaternary ammonium drug for the purpose of saturating the binding sites of mucin and other mucopolysaccharides and, thereby, enhancing the absorption of drug.

In addition to mucopolysaccharides in mucous secretions, food in the gastrointestinal track binds many drugs and slows absorption. Antacids, especially aluminum hydroxide plus other basic aluminum compounds and magnesium trisilicate, bind amine and ammonium drugs and interfere with absorption.

Donnan Effect—The presence of a charged macromolecule on one side of a semipermeable membrane (impermeable to the macromolecule) will alter the concentration of permeant ionized particles according to the Donnan equilibrium (page 716). Accordingly, drug molecules of the same charge as the macromolecule will be constrained to the opposite side of the membrane. The presence of appropriately charged macromolecules not only will influence the distribution of drug ions in accordance with the Donnan equation but also increase the rate of transfer of the drug across the membrane, because of mutual ionic repulsion. This effect is sometimes used to facilitate the absorption of ionizable drugs from the gastrointestinal tract. The Donnan effect also operates to retard the absorption of drug ions of opposite charge; however, the mutual electrostatic attraction of a macromolecule and drug ion generally results in actual binding, which is more important than the Donnan effect.

Vehicles and Absorption Adjuvants—Drugs that are to be applied topically to the skin and mucous membranes often are dissolved in vehicles that are thought to enhance penetrance. For a long time it was thought that oleaginous vehicles promoted the absorption of lipid-soluble drugs. However, the role and effect of the vehicle has proven to be quite complex. In the skin at least five factors are involved:

1. The effect of the vehicle to alter the hydration of the keratin in the barrier layer.
2. The effect of the vehicle to promote or prevent the collection of sweat at the surface of the skin.
3. The partition coefficient of the drug in a vehicle-water system.
4. The permeability of the skin to the undissolved drug.
5. The permeability of the skin to the vehicle.

The effect of the vehicle to aid in the access of the drug to the hair follicles and sebaceous glands also may be involved, although in man the follicles and glands are probably ordinarily of minor importance to absorption.

A layer of oleaginous material over the skin prevents the evaporation of water, so that the stratum corneum may become macerated and more permeable to drugs. In dermatology it is sometimes the practice to wrap the site of application with plastic wrap or some other waterproof material for the purpose of increasing the maceration of the stratum

corneum. However, the layer of perspiration that forms under an occlusive vehicle may become a barrier to the movement of lipid-soluble drugs from the vehicle to the skin, but it may facilitate the movement of water-soluble drugs. Conversely, polyethylene glycol vehicles remove the perspiration and dehydrate the barrier, which decreases the permeability to drugs; such vehicles remove the aqueous medium through which water-soluble drugs may pass down into the stratum corneum but at the same time facilitate the transfer of lipid-soluble drugs from the vehicle to the skin.

Even in the absence of a vehicle, it is not clear what physicochemical properties of a drug favor cutaneous penetration, high lipid-solubility being a prerequisite, according to some authorities, and an ether–water partition coefficient of approximately one, according to others. Yet, the penetration of ethanol and dibromomethane are nearly equal, and other such enigmas exist. It is not surprising, then, that the effects of vehicles are not altogether predictable.

A general statement might be made that if a drug is quite soluble in a poorly absorbed vehicle, the vehicle will retard the movement of the drug into the skin. For example, salicylic acid is 100 times as permeant when absorbed from water than from polyethylene glycol and pentanol is 5 times as permeant from water as from olive oil. Yet, ethanol penetrates 5 times faster from olive oil than from either water or ethanol, all of which denies the trustworthiness of generalizations about vehicles.

Since the 1960s, there has been much interest in certain highly dielectric aprotic solvents, especially dimethyl sulfoxide (DMSO). Such substances generally prove to be excellent solvents for both water- and lipid-soluble compounds and for some compounds not soluble in either water or lipid solvents. The extraordinary solvent properties probably are due to a high polarizability and van der Waals bonding capacity, a high degree of polarization (dipole moment) and a lack of association through hydrogen bonding. As a vehicle, DMSO greatly facilitates the permeation of the skin and other biological membranes by numerous drugs, including such large molecules as insulin. The mechanism is understood poorly. Such vehicles have a potential for many important uses, but they are at present only experimental, pending further investigations on toxicity.

From time to time, a claim is made that a new ingredient of a tablet or elixir enhances the absorption of a drug, and a comparison of plasma levels of the old and new preparations seems to support the claim. Upon further investigation, however, it may be revealed that the new so-called absorption adjuvant is replacing an ingredient that previously bound the drug or delayed its absorption; thus, the new "adjuvant" is not an adjuvant but rather it is only a nondeterrent.

Other Factors—A number of other less-well-defined factors affect the absorption of drugs, some of which may operate, in part, through factors already cited above. Disease or injury has a considerable effect upon absorption. For example, debridement of the stratum corneum increases the permeability to topical agents, meningitis increases the permeability of the blood–brain barrier, biliary insufficiency decreases the absorption of lipid-soluble substances from the intestine and acid–base disturbances can affect the absorption of weak acids or bases. Certain drugs, such as ouabain, that affect active transport processes may interfere with the absorption of certain other drugs. The condition of the *ground substance*, or "intracellular cement," probably bears on the absorption of certain types of molecules. Hyaluronidase, which depolymerizes the mucopolysaccharide ground substance, can be demonstrated to facilitate the absorption of some, but not all, drugs from subcutaneous sites.

Drug Disposition

The term *drug disposition* is used here to include all processes which tend to lower the plasma concentration of drug, as opposed to drug absorption, which elevates the plasma level. Consequently, the distribution of drugs to the various tissues will be considered under *Disposition*. Some authors use the term disposition synonymously with elimination, that is, to include only those processes which decrease the amount of drug in the body. In the present context, disposition comprises three categories of processes: distribution, biotransformation and excretion.

Distribution, Biotransformation and Excretion

The term *distribution* is self-explanatory. It denotes the partitioning of a drug among the numerous locations where a drug may be contained within the body. *Biotransformations* are the alterations in the chemical structure of a drug that are imposed upon it by the life processes. *Excretion* is, in a sense, the converse of absorption, namely, the transportation of the drug, or its products, out of the body. The term applies whether or not special organs of excretion are involved.

Distribution

The body may be considered to comprise a number of *compartments:* enteric (gastrointestinal), plasma, interstitial, cerebrospinal fluid, bile, glandular secretions, urine, storage vesicles, cytoplasm or intracellular space, etc. Some of these "compartments," such as urine and secretions, are open-ended, but since their contents relate to those in the closed compartments, they also must be included.

At first thought, it may seem that if a drug were distributed passively (ie, by simple diffusion) and the plasma concentration could be maintained at a steady level, the concentration of a drug in the water in all compartments ought to become equal. It is true that some substances, such as ethanol and antipyrine, are distributed nearly equally throughout the body water, but they are more the exception than the rule. Such substances are mainly small, uncharged, nondissociable, highly water-soluble molecules.

The condition of small size and high water solubility allows for passage through the pores without the necessity of carrier or active transport. Small size also places a limit on van der Waals binding energy and configurational complementariness, so that binding to proteins in plasma, or cells, is slight. The presence of a charge on a drug molecule makes for unequal distribution across charged membranes, in accordance with the Donnan distribution (see below). Dissociability causes unequal distribution when there is a pH differential between compartments, as discussed under *The pH Partition Principle* (see below). Thus, even if a drug is distributed passively, its distribution may be uneven throughout the body. When active transport into, or a rapid metabolic destruction occurs within, some compartments, uneven distribution is also inevitable.

The pH Partition Principle—An important consequence of nonionic diffusion is that a difference in pH between two compartments will have an important influence upon the partitioning of a weakly acidic or basic drug between those compartments. The partition is such that the un-ionized form of the drug has the same concentration in both compartments, since it is the form that is freely diffusible; the ionized form in each compartment will have the

concentration that is determined by the pH in that compartment, the pK and the concentration of the un-ionized form. The governing effect of pH and pK on the partition is known as the *pH partition principle*.

To illustrate the principle, consider the partition of salicylic acid between the gastric juice and the interior of a gastric mucosal cell. Assume the pH of the gastric juice to be 1.0, which it occasionally becomes. The pK_a of salicylic acid is 3.0 (Martin[10] provides one source of pK values of drugs). With the Henderson–Hasselbach equation (see page 242) it may be calculated that the drug is only 1% ionized at pH 1.0.* The intracellular pH of most cells is about 7.0. Assuming the pH of the mucosal cell to be the same, it may be calculated that salicylic acid will be 99.99% ionized within the cells. Since the concentration of the un-ionized form is theoretically the same in both gastric juice and mucosal cells, it follows that the total concentration of the drug (ionized + un-ionized) within the mucosal cell will be 10,000 times greater than in gastric juice. This is illustrated in Fig 35-11. Such a relatively high intracellular concentration can have important osmotic and toxicologic consequences.

Had the drug been a weak base instead of an acid, the high concentration would have been in the gastric juice. In the small intestine, where the pH may range from 7.5 to 8.1, the partition of a weak acid or base will be the reverse of that in the stomach, but the concentration differential will be less, because the pH differential from lumen to mucosal cells, etc, will be less. The reversal of partition as the drug moves from the stomach to the small intestine accounts for the phenomenon that some drugs may be absorbed from one gastrointestinal segment and returned to another. The weak base, atropine, is absorbed from the small intestine, but, because of pH partition, it is "secreted" into the gastric juice.

The pH partition of drugs has never been demonstrated to be as marked as that illustrated in Fig 35-11 and in the text. Not only do many drug ions probably pass through the pores of the membrane to a significant extent, but also some may pass through the lipid phase, as explained above for the morphinans and mecamylamine. Furthermore, ion-pair formation in carrier transport also bypasses nonionic diffusion. All processes that tend toward an equal distribution of drugs across membranes, and among compartments, will cause further deviations from theoretical predictions of pH partition.

Electrochemical and Donnan Distribution—A drug ion may be distributed passively across a membrane in accordance with the membrane potential, the charge on the drug ion and the Donnan effect. The relationship of the membrane potential to the passive distribution of ions is expressed quantitatively by the Nernst equation (Eq 7, page 709) and already has been discussed. Barring active transport, pH partition and binding, the drug will be said to be distributed according to the electrical gradient or to its "equilibrium" potential. If the membrane potential is 90 mv, the concentration of a univalent cation will be 30 times as high within the cell as without; if the drug cation is divalent, the ratio will be 890. The distribution of anions would be just the reverse. If the membrane potential is but 9 mv, the ratio for a univalent cation will be only 1.4 and for a divalent cation only 2.0. It, thus, can be seen how important membrane potential may be to the distribution of ionized drugs.

It was pointed out under *Membrane Potentials*, page 707, that large potentials derive from active transport of ions but that small potentials may result from Donnan distribution. Donnan membrane theory is discussed in Chapter 14. According to the theory, the ratio of the intracellular/extracellular concentration of a permeant univalent anion is equal to the ratio of extracellular/intracellular concentration of a permeant univalent cation. A more general mathematical expression that includes ions of any valence is

$$\left(\frac{A_i}{A_e}\right)^{1/Z_a} = \left(\frac{C_e}{C_i}\right)^{1/Z_c} = r \qquad (8)$$

where A_i is the intracellular and A_e the extracellular concentration of anion, Z_c is the valence of cation, Z_a is the valence of anion, C_i is the intracellular and C_e the extracellular concentration of cation and r is the Donnan factor. The value of r depends upon the average molecular weight and valence of the macromolecules (mostly protein) within the cell and the intracellular and extracellular volumes. Since the macromolecules within the cell are charged negatively, the cation concentration will be higher within the cell, that is, $C_i > C_e$. Since a Donnan distribution results in a membrane potential, the distribution of drug ion also will be in keeping with the membrane potential.

The Donnan distribution also applies to the distribution of a charged drug between the plasma and interstitial compartment, because of the presence of anionic proteins in the plasma. Eq 8 applies by changing the subscript i to p, for plasma, and e to i, for interstitial. The Donnan factor, r, for plasma–interstitial space partition is about 1.05:1.

Binding and Storage—Drugs frequently are bound to plasma proteins (especially albumin), interstitial substances, intracellular constituents and bone and cartilage. If binding is extensive and firm, it will have a considerable impact upon the distribution, excretion and sojourn of the drug in the body. Obviously, a drug that is bound to a protein or any other macromolecule will not pass through the membrane in the bound form; only the unbound form can negotiate among the various compartments.

The partition among compartments is determined by the binding capacity and binding constant in each compartment. As long as the binding capacity exceeds the quantity of drug in the compartment, the following equation generally applies:

* The relationship of ionization and partition to pH and pK has been formulated in several different ways, but the student may calculate the concentrations from simple mass law equations. More sophisticated calculations and reviews of this subject are available.[6,11-16]

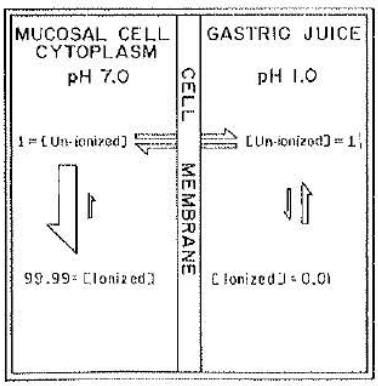

Fig 35-11. Hypothetical partition of salicylic acid between gastric juice and the cytoplasm of a gastric mucosal cell. It is assumed that the ionized form cannot pass through the cell membrane. The intragastric concentration of salicylic acid is arranged arbitrarily to provide unit concentration of the un-ionized form. *Bracketed values:* concentration; *arrows:* relative size depicts the direction in which dissociation-association is favored at equilibrium.

$$\log D_b = \log K + a \log D_f \quad (9)$$

where D_b is the concentration of bound drug, D_f is the concentration of free drug and a and K are constants characteristic of the drug and binding macromolecule. The equation is that of a Freundlich isotherm. As the binding capacity is approached, the relationship no longer holds. For a nondissociable drug at equilibrium, D_f will be the same in all communicating compartments, so that it would be possible to calculate the partition if K and a are known for each compartment. Except for plasma, the values of K and a are generally unknown, hut the percent bound is often known. From the percent bound, the partition also can be calculated, as in Fig 35-12. However, the logarithmic relationships shown in Eq 9 serve as a reminder that the percent bound changes with the concentration, so that the partition will vary with the dose. If the drug is a weak acid or base, the un-ionized free form negotiates among the compartments, but the ionized form is often the more firmly bound, and calculations must take into account the dissociation constant and the different K's and a's of the ionized and un-ionized forms.

It is misbelieved commonly that binding in the plasma interferes with the activity of a drug and the intracellular binding in a responsive cell increases activity or toxicity. Both binding in plasma and in the tissues decreases the concentration of free drug, but this is easily remedied by adjusting the dose to give a sufficient concentration for pharmacological activity. The distribution and activity of the free form is not affected by binding. The principal effect of binding is to increase the initial dose requirement for the drug and create a reservoir of drug from which the drug may be withdrawn as the free form is excreted or metabolized. However, if the binding is extremely firm and release is slow, the rate of release may not be enough to sustain the free form at a sufficient level for pharmacological activity; in such instances the bound drug cannot be considered a reserve.

The effect of binding upon the sojourn of a drug may be considerable. For example, quinacrine, which may he concentrated in the liver to as much as several thousand times the concentration in plasma, may remain in the body for months. Some iodine-containing radiopaque diagnostic agents are bound strongly to plasma protein and may remain in the plasma for as long as 2 yr. In pathological conditions, such as nephrosis, diabetes or cirrhosis, in which plasma protein levels may be decreased, the plasma protein binding, loading dose and duration of action all may be decreased.

If a drug is hound to a functional macromolecule, binding may relate to pharmacological activity and toxicity, providing that the binding is at a critical center of the macromolecule. The binding by nucleic acids of certain antimalarials, such as quinacrine, undoubtedly contributes to the parasiticidal actions as well as to toxicity.

Most drugs are bound to proteins by relatively weak forces, such as van der Waals (London, Keesom or Debye) forces, or hydrogen or ionic bonds. Consequently, binding constants generally are small and binding is usually readily reversible. The larger the molecule, the greater the van der Waals bonding, so that large drug molecules are more likely to be bound strongly than are small ones.

Just as shape and the nature of functional groups is important to drug–receptor combination, so they also are to binding. Drugs of similar shape and/or chemical affinities may bind at the same sites on a binding protein and hence compete with one another. For example, phenylbutazone displaces warfarin from human plasma albumin, which may cause an increase in the anticoagulant effect of warfarin. Some drugs also may displace protein-bound endogenous constituents. For example, sulfisoxazole displaces bilirubin from plasma proteins; in infants with kernicterus the freed bilirubin floods the central nervous system and causes sometimes fatal toxicity.

Depending on the lipid–water coefficient, a drug may be taken up into fatty tissue. The ratio of concentration in fat, to that in the plasma, will not be the same as dictated by the partition coefficient because of the content of water and nonlipids in adipose tissue, and because electrolytes and other solutes alter the dielectric constant and hence solubilities from those of pure water. Lipoproteins, and even nonpolar substituents on plasma proteins, also take up lipid-soluble molecules, so that solubility in plasma can be considerably higher than that in water. The relatively high solubility of ether in plasma makes plasma a pool for ether, the filling of which delays the onset of anesthesia. However, ether and other volatile anesthetics are taken up gradually into the adipose tissue, which acts as a store of the anesthetic. The longer the anesthetic is administered, the greater the store and the longer it takes for anesthesia to terminate when inhalation has been discontinued.

Another notable substance that is taken up readily into fat is thiopental. Even though there is a high solubility of this barbiturate in fat, the low rate of blood flow in fat limits the rate of uptake. Because the blood flow in the brain is very high, thiopental rapidly enters brain tissue. However, it soon equilibrates with the other tissues, and the brain concentration falls as that in the other tissues (eg, muscle or liver) increases. Gradually, however, the fat accumulates the drug at the expense of other compartments. The gradual entry of thiopental into fat at the expense of plasma, muscle or liver is illustrated in Fig 35-13.

Nonequilibrium and Redistribution—Thus far, the distribution of drugs has been discussed mainly as though equilibrium or steady state conditions exist after a drug is absorbed and distributed. However, since most drugs are administered at intervals and the body content of drug rises and falls with absorption and destruction-excretion, neither

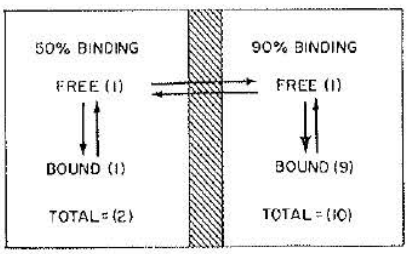

Fig 35-12. Distribution of a drug between two compartments in which the degrees of binding to protein differ. The percent of binding is indicated. Only the unbound drug can pass through the membrane. *Bracketed values:* concentration (courtesy, Schanker[12]).

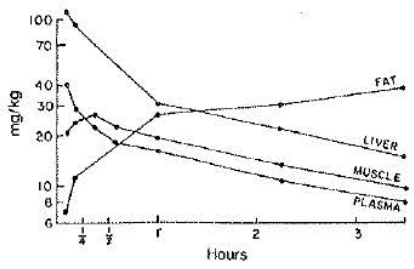

Fig 35-13. Predisposition of thiopental for fat. 25 mg/kg was given to a dog. After a brief sojourn in the more vascular tissues, thiopental gradually transfers to fat, where the lipid-soluble drug dissolves in fat droplets (courtesy, Brodie and Hogben[13]).

a true equilibrium among the body compartments nor a steady state exists.

The term equilibrium is used misleadingly to describe the conditions that exist when the plasma concentration and the concentration in a tissue are equal, as exemplified at the point of intersection of the curves for plasma and muscle or plasma and fat in Fig 35-13. But such "equilibrium" with fat occurs much later than "equilibrium" with muscle, so that no true equilibrium really exists among all the compartments. Furthermore, the crossover point for plasma and any one tissue is not necessarily an equilibrium point, because the rates of ingress and egress from the tissue are not necessarily equal when the internal and external concentrations are equal, since there are numerous factors that make for unequal distribution (pH partition, Donnan effect, electrochemical distribution, active transport, binding, etc).

A study of Fig 35-13 shows that the distribution of thiopental continually changed during the 3.5 hr of observation. At the end of the period, the content in fat was still increasing while that in each of the other compartments was decreasing. This time-dependent shift in partition is called *redistribution*. Eventually, the content in fat would have reached a peak, which would represent as nearly a true equilibrium point as could be achieved in the dynamic situation where metabolic destruction and a slight amount of excretion of the drug was taking place. Once the concentration in the fat had reached its peak, its content would have declined in parallel with that in the other tissues and the partition among the compartments would have remained essentially constant. Redistribution, then, takes place only until the concentration in the slowest filling compartment reaches its peak, so long as the kinetics of elimination are constant.

An index of distribution known as the *volume of distribution* (amount of drug in the body divided by plasma concentration) is of considerable usefulness in pharmacokinetics but is of limited value in defining the way in which a drug is partitioned in the body. Volume of distribution is discussed on page 727.

The word *space* is often used synonymously with volume of distribution. It is employed especially when the distributed substance has a volume of distribution that is essentially identical to a physical real space or body compartment. *N*-acetyl-4-aminoantipyrine is distributed evenly throughout the total body water and is not bound to proteins or other tissue constituents. Thus, the acetylaminoantipyrine space, or volume of distribution, coincides with that of total body water. Inulin, sucrose, sulfate and a number of other substances essentially are confined to extracellular water, so that an inulin space, for example, measures the extracellular fluid volume. Evans blue is confined to the plasma, so that the Evans blue space is the plasma volume. Such space measurements with standard space indicators are a necessary part of studies on the distribution of drugs, since it is desirable to compare the volume of distribution to a drug to the physiological spaces.

Biotransformations

Most drugs are acted upon by enzymes in the body and converted to metabolic derivatives called metabolites. The process of conversion is called *biotransformation*. Metabolites are usually more polar and less lipid-soluble than the parent drug because of the introduction of oxygen into the molecule, hydrolysis to yield more highly polar groups or conjugation with a highly polar substance. As a consequence, metabolites often show less penetrance into tissues and less renal tubular resorption than the parent drug, in accordance with the principle of the low penetrance of polar and high penetrance of lipid-soluble substances. For similar reasons, metabolites are usually less active than the parent drug, often inactive; even if they are appreciably active, they generally are excreted more rapidly. Therefore, the usual net effect of biotransformation may be said to be one of *inactivation* or *detoxication*.

There are, however, numerous examples in which biotransformation does not result in inactivation. Table III (page 742) lists a number of drugs that generate active metabolites; in a few instances activity derives entirely from the metabolite.

There are also examples in which the parent drug has no activity of its own but is converted to an active metabolite: parathion, malathion and certain other anticholinesterases require metabolic activation; inactive chloroguanide is converted to an active triazine derivative; phenylbutazone is hydroxylated to the antirheumatic hydroxyphenylbutazone; inactive pentavalent arsenicals are reduced to their active trivalent metabolites and there are other examples of an activating biotransformation.

When a delayed or prolonged response to a drug is desired or an unpleasant taste or local reaction is to be avoided, it is a common pharmaceutical practice to prepare an inactive or nonoffending precursor, such that the active form may be generated in the body. This practice has been termed *drug latentiation*. Chloramphenicol palmitate, dichloralphenazone and the estolates of various steroid hormones are examples of deliberately latentiated drugs. Because inactive metabolites do not always result from biotransformation, the term detoxication should not be used as a synonym for biotransformation. See Chapter 25.

Biotransformations take place principally in the liver, although the kidney, skeletal muscle, intestine or even plasma may be important sites of the enzymatic attack of some drugs. Since plasma lacks the enzymes and structures required for electron transport, biotransformations in plasma are mostly hydrolytic.

Endoplasmic Reticulum and Microsomal System—Biotransformations in the liver mainly occur in *smooth endoplasmic reticulum*. The endoplasmic reticulum is a tubular system which courses through the interior of the cell but also appears to communicate with the interstitial space, and its membrane is continuous with the cell membrane. Some of the reticulum is lined with ribonucleoprotein particles, called ribosomes, which are engaged in protein synthesis; this is the *rough* endoplasmic reticulum. Although the smooth endoplasmic reticulum lacks such a granular appearance, it is invested heavily with numerous enzymes which biotransform many drugs and some endogenous substances.

When a broken-cell homogenate of the liver is prepared, the reticulum becomes fragmented and the fragments form vesicular structures called *microsomes*. Although the microsomes are artifacts, it is the practice to refer to the *microsomal drug metabolizing system* rather than to the smooth endoplasmic reticulum.

The microsomal system is peculiar in that both oxidations and reductions usually require the reducing cofactor, reduced nicotinamide adenine dinucleotide phosphate (NADPH). This is because microsomal oxidations proceed by way of the introduction of oxygen rather than by dehydrogenation and NADPH is essential to reduce one of the atoms of oxygen. The drug first binds to an oxidized cytochrome P-450. The drug–cytochrome complex then is reduced by NADPH-cytochrome reductase; the reduced complex then combines with oxygen, after which the metabolite is released and oxidized cytochrome P-450 is regenerated. Cytochrome P-450 is a generic term that includes at least 30 and perhaps as many as 100 separate enzymes.[17]

Some of the enzymes of the microsomal system are quite easily *induced;* that is, a substrate of the enzyme may increase considerably the activity of that enzyme by increasing

the biosynthesis of that enzyme. An increase in the amount of smooth endoplasmic reticulum sometimes also occurs concomitantly with enzyme induction.

Treatment of an experimental subject with phenobarbital greatly will increase the rate of metabolism of phenobarbital, which necessitates larger and more frequent doses of the drug in order to maintain a constant sedative effect. Moreover, phenobarbital may induce an increased metabolism of some other, but not all, barbiturates as well as some unrelated drugs, such as strychnine and warfarin. Oddly, warfarin does not induce its own biotransformation readily. At the present time, both self-induction and cross-induction appear capricious and unpredictable.

Induction may create therapeutic problems. For example, the use of phenobarbital during treatment with warfarin increases the dose requirement for warfarin. If the physician is unaware of this interaction and fails to increase the dose, the patient may suffer a thrombotic episode. If the dose of warfarin has been increased and the phenobarbital is then discontinued, the rate of metabolism of warfarin may drop to its previous level, so that the patient is overdosed, with hemorrhagic consequences. Some drugs inhibit rather than induce the microsomal system, which reduces the dose requirement and may lead to toxicity. Cimetidine is an example of a drug that inhibits the hepatic metabolism of a number of other drugs.

The activity of the microsomal system is affected by many factors other than the presence of drugs. Age, sex, nutritional states, pathological conditions, body temperature and genetic factors are among the influences that have been identified. Age, particularly, has received considerable attention. Infants have a poorly developed microsomal system, which accounts for the low dose requirement for morphine and also explains the high toxicity of chloramphenicol in infants.

The activity and selectivity of the microsomal system varies greatly from species to species, so that care must be exercised in extrapolating experimental findings in laboratory animals to man.

Types of Biotransformations—Biotransformations may be *degradative*, wherein the drug molecule is diminished to a smaller structure, or *synthetic*, wherein one or more atoms or groups may be added to the molecule. Very few drugs are degraded completely. However, it is more useful to categorize biotransformations with respect to "metabolic" (nonconjugative) biotransformations and conjugative biotransformations. The former is called phase I and the latter phase II. In phase I, pharmacodynamic activity may be lost; however, active and chemically reactive intermediates also may be generated. The polarity of the molecule may or may not be increased sufficiently to increase excretion markedly. In phase II, metabolites from phase I may be conjugated and sometimes the original drug may be conjugated, thus bypassing phase I. Phase II generates metabolites of high polarity which are excreted readily.

Biotransformations may be placed into five categories: (1) oxidation, (2) reduction, (3) hydrolysis, (4) conjugation and (5) miscellaneous. Oxidation, reduction and hydrolysis comprise phase I. Conjugation comprises phase II. Most of the miscellaneous biotransformations belong in phase I.

Oxidation—Oxidation is more common than any other type of biotransformation. Oxidations that occur primarily in the liver microsomal system include side-chain hydroxylation, aromatic hydroxylation, deamination (which is oxidative and results in the intermediate formation of RCHO), *N*-, *O*-, and *S*-dealkylation (which probably involves hydroxylation of the alkyl group followed by oxidation to the aldehyde) and sulfoxide formation. *N*-Demethylation involves a different system from *N*-dealkylation of higher radicals.

Oxidations that occur elsewhere, other than the microsomes, are generally dehydrogenations followed by the addition of oxygen or water. Examples are the oxidation of alcohols by alcohol dehydrogenase, the oxidation of aldehyde by aldehyde dehydrogenase and the deamination of monoamines by monoamine oxidase and diamines by diamine oxidase. The oxidation of purines, like caffeine and theophylline, is also extramicrosomal.

Reduction—Reductions are relatively uncommon. They mainly occur in liver microsomes, but they occasionally take place in other tissues. Examples are the reduction of nitro and nitroso groups (as in chloramphenicol, nitroglycerin and organic nitrites), of the azo group (as in prontosil) and of certain aldehydes to the corresponding alcohols (as with the deaminated serotonin metabolite, 5-hydroxytryptophal, to 5-hydroxytryptophol).

Hydrolysis—Hydrolysis is a common biotransformation among esters and amides. Esterases are located in many structures besides the microsomes. For example, cholinesterases are found in plasma, erythrocytes, liver, nerve terminals, junctional interstices and postjunctional structures, and procaine esterases are found in plasma. Various phosphatases and sulfatases also are distributed widely in tissues and plasma, although few drugs are appropriate substrates. The hydrolytic deamidation of meperidine occurs primarily in the hepatic microsomes.

Conjugation—A large number of drugs, or their metabolites, are conjugated. Conjugation is the biosynthetic process of combining a chemical compound with a highly polar and water-soluble natural substance to yield a water-soluble, usually inactive, product. Conjugations generally involve either esterification, amidation, mixed anhydride formation, hemiacetal formation or etherization.

Glucuronic acid is the most frequent partner to the drug in conjugation. Actually, the drug reacts with uridine diphosphoglucuronic acid rather than with simple glucuronic acid. The drug or drug metabolite combines at the number 1 carbon (aldehyde end) and not at the carboxyl end of glucuronic acid. The hydroxyl group of an alcohol or a phenol attacks the number 1 carbon of the pyran ring to replace uridine diphosphate. The product is a hemiacetal-like derivative. Since the product is not an ester, the term *glucuronide* is appropriate. Rarely, thiols and amines may form analogous glucuronides.

Carboxyl compounds form esters, appropriately called *glucuronates*, in replacing the uridine diphosphate. *Sulfuric acid* is also a frequent conjugant, especially with phenols and to a lesser extent with simple alcohols. The sulfurated product is called an *ethereal sulfate*. Occasionally sulfuric acid conjugates with aromatic amines to form *sulfamates*. *Phosphoric acid* also conjugates with phenols and aromatic amines. The conjugation of benzoic acid with glycine to yield hippuric acid is a classical example of an *amidation* conjugative process. Cysteine may take the place of glycine, through the intermediation of glutathione, to yield mercapturic acids with certain aromatic acids.

Amidations with amino acids are less frequent than *acetylation*, partly because few drugs are carboxylic compounds. Aromatic amines and occasionally aliphatic amines or heterocyclic nitrogen frequently are acetylated. Acetyl-CoA is the biological reagent rather than acetic acid itself. Unlike most other conjugates, the acetylate (amide) is usually less water-soluble than the parent compound. The acetylation of the para-amino group of the sulfonamides is a prime example of this type of conjugation.

Although most conjugations occur in the liver, the microsomal system is not involved. Some conjugations occur in the kidney or in other tissues.

Miscellaneous—Many amines, especially derivatives of β-phenylethylamine and heterocyclic compounds, are methylated in the body. The products are usually biologically active, sometimes more so than the parent compound. *N-Methylation* may occur in the cell sap of the liver and elsewhere, especially in chromaffin tissue in the case of phenylethylamines.

Phenolic compounds may be *O*-methylated. *O-Methylation* is the principal route of biotransformation of catecholamines such as epinephrine and norepinephrine, the methyl group being introduced on the *meta*-hydroxy substituent. Both *N*- and *O*-methylation require *S*-methyladenosyl cysteine.

Desulfuration, in which oxygen may replace sulfur, takes place in the liver. Thiopental is converted in part to pentobarbital by desulfuration, and parathion is transformed to paraoxon.

Dehalogenation of certain insecticides and various halogenated hydrocarbons may take place, principally in the liver but not in the microsomes.

Excretion

Some drugs are not biotransformed in the body. Others may be biotransformed, but their products still remain to be eliminated. It follows that excretion is involved in the elimination of all drugs and/or their metabolites. Although the kidney is the most important organ of excretion, some substances are excreted in bile, sweat, saliva, gastric juice or from the lungs.

Renal Excretion—The excretory unit of the kidney is called the *nephron* (Fig 35-14). There are several million nephrons in the human kidney. The nephron is essentially

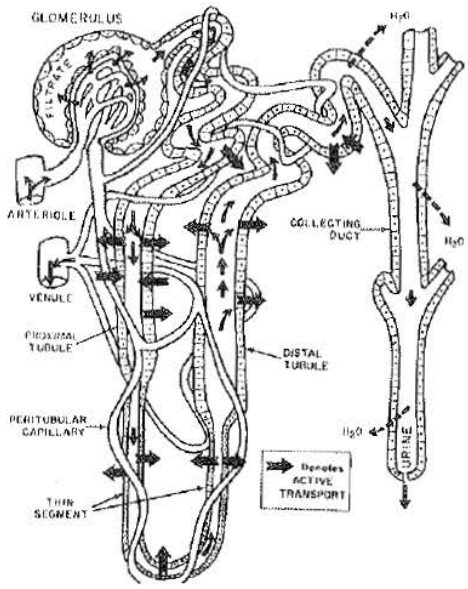

Fig 35-14. Diagram of a mammalian nephron. Note how the lower loops of the postglomerular capillaries course downward and double back along with the tubule. This allows countercurrent distribution to maintain hyperosmolar urine within the thin segment.

a filter funnel, called *Bowman's capsule*, with a long stem, called a *renal tubule*. It also is recognized now that the collecting duct is functionally a part of the nephron. The *blood vessels* that invest the capsule and the tubule are also an essential part of the nephron.

Bowman's capsule is packed with a tuft of branching interconnected capillaries (*glomerular tuft*), which provide a large surface area of capillary endothelium ("filter paper") through which fluid and small molecules may filter into the capsule and begin passage down the tubule. The glomerular tuft, together with Bowman's capsule, constitute the *glomerulus*. The glomerular capillary endothelium and the supporting layer of Bowman's capsule have channels ranging upward to 40 Å. Consequently, all unbound crystalloid solutes in plasma, and even a little albumin, pass into the glomerular filtrate.

The postglomerular vessels, which lie close to the tubules, are critically important to renal function in that substances resorbed from the filtrate by the tubule are returned to the blood along these vessels. The tubule is not straight but rather first makes a number of convolutions (called a *proximal convoluted tubule*), then courses down and back up a long loop (called the *loop of Henle*), makes more convolutions (the *distal convoluted tubule*) and finally joins the collecting duct. The loop of Henle is divided into a *proximal (descending) tubule*, a thin segment and a *distal (ascending) tubule*.

As the glomerular filtrate passes through the proximal tubule, some solute may be resorbed (*tubular resorption*) through the tubular epithelium and returned to the blood. Resorption occurs in part by passive diffusion and in part by active transport, especially with sodium and glucose. Chloride follows sodium obligatorily.

In the proximal region, the tubule is quite permeable to water, so that resorbed solutes are accompanied by enough water to keep the resorbate isotonic. Consequently, although the filtrate becomes diminished in volume by approximately 80% in the proximal tubule, it is not concentrated.

Some *acidification* occurs in the proximal tubule as the result of carbonic anhydrase activity in the tubule cells and the diffusion of hydronium ions into the lumen. In the lumen the hydronium ion reacts with bicarbonate ion, which is converted to resorbable nonionic CO_2.

There is also active transport of organic cations and anions into the lumen (*tubular secretion*), each by a separate system. These active transport systems are extremely important in the excretion of a number of drugs; for example, penicillin G is secreted rapidly by the anion transport system and tetraethylammonium ion by the cation transport system. Probenecid is an inhibitor of anion secretion and, hence, decreases the rate of loss of penicillin from the body.

As the filtrate travels through the thin segment it becomes concentrated, especially at the bottom, as a result of active resorption and a countercurrent distribution effect enabled by the recurrent and parallel arrangement of the ascending segment, the parallel orientation of the collecting duct and the similar recurrent geometry of the associated capillaries.

In the thick segment of the ascending loop of Henle, both sodium and chloride are transported actively.

In the distal tubule, sodium resorption occurs partly in *exchange* for potassium (*potassium secretion*) and for hydronium ions. Adrenal mineralocorticoids promote distal tubular sodium resorption and potassium and hydronium secretion. *Ammonia secretion* also occurs, so that the urine either may be acidified or alkalinized, according to acid–base and electrolyte requirements.

Water is resorbed selectively from the distal end of the distal convoluted tubule and the collecting ducts; water resorption is under the control of the antidiuretic hormone.

Drugs also may be resorbed in the distal tubule; the pH of the urine there is extremely important in determining the rate of resorption, in accordance with the principle of nonionic diffusion and pH partition. The pH of the tubular fluid also affects the tubular secretion of drugs.

As an example of the importance of urine pH, in humans the secondary amine, mecamylamine, is excreted more than four times faster when the urine pH is less than 5.5 than when it is above 7.5; Fig 35-15 illustrates the effect of urine pH on the excretion of this amine. The effect of urine pH on the excretion of a weak acid, sulfaethidole (for the structure, see page 1109, RPS-15), is shown in Fig 35-16.

The urine pH and, hence, drug excretion may fluctuate widely according to the diet, exercise, drugs, time of day and other factors. Obviously, the excretion of weak acids and bases can be controlled partly with acidifying or alkalinizing salts, such as ammonium chloride or sodium bicarbonate, respectively. Comparative studies on potency and efficacy

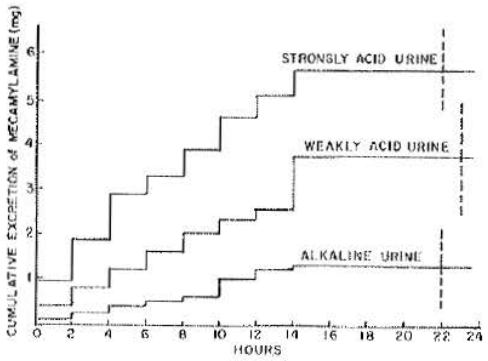

Fig 35-15. The effect of urinary pH on the mean cumulative excretion in man of mecamylamine during the first day after oral administration of 10 mg. *Vertical broken lines:* standard deviation (courtesy, Milne, *et al*[18]).

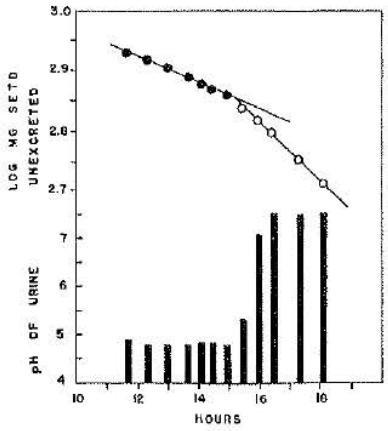

Fig 35-16. The effect of urinary pH on the excretion of sulfaethidole in a human subject after oral administration of 2 g. *Bars* (lower half): urinary pH; *circles* (open and closed, top): log of the amount of drug remaining in the body; *negative slopes* (of lines defined by the circles): a function of the rate constant of excretion. Note the abrupt increase in rate when the urinary pH is changed from acidic to neutral or slightly alkaline (courtesy, Kostenbauder, *et al*[19]).

in man have demonstrated the importance of controlling urinary pH. Urine pH is important only when the drug in question is a weak acid or base of which a significant fraction is excreted. The plasma levels will change inversely to the excretory rate. For example, it has been shown clinically with quinidine that alkalinization of the urine not only decreases the urine concentration but also increases the plasma concentration and toxicity.

The collecting duct also resorbs sodium and water, secretes potassium, acidifies and concentrates the urine. Antidiuretic hormone (ADH) controls the permeability to water of both the collecting duct and the distal tubule.

Renal clearance and the kinetics of renal elimination are discussed in Chapter 36 (page 730).

Biliary Excretion and Fecal Elimination—Many drugs are secreted into the bile and, thence, pass into the intestine. A drug that is passed into the intestine via the bile may be reabsorbed and not lost from the body. This cycle of biliary secretion and intestinal resorption is called *enterohepatic circulation*. Examples of drugs enterohepatically circulated are morphine and the penicillins. The biliary secretory systems greatly resemble those of the kidney tubules. The enterohepatic system may provide a considerable reservoir for a drug.

If a drug is not absorbed completely from the intestine, the unabsorbed fraction will be eliminated in the feces. An unabsorbable drug that is secreted into the bile will likewise be eliminated in the feces. Such fecal elimination is called *fecal excretion*. Only rarely are drugs secreted into the intestine through the succus entericus (intestinal secretions), although a number of amines are secreted into gastric juice.

Alveolar Excretion—The large alveolar area and high blood flow make the lungs ideal for the excretion of appropriate substances. Only volatile liquids or gases are eliminated from the lungs. Gaseous and volatile anesthetics essentially are eliminated completely by this route. Only a small amount of ethanol is eliminated by the lungs, but the concentration in the alveolar air is related so constantly to the blood alcohol concentration that the analysis of expired air is acceptable for legal purposes. The high aqueous solubility and relatively low vapor pressure of ethanol at body temperature account for the retention of most of the substance in the blood. Carbon dioxide from those drugs that are partly degraded also is excreted in the lungs.

Pharmacokinetics

Pharmacokinetics is the science that treats of the rate of absorption, extent of absorption, rates of distribution among body compartments, rate of elimination and related phenomena. Because of its importance, two chapters, *Basic Pharmacokinetics* (page 725) and *Clinical Pharmacokinetics* (page 746), have been devoted to the subject.

Drug Interaction and Combination

Frequently a patient may receive more than one drug concurrently. Case records show that surgical patients commonly receive more than ten, and sometimes as many as 30, drugs and the patient is often under the influence of several drugs at once, sometimes unnecessarily. Multiple-drug administration also is common for patients hospitalized for infections and other disorders. Furthermore, a patient may be suffering from more than one unrelated disorder which demands simultaneous treatment with two or more drugs. In such instances, interactions are unsolicited and often unexpected.

In addition to the administration of drugs concurrently for their independent and unrelated effects, drugs are sometimes administered concurrently deliberately to make use of expected interactions.

Types of Interaction and Reasons for Combination Therapy

A drug may affect the response to another drug in a quantitative way. On the one hand, the intensity of either the therapeutic effect, or side effect, may be augmented or suppressed. On the other hand, a qualitatively different effect may be brought out. The mechanisms of such interactions are many and not always are understood. A drug may not necessarily affect either the quality or initial intensity or effect of another drug, but may cause significant to profound changes in the duration of action. The nature of this type of interaction generally is understood fairly well, although it may not yet have been ascertained for any particular drug combination. The deliberate use of combined interacting drugs is most valid when the mechanism of the interaction is understood and the combined effects are both quantifiable and predictable. The rationales of drug combination and the principles involved are discussed below.

Combinations to Increase Intensity of Response or Efficacy—Sometimes the basis for the action of one drug to increase the intensity of response to another is well understood, but often the reason for a positive interaction is obscure. A terminology has arisen that frequently is not only enlightening as to mechanisms and principles but also which is somewhat confusing.

Drugs that elicit the same quality of effect and are mutually interactive are called *homergic*, regardless of whether there is anything in common between the separate response systems. Thus, the looseness of the term admits a pressor response consequent to an increase in cardiac output to be

homergic with one resulting from arteriolar constriction, even though there is not one common responsive element, the blood pressure itself being but a passive indicator. However, homergic drugs usually have in common at least part of a response system. Thus, both norepinephrine and pitressin stimulate some of the same vascular smooth muscle, even though they do not excite the same receptors.

Two homergic drugs can be agonists of the same receptor, so that the entire response system is common to both. Such drugs are called *homodynamic*. As discussed under *Drug Receptors and Receptor Theory* (page 702), homodynamic drugs will generate dose–intensity of effect curves with parallel slopes, but not necessarily with identical maxima or efficacies, if one of the drugs is a partial agonist.

From mass-law kinetics and dose–effect data of the separate drugs, it is possible to predict the combined effects of two agonists to the same receptor. If both drugs are full agonists, theory predicts that an *EDx* of Drug *A* added to an *EDy* of Drug *B* should elicit the same effect as that of an *EDy* of Drug *A* added to an *EDx* of Drug *B*. An example is shown in Fig 35-17. Dose–percent data with homodynamic drugs can be treated in the same way.[21]

Drugs whose combined effects fit the above conditions are called *additive*. If the response to the combination exceeds the expected value for additivity, the drugs are considered to be *supra-additive*. Purely homodynamic drugs do not show supra-additivity; however, if one drug in the pair has an additional action to affect the concentration or penetrance of the other or to prime the response system in some way, two agonists to the same receptor may exhibit supra-additivity. Two homergic drugs are *infra-additive* if their combined effect is less than expected from additivity. As with supra-additivity, infra-additivity must involve an action elsewhere than on a common receptor.

Two drugs are said to be *summative* if a dose of drug that elicits response x added to a dose of another drug that elicits response y gives the combined response $x + y$. Very little significance usually can be attached to summation. Unless the dose–intensity curve of each drug is linear, rather than log-linear, summation cannot be predicted from the two curves. When summation does occur with the usual clinical doses of two drugs, it almost never occurs over the entire dose range; indeed, if the dose of each of the two drugs is greater than an ED50, summation is theoretically impossible unless it is possible to increase the maximal response. At best, summation is an infrequent clinical finding limited to one or two doses.

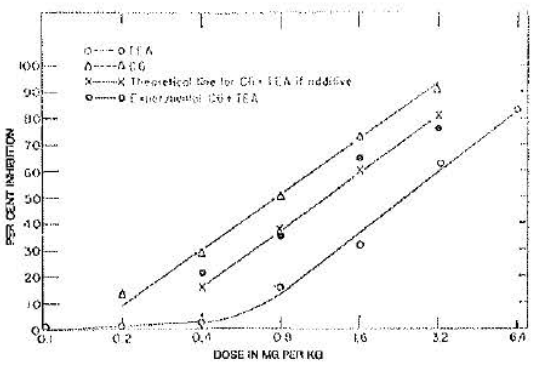

Fig 35-17. Additive inhibitory effects of tetraethylammonium (TEA) and hexamethonium (C6) on the superior cervical ganglion of the cat. The theoretical line for additivity was calculated on the basis that an increment of TEA added to an EDx of C6 should have the same effect as if it were added to an EDx of TEA. When TEA and C6 were administered together, an equal amount of each was given. The dose is the sum of the doses of the two components (courtesy, Harvey[20]).

Two drugs are said to be *heterergic* if the drugs do not cause responses of the same quality. When heterergy is positive, ie, the response to one drug is enhanced by the other, *synergism* is said to occur. The word often has been used to describe any positive interaction, but it should be used only to describe a positive interaction between heterergic drugs. The term *potentiation* has been used synonymously with synergism, but misuse of the term has led to the recommendation that the term be dropped. Synergism is often the result of an effect to interfere with the elimination of a drug and, thus, to increase the concentration; synergism also may result from an effect on penetrance or on the responsivity of the effector system. Examples of a synergistic effect, in which responsivity is enhanced, are the action of adrenalcorticoids to enhance the vasoconstrictor response to epinephrine and the increase of epinephrine-induced hyperglycemia consequent to impairment by theophylline of the enzymatic destruction of the cAMP which mediates the response.

In clinical practice two homodynamic drugs rarely are coadministered for the purpose of increasing the response, since a sufficient dose of either drug should be able to achieve the same effect as a combination of the two. Most clinical combinations with positively interacting drugs are with heterergic drugs.

Combinations to Decrease Individual Doses and Toxicity—When homodynamic drugs are coadministered, it is usually for the purpose of decreasing toxicity. If the toxicities of two homodynamic drug are infra-additive, the toxicity of combined partial doses of the two drugs often will be less than with full doses of either drug. This principle is valid for trisulfapyrimidines mixture (see page 1181).

Combinations to Attack a Disease Complex at Different Points—With many diseases, more than one organ or tissue may be affected or events at more than one locus may bear upon the ultimate perturbation. For example, in duodenal ulcer, psychic factors appear to increase activity in the vagus nerve, which modulates gastric secretion, so that it is rational to explore the effects of sedatives, ganglionic blocking drugs, antimuscarinic drugs and antacids, singly and in combination. In heart failure the decrement in renal plasma flow and changes in aldosterone levels promote the retention of salt and water, so that diuretics and digitalis usually are employed concomitantly. Pain, anxiety and agitation or depression are frequent accompaniments of various pathologic processes, so that it is to be expected that analgesics, tranquilizers, sedatives or antidepressives frequently will be given at the same time, along with other drugs intended to correct the specific pathology.

Combinations to Antagonize Untoward Actions—The side effects of a number of drugs can be prevented, or suppressed, by other drugs. An antagonist may compete with the drug at the receptor that initiates the side effect, depress the side-effector system at a point other than the receptor, or stimulate an opposing system.

Antagonism at the receptor is *competitive antagonism* if the antagonist attaches at the same receptor group as the agonist (see page 703). Antagonism at a different receptor group, or inhibition elsewhere in the response system, is *noncompetitive antagonism*. Both competitive and noncompetitive antagonism are classified as *pharmacological antagonism*. The stimulation of an opposing system is *physiological antagonism*.

Examples of pharmacological antagonism are the use of atropine to suppress the muscarinic effects of excess acetylcholine consequent to the use of neostigmine and the use of antihistaminics to prevent the effects of histamine liberated by tubocurarine. Examples of physiological antagonism are

the use of amphetamine to correct partially the sedation caused by anticonvulsant doses of phenobarbital and the administration of ephedrine to correct hypotension resulting from spinal anesthesia.

Combinations That Affect Elimination—Only a few drugs presently are used purposefully to elevate or prolong plasma levels by interfering with elimination, although continued interest in such drugs probably will increase the number.

Probenecid, which already has been mentioned to antagonize the renal secretion of penicillin, was introduced originally for this purpose. However, because penicillin G is inexpensive and available in repository forms, as well as oral forms (obviating the need for injection), it is less imperative to retard the excretion of penicillin. The low nonallergenic toxicity of penicillin permits very large doses to be given without concern for the high plasma concentrations that result, which also means that there is little necessity for increasing the biological half-life of the drug. Consequently, probenecid is not used routinely today in combination with penicillin.

The use of vasoconstrictors to increase the sojourn of local anesthetics at the site of infiltration continues, but few other clinical examples of the deliberate use of one drug to interfere with either the distribution or elimination of another can be cited. Nevertheless, the subject of the effect of one drug on the elimination of another has become immensely active. Innumerable drugs affect the fate of others and the therapist must be aware of such interactions.

Drugs that induce cytochrome P-450s enhance the elimination of drugs that are metabolized by the liver microsomes. There would be very little point ordinarily to solicit combinations that would shorten the duration of action or lower plasma levels, unless it were to reduce an overdosage. However, since such combinations are used unwittingly or unavoidably, this type of interaction is of great clinical importance.

Combinations to Alter Absorption—In the section on *Vehicles and Absorption Adjuvants* (page 714) it was mentioned that certain substances facilitate the absorption of others. The use of such absorption adjuvants generally is included under the subject of formulation rather than under drug combination. Although drugs which increase blood flow, motility, etc have an effect to increase the rate of absorption, the use of such drugs so far has not proved to be very practical. When it is desired to slow the absorption of drugs, various physical or physicochemical means prove to be more effective and less troublesome than drug combinations.

Fixed Combinations of Drugs

Concomitant treatment with two or more drugs frequently is unnecessary, and it, generally, immeasurably complicates therapy and the evaluation of response and toxicity. Nevertheless, it is often warranted, even essential and cannot be condemned categorically. However, with fixed-dose or fixed-ratio combinations, in which the drugs are together in the same preparation, there are certain disadvantages, except for a few rare instances like trisulfapyrimidines.

The disadvantages are as follows: patients differ in their responsivity or sensitivity to drugs and adjustments in dosage or dose-interval may be necessary. If adjustment of only one component of the mixture is required, it is undesirable that the schedule of the second component be adjusted obligatorily, as it is in a fixed combination. According to which way the dose is adjusted, either toxicity or loss of the therapeutic effect may result. Furthermore, when adverse effects to either component occur, both drugs must be discontinued. The fixed combination denies the physician flexible control of therapy. Especially when one component in a mixture is superfluous yet potentially toxic, as is often the case, the promotion of fixed combinations is reprehensible. However, the separate administration of drugs used in combination often complicates treatment for the patient, who, in an outpatient situation and sometimes in the hospital, may not take all of his medication or who may take it at inappropriate intervals. The resulting consequences may be worse than those of fixed combinations in certain instances. Consequently, a summary dismissal of fixed combinations is unwarranted. Rather, the fundamentals of pharmacokinetics and clinical experience must be brought together with biopharmaceutics to analyze present combinations and to predict possible new allowable combinations.

Dangers in Multiple-Drug Therapy

Some objections to fixed-dose combinations were stated above. Also the unanticipated effects of drug combinations have been touched upon, particularly with respect to effects upon elimination. But it should be made clear that more is at stake than simply the biological half-life of a drug. On page 717 an example was given of the grave clinical consequences of the effect of phenobarbital to enhance the biotransformation of warfarin. Other examples of dangerous interactions, such as the effect of several antidepressants greatly to synergize catecholamines, may be cited. Even some antibiotics antagonize each other and increase mortality.

In addition to the obvious pitfalls posed by the interactions themselves, the use of multiple-drug therapy fosters careless diagnosis and a false sense of security in the number of drugs employed. Multiple-drug therapy should never be employed without a convincing indication that each drug is beneficial beyond the possible detriments or without proof that a therapeutically equivocal combination is definitely harmless. Finally, the expense to the patient warrants consideration.

References

1. Clark AJ: *J Physiol (London) 61:* 547, 1926.
2. Ariens EJ, ed: *Molecular Pharmacology,* vol 1, Academic, New York, 176–193, 1964.
3. Stephenson RP: *Brit J Pharmacol 11:* 379, 1956.
4. Rang HP: *Brit J Pharmacol 48:* 475, 1973.
5. Colquhoun D: The relation between classical and cooperative models for drug action. In Rang HP, ed: *Drug Receptors:* University Park, Baltimore, 1973.
6. Schanker LS: *Advan Drug Res 1:* 71, 1964.
7. Brodie BB, *et al: J Pharmacol Exp Ther 130:* 20, 1960.
8. Truitt EB, *et al: J Pharmacol Exp Ther 100:* 309, 1950.
9. Lillehei JP: *JAMA 205:* 531, 1968.
10. Martin AN, *et al: Physical Pharmacy,* 2nd ed, Lea & Febiger, Philadelphia, 247, 253, 1969.
11. Jacobs MH: *Cold Spring Harbor Symp Quant Biol 8:* 30, 1940.
12. Schanker LS: *Pharmacol Rev 14:* 501, 1961.
13. Brodie BB, Hogben CA: *J Pharm Pharmacol 9:* 345, 1957.
14. Hogben CA: *Fed Proc 19:* 864, 1960.
15. Albert A: *Pharmacol Rev 4:* 136, 1952.
16. Ariens EJ, *et al.* In Ariens, EJ, ed. *Molecular Pharmacology,* vol 1, Academic, New York, 7–52, 1964.
17. Whitlock JP Jr: *Ann Rev Pharmacol Toxicol 26:* 333, 1986.
18. Milne MD, *et al: Clin Sci 16:* 599, 1957.
19. Kostenbauder HB, *et al: J Pharm Sci 51:* 1084, 1962.
20. Harvey SC: *Arch Intern Pharmacodyn 114:* 232, 1958.
21. Weaver LC, *et al: J Pharmacol Exp Ther 113:* 359, 1955.

Bibliography

Albert A: *Selective Toxicity,* 7th ed, Chapman and Hall, London, 1985.

Aronow L, *et al: Principles of Drug Action,* 3rd ed, Wiley, New York, 1988.

Barlow RB: *Quantitative Aspects of Chemical Pharmacology,* University Park, Baltimore, 1980.

Benford D, *et al: Drug Metabolism. From Molecules to Man,* Taylor & Francis, New York, 1987.

Black, JW, *et al*, eds: *Perspectives on Receptor Classification*, Liss, New York, 1987.
Boeynaems JM, Dumont JE, eds: *Outlines of Receptor Theory*, Elsevier/North Holland, Amsterdam, 1980.
Boobis AR, *et al*, eds: *Microsomes and Drug Oxidations*, Taylor & Francis, New York, 1985.
Brodie BB: Physicochemical factors in drug absorption. In Binns TB, ed: *Absorption and Distribution of Drugs*, Williams & Wilkins, Baltimore, 1964.
Burgen ASV, Roberts GCK, eds: *Topics in Molecular Pharmacology*, vols 1 & 2, Elsevier, Amsterdam, 1981, 1983.
Colquhoun D: The link between drug binding and response: theories and observations. In O'Brien RD, ed: *The Receptors: A Comprehensive Treatise*, vol 1, Plenum, New York, 1979.
Coulson CJ: *Mechanisms of Drug Action*, Taylor & Francis, New York, 1987.
DeRobertis E: *Synaptic Receptors, Isolation and Biology*, Dekker, New York, 1975.
Featherstone RM, ed: *A Guide to Molecular Pharmacology*, Parts I and II, Dekker, New York, 1973.
Finean JB, Michell RH, eds: *Membrane Structure*, Elsevier/North Holland, Amsterdam, 1981.
Gilman AG: Guanine nucleotide regulatory proteins and dual control of adenylate cyclase. *J Clin Invest 73*: 93, 1979.
Gregoriadis G, Senior J: *Targeting of Drugs with Synthetic Systems*, Plenum, New York, 1986.
Jakoby WB, *et al*: *Metabolic Basis of Detoxification*, Academic, New York, 1982.
Jenner P, Testa B, eds: *Concepts in Drug Metabolism*, Parts A & B, Dekker, New York, 1980, 1981.
Jolles G, Woolridge KRH, eds: *Drug Design: Fact or Fancy?* Academic, London, 1984.
Kalow W, *et al*, eds: *Ethnic Differences in Reactions to Drugs and Xenobiotics*, Liss, New York, 1986.
Karlin A: Anatomy of a receptor. *Neurosci Comment 1*: 111, 1983.
Kenakin TP: The classification of drugs and drug receptors in isolated tissues. *Pharmacol Rev 36*: 165, 1984.
Kenakin TP: *Pharmacological Analysis of Drug Receptor Interaction*, Raven, New York, 1987.
Lamble JW, ed: *Towards Understanding Receptors*, Elsevier/North Holland, Amsterdam, 1981.
Lamble JW, ed: *More About Receptors*, Elsevier/North Holland, Amsterdam, 1982.
Lamble JW, Abbott AC, eds: *Receptors Again!* Elsevier, Amsterdam, 1984.
Lefkowitz RJ, ed: *Receptor Regulation*, Chapman & Hall, London, 1981.
Lefkowitz RJ, *et al*: Mechanism of hormone-receptor-effector coupling: the β-adrenergic receptor and adenylate cyclase. *Fed Proc 41*: 2664, 1982.
Levine RR: Pharmacology. In *Drug Actions and Reactions*, 3rd ed, Little, Brown, New York, 1983.
Limbird LE: *Cell Surface Receptors: A Short Course on Theory and Methods*, Nijhoff, Boston, 1986.
Mukrikanis A: *New Methods in Drug Research*, Prous, Barcelona, 1985.
Martonosi AN: *Membranes and Transport*, Plenum, New York, 1982.
Meyer UA: Role of genetic factors in the rational use of drugs (Chap 18). In Melmon KL, Morrelli HF, eds: *Clinical Pharmacology*, 2nd ed, Macmillan, New York, 1978.
Mitchell JR, Horning MG, eds: *Drug Metabolism and Drug Toxicity*, Raven, New York, 1984.
Nebert DW, *et al*: Genetic mechanisms controlling the induction of polysubstrate monooxygenase (P-450) activities. *Ann Rev Pharmacol Toxicol 21*: 431, 1981.
Olson RW, Venter JC, eds: *Benzodiazepine/GABA Receptors and Chloride Channels*, Liss, New York, 1986.
Ortiz de Montellano PR, ed: *Cytochrome P450. Structure, Mechanism, and Biochemistry*, Plenum, New York, 1986.
Post G, Crooke ST, eds: *Mechanisms of Receptor Regulation*, Plenum, New York, 1986.
Putney JW Jr, ed: *Phosphoinositides and Receptor Mechanisms*, Liss, New York, 1986.
Rietbrock N, Woodcock BG: *Progress in Drug Protein Binding*, Heyden, Philadelphia, 1981.
Roberts, GCK: *Drug Action at the Molecular Level*, University Park, Baltimore, 1977.
Roche EB, ed: *Bioreversible Carriers in Drug Design*, Pergamon, New York, 1987.
Roth SH, Miller KW, eds: *Molecular and Cellular Mechanisms of Anesthetics*, Plenum, New York, 1986.
Sandler M, ed: *Enzyme Inhibitors as Drugs*, University Park, Baltimore, 1980.
Schanker L: Drug absorption. In La Du, *et al*, eds: *Fundamentals of Drug Metabolism and Drug Disposition*, Williams & Wilkins, Baltimore, 1971.
Schmucker DL: Aging and drug disposition. *Pharmacol Rev 37*: 133, 1985.
Schou JS *et al*, eds: *Drug Receptors and Dynamic Processes in Cells*, Raven, New York, 1986.
Stein WD: *Transport and Diffusion Across Cell Membranes*, Academic, Orlando, 1986.
Stenlake JB: *The Chemical Basis of Drug Action*, Athlone, London, 1979.
Strange PG, ed: *Cell Surface Receptors*, Norwood, Chichester, 1983.
Stroud RM: Acetylcholine receptor structure. *Neurosci Comment 1*: 124, 1983.
Testa, B, ed: *Advances in Drug Research*, vols 14, 15, Academic London, 1985, 1986.
Triggle DJ, Janis RA: Calcium channel ligands. *Ann Rev Pharmacol Toxicol 27*: 347, 1987.
Usdin E, *et al*: *Neuroreceptors*, Wiley & Sons, Chichester, 1981.
Van Rossum JM, ed: *Kinetics of Drug Action*, Springer-Verlag, Berlin, 1977.
Venter, JC, Harrison LC, eds: *Molecular and Chemical Characterization of Membrane Receptors*, Liss, New York, 1984.
Vesell E: Pharmacogenetics. *Biochem Pharmacol 24*: 445, 1975.
Vesell ES: Pharmacogenetics—multiple interactions between genes and environment as determinants of drug response. *Am J Med 61*: 83, 1979.
Vesell ES: The influence of host factors on drug response. I. ethnic background. *Ration Drug Ther 13(8)*: 1, 1979.
Vesell ES: The influence of host factors on drug response. III. diet. *Ration Drug Ther 14(5)*: 1, 1980.
Wardle EN: *Cell Surface Science in Medicine and Pathology*, Elsevier, New York, 1985.

CHAPTER 36

Basic Pharmacokinetics

Stewart C Harvey, PhD
Professor of Pharmacology

and C Dean Withrow, PhD
Associate Professor of Pharmacology
School of Medicine, University of Utah
Salt Lake City, UT 84132

Pharmacokinetics is the discipline which is concerned with the rates of movement of a drug or its metabolites into the body, among its many compartments, out of the body and also which attempts to evaluate the rates of biotransformations of the drug and its metabolites. As in chemistry, it involves primarily following the rate of change in concentration in the appropriate compartment(s), most often in the extracellular fluid (plasma) and/or urine. However, pharmacokinetics is by no means limited to observations of concentration; rates of movement of a drug can be followed by isotopes or other means. The application of pharmacokinetics to drug formulation and treatment regimens also is within the scope of this title. The applications to treatment regimens and other clinical uses of pharmacokinetics are treated in Chapter 37, *Principles of Clinical Pharmacokinetics.*

Orders of Processes

The order of any process is determined by the probability that the appropriate unit events will occur in a given population within a given time. Processes may be zero-order, first-order, second-order, etc, depending upon the number of variables that determine the probability. In pharmacokinetics, only zero-order and first-order processes are important, the latter being of overwhelming significance; consequently, only the kinetics of these two processes will be treated in this chapter.

First-Order Processes

When activity is random within a population of a single species, the probability that a given event will occur is directly proportional to the size of the population. For example, the probability that some atom in a population of radionuclides will disintegrate in any instant is directly proportional to the number of radionuclide atoms in the population. Similarly, the number of molecules of drug that diffuse across a given boundary (eg, the vascular endothelium) per unit time will be proportional directly to the number of molecules near the boundary, which, in turn, is proportional to the concentration. This is the basis of Fick's Law of Diffusion (page 208). Any process in which the rate of change in a population is directly proportional to the population is known as a *first-order* process. In such a process, the time-dependent change in concentration is defined by the equation

$$C = C_0 e^{-kt} \qquad \text{[units of wt} \cdot \text{vol}^{-1} \text{ or molar, etc]} \qquad (1)$$

where C is the concentration at time t, C_0 is the initial concentration (time zero), t is time, e is the natural (Naperian) log base and k is a proportionality constant known as the rate constant. (For a derivation of Eq 1, see page 247.) In a diffusion process, the magnitude of k is determined by the temperature, mobility, permeability and other factors. The numerical value of k also will depend upon the time units (min vs hr, etc) chosen.

Eq 1 predicts that as t approaches infinity, C approaches zero, which would be true for irreversible processes like radioactive decay, diffusion into infinite space, some exentropic SN_1 chemical decompositions and certain enzymatic reactions. However, in a confining space, diffusion and many chemical reactions reach an equilibrium state in which C approaches a finite asymptote as t approaches infinity. Figure 36-1 illustrates a simple situation in which the asymptote is necessarily finite. To satisfy the conditions of this

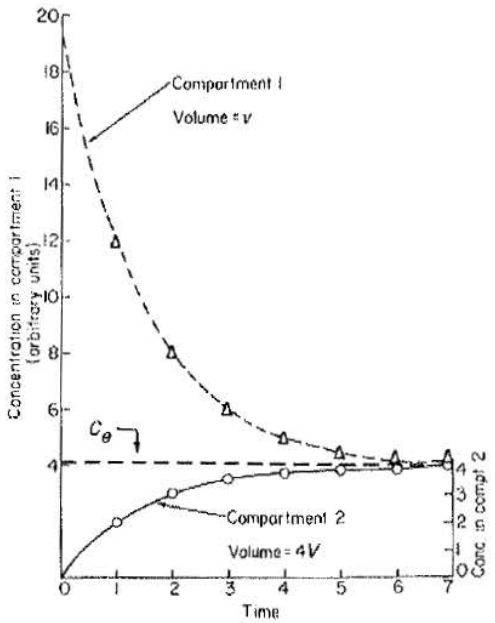

Fig 36-1. Idealized diffusion kinetics of a hypothetical drug that equilibrates between two compartments. Transfer is from compartment 1 into compartment 2. The equilibrium concentration is ⅕ of that initially in compartment 1, because the final volume of distribution is 5 times that of compartment 1.

closed system, $(C_0 - C_e)$ must be substituted for C_0 in Eq 1, C_e being the equilibrium concentration.

In Eq 1, the algebraic sign of k is usually negative, which indicates a diminishing concentration with time. However, in Fig 36-1 the concentration in compartment 2 rises logarithmically with time; nevertheless, k is negative, since the rate diminishes exponentially with time. The equation for the logarithmically rising concentration in compartment 2 will take the form of Eq 5 (page 727), in which C_e would be used in lieu of C_p^∞.

Eq 1 can be written in the log form,

$$\log C = \log C_0 - 0.434kt \qquad [\text{no units}] \qquad (2)$$

The coefficient 0.434 results from the conversion of the natural log base, e, to log base 10 (0.434 = 1/2.303). The equation determines that a plot of log C against t will be rectilinear (bottom of Fig 36-1) with a slope of $-0.434k$ and an ordinate-intercept of C_0. For pharmacokinetics, this is a useful type of plot, because, in the straight-line form, back extrapolation to estimate C_0 is easier and more accurate than from a curve, and k can also be determined graphically.

Rate Constants and Half-Life—Since first-order processes are characterized by exponential or logarithmic kinetics, it follows that a constant fraction of the present or instantaneous population (eg, concentration) changes per unit time, that fraction being equal to $0.434k$; k has the units of t^{-1}. Another way of expressing the rate of change is that of half-time (or especially *half-life*, if the population is decreasing), with the notation $t_{1/2}$. The half-time is the time that it takes the population to decrease (or increase) by 50% of the total possible change. By setting C equal to $\frac{1}{2}C_0$ in either Eqs 1 or 2 and solving for t (which is $t_{1/2}$ under these constraints),

$$t_{1/2} = \frac{0.693}{k} \qquad [\text{units of time}] \qquad (3)$$

Zero-Order Processes

When an enzyme or transport system is saturated, the activity cannot be increased further by increases in the concentration of substrate. Consequently, the rate remains constant so long as the concentration of substrate is in excess of the saturating concentration. In this situation, the rate is independent of the concentration. The kinetics are described as being of *zero-order*, and it is customary to speak of the process as being a zero-order process. The equation describing zero-order kinetics is

$$C = C_0 - kt \qquad [\text{conc vol}^{-1}] \qquad (4)$$

where k has the units of amount/unit time. A plot of C against t on Cartesian coordinates will yield a straight line; a plot of log C against t will yield a curved line. As the process continues, the concentration eventually will fall to subsaturation levels, and the kinetics will change, usually to first-order kinetics, so that it is more appropriate to speak of the initial kinetics and not the process as being zero-order.

Pharmacokinetic Models

The plasma, cerebrospinal fluid, interstitial space, glandular or renal tubular lumina, gall bladder, etc and each cell are all compartments which a drug may or may not enter or leave with different rate constants. In addition, binding to protein or other sequestration also is governed by characteristic rate processes. Consequently, it might be expected that the kinetics of absorption, distribution and elimination would be very complex and perhaps beyond analysis and mathematical description. Fortunately, the rates of distribution among the various tissues and myriad cells generally are not dispersed greatly, and most such processes are first-order. Thus, the kinetics behave as though the drug were being distributed among one, two or, at the most, a few compartments, and they are amenable to mathematical modeling. Like the volume of distribution (page 727), a pharmacokinetic compartment is fictive or virtual and may be difficult to define in precise anatomical terms. Therefore, a compartment is defined mainly by its pharmacokinetic parameters.

Open One-Compartment Model

In this model, the body is assumed to behave as though it were a single compartment, that is, as though there were no barriers to movement of a drug within the total body space and as though the final equilibrium distribution is attained instantaneously. In practice, the model adequately describes the pharmacokinetic behavior of a drug if the final equilibrium distribution is attained rapidly in comparison to the rates of absorption and elimination. The term *open* indicates that input and output (from any and all routes of administration and elimination, respectively) are unidirectional and that the one compartment (ie, body) is not within a confined space and hence does not come into chemical equilibrium with its external environment. In simple diagram, such an open one-compartment model is depicted in Fig 36-2. In the diagram, the compartment represents the entire body (excluding the lumina of the gastrointestinal tract, urinary tract, pulmonary alveoli, etc, which communicate with the open environment). The term, V_d is the *volume of distribution* (see page 727). However, V_d is not necessarily that of the body or even total body water; as noted on page 728, the volume of distribution, V_d, is a fictive one considered to be equal to fD/C_p (where f is the fraction absorbed, D is the dose and C_p is the plasma concentration) in which it hypothetically is assumed that the concentration is the same throughout the volume and is equal to the plasma concentration. In reality, concentration is not homogeneous throughout, but this cannot be determined from C_p alone (which simply averages all inputs and outputs); as long as distribution equilibrium is achieved rapidly, the kinetics as perceived through blood or urine concentrations are the same whether distribution is homogeneous or heterogeneous.

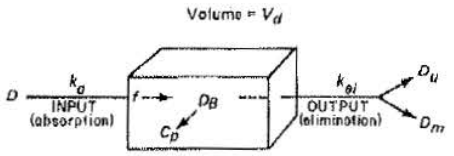

Fig 36-2. The open one-compartment pharmacokinetic model. An amount of drug, D_B, is absorbed from the administered dose, D, with a rate constant of k_a into a compartment with volume V_d and is distributed instantaneously to reach a plasma concentration C_p. V_d is obtained by dividing D_B by C_p. D_B = dose D times f, the fraction absorbed. Drug is eliminated from the compartment with a rate constant k_{ei}. D_u is the amount excreted into urine, feces, expired air, sweat, milk, etc; D_m is the amount of drug metabolized.

In order to derive formulae to describe time-related changes in C_p, it is convenient to consider absorption and elimination separately, as though each were occurring in the absence of the other, then to add them algebraically to determine the total integral kinetics.

Absorption—If a drug is administered intravenously in a single, rapid injection, absorption is bypassed. The time for such injections is usually so short compared to other pharmacokinetic processes that it is customary to consider the peak plasma concentration and equilibrium distribution to occur instantaneously in one-compartment systems. This is depicted in panel A of Fig 36-3. In the model for the figure, there is no elimination and C_p remains constant once injection is accomplished. With constant intravenous infusion (panel B), C_p rises rectilinearly so long as infusion continues at a constant rate. With other routes of administration, absorption usually manifests first-order kinetics, since most drugs are absorbed by simple diffusion. Thus, the drug disappears exponentially from the site of administration (as from compartment 1 in Fig 36-1). The equation for the concentration of a drug in the plasma after a single extravascular dose of a drug, assuming no elimination takes place, is

$$C_p = C_p^\infty - C_p^\infty e^{-k_a t} \qquad \text{[units: wt} \cdot \text{vol}^{-1}\text{, etc]} \qquad (5)$$

where C_p is the concentration at time t, C_p^∞ is the final concentration at "infinite" time and k_a is the absorption rate constant (units: time^{-1}). Absorption is characterized by a half-time equal to $0.693/k_a$. Bimolecular absorption processes, such as facilitated diffusion or active transport, also often show first-order kinetics, especially at drug concentrations well below those at which the carrier system will become saturated. At saturation, the kinetics become zero-order. Even the rate of dissolution of a drug approximates a first-order process, provided that the drug is soluble readily and diffuses rapidly. If the solubility and diffusibility are low, it will approximate a zero-order process so long as there is saturation around the solid phase. Some sustained-release dosage forms are designed to release drugs at a constant rate (zero-order) over long periods of time.

Absorption by the oral route rarely conforms to simple first-order kinetics. A drug is absorbed at different rates from the stomach and the three segments of the intestine, partly simultaneously and partly sequentially. Absorption from the stomach usually is quite slow compared to that from the small intestine, and it is sometimes so slow that a significant amount of drug appears in the blood only after the stomach contents are emptied. Thus, there may be a *lag* between the time of drug administration and the appearance of drug in the blood. That is, the curve describing the time-dependent rise in C_p does not pass through the origin. An example of lag in the absorption of pentobarbital is shown in Fig 36-4. Enteric-coated or other delayed-release dosage forms also cause lag. The mathematical formulation of lag

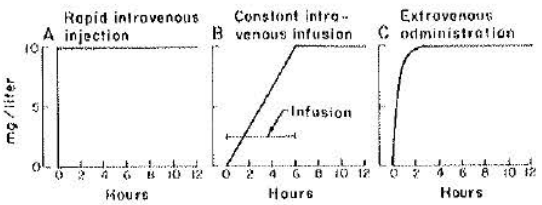

Fig 36-3. Time-concentration curves for injection (A), infusion (B) and extravenous (C) administration of drug in the one-compartment model. The volume of the compartment is 100 L (V_d = 100 L); the amount of drug administered in each instance is 1000 mg. Drug elimination has been set to zero, so that the time-concentration curve for each model of administration can be examined without the complication of simultaneous elimination (courtesy, Bigger,[1] adapted).

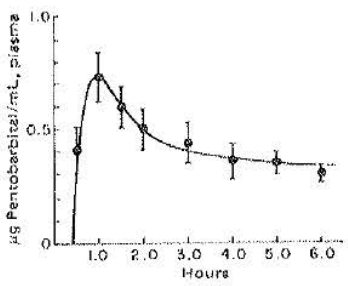

Fig 36-4. The time course of pentobarbital in the blood of a fasting human subject following the oral administration of 50 mg. The figure shows a lag-time of about 20 min, approximately the emptying time of the fasting stomach (courtesy, Dittert[2]).

will be deferred to the next section in connection with Eq 28. Factors affecting absorption are enumerated on page 713. Some changes in gastrointestinal conditions during the course of absorption are part of diurnal rhythms or are caused by the drug itself, which make it impossible to establish a steady basal state for description; others may result from emotionality, ingestion of foodstuffs, water, other drugs, etc, and can be controlled adequately for scientific purposes but may vary greatly in practical circumstances. Absorption by other routes is also subject to variability. Some drugs that are completely absorbed in normal patients may not be absorbed in persons with abnormal gastrointestinal function, as the result of genetic, pathological or surgical factors. Many drugs are not absorbed completely even when gastrointestinal function is optimal. Absorption can be limited by the physical state of the drug and by other substances in the dosage form. The amount of drug absorbed into the body (D_B) is related to the dose as follows:

$$D_B = fD \qquad \text{[units: wt]} \qquad (6)$$

where D_B is the amount absorbed (drug in the body), f is the fraction absorbed and D is the dose administered. The property of a drug to be absorbed from its dosage form is known as *bioavailability*, and f is the bioavailability factor. The bioavailability factor often is determined by comparison of the area under the concentration curve (AUC) of a given dose of drug given orally with that of the same dose given intravenously (see page 736).

Distribution—In the open one-compartment model, the body is treated as though it were a single compartment in which the absorbed drug is mixed instantaneously and homogeneously. Clearly, the assumption of instantaneous equilibrium establishes only an ideal mathematical boundary condition to facilitate pharmacokinetic calculations. At best, no drug could be equilibrated in less than one circulation time, and no drug has been shown to distribute so rapidly. However, for practical purposes, a distribution time of a few minutes is negligible compared to absorption and elimination times. Only water-soluble drugs of small molecular size which are confined completely to the extracellular space equilibrate rapidly enough to meet the requirements of the ideal one-compartment model, but, for clinical purposes, the one-compartment model is adequate to describe the pharmacokinetics of a large number of drugs.

Volume of Distribution and Distribution Coefficient—The hypothetical volume within which a drug is distributed is known as the *volume of distribution*, V_d. It may be calculated by dividing the amount of drug in the body, D_B, by the plasma concentration, C_p, where C_p is the concentration in plasma. It is important to note that C_p is usually the total concentration of unbound plus bound drug. Under real conditions, D_B and C_p vary with time, and computation

must be made in such a way as to eliminate the time variable. One such way is to extrapolate C_p to zero time (eg, see Figs 36-6 and 36-8), in which case

$$V_d = f\,D/C_p^0 \qquad (7)$$

where D is the dose administered, f is the bioavailability factor (fraction that reaches the systemic circulation) and C_p^0 is the plasma concentration at zero time, determined by extrapolation. When the drug is given intravenously, $D_B = D$.

Of course, V_d will vary with body weight, so that it needs to be normalized in a way that allows comparisons among individuals of different body weights. Such a normalized V_d is the *distribution coefficient*,* Δ', calculated by the equation

$$\Delta' = V_d/BW \qquad (8)$$

where BW is body weight. Units are usually mL/g or L/kg, and care must be taken to employ the appropriate units of weight, concentration and volume in Eqs 7 and 8. The notation Δ' is a more serviceable parameter than V_d and is the form of V_d usually found in tables of pharmacokinetic data, usually under the heading, "Volume of Distribution," rather then Δ'.

Although V_d and Δ' are derived as though the concentration was equal to C_p throughout the volume, concentration is, in fact, almost never homogeneous, and consequently V_d and Δ' are only imaginary (fictive, virtual) volumes. Factors that make for nonhomogeneous distribution are: binding to proteins, dissolution into body lipids, pH partition, active transport, electrochemical and Donnan distributions, etc. Even if C_p (free) rather than C_p (total) is used to calculate V_d, V_d would not represent a real space, because of these manifold factors that cause uneven distribution. Consequently, the principal utility of V_d or Δ' is not so much in permitting an estimation of where the drug is distributed but rather as a measure of the reservoir from which a drug is being delivered and/or cleared (see page 729 and Table II, page 731). However, with appropriate considerations, V_d or Δ' also may indicate the general ability of a drug to penetrate membranes, dissolve in fat or bind extensively to extravascular macromolecules.

Highly polar, poorly penetrant drugs tend to be confined mostly to the extracellular space; if these drugs are little bound to plasma proteins, they will have Δ's of about 0.3 mL/g, less if there is significant binding to plasma proteins. The lower limit to Δ' is about 0.04 mL/g, which approximately is equal to the plasma volume. Drugs that are distributed throughout body water and are not bound or concentrated have Δ's of approximately 0.7 mL/g, the Δ' of body water. Lipid-soluble drugs that are bound negligibly to plasma protein have Δ's that range usually from about 0.7 to 3–4 mL/g, depending upon water-lipid distribution coefficients. Some drugs that bind strongly to chromatin have Δ's that approach 1000 mL/g. However, many drugs combine penetrance, lipid solubility and protein binding in such proportions to make it difficult to interpret the meaning of Δ' without ancillary information.

Since, by definition, V_d varies reciprocally with C_p, it is essential to recognize that binding to plasma proteins, by increasing C_p, will decrease V_d. Despite this, plasma protein binding has no *real* effect on extravascular distribution. Since it is only the free form that moves among the spaces and tissues, it follows that alterations in plasma protein binding alone will not alter the extravascular (indeed, extraplasma) distribution. Only the calculated, fictive quantity, Δ', is affected. For example, nafcillin has a Δ' of 0.29 mL/g and is 90% bound to plasma proteins. If there was no protein binding, Δ' would equal 2.9 mL/g, a volume sufficiently larger than that of water, to suggest considerable extravascular binding. However, it is not the masking of the degree of extravascular distribution that is the source of difficulty when there is significant binding to plasma proteins, but rather because the extent of protein binding is not always constant. Both the quantity and binding properties (affinity and capacity) of human plasma proteins can vary in health, disease and the presence of other drugs (see pages 195 and 716). If the degree of binding of nafcillin to plasma proteins was to change to 50% as the result of hypoalbuminuria, Δ' would become nearly 0.48 mL/g. The Δ' of ampicillin, which is bound only to the extent of 18%, would not be affected so greatly. A further complication of binding to plasma proteins is occasioned when the degree of binding, and hence the magnitude of Δ', is dose-dependent. There are a number of known examples in which Δ' varies with the dose.

Elimination—Once a drug is absorbed, it is transported by the blood to the tissues, among which it is distributed, metabolized and/or excreted; all of these processes lower the plasma concentration of the drug. Each separate process ordinarily has first-order kinetics, and the overall change in plasma concentration is described by the linear combination (or algebraic addition) of the separate equations. In the one-compartment model, the kinetics of distribution are ignored, since distribution occurs so rapidly that distribution occurs before any practical blood-sampling or repetitive dosing occurs. Thus, after intravascular administration the plasma concentration, C_p, will fall exponentially according to Eq 1. Such an exponential elimination of theophylline, given intravenously, is shown in Fig 36-5. According to Eq 2, if the data of Fig 36-5 are plotted semilogarithmically, as in Fig 36-6, a straight line should result. Several derived data can be obtained from such a plot. Extrapolation to zero time (ie, the y intercept) gives C_p^0, the theoretical plasma concentration at time zero. It is a theoretical concentration, because neither injection nor distribution actually is instantaneous. Nevertheless, C_p^0 is a very practical figure. For example, from it may be derived the volume of distribution, V_d, simply by dividing the dose, D, by C_p^0 (see page 727). In the figure, C_p^0 = 0.0115 mg/mL, so that V_d is 43.5 L, or about 89% of the volume of total body water in a 70-kg adult. The plasma half-time, $t_{1/2}$, can be determined directly from the graph or from the elimination rate constant, k_{el}, by means of Eq 3. (Conversely, k_{el} could be derived from $t_{1/2}$, determined visually from a graph.) When determining k_{el} from the slope, it must be kept in mind that the log of the concentration must be used rather than the antilog that is

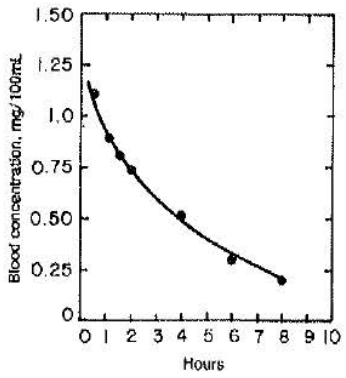

Fig 36-5. Elimination curve of average blood levels of theophylline in 11 human subjects after intravenous administration of 0.5 g aminophylline per 70 kg to each (courtesy, data, Truitt, *et al* [3]).

* Δ' is not to be confused with water-lipid distribution coefficients.

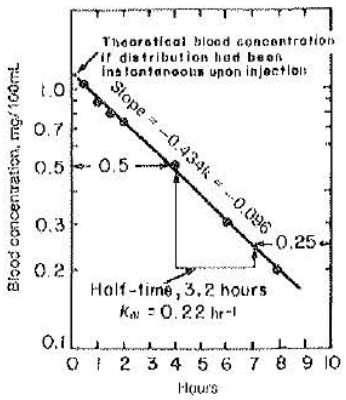

Fig 36-6. Semilog plot of the elimination curve in Fig 36-5. Note the log scale of the ordinate.

plotted on the log-scaled ordinate in the figure. In natural logarithms, the slope $(\ln C_{p1} - \ln C_{p2})/(t_2 - t_1)$ is equal to k_{el}; in decilogarithms, the slope $(\log C_{p1} - \log C_{p2})/(t_2 - t_1)$ is equal to $0.434k_{el}$. From the figure, k_{el} is found to be 0.22 hr^{-1}. However, this is an *instantaneous* rate and is not the same as the fraction that disappears over a finite interval, eg, 1 hr. Nevertheless, it is sometimes convenient to use k_{el} for calculation of the amount of drug lost per unit time, eg, for the calculation of maintenance doses during chronic drug therapy. If the time interval under consideration is very short (at least ½) as compared to the half-life of the drug, k_{el} may be considered to be an indication of the percentage of the drug lost during the time interval. For example, if the half-life of a drug is 6.93 hr, k_{el} would equal 0.1 hr^{-1}, and approximately 10% of the drug would be lost in 1 hr. Thus, if 100 mg of drug were present at the beginning of the hour, 10 mg would be lost by the end of the hour. (The exact amount of drug lost is determined by use of Eq 2 and is found to be 9.51 mg in 1 hr, a difference of about 5% from the approximated value.) If such an approximation was used in the example given in Fig 36-6 (1 hr is a little less than ½ of the half-life) and the amount of drug present is 100 mg, the figures comparable to the illustration above would be 22 versus 19.7 mg, respectively, almost a 12% difference.

Within the group of 11 subjects in the above study, there were considerable differences in k_{el} or $t_{1/2}$ among the members. One cannot overemphasize the caveat not to take too literally the average half-life data found in various tables or other literature but rather to assume a probability that the half-life in a given patient may depart considerably from that average value. The half-lives of some drugs vary over a wide range even in normal individuals. The half-life of amitriptyline, a drug with a complex metabolic and excretory elimination, varies nearly tenfold; even the half-life of penicillin, a drug with a simple excretory elimination, varies twofold. In persons with hepatic or renal failure, the published mean half-life data may not even be in a range applicable to such persons.

The half-life also may vary widely from species to species; for example, in man, the half-life of sulfaethidole is about 8 hr, whereas in cattle it is less than 2 hr. Half-lives also vary considerably even among congeneric drugs, as may be seen with the sulfonamides shown in Table I.

The biological half-life must not be confused with the time for the response to decline by 50%, since dose, the requirement for a threshold concentration, latency of response and other factors may cause a nonparallelism between blood concentration and intensity of response. In fact, because the relationship between effect and plasma concentration is usually logarithmic, effect tends to decline in a linear, not loglinear, fashion.

Table I—The Approximate Biologic Half-Life in Man of Several Sulfonamides[4]

Drug	$t_{1/2}$ (hr)
Sulfamethylthiadiazole	2
Sulfaethidole	8
Sulfisoxazole	8
Sulfamethoxypyridazine	34

In Fig 36-2, the rate constant for elimination is designated k_{el}, without reference to the mode or route of elimination. However, k_{el} may be a compound constant equal to the sum of the rate constants that define the various simultaneous (ie, parallel) contributory processes, such as biotransformation, renal excretion, biliary secretion, etc. Thus, the compound or overall constant is $K = k_1 + k_2 + k_3 \ldots k_n$, where k_1, $k_2 \ldots k_n$ are the rate constants of the separate contributory processes. Consider the case in which a drug simultaneously is biotransformed and excreted unchanged in the urine. The initial concentration, C_p^0, therefore will be diminished by both $C_p^0 e^{-k_m t}$ and $C_p^0 e^{-k_u t}$, where m designates metabolism and u renal excretion. (In some notations, k_u is designated k_r, k_{10}, k_3 or k_e.) Therefore, Eq 2 adapted for the two processes becomes

$$\log C_p = \log C_p^0 - 0.434k_m t - 0.434k_u t$$
$$= \log C_p^0 - 0.434(k_m + k_u)t \qquad \text{[no units]} \quad (9)$$

Thus, $k_m + k_u$ combine to make a single constant, which is the overall elimination rate constant. In order to identify it as a compound, or overall, constant, it is sometimes designated as K, rather than k_{el}.

Clearance and Routes of Elimination—The half-life of a drug is a useful pharmacokinetic parameter. Since half-life is expressed in units of time, it is an easily understood, concise indication of the rate of disappearance or accumulation of a drug. Further, it is used to estimate the time necessary to attain a new steady state whenever a steady state is altered by a change in the factors determining dose regimen, namely, drug dose, bioavailability, the dose interval, rate of elimination and volume of distribution (see page 740). However, the elimination half-life of a drug is a complex function of drug distribution, biotransformation and elimination. A more direct expression of the rate of drug elimination is drug clearance.

Clearance is the rate of removal of a drug or other substance from the body, expressed as the *in vivo* volume equivalent of the substance being removed per unit time. In order to illustrate the concept, assume that drug D is being eliminated from the body at a rate of 0.1% per minute. The absolute amount of drug that was eliminated would therefore be equal to 0.1% D_B per minute. Since D_B is distributed as though it were in a volume V_d (volume of distribution, page 740), one can calculate the fictive volume equivalent of the amount of drug lost per minute, which in this instance would be 0.1% V_d/min. Since the relative rate of loss, 0.1%/min or 0.001/min is, in fact, k_{el}, it may be seen that

$$Cl_t = k_{el} \times V_d \qquad \text{[vol/unit time]} \quad (10)$$

where Cl_t is total body clearance. It may be expressed in units of mL/min, L/hr or mL/kg/min; the reader must be alert to the units in which given clearance data are expressed. It must be emphasized that clearance is a hypothetical or fictive quantity, since the body rarely clears a drug completely from a specific volume of body fluid. Only when elimination is flow-limited is the blood that passes through the eliminating organs(s) totally cleared, so that the effluent blood is essentially devoid of drug; in such an in-

stance the clearance approximates the rate of blood flow. If the concentration in the effluent blood were to be only 0.5 of the affluent blood, the clearance would be said to be 0.5 that of the blood flow.

Although clearance is the dV/dt equivalent of dD_b/dt, or the volume equivalent of the drug lost per unit time, the hypothetical volume cannot be regarded as also having been eliminated. Just as the depleted effluent blood from the eliminating organ is returned to the systemic circulation to mix with all the blood and as drug is redistributed and re-equilibrated among the vascular and extravascular components of V_d, the fictive volume that is "cleared" remains a part of V_d, so that the only change that is effected is one in concentration, of which C_p is the index. Since V_d and k_{el} are both constant, it follows that Cl_t is also constant.

The concept of clearance can be applied to the whole body or to specific organs. The former application is a convenient way to indicate overall drug elimination; the latter application is used to indicate the contribution of a specific organ to drug disappearance.

Total Systemic (Whole Body) Clearance—Total body clearance is the sum of all the separate clearances that contribute to drug elimination, ie, $Cl_{tot} = Cl_{metab} + Cl_{renal}$, etc. It is essential that k_{el} be expressed in the same time units as are used in clearance (usually min). In Eq 10, dividing by 60 converts k_{el} in hr^{-1} to min^{-1}, so that clearance can be expressed in mL/min. Whole-body clearance in a one-compartment system is also equal to dose divided by the area under the curve:

$$Cl_{tot} = D/AUC_0 \qquad [mL \cdot min^{-1}] \qquad (11)$$

where AUC_0 is the area under the curve (AUC), discussed on page 736. The determination of Cl_{tot} in the two-compartment system is discussed on page 739.

Renal Excretion and Clearance—The principles of renal excretion and clearance have been used for approximately 50 years as tools for studying renal physiology and pathology and hence were adapted early to pharmacokinetics. Consequently, renal clearance of drugs is a classic illustration of the general subject of clearance. As discussed in Chapter 35, all drugs are filtered in the glomerulus and some also are secreted into the urine by renal tubular cells; there is also resorption of drugs from the tubular luminal fluid back into the blood as the fluid passes along the tubule. Glomerular filtration is the passage through the glomerular vascular endothelium of the plasma fluid and all solutes therein small enough to pass through the pores; that is, it is the filtration of water and all micromolecular solutes. Thus, it is independent of the presence of drug and is a function of the filtration pressure (which relates to blood pressure) and the mean transit time across the glomerular capillaries. The rate of filtration is known as the *glomerular filtration rate*, *GFR*, and it has the units of vol/min (usually mL/min). In turn, the transit time is determined by the rate of flow of blood through the glomeruli; this rate of blood flow is known as the *renal plasma flow*, *RPF*. Since only a fraction of the plasma is filtered during passage through the glomerulus, it is useful to designate this fraction as the filtration fraction, *FF*, where $FF = GFR/RPF$. The average renal *plasma* (not blood) flow in the adult human male is approximately 600–700 mL/min, and the *GFR* is approximately 100–125 mL/min (of which 99% of the water is resorbed and returned to the blood); thus, the filtration fraction is approximately 0.2.

Under basal conditions, the *GFR* is roughly constant in time. Therefore, the only major variable that determines the rate of filtration of free drug is the concentration of drug in the plasma. Thus,

$$F = C_{pf} \cdot GFR \qquad [\text{units: } mL \cdot min^{-1}] \qquad (12)$$

where F is the filtration rate of the drug, usually in units of mg/min, and C_{pf} is the amount of *free* drug in the plasma. If the drug is unbound, $C_{pf} = C_p$. If the drug is bound to plasma protein, then

$$F = [C_p(1 - p)] \cdot GFR \qquad [mL \cdot min^{-1}] \qquad (13)$$

where p is the fraction bound to plasma protein.

The *GFR* may be determined by the steady-state rate of excretion of any nonbound chemical substance that is not secreted subsequently and/or resorbed by the renal tubules, so that the amount of substance which appears in the urine is all of that which was filtered and no more. Two such substances are *creatinine* and *inulin*. With creatinine, the endogenous plasma levels are nearly constant, and thus creatinine lends itself readily to the determination of *GFR*. Either inulin or creatinine may be given by constant intravenous infusion; usually, creatinine is used. However, it is not customary to express the *GFR* of creatinine or of drugs as F, in terms of mg/min, but rather in terms of *clearance*. As discussed above, clearance is a hypothetical volume of plasma which, if completely cleared of its content of drug in unit time, would be equivalent to the amount of drug that disappears in unit time. In the instance of filtration, it is easy to visualize clearance as that volume filtered/min, since the filtered volume actually is separated physically from the blood. Thus, the *creatinine clearance*, or *GFR*, is equal to the total amount of creatinine found in the urine (equal to urine concentration times urine volume) divided by the plasma concentration.

The general concept of clearance can be applied to the kidney according to the equation

$$Cl_{ren} = \frac{\bar{C}_u V}{\bar{C}_p t} \qquad [mL \cdot min^{-1}] \qquad (14)$$

where Cl_{ren} is renal clearance, $\bar{C}_u$ is concentration in mg/mL of drug in urine collected during time t, V is urine volume in mL generated in time t (min) and $\bar{C}_p$ is the mean concentration (during the collection interval, t) of drug in the plasma in mg/mL; the units are thus mL/min. Urine is collected from the bladder by catheter or by voiding. At the beginning of the collection interval (time from last voiding) both $\bar{C}_p$ and $\bar{C}_u$ are higher than at the end. Consequently, $\bar{C}_p$ must be calculated; $\bar{C}_u$ is automatically the mean of the instantaneous collecting duct concentrations. Equation 14 is valid whether the drug is "cleared" by filtration or by tubular secretion and whether or not tubular resorption occurs. If the drug is protein-bound, the formula becomes

$$Cl_{ren(corr)} = \frac{\bar{C}_u V}{\bar{C}_p t(1 - p)} \qquad [mL/min] \qquad (15)$$

where $Cl_{ren(corr)}$ is the corrected renal clearance.

The ratio between Cl_{ren} and Cl_{creat}, Cl_{ren}/Cl_{creat} (or Cl_{ren}/Cl_{inulin}), is known as the *clearance ratio*. If the drug is protein-bound and the *corrected clearance* is used, the ratio $Cl_{ren(corr)}/Cl_{creat}$ is known as the *excretion ratio*.

If an unbound drug is filtered only and not resorbed, the excretion ratio will be 1 and the clearance about 125 mL/min; if the drug subsequently is resorbed, the excretion ratio will be less than 1 and the clearance will lie between 125 and 1 mL/min, the values depending upon the degree of resorption. A clearance of 1 mL/min suggests distribution and elimination like those of water. If there is tubular secretion (plus obligatory filtration), the excretion ratio may exceed 1, and the clearance could be as high as 600–700 mL/min, depending upon the extent of tubular secretion and resorption. *Para*-aminohippuric acid (PAHA) is not bound to plasma protein, is not tubularly resorbed and is secreted so fast by the renal tubules that the plasma passing through the kidney is 90% cleared of PAHA. Thus Cl_{PAHA} is

equal to 0.90 *RPF*. This is called the *effective renal plasma flow, ERPF*. The excretion ratio of PAHA is about 5 to 6.

Eq 14 can be rearranged so that

$$\frac{\bar{C}_u}{t} = \frac{\bar{C}_p Cl_{ren}}{v} \qquad [\text{wt} \cdot \text{vol}^{-1} \cdot \text{min}^{-1}] \qquad (16)$$

Thus, it may be seen that the concentration of drug in newly formed urine is directly proportional to the plasma concentration. Since the plasma concentration falls exponentially during the collection interval, t, it follows that the instantaneous urine concentration in the collecting ducts, likewise, must fall exponentially and hence the rate of fall can be expressed by a first-order rate constant, k_u. This constant relates to renal clearance as follows:

$$k_u = \frac{Cl_{ren}}{V_d} \qquad [\text{min}^{-1}] \qquad (17)$$

The excretory rate constant may be simple, as with a drug like creatinine, or compound, as with a drug that is secreted tubularly and/or resorbed.

The overall renal elimination constant, k_r, is defined by

$$k_r = k_g + k_{ts} - k_{tr} \qquad [\text{min}^{-1}] \qquad (18)$$

where k_g is the constant for glomerular filtration, k_{ts} for tubular secretion and k_{tr} for tubular resorption. Although k_r might be thought to be the same as k_u on page 729, in practice it is not, because clearance data are obtained from time-averaged concentrations and cannot provide instantaneous rates. However, creatinine-derived k_r is close to the instantaneous k_u at the midpoint of the collection period.

By combining Eqs 3 and 17 and assuming that there is no other route of elimination,

$$t_{1/2} = 0.693 \frac{V_d}{Cl_{ren}} \qquad [\text{time}] \qquad (19)$$

The units of time must be the same for both $t_{1/2}$ and Cl_{ren}. The equation enables the calculation of some thought-provoking information about the biological half-lives of nonmetabolized drugs of different excretion profiles and volumes of distribution. Approximate hypothetical half-lives of drugs of different volumes of distribution and renal clearance are shown in Table II. The drugs are assumed to be eliminated only by renal excretion. A volume of distribution of 50 L is that of total body water, 15 L is that of extracellular water and 50,000 L is that of a drug strongly bound in the tissues. Because of biotransformations, few drugs have half-lives longer than 1 yr. However, a few radioopaque iodine-containing diagnostic agents are so tightly bound that their half-lives exceed 1 yr. At the other extreme, a half-life of 15 min by renal elimination is uncommon, because few drugs that are totally cleared have volumes of distribution as small as that of extracellular water. However, the half-life of penicillin G is about 30 min.

Although data from collected urine cannot provide instantaneous rates, it does allow the calculation of the plasma half-life. The instantaneous excretion rate, dD_u/dt (where D_u is the amount of drug in urine), is directly proportional to the body burden, D_B, such that

$$dD_u/dt = k_u D_B \qquad [\text{wt} \cdot \text{min}^{-1}] \qquad (20)$$

But, D_B is falling exponentially with a rate constant k, so that $D_B = D_B^0 e^{-kt}$; therefore, $dA/dt = k_u D_B^0 e^{-kt}$. It follows that the slope of a plot of the log of the excretion rate versus time will have a slope of $-0.434k$, analogous to Eq 2 (adapted to total content rather than concentration). The y intercept of such a plot is $\log k_u = D_B^0$, where D_B^0 is the amount of drug in the body at zero time. However, data on excretion rates require renal catheterization and are subject to considerable error. An alternative, usually more accurate, method of estimating k from urine concentration is to employ the cumulative amount excreted. In this method,

$$D_u = \frac{D_B k_u}{k}(1 - e^{-kt}) \qquad [\text{wt}] \qquad (21)$$

Since k_u/k expresses the proportion of D_B being transferred to the urine, $D_B^0 k_u/k$ represents the total amount of drug excreted, or D_u^∞, where ∞ designates infinite time. Eq 21 in log form, with the above substitution and transposition, becomes

$$\log (D_u^\infty - D_u) = \log D_B^0 \frac{k_u}{k} - 0.434kt$$
$$= \log D_u^\infty - 0.434kt \qquad [\text{no units}] \qquad (22)$$

The slope of the plot against time is also $-0.434k$ and $(D^\infty - D_u)$ is the amount of drug that remains in the body. The equation applies if the drug is administered intravascularly. This is known as the sigma minus method (sigma for the integral D_u^∞ and minus for the $-D_u$). Urine needs to be collected for only 3 or 4 half-lives in order for the semilog plot to yield a reliable slope and $t_{1/2}$. The method is useful especially when plasma concentrations are low.

Hepatic Clearance—The concept of hepatic clearance is like that of renal clearance, and hepatic clearance is likewise a hypothetical volume of blood per min imagined to be totally cleared of drug during passage through the liver. Unlike renal clearance, the input is both portal venous and hepatic arterial blood and the output is both hepatic venous blood and bile, rather than arterial blood and urine, respectively. Portal venous blood and bile cannot be sampled readily, so that the concepts involved in hepatic clearance serve better to provide a model for understanding the role of the liver in pharmacokinetics than a clinical methodology for its direct measurement.

Although the mathematical treatment of hepatic clearance has been developed for steady-state conditions, rather than for exponentially falling drug concentrations in the inputs and outputs to the liver, the subject is appropriate at this place, in conjunction with other clearances.

The *hepatic clearance*, Cl_H, can be defined by the equation

$$Cl_H = HBF\left(\frac{C_{ap} - C_v}{C_{ap}}\right) = HBF \cdot E \qquad [\text{mL} \cdot \text{min}^{-1}] \qquad (23)$$

Table II—Hypothetical Half-Lives of Drugs of Differing Volumes of Distribution and Clearances

Drug No.	Distribution	V_d L	Renal Disposition	Clearance, mL/min	Half-Life
1	Total body water	50	Filtered and resorbed with water	1	24 days
2	Total body water	50	Filtered, no resorption	125	4.67 hr
3	Total body water	50	Tubular secretion, total clearance	700	50 min
4	Extracellular water	15	Tubular secretion, total clearance	700	15 min
5	Strongly bound in tissues	50,000	Filtered and resorbed with water	1	66 yr
6	Strongly bound in tissues	50,000	Tubular secretion, total clearance	700	35 days

where HBF is the total hepatic blood flow, C_{ap} the hypothetical mean of mixed hepatic arterial and portal venous concentrations and C_v is the hepatic venous concentration. The ratio, $(C_{ap} - C_v)/C_{ap}$, is the *extraction ratio, E*. Unlike glomerular filtration, there is an upper limit to the absolute quantity of drug that can be cleared and hence to the extraction ratio. Extraction is flow-limited only so long as the biotransforming enzyme system is not approaching saturation. The maximal clearance in the presence of normal blood flow has been called the *total intrinsic clearance*, Cl_{intr}. The extraction ratio expressed in terms of Cl_{intr} is

$$E = \frac{Cl_{intr}}{HBF + Cl_{intr}} \qquad \text{[no units]} \tag{24}$$

which may be substituted into Eq 23, to yield

$$Cl_H = HBF\left(\frac{Cl_{intr}}{HBF + Cl_{intr}}\right) = HBF \cdot E \qquad [\text{mL} \cdot \text{min}^{-1}] \tag{25}$$

The intrinsic clearance becomes

$$Cl_{intr} = \frac{HBF \cdot E}{1 - E} \qquad [\text{mL} \cdot \text{min}^{-1}] \tag{26}$$

Cl_{intr} is thus somewhat analogous to V_{max}/K_m in enzyme kinetics.

Eqs 23 through 26 emphasize that hepatic clearance and extraction are functions both of hepatic blood flow and the capacity of hepatic enzymes to biotransform (or secrete into bile) the drug that is delivered. In order to appreciate the relative dependencies on Cl_{intr} and HBF, various assumed values may be substituted into the equations. What will be found is that the larger the Cl_{intr}, the more Cl_H tends to be flow-limited (ie, dependent upon the rate of delivery of blood), whereas when Cl_{intr} is small, Cl_H is metabolism-limited. At constant blood flow with a drug in which elimination is predominately hepatic, when intrinsic clearance and hence extraction ratios are small, a significant change in intrinsic clearance will be accompanied by a significant change in $t_{1/2}$; when intrinsic clearance is high, a significant change may be accompanied by a small, often insignificant, change in $t_{1/2}$ but a significant decrease in bioavailability. In the latter instance, $t_{1/2}$ is determined mostly by the fraction of the cardiac output that passes through the liver. Figures illustrating these features, an excellent discussion of hepatic clearance and various models of hepatic elimination are available,[5] as well as a treatment of the effect of binding of drug to plasma protein. Binding to plasma protein limits clearance when intrinsic clearance is low but not when it is high.

Although the determination of Cl_{intr} is too involved for routine investigative purposes, it may be estimated according to the equation

$$Cl_{intr} = \frac{\left(1 - \frac{D_{un}}{fD}\right)D}{AUC_0} \qquad [\text{mL} \cdot \text{min}^{-1}] \tag{27}$$

where D_{un} is the total quantity of drug excreted unchanged, f is the fraction absorbed, D is the dose administered and AUC_0 is the total area under the blood concentration-time curve after intravenous administration. The meaning of AUC will be discussed later (page 737).

Some drugs may be used to illustrate some of the points emphasized by the model. For example, at blood concentrations of ethanol above 0.02–0.04%, the hepatic alcohol dehydrogenase system is saturated, and hence hepatic blood flow will have little effect on Cl_H of ethanol above the concentration indicated. This implies that liver disease or injury will not much affect the rate that ethanol is cleared from the blood, a fact of some forensic importance. The hepatic biotransformations of pentobarbital and phenytoin are relatively slow, ie, Cl_{intr} are low; consequently, the induction of hepatic cytochrome P450 will increase Cl_H, almost in proportion to the degree of induction, and the $t_{1/2}$ will be shortened accordingly. The hepatic biotransformation of lidocaine is extremely rapid, ie, the Cl_{intr} is very high, so that Cl_H is limited by hepatic blood flow. This means that by the oral route, in which all of the absorbed drug obligatorily passes through the liver, only very small amounts will survive the pass through the liver into the systemic circulation. This nearly total clearance as the drug passes through the liver into the rest of the body is known as the *first-pass effect*. The clinical significance of the first-pass effect is discussed in Chapter 37. The flow-limitation in the hepatic metabolism of lidocaine also means that in congestive heart failure or shock, in which hepatic blood flow is diminished, the rate of biotransformation will decrease and $t_{1/2}$ will increase.

Biliary secretion contributes to hepatic clearance and hence is included in the above pharmacokinetic considerations. However, drugs that are excreted intact or in a form from which the drug can be sequestered in the intestines and subsequently resorbed (enterohepatic recirculation) may have complex pharmacokinetics if the rate of biliary secretion is an appreciable fraction of the hepatic clearance and if the enterohepatic reservoir is large.

Other Routes and Clearances—The kidney and liver are usually the major organs in the elimination of drugs, and all other routes combined often contribute negligibly. However, with the volatile anesthetics, pulmonary clearance is the major route, and pulmonary clearance becomes dominant; pulmonary clearance of gases is flow-limited. With some drugs, mammary secretion is appreciable, and the presence of drug in milk may present hazards to nursing children; however, pharmacokinetics in the mother usually is not affected by lactation. Salivary secretion is too small to affect systemic pharmacokinetics, but the concentration of drug in saliva usually parallels that in plasma, so that, with some drugs, it is possible to follow systemic pharmacokinetics by sampling saliva.

Absorption Plus Elimination—The kinetics of absorption and disposition must now be put together to define the time-related curve which describes the plasma concentration of a drug administered *extra*vascularly. The curve is determined by the algebraic sum of all processes involved in absorption, distribution and elimination. Since disposition (distribution plus elimination) begins as soon as the drug enters the blood stream, the plasma concentration reflects all these processes from the outset.

The time course of the plasma concentration of a drug in a one-compartment body can be obtained by combining algebraically Eqs 1 and 5, with appropriate rate constants, and substituting fD/V_d for C_p^∞. When the equations for absorption and elimination are thus combined

$$C_p = \frac{fDk_a}{V_d(k_a + k_{el})}\left(e^{-k_{el}t} - e^{-k_a t}\right) \qquad [\text{wt} \cdot \text{vol}^{-1}] \tag{28}$$

where f is the fraction absorbed, D is the dose, etc.

This equation simplifies to

$$C_p = C_p^0 e^{-k_{el}t} - C_p^0 e^{-k_a t} \qquad [\text{wt} \cdot \text{vol}^{-1}] \tag{29}$$

where C_p^0 and the C_p^∞ of Eq 5 are the same, since they both represent all of dose, D, distributed in V_d. If there is a lag, the t-factor in the exponents of e should be $t - t_l$, where t_l is the lag time. Figure 36-7 shows a plot of the plasma concentration for each of absorption and elimination separately and when the two are combined.

In Fig 36-7 the parameters of absorption and elimination were assumed, in order to construct the figure. In practice, drug concentration-time data are obtained empirically, and the parameters are obtained from a semilog plot of the data,

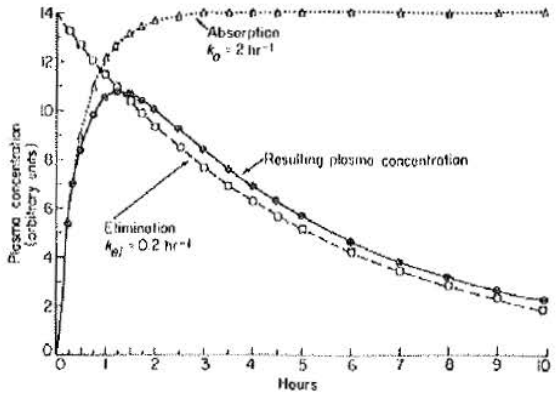

Fig 36-7. Time course of the plasma concentration of a hypothetical drug with simple first-order absorption and elimination kinetics. The rate constants are shown in the figure. The half-time for elimination is 3.47 hr.

as in Fig 36-8. The rising phase of the plot is not log-linear, since that which is added by absorption is diminished by elimination. Only after absorption is complete does the plot become log-linear, since now there is no opposing process at work against the mono-exponential decline in concentration. The time at which absorption essentially is complete is called the *absorption time* and is detected as that time at which the plot becomes log-linear. However, prior to the absorption time, the concentration at the site of deposition becomes equal to that in plasma. This is called the *equilibrium time*. It is also the *peak-time* for plasma concentration. Because of the interplay of physicochemical and active transport factors that affect the distribution of a drug, true chemical equilibrium is not reached necessarily at the pharmacokinetic equilibrium point. The log-linear line described by the elimination phase, when back-extrapolated to the y-axis, yields a theoretical C_p^0, just as with intravascular injection, and $V_{d(extrap)}$ can be calculated accordingly. Furthermore, the slope of the log-linear elimination segment of the semilog plot is equal to $-0.434k_{el}$, as with intravascular injection. The absorption rate constant, k_a, also can be obtained from the plot, if the empirical curve is subtracted from the back-extrapolated elimination line. This is done by subtracting the real values for C_p (C_1, C_2, etc) at various times during the absorption phase from the extrapolated values for C_p, designated C', on the back-extrapolated elimination line. It must be remembered that the antilog and not the log of C must be used if *log* C is plotted in Cartesian coordinates. This method of dissecting a compound function into its separate components is known as the *method of residuals*, or *back-feathering*. The back-feathered *absorption* line is the dashed line; its slope is negative, as though it were being seen from the site of administration.

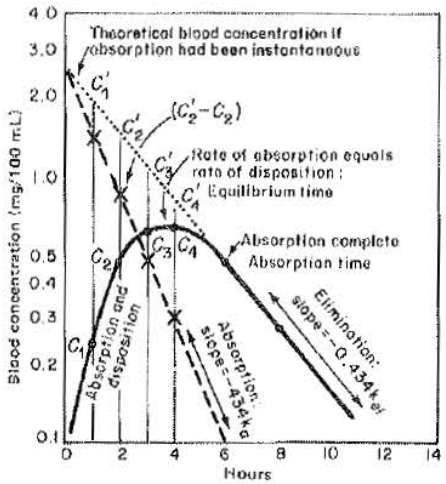

Fig 36-8. Kinetics of absorption and disposition of theophylline in the human subject after oral administration of 0.5 g of aminophylline per 70 kg. Blood concentration is plotted on a log scale (courtesy, data, Truitt, *et al* [3]).

The *peak concentration*, *time of peak concentration* and *duration of action* are affected by various factors, some of which are discussed below.

Peak Concentration—That the peak concentration should vary with the dose is self-evident; according to Eq 28, it is directly proportional to the dose (assuming that absorption and elimination are first-order processes). Figure 36-9 shows how peak concentration varies directly with dose. Note that the time of peak concentration is the same for all doses; this independence of peak time from dose is approximately true in all multicompartment systems. Departures from the generalization occur especially when the rate of absorption or elimination is different at high from those at low concentration; ie, when it is dose-dependent (see page 744).

Time of Peak Concentration—The time of peak concentration must not be confused with the time of peak effect. Effect often lags behind plasma concentration, sometimes because the tissue concentration at the point of action has not yet reached its peak and sometimes because a response may have a considerable latency. The latency of effect of reserpine or phenytoin (in its anticonvulsant effect) is measured in hours to days. Occasionally, the time of peak effect may precede the time of peak concentration because of a reflex or other compensatory process which limits effect before the concentration becomes maximal. This is often true with oral administration of ethanol or ephedrine. Both the peak concentration and time of peak concentration are considerably affected by the rate constants for absorption and elimination. In Fig 36-10, the effect of differences in absorption rate is shown indicating that the higher the absorption rate, the higher the peak concentration and the earlier the time of peak concentration. Figure 36-11 shows the effect of differences in rate of elimination depicting that the higher the elimination rate, the lower the peak concentration but the earlier the time of peak concentration. The two effects of absorption rate and elimination rate can be

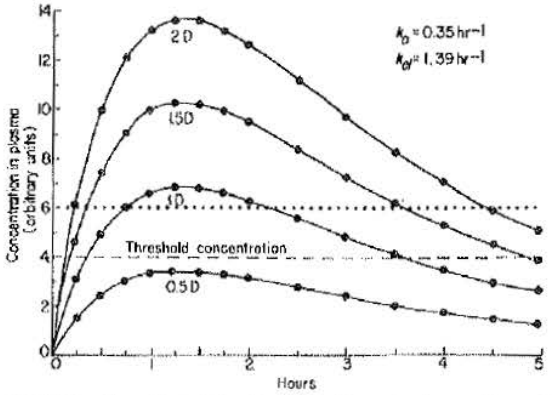

Fig 36-9. The effect of the size of the dose of a drug on the peak concentration, time of peak concentration and duration of action. The data are calculated from a one-compartment model.