2 1ST EDITION

Remington

The Science and Practice
of Pharmacy

Remington: The Science and Practice of Pharmacy . . . *A treatise on the theory and practice of the pharmaceutical sciences, with essential information about pharmaceutical and medicinal agents; also, a guide to the professional responsibilities of the pharmacist as the drug information specialist of the health team . . . A textbook and reference work for pharmacists, physicians, and other practitioners of the pharmaceutical and medical sciences.*

AUTHORS

The 133 chapters of this edition of *Remington* were written by the editors, by members of the Editorial Board, and by the authors listed on pages xi to xv.

Director

Philip P Gerbino 1995-2005

Twenty-first Edition—2005

Published in the 185th year of the
PHILADELPHIA COLLEGE OF PHARMACY AND SCIENCE

Drug Absorption, Action, and Disposition

Michael R Franklin, PhD
Donald N Franz, PhD

Although drugs differ widely in their pharmacodynamic effects and clinical applications; in penetration, 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 differences among the members of that class.

For a drug to act it must be absorbed, transported to the appropriate tissue or organ, penetrate to the responding cell surface or 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 Figure 57-1. 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'être*, *drug action*, and *effect* are discussed first in the text that follows, even though they are preceded by other events.

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 to a noxious stimulus in a certain way, called pain. (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.) 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 64 through 89 and are not 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 that 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 region of the hypothalamus, which regulates the release of antidiuretic hormone from the posterior pituitary gland. The alteration of hypothalamic 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.
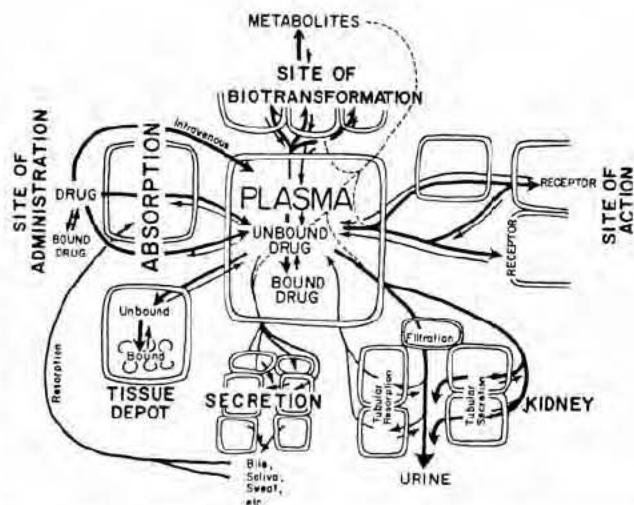
Figure 57-1. The absorption, distribution, action, and elimination of a drug (arrows represent drug movement). Intravenous administration is the only process by which a drug may enter a compartment without passing through a biological membrane. Note that drugs excreted in bile and saliva may be resorbed.

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 phenylephrine that accrue to its structural similarity to norepinephrine and its ability to exert sympathetic responses, 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 immunological (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 that subserve the function of sweating than it does for the nicotinic receptors 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 inactivating 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 that 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 systems involved in the inactivation and elimination of 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, and 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 61). 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 not a rigid quantity but rather that which is necessary and can be tolerated and individualizes the regimen accordingly. The wise pharmacist also recognizes that official or manufacturer's recommended doses are sometimes quite narrowly defined and 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 prototypic 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 important 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
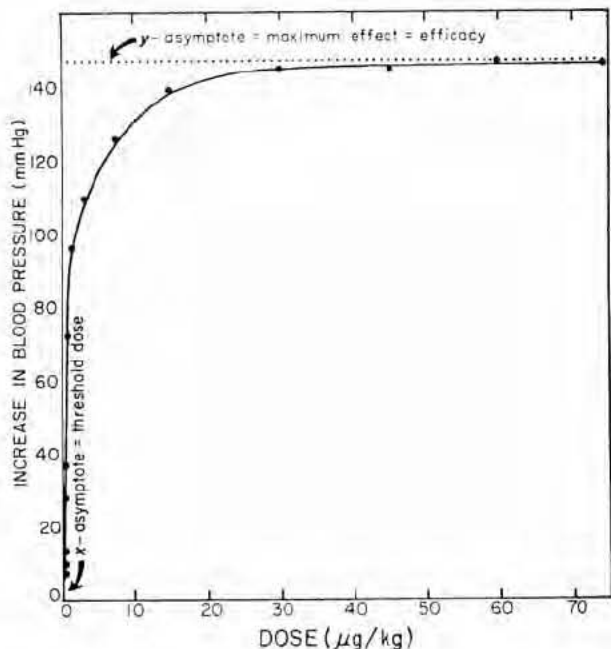
Figure 57-2. The relationship of the intensity of the blood-pressure response of the cat to the intravenous dose of norepinephrine.

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 the-

ory 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 Figure 57-2, 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 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 that 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 very low and very high doses. The example shown represents an unusually good set of data.

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 57-3 is such a semilogarithmic plot of the same data used in Figure 57-2. In the figure the intensity of effect is plotted both in absolute units (at the left) or in relative units, as percentages (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
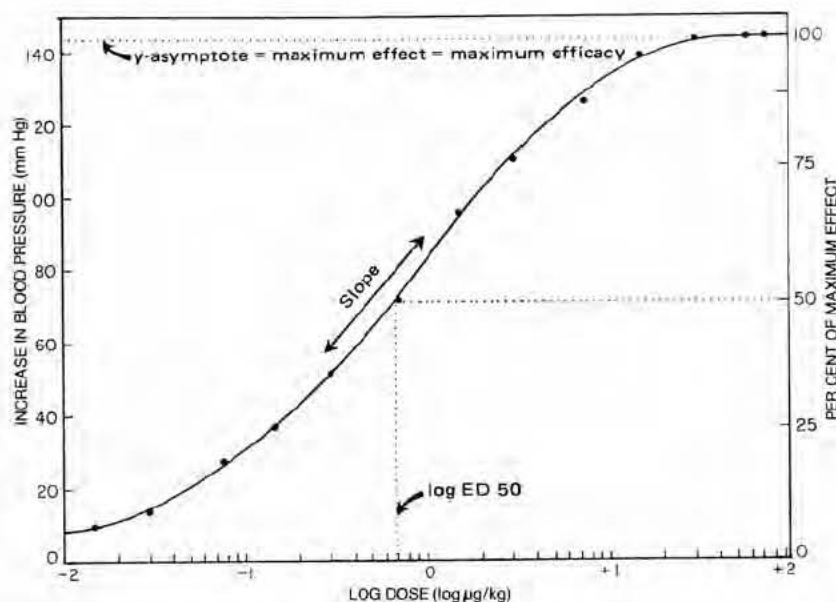

Figure 57-3. The relationship of the intensity of the blood-pressure response of the cat to the log of the intravenous dose of norepinephrine.

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 that 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 (see Fig 57-2). 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*).

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.

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. 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 that elicits a given intensity of effect. The intensity of effect that is generally designated is 50% of maximum intensity. The corresponding dose is called the *50% effective dose*, or *individual ED50* (see Fig 57-3). 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 chapter).

Drugs that elicit the same quality of effect may be compared graphically. In Figure 57-4, 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*) 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.

The relative potency of any drug may be obtained by dividing the ED50 of the standard, or prototypic, drug by that of the drug
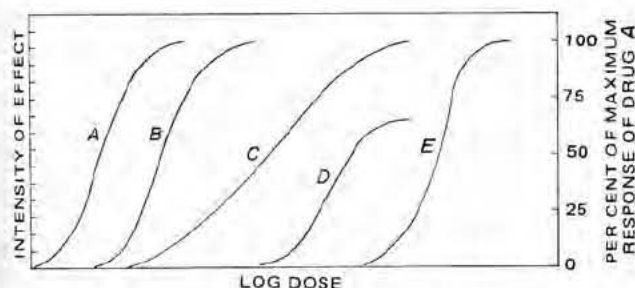
in question. Any level of effect other than 50% may be used, but it should be recognized that when the slopes are not parallel, the relative potency depends upon the intensity of effect chosen. Thus, the potency of *A* relative to *C* (see Fig 57-4) calculated from the ED50 will be smaller than that 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 percentage 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 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 percentage of maximum effect; thus, all curves can share a common scale of ordinates in addition to a 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 percentage 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 and that for another effect, or a ratio at some other intensity of effect. As with 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. 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 large enough 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) endpoint and plot the frequency or cumulative frequency of response over the dose range, as discussed below.

DOSE–FREQUENCY OF RESPONSE RELATIONSHIPS—When an endpoint 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 endpoint. If the endpoint is taken as an increase in heart rate of 20 beats/min, all individuals whose tachycardia is less than 20 beats/min would be recorded as nonresponders, while all those with 20 or above would be recorded as responders. When the percentage of responders in the population is plotted against the dose, a characteristic dose-response curve, more
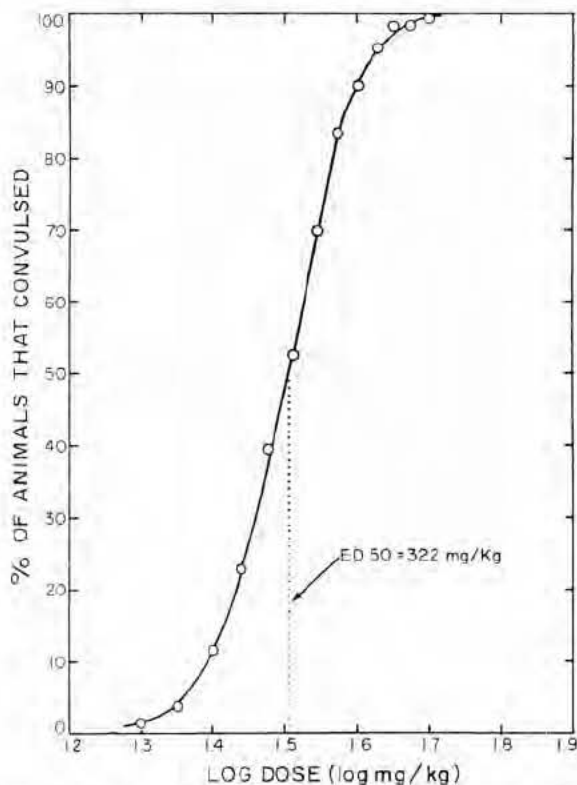

Figure 57-4. Log dose–intensity of effect curves of five different hypothetical drugs (see text for explanation).

Figure 57-5. The relationship of the number of responders in a population of mice to the dose of pentylenetetrazole.

properly called a *dose–cumulative frequency* or *dose-percentage* curve, is generated. Such a curve is, in fact, a cumulative frequency-distribution curve, the percentage 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 (Fig 57-5). The tendency of the cumulated frequency of response (ie, percentage) 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 percentage 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.

If only the increase in the number of responders with each new dose is plotted, instead of the cumulative percentage 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. The distribution will be symmetrical—ie, *normal* or Gaussian (see Chapter 12)—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 57-5). If the frequency distribution is normal, the ED50 is both the arithmetic mean and the 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 the test population. Even two drugs of identical mechanism may give rise to different slopes in dose-percentage 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 12).

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 12. 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 percentage response may be either the percentage of maximum intensity or the percentage of a population responding. The equation is thus basically the same for both log normal dose-intensity and log normal dose-percentage 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-percentage relationships among drugs is subject to the pitfalls indicated for dose-intensity comparisons, namely, that when the slopes of the curves are not the same (ie, the dose-percentage 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 percentage response. The expression of selectivity is, likewise, subject to similar qualifications, inasmuch as the dose-percentage 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 T (for toxic) and the therapeutic effect, 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 percentage 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 percentage incidence of a given side effect.

VARIATIONS IN RESPONSE AND RESPONSIVENESS—From the above discussion of dose-frequency relationships and Chapter 12, 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-percentage 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 Figure 57-5, 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 percentage would respond to 0.7 of the ED50. The physician who is unfamiliar with statistics is apt to consider the 2.25% at either extreme to be abnormal reactors. The statistician will argue that these 4.5% are within the normal population and that only those who respond well outside 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 system (CNS) 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 from those 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 *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 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 CNS, and ACh receptors elsewhere, are coupled to ion channels that, when opened, permit the rapid ingress, especially of sodium ions. Each ion channel is composed of five subunits, and each subunit has four transmembrane, spanning regions. GABA (γ-aminobutyric acid) and glycine are coupled to inhibitory chloride channels. Each of these receptors is composed of pentameric proteins, each of which has two to four different types of subunits. Benzodiazepine receptors are coupled to the GABA-receptor. Beta-adrenergic receptors, histamine (H2) receptors, and a number of receptors for polypeptide hormones interact with a stimula-

tory GDP/GTP-binding protein (G-protein) that 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 α_1, some muscarinic (M_1 and M_3), 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 calcium, 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 be shown not only in the types of coupling systems and effects but 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; three each 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. 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. 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 the 20th 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 that 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 be 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 with 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 percentage of maximum response. Clark assumed that y was a linear function of the percentage of receptors occupied by the drug, so that y could also symbolize the percentage 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 Figure 57-2. 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-percentage curves have the same shape shows that they involve similar statistics.

If Equation 2 is put into log form

$$\log K + n \log x = \log \frac{y}{100 - y} \tag{4}$$

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 that 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 equal 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 as the concentration of a drug can. 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 percentage 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 (Fig 57-6). 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 (Fig 57-7).
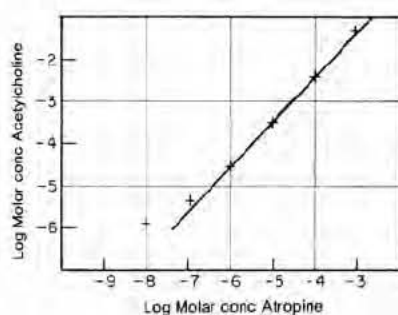
Figure 57-6. Direct proportionality of the dose of agonist (acetylcholine) to the dose of antagonist (atropine) necessary to cause a constant degree of inhibition (50%) of the response of the frog heart. (Adapted from Clark AJ. *J Physiol (London)* 1926; 61:547.)

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 in 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 that the maximal response occurred 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 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.
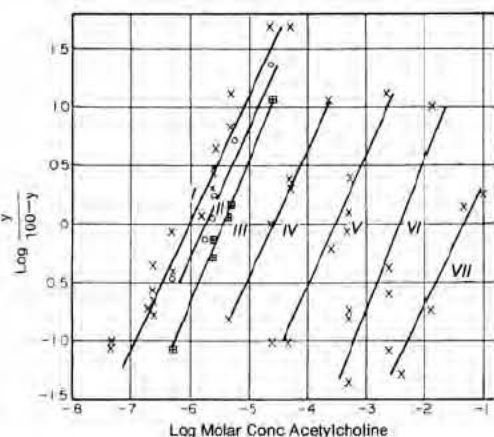

Figure 57-7. Effect of an antagonist to shift the log dose-intensity curve to the right without altering the slope. The effector is the isolated heart. *I:* no atropine; *II:* atropine, 10^{-8} M; *III:* 10^{-7} M; *IV:* 10^{-6} M; *V:* 10^{-5} M; *VI:* 10^{-4} M; *VII:* 10^{-3} M; *Y:* % of maximum intensity of response. The function log $y/(100 - y)$ converts the log dose-intensity relationship to a straight line. (Adapted from Clark AJ. *J Physiol (London)* 1926; 61:547.)

The phenomena that neither the classical occupation nor the 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 that interact cooperatively.[5]

MECHANISMS OF DRUG ACTION

Drugs are distributed to many or all parts of the body by the circulation. However, they do not act everywhere; they would have extremely limited usefulness if they did. Clinically useful drugs act only on certain existing biological systems. Although drugs cannot create new systems, some drugs can temporarily or permanently damage existing functional systems that are susceptible to them, thereby producing toxic effects. Almost all drugs act more or less *selectively* on large specific proteins, glycoproteins, or lipoproteins located on the cell membrane or in the cell cytoplasm, nuclei, or other intracellular organelles. These specific proteins are referred to as *receptors*. Although they often are regarded as drug receptors, they are in reality receptors for *endogenous* substances that mediate normal biological and physiological regulatory processes.

Virtually all cells of the body have multiple receptors, since they are regulated by a variety of endogenous substances that act continuously, intermittently, or only occasionally. Similarly, cells theoretically can be influenced by a variety of drugs that act on the different receptors that the cells contain. The chemical nature of many of the endogenous substances that activate receptors is known, but new ones continue to be identified and sought. For example, the former mystery of why animals have receptors for morphine, which is produced by some species of poppy plants, was solved when endogenous opioid peptides were identified in the brain and some peripheral tissues in the mid-1970s.

Drugs that selectively activate receptors and produce the same *effects* normally produced by a respective endogenous substance are called *agonists*. Drugs that selectively block receptors are called *antagonists* because they antagonize, or block, the normal effects of the respective endogenous substance. Pure antagonists do not activate their receptors. Some experimental drugs stimulate or activate certain enzymes, but none are useful therapeutic agents because their effects are too widespread. Forskolin is one such example; it directly stimulates the enzyme adenylyl cyclase to synthesize cyclic AMP, which is a second messenger in many cellular systems throughout the body.

On the other hand, many very useful therapeutic drugs are *enzyme inhibitors*, which selectively inhibit the normal activity of only one type of enzyme, thereby reducing the ability of the enzyme to act on its normal biochemical substrate. In this context, the enzymes are the drug receptors. Although the chemical nature of receptors and enzymes and their interactions with drugs was often vague in the past, the application of new techniques in molecular biology, biochemistry, and pharmacology since the mid-1980s has made unprecedented progress in defining the structures of receptors and enzymes and the consequences of drug-receptor interactions.

TYPES OF TARGETS FOR DRUG ACTION

Drug *effects* are the result of drug *actions*. Drug action may be defined as the drug-receptor interaction, whereas drug effects are the consequences of that action. For example, the interaction of epinephrine with β-receptors in the heart sets into mo-

tion a cascade of intracellular events (actions) that lead to increases in heart rate and strength of contraction (effects). The interaction of epinephrine with α-receptors in the vasculature sets into motion a cascade of intracellular events (actions) that lead to vasoconstriction and increased blood pressure (effects).

Typical responses that involve drug-receptor interactions are those that involve agonist or antagonist interactions at a receptor. Agonists also can act through various transduction mechanisms to produce a variety of intracellular changes that alter cellular activity. Transduction mechanisms are considered in more detail near the end of this section. Agonist actions may be direct, as with acetylcholine acting on the nicotinic receptors at the neuromuscular junction to briefly open sodium channels. This produces rapid depolarization of skeletal muscle, leading to muscle contraction. Drugs also can act directly on ion channels to block their activity. For example, lidocaine (Xylocaine) and other local anesthetics block sodium channels in nerve fibers (axons) so that the conduction of action potentials is blocked, and the area served by those nerve fibers is anesthetized. Drugs also can act directly on ion channels to modulate their activity. The benzodiazepines, characterized by diazepam (Valium), produces multiple effects (sedation, hypnosis, anticonvulsant and antianxiety activity, and muscle relaxation) by *modifying* the actions of GABA on its receptors in the CNS. GABA is the predominant inhibitory neurotransmitter in the CNS, and it acts on GABA$_A$-receptor complexes by opening chloride channels on neurons to hyperpolarize them and render them less excitable. The benzodiazepines act on a different receptor on the GABA$_A$-receptor complex to enhance the actions of GABA on its receptors, thereby rendering target neurons even less excitable.

Many drugs act by inhibiting enzymes so that they cannot perform their normal functions as efficiently. One such drug, omeprazole (Prilosec), reduces the ability of parietal cells in the stomach to produce hydrochloric acid by inhibiting the enzyme, or proton pump, H$^+$, K$^+$-ATPase, which is found only in these parietal cells. It is used to facilitate healing of peptic ulcers and control esophageal reflux (heartburn). The body's normal enzymes also can convert false substrates into active drugs. For example, α-methyldopa (Aldomet) is converted into α-methylnorepinephrine by the enzymes that normally synthesize dopamine and norepinephrine from dopa. α-Methylnorepinephrine acts on brain receptors to reduce sympathetic activity to blood vessels, thereby reducing blood pressure in hypertensive patients. Antimetabolites used to treat cancer are also false substrates, which are similar in structure to endogenous metabolites involved in cell-cycle reactions but function abnormally to interfere with synthesis of essential metabolites. Some drugs are, or have been, designed to be inactive until they are converted, usually by liver drug-metabolizing enzymes such as cytochrome P450, to active drug; the inactive drug is called a *prodrug*.

Various *carriers* are used by cells to take up neurotransmitters that have been released. The actions of dopamine released from dopamine nerve terminals in the brain are terminated by reuptake into the nerve terminals by a dopamine carrier. The dopamine then is reused for neurotransmission. If the carrier is blocked by a reuptake blocker such as cocaine, dopamine concentrations between the nerve terminals and the dopamine receptors build up for a time and produce greater effects.

Finally, antibiotics and antiviral, antifungal, and antiparasitic drugs owe their selectivities to selective actions on certain biochemical processes that are essential to the offending organism but are not shared by the mammalian host. The penicillins and related antibiotics interfere with the synthesis of rigid cell walls by growing bacteria, but mammalian cells are contained only by plasma membranes and, therefore, are not affected by penicillins. Antiparasitic drugs target enzymes found only in parasites, enzymes that are indispensable only in parasites, or biochemical functions with different pharmacological properties in the parasite and the host.

RECEPTOR BINDING

Drugs that bind to certain receptors selectively at pharmacological concentrations are known as *receptor ligands;* they can be agonists or antagonists. Many drugs also bind nonselectively to nonreceptor proteins throughout the body where they exert no pharmacological actions or effects. Many drugs bind to plasma proteins, especially albumin. Albumin-bound drug can act as a reservoir for free drug, with which it is in equilibrium, and competition among drugs for plasma protein binding can lead to increased free drug levels and drug interactions as they displace one another.

Drugs and endogenous ligands or substrates bind selectively to certain receptors because of both a chemical attraction and a proper *fit* to the protein. The lock-and-key analogy provides a useful concept of proper fit. Carried a step further, an agonist fits the lock and turns it, but an antagonist only fits the lock but cannot turn it; yet, it does block entry of the agonist key. Generally, a number of drugs with both characteristics can combine with the same receptor. The study of structure-activity relationships among similar drugs and their receptors always has been an important and fruitful approach of both pharmacology and medicinal chemistry. Highly selective drugs tend to bind to only one or several closely related receptors. However, some drugs can combine with and activate or inactivate a number of different receptors that have similar structures, thereby diminishing selectivity and magnifying side effects.

The types of chemical bonds by which drugs bind to their receptors are, in decreasing order of bond strength: covalent, ionic, hydrogen, hydrophobic, and van der Waals bonds. Relatively few drugs form covalent bonds with their receptors. Covalent bonds are *irreversible* and very long-lasting; new receptors or enzymes must be synthesized to restore function, and this process takes a week or two. Most drugs rely on combinations of the other weaker bonds to bind tightly but *reversibly* to receptors. For example, the binding of acetylcholine, a relatively simple molecule, to nicotinic receptors at the neuromuscular junction, involves ionic, hydrogen, and van der Waals bonds, with ionic and hydrogen bonds being the most important. It is no accident that receptor-binding drugs are partially ionized at body pH, because receptor proteins also are partially ionized. Drugs and proteins contain positively charged nitrogen groups and negatively charged carboxyl groups that strongly attract one another and usually provide the initial drug-receptor bonds. Hydrogen bonds, formed between bound hydrogen atoms and oxygen, nitrogen, fluorine, or sulfur atoms, further orient the drug molecule to its receptor to enhance the proper fit. One or several hydrogen bonds can be involved. Hydrophobic bonds form among nonpolar ring structures (eg, benzene) or chains of methylene groups to stabilize orientation further. Finally, the very weak van der Waals forces provide some additional, electrostatic bonding over very short distances.

Drug molecules that contain asymmetrical carbon atoms can exist as stereoisomers, only one of which is oriented to bond well with its receptors. For example, the side chain of epinephrine contains an asymmetrical carbon atom in the alpha position of the side chain, with a hydroxyl group attached, permitting epinephrine to exist in D- and L- forms (mirror images). The endogenous L-form is about 1000 times more potent than the synthesized D-form because the L-form has a much greater binding affinity for its receptors because of its preferred configuration (see Chapter 28). In the past, drugs synthesized as mixtures of stereoisomers were formulated as racemic mixtures, but improved chemical separation techniques now often allow isolation of the more active isomer for formulation.

RECEPTOR STRUCTURE AND FUNCTION

The number of receptors and their subtypes continues to grow at a rapid pace as a result of identifying new endogenous ligands and applying advancing techniques to study them. De-

spite this large number, most receptors can be classified structurally and functionally into only a few basic types that are described below. No attempt is made to provide detailed descriptions of individual receptors within each category. Rather, one or two examples will suffice for each, with brief reference to some prominent types that are therapeutically relevant.

VOLTAGE-SENSITIVE CHANNELS—While not generally classified as receptors, voltage-sensitive channels contain receptors that are acted upon by drugs or toxins to block or modify their normal function. The voltage-sensitive sodium channels in axons allow initiation and conduction of action potentials in response to a voltage change in the plasma membrane. When sodium channels open, sodium ions rush into the cytoplasm, thereby causing depolarization and propagation of the action potential. The crucial component of the sodium channel is a single protein composed of a chain of about 2000 amino acids and called the α subunit. Several β subunits with minor roles are also associated with the α subunit. The α subunit has four repeating domains composed of about 250 amino acids each, and each domain contains six, α-helical, 22–to 25–amino acid, transmembrane, spanning segments. Each domain forms one of four clusters of the six membrane-spanning regions to encircle the sodium channel so formed. On end, the channel resembles 24 cylinders neatly arranged around the sodium channel that, at rest, is charged positively due to positive charges on the four transmembrane helices that surround the channel. Upon activation, these particular helices are thought to rotate upward, thereby moving the positive charges away from the channel and allowing the positive sodium ions to rush through. The channel remains open for only about 1 msec because the voltage changes attract a protein loop of the channel in the cytoplasm to shut the channel like a tether ball. Local anesthetics block the sodium channel from the cytoplasmic side by binding to receptors inside the channel. Several neurotoxins block from the outside.

Axons are repolarized by brief (~1 msec) opening of voltage-activated potassium channels that are constructed similarly to sodium channels but are composed of four identical subunits of peptide that associate in the membrane to form the potassium channel. Each subunit spans the membrane six times. It probably functions much like the sodium channel, including inactivation by a tether-ball segment of cytoplasmic peptide. Quinidine, an antiarrhythmic drug, will block this potassium channel in the heart.

Voltage-activated calcium channels of the L-type are composed of five similar protein subunits that assemble across heart muscle and vascular smooth muscle membranes to form the calcium channel. Its arrangement in the membrane is similar to that of the sodium and potassium channels. Calcium channel blockers such as verapamil *(Calan)* and nifedipine *(Procardia)* are used to treat several cardiovascular conditions by virtue of their ability to block calcium channels in the heart and blood vessels.

LIGAND-ACTIVATED ION CHANNELS—The best-characterized ligand-activated ion channel is the nicotinic receptor complex at the neuromuscular junction. As the name implies, these channels are activated by receptor ligands, in this case acetylcholine. The nicotinic receptor complex is composed of five subunit proteins with similar structures that associate across the plasma membrane to form a sodium channel. The receptor complex is formed from two α and one each of β, γ, and δ subunits (Fig 57-8). In contrast to the voltage-activated ion channels, each of the five proteins crosses the membrane only four times. The two α subunits contain the nicotinic receptors, which acetylcholine activates, and both must be activated to open the sodium channel to 6.5 Å for about 4 msec. The receptors can be blocked by neuromuscular blocking agents such as curare. The nicotinic receptors on autonomic ganglia are similar in structure but are composed of a different set of subunits, which accounts for the long-known differences in selective antagonists at the two sites.
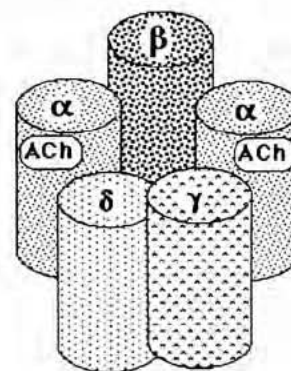
Figure 57-8. Nicotinic receptor complex.

Other ligand-activated ion channels, $GABA_A$, glycine, and glutamate, have structures that are similar to that of the nicotinic receptor complex. GABA and glycine channels are chloride channels, which permit chloride influx into neurons to produce hyperpolarization and decreased neuronal excitability. Glutamate channels are primarily sodium channels, and they also contain modifying receptors for glycine and polyamines. The $GABA_A$-receptor complex contains receptors not only for GABA but also separate receptors for benzodiazepines (eg, *Valium*), barbiturates, and steroids, which modify the actions of GABA on the chloride channel. The convulsant activity of strychnine is due solely to its ability to block glycine receptors, primarily in the brainstem and spinal cord.

G PROTEIN–COUPLED RECEPTORS—These receptors comprise a very large family of receptors that are activated by monoamines (epinephrine, norepinephrine, dopamine, and serotonin), acetylcholine (muscarinic receptors), opioids, and a host of active peptides including a number of hormones. Structurally, these receptors are single proteins, most of which are composed of chains of 350 to 550 amino acids and cross the plasma membrane seven times in a *serpentine* arrangement (Fig 57-9). Each
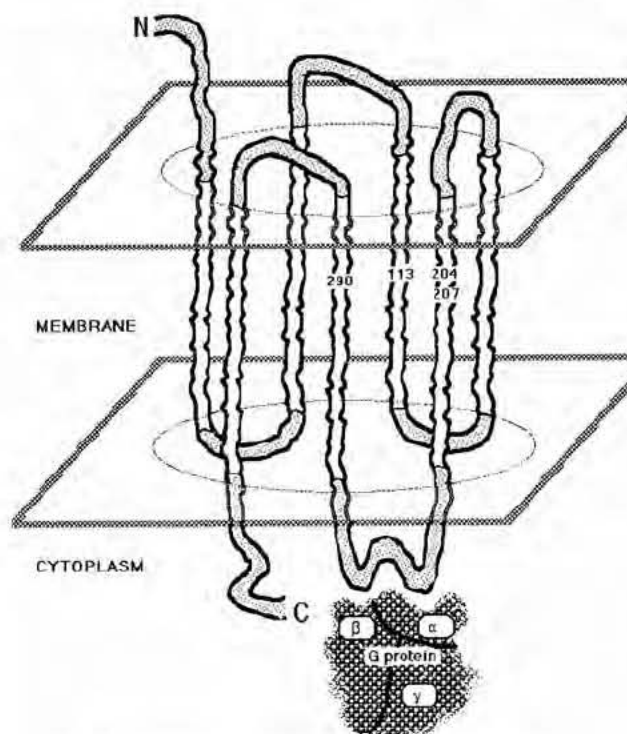
Figure 57-9. G-Protein coupled receptor complex.

of the seven transmembrane domains is composed of 22 to 30 amino acids configured into an α-helix. The third of three intracellular (cytoplasmic) loops is much longer than the other two and is responsible for coupling with the G proteins. Rather than residing at the extracellular surface of the receptor, the actual receptor-binding sites often lie *within* the membrane between the seven transmembrane domains. For example, the β-adrenergic receptor lies 11 Å below the extracellular surface, or about one-third of the distance through the membrane. The positively charged nitrogen on the side chain of the epinephrine molecule forms an ionic bond with the negatively charged carboxyl group on an aspartate amino acid (residue 113) in the third transmembrane domain (TM3). The two catechol hydroxyl groups of epinephrine form hydrogen bonds with the free hydroxyl groups of two serine amino acids at residues 204 and 207 in TM5, and the aromatic ring of epinephrine forms a hydrophobic bond with that of a phenylalanine at residue 290 in TM6. The location of G protein-coupled receptors within the membrane underscores the importance of *size and configuration* in the molecular structure of both agonists and antagonists for these receptors. Some negatively charged and peptide ligands do bind to an extracellular domain, however.

Among some families of G protein-coupled receptors there is considerable structural homology; ie, the same amino acids and the same sequences make up large portions of a number of different receptors. Consequently, a number of antagonist receptor ligands bind to these similar arrangements of amino acids in the transmembrane domains. For example, many of the antipsychotic drugs (neuroleptics) are antagonists not only at dopamine receptors, where they are thought to exert their therapeutic effects, but also at α1-adrenergic, serotonin, histamine, and muscarinic receptors, thereby producing hypotension, sedation, blurred vision, dry mouth, and constipation as side effects.

The G proteins closely associated with the third cytoplasmic loop of the receptors are heterotrimers composed of three different subunits, α, β, and γ. Upon receptor activation, the α subunit exchanges a bound GDP for a GTP and dissociates from the βγ subunits to activate a membrane enzyme such as adenylyl cyclase or to influence an ion channel. In some cases, the βγ subunits may interact with the same or a different intracellular effector. The duration of action of the active GTP-α subunit is determined by the hydrolysis of GTP to GDP by a GTPase, which is intrinsic to the α subunit, and its reassociation with the βγ subunits. This process is of longer duration than the association of the ligand with the ligand-G protein-coupled receptor, resulting in *amplification* of the original signal.

In the case of adenylyl cyclase activation, this enzyme synthesizes cyclic adenosine-3',5'-monophosphate (cAMP) from ATP. As a *second messenger*, cAMP then goes on to activate one or several protein kinase As that phosphorylate one or several other proteins to produce the appropriate cellular effects. The targeted protein may be an enzyme, a transport protein, a contractile protein, or an ion channel. The specificity of these regulatory effects depends on the distinct protein substrates that are expressed in different cells (eg, liver vs smooth muscle). The actions of cAMP are terminated by several types of intracellular phosphodiesterases that convert cAMP to 5'-AMP. Competitive inhibition of phosphodiesterases to prolong the actions of cAMP is one of the mechanisms by which caffeine produces its effects.

As if the foregoing is not sufficiently complicated, the activity of adenylyl cyclase can also be inhibited by activation of different G protein-coupled receptors. The G proteins coupled to inhibitory receptors are designated Gi proteins, as opposed to those coupled to stimulatory receptors and designated Gs proteins. Gi proteins are also heterotrimers, and receptor activation of Gi also leads to GTP binding to the α subunit and its dissociation from the βγ, but Gi proteins differ structurally from Gs proteins. Examples of Gs-coupled receptors are β-adrenergic, dopamine-1, histamine-2, glucagon, and ACTH. Examples of Gi-coupled receptors are α2-adrenergic, dopamine-2, mus-

carinic, and opioid. A number of different Gs and Gi protein-coupled receptors can exist on the same cell, so that the activity of adenylyl cyclase can be fine-tuned between zero and maximum.

Another important group of G protein-coupled receptors activate the enzyme phospholipase C (PLC) to hydrolyze a minor component of the plasma membrane, phosphatidylinositol-4,5-biphosphate, into two second messengers, diacylglycerol (DAG) and inositol-1,4,5-triphosphate (IP3). In contrast to the cAMP systems, receptors coupled to PLC are only excitatory. Examples are α1-adrenergic, muscarinic, Substance P, and thyrotropin-releasing hormone receptors. The second messenger DAG is confined to the membrane, where it activates a protein kinase C, of which nine distinct types have been identified. The other second messenger, IP3, diffuses through the cytosol to release calcium from intracellular stores. Calcium is involved in many cellular regulatory activities including activation of calcium-calmodulin, which regulates the activities of other enzymes including other kinases. The kinases in turn phosphorylate enzymes, ion channels, or other proteins to produce cellular effects. When the phosphoinositide and cAMP signaling systems coexist, they can oppose or complement one another in complex ways.

A third second-messenger system uses cyclic guanosine-3',5'-monophosphate (cGMP) in intestinal mucosa and vascular smooth muscle. It is synthesized from GTP by activation of guanylyl cyclase and activates protein kinase G, which then dephosphorylates myosin light chains in vascular smooth muscle, thereby producing muscle relaxation. Agonists, eg, acetylcholine and histamine, cause the release of nitric oxide from vascular endothelial cells, which then diffuses into the smooth muscle cells to activate guanylyl cyclase. A direct receptor-mediated activation is produced by atrial natriuretic factor (ANF), a blood-borne peptide hormone. In this case, the receptor domain is outside the membrane and is connected through a single transmembrane domain to the intracellular guanylyl cyclase enzyme, which is activated by receptor binding.

TYROSINE KINASE-LINKED RECEPTORS—These receptors are composed of an extracellular receptor domain, a *single* transmembrane domain, and an intracellular catalytic domain that catalyzes phosphorylation of tyrosine residues on target proteins. Some receptors are composed of single proteins, whereas others are assembled from two subunits (eg, insulin receptors). Activation of insulin receptors triggers increased uptake of glucose and amino acids and regulates metabolism of glycogen and lipids in the cell. The catalytic actions persist for a number of minutes after insulin leaves the binding site. Several growth factors also exert their complex cellular effects by activating tyrosine kinase or similar receptors. Growth factors trigger changes in membrane transport and other metabolic events including regulation of DNA synthesis.

INTRACELLULAR RECEPTORS THAT CONTROL DNA TRANSCRIPTION—Activation of intracellular receptors for steroids (glucocorticoids, mineralocorticoids, sex steroids, vitamin D) and thyroid hormones stimulates the transcription of certain genes by binding to specific DNA sequences in the nucleus. The receptors generally are composed of a single protein with a ligand-binding domain, a DNA-binding domain, and a transcription-activating domain. In the inactivated state, the receptor protein is bound to another protein, a heat shock protein (hsp 90), which dissociates upon activation by a hormone, permitting DNA binding and transcription of mRNA, which then is translated into new protein. This process typically takes several hours, and the effects can last for days or weeks if there is a slow turnover of the newly synthesized proteins. A similar process accounts for the induction of drug-metabolizing enzymes in the liver by certain drugs and other chemicals. In this process, formation of a heterodimeric complex between a second protein and the ligand-bound receptor is required for DNA binding.

ENZYME INHIBITION—Enzymes are very large, complex proteins or associated proteins that evolved to catalyze specific

biochemical reactions that are essential to normal cellular function. A number of very selective drugs exert their effects by inhibiting particular enzymes, so that their abilities to process their normal substrates are blocked or impaired. Enzyme inhibitors can produce competitive blockade at a substrate or cofactor binding site on the enzyme. For example, the stimulant effect of digitalis glycosides on cardiac muscle contraction is mediated by competitive inhibition of a sodium pump, Na^+,K^+-ATPase, which leads indirectly to an increase in intracellular calcium to interact with contractile proteins. Other enzyme inhibitors act noncompetitively at allosteric sites (sites remote from the substrate binding site), which prevent the enzyme from performing its catalytic function. For example, aspirin binds irreversibly to a site on cyclooxygenase that is remote from the binding site for arachidonic acid, which is normally converted to prostaglandins by the enzyme. The binding of related drugs such as ibuprofen (*Advil*) is reversible. Irreversible inhibition by the formation of covalent bonds between a drug and an enzyme is typically long lasting because new enzyme must be synthesized to restore function.

RECEPTOR REGULATION—The regulation of receptor numbers or density is normally constant, as synthesis keeps pace with degradation of the proteins. However, continuous stimulation of receptors with agonists can lead to desensitization or *down-regulation* of receptor sensitivity or number. Desensitization can occur rapidly without a change in receptor number, whereas down-regulation usually implies a decline in receptor number. For example, excess use of β-adrenergic agonists for treating bronchial asthma can lead to loss of receptor sensitivity to the agonist, caused by changes in coupling mechanisms to the G proteins. Chronic blockade of receptors can lead to *up-regulation*, which, in some cases, is due to synthesis of new receptors. An example is chronic blockade of β-adrenergic receptors in the heart, in which new β-receptors are synthesized, leading to supersensitivity upon abrupt withdrawal of the blocker. Another form of supersensitivity is demonstrated by denervation of skeletal muscle, which is followed by a proliferation of nicotinic receptors within and adjacent to the neuromuscular junction.

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, this section begins with a brief description of biological membranes and membrane processes and the relationship of the physicochemical properties of a drug molecule to penetration 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 physicochemical properties of cells seemed in accord with this view. Microchemical, x-ray diffraction, electron microscopic, nuclear magnetic resonance, electron spin resonance, and other investigations have established the 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 drug penetration into and through membranes.

STRUCTURE AND COMPOSITION—The cell membrane has been described as a bimolecular layer of lipid material entrained between two parallel monomolecular layers of protein. However, the protein does not make continuous layers, but rather is sporadically scattered over the surfaces, like icebergs; ie, much of the protein is below the surface. In Figure 57-10 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 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 co-

extensive. 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, most often to the asparagine residue. 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, most 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
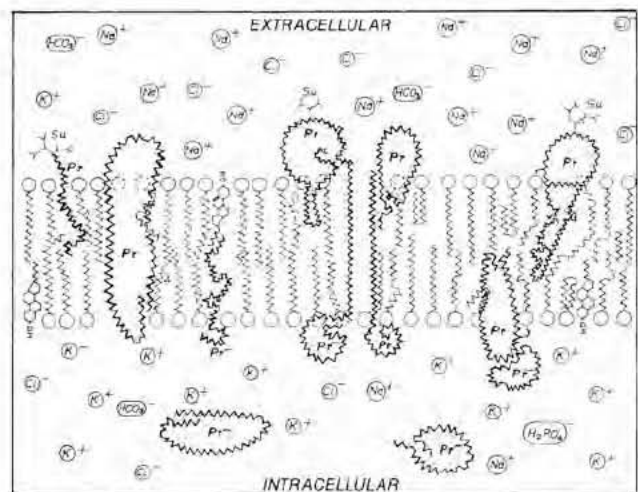

Figure 57-10. Simplified cross section of a cell membrane (components are not to scale). The lipid interior of the lamellar portion of the membrane consists of various phospholipids, fatty acids, cholesterol, and other steroids. Ions are indicated to illustrate differences in size relative to tein; Su, sugar.

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 α-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, which 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 Figure 57-10. 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 upward 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 be 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,

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

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 that on the other, x is the thickness of the membrane, A is the area, and $\overline{D}$ is the diffusion coefficient, related to permeability. Since a biological membrane is heterogeneous, with pores of different sizes and probably with varying thickness and composition, both $\overline{D}$ and x probably vary from place to place. 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 $P = \overline{D}/x$, and A in Equation 5 has unit value. The rate of net transport (diffusion) across the membrane then becomes

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

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 also is 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 toward 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°), ionized drug molecules will be distributed according to the Nernst equation,

$$\pm \log \frac{C_o}{C_i} = \frac{ZE}{61} \tag{7}$$

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 acceler-

ated 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 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 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

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 blood, from intestinal cells into the lumen of the GI tract, 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 documented. Active transport is required for the penetration 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 with most active transport, except in GI absorption, in which pinocytosis can be 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 PENETRATION

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 differ 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, although 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 penetration 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 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 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 penetration 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 penetration unless the water solubility is low enough so that the drug is not entrained in the aqueous phase.

In Figure 57-11 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, virtually are 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, penetrate poorly. 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 10 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*).

It often is 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 even though they are ionized completely at
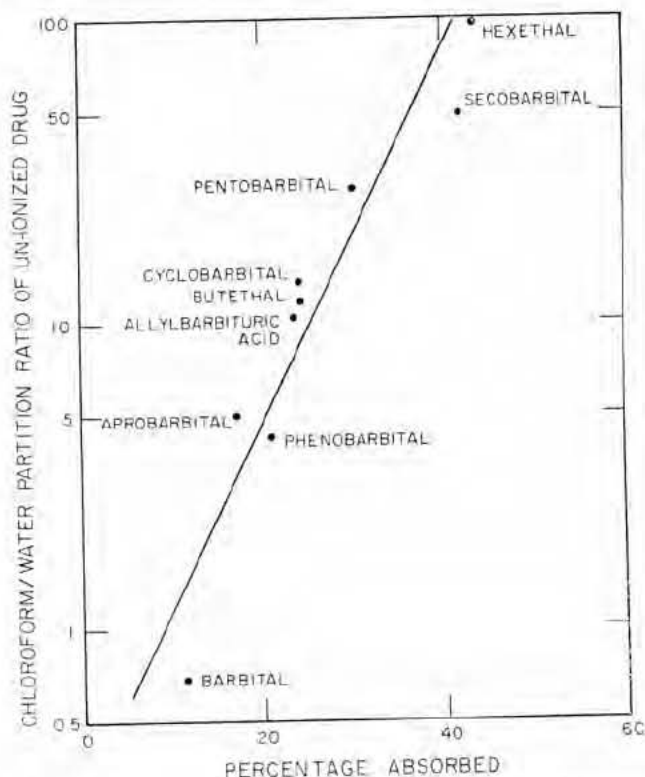
Figure 57-11. The relationship of absorption of the un-ionized forms of drugs from the colon of the rat to the chloroform:water partition coefficient. (From Schanker LS. *Adv Drug Res* 1964; 1:71.)

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 Figure 57-11.

For the purpose of further illustrating the principle, Table 57-1 is provided.[7] In the table, the permeability constants for penetration into the cerebral spinal fluid of rats are higher for un-ionized drugs than for ionized ones. The apparent excep-

tions—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 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 GI intolerance or who are in preparation for anesthesia or who have had GI 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

Table 57-1. Rates of Entry of Drugs in CSF and the Degrees of Ionization of Drugs at pH 7.4[7]

DRUG/CHEMICAL	% BINDING TO PLASMA PROTEIN	PK$_a$[a]	% UN-IONIZED AT pH 7.4	PERMEABILITY CONSTANT (P 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 the pK$_a$, the 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.
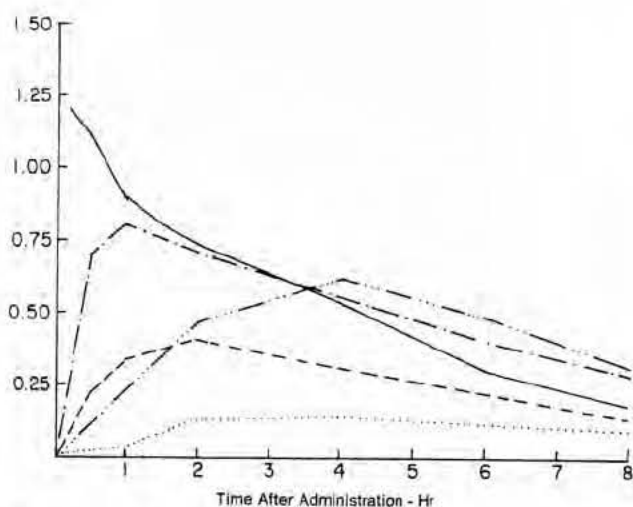
Figure 57-12. Blood concentration in mg/100 mL of theophylline (ordinate) following administration to humans of aminophylline in the amounts and by the routes indicated. Doses; per 70 kg. Theophylline-ethylenediamine by various routes:——intravenous, 0.5 g;—·—·—retention enema, 0.5 g;—•·—·•·•—oral tablets-PI, 0.5 g; - - - oral tablets-PI, 0.3 g; •••• rectal suppository, 0.5 g. (Adapted Truitt EB, et al. *J Pharmacol Exp Ther* 1950; 100:309.)

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 Figure 57-12[8] the availability of a drug by retention enema may be compared with that by the intravenous and oral routes 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 when 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,[9] 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 when 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-GI (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 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 when 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 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, 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 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 that 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 that is too large to be in rapid diffusion equilibrium with the blood. *Intracavitary* administration, thus, may be required. It is extremely important that sterile, 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 that 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 larger than 1 μm in diameter tend to settle in the bronchioles and bronchi, whereas particles smaller 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 investigations with aprotic solvent vehicles such as dimethyl sulfoxide (DMSO) also have 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, 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, 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 padministration 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 GI 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 GI 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 always show 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 that 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 38 and dissolution, Chapter 35) or retard it (eg, coatings, Chapter 46 and slow-release formulations, Chapter 47). There also are 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 47).

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 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 e large alveolar surface of the lungs allows

extremely rapid absorption of gases, vapors, and properly aerosolized solutions; with some drugs the rate of absorption may be nearly as fast as with 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 predominantly in the small intestine despite a pH partition factor that should favor absorption from the stomach (see *The pH Partition Principle*).

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 that 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 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/mm^2 of cross-section, a drug molecule would not have to travel more than a few microns, hence less than a second on 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. Overall, 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 hyperemia 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, on World War II battlegrounds 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 about 10 min, though it varies according to its contents, reflex, and psychological factors, and 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 depends 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 from 5.1, and the complex can be used as a low-solubility, prolonged-action drug. For more details, see Chapter 77.

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 retard 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 GI tract 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. Accordingly, drug molecules of the same charge as the macromolecule will be constrained to the opposite side of the mem-
brane. 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 GI 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 penetration. 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 five times as permeant from water as from olive oil. Yet, ethanol penetrates five times faster from olive oil than from either water or ethanol, all of which denies the trustworthiness of generalizations about vehicles.

For several decades 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 pre-sent only experimental, pending continuing 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 that 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 is considered under *Disposition*. Some authors use the term disposition synonymously with elimination, that is, to include only those processes that 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* 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 (GI), 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 was 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 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 rapid biotransformation occurs within, some compartments, uneven distribution also is 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, which it occasionally becomes. The pK_a of salicylic acid is 3 (Martin[10] provides one source of pK values of drugs). With the Henderson-Hasselbalch equation (see Chapter 17) it may be calculated that the drug is only 1% ionized at pH 1. (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]) The intracellular pH of most cells is about 7. 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 that in gastric juice. This is illustrated in Figure 57-13. Such a relatively high intracellular concentration can have important osmotic and toxicological consequences.
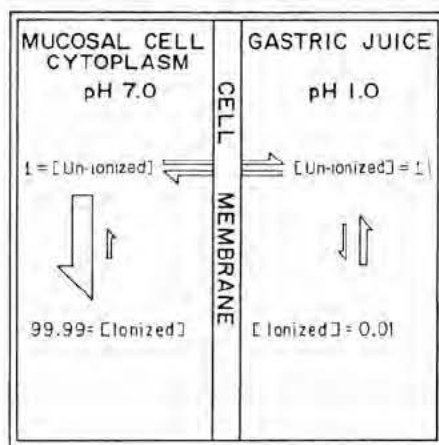

Figure 57-13. Hypothetical partition of salicylic acid between gastric juice and the cytoplasm of a gastric mucosal cell. It is assumed that the ionized form cannot pass through the cell membrane. The intragastric concentration of salicylic acid is arranged arbitrarily to provide unit concentration of the un-ionized form. *Bracketed values*, concentration; *arrows*, relative size depicts the direction in which dissociation-association is favored at equilibrium.

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 lower, because the pH differential from lumen to mucosal cells, etc, will be lower. 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 GI 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 Figure 57-13 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) 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*, 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 20. According to the theory, the ratio of 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. Equation 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 (especially albumin), interstitial substances,
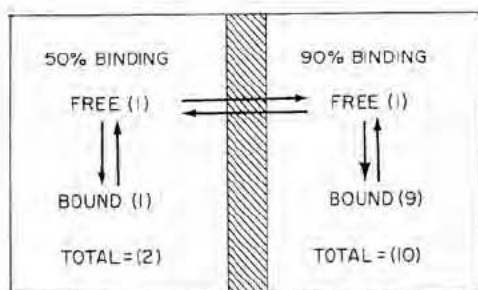
Figure 57-14. Distribution of a drug between two compartments in which the degrees of binding to protein differ. The percentage of binding is indicated. Only the unbound drug can pass through the membrane. *Bracketed values:* concentration. (From Schanker LS. *Pharmacol Rev* 1961; 14:501.)

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:

$$\log D_b = \log K + a \log D_f$$

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, but the percentage bound is often known.

From the percentage bound, the partition also can be calculated, as in Figure 57-14.[12] However, the logarithmic relationships shown in Equation 9 serve as a reminder that the percentage 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 Ks and as 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 are 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 level sufficient 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 be 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 bound 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 are 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 CNS and causes sometimes fatal toxicity.

Depending on the lipid-water partition 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. As the brain concentration falls, anesthesia ceases. Gradually, 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 Figure 57-15.
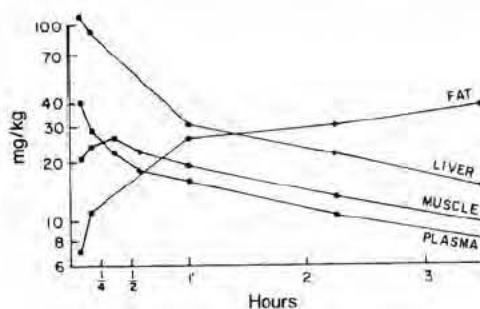

Figure 57-15. Predisposition of thiopental for fat; 25 mg/kg was given to a dog. After a brief sojourn in the more vascular tissues, thiopental gradually transfers to fat, where the lipid-soluble drug dissolves in fat droplets. (From *J Pharm Pharmacol* 1957; 9:345.)

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 biotransformation-excretion, neither a true equilibrium among the body compartments nor a steady state exist.

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 Figure 57-15. 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 Figure 57-15 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 biotransformation 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.

The word *space* often is 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 of a drug with 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 penetration into tissues and less renal tubular resorption than the parent drug, in accordance with the principle of the low penetration of polar and high penetration of lipid-soluble substances. For similar reasons, metabolites, particularly conjugates, are usually less active than the parent drug and 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.

There are also examples in which the parent drug has little or 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.

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. Biotransformations in plasma are mostly hydrolytic.

ENDOPLASMIC RETICULUM AND MICROSOMAL SYSTEM—Many biotransformations in the liver occur in the *endoplasmic reticulum*. The endoplasmic reticulum is a tubular system that 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. The smooth endoplasmic reticulum lacks such a granular appearance. The endoplasmic reticulum 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 often the practice to refer to drug metabolism as occurring in microsomes rather than in the 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 P450. The drug-cytochrome complex then is reduced by NADPH–cytochrome P450 reductase; the reduced complex then combines with oxygen, after which the metabolite is released and oxidized cytochrome P450 is regenerated. Cytochrome P450 is a generic term for a superfamily of enzymes.[17]

The general designation of the cytochromes P450 is *CYP* followed by number (the family) and letter (the subfamily) subdivisions. The classification is based on amino acid sequence homology. To belong to the same family, the homology must be greater than 40% and to the same subfamily greater than 59%. The form is indicated by a number that is based upon the chronological discovery order. The major human forms involved in drug metabolism are CYP1A1 and CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9/10, CYP2C18/19, CYP2D6, CYP2E1, CYP3A4, CYP3A5, and CYP3A7. In concentration, CYP3As comprise 40% of the liver P450, CYP2Cs comprise 25%, and CYP1A2 about 15%. Despite its limited concentration, ... metabolizes about one-fourth of currently used ...

phism in which 5 to 10% of the population are poor metabolizers. The different isozymes present in humans, together with which drugs they metabolize, are of increasing importance in understanding drug interactions and toxicities and individual responses to standardized doses.

In addition to cytochrome P450, the endoplasmic reticulum contains flavoprotein monooxygenases, which are also responsible for the oxidative metabolism of drugs. The mechanism of oxidation differs from that of cytochrome P450, and their substrate (any drug containing a nucleophilic heteroatom) selectivity is much less. *FMO3* is the major human liver form.

Some of the enzymes of the microsomal system are quite easily *induced;* that is, a drug may increase considerably the activity of the enzyme by increasing the biosynthesis of the enzyme. An increase in the amount of endoplasmic reticulum sometimes occurs concomitantly with enzyme induction.

The mechanism of induction is best documented for polycyclic aromatic hydrocarbon (Ah)-type inducers but is thought to be similar for all agents; however, it involves different receptors, which interact with different regulatory elements on the DNA (Fig 57-16). The cytosol contains proteins that have a high affinity for the inducing agents. In normal drug therapy, the drug (D) enters the liver cell and, if adequately metabolized, is discharged as metabolites. Inefficient clearance from the cell, possibly due to high dosage, results in accumulation (ie, excess), and some is able to bind to the protein, which has a high affinity for the accumulating drug. When the inducing agent binds to its receptor, there is a conformational change (for an Ah receptor, chaperone proteins are displaced) allowing the receptor-inducer complex to translocate into the nucleus, link with additional nuclear factors, and initiate the transcription of mRNA to a limited number of proteins, by binding to DNA regions termed a drug-response element (DRE) (xenobiotic response element for the Ah receptor complex) that activate gene transcription. (For polycyclic aromatic hydrocarbons, the activated genes including specific isozymes of cytochrome P450, glutathione S-transferase, and UDP-glucuronosyltransferase.) These mRNA molecules move out of the nucleus and are translated into new proteins on the ribosomes attached to the endoplasmic reticulum.

The drug-metabolizing enzymes differ in their ability to be induced. For cytochrome P450s, CYP1A2 is induced preferentially by polycyclic aromatic hydrocarbons and other chemicals contained in cigarette smoke and charcoal-broiled meats, as well as by components in cruciferous vegetables. CYP2A6 is induced by barbiturates as are CYP2C9 and CYP3A4. CYP2C9, CYP2C19, and CYP3A4 are all induced by rifampicin, but CYP3A4 is additionally induced by many drugs including carbamazepine, phenytoin, glucocorticoids (dexamethasone), clotrimazole, sulfinpyrazone, and macrolide antibiotics such as troleandomycin. CYP2E1 can be induced by ethanol and isoniazid. There are no known inducers of CYP2D6.

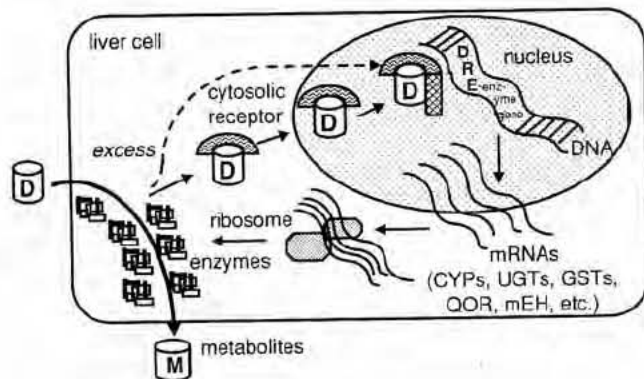Treatment of an experimental subject with phenobarbital will increase the rate of metabolism of phenobarbital, which necessitates larger and more frequent doses of the drug to maintain a constant sedative effect. Moreover, phenobarbital may induce an increased metabolism of some other, but not all, barbiturates as well as some unrelated drugs, such as strychnine and warfarin. Oddly, warfarin does not induce its own biotransformation readily.

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 enzymes, 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 biotransformation enzymes is affected by many factors other than the presence of drugs. Age, sex, nutritional states, pathological conditions, and genetic factors are among the influences that have been identified. Age, particularly, has received considerable attention. Infants have a poorly developed microsomal biotransformation 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 biotransformation 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 four main categories: (1) oxidation, (2) reduction, (3) hydrolysis, and (4) conjugation. Oxidation, reduction, and hydrolysis comprise Phase I. Conjugation comprises Phase II.

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.

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.

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.

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 idine occurs primarily in the hepatic microsomes.

Figure 57-16.

The hydrolysis of epoxides, often generated by cytochrome P450 oxidations, to form dihydrodiols is an important detoxification reaction.

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.

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.

Many electrophilic compounds conjugate with the nucleophilic tripeptide, glutathione. Through a series of enzymatic reactions, the γ-glutamyl and glycyl residues are removed, the remaining cysteine conjugate is *N*-acetylated, and the product spontaneously dehydrates to form a mercapturic acid.

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, some occur in the kidney or in other tissues.

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 cytoplasm 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*-adenosylmethionine.

All the drug conjugation reactions are catalyzed by specialized enzymes present in multiple forms. Glucuronidation is catalyzed by UDP-glucuronosyltransferases, *UGTs*, located in the endoplasmic reticulum. UGTs are classified in two major classes, UGT1As and UGT2Bs, based on amino acid homology, but the two classes also differ in substrate selectivity, with UGT1As preferring planar drugs and UGT2Bs preferring bulkier molecules. As with cytochrome P450s, these enzymes are inducible, and the two classes differ in their response to various drugs and other chemicals.

Sulfation is catalyzed by sulfotransferases, *SULTs*, located in the cytoplasm. The many isozymes exhibit substrate selectivity, and some differ in thermal stability. Unlike most major drug-metabolizing enzymes, SULTs are refractory to induction by drugs.

Glutathione conjugations are catalyzed by glutathione-*S*-transferases, *GSTs*, also located in the cytoplasm. The multiple isozymes are designated into four major classes: alpha, mu, pi and theta. The isozymes have relatively low substrate (electrophile) selectivity. Methylation reactions are catalyzed by cytoplasmic *O*-, *N*-, and *S*-methyltransferases, and each exists in multiple forms.

Acetylation is catalyzed by cytoplasmic *N*-acetyltransferases, *NAT*1, and in the liver, *NAT*2. NAT2 exhibits a genetic polymorphism, giving *fast and slow* acetylator phenotypes with differing incidences in various populations (slow is high in Middle Eastern, low in Asian).

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, or gastric juice or from the lungs.

RENAL EXCRETION—The excretory unit of the kidney is called the *nephron* (Fig 57-17). There are several million nephrons in the human kidney. The nephron is essentially 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 or are forced by pressure 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
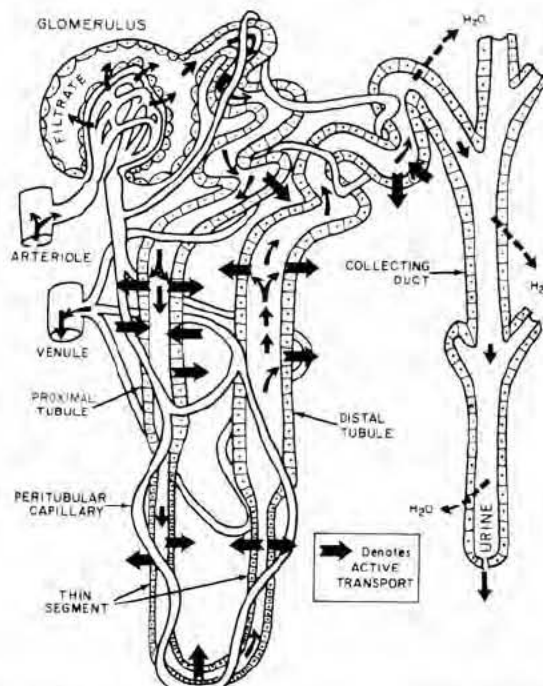

Figure 57-17. Diagram of a mammalian nephron. Note how the lower loops of the postglomerular capillaries course downward and double back along with the tubule. This allows countercurrent distribution to lar urine within the thin segment.

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

Figure 57-19. 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. (From Kostenbauder HB, et al. *J Pharm Sci* 1962; 51:1084.)

of resorption, in accordance with the principle of non-ionic diffusion and pH partition. The pH of the tubular fluid also affects the tubular secretion of drugs.

As an example of the importance of urinary pH, in humans the secondary amine mecamylamine is excreted more than four times faster when the urinary pH is below 5.5 than when it is above 7.5; Figure 57-18 illustrates the effect of urinary pH on the excretion of this amine. The effect of urinary pH on the excretion of a weak acid, sulfaethidole, is shown in Figure 57-19.

The urinary 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 in man have demonstrated the importance of controlling urinary pH. Urinary 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, and 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 58.

BILIARY EXCRETION AND FECAL ELIMINATION— Many drugs are secreted into the bile and then pass into the intestine. A drug that is passed into the intestine via the bile may be reabsorbed and not lost from the body. A drug conjugate entering the intestine may be deconjugated by enzymes and the parent drug reabsorbed. 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*
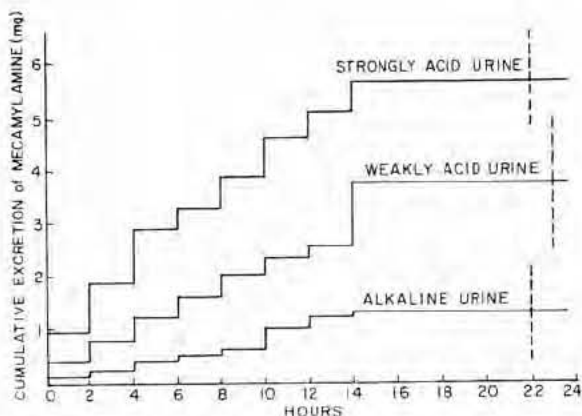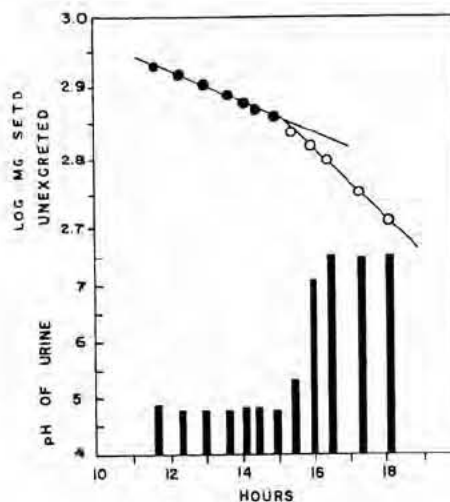
Figure 57-18. The effect of urinary pH on the mean cumulative excretion in man of mecamylamine during the first day after oral administration of 10 mg. *Vertical broken lines:* standard deviation. (From Milne MD, et al. *Clin Sci* 1957; 16:599.)

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 reten-

tion 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 the rate of absorption, extent of absorption, rates of distribution among body compartments, rate of elimination, and related phenomena. Because of its importance. Chapters 58 and 59, *Basic Pharmacokinetics* and *Clinical Pharmacokinetics*, 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 10 drugs, and the patient is often under the influence of several drugs at once. 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 that 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 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 elicited. The mechanisms of such interactions are many and are not always well 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 which also 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 vasopressin 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 Receptor*.

and Receptor Theory, 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 ED_x of Drug A added to an ED_y of Drug B should elicit the same effect as an ED_y of Drug A added to an ED_x of Drug B. An example is shown in Figure 57-20. Dose-percentage 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 penetration 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,
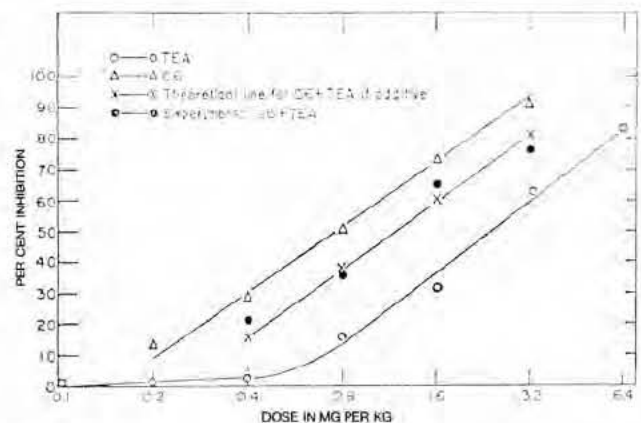

Figure 57-20. Additive inhibitory effects of tetraethylammonium (TEA) and hexamethonium (C6) on the superior cervical ganglion of the cat. The theoretical line for additivity was calculated on the basis that an increment of TEA added to an ED_x of C6 should have the same effect as if it were added to an ED_x 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. (From Harvey SC. *Arch Intern Pharmaco-*

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.

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 only 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 penetration or on the responsivity of the effector system. Examples of a synergistic effect, in which responsivity is enhanced, are the action of adrenal-corticoids 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 that 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 involve 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 drugs 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 RPS-18, 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 pathological 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 1104). 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 ad-

ministration 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 P450s and other drug-metabolizing enzymes enhance the elimination of drugs that are metabolized by the liver. There would be very little point ordinarily in soliciting 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.

Drugs that inhibit cytochrome P450 will, of course, reduce the metabolism of a wide range of additional drugs and serve to prolong or elevate plasma concentration.

COMBINATIONS TO ALTER ABSORPTION—In the section *Vehicles and Absorption Adjuvants*, 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 that 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 generally, it 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 such as 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 patients, who, in an outpatient situation and sometimes in the hospital, may not take all of their medication or 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. An example is given of the grave clinical consequences of the effect of phenobarbital enhancing the biotransformation of warfarin. Other examples of dangerous interactions, such as the effect of several antidepressants in greatly synergizing 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)* 1926; 61:547.
2. Ariens EJ, ed. *Molecular Pharmacology*, vol 1. New York: Academic, 1964, p 176.
3. Stephenson RP. *Br J Pharmacol* 1956; 11:379.
4. Rang HP. *Br J Pharmacol* 1973; 48:475.
5. Colquhoun D. In: Rang HP, ed. *Drug Receptors.* Baltimore: University Park, 1973.
6. Schanker LS. *Adv Drug Res* 1964; 1:71.
7. Brodie BB, et al. *J Pharmacol Exp Ther* 1960; 130:20.
8. Truitt EB, et al. *J Pharmacol Exp Ther* 1950; 100:309.
9. Lillehei JP. *JAMA* 1968; 205:531.
10. Martin AN, et al. *Physical Pharmacy*, 2nd ed. Philadelphia: Lea & Febiger, 1969, pp 247, 253.
11. Jacobs MH. *Cold Spring Harbor Symp Quant Biol* 1940; 8:30.
12. Schanker LS. *Pharmacol Rev* 1961; 14:501.
13. Brodie BB, Hogben CA. *J Pharm Pharmacol* 1957; 9:345.
14. Hogben CA. *Fed Proc* 1960; 19:864.
15. Albert A. *Pharmacol Rev* 1952; 4:136.
16. Ariens EJ, et al. In: *Molecular Pharmacology*, vol 1. Ariens EJ, ed. New York: Academic, 1964, p 7.
17. Nelson DR, et al. *DNA Cell Biol* 1993; 12:1.
18. Milne MD, et al. *Clin Sci* 1957; 16:599.
19. Kostenbauder HB, et al. *J Pharm Sci* 1962; 51:1084.
20. Harvey SC. *Arch Intern Pharmacodyn* 1958; 114:232.
21. Weaver LC, et al. *J Pharmacol Exp Ther* 1955; 113:359.

BIBLIOGRAPHY

Anders MW. *Bioactivation of Foreign Compounds.* New York: Academic, 1985.
Bend JR, Serabjit-Singh CJ, Philpot RM. The pulmonary uptake accumulation and metabolism of xenobiotics. *Annu Rev Pharmacol Toxicol* 1985; 25:97.
Benford D, et al. *Drug Metabolism: From Molecules to Man.* New York: Taylor & Francis, 1987.
Bertolino M, Llinas R. The central role of voltage activated and receptor operated calcium channels in neuronal cells. *Annu Rev Pharmacol Toxicol* 1992; 32:399.
Black JW, et al, eds. *Perspectives on Receptor Classification.* New York: Liss, 1987.

Boelsterli U. *Mechanistic Toxicology: The Molecular Basis of How Chemicals Disrupt Biological Targets.* New York: Taylor and Francis, 2003.
Caldwell J, Jakoby WB. *Biological Basis of Detoxification.* New York: Academic, 1983.
Coulson CJ. *Mechanisms of Drug Action.* New York: Taylor & Francis, 1987.
Dean PM. *Molecular Foundations of Drug-Receptor Interaction.* Cambridge: Cambridge University Press, 1987.
Denison MS, Nagy SR. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals *Annu Rev Pharmacol Toxicol* 2003; 43:309.
Ding X, Kaminsky LS. Human extrahepatic cytochromes P450: function in xenobiotic metabolism and tissue selective chemical toxicity in the respiratory and gastrointestinal tract. *Annu Rev Pharmacol Toxicol* 2003; 43:149.
Finean JB, Michell RH, eds. *Membrane Structure.* Amsterdam: Elsevier/North Holland, 1981.
Gibson GG, Skett PL. *Introduction to Drug Metabolism*, 2nd ed. London: Chapman & Hall, 1994.
Gilman AG. G proteins: transducers of receptor generated signals. *Annu Rev Biochem* 1987; 56:615.
Lewis DFV. *Guides to Cytochromes P450.* New York: Taylor & Francis, 2002.
Gregoriadis G, Senior J. *Targeting of Drugs with Synthetic Systems.* New York: Plenum, 1986.
Hulme EC, Birdsall NJM, Buckley NJ. Muscarinic receptor subtypes. *Annu Rev Pharmacol Toxicol* 1990; 30:633.
Ioannides C, ed. *Cytochromes P450, Metabolic and Toxicological Aspects.* Boca Raton, FL: CRC Press, 1996.
Jakoby WB, et al, eds. *Metabolic Basis of Detoxification.* New York: Academic, 1982.
Kalow W. Pharmacogenetics in biological perspective. *Pharmacol Rev* 1997; 49:369.
Karlin A. Anatomy of a receptor. *Neurosci Comment* 1983; 1:111.
Kenakin TP. The classification of drugs and drug receptors in isolated tissues. *Pharmacol Rev* 1984; 36:165.
Kenakin TP. *Pharmacological Analysis of Drug Receptor Interaction.* New York: Raven Press, 1987.
La Du B, et al. *Fundamentals of Drug Metabolism and Drug Disposition.* Baltimore: Williams & Wilkins, 1971.
Lamble JW, Abbott AC, eds. *Receptors Again!* Amsterdam: Elsevier, 1984.
Lefkowitz RJ, ed. *Receptor Regulation.* London: Chapman & Hall, 1981.
Levine RR. *Pharmacology: Drug Actions and Reactions*, 4th ed. Boston: Little, Brown, 1990.
Limbird LE. *Cell Surface Receptors: A Short Course on Theory and Methods.* Boston: Nijhoff, 1986.
Loh HH, Smith AP, Birnbammer L. Molecular characterization of opioid receptors G proteins in signal transduction. *Annu Rev Pharmacol Toxicol* 1990; 30:123.
Martonosi AN. *Membranes and Transport.* New York: Plenum, 1982.
Meyer UA. Drugs in special patient groups: clinical importance of genetics in drug effects. In: Melman KL, et al, eds. *Clinical Pharmacology*, 3rd ed. New York: McGraw-Hill, 1992.
Mulder GJ, ed. *Conjugation Reactions in Drug Metabolism.* New York: Taylor & Francis, 1990.
Mulder GJ. Glucuronidation and its role in regulation of biological activity of drugs. *Annu Rev Pharmacol Toxicol* 1992; 32:25.
Nguyen T, Sherratt PJ, Pickett CB. Regulatory mechanisms controlling gene expression mediated by the antioxidant response element *Annu Rev Pharmacol Toxicol* 2003; 43:233.
O'Dowd BF, et al. Structure of the adrenergic and related receptors. *Annu Rev Neurosci* 1989; 12:67.
Olson RW, Venter JC, eds. *Benzodiazepine/GABA Receptors and Chloride Channels.* New York: Liss, 1986.
Ortiz de Montellano PR, ed. *Cytochrome P450. Structure, Mechanism, and Biochemistry*, 3rd ed. New York: Kluwer Academic Plenum, 2003.
Parkinson A. Biotransformation of xenobiotics. In: *Casarett and Doull's Toxicology: The Basic Science of Poisons*, 6th ed. New York: McGraw Hill, 2001.
Post G, Crooke ST, eds. *Mechanisms of Receptor Regulation.* New York: Plenum, 1986.
Pratt WB, Taylor P. *Principles of Drug Action.* New York: Churchill Livingstone, 1990.
Putney JW Jr, ed. *Phosphoinositides and Receptor Mechanisms.* New York: Liss, 1986.
Roche EB, ed. *Bioreversible Carriers in Drug Design.* New York: Pergamon, 1987.
Roth SH, Miller KW, eds. *Molecular and Cellular Mechanisms of Anesthetics.* New York: Plenum, 1986.

Sandler M, ed. *Enzyme Inhibitors As Drugs*. Baltimore: University Park, 1980.

Schmucker DL. Aging and drug disposition. *Pharmacol Rev* 1985; 37:133.

Schou JS, et al, eds. *Drug Receptors and Dynamic Processes in Cells*. New York: Raven, 1986.

Stein WD. *Transport and Diffusion Across Cell Membranes*. Orlando: Academic, 1986.

Stoughton RB. Percutaneous absorption of drugs. *Annu Rev Pharmacol Toxicol* 1989; 29:55.

Stroud RM. Acetylcholine receptor structure. *Neurosci Comment* 1983; 1:124.

Sueyoshi T, Negishi M. Phenobarbital response elements of cytochrome P450 genes and nuclear receptors. *Annu Rev Pharmacol Toxicol* 2001; 41:123.

Testa, B, ed. *Advances in Drug Research*, vols 14, 15. London: Academic, 1985, 1986.

Thummel KE, Wilkinson GR. In vitro and in vivo drug interactions involving human CYP3A. *Annu Rev Pharmacol Toxicol* 1998; 38:389.

Triggle DJ, Janis RA. Calcium channel ligands. *Annu Rev Pharmacol Toxicol* 1987; 27:347.

Tukey RH, Strassburg CP. Human UDP-glucuronosyltransferases: metabolism, expression and disease. *Annu Rev Pharmacol Toxicol* 2000; 40:581.

Venter JC, Harrison LC, eds. *Molecular and Chemical Characterization of Membrane Receptors*. New York: Liss, 1984.

Wardle EN. *Cell Surface Science in Medicine and Pathology*. New York: Elsevier, 1985.

Yamazaki M, Suzuki H, Sugiyama Y. Recent advances in carrier mediated hepatic uptake and biliary excretion of xenobiotics. *Pharmacut Res* 1996; 13:497.

Zaki Y, et al. Opioid receptor types and subtypes: the δ receptor as a model. *Annu Rev Pharmacol Toxicol* 1996; 36:379.