SECTION I BASIC PRINCIPLES

CHAPTER 1

Introduction

Bertram G. Katzung, MD, PhD

CASE STUDY

A 26-year-old man is brought by friends to the emergency department of the city hospital because he has been behaving strangely for several days. A known user of methamphetamine, he has not eaten or slept in 48 hours. He threatened to shoot one of his friends because he believes this friend is plotting against him. On admission, the man is extremely agitated, appears to be underweight, and is unable to give a coherent history. He has to be restrained to prevent him from walking out of the emergency department and into traffic on the street. His blood pressure is 160/100 mm Hg, heart rate 100, temperature 39°C, and respirations 30/min. His arms show evidence of numerous intravenous injections. The remainder of his physical examination is unremarkable. After evaluation, the man is given a sedative, fluids, a diuretic, and ammonium chloride parenterally. What is the purpose of the ammonium chloride?

Pharmacology can be defined as the study of substances that interact with living systems through chemical processes, especially by binding to regulatory molecules and activating or inhibiting normal body processes. These substances may be chemicals administered to achieve a beneficial therapeutic effect on some process within the patient or for their toxic effects on regulatory processes in parasites infecting the patient. Such deliberate therapeutic applications may be considered the proper role of **medical pharmacology,** which is often defined as the science of substances used to prevent, diagnose, and treat disease. **Toxicology** is the branch of pharmacology that deals with the undesirable effects of chemicals on living systems, from individual cells to humans to complex ecosystems (Figure 1–1).

THE HISTORY OF PHARMACOLOGY

Prehistoric people undoubtedly recognized the beneficial or toxic effects of many plant and animal materials. Early written records from China and Egypt and the traditions of India list remedies of many types, including a few that are still recognized as useful drugs today. Most, however, were worthless or actually harmful. In the 1500 years or so preceding the present, there were sporadic attempts to introduce rational methods into medicine, but none was successful owing to the dominance of systems of thought that purported to explain all of biology and disease without the need for experimentation and observation. These schools promulgated bizarre notions such as the idea that

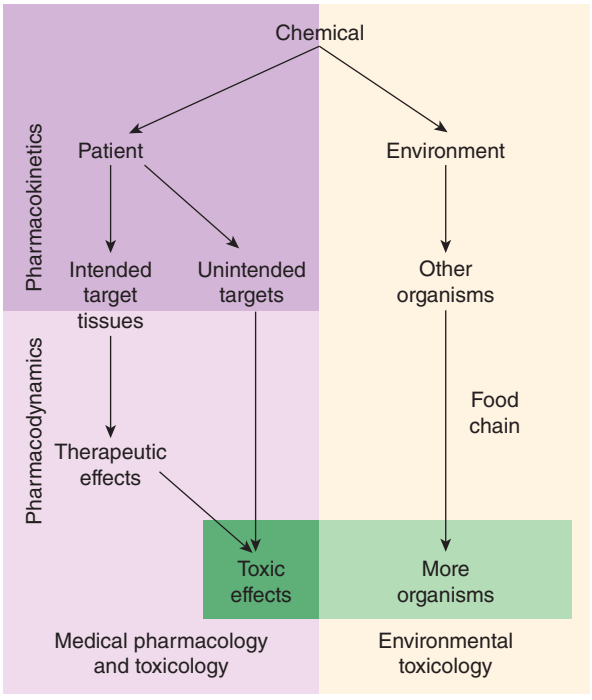


FIGURE 1–1 Major areas of study in pharmacology. The actions of chemicals can be divided into two large domains. The first (*left side*) is that of medical pharmacology and toxicology, which is aimed at understanding the actions of drugs as chemicals on individual organisms, especially humans and domestic animals. Both beneficial and toxic effects are included. Pharmacokinetics deals with the absorption, distribution, and elimination of drugs. Pharmacodynamics concerns the actions of the chemical on the organism. The second domain (*right side*) is that of environmental toxicology, which is concerned with the effects of chemicals on all organisms and their survival in groups and as species.

disease was caused by excesses of bile or blood in the body, that wounds could be healed by applying a salve to the weapon that caused the wound, and so on.

Around the end of the 17th century, and following the example of the physical sciences, reliance on observation and experimentation began to replace theorizing in medicine. As the value of these methods in the study of disease became clear, physicians in Great Britain and on the Continent began to apply them to the effects of traditional drugs used in their own practices. Thus, **materia medica**—the science of drug preparation and the medical use of drugs—began to develop as the precursor to pharmacology. However, any real understanding of the mechanisms of action of drugs was prevented by the absence of methods for purifying active agents from the crude materials that were available and—even more—by the lack of methods for testing hypotheses about the nature of drug actions.

In the late 18th and early 19th centuries, François Magendie, and later his student Claude Bernard, began to develop the methods of **experimental physiology** and **pharmacology.** Advances in chemistry and the further development of physiology in the 18th, 19th, and early 20th centuries laid the foundation needed for understanding how drugs work at the organ and tissue levels. Paradoxically, real advances in basic pharmacology during this time were accompanied by an outburst of unscientific claims by manufacturers and marketers of worthless "patent medicines." Not until the concepts of rational therapeutics, especially that of the **controlled clinical trial,** were reintroduced into medicine—only about 60 years ago—did it become possible to accurately evaluate therapeutic claims.

Around the same time, a major expansion of research efforts in all areas of biology began. As new concepts and new techniques were introduced, information accumulated about drug action and the biologic substrate of that action, the **drug receptor.** During the last half-century, many fundamentally new drug groups and new members of old groups were introduced. The last three decades have seen an even more rapid growth of information and understanding of the molecular basis for drug action. The molecular mechanisms of action of many drugs have now been identified, and numerous receptors have been isolated, structurally characterized, and cloned. In fact, the use of receptor identification methods (described in Chapter 2) has led to the discovery of many orphan receptors—receptors for which no ligand has been discovered and whose function can only be surmised. Studies of the local molecular environment of receptors have shown that receptors and effectors do not function in isolation; they are strongly influenced by other receptors and by companion regulatory proteins.

Pharmacogenomics—the relation of the individual's genetic makeup to his or her response to specific drugs—is close to becoming a practical area of therapy (see Box: Pharmacology & Genetics). Decoding of the genomes of many species—from bacteria to humans—has led to the recognition of unsuspected relationships between receptor families and the ways that receptor proteins have evolved. Discovery that small segments of RNA can interfere with protein synthesis with extreme selectivity has led to investigation of **small interfering RNAs (siRNAs)** and **microRNAs (miRNAs)** as therapeutic agents. Similarly, short nucleotide chains called **antisense oligonucleotides (ANOs)** synthesized to be complementary to natural RNA or DNA can interfere with the readout of genes and the transcription of RNA. These intracellular targets may provide the next major wave of advances in therapeutics.

The extension of scientific principles into everyday therapeutics is still going on, although the medication-consuming public is still exposed to vast amounts of inaccurate, incomplete, or unscientific information regarding the pharmacologic effects of chemicals. This has resulted in the irrational use of innumerable expensive, ineffective, and sometimes harmful remedies and the growth of a huge "alternative health care" industry. Unfortunately, manipulation of the legislative process in the United States has allowed many substances promoted for health—but not promoted specifically as "drugs"—to avoid meeting the Food and Drug Administration (FDA) standards described in Chapter 5. Conversely, lack of understanding of basic scientific principles in biology and statistics and the absence of critical thinking about public health issues have led to rejection of medical

Pharmacology & Genetics

It has been known for centuries that certain diseases are inherited, and we now understand that individuals with such diseases have a heritable abnormality in their DNA. During the last 10 years, the genomes of humans, mice, and many other organisms have been decoded in considerable detail. This has opened the door to a remarkable range of new approaches to research and treatment. It is now possible in the case of some inherited diseases to define exactly which DNA base pairs are anomalous and in which chromosome they appear. In a small number of animal models of such diseases, it has been possible to correct the abnormality by gene therapy, ie, insertion of an appropriate "healthy" gene into somatic cells. Human somatic cell **gene therapy** has been attempted, but the technical difficulties are great.

Studies of a newly discovered receptor or endogenous ligand are often confounded by incomplete knowledge of the exact role of that receptor or ligand. One of the most powerful of the new genetic techniques is the ability to breed animals (usually mice) in which the gene for the receptor or its endogenous ligand has been "knocked out," ie, mutated so that the gene product is absent or nonfunctional. Homozygous **knockout** mice usually have complete suppression of that function, whereas heterozygous animals usually have partial suppression. Observation of the behavior, biochemistry, and physiology of the knockout mice often defines the role of the missing gene product very clearly. When the products of a particular gene are so essential that even heterozygotes do not survive to birth, it is sometimes possible to breed "knockdown" versions with only limited suppression of function. Conversely, "knockin" mice, which overexpress certain proteins of interest, have been bred.

Some patients respond to certain drugs with greater than usual sensitivity to standard doses. It is now clear that such increased sensitivity is often due to a very small genetic modification that results in decreased activity of a particular enzyme responsible for eliminating that drug. (Such variations are discussed in Chapter 4.) **Pharmacogenomics** (or pharmacogenetics) is the study of the genetic variations that cause differences in drug response among individuals or populations. Future clinicians may screen every patient for a variety of such differences before prescribing a drug.

science by a segment of the public and to a common tendency to assume that all adverse drug effects are the result of malpractice.

Two general principles that the student should remember are (1) that *all* substances can under certain circumstances be toxic, and the chemicals in botanicals (herbs and plant extracts) are no different from chemicals in manufactured drugs except for the proportion of impurities (greater in botanicals); and, (2) that all dietary supplements and all therapies promoted as health-enhancing should meet the same standards of efficacy and safety as conventional drugs and medical therapies. That is, there should be no artificial separation between scientific medicine and "alternative" or "complementary" medicine.

PHARMACOLOGY & THE PHARMACEUTICAL INDUSTRY

A truly new drug (one that does not simply mimic the structure and action of previously available drugs) requires the discovery of a new drug *target*, ie, the pathophysiologic process or substrate of a disease. Such discoveries are usually made in public sector institutions (universities and research institutes), and the molecules that have beneficial effects on such targets are often discovered in the same laboratories. However, the *development* of new drugs usually takes place in industrial laboratories because optimization of a class of new drugs requires painstaking and expensive chemical, pharmacologic, and toxicologic research. In fact, much of the recent progress in the application of drugs to disease problems can be ascribed to the pharmaceutical industry including "big pharma," the multibillion-dollar corporations that specialize in drug discovery and development. As described in Chapter 5, these companies are uniquely skilled in exploiting discoveries from academic and governmental laboratories and translating these basic findings into commercially successful therapeutic breakthroughs.

Such breakthroughs come at a price, however, and the escalating cost of drugs has become a significant contributor to the inflationary increase in the cost of health care. Development of new drugs is enormously expensive, and to survive and prosper, big pharma must pay the costs of drug development and marketing and return a profit to its shareholders. Today, considerable controversy surrounds drug pricing. Critics claim that the costs of development and marketing are grossly inflated by marketing activities, which may consume as much as 25% or more of a company's budget in advertising and other promotional efforts. Furthermore, profit margins for big pharma have historically exceeded all other industries by a significant factor. Finally, pricing schedules for many drugs vary dramatically from country to country and even within countries, where large organizations can negotiate favorable prices and small ones cannot. Some countries have already addressed these inequities, and it seems likely that all countries will have to do so during the next few decades.

GENERAL PRINCIPLES OF PHARMACOLOGY

THE NATURE OF DRUGS

In the most general sense, a drug may be defined as any substance that brings about a change in biologic function through its chemical actions. In most cases, the drug molecule interacts as an

agonist (activator) or **antagonist** (inhibitor) with a specific molecule in the biologic system that plays a regulatory role. This target molecule is called a **receptor**. The nature of receptors is discussed more fully in Chapter 2. In a very small number of cases, drugs known as **chemical antagonists** may interact directly with other drugs, whereas a few drugs (osmotic agents) interact almost exclusively with water molecules. Drugs may be synthesized within the body (eg, **hormones**) or may be chemicals *not* synthesized in the body (ie, **xenobiotics,** from the Greek *xenos*, meaning "stranger"). **Poisons** are drugs that have almost exclusively harmful effects. However, Paracelsus (1493–1541) famously stated that "the dose makes the poison," meaning that any substance can be harmful if taken in the wrong dosage. **Toxins** are usually defined as poisons of biologic origin, ie, synthesized by plants or animals, in contrast to inorganic poisons such as lead and arsenic.

To interact chemically with its receptor, a drug molecule must have the appropriate size, electrical charge, shape, and atomic composition. Furthermore, a drug is often administered at a location distant from its intended site of action, eg, a pill given orally to relieve a headache. Therefore, a useful drug must have the necessary properties to be transported from its site of administration to its site of action. Finally, a practical drug should be inactivated or excreted from the body at a reasonable rate so that its actions will be of appropriate duration.

The Physical Nature of Drugs

Drugs may be solid at room temperature (eg, aspirin, atropine), liquid (eg, nicotine, ethanol), or gaseous (eg, nitrous oxide). These factors often determine the best route of administration. The most common routes of administration are described in Table 3–3. The various classes of organic compounds—carbohydrates, proteins, lipids, and their constituents—are all represented in pharmacology. As noted above, oligonucleotides, in the form of small segments of RNA, have entered clinical trials and are on the threshold of introduction into therapeutics.

A number of useful or dangerous drugs are inorganic elements, eg, lithium, iron, and heavy metals. Many organic drugs are weak acids or bases. This fact has important implications for the way they are handled by the body, because pH differences in the various compartments of the body may alter the degree of ionization of such drugs (see text that follows).

Drug Size

The molecular size of drugs varies from very small (lithium ion, MW 7) to very large (eg, alteplase [t-PA], a protein of MW 59,050). However, most drugs have molecular weights between 100 and 1000. The lower limit of this narrow range is probably set by the requirements for specificity of action. To have a good "fit" to only one type of receptor, a drug molecule must be sufficiently unique in shape, charge, and other properties, to prevent its binding to other receptors. To achieve such selective binding, it appears that a molecule should in most cases be at least 100 MW units in size. The upper limit in molecular weight is determined primarily by the requirement that drugs must be able to move within the body (eg, from the site of administration to the site of action). Drugs much larger than MW 1000 do not diffuse readily between compartments of the body (see Permeation, in following text). Therefore, very large drugs (usually proteins) must often be administered directly into the compartment where they have their effect. In the case of alteplase, a clot-dissolving enzyme, the drug is administered directly into the vascular compartment by intravenous or intra-arterial infusion.

Drug Reactivity and Drug-Receptor Bonds

Drugs interact with receptors by means of chemical forces or bonds. These are of three major types: **covalent, electrostatic,** and **hydrophobic.** Covalent bonds are very strong and in many cases not reversible under biologic conditions. Thus, the covalent bond formed between the acetyl group of acetylsalicylic acid (aspirin) and cyclooxygenase, its enzyme target in platelets, is not readily broken. The platelet aggregation–blocking effect of aspirin lasts long after free acetylsalicylic acid has disappeared from the bloodstream (about 15 minutes) and is reversed only by the synthesis of new enzyme in new platelets, a process that takes several days. Other examples of highly reactive, covalent bond-forming drugs are the DNA-alkylating agents used in cancer chemotherapy to disrupt cell division in the tumor.

Electrostatic bonding is much more common than covalent bonding in drug-receptor interactions. Electrostatic bonds vary from relatively strong linkages between permanently charged ionic molecules to weaker hydrogen bonds and very weak induced dipole interactions such as van der Waals forces and similar phenomena. Electrostatic bonds are weaker than covalent bonds.

Hydrophobic bonds are usually quite weak and are probably important in the interactions of highly lipid-soluble drugs with the lipids of cell membranes and perhaps in the interaction of drugs with the internal walls of receptor "pockets."

The specific nature of a particular drug-receptor bond is of less practical importance than the fact that drugs that bind through weak bonds to their receptors are generally more selective than drugs that bind by means of very strong bonds. This is because weak bonds require a very precise fit of the drug to its receptor if an interaction is to occur. Only a few receptor types are likely to provide such a precise fit for a particular drug structure. Thus, if we wished to design a highly selective short-acting drug for a particular receptor, we would avoid highly reactive molecules that form covalent bonds and instead choose a molecule that forms weaker bonds.

A few substances that are almost completely inert in the chemical sense nevertheless have significant pharmacologic effects. For example, xenon, an "inert" gas, has anesthetic effects at elevated pressures.

Drug Shape

The shape of a drug molecule must be such as to permit binding to its receptor site via the bonds just described. Optimally, the drug's shape is complementary to that of the receptor site in the same way that a key is complementary to a lock. Furthermore, the phenomenon of **chirality (stereoisomerism)** is so common in biology that more than half of all useful drugs are chiral molecules; that is, they

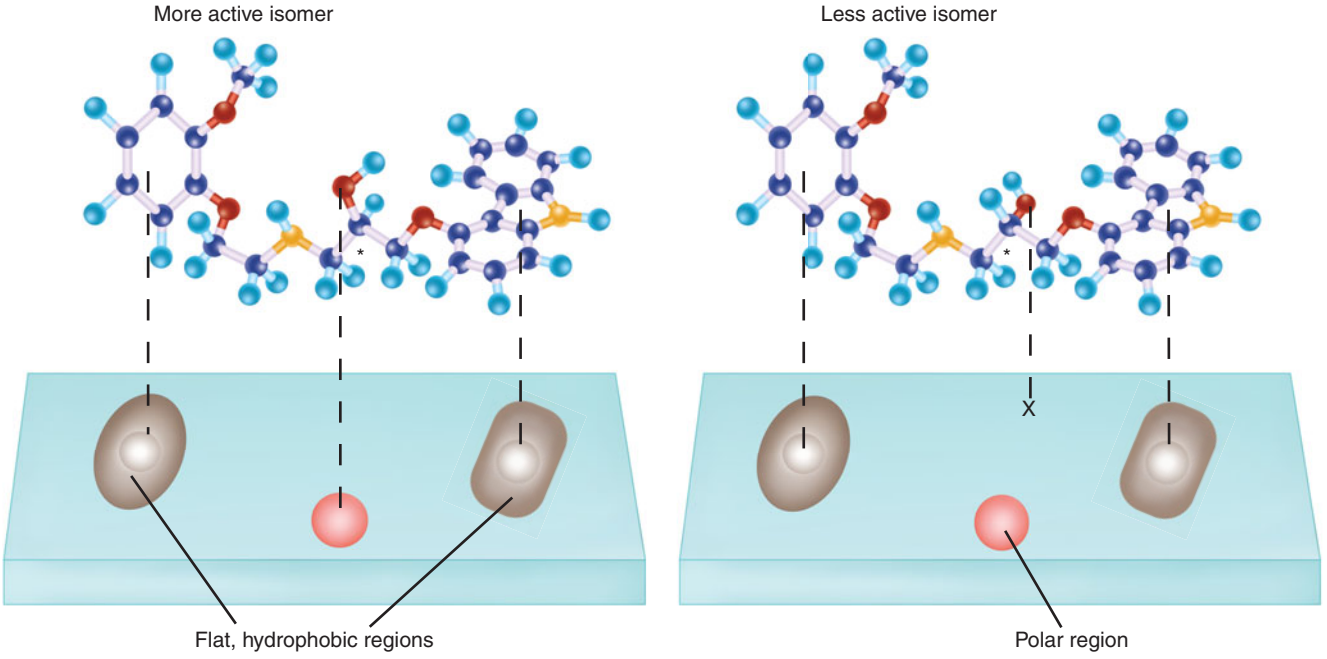


FIGURE 1–2 Cartoon illustrating the nonsuperimposibility of the two stereoisomers of carvedilol on the β receptor. The "receptor surface" has been grossly oversimplified. The chiral center carbon is denoted with an asterisk. One of the two isomers fits the three-dimensional configuration of binding site of the β-adrenoceptor molecule very well (*left*), and three groups, including an important polar moiety (an hydroxyl group, indicated by the central dashed line), bind to key areas of the surface. The less active isomer cannot orient all three binding areas to the receptor surface (*right*). (Molecule generated by means of Jmol, an open-source Java viewer for chemical structures in 3D [http://jmol.sourceforge.net/] with data from DrugBank [http://www.drugbank.ca].)

can exist as enantiomeric pairs. Drugs with two asymmetric centers have four diastereomers, eg, ephedrine, a sympathomimetic drug. In most cases, one of these enantiomers is much more potent than its mirror image enantiomer, reflecting a better fit to the receptor molecule. If one imagines the receptor site to be like a glove into which the drug molecule must fit to bring about its effect, it is clear why a "left-oriented" drug is more effective in binding to a left-hand receptor than its "right-oriented" enantiomer.

The more active enantiomer at one type of receptor site may not be more active at another receptor type, eg, a type that may be responsible for some other effect. For example, carvedilol, a drug that interacts with adrenoceptors, has a single chiral center and thus two enantiomers (Figure 1–2, Table 1–1). One of these enantiomers, the *(S)*(–) isomer, is a potent β-receptor blocker. The *(R)*(+) isomer is 100-fold weaker at the β receptor. However, the isomers are approximately equipotent as α-receptor blockers. Ketamine is an intravenous anesthetic. The (+) enantiomer is a more potent anesthetic and is less toxic than the (–) enantiomer. Unfortunately, the drug is still used as the racemic mixture.

Finally, because enzymes are usually stereoselective, one drug enantiomer is often more susceptible than the other to drug-metabolizing enzymes. As a result, the duration of action of one enantiomer may be quite different from that of the other. Similarly, drug transporters may be stereoselective.

Unfortunately, most studies of clinical efficacy and drug elimination in humans have been carried out with racemic mixtures of drugs rather than with the separate enantiomers. At present, only a small percentage of the chiral drugs used clinically are marketed as the active isomer—the rest are available only as racemic mixtures. As a result, many patients are receiving drug doses of which 50% is less active, inactive, or actively toxic. Some drugs are currently available in both the racemic and the pure, active isomer forms. Unfortunately, the hope that administration of the pure, active enantiomer would decrease adverse effects relative to those produced by racemic formulations has not been firmly established. However, there is increasing interest at both the scientific and the regulatory levels in making more chiral drugs available as their active enantiomers.

TABLE 1–1 Dissociation constants (K_d) of the enantiomers and racemate of carvedilol.

Form of Carvedilol	α Receptors (K_d, nmol/L[1])	β Receptors (K_d, nmol/L)
R(+) enantiomer	14	45
S(–) enantiomer	16	0.4
R,S(±) enantiomers	11	0.9

[1]The K_d is the concentration for 50% saturation of the receptors and is inversely proportionate to the affinity of the drug for the receptors.

Data from Ruffolo RR et al: The pharmacology of carvedilol. Eur J Pharmacol 1990;38:S82.

Rational Drug Design

Rational design of drugs implies the ability to predict the appropriate molecular structure of a drug on the basis of information about its biologic receptor. Until recently, no receptor was known in sufficient detail to permit such drug design. Instead, drugs were developed through random testing of chemicals or modification of drugs already known to have some effect (see Chapter 5). However, the characterization of many receptors during the past three decades has changed this picture. A few drugs now in use were developed through molecular design based on knowledge of the three-dimensional structure of the receptor site. Computer programs are now available that can iteratively optimize drug structures to fit known receptors. As more becomes known about receptor structure, rational drug design will become more common.

Receptor Nomenclature

The spectacular success of newer, more efficient ways to identify and characterize receptors (see Chapter 2) has resulted in a variety of differing, and sometimes confusing, systems for naming them. This in turn has led to a number of suggestions regarding more rational methods of naming receptors. The interested reader is referred for details to the efforts of the International Union of Pharmacology (IUPHAR) *Committee on Receptor Nomenclature and Drug Classification* (reported in various issues of *Pharmacological Reviews*) and to Alexander SPH, Mathie A, Peters JA: Guide to receptors and channels (GRAC), 4th edition. *Br J Pharmacol* 2009;158(Suppl 1):S1–S254. The chapters in this book mainly use these sources for naming receptors.

DRUG-BODY INTERACTIONS

The interactions between a drug and the body are conveniently divided into two classes. The actions of the drug on the body are termed **pharmacodynamic** processes (Figure 1–1); the principles of pharmacodynamics are presented in greater detail in Chapter 2. These properties determine the group in which the drug is classified, and they play the major role in deciding whether that group is appropriate therapy for a particular symptom or disease. The actions of the body on the drug are called **pharmacokinetic** processes and are described in Chapters 3 and 4. Pharmacokinetic processes govern the absorption, distribution, and elimination of drugs and are of great practical importance in the choice and administration of a particular drug for a particular patient, eg, a patient with impaired renal function. The following paragraphs provide a brief introduction to pharmacodynamics and pharmacokinetics.

Pharmacodynamic Principles

Most drugs must bind to a receptor to bring about an effect. However, at the cellular level, drug binding is only the first in what is often a complex sequence of steps:

- Drug (D) + receptor-effector (R) → drug-receptor-effector complex → effect
- D + R → drug-receptor complex → effector molecule → effect
- D + R → D-R complex → activation of coupling molecule → effector molecule → effect
- Inhibition of metabolism of endogenous activator → increased activator action on an effector molecule → increased effect

Note that the final change in function is accomplished by an **effector** mechanism. The effector may be part of the receptor molecule or may be a separate molecule. A very large number of receptors communicate with their effectors through coupling molecules, as described in Chapter 2.

A. Types of Drug-Receptor Interactions

Agonist drugs bind to and *activate* the receptor in some fashion, which directly or indirectly brings about the effect (Figure 1–3A). Receptor activation involves a change in conformation in the cases that have been studied at the molecular structure level. Some receptors incorporate effector machinery in the same molecule, so that drug binding brings about the effect directly, eg, opening of an ion channel or activation of enzyme activity. Other receptors are linked through one or more intervening coupling molecules to a separate effector molecule. The five major types of drug-receptor-effector coupling systems are discussed in Chapter 2. **Pharmacologic antagonist** drugs, by binding to a receptor, compete with and prevent binding by other molecules. For example, acetylcholine receptor blockers such as atropine are antagonists because they prevent access of acetylcholine and similar agonist drugs to the acetylcholine receptor site and they stabilize the receptor in its inactive state (or some state other than the acetylcholine-activated state). These agents reduce the effects of acetylcholine and similar molecules in the body (Figure 1–3B), but their action can be overcome by increasing the dosage of agonist. Some antagonists bind very tightly to the receptor site in an irreversible or pseudoirreversible fashion and cannot be displaced by increasing the agonist concentration. Drugs that bind to the same receptor molecule but do not prevent binding of the agonist are said to act **allosterically** and may enhance (Figure 1–3C) or inhibit (Figure 1–3D) the action of the agonist molecule. Allosteric inhibition is not overcome by increasing the dose of agonist.

B. Agonists That Inhibit Their Binding Molecules

Some drugs mimic agonist drugs by inhibiting the molecules responsible for terminating the action of an endogenous agonist. For example, acetylcholinesterase *inhibitors*, by slowing the destruction of endogenous acetylcholine, cause cholinomimetic effects that closely resemble the actions of cholinoceptor *agonist* molecules even though cholinesterase inhibitors do not bind or only incidentally bind to cholinoceptors (see Chapter 7). Because they amplify the effects of physiologically released agonist ligands, their effects are sometimes more selective and less toxic than those of exogenous agonists.

C. Agonists, Partial Agonists, and Inverse Agonists

Figure 1–4 describes a useful model of drug-receptor interaction. As indicated, the receptor is postulated to exist in the inactive,

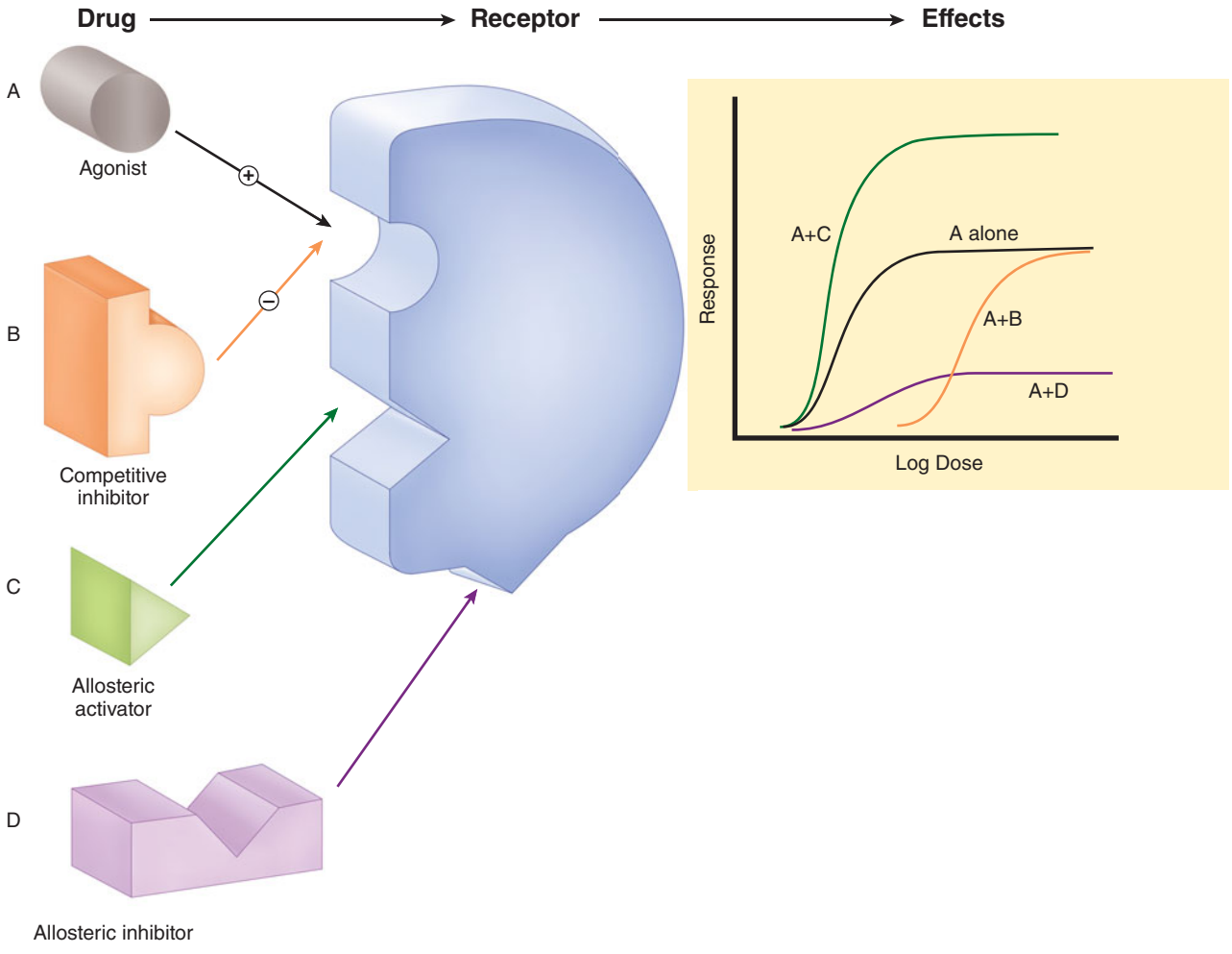


FIGURE 1–3 Drugs may interact with receptors in several ways. The effects resulting from these interactions are diagrammed in the dose-response curves at the right. Drugs that alter the agonist (**A**) response may activate the agonist binding site, compete with the agonist (competitive inhibitors, **B**), or act at separate (allosteric) sites, increasing (**C**) or decreasing (**D**) the response to the agonist. Allosteric activators (**C**) may increase the efficacy of the agonist or its binding affinity. The curve shown reflects an increase in efficacy; an increase in affinity would result in a leftward shift of the curve.

nonfunctional form (R_i) and in the activated form (R_a). Thermodynamic considerations indicate that even in the absence of any agonist, some of the receptor pool must exist in the R_a form some of the time and may produce the same physiologic effect as agonist-induced activity. This effect, occurring in the absence of agonist, is termed **constitutive activity.** Agonists are those drugs that have a much higher affinity for the R_a configuration and stabilize it, so that a large percentage of the total pool resides in the R_a–D fraction and a large effect is produced. The recognition of constitutive activity may depend on the receptor density, the concentration of coupling molecules (if a coupled system), and the number of effectors in the system.

Many agonist drugs, when administered at concentrations sufficient to saturate the receptor pool, can activate their receptor-effector systems to the maximum extent of which the system is capable; that is, they cause a shift of almost all of the receptor pool to the R_a–D pool. Such drugs are termed **full agonists.** Other drugs, called **partial agonists,** bind to the same receptors and activate them in the same way but do not evoke as great a response, no matter how high the concentration. In the model in Figure 1–4, partial agonists do not stabilize the R_a configuration as fully as full agonists, so that a significant fraction of receptors exists in the R_i–D pool. Such drugs are said to have low **intrinsic efficacy.** Thus, pindolol, a β-adrenoceptor partial agonist, may act either as an agonist (if no full agonist is present) or as an antagonist (if a full agonist such as epinephrine is present). (See Chapter 2.) Intrinsic efficacy is independent of affinity (as usually measured) for the receptor.

In the same model, conventional antagonist action can be explained as fixing the fractions of drug-bound R_i and R_a in the same relative amounts as in the absence of any drug. In this situation, no change will be observed, so the drug will appear to be without effect. However, the presence of the antagonist at the receptor site will block access of agonists to the receptor and prevent the usual agonist effect. Such blocking action can be termed **neutral antagonism.**

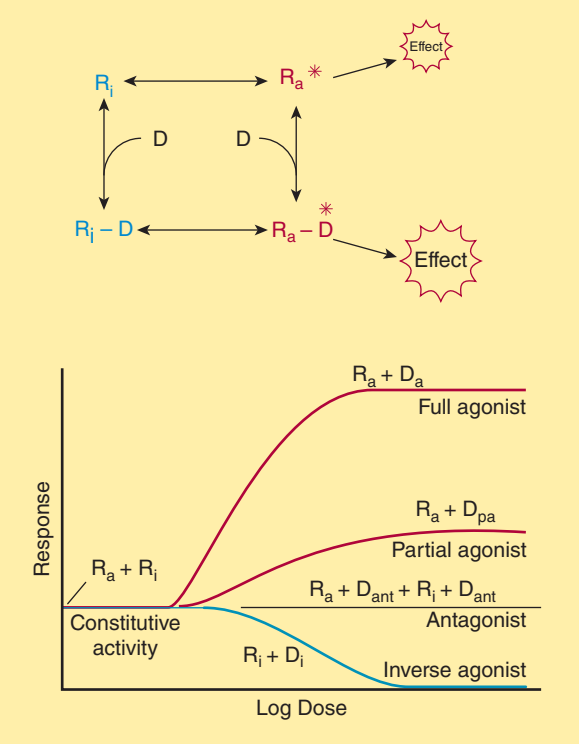


FIGURE 1–4 A model of drug-receptor interaction. The receptor is able to assume two conformations. In the R_i conformation, it is inactive and produces no effect, even when combined with a drug molecule. In the R_a conformation, the receptor can activate downstream mechanisms that produce a small observable effect, even in the absence of drug (constitutive activity). In the absence of drugs, the two isoforms are in equilibrium, and the R_i form is favored. Conventional full agonist drugs have a much higher affinity for the R_a conformation, and mass action thus favors the formation of the R_a–D complex with a much larger observed effect. Partial agonists have an intermediate affinity for both R_i and R_a forms. Conventional antagonists, according to this hypothesis, have equal affinity for both receptor forms and maintain the same level of constitutive activity. Inverse agonists, on the other hand, have a much higher affinity for the R_i form, reduce constitutive activity, and may produce a contrasting physiologic result.

What will happen if a drug has a much stronger affinity for the R_i than for the R_a state and stabilizes a large fraction in the R_i–D pool? In this scenario the drug would reduce any constitutive activity, thus resulting in effects that are the opposite of the effects produced by conventional agonists at that receptor. Such drugs have been termed **inverse agonists** (Figure 1–4). One of the best documented examples of such a system is the γ-aminobutyric acid ($GABA_A$) receptor-effector (a chloride channel) in the nervous system. This receptor is activated by the endogenous transmitter GABA and causes inhibition of postsynaptic cells. Conventional exogenous agonists such as benzodiazepines also facilitate the receptor-effector system and cause GABA-like inhibition with sedation as the therapeutic result. This inhibition can be blocked by conventional neutral antagonists such as flumazenil. In addition, inverse agonists have been found that cause anxiety and agitation, the inverse of sedation (see Chapter 22). Similar inverse agonists have been found for β-adrenoceptors, histamine H_1 and H_2 receptors, and several other receptor systems.

D. Duration of Drug Action

Termination of drug action is a result of one of several processes. In some cases, the effect lasts only as long as the drug occupies the receptor, and dissociation of drug from the receptor automatically terminates the effect. In many cases, however, the action may persist after the drug has dissociated because, for example, some coupling molecule is still present in activated form. In the case of drugs that bind covalently to the receptor site, the effect may persist until the drug-receptor complex is destroyed and new receptors or enzymes are synthesized, as described previously for aspirin. In addition, many receptor-effector systems incorporate desensitization mechanisms for preventing excessive activation when agonist molecules continue to be present for long periods. (See Chapter 2 for additional details.)

E. Receptors and Inert Binding Sites

To function as a receptor, an endogenous molecule must first be **selective** in choosing ligands (drug molecules) to bind; and second, it must **change its function** upon binding in such a way that the function of the biologic system (cell, tissue, etc) is altered. The selectivity characteristic is required to avoid constant activation of the receptor by promiscuous binding of many different ligands. The ability to change function is clearly necessary if the ligand is to cause a pharmacologic effect. The body contains a vast array of molecules that are capable of binding drugs, however, and not all of these endogenous molecules are regulatory molecules. Binding of a drug to a nonregulatory molecule such as plasma albumin will result in no detectable change in the function of the biologic system, so this endogenous molecule can be called an **inert binding site.** Such binding is not completely without significance, however, because it affects the distribution of drug within the body and determines the amount of free drug in the circulation. Both of these factors are of pharmacokinetic importance (see also Chapter 3).

Pharmacokinetic Principles

In practical therapeutics, a drug should be able to reach its intended site of action after administration by some convenient route. In many cases, the active drug molecule is sufficiently lipid-soluble and stable to be given as such. In some cases, however, an inactive precursor chemical that is readily absorbed and distributed must be administered and then converted to the active drug by biologic processes—inside the body. Such a precursor chemical is called a **prodrug.**

In only a few situations is it possible to apply a drug directly to its target tissue, eg, by topical application of an anti-inflammatory agent to inflamed skin or mucous membrane. Most often, a drug is administered into one body compartment, eg, the gut, and must move to its site of action in another compartment, eg, the brain in the case of an antiseizure medication. This requires that the drug be **absorbed** into the blood from its site of administration and **distributed** to its site of action, **permeating** through the various

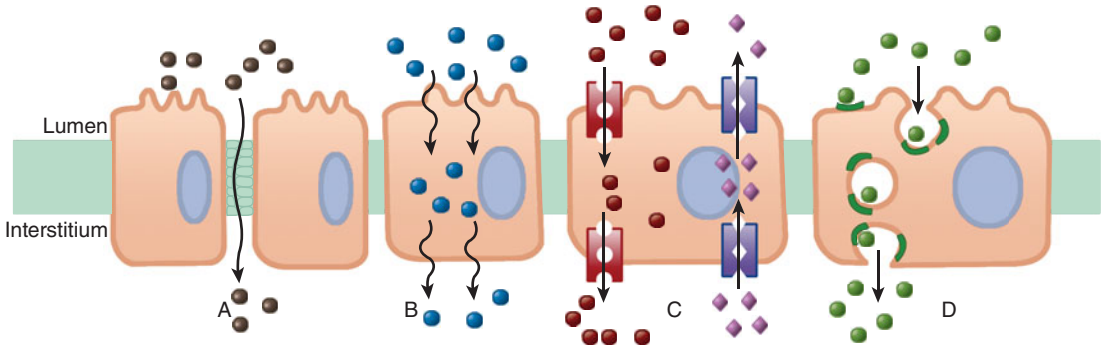


FIGURE 1–5 Mechanisms of drug permeation. Drugs may diffuse passively through aqueous channels in the intercellular junctions (eg, tight junctions, **A**), or through lipid cell membranes (**B**). Drugs with the appropriate characteristics may be transported by carriers into or out of cells (**C**). Very impermeant drugs may also bind to cell surface receptors (dark binding sites), be engulfed by the cell membrane (endocytosis), and then released inside the cell or expelled via the membrane-limited vesicles out of the cell into the extracellular space (exocytosis, **D**).

barriers that separate these compartments. For a drug given orally to produce an effect in the central nervous system, these barriers include the tissues that make up the wall of the intestine, the walls of the capillaries that perfuse the gut, and the blood-brain barrier, the walls of the capillaries that perfuse the brain. Finally, after bringing about its effect, a drug should be **eliminated** at a reasonable rate by metabolic inactivation, by excretion from the body, or by a combination of these processes.

A. Permeation

Drug permeation proceeds by several mechanisms. Passive diffusion in an aqueous or lipid medium is common, but active processes play a role in the movement of many drugs, especially those whose molecules are too large to diffuse readily (Figure 1–5).

***1. Aqueous diffusion*—**Aqueous diffusion occurs within the larger aqueous compartments of the body (interstitial space, cytosol, etc) and across epithelial membrane tight junctions and the endothelial lining of blood vessels through aqueous pores that—in some tissues—permit the passage of molecules as large as MW 20,000–30,000.* See Figure 1–5A.

Aqueous diffusion of drug molecules is usually driven by the concentration gradient of the permeating drug, a downhill movement described by Fick's law (see below). Drug molecules that are bound to large plasma proteins (eg, albumin) do not permeate most vascular aqueous pores. If the drug is charged, its flux is also influenced by electrical fields (eg, the membrane potential and—in parts of the nephron—the transtubular potential).

***2. Lipid diffusion*—**Lipid diffusion is the most important limiting factor for drug permeation because of the large number of lipid barriers that separate the compartments of the body. Because these lipid barriers separate aqueous compartments, the **lipid:aqueous partition coefficient** of a drug determines how readily the molecule moves between aqueous and lipid media. In the case of weak acids and weak bases (which gain or lose electrical charge-bearing protons, depending on the pH), the ability to move from aqueous to lipid or vice versa varies with the pH of the medium, because charged molecules attract water molecules. The ratio of lipid-soluble form to water-soluble form for a weak acid or weak base is expressed by the Henderson-Hasselbalch equation (described in the following text). See Figure 1–5B.

***3. Special carriers*—**Special carrier molecules exist for many substances that are important for cell function and too large or too insoluble in lipid to diffuse passively through membranes, eg, peptides, amino acids, and glucose. These carriers bring about movement by active transport or facilitated diffusion and, unlike passive diffusion, are selective, saturable, and inhibitable. Because many drugs are or resemble such naturally occurring peptides, amino acids, or sugars, they can use these carriers to cross membranes. See Figure 1–5C.

Many cells also contain less selective membrane carriers that are specialized for expelling foreign molecules. One large family of such transporters binds adenosine triphosphate (ATP) and is called the ABC (ATP-binding cassette) family. This family includes the **P-glycoprotein** or **multidrug resistance type 1 (MDR1) transporter** found in the brain, testes, and other tissues, and in some drug-resistant neoplastic cells, Table 1–2. Similar transport molecules from the ABC family, the **multidrug resistance-associated protein (MRP)** transporters, play important roles in the excretion of some drugs or their metabolites into urine and bile and in the resistance of some tumors to chemotherapeutic drugs. Several other transporter families have been identified that do not bind ATP but use ion gradients to drive transport. Some of these (the solute carrier [SLC] family) are particularly important in the uptake of neurotransmitters across nerve-ending membranes. The latter carriers are discussed in more detail in Chapter 6.

***4. Endocytosis and exocytosis*—**A few substances are so large or impermeant that they can enter cells only by endocytosis, the process by which the substance is bound at a cell-surface receptor,

*The capillaries of the brain, the testes, and some other tissues are characterized by the absence of pores that permit aqueous diffusion. They may also contain high concentrations of drug export pumps (MDR pumps; see text). These tissues are therefore protected or "sanctuary" sites from many circulating drugs.

TABLE 1–2 Some transport molecules important in pharmacology.

Transporter	Physiologic Function	Pharmacologic Significance
NET	Norepinephrine reuptake from synapse	Target of cocaine and some tricyclic antidepressants
SERT	Serotonin reuptake from synapse	Target of selective serotonin reuptake inhibitors and some tricyclic antidepressants
VMAT	Transport of dopamine and norepinephrine into adrenergic vesicles in nerve endings	Target of reserpine and tetrabenazine
MDR1	Transport of many xenobiotics out of cells	Increased expression confers resistance to certain anticancer drugs; inhibition increases blood levels of digoxin
MRP1	Leukotriene secretion	Confers resistance to certain anticancer and antifungal drugs

MDR1, multidrug resistance protein-1; MRP1, multidrug resistance-associated protein-1; NET, norepinephrine transporter; SERT, serotonin reuptake transporter; VMAT, vesicular monoamine transporter.

engulfed by the cell membrane, and carried into the cell by pinching off of the newly formed vesicle inside the membrane. The substance can then be released inside the cytosol by breakdown of the vesicle membrane, Figure 1–5D. This process is responsible for the transport of vitamin B_{12}, complexed with a binding protein (intrinsic factor) across the wall of the gut into the blood. Similarly, iron is transported into hemoglobin-synthesizing red blood cell precursors in association with the protein transferrin. Specific receptors for the transport proteins must be present for this process to work.

The reverse process (exocytosis) is responsible for the secretion of many substances from cells. For example, many neurotransmitter substances are stored in membrane-bound vesicles in nerve endings to protect them from metabolic destruction in the cytoplasm. Appropriate activation of the nerve ending causes fusion of the storage vesicle with the cell membrane and expulsion of its contents into the extracellular space (see Chapter 6).

B. Fick's Law of Diffusion

The passive flux of molecules down a concentration gradient is given by Fick's law:

$$\text{Flux (molecules per unit time)} = (C_1 - C_2) \times \frac{\text{Area} \times \text{Permeability coefficient}}{\text{Thickness}}$$

where C_1 is the higher concentration, C_2 is the lower concentration, area is the cross-sectional area of the diffusion path, permeability coefficient is a measure of the mobility of the drug molecules in the medium of the diffusion path, and thickness is the thickness (length) of the diffusion path. In the case of lipid diffusion, the lipid:aqueous partition coefficient is a major determinant of mobility of the drug, because it determines how readily the drug enters the lipid membrane from the aqueous medium.

C. Ionization of Weak Acids and Weak Bases; the Henderson-Hasselbalch Equation

The electrostatic charge of an ionized molecule attracts water dipoles and results in a polar, relatively water-soluble and lipid-insoluble complex. Because lipid diffusion depends on relatively high lipid solubility, ionization of drugs may markedly reduce their ability to permeate membranes. A very large percentage of the drugs in use are weak acids or weak bases (Table 1–3). For drugs, a weak acid is best defined as a neutral molecule that can reversibly dissociate into an anion (a negatively charged molecule) and a proton (a hydrogen ion). For example, aspirin dissociates as follows:

$$\underset{\text{Neutral aspirin}}{C_8H_7O_2COOH} \rightleftharpoons \underset{\text{Aspirin anion}}{C_8H_7O_2COO^-} + \underset{\text{Proton}}{H^+}$$

A drug that is a weak base can be defined as a neutral molecule that can form a cation (a positively charged molecule) by combining with a proton. For example, pyrimethamine, an antimalarial drug, undergoes the following association-dissociation process:

$$\underset{\text{Pyrimethamine cation}}{C_{12}H_{11}ClN_3NH_3^+} \rightleftharpoons \underset{\text{Neutral pyrimethamine}}{C_{12}H_{11}ClN_3NH_2} + \underset{\text{Proton}}{H^+}$$

Note that the protonated form of a weak acid is the neutral, more lipid-soluble form, whereas the unprotonated form of a weak base is the neutral form. The law of mass action requires that these reactions move to the left in an acid environment (low pH, excess protons available) and to the right in an alkaline environment. The Henderson-Hasselbalch equation relates the ratio of protonated to unprotonated weak acid or weak base to the molecule's pK_a and the pH of the medium as follows:

$$\log \frac{(\text{Protonated})}{(\text{Unprotonated})} = pK_a - pH$$

This equation applies to both acidic and basic drugs. Inspection confirms that the lower the pH relative to the pK_a, the greater will be the fraction of drug in the protonated form. Because the uncharged form is the more lipid-soluble, more of a weak acid will be in the lipid-soluble form at acid pH, whereas more of a basic drug will be in the lipid-soluble form at alkaline pH.

TABLE 1–3 Ionization constants of some common drugs.

Drug	pK_a[1]	Drug	pK_a[1]	Drug	pK_a[1]
Weak acids		**Weak bases**		**Weak bases (cont'd)**	
Acetaminophen	9.5	Albuterol (salbutamol)	9.3	Isoproterenol	8.6
Acetazolamide	7.2	Allopurinol	9.4, 12.3[2]	Lidocaine	7.9
Ampicillin	2.5	Alprenolol	9.6	Metaraminol	8.6
Aspirin	3.5	Amiloride	8.7	Methadone	8.4
Chlorothiazide	6.8, 9.4[2]	Amiodarone	6.6	Methamphetamine	10.0
Chlorpropamide	5.0	Amphetamine	9.8	Methyldopa	10.6
Ciprofloxacin	6.1, 8.7[2]	Atropine	9.7	Metoprolol	9.8
Cromolyn	2.0	Bupivacaine	8.1	Morphine	7.9
Ethacrynic acid	2.5	Chlordiazepoxide	4.6	Nicotine	7.9, 3.1[2]
Furosemide	3.9	Chloroquine	10.8, 8.4	Norepinephrine	8.6
Ibuprofen	4.4, 5.2[2]	Chlorpheniramine	9.2	Pentazocine	7.9
Levodopa	2.3	Chlorpromazine	9.3	Phenylephrine	9.8
Methotrexate	4.8	Clonidine	8.3	Physostigmine	7.9, 1.8[2]
Methyldopa	2.2, 9.2[2]	Cocaine	8.5	Pilocarpine	6.9, 1.4[2]
Penicillamine	1.8	Codeine	8.2	Pindolol	8.6
Pentobarbital	8.1	Cyclizine	8.2	Procainamide	9.2
Phenobarbital	7.4	Desipramine	10.2	Procaine	9.0
Phenytoin	8.3	Diazepam	3.0	Promethazine	9.1
Propylthiouracil	8.3	Diphenhydramine	8.8	Propranolol	9.4
Salicylic acid	3.0	Diphenoxylate	7.1	Pseudoephedrine	9.8
Sulfadiazine	6.5	Ephedrine	9.6	Pyrimethamine	7.0–7.3[3]
Sulfapyridine	8.4	Epinephrine	8.7	Quinidine	8.5, 4.4[2]
Theophylline	8.8	Ergotamine	6.3	Scopolamine	8.1
Tolbutamide	5.3	Fluphenazine	8.0, 3.9[2]	Strychnine	8.0, 2.3[2]
Warfarin	5.0	Hydralazine	7.1	Terbutaline	10.1
		Imipramine	9.5	Thioridazine	9.5

[1]The pK_a is that pH at which the concentrations of the ionized and nonionized forms are equal.

[2]More than one ionizable group.

[3]Isoelectric point.

Application of this principle is made in the manipulation of drug excretion by the kidney. Almost all drugs are filtered at the glomerulus. If a drug is in a lipid-soluble form during its passage down the renal tubule, a significant fraction will be reabsorbed by simple passive diffusion. If the goal is to accelerate excretion of the drug (eg, in a case of drug overdose), it is important to prevent its reabsorption from the tubule. This can often be accomplished by adjusting urine pH to make certain that most of the drug is in the ionized state, as shown in Figure 1–6. As a result of this partitioning effect, the drug is "trapped" in the urine. Thus, weak acids are usually excreted faster in alkaline urine; weak bases are usually excreted faster in acidic urine. Other body fluids in which pH differences from blood pH may cause trapping or reabsorption are the contents of the stomach and small intestine; breast milk; aqueous humor; and vaginal and prostatic secretions (Table 1–4).

As suggested by Table 1–3, a large number of drugs are weak bases. Most of these bases are amine-containing molecules. The nitrogen of a neutral amine has three atoms associated with it plus a pair of unshared electrons (see the display that follows). The three atoms may consist of one carbon (designated "R") and two hydrogens (a **primary amine**), two carbons and one hydrogen (a **secondary amine**), or three carbon atoms (a **tertiary amine**). Each of these three forms may reversibly bind a proton with the unshared electrons. Some drugs have a fourth carbon-nitrogen bond; these are **quaternary amines**. However, the quaternary amine is permanently charged and has no unshared electrons with which to reversibly bind a proton. Therefore, primary, secondary,

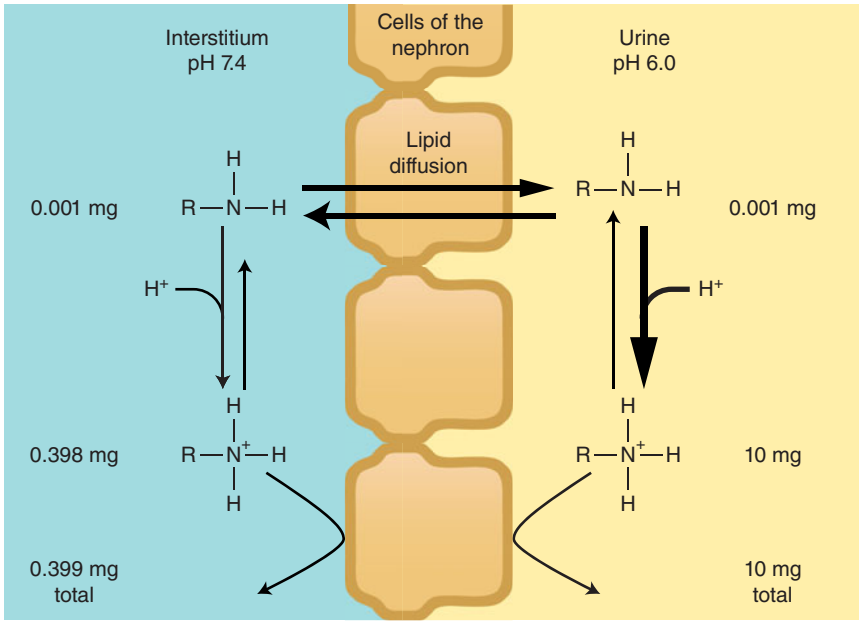
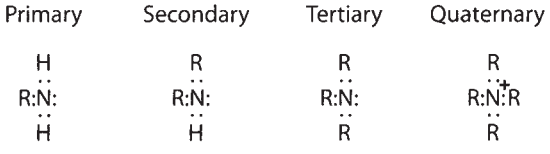


FIGURE 1–6 Trapping of a weak base (methamphetamine) in the urine when the urine is more acidic than the blood. In the hypothetical case illustrated, the diffusible uncharged form of the drug has equilibrated across the membrane, but the total concentration (charged plus uncharged) in the urine (more than 10 mg) is 25 times higher than in the blood (0.4 mg).

and tertiary amines may undergo reversible protonation and vary their lipid solubility with pH, but quaternary amines are always in the poorly lipid-soluble charged form.



DRUG GROUPS

To learn each pertinent fact about each of the many hundreds of drugs mentioned in this book would be an impractical goal and, fortunately, is unnecessary. Almost all the several thousand drugs currently available can be arranged into about 70 groups. Many of the drugs within each group are very similar in pharmacodynamic actions and in their pharmacokinetic properties as well. For most groups, one or more **prototype drugs** can be identified that typify

TABLE 1–4 Body fluids with potential for drug "trapping" through the pH-partitioning phenomenon.

Body Fluid	Range of pH	Total Fluid: Blood Concentration Ratios for Sulfadiazine (acid, pK_a 6.5)[1]	Total Fluid: Blood Concentration Ratios for Pyrimethamine (base, pK_a 7.0)[1]
Urine	5.0–8.0	0.12–4.65	72.24–0.79
Breast milk	6.4–7.6[2]	0.2–1.77	3.56–0.89
Jejunum, ileum contents	7.5–8.0[3]	1.23–3.54	0.94–0.79
Stomach contents	1.92–2.59[2]	0.11[4]	85,993–18,386
Prostatic secretions	6.45–7.4[2]	0.21	3.25–1.0
Vaginal secretions	3.4–4.2[3]	0.11[4]	2848–452

[1]Body fluid protonated-to-unprotonated drug ratios were calculated using each of the pH extremes cited; a blood pH of 7.4 was used for blood:drug ratio. For example, the steady-state urine:blood ratio for sulfadiazine is 0.12 at a urine pH of 5.0; this ratio is 4.65 at a urine pH of 8.0. Thus, sulfadiazine is much more effectively trapped and excreted in alkaline urine.

[2]Lentner C (editor): *Geigy Scientific Tables,* vol 1, 8th ed. Ciba Geigy, 1981.

[3]Bowman WC, Rand MJ: *Textbook of Pharmacology,* 2nd ed. Blackwell, 1980.

[4]Insignificant change in ratios over the physiologic pH range.

the most important characteristics of the group. This permits classification of other important drugs in the group as variants of the prototype, so that only the prototype must be learned in detail and, for the remaining drugs, only the differences from the prototype.

SOURCES OF INFORMATION

Students who wish to review the field of pharmacology in preparation for an examination are referred to *Pharmacology: Examination and Board Review,* by Trevor, Katzung, and Masters (McGraw-Hill, 2010). This book provides over 1000 questions and explanations in USMLE format. A short study guide is *USMLE Road Map: Pharmacology,* by Katzung and Trevor (McGraw-Hill, 2006). *Road Map* contains numerous tables, figures, mnemonics, and USMLE-type clinical vignettes.

The references at the end of each chapter in this book were selected to provide reviews or classic publications of information specific to those chapters. More detailed questions relating to basic or clinical research are best answered by referring to the journals covering general pharmacology and clinical specialties. For the student and the physician, three periodicals can be recommended as especially useful sources of current information about drugs: *The New England Journal of Medicine*, which publishes much original drug-related clinical research as well as frequent reviews of topics in pharmacology; *The Medical Letter on Drugs and Therapeutics*, which publishes brief critical reviews of new and old therapies, mostly pharmacologic; and *Drugs*, which publishes extensive reviews of drugs and drug groups.

Other sources of information pertinent to the United States should be mentioned as well. The "package insert" is a summary of information that the manufacturer is required to place in the prescription sales package; *Physicians' Desk Reference (PDR)* is a compendium of package inserts published annually with supplements twice a year. It is sold in bookstores and distributed free to licensed physicians. The package insert consists of a brief description of the pharmacology of the product. This brochure contains much practical information, and it is also used as a means of shifting liability for untoward drug reactions from the manufacturer onto the practitioner. Therefore, the manufacturer typically lists every toxic effect ever reported, no matter how rare. *Micromedex* is an extensive subscription website maintained by the Thomson Corporation (http://clinical.thomsonhealthcare.com/products/physicians/). It provides downloads for personal digital assistant devices, online drug dosage and interaction information, and toxicologic information. A useful and objective quarterly handbook that presents information on drug toxicity and interactions is *Drug Interactions: Analysis and Management.* Finally, the FDA maintains an Internet website that carries news regarding recent drug approvals, withdrawals, warnings, etc. It can be accessed at http://www.fda.gov. The MedWatch drug safety program is a free e-mail notification service that provides news of FDA drug warnings and withdrawals. Subscriptions may be obtained at https://service.govdelivery.com/service/user.html?code=USFDA.

REFERENCES

Drug Interactions: Analysis and Management (quarterly). Wolters Kluwer Publications.

Pharmacology: Examination & Board Review, 9th ed. McGraw-Hill Companies, Inc.

Symposium: Allosterism and collateral efficacy. Trends Pharmacol Sci 2007;28(8):entire issue.

USMLE Road Map: Pharmacology, McGraw-Hill Companies, Inc.

The Medical Letter on Drugs and Therapeutics. The Medical Letter, Inc.

CASE STUDY ANSWER

In the case study, the patient intravenously self-administered an overdose of methamphetamine, a weak base. This drug is freely filtered at the glomerulus, but can be rapidly reabsorbed in the renal tubule. Administration of ammonium chloride acidifies the urine, converting a larger fraction of the drug to the protonated, charged form, which is poorly reabsorbed and thus more rapidly eliminated. Note that not all experts recommend forced diuresis and urinary pH manipulation after methamphetamine overdose because of the risk of renal damage (see Figure 1–6).

This page intentionally left blank

CHAPTER 2

Drug Receptors & Pharmacodynamics

Mark von Zastrow, MD, PhD*

CASE STUDY

A 51-year-old man presents to his medical clinic due to difficulty breathing. The patient is afebrile and normotensive, but tachypneic. Auscultation of the chest reveals diffuse wheezes. The physician provisionally makes the diagnosis of bronchial asthma and administers epinephrine by intramuscular injection, improving the patient's breathing over several minutes. A normal chest X-ray is subsequently obtained, and the medical history is remarkable only for mild hypertension that was recently treated with propranolol. The physician instructs the patient to discontinue use of propranolol, and changes the patient's antihypertensive medication to verapamil. Why is the physician correct to discontinue propranolol? Why is verapamil a better choice for managing hypertension in this patient?

Therapeutic and toxic effects of drugs result from their interactions with molecules in the patient. Most drugs act by associating with specific macromolecules in ways that alter the macromolecules' biochemical or biophysical activities. This idea, more than a century old, is embodied in the term **receptor:** the component of a cell or organism that interacts with a drug and initiates the chain of events leading to the drug's observed effects.

Receptors have become the central focus of investigation of drug effects and their mechanisms of action (pharmacodynamics). The receptor concept, extended to endocrinology, immunology, and molecular biology, has proved essential for explaining many aspects of biologic regulation. Many drug receptors have been isolated and characterized in detail, thus opening the way to precise understanding of the molecular basis of drug action.

The receptor concept has important practical consequences for the development of drugs and for arriving at therapeutic decisions in clinical practice. These consequences form the basis for understanding the actions and clinical uses of drugs described in almost every chapter of this book. They may be briefly summarized as follows:

1. **Receptors largely determine the quantitative relations between dose or concentration of drug and pharmacologic effects.** The receptor's affinity for binding a drug determines the concentration of drug required to form a significant number of drug-receptor complexes, and the total number of receptors may limit the maximal effect a drug may produce.
2. **Receptors are responsible for selectivity of drug action.** The molecular size, shape, and electrical charge of a drug determine whether—and with what affinity—it will bind to a particular receptor among the vast array of chemically different binding sites available in a cell, tissue, or patient. Accordingly, changes in the chemical structure of a drug can dramatically increase or decrease a new drug's affinities for different classes of receptors, with resulting alterations in therapeutic and toxic effects.
3. **Receptors mediate the actions of pharmacologic agonists and antagonists.** Some drugs and many natural ligands, such as hormones and neurotransmitters, regulate the function of receptor macromolecules as **agonists;** this means that they activate the receptor to signal as a direct result of binding to it. Some agonists activate a single kind of receptor to produce all their biologic functions, whereas others selectively promote one receptor function more than another.

   Other drugs act as pharmacologic **antagonists;** that is, they bind to receptors but do not activate generation of a signal; consequently, they interfere with the ability of an agonist to activate the receptor. The effect of a so-called "pure" antagonist on a cell or in a patient depends entirely on its preventing the binding of agonist molecules and blocking their biologic actions. Other

*The author thanks Henry R. Bourne, MD, for major contributions to this chapter.

antagonists, in addition to preventing agonist binding, suppress the basal signaling ("constitutive") activity of receptors. Some of the most useful drugs in clinical medicine are pharmacologic antagonists.

MACROMOLECULAR NATURE OF DRUG RECEPTORS

Most receptors are proteins, presumably because the structures of polypeptides provide both the necessary diversity and the necessary specificity of shape and electrical charge. Receptors vary greatly in structure and can be identified in many ways. Traditionally, drug binding was used to identify or purify receptors from tissue extracts; consequently, receptors were discovered after the drugs that bind to them. However, advances in molecular biology and genome sequencing have effectively reversed this order. Now receptors are being discovered by predicted structure or sequence homology to other (known) receptors, and drugs that bind to them are developed later using chemical screening methods. This effort has revealed, for many known drugs, a larger diversity of receptors than previously anticipated. It has also identified a number of **"orphan" receptors,** so-called because their ligands are presently unknown, which may prove to be useful targets for the development of new drugs.

The best-characterized drug receptors are **regulatory proteins,** which mediate the actions of endogenous chemical signals such as neurotransmitters, autacoids, and hormones. This class of receptors mediates the effects of many of the most useful therapeutic agents. The molecular structures and biochemical mechanisms of these regulatory receptors are described in a later section entitled Signaling Mechanisms & Drug Action.

Other classes of proteins that have been clearly identified as drug receptors include **enzymes,** which may be inhibited (or, less commonly, activated) by binding a drug (eg, dihydrofolate reductase, the receptor for the antineoplastic drug methotrexate); **transport proteins** (eg, Na^+/K^+-ATPase, the membrane receptor for cardioactive digitalis glycosides); and **structural proteins** (eg, tubulin, the receptor for colchicine, an anti-inflammatory agent).

This chapter deals with three aspects of drug receptor function, presented in increasing order of complexity: (1) receptors as determinants of the quantitative relation between the concentration of a drug and the pharmacologic response, (2) receptors as regulatory proteins and components of chemical signaling mechanisms that provide targets for important drugs, and (3) receptors as key determinants of the therapeutic and toxic effects of drugs in patients.

RELATION BETWEEN DRUG CONCENTRATION & RESPONSE

The relation between dose of a drug and the clinically observed response may be complex. In carefully controlled in vitro systems, however, the relation between concentration of a drug and its effect is often simple and can be described with mathematical precision. This idealized relation underlies the more complex relations between dose and effect that occur when drugs are given to patients.

Concentration-Effect Curves & Receptor Binding of Agonists

Even in intact animals or patients, responses to low doses of a drug usually increase in direct proportion to dose. As doses increase, however, the response increment diminishes; finally, doses may be reached at which no further increase in response can be achieved. In idealized or in vitro systems, the relation between drug concentration and effect is described by a hyperbolic curve (Figure 2–1A) according to the following equation:

$$E = \frac{E_{max} \times C}{C + EC_{50}}$$

where E is the effect observed at concentration C, E_{max} is the maximal response that can be produced by the drug, and EC_{50} is the concentration of drug that produces 50% of maximal effect.

This hyperbolic relation resembles the mass action law, which describes association between two molecules of a given affinity. This resemblance suggests that drug agonists act by binding to ("occupying") a distinct class of biologic molecules with a characteristic affinity for the drug receptor. Radioactive receptor ligands have been used to confirm this occupancy assumption in many drug-receptor systems. In these systems, drug bound to receptors (B) relates to the concentration of free (unbound) drug (C) as depicted in Figure 2–1B and as described by an analogous equation:

$$B = \frac{B_{max} \times C}{C + K_d}$$

in which B_{max} indicates the total concentration of receptor sites (ie, sites bound to the drug at infinitely high concentrations of free drug) and K_d (the equilibrium dissociation constant) represents the concentration of free drug at which half-maximal binding is observed. This constant characterizes the receptor's affinity for binding the drug in a reciprocal fashion: If the K_d is low, binding affinity is high, and vice versa. The EC_{50} and K_d may be identical, but need not be, as discussed below. Dose-response data are often presented as a plot of the drug effect (ordinate) against the *logarithm* of the dose or concentration (abscissa). This mathematical maneuver transforms the hyperbolic curve of Figure 2–1 into a sigmoid curve with a linear midportion (eg, Figure 2–2). This expands the scale of the concentration axis at low concentrations (where the effect is changing rapidly) and compresses it at high concentrations (where the effect is changing slowly), but has no special biologic or pharmacologic significance.

Receptor-Effector Coupling & Spare Receptors

When a receptor is occupied by an agonist, the resulting conformational change is only the first of many steps usually required to

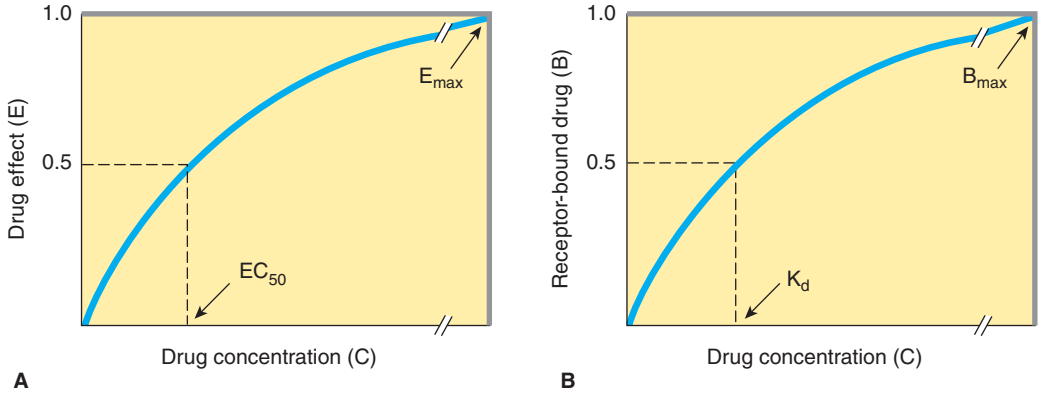


FIGURE 2–1 Relations between drug concentration and drug effect (**A**) or receptor-bound drug (**B**). The drug concentrations at which effect or receptor occupancy is half-maximal are denoted by EC_{50} and K_d, respectively.

produce a pharmacologic response. The transduction process that links drug occupancy of receptors and pharmacologic response is often termed **coupling.** The relative efficiency of occupancy-response coupling is partially determined by the initial conformational change in the receptor; thus, the effects of full agonists can be considered more efficiently coupled to receptor occupancy than can the effects of partial agonists (described in text that follows). Coupling efficiency is also determined by the biochemical events that transduce receptor occupancy into cellular response. Sometimes the biologic effect of the drug is linearly related to the number of receptors bound. This is often true for drug-regulated ion channels, eg, in which the ion current produced by the drug is directly proportional to the number of receptors (ion channels) bound. In other cases, the biologic response is a more complex function of drug binding to receptors. This is often true for receptors linked to enzymatic signal transduction cascades, eg, in which the biologic response often increases disproportionately to the number of receptors occupied by drug.

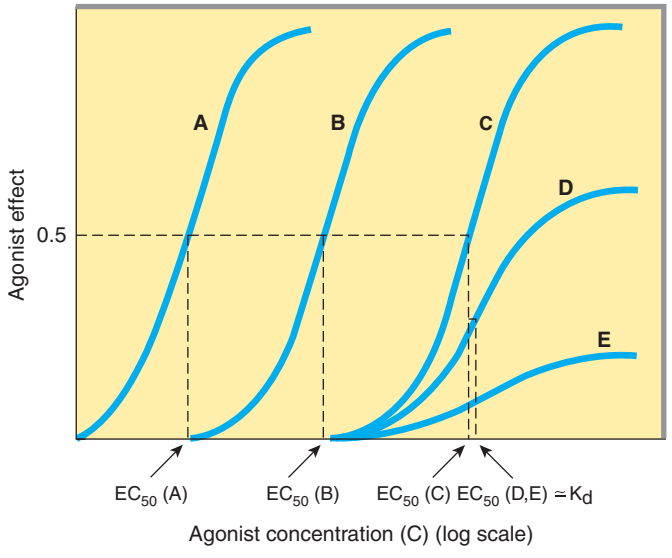


FIGURE 2–2 Logarithmic transformation of the dose axis and experimental demonstration of spare receptors, using different concentrations of an irreversible antagonist. Curve **A** shows agonist response in the absence of antagonist. After treatment with a low concentration of antagonist (curve **B**), the curve is shifted to the right. Maximal responsiveness is preserved, however, because the remaining available receptors are still in excess of the number required. In curve **C,** produced after treatment with a larger concentration of antagonist, the available receptors are no longer "spare"; instead, they are just sufficient to mediate an undiminished maximal response. Still higher concentrations of antagonist (curves **D** and **E**) reduce the number of available receptors to the point that maximal response is diminished. The apparent EC_{50} of the agonist in curves **D** and **E** may approximate the K_d that characterizes the binding affinity of the agonist for the receptor.

Many factors can contribute to nonlinear occupancy-response coupling, and often these factors are only partially understood. The concept of **"spare" receptors,** regardless of the precise biochemical mechanism involved, can help us to think about these effects. Receptors are said to be "spare" for a given pharmacologic response if it is possible to elicit a maximal biologic response at a concentration of agonist that does not result in occupancy of the full complement of available receptors. Experimentally, spare receptors may be demonstrated by using irreversible antagonists to prevent binding of agonist to a proportion of available receptors and showing that high concentrations of agonist can still produce an undiminished maximal response (Figure 2–2). Thus, the same maximal inotropic response of heart muscle to catecholamines can be elicited even under conditions in which 90% of the β adrenoceptors are occupied by a quasi-irreversible antagonist. Accordingly, myocardial cells are said to contain a large proportion of spare β adrenoceptors.

How can we account for the phenomenon of spare receptors? In the example of the β adrenoceptor, receptor activation promotes binding of guanosine triphosphate (GTP) to an intermediate signaling protein and activation of the signaling intermediate may greatly outlast the agonist-receptor interaction (see the following section on G Proteins & Second Messengers). In such a case, the "spareness" of receptors is *temporal.* Maximal

response can be elicited by activation of relatively few receptors because the response initiated by an individual ligand-receptor binding event persists longer than the binding event itself.

In other cases, in which the biochemical mechanism is not understood, we imagine that the receptors might be *spare in number*. If the concentration or amount of cellular components other than the receptors limits the coupling of receptor occupancy to response, then a maximal response can occur without occupancy of all receptors. Thus, the sensitivity of a cell or tissue to a particular concentration of agonist depends not only on the *affinity* of the receptor for binding the agonist (characterized by the K_d) but also on the *degree of spareness*—the total number of receptors present compared with the number actually needed to elicit a maximal biologic response.

The concept of spare receptors is very useful clinically because it allows one to think precisely about the effects of drug dosage without needing to consider biochemical details of the signaling response. The K_d of the agonist-receptor interaction determines what fraction (B/B_{max}) of total receptors will be occupied at a given free concentration (C) of agonist regardless of the receptor concentration:

$$\frac{B}{B_{max}} = \frac{C}{C + K_d}$$

Imagine a responding cell with four receptors and four effectors. Here the number of effectors does not limit the maximal response, and the receptors are *not* spare in number. Consequently, an agonist present at a concentration equal to the K_d will occupy 50% of the receptors, and half of the effectors will be activated, producing a half-maximal response (ie, two receptors stimulate two effectors). Now imagine that the number of receptors increases 10-fold to 40 receptors but that the total number of effectors remains constant. Most of the receptors are now spare in number. As a result, a much lower concentration of agonist suffices to occupy 2 of the 40 receptors (5% of the receptors), and this same low concentration of agonist is able to elicit a half-maximal response (two of four effectors activated). Thus, it is possible to change the sensitivity of tissues with spare receptors by changing receptor number.

Competitive & Irreversible Antagonists

Receptor antagonists bind to receptors but do not activate them. The primary action of antagonists is to prevent agonists (other drugs or endogenous regulatory molecules) from activating receptors. Some antagonists (so-called "inverse agonists," see Chapter 1), also reduce receptor activity below basal levels observed in the absence of bound ligand. Antagonists are divided into two classes depending on whether or not they *reversibly compete* with agonists for binding to receptors.

In the presence of a fixed concentration of agonist, increasing concentrations of a reversible **competitive antagonist** progressively inhibit the agonist response; high antagonist concentrations prevent response completely. Conversely, sufficiently high concentrations of agonist can surmount the effect of a given concentration of the antagonist; that is, the E_{max} for the agonist remains the same for any fixed concentration of antagonist (Figure 2–3A). Because the antagonism is competitive, the presence of antagonist increases the agonist concentration required

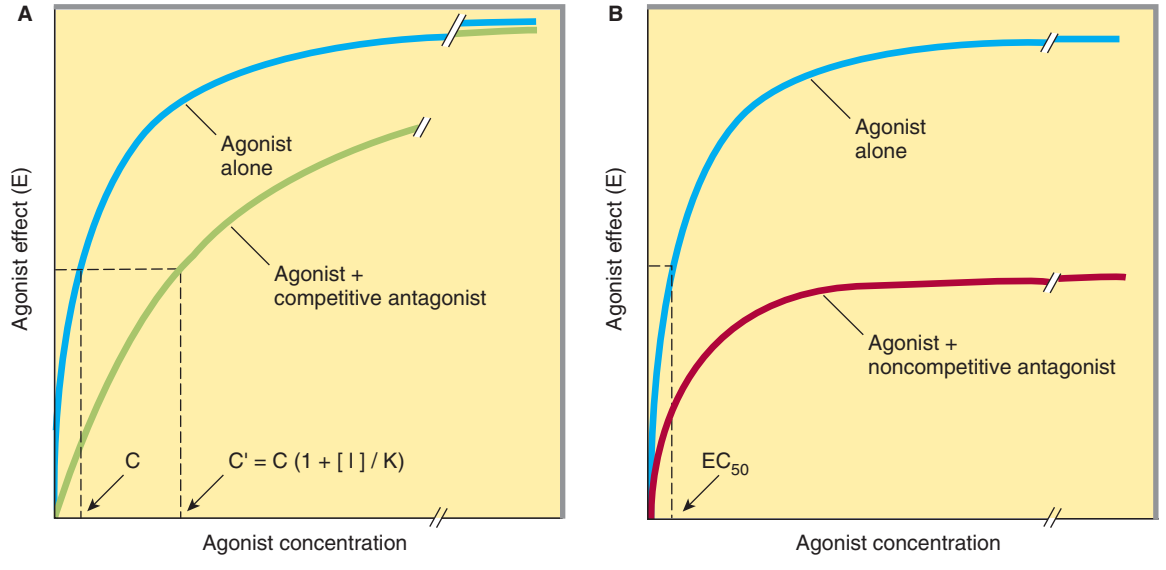


FIGURE 2–3 Changes in agonist concentration-effect curves produced by a competitive antagonist (**A**) or by an irreversible antagonist (**B**). In the presence of a competitive antagonist, higher concentrations of agonist are required to produce a given effect; thus the agonist concentration (C′) required for a given effect in the presence of concentration [I] of an antagonist is shifted to the right, as shown. High agonist concentrations can overcome inhibition by a competitive antagonist. This is not the case with an irreversible (or noncompetitive) antagonist, which reduces the maximal effect the agonist can achieve, although it may not change its EC_{50}.

for a given degree of response, and so the agonist concentration-effect curve is shifted to the right.

The concentration (C′) of an agonist required to produce a given effect in the presence of a fixed concentration ([I]) of competitive antagonist is greater than the agonist concentration (C) required to produce the same effect in the absence of the antagonist. The ratio of these two agonist concentrations (dose ratio) is related to the dissociation constant (K_i) of the antagonist by the Schild equation:

$$\frac{C'}{C} = 1 + \frac{[I]}{K_i}$$

Pharmacologists often use this relation to determine the K_i of a competitive antagonist. Even without knowledge of the relation between agonist occupancy of the receptor and response, the K_i can be determined simply and accurately. As shown in Figure 2–3, concentration-response curves are obtained in the presence and in the absence of a fixed concentration of competitive antagonist; comparison of the agonist concentrations required to produce identical degrees of pharmacologic effect in the two situations reveals the antagonist's K_i. If C′ is twice C, for example, then $[I] = K_i$.

For the clinician, this mathematical relation has two important therapeutic implications:

1. The degree of inhibition produced by a competitive antagonist depends on the concentration of antagonist. The competitive β-adrenoceptor antagonist propranolol provides a useful example. Patients receiving a fixed dose of this drug exhibit a wide range of plasma concentrations, owing to differences among individuals in clearance of propranolol. As a result, inhibitory effects on physiologic responses to norepinephrine and epinephrine (endogenous adrenergic receptor agonists) may vary widely, and the dose of propranolol must be adjusted accordingly.
2. Clinical response to a competitive antagonist also depends on the concentration of agonist that is competing for binding to receptors. Again, propranolol provides a useful example: When this drug is administered at moderate doses sufficient to block the effect of basal levels of the neurotransmitter norepinephrine, resting heart rate is decreased. However, the increase in the release of norepinephrine and epinephrine that occurs with exercise, postural changes, or emotional stress may suffice to overcome this competitive antagonism. Accordingly, the same dose of propranolol may have little effect under these conditions, thereby altering therapeutic response.

Some receptor antagonists bind to the receptor in an **irreversible** or nearly irreversible fashion, either by forming a covalent bond with the receptor or by binding so tightly that, for practical purposes, the receptor is unavailable for binding of agonist. After occupancy of some proportion of receptors by such an antagonist, the number of remaining unoccupied receptors may be too low for the agonist (even at high concentrations) to elicit a response comparable to the previous maximal response (Figure 2–3B). If spare receptors are present, however, a lower dose of an irreversible antagonist may leave enough receptors unoccupied to allow achievement of maximum response to agonist, although a higher agonist concentration will be required (Figure 2–2B and C; see Receptor-Effector Coupling & Spare Receptors).

Therapeutically, irreversible antagonists present distinct advantages and disadvantages. Once the irreversible antagonist has occupied the receptor, it need not be present in unbound form to inhibit agonist responses. Consequently, the duration of action of such an irreversible antagonist is relatively independent of its own rate of elimination and more dependent on the rate of turnover of receptor molecules.

Phenoxybenzamine, an irreversible α-adrenoceptor antagonist, is used to control the hypertension caused by catecholamines released from pheochromocytoma, a tumor of the adrenal medulla. If administration of phenoxybenzamine lowers blood pressure, blockade will be maintained even when the tumor episodically releases very large amounts of catecholamine. In this case, the ability to prevent responses to varying and high concentrations of agonist is a therapeutic advantage. If overdose occurs, however, a real problem may arise. If the α-adrenoceptor blockade cannot be overcome, excess effects of the drug must be antagonized "physiologically," ie, by using a pressor agent that does not act via α receptors.

Antagonists can function noncompetitively in a different way; that is, by binding to a site on the receptor protein separate from the agonist binding site, and thereby modifying receptor activity without blocking agonist binding (see Figure 1–3C and D). Although these drugs act noncompetitively, their actions are reversible if they do not bind covalently. Such drugs are often called *allosteric modulators*. For example, benzodiazepines bind noncompetitively to ion channels activated by the neurotransmitter γ-aminobutyric acid (GABA), enhancing the net activating effect of GABA on channel conductance.

Partial Agonists

Based on the maximal pharmacologic response that occurs when all receptors are occupied, agonists can be divided into two classes: **partial agonists** produce a lower response, at full receptor occupancy, than do **full agonists.** Partial agonists produce concentration-effect curves that resemble those observed with full agonists in the presence of an antagonist that irreversibly blocks some of the receptor sites (compare Figures 2–2 [curve D] and 2–4B). It is important to emphasize that the failure of partial agonists to produce a maximal response is not due to decreased affinity for binding to receptors. Indeed, a partial agonist's inability to cause a maximal pharmacologic response, even when present at high concentrations that saturate binding to all receptors, is indicated by the fact that partial agonists competitively inhibit the responses produced by full agonists (Figure 2–4C). Many drugs used clinically as antagonists are actually weak partial agonists. Partial agonism can be useful in some clinical circumstances. For example, buprenorphine, a partial agonist of μ-opioid receptors, is a generally safer analgesic drug than morphine because it produces less respiratory depression in overdose. Buprenorphine is effectively antianalgesic when administered to

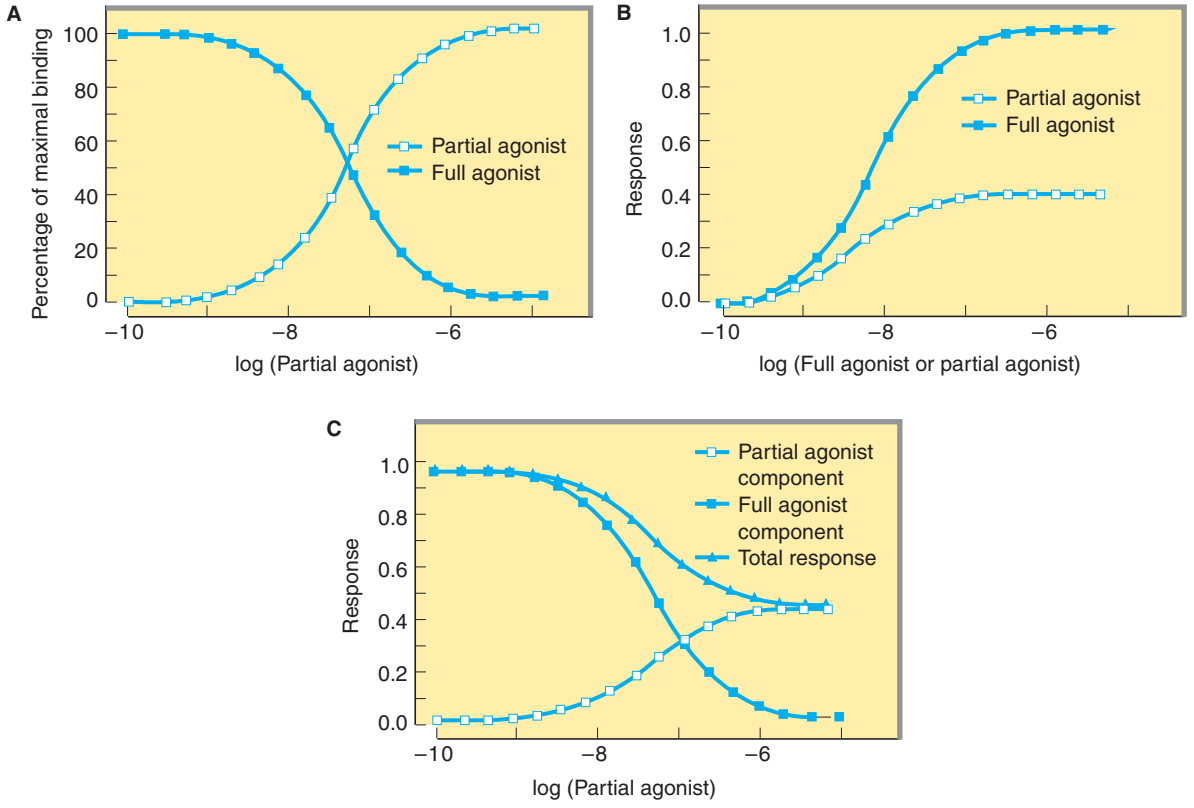


FIGURE 2–4 **A:** The percentage of receptor occupancy resulting from full agonist (present at a single concentration) binding to receptors in the presence of increasing concentrations of a partial agonist. Because the full agonist (filled squares) and the partial agonist (open squares) compete to bind to the same receptor sites, when occupancy by the partial agonist increases, binding of the full agonist decreases. **B:** When each of the two drugs is used alone and response is measured, occupancy of all the receptors by the partial agonist produces a lower maximal response than does similar occupancy by the full agonist. **C:** Simultaneous treatment with a single concentration of full agonist and increasing concentrations of the partial agonist produces the response patterns shown in the bottom panel. The fractional response caused by a single high concentration of the full agonist (filled squares) decreases as increasing concentrations of the partial agonist compete to bind to the receptor with increasing success; at the same time the portion of the response caused by the partial agonist (open squares) increases, while the total response—ie, the sum of responses to the two drugs (filled triangles)—gradually decreases, eventually reaching the value produced by partial agonist alone (compare with B).

morphine-dependent individuals, however, and may precipitate a drug withdrawal syndrome due to competitive inhibition of morphine's agonist action.

Other Mechanisms of Drug Antagonism

Not all the mechanisms of antagonism involve interactions of drugs or endogenous ligands at a single type of receptor, and some types of antagonism do not involve a receptor at all. For example, protamine, a protein that is positively charged at physiologic pH, can be used clinically to counteract the effects of heparin, an anticoagulant that is negatively charged. In this case, one drug acts as a **chemical antagonist** of the other simply by ionic binding that makes the other drug unavailable for interactions with proteins involved in blood clotting.

Another type of antagonism is **physiologic antagonism** between endogenous regulatory pathways mediated by different receptors. For example, several catabolic actions of the glucocorticoid hormones lead to increased blood sugar, an effect that is physiologically opposed by insulin. Although glucocorticoids and insulin act on quite distinct receptor-effector systems, the clinician must sometimes administer insulin to oppose the hyperglycemic effects of a glucocorticoid hormone, whether the latter is elevated by endogenous synthesis (eg, a tumor of the adrenal cortex) or as a result of glucocorticoid therapy.

In general, use of a drug as a physiologic antagonist produces effects that are less specific and less easy to control than are the effects of a receptor-specific antagonist. Thus, for example, to treat bradycardia caused by increased release of acetylcholine from vagus nerve endings, the physician could use isoproterenol, a β-adrenoceptor agonist that increases heart rate by mimicking sympathetic stimulation of the heart. However, use of this physiologic antagonist would be less rational—and potentially more dangerous—than would use of a receptor-specific antagonist such as atropine (a competitive antagonist at the receptors at which acetylcholine slows heart rate).

SIGNALING MECHANISMS & DRUG ACTION

Until now we have considered receptor interactions and drug effects in terms of equations and concentration-effect curves. We must also understand the molecular mechanisms by which a drug acts. Such understanding allows us to ask basic questions with important clinical implications:

- Why do some drugs produce effects that persist for minutes, hours, or even days after the drug is no longer present?
- Why do responses to other drugs diminish rapidly with prolonged or repeated administration?
- How do cellular mechanisms for amplifying external chemical signals explain the phenomenon of spare receptors?
- Why do chemically similar drugs often exhibit extraordinary selectivity in their actions?
- Do these mechanisms provide targets for developing new drugs?

Most transmembrane signaling is accomplished by a small number of different molecular mechanisms. Each type of mechanism has been adapted, through the evolution of distinctive protein families, to transduce many different signals. These protein families include receptors on the cell surface and within the cell, as well as enzymes and other components that generate, amplify, coordinate, and terminate postreceptor signaling by chemical second messengers in the cytoplasm. This section first discusses the mechanisms for carrying chemical information across the plasma membrane and then outlines key features of cytoplasmic second messengers.

Five basic mechanisms of transmembrane signaling are well understood (Figure 2–5). Each uses a different strategy to circumvent the barrier posed by the lipid bilayer of the plasma membrane. These strategies use (1) a lipid-soluble ligand that crosses the membrane and acts on an intracellular receptor; (2) a transmembrane receptor protein whose intracellular enzymatic activity is allosterically regulated by a ligand that binds to a site on the protein's extracellular domain; (3) a transmembrane receptor that binds and stimulates a protein tyrosine kinase; (4) a ligand-gated transmembrane ion channel that can be induced to open or close by the binding of a ligand; or (5) a transmembrane receptor protein that stimulates a GTP-binding signal transducer protein (G protein), which in turn modulates production of an intracellular second messenger.

Although the five established mechanisms do not account for all the chemical signals conveyed across cell membranes, they do transduce many of the most important signals exploited in pharmacotherapy.

Intracellular Receptors for Lipid-Soluble Agents

Several biologic ligands are sufficiently lipid-soluble to cross the plasma membrane and act on intracellular receptors. One class of such ligands includes steroids (corticosteroids, mineralocorticoids, sex steroids, vitamin D), and thyroid hormone, whose receptors stimulate the transcription of genes by binding to specific DNA sequences near the gene whose expression is to be regulated. Many of the target DNA sequences (called **response elements**) have been identified.

These "gene-active" receptors belong to a protein family that evolved from a common precursor. Dissection of the receptors by

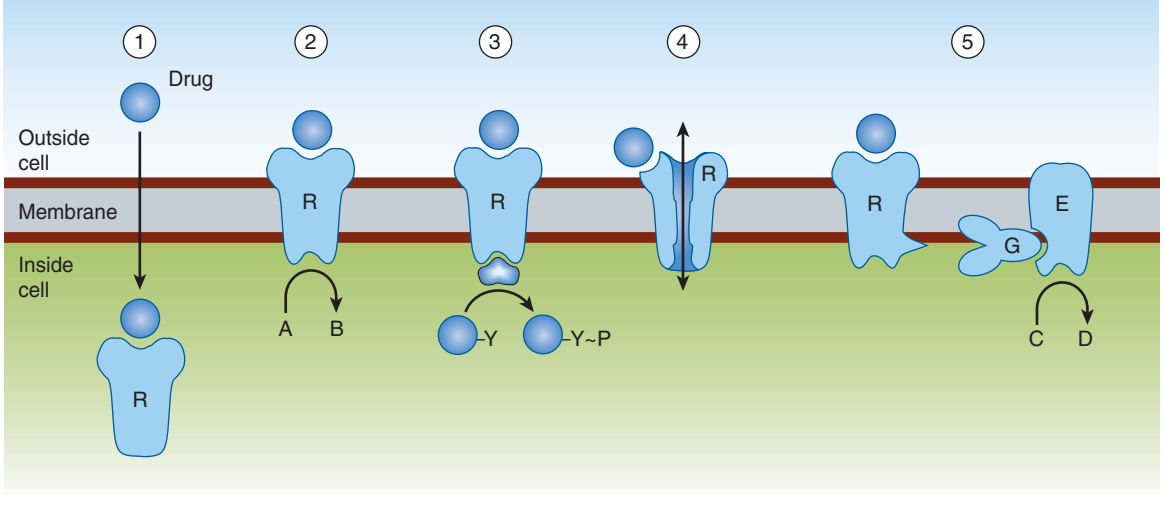


FIGURE 2–5 Known transmembrane signaling mechanisms: **1:** A lipid-soluble chemical signal crosses the plasma membrane and acts on an intracellular receptor (which may be an enzyme or a regulator of gene transcription); **2:** the signal binds to the extracellular domain of a transmembrane protein, thereby activating an enzymatic activity of its cytoplasmic domain; **3:** the signal binds to the extracellular domain of a transmembrane receptor bound to a separate protein tyrosine kinase, which it activates; **4:** the signal binds to and directly regulates the opening of an ion channel; **5:** the signal binds to a cell-surface receptor linked to an effector enzyme by a G protein. (A, C, substrates; B, D, products; R, receptor; G, G protein; E, effector [enzyme or ion channel]; Y, tyrosine; P, phosphate.)

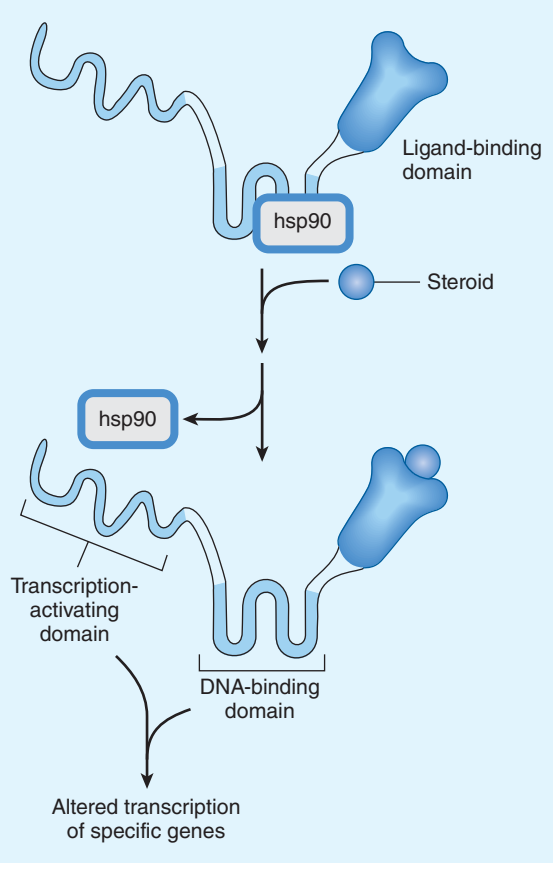


FIGURE 2–6 Mechanism of glucocorticoid action. The glucocorticoid receptor polypeptide is schematically depicted as a protein with three distinct domains. A heat-shock protein, hsp90, binds to the receptor in the absence of hormone and prevents folding into the active conformation of the receptor. Binding of a hormone ligand (steroid) causes dissociation of the hsp90 stabilizer and permits conversion to the active configuration.

recombinant DNA techniques has provided insights into their molecular mechanism. For example, binding of glucocorticoid hormone to its normal receptor protein relieves an inhibitory constraint on the transcription-stimulating activity of the protein. Figure 2–6 schematically depicts the molecular mechanism of glucocorticoid action: In the absence of hormone, the receptor is bound to hsp90, a protein that appears to prevent normal folding of several structural domains of the receptor. Binding of hormone to the ligand-binding domain triggers release of hsp90. This allows the DNA-binding and transcription-activating domains of the receptor to fold into their functionally active conformations, so that the activated receptor can initiate transcription of target genes.

The mechanism used by hormones that act by regulating gene expression has two therapeutically important consequences:

1. All of these hormones produce their effects after a characteristic lag period of 30 minutes to several hours—the time required for the synthesis of new proteins. This means that the gene-active hormones cannot be expected to alter a pathologic state within minutes (eg, glucocorticoids will not immediately relieve the symptoms of acute bronchial asthma).
2. The effects of these agents can persist for hours or days after the agonist concentration has been reduced to zero. The persistence of effect is primarily due to the relatively slow turnover of most enzymes and proteins, which can remain active in cells for hours or days after they have been synthesized. Consequently, it means that the beneficial (or toxic) effects of a gene-active hormone usually decrease slowly when administration of the hormone is stopped.

Ligand-Regulated Transmembrane Enzymes Including Receptor Tyrosine Kinases

This class of receptor molecules mediates the first steps in signaling by insulin, epidermal growth factor (EGF), platelet-derived growth factor (PDGF), atrial natriuretic peptide (ANP), transforming growth factor-β (TGF-β), and many other trophic hormones. These receptors are polypeptides consisting of an extracellular hormone-binding domain and a cytoplasmic enzyme domain, which may be a protein tyrosine kinase, a serine kinase, or a guanylyl cyclase (Figure 2–7). In all these receptors, the two domains are connected by a hydrophobic segment of the polypeptide that crosses the lipid bilayer of the plasma membrane.

The receptor tyrosine kinase signaling pathway begins with binding of ligand, typically a polypeptide hormone or growth factor, to the receptor's extracellular domain. The resulting change in receptor conformation causes two receptor molecules to bind to one another *(dimerize)*, which in turn brings together the tyrosine kinase domains, which become enzymatically active, and phosphorylate one another as well as additional downstream signaling proteins. Activated receptors catalyze phosphorylation of tyrosine residues on different target signaling proteins, thereby allowing a single type of activated receptor to modulate a number of biochemical processes. (Some receptor tyrosine kinases form oligomeric complexes larger than dimers upon activation by ligand, but the pharmacologic significance of such higher-order complexes is presently unclear.)

Insulin, for example, uses a single class of receptors to trigger increased uptake of glucose and amino acids and to regulate metabolism of glycogen and triglycerides in the cell. Similarly, each of the growth factors initiates in its specific target cells a complex program of cellular events ranging from altered membrane transport of ions and metabolites to changes in the expression of many genes.

Inhibitors of receptor tyrosine kinases are finding increased use in neoplastic disorders in which excessive growth factor signaling is often involved. Some of these inhibitors are monoclonal antibodies (eg, trastuzumab, cetuximab), which bind to the extracellular domain of a particular receptor and interfere with binding of growth factor. Other inhibitors are membrane-permeant "small molecule" chemicals (eg, gefitinib, erlotinib), which inhibit the receptor's kinase activity in the cytoplasm.

The intensity and duration of action of EGF, PDGF, and other agents that act via receptor tyrosine kinases are limited by a process

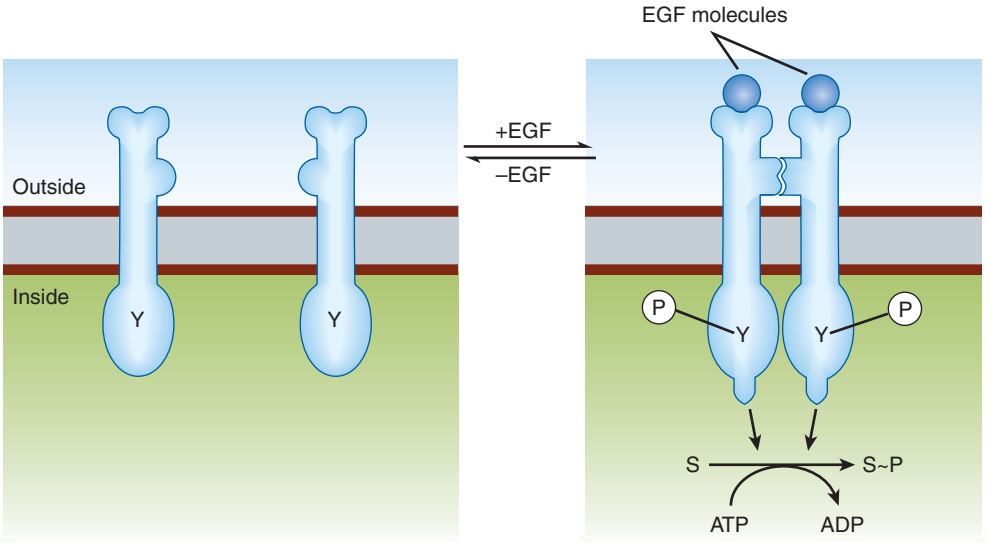


FIGURE 2–7 Mechanism of activation of the epidermal growth factor (EGF) receptor, a representative receptor tyrosine kinase. The receptor polypeptide has extracellular and cytoplasmic domains, depicted above and below the plasma membrane. Upon binding of EGF (circle), the receptor converts from its inactive monomeric state (*left*) to an active dimeric state (*right*), in which two receptor polypeptides bind noncovalently. The cytoplasmic domains become phosphorylated (P) on specific tyrosine residues (Y), and their enzymatic activities are activated, catalyzing phosphorylation of substrate proteins (S).

called receptor **down-regulation.** Ligand binding often induces accelerated endocytosis of receptors from the cell surface, followed by the degradation of those receptors (and their bound ligands). When this process occurs at a rate faster than de novo synthesis of receptors, the total number of cell-surface receptors is reduced (down-regulated), and the cell's responsiveness to ligand is correspondingly diminished. A well-understood example is the EGF receptor tyrosine kinase, which undergoes rapid endocytosis followed by proteolysis in lysosomes after EGF binding; genetic mutations that interfere with this process cause excessive growth factor-induced cell proliferation and are associated with an increased susceptibility to certain types of cancer. Endocytosis of other receptor tyrosine kinases, most notably receptors for nerve growth factor, serves a very different function. Internalized nerve growth factor receptors are not rapidly degraded and are translocated in endocytic vesicles from the distal axon, where receptors are activated by nerve growth factor released from the innervated tissue, to the cell body. In the cell body, the growth factor signal is transduced to transcription factors regulating the expression of genes controlling cell survival. This process effectively transports a critical survival signal from its site of release to its site of signaling effect, and does so over a remarkably long distance—up to 1 meter in certain sensory neurons.

A number of regulators of growth and differentiation, including TGF-β, act on another class of transmembrane receptor enzymes that phosphorylate serine and threonine residues. ANP, an important regulator of blood volume and vascular tone, acts on a transmembrane receptor whose intracellular domain, a guanylyl cyclase, generates cGMP (see below). Receptors in both groups, like the receptor tyrosine kinases, are active in their dimeric forms.

Cytokine Receptors

Cytokine receptors respond to a heterogeneous group of peptide ligands, which include growth hormone, erythropoietin, several kinds of interferon, and other regulators of growth and differentiation. These receptors use a mechanism (Figure 2–8) closely resembling that of receptor tyrosine kinases, except that in this case, the protein tyrosine kinase activity is not intrinsic to the receptor molecule. Instead, a separate protein tyrosine kinase, from the Janus-kinase (JAK) family, binds noncovalently to the receptor. As in the case of the EGF receptor, cytokine receptors dimerize after they bind the activating ligand, allowing the bound JAKs to become activated and to phosphorylate tyrosine residues on the receptor. Phosphorylated tyrosine residues on the receptor's cytoplasmic surface then set in motion a complex signaling dance by binding another set of proteins, called STATs (signal transducers and activators of transcription). The bound STATs are themselves phosphorylated by the JAKs, two STAT molecules dimerize (attaching to one another's tyrosine phosphates), and finally the STAT/STAT dimer dissociates from the receptor and travels to the nucleus, where it regulates transcription of specific genes.

Ligand- and Voltage-Gated Channels

Many of the most useful drugs in clinical medicine act by mimicking or blocking the actions of endogenous ligands that regulate the flow of ions through plasma membrane channels. The natural ligands are acetylcholine, serotonin, GABA, and glutamate. All of these agents are synaptic transmitters.

Each of their receptors transmits its signal across the plasma membrane by increasing transmembrane conductance of the

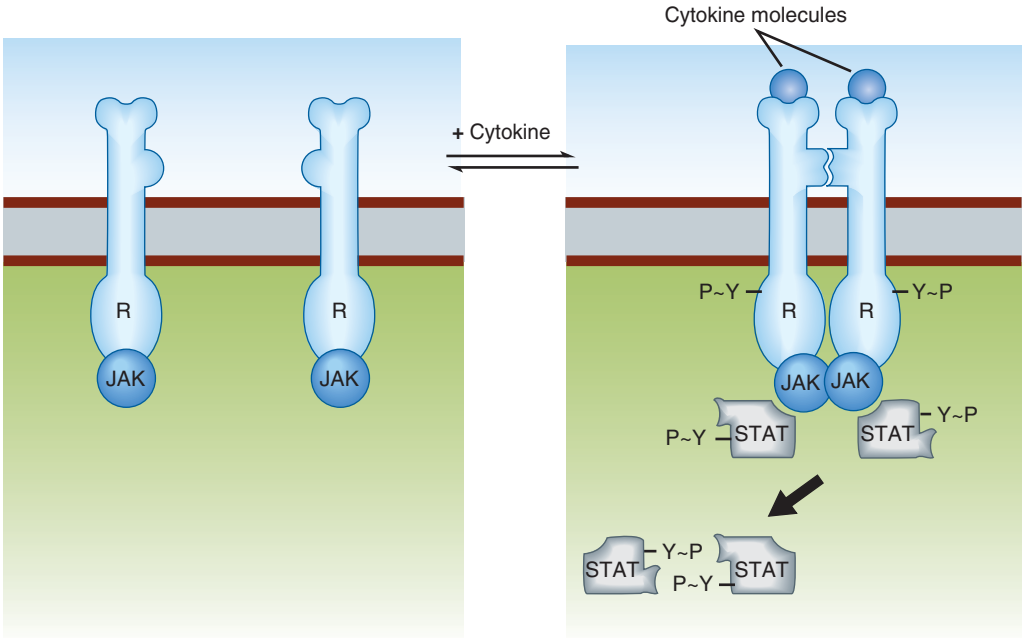


FIGURE 2–8 Cytokine receptors, like receptor tyrosine kinases, have extracellular and intracellular domains and form dimers. However, after activation by an appropriate ligand, separate mobile protein tyrosine kinase molecules (JAK) are activated, resulting in phosphorylation of signal transducers and activation of transcription (STAT) molecules. STAT dimers then travel to the nucleus, where they regulate transcription.

relevant ion and thereby altering the electrical potential across the membrane. For example, acetylcholine causes the opening of the ion channel in the nicotinic acetylcholine receptor (nAChR), which allows Na^+ to flow down its concentration gradient into cells, producing a localized excitatory postsynaptic potential—a depolarization.

The nAChR is one of the best characterized of all cell-surface receptors for hormones or neurotransmitters (Figure 2–9). One form of this receptor is a pentamer made up of four different polypeptide subunits (eg, two α chains plus one β, one γ, and one δ chain, all with molecular weights ranging from 43,000 to 50,000). These polypeptides, each of which crosses the lipid bilayer four times, form a cylindrical structure that is 8 nm in diameter. When acetylcholine binds to sites on the α subunits, a conformational change occurs that results in the transient opening of a central aqueous channel through which sodium ions penetrate from the extracellular fluid into the cell.

The time elapsed between the binding of the agonist to a ligand-gated channel and the cellular response can often be measured in milliseconds. The rapidity of this signaling mechanism is crucially important for moment-to-moment transfer of information across synapses. Ligand-gated ion channels can be regulated by multiple mechanisms, including phosphorylation and endocytosis. In the central nervous system, these mechanisms contribute to synaptic plasticity involved in learning and memory.

Voltage-gated ion channels do not bind neurotransmitters directly but are controlled by membrane potential; such channels are also important drug targets. For example, verapamil inhibits voltage-gated calcium channels that are present in the heart and in vascular smooth muscle, producing antiarrhythmic effects and reducing blood pressure without mimicking or antagonizing any known endogenous transmitter.

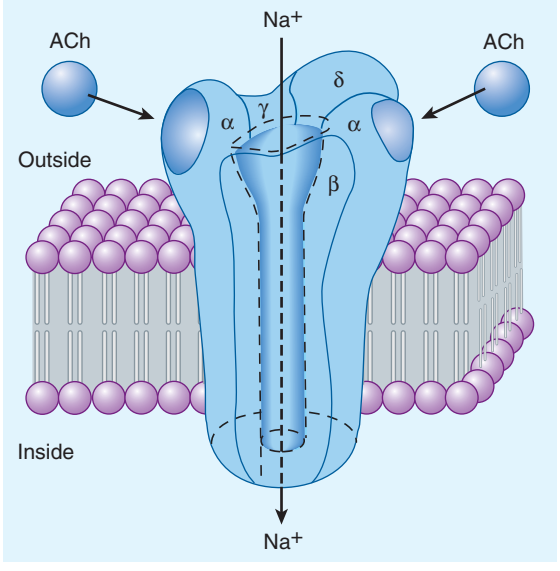


FIGURE 2–9 The nicotinic acetylcholine (ACh) receptor, a ligand-gated ion channel. The receptor molecule is depicted as embedded in a rectangular piece of plasma membrane, with extracellular fluid above and cytoplasm below. Composed of five subunits (two α, one β, one γ, and one δ), the receptor opens a central transmembrane ion channel when ACh binds to sites on the extracellular domain of its α subunits.

G Proteins & Second Messengers

Many extracellular ligands act by increasing the intracellular concentrations of second messengers such as **cyclic adenosine-3′,5′-monophosphate (cAMP), calcium ion,** or the **phosphoinositides** (described below). In most cases, they use a transmembrane signaling system with three separate components. First, the extracellular ligand is selectively detected by a cell-surface receptor. The receptor in turn triggers the activation of a G protein located on the cytoplasmic face of the plasma membrane. The activated G protein then changes the activity of an effector element, usually an enzyme or ion channel. This element then changes the concentration of the intracellular second messenger. For cAMP, the effector enzyme is adenylyl cyclase, a membrane protein that converts intracellular adenosine triphosphate (ATP) to cAMP. The corresponding G protein, G_s, stimulates adenylyl cyclase after being activated by hormones and neurotransmitters that act via specific G_s-coupled receptors. There are many examples of such receptors, including β adrenoceptors, glucagon receptors, thyrotropin receptors, and certain subtypes of dopamine and serotonin receptors.

G_s and other G proteins use a molecular mechanism that involves binding and hydrolysis of GTP (Figure 2–10). This mechanism allows the transduced signal to be amplified. For example, a neurotransmitter such as norepinephrine may encounter its membrane receptor for only a few milliseconds. When the encounter generates a GTP-bound G_s molecule, however, the duration of activation of adenylyl cyclase depends on the longevity of GTP binding to G_s rather than on the receptor's affinity for norepinephrine. Indeed, like other G proteins, GTP-bound G_s may remain active for tens of seconds, enormously amplifying the original signal. This mechanism also helps explain how signaling by G proteins produces the phenomenon of spare receptors. The family of G proteins contains several functionally diverse subfamilies (Table 2–1), each of which mediates effects of a particular set of receptors to a distinctive group of effectors. Note that an endogenous ligand (eg, norepinephrine, acetylcholine, serotonin, many others not listed in Table 2–1) may bind and stimulate receptors that couple to different subsets of G proteins. The apparent promiscuity of such a ligand allows it to elicit different G protein-dependent responses in different cells. For instance, the body responds to danger by using catecholamines (norepinephrine and epinephrine) both to increase heart rate and to induce constriction of blood vessels in the skin, by acting on G_s-coupled β adrenoceptors and G_q-coupled α_1 adrenoceptors, respectively. Ligand promiscuity also offers opportunities in drug development (see Receptor Classes & Drug Development in the following text).

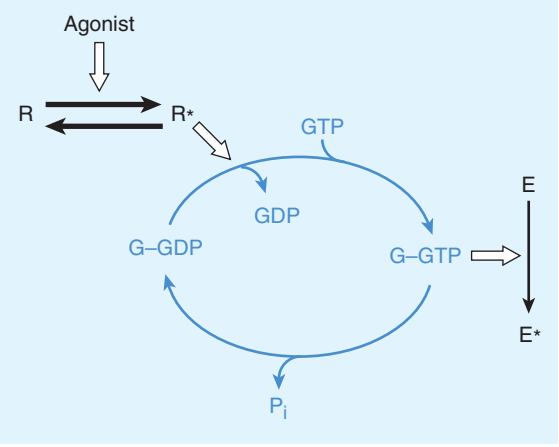


FIGURE 2–10 The guanine nucleotide-dependent activation-inactivation cycle of G proteins. The agonist activates the receptor (R→R*), which promotes release of GDP from the G protein (G), allowing entry of GTP into the nucleotide binding site. In its GTP-bound state (GGTP), the G protein regulates activity of an effector enzyme or ion channel (E→E*). The signal is terminated by hydrolysis of GTP, followed by return of the system to the basal unstimulated state. Open arrows denote regulatory effects. (P_i, inorganic phosphate.)

Receptors coupled to G proteins are often called "G protein-coupled receptors" **(GPCRs)**, "seven-transmembrane" (7-TM), or "serpentine" receptors. GPCRs make up the largest receptor family and are so-named because the receptor polypeptide chain "snakes" across the plasma membrane seven times (Figure 2–11). Receptors for adrenergic amines, serotonin, acetylcholine (muscarinic but not nicotinic), many peptide hormones, odorants, and even visual receptors (in retinal rod and cone cells) all belong to the GPCR family. All were derived from a common evolutionary precursor. A few GPCRs (eg, $GABA_B$ and metabotropic glutamate receptors) require stable assembly into either *homodimers* (complexes of two identical receptor polypeptides) or *heterodimers* (complexes of different isoforms) for functional activity. However, in contrast to tyrosine kinase and cytokine receptors, most GPCRs are thought to be able to function as monomers.

All GPCRs transduce signals across the plasma membrane in essentially the same way. Often the agonist ligand—eg, a catecholamine or acetylcholine—is bound in a pocket enclosed by the transmembrane regions of the receptor (as in Figure 2–11). The resulting change in conformation of these regions is transmitted to cytoplasmic loops of the receptor, which in turn activate the appropriate G protein by promoting replacement of GDP by GTP, as described above. Amino acids in the third cytoplasmic loop of the GPCR polypeptide are generally thought to play a key role in mediating receptor interaction with G proteins (shown by arrows in Figure 2–11). The structural basis for ligand binding to β adrenoceptors was determined recently using X-ray crystallography.

Receptor Regulation

G protein-mediated responses to drugs and hormonal agonists often attenuate with time (Figure 2–12, top). After reaching an initial high level, the response (eg, cellular cAMP accumulation, Na^+ influx, contractility, etc) diminishes over seconds or minutes, even in the continued presence of the agonist. This **"desensitization"** is often rapidly reversible; a second exposure to agonist, if provided a few minutes after termination of the first exposure, results in a response similar to the initial response.

TABLE 2–1 G proteins and their receptors and effectors.

G Protein	Receptors for	Effector/Signaling Pathway
G_s	β-Adrenergic amines, glucagon, histamine, serotonin, and many other hormones	↑ Adenylyl cyclase →↑ cAMP
G_{i1}, G_{i2}, G_{i3}	α_2-Adrenergic amines, acetylcholine (muscarinic), opioids, serotonin, and many others	Several, including: ↓ Adenylyl cyclase →↓ cAMP Open cardiac K^+ channels →↓ heart rate
G_{olf}	Odorants (olfactory epithelium)	↑ Adenylyl cyclase →↑ cAMP
G_o	Neurotransmitters in brain (not yet specifically identified)	Not yet clear
G_q	Acetylcholine (muscarinic), bombesin, serotonin (5-HT_2), and many others	↑ Phospholipase C →↑ IP_3, diacylglycerol, cytoplasmic Ca^{2+}
G_{t1}, G_{t2}	Photons (rhodopsin and color opsins in retinal rod and cone cells)	↑ cGMP phosphodiesterase →↓ cGMP (phototransduction)

cAMP, cyclic adenosine monophosphate; cGMP, cyclic guanosine monophosphate.

Many GPCRs are regulated by phosphorylation, as illustrated by rapid desensitization of the β adrenoceptor. The agonist-induced change in conformation of the receptor causes it to bind, activate, and serve as a substrate for a family of specific receptor kinases, called G protein-coupled receptor kinases (GRKs). The activated GRK then phosphorylates serine residues in the receptor's carboxyl terminal tail (Figure 2-12, panel B). The presence of phosphoserines increases the receptor's affinity for binding a third protein, β-arrestin. Binding of β-arrestin to cytoplasmic loops of the receptor diminishes the receptor's ability to interact with G_s, thereby reducing the agonist response (ie, stimulation of adenylyl cyclase). Upon removal of agonist, GRK activation is terminated, and the desensitization process can be reversed by cellular phosphatases.

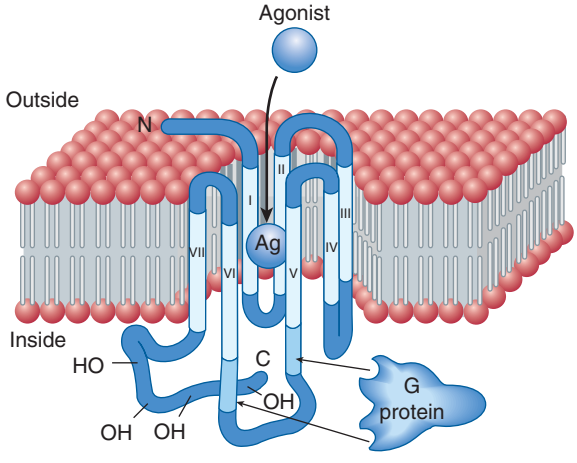


FIGURE 2–11 Transmembrane topology of a typical "serpentine" GPCR. The receptor's amino (N) terminal is extracellular (above the plane of the membrane), and its carboxyl (C) terminal intracellular. The terminals are connected by a polypeptide chain that traverses the plane of the membrane seven times. The hydrophobic transmembrane segments (light color) are designated by Roman numerals (I–VII). The agonist (Ag) approaches the receptor from the extracellular fluid and binds to a site surrounded by the transmembrane regions of the receptor protein. G proteins interact with cytoplasmic regions of the receptor, especially with portions of the third cytoplasmic loop between transmembrane regions V and VI. The receptor's cytoplasmic terminal tail contains numerous serine and threonine residues whose hydroxyl (-OH) groups can be phosphorylated. This phosphorylation may be associated with diminished receptor-G protein interaction.

For β adrenoceptors, and many other GPCRs, β-arrestin binding also accelerates endocytosis of receptors from the plasma membrane. Endocytosis of receptors promotes their dephosphorylation by a receptor phosphatase that is present at high concentration on endosome membranes, and receptors then return to the plasma membrane. This helps explain the ability of cells to recover receptor-mediated signaling responsiveness very efficiently after agonist-induced desensitization. Several GPCRs—including β adrenoceptors if persistently activated—instead traffic to lysosomes after endocytosis and are degraded. This process effectively attenuates (rather than restores) cellular responsiveness, similar to the process of down-regulation described above for the epidermal growth factor receptor. Thus, depending on the particular receptor and duration of activation, endocytosis can contribute to either rapid recovery or prolonged attenuation of cellular responsiveness (Figure 2–12).

Well-Established Second Messengers

A. Cyclic Adenosine Monophosphate (cAMP)

Acting as an intracellular second messenger, cAMP mediates such hormonal responses as the mobilization of stored energy (the breakdown of carbohydrates in liver or triglycerides in fat cells stimulated by β-adrenomimetic catecholamines), conservation of water by the kidney (mediated by vasopressin), Ca^{2+} homeostasis (regulated by parathyroid hormone), and increased rate and contractile force of heart muscle (β-adrenomimetic catecholamines). It also regulates the production of adrenal and sex steroids (in response to corticotropin or follicle-stimulating hormone), relaxation of smooth muscle, and many other endocrine and neural processes.

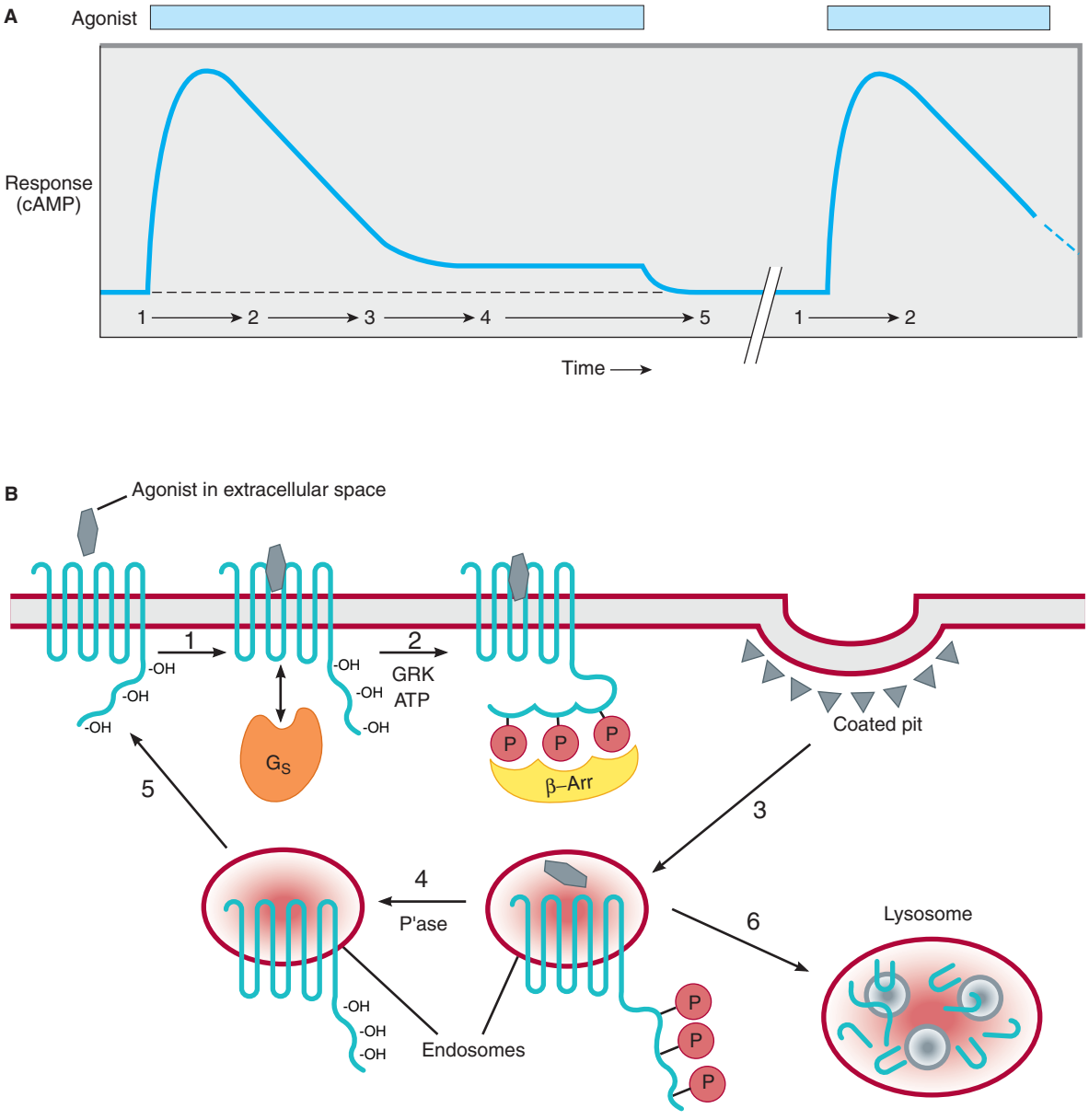


FIGURE 2–12 Rapid desensitization, resensitization, and down-regulation of β adrenoceptors. **A:** Response to a β-adrenoceptor agonist (ordinate) versus time (abscissa). (Numbers refer to the phases of receptor function in B.) Exposure of cells to agonist (indicated by the light-colored bar) produces a cyclic AMP response. A reduced cAMP response is observed in the continued presence of agonist; this "desensitization" typically occurs within a few minutes. If agonist is removed after a short time (typically several to tens of minutes, indicated by broken line on abscissa), cells recover full responsiveness to a subsequent addition of agonist (second light-colored bar). This "resensitization" fails to occur, or occurs incompletely, if cells are exposed to agonist repeatedly or over a more prolonged time period. **B:** Agonist binding to receptors initiates signaling by promoting receptor interaction with G proteins (G_s) located in the cytoplasm (step 1 in the diagram). Agonist-activated receptors are phosphorylated by a G protein-coupled receptor kinase (GRK), preventing receptor interaction with G_s and promoting binding of a different protein, β-arrestin (β-Arr), to the receptor (step 2). The receptor-arrestin complex binds to coated pits, promoting receptor internalization (step 3). Dissociation of agonist from internalized receptors reduces β-Arr binding affinity, allowing dephosphorylation of receptors by a phosphatase (P'ase, step 4) and return of receptors to the plasma membrane (step 5); together, these events result in the efficient resensitization of cellular responsiveness. Repeated or prolonged exposure of cells to agonist favors the delivery of internalized receptors to lysosomes (step 6), promoting receptor down-regulation rather than resensitization.

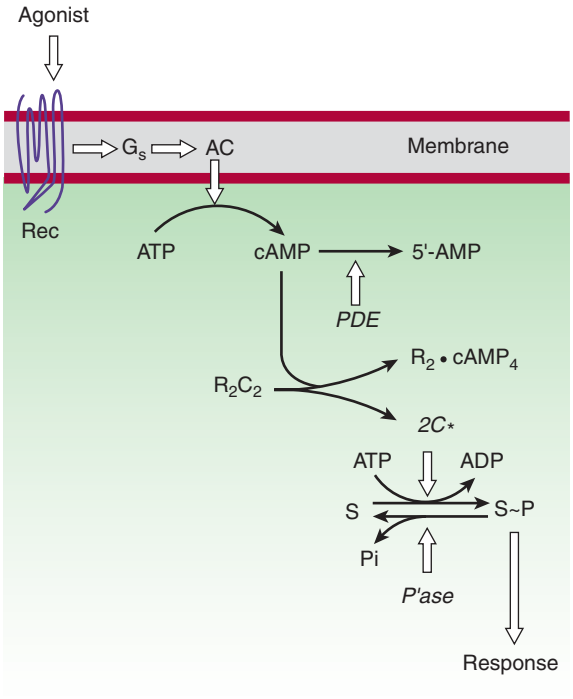


FIGURE 2–13 The cAMP second messenger pathway. Key proteins include hormone receptors (Rec), a stimulatory G protein (G_s), catalytic adenylyl cyclase (AC), phosphodiesterases (PDE) that hydrolyze cAMP, cAMP-dependent kinases, with regulatory (R) and catalytic (C) subunits, protein substrates (S) of the kinases, and phosphatases (P'ase), which remove phosphates from substrate proteins. Open arrows denote regulatory effects.

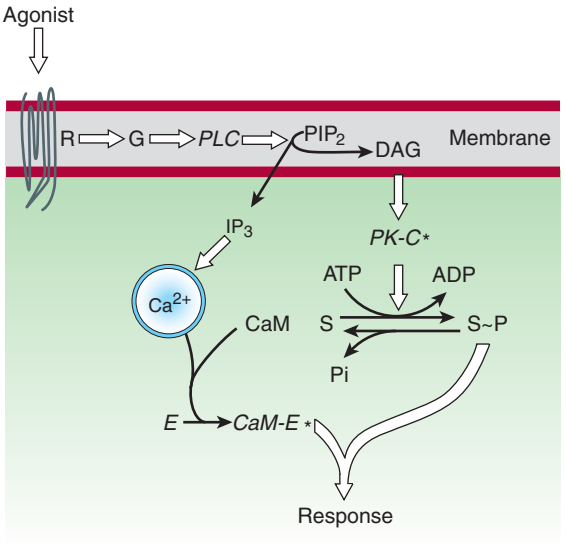


FIGURE 2–14 The Ca^{2+}-phosphoinositide signaling pathway. Key proteins include hormone receptors (R), a G protein (G), a phosphoinositide-specific phospholipase C (PLC), protein kinase C substrates of the kinase (S), calmodulin (CaM), and calmodulin-binding enzymes (E), including kinases, phosphodiesterases, etc. (PIP_2, phosphatidylinositol-4,5-bisphosphate; DAG, diacylglycerol; IP_3, inositol trisphosphate. Asterisk denotes activated state. Open arrows denote regulatory effects.)

cAMP exerts most of its effects by stimulating cAMP-dependent protein kinases (Figure 2–13). These kinases are composed of a cAMP-binding regulatory (R) dimer and two catalytic (C) chains. When cAMP binds to the R dimer, active C chains are released to diffuse through the cytoplasm and nucleus, where they transfer phosphate from ATP to appropriate substrate proteins, often enzymes. The specificity of the regulatory effects of cAMP resides in the distinct protein substrates of the kinases that are expressed in different cells. For example, liver is rich in phosphorylase kinase and glycogen synthase, enzymes whose reciprocal regulation by cAMP-dependent phosphorylation governs carbohydrate storage and release.

When the hormonal stimulus stops, the intracellular actions of cAMP are terminated by an elaborate series of enzymes. cAMP-stimulated phosphorylation of enzyme substrates is rapidly reversed by a diverse group of specific and nonspecific phosphatases. cAMP itself is degraded to 5′-AMP by several cyclic nucleotide phosphodiesterases (PDE; Figure 2–13). Milrinone, a selective inhibitor of type 3 phosphodiesterases that are expressed in cardiac muscle cells, has been used as an adjunctive agent in treating acute heart failure. Competitive inhibition of cAMP degradation is one way that caffeine, theophylline, and other methylxanthines produce their effects (see Chapter 20).

B. Phosphoinositides and Calcium

Another well-studied second messenger system involves hormonal stimulation of phosphoinositide hydrolysis (Figure 2–14). Some of the hormones, neurotransmitters, and growth factors that trigger this pathway bind to receptors linked to G proteins, whereas others bind to receptor tyrosine kinases. In all cases, the crucial step is stimulation of a membrane enzyme, phospholipase C (PLC), which splits a minor phospholipid component of the plasma membrane, phosphatidylinositol-4,5-bisphosphate (PIP_2), into two second messengers, **diacylglycerol (DAG)** and **inositol-1,4,5-trisphosphate (IP_3 or $InsP_3$).** Diacylglycerol is confined to the membrane, where it activates a phospholipid- and calcium-sensitive protein kinase called protein kinase C. IP_3 is water-soluble and diffuses through the cytoplasm to trigger release of Ca^{2+} by binding to ligand-gated calcium channels in the limiting membranes of internal storage vesicles. Elevated cytoplasmic Ca^{2+} concentration resulting from IP_3-promoted opening of these channels promotes the binding of Ca^{2+} to the calcium-binding protein calmodulin, which regulates activities of other enzymes, including calcium-dependent protein kinases.

With its multiple second messengers and protein kinases, the phosphoinositide signaling pathway is much more complex than the cAMP pathway. For example, different cell types may contain one or more specialized calcium- and calmodulin-dependent kinases with limited substrate specificity (eg, myosin light-chain kinase) in addition to a general calcium- and calmodulin-dependent kinase that can phosphorylate a wide variety of protein

substrates. Furthermore, at least nine structurally distinct types of protein kinase C have been identified.

As in the cAMP system, multiple mechanisms damp or terminate signaling by this pathway. IP_3 is inactivated by dephosphorylation; diacylglycerol is either phosphorylated to yield phosphatidic acid, which is then converted back into phospholipids, or it is deacylated to yield arachidonic acid; Ca^{2+} is actively removed from the cytoplasm by Ca^{2+} pumps.

These and other nonreceptor elements of the calcium-phosphoinositide signaling pathway are of considerable importance in pharmacotherapy. For example, lithium ion, used in treatment of bipolar (manic-depressive) disorder, affects the cellular metabolism of phosphoinositides (see Chapter 29).

C. Cyclic Guanosine Monophosphate (cGMP)

Unlike cAMP, the ubiquitous and versatile carrier of diverse messages, cGMP has established signaling roles in only a few cell types. In intestinal mucosa and vascular smooth muscle, the cGMP-based signal transduction mechanism closely parallels the cAMP-mediated signaling mechanism. Ligands detected by cell-surface receptors stimulate membrane-bound guanylyl cyclase to produce cGMP, and cGMP acts by stimulating a cGMP-dependent protein kinase. The actions of cGMP in these cells are terminated by enzymatic degradation of the cyclic nucleotide and by dephosphorylation of kinase substrates.

Increased cGMP concentration causes relaxation of vascular smooth muscle by a kinase-mediated mechanism that results in dephosphorylation of myosin light chains (see Figure 12–2). In these smooth muscle cells, cGMP synthesis can be elevated by two transmembrane signaling mechanisms utilizing two different guanylyl cyclases. Atrial natriuretic peptide, a blood-borne peptide hormone, stimulates a transmembrane receptor by binding to its extracellular domain, thereby activating the guanylyl cyclase activity that resides in the receptor's intracellular domain. The other mechanism mediates responses to nitric oxide (NO; see Chapter 19), which is generated in vascular endothelial cells in response to natural vasodilator agents such as acetylcholine and histamine. After entering the target cell, nitric oxide binds to and activates a cytoplasmic guanylyl cyclase (see Figure 19–2). A number of useful vasodilating drugs, such as nitroglycerin and sodium nitroprusside used in treating cardiac ischemia and acute hypertension, act by generating or mimicking nitric oxide. Other drugs produce vasodilation by inhibiting specific phosphodiesterases, thereby interfering with the metabolic breakdown of cGMP. One such drug is sildenafil, used in treating erectile dysfunction and pulmonary hypertension (see Chapter 12).

Interplay among Signaling Mechanisms

The calcium-phosphoinositide and cAMP signaling pathways oppose one another in some cells and are complementary in others. For example, vasopressor agents that contract smooth muscle act by IP_3-mediated mobilization of Ca^{2+}, whereas agents that relax smooth muscle often act by elevation of cAMP. In contrast, cAMP and phosphoinositide second messengers act together to stimulate glucose release from the liver.

Phosphorylation: A Common Theme

Almost all second messenger signaling involves reversible phosphorylation, which performs two principal functions in signaling: amplification and flexible regulation. In **amplification,** rather like GTP bound to a G protein, the attachment of a phosphoryl group to a serine, threonine, or tyrosine residue powerfully amplifies the initial regulatory signal by recording a molecular memory that the pathway has been activated; dephosphorylation erases the memory, taking a longer time to do so than is required for dissociation of an allosteric ligand. In **flexible regulation,** differing substrate specificities of the multiple protein kinases regulated by second messengers provide branch points in signaling pathways that may be independently regulated. In this way, cAMP, Ca^{2+}, or other second messengers can use the presence or absence of particular kinases or kinase substrates to produce quite different effects in different cell types. Inhibitors of protein kinases have great potential as therapeutic agents, particularly in neoplastic diseases. Trastuzumab, an antibody that antagonizes growth factor receptor signaling (discussed earlier), is a useful therapeutic agent for breast cancer. Another example of this general approach is imatinib, a small molecule inhibitor of the cytoplasmic tyrosine kinase Abl, which is activated by growth factor signaling pathways. Imatinib is effective for treating chronic myelogenous leukemia, which is caused by a chromosomal translocation event that produces an active Bcr/Abl fusion protein in hematopoietic cells.

RECEPTOR CLASSES & DRUG DEVELOPMENT

The existence of a specific drug receptor is usually inferred from studying the **structure-activity relationship** of a group of structurally similar congeners of the drug that mimic or antagonize its effects. Thus, if a series of related agonists exhibits identical relative potencies in producing two distinct effects, it is likely that the two effects are mediated by similar or identical receptor molecules. In addition, if identical receptors mediate both effects, a competitive antagonist will inhibit both responses with the same K_i; a second competitive antagonist will inhibit both responses with its own characteristic K_i. Thus, studies of the relation between structure and activity of a series of agonists and antagonists can identify a species of receptor that mediates a set of pharmacologic responses.

Exactly the same experimental procedure can show that observed effects of a drug are mediated by *different* receptors. In this case, effects mediated by different receptors may exhibit different orders of potency among agonists and different K_i values for each competitive antagonist.

Wherever we look, evolution has created many different receptors that function to mediate responses to any individual chemical signal. In some cases, the same chemical acts on completely different structural receptor classes. For example, acetylcholine uses ligand-gated ion channels (nicotinic AChRs) to initiate a fast (in milliseconds) excitatory postsynaptic potential (EPSP) in postganglionic neurons. Acetylcholine also activates a separate class of G

protein-coupled receptors (muscarinic AChRs), which mediate slower (seconds to minutes) modulatory effects on the same neurons. In addition, each structural class usually includes multiple subtypes of receptor, often with significantly different signaling or regulatory properties. For example, many biogenic amines (eg, norepinephrine, acetylcholine, and serotonin) activate more than one receptor, each of which may activate a different G protein, as previously described (see also Table 2–1). The existence of many receptor classes and subtypes for the same endogenous ligand has created important opportunities for drug development. For example, propranolol, a selective antagonist of β adrenoceptors, can reduce an accelerated heart rate without preventing the sympathetic nervous system from causing vasoconstriction, an effect mediated by α_1 receptors.

The principle of drug selectivity may even apply to structurally identical receptors expressed in different cells, eg, receptors for steroids such as estrogen (Figure 2–6). Different cell types express different accessory proteins, which interact with steroid receptors and change the functional effects of drug-receptor interaction. For example, tamoxifen acts as an *antagonist* on estrogen receptors expressed in mammary tissue but as an *agonist* on estrogen receptors in bone. Consequently, tamoxifen may be useful not only in the treatment and prophylaxis of breast cancer but also in the prevention of osteoporosis by increasing bone density (see Chapters 40 and 42). Tamoxifen may also create complications in postmenopausal women, however, by exerting an agonist action in the uterus, stimulating endometrial cell proliferation.

New drug development is not confined to agents that act on receptors for extracellular chemical signals. Increasingly, pharmaceutical chemists are determining whether elements of signaling pathways distal to the receptors may also serve as targets of selective and useful drugs. We have already discussed drugs that act on phosphodiesterase and some intracellular kinases. There are several additional kinase inhibitors presently in clinical trials, as well as preclinical efforts directed at developing inhibitors of G proteins.

RELATION BETWEEN DRUG DOSE & CLINICAL RESPONSE

We have dealt with receptors as molecules and shown how receptors can quantitatively account for the relation between dose or concentration of a drug and pharmacologic responses, at least in an idealized system. When faced with a patient who needs treatment, the prescriber must make a choice among a variety of possible drugs and devise a dosage regimen that is likely to produce maximal benefit and minimal toxicity. To make rational therapeutic decisions, the prescriber must understand how drug-receptor interactions underlie the relations between dose and response in patients, the nature and causes of variation in pharmacologic responsiveness, and the clinical implications of selectivity of drug action.

Dose & Response in Patients

A. Graded Dose-Response Relations

To choose among drugs and to determine appropriate doses of a drug, the prescriber must know the relative **pharmacologic potency** and **maximal efficacy** of the drugs in relation to the desired therapeutic effect. These two important terms, often confusing to students and clinicians, can be explained by referring to Figure 2–15, which depicts graded dose-response curves that relate the dose of four different drugs to the magnitude of a particular therapeutic effect.

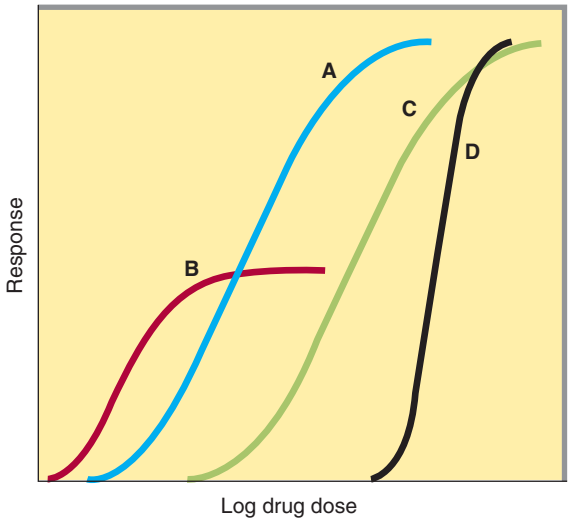


FIGURE 2–15 Graded dose-response curves for four drugs, illustrating different pharmacologic potencies and different maximal efficacies. (See text.)

1. Potency—Drugs A and B are said to be more potent than drugs C and D because of the relative positions of their dose-response curves along the **dose axis** of Figure 2–15. Potency refers to the concentration (EC_{50}) or dose (ED_{50}) of a drug required to produce 50% of that drug's maximal effect. Thus, the pharmacologic potency of drug A in Figure 2–15 is less than that of drug B, a partial agonist because the EC_{50} of A is greater than the EC_{50} of B. Potency of a drug depends in part on the affinity (K_d) of receptors for binding the drug and in part on the efficiency with which drug-receptor interaction is coupled to response. Note that some doses of drug A can produce larger effects than any dose of drug B, despite the fact that we describe drug B as pharmacologically more potent. The reason for this is that drug A has a larger maximal efficacy (as described below).

For clinical use, it is important to distinguish between a drug's potency and its efficacy. The clinical effectiveness of a drug depends not on its potency (EC_{50}), but on its maximal efficacy (see below) and its ability to reach the relevant receptors. This ability can depend on its route of administration, absorption, distribution through the body, and clearance from the blood or site of action. In deciding which of two drugs to administer to a patient, the prescriber must usually consider their relative effectiveness rather than their relative potency. Pharmacologic potency can largely determine the administered dose of the chosen drug.

For therapeutic purposes, the potency of a drug should be stated in dosage units, usually in terms of a particular therapeutic

end point (eg, 50 mg for mild sedation, 1 mcg/kg/min for an increase in heart rate of 25 bpm). Relative potency, the ratio of equi-effective doses (0.2, 10, etc), may be used in comparing one drug with another.

2. Maximal efficacy—This parameter reflects the limit of the dose-response relation on the **response axis.** Drugs A, C, and D in Figure 2–15 have equal maximal efficacy, and all have greater maximal efficacy than drug B. The maximal efficacy (sometimes referred to simply as efficacy) of a drug is obviously crucial for making clinical decisions when a large response is needed. It may be determined by the drug's mode of interactions with receptors (as with partial agonists[*] or by characteristics of the receptor-effector system involved.

Thus, diuretics that act on one portion of the nephron may produce much greater excretion of fluid and electrolytes than diuretics that act elsewhere. In addition, the *practical* efficacy of a drug for achieving a therapeutic end point (eg, increased cardiac contractility) may be limited by the drug's propensity to cause a toxic effect (eg, fatal cardiac arrhythmia) even if the drug could otherwise produce a greater therapeutic effect.

B. Shape of Dose-Response Curves

Although the responses depicted in curves A, B, and C of Figure 2–15 approximate the shape of a simple Michaelis-Menten relation (transformed to a logarithmic plot), some clinical responses do not. Extremely steep dose-response curves (eg, curve D) may have important clinical consequences if the upper portion of the curve represents an undesirable extent of response (eg, coma caused by a sedative-hypnotic). Steep dose-response curves in patients can result from cooperative interactions of several different actions of a drug (eg, effects on brain, heart, and peripheral vessels, all contributing to lowering of blood pressure).

C. Quantal Dose-Effect Curves

Graded dose-response curves of the sort described above have certain limitations in their application to clinical decision making. For example, such curves may be impossible to construct if the pharmacologic response is an either-or (quantal) event, such as prevention of convulsions, arrhythmia, or death. Furthermore, the clinical relevance of a quantitative dose-response relation in a single patient, no matter how precisely defined, may be limited in application to other patients, owing to the great potential variability among patients in severity of disease and responsiveness to drugs.

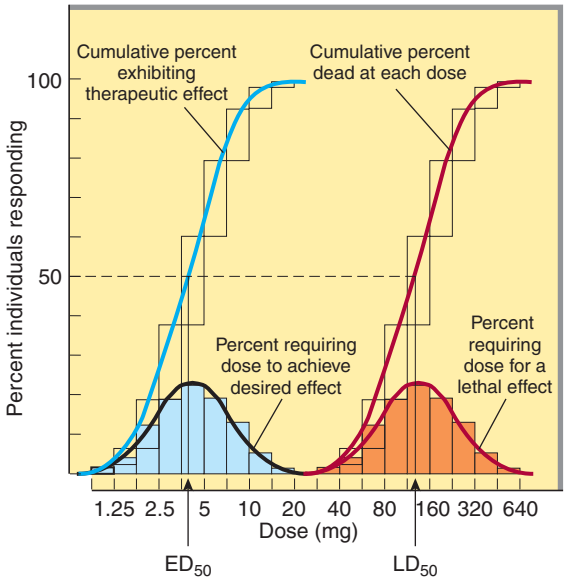


FIGURE 2–16 Quantal dose-effect plots. Shaded boxes (and the accompanying bell-shaped curves) indicate the frequency distribution of doses of drug required to produce a specified effect; that is, the percentage of animals that required a particular dose to exhibit the effect. The open boxes (and the corresponding colored curves) indicate the cumulative frequency distribution of responses, which are lognormally distributed.

Some of these difficulties may be avoided by determining the dose of drug required to produce a specified magnitude of effect in a large number of individual patients or experimental animals and plotting the cumulative frequency distribution of responders versus the log dose (Figure 2–16). The specified quantal effect may be chosen on the basis of clinical relevance (eg, relief of headache) or for preservation of safety of experimental subjects (eg, using low doses of a cardiac stimulant and specifying an increase in heart rate of 20 bpm as the quantal effect), or it may be an inherently quantal event (eg, death of an experimental animal). For most drugs, the doses required to produce a specified quantal effect in individuals are lognormally distributed; that is, a frequency distribution of such responses plotted against the log of the dose produces a gaussian normal curve of variation (colored areas, Figure 2–16). When these responses are summated, the resulting cumulative frequency distribution constitutes a quantal dose-effect curve (or dose-percent curve) of the proportion or percentage of individuals who exhibit the effect plotted as a function of log dose.

The quantal dose-effect curve is often characterized by stating the **median effective dose (ED_{50}),** which is the dose at which 50% of individuals exhibit the specified quantal effect. (Note that the abbreviation ED_{50} has a different meaning in this context from its meaning in relation to graded dose-effect curves, described in previous text). Similarly, the dose required to produce a particular toxic effect in 50% of animals is called the **median toxic dose (TD_{50}).** If the toxic effect is death of the animal, a **median lethal dose (LD_{50})** may be experimentally defined. Such values provide a convenient way of comparing the potencies of drugs in

[*]Note that "maximal efficacy," used in a therapeutic context, does not have exactly the same meaning that the term denotes in the more specialized context of drug-receptor interactions described earlier in this chapter. In an idealized in vitro system, efficacy denotes the relative maximal efficacy of agonists and partial agonists that act via the same receptor. In therapeutics, efficacy denotes the extent or degree of an effect that can be achieved in the intact patient. Thus, therapeutic efficacy may be affected by the characteristics of a particular drug-receptor interaction, but it also depends on a host of other factors as noted in the text.

experimental and clinical settings: Thus, if the ED_{50}s of two drugs for producing a specified quantal effect are 5 and 500 mg, respectively, then the first drug can be said to be 100 times more potent than the second for that particular effect. Similarly, one can obtain a valuable index of the selectivity of a drug's action by comparing its ED_{50}s for two different quantal effects in a population (eg, cough suppression versus sedation for opioid drugs).

Quantal dose-effect curves may also be used to generate information regarding the margin of safety to be expected from a particular drug used to produce a specified effect. One measure, which relates the dose of a drug required to produce a desired effect to that which produces an undesired effect, is the **therapeutic index.** In animal studies, the therapeutic index is usually defined as the ratio of the TD_{50} to the ED_{50} for some therapeutically relevant effect. The precision possible in animal experiments may make it useful to use such a therapeutic index to estimate the potential benefit of a drug in humans. Of course, the therapeutic index of a drug in humans is almost never known with real precision; instead, drug trials and accumulated clinical experience often reveal a range of usually effective doses and a different (but sometimes overlapping) range of possibly toxic doses. The clinically acceptable risk of toxicity depends critically on the severity of the disease being treated. For example, the dose range that provides relief from an ordinary headache in the majority of patients should be very much lower than the dose range that produces serious toxicity, even if the toxicity occurs in a small minority of patients. However, for treatment of a lethal disease such as Hodgkin's lymphoma, the acceptable difference between therapeutic and toxic doses may be smaller.

Finally, note that the quantal dose-effect curve and the graded dose-response curve summarize somewhat different sets of information, although both appear sigmoid in shape on a semilogarithmic plot (compare Figures 2–15 and 2–16). Critical information required for making rational therapeutic decisions can be obtained from each type of curve. Both curves provide information regarding the **potency** and **selectivity** of drugs; the graded dose-response curve indicates the **maximal efficacy** of a drug, and the quantal dose-effect curve indicates the potential **variability** of responsiveness among individuals.

Variation in Drug Responsiveness

Individuals may vary considerably in their response to a drug; indeed, a single individual may respond differently to the same drug at different times during the course of treatment. Occasionally, individuals exhibit an unusual or **idiosyncratic** drug response, one that is infrequently observed in most patients. The idiosyncratic responses are usually caused by genetic differences in metabolism of the drug or by immunologic mechanisms, including allergic reactions.

Quantitative variations in drug response are in general more common and more clinically important. An individual patient is **hyporeactive** or **hyperreactive** to a drug in that the intensity of effect of a given dose of drug is diminished or increased compared with the effect seen in most individuals. (**Note:** The term **hypersensitivity** usually refers to allergic or other immunologic responses to drugs.) With some drugs, the intensity of response to a given dose may change during the course of therapy; in these cases, responsiveness usually decreases as a consequence of continued drug administration, producing a state of relative **tolerance** to the drug's effects. When responsiveness diminishes rapidly after administration of a drug, the response is said to be subject to **tachyphylaxis.**

Even before administering the first dose of a drug, the prescriber should consider factors that may help in predicting the direction and extent of possible variations in responsiveness. These include the propensity of a particular drug to produce tolerance or tachyphylaxis as well as the effects of age, sex, body size, disease state, genetic factors, and simultaneous administration of other drugs.

Four general mechanisms may contribute to variation in drug responsiveness among patients or within an individual patient at different times.

A. Alteration in Concentration of Drug That Reaches the Receptor

Patients may differ in the rate of absorption of a drug, in distributing it through body compartments, or in clearing the drug from the blood (see Chapter 3). By altering the concentration of drug that reaches relevant receptors, such pharmacokinetic differences may alter the clinical response. Some differences can be predicted on the basis of age, weight, sex, disease state, and liver and kidney function, and by testing specifically for genetic differences that may result from inheritance of a functionally distinctive complement of drug-metabolizing enzymes (see Chapters 3 and 4). Another important mechanism influencing drug availability is active transport of drug from the cytoplasm, mediated by a family of membrane transporters encoded by the so-called multidrug resistance (*MDR*) genes. For example, up-regulation of *MDR* gene-encoded transporter expression is a major mechanism by which tumor cells develop resistance to anticancer drugs.

B. Variation in Concentration of an Endogenous Receptor Ligand

This mechanism contributes greatly to variability in responses to pharmacologic antagonists. Thus, propranolol, a β-adrenoceptor antagonist, markedly slows the heart rate of a patient whose endogenous catecholamines are elevated (as in pheochromocytoma) but does not affect the resting heart rate of a well-trained marathon runner. A partial agonist may exhibit even more dramatically different responses: Saralasin, a weak partial agonist at angiotensin II receptors, lowers blood pressure in patients with hypertension caused by increased angiotensin II production and raises blood pressure in patients who produce normal amounts of angiotensin.

C. Alterations in Number or Function of Receptors

Experimental studies have documented changes in drug response caused by increases or decreases in the number of receptor sites or by alterations in the efficiency of coupling of receptors to distal effector mechanisms. In some cases, the change in receptor number is caused by other hormones; for example, thyroid hormones increase both the number of β receptors in rat heart muscle and

cardiac sensitivity to catecholamines. Similar changes probably contribute to the tachycardia of thyrotoxicosis in patients and may account for the usefulness of propranolol, a β-adrenoceptor antagonist, in ameliorating symptoms of this disease.

In other cases, the agonist ligand itself induces a decrease in the number (eg, down-regulation) or coupling efficiency (eg, desensitization) of its receptors. These mechanisms (discussed previously under Signaling Mechanisms & Drug Actions) may contribute to two clinically important phenomena: first, tachyphylaxis or tolerance to the effects of some drugs (eg, biogenic amines and their congeners), and second, the "overshoot" phenomena that follow withdrawal of certain drugs. These phenomena can occur with either agonists or antagonists. An antagonist may increase the number of receptors in a critical cell or tissue by preventing down-regulation caused by an endogenous agonist. When the antagonist is withdrawn, the elevated number of receptors can produce an exaggerated response to physiologic concentrations of agonist. Potentially disastrous withdrawal symptoms can result for the opposite reason when administration of an agonist drug is discontinued. In this situation, the number of receptors, which has been decreased by drug-induced down-regulation, is too low for endogenous agonist to produce effective stimulation. For example, the withdrawal of clonidine (a drug whose α_2-adrenoceptor agonist activity reduces blood pressure) can produce hypertensive crisis, probably because the drug down-regulates α_2 adrenoceptors (see Chapter 11).

Genetic factors also can play an important role in altering the number or function of specific receptors. For example, a specific genetic variant of the α_{2C} adrenoceptor—when inherited together with a specific variant of the α_1 adrenoceptor—confers increased risk for developing heart failure, which may be reduced by early intervention using antagonist drugs. The identification of such genetic factors, part of the rapidly developing field of pharmacogenetics, holds promise for clinical diagnosis and in the future may help physicians design the most appropriate pharmacologic therapy for individual patients.

Another interesting example of genetic determination of effects on drug response is seen in the treatment of cancers involving excessive growth factor signaling. Somatic mutations affecting the tyrosine kinase domain of the epidermal growth factor receptor confer enhanced sensitivity to kinase inhibitors such as gefitinib in certain lung cancers. This effect enhances the antineoplastic effect of the drug and, because the somatic mutation is specific to the tumor and not present in the host, the therapeutic index of these drugs can be significantly enhanced in patients whose tumors harbor such mutations.

D. Changes in Components of Response Distal to the Receptor

Although a drug initiates its actions by binding to receptors, the response observed in a patient depends on the functional integrity of biochemical processes in the responding cell and physiologic regulation by interacting organ systems. Clinically, changes in these postreceptor processes represent the largest and most important class of mechanisms that cause variation in responsiveness to drug therapy.

Before initiating therapy with a drug, the prescriber should be aware of patient characteristics that may limit the clinical response. These characteristics include the age and general health of the patient and—most importantly—the severity and pathophysiologic mechanism of the disease. The most important potential cause of failure to achieve a satisfactory response is that the diagnosis is wrong or physiologically incomplete. Drug therapy is always most successful when it is accurately directed at the pathophysiologic mechanism responsible for the disease.

When the diagnosis is correct and the drug is appropriate, an unsatisfactory therapeutic response can often be traced to compensatory mechanisms in the patient that respond to and oppose the beneficial effects of the drug. Compensatory increases in sympathetic nervous tone and fluid retention by the kidney, for example, can contribute to tolerance to antihypertensive effects of a vasodilator drug. In such cases, additional drugs may be required to achieve a useful therapeutic result.

Clinical Selectivity: Beneficial versus Toxic Effects of Drugs

Although we classify drugs according to their principal actions, it is clear that *no drug causes only a single, specific effect.* Why is this so? It is exceedingly unlikely that any kind of drug molecule will bind to only a single type of receptor molecule, if only because the number of potential receptors in every patient is astronomically large. Even if the chemical structure of a drug allowed it to bind to only one kind of receptor, the biochemical processes controlled by such receptors would take place in many cell types and would be coupled to many other biochemical functions; as a result, the patient and the prescriber would probably perceive more than one drug effect. Accordingly, drugs are only *selective*—rather than specific—in their actions, because they bind to one or a few types of receptor more tightly than to others and because these receptors control discrete processes that result in distinct effects.

It is only because of their selectivity that drugs are useful in clinical medicine. Selectivity can be measured by comparing binding affinities of a drug to different receptors or by comparing ED_{50}s for different effects of a drug in vivo. In drug development and in clinical medicine, selectivity is usually considered by separating effects into two categories: **beneficial** or **therapeutic effects** versus **toxic** or **adverse effects.** Pharmaceutical advertisements and prescribers occasionally use the term **side effect,** implying that the effect in question is insignificant or occurs via a pathway that is to one side of the principal action of the drug; such implications are frequently erroneous.

A. Beneficial and Toxic Effects Mediated by the Same Receptor-Effector Mechanism

Much of the serious drug toxicity in clinical practice represents a direct pharmacologic extension of the therapeutic actions of the drug. In some of these cases (eg, bleeding caused by anticoagulant therapy; hypoglycemic coma due to insulin), toxicity may be avoided by judicious management of the dose of drug administered, guided by careful monitoring of effect (measurements of

blood coagulation or serum glucose) and aided by ancillary measures (avoiding tissue trauma that may lead to hemorrhage; regulation of carbohydrate intake). In still other cases, the toxicity may be avoided by not administering the drug at all, if the therapeutic indication is weak or if other therapy is available.

In certain situations, a drug is clearly necessary and beneficial but produces unacceptable toxicity when given in doses that produce optimal benefit. In such situations, it may be necessary to add another drug to the treatment regimen. In treating hypertension, for example, administration of a second drug often allows the prescriber to reduce the dose and toxicity of the first drug (see Chapter 11).

B. Beneficial and Toxic Effects Mediated by Identical Receptors but in Different Tissues or by Different Effector Pathways

Many drugs produce both their desired effects and adverse effects by acting on a single receptor type in different tissues. Examples discussed in this book include: digitalis glycosides, which act by inhibiting Na^+/K^+-ATPase in cell membranes; methotrexate, which inhibits the enzyme dihydrofolate reductase; and glucocorticoid hormones.

Three therapeutic strategies are used to avoid or mitigate this sort of toxicity. First, the drug should always be administered at the lowest dose that produces acceptable benefit. Second, adjunctive drugs that act through different receptor mechanisms and produce different toxicities may allow lowering the dose of the first drug, thus limiting its toxicity (eg, use of other immunosuppressive agents added to glucocorticoids in treating inflammatory disorders). Third, selectivity of the drug's actions may be increased by manipulating the concentrations of drug available to receptors in different parts of the body, for example, by aerosol administration of a glucocorticoid to the bronchi in asthma.

C. Beneficial and Toxic Effects Mediated by Different Types of Receptors

Therapeutic advantages resulting from new chemical entities with improved receptor selectivity were mentioned earlier in this chapter and are described in detail in later chapters. Such drugs include α- and β-selective adrenoceptor agonists and antagonists, H_1 and H_2 antihistamines, nicotinic and muscarinic blocking agents, and receptor-selective steroid hormones. All these receptors are grouped in functional families, each responsive to a small class of endogenous agonists. The receptors and their associated therapeutic uses were discovered by analyzing effects of the physiologic chemical signals—catecholamines, histamine, acetylcholine, and corticosteroids.

Several other drugs were discovered by exploiting therapeutic or toxic effects of chemically similar agents observed in a clinical context. Examples include quinidine, the sulfonylureas, thiazide diuretics, tricyclic antidepressants, opioid drugs, and phenothiazine antipsychotics. Often the new agents turn out to interact with receptors for endogenous substances (eg, opioids and phenothiazines for endogenous opioid and dopamine receptors, respectively). It is likely that other new drugs will be found to do so in the future, perhaps leading to the discovery of new classes of receptors and endogenous ligands for future drug development.

Thus, the propensity of drugs to bind to different classes of receptor sites is not only a potentially vexing problem in treating patients, it also presents a continuing challenge to pharmacology and an opportunity for developing new and more useful drugs.

REFERENCES

Berridge MJ: Unlocking the secrets of cell signaling. Ann Rev Physiol 2005;67:1.

Cabrera-Vera TM et al: Insights into G protein structure, function, and regulation. Endocr Rev 2003;24:765.

Civelli O et al: Orphan GPCRs and their ligands. Pharmacol Ther 2006;110:525.

Davies MA, Samuels Y: Analysis of the genome to personalize therapy for melanoma. Oncogene 2010;29:5545.

Feng XH, Derynck R: Specificity and versatility in TGF-beta signaling through Smads. Annu Rev Cell Dev Biol 2005;21:659.

Galandrin S, Oligny-Longpre G, Bouvier M: The evasive nature of drug efficacy: Implications for drug discovery. Trends Pharmacol Sci 2007;28:423.

Ginty DD, Segal RA: Retrograde neurotrophin signaling: Trk-ing along the axon. Curr Opin Neurobiol 2002;12:268.

Gouaux E, MacKinnon R: Principles of selective ion transport in channels and pumps. Science 2005;310:1461.

Hermiston ML et al: Reciprocal regulation of lymphocyte activation by tyrosine kinases and phosphatases. J Clin Invest 2002;109:9.

Kenakin T: Principles: Receptor theory in molecular pharmacology. Trends Pharmacol Sci 2004;25:186.

Mosesson Y, Yarden Y: Oncogenic growth factor receptors: Implications for signal transduction therapy. Semin Cancer Biol 2004;14:262.

Pawson T: Dynamic control of signaling by modular adaptor proteins. Curr Opin Cell Biol 2007;19:112.

Rajagopal S, Rajagopal K, Lefkowitz RJ: Teaching old receptors new tricks- biasing seven-transmembrane receptors. Nat Rev Drug Discov 2010;9:373.

Roden DM, George AL Jr: The genetic basis of variability in drug responses. Nat Rev Drug Discov 2002;1:37.

Rosenbaum DM, Rasmussen SG, Kobilka BK: The structure and function of G-protein-coupled receptors. Nature 2009;459:356.

Rotella DP: Phosphodiesterase 5 inhibitors: Current status and potential applications. Nat Rev Drug Discov 2002;1:674.

Small KM, McGraw DW, Liggett SB: Pharmacology and physiology of human adrenergic receptor polymorphisms. Ann Rev Pharmacol Toxicol 2003;43:381.

Sorkin A, von Zastrow M: Endocytosis and signaling: Intertwining molecular networks. Nat Rev Mol Cell Biol 2009;10:609.

Yu FH et al: Overview of molecular relationships in the voltage-gated ion channel superfamily. Pharmacol Rev 2005;57:387.

Yuan TL, Cantley LC: PI3K pathway alterations in cancer: Variations on a theme. Oncogene 2008;27:5497.

CASE STUDY ANSWER

Propranolol, a nonselective β-adrenoceptor blocker, is a useful antihypertensive agent because it reduces cardiac output and probably vascular resistance as well. However, it also prevents β_2-receptor-induced bronchodilation and may precipitate bronchoconstriction in susceptible individuals. Calcium channel blockers such as verapamil also reduce blood pressure but do not cause bronchoconstriction or prevent bronchodilation. Selection of the most appropriate drug or drug group for one condition requires awareness of the other conditions a patient may have and the receptor selectivity of the drug groups available.

This page intentionally left blank

CHAPTER

3

Pharmacokinetics & Pharmacodynamics: Rational Dosing & the Time Course of Drug Action

Nicholas H. G. Holford, MB, ChB, FRACP

CASE STUDY

An 85-year-old, 60-kg woman with a serum creatinine of 1.8 mg/dL has atrial fibrillation. A decision has been made to use digoxin to control the rapid heart rate. The target concentration of digoxin for the treatment of atrial fibrillation is 2 ng/mL. Tablets of digoxin contain 62.5 micrograms and 250 micrograms (mcg). What maintenance dose would you recommend?

The goal of therapeutics is to achieve a desired beneficial effect with minimal adverse effects. When a medicine has been selected for a patient, the clinician must determine the dose that most closely achieves this goal. A rational approach to this objective combines the principles of pharmacokinetics with pharmacodynamics to clarify the dose-effect relationship (Figure 3–1). Pharmacodynamics governs the concentration-effect part of the interaction, whereas pharmacokinetics deals with the dose-concentration part (Holford & Sheiner, 1981). The pharmacokinetic processes of absorption, distribution, and elimination determine how rapidly and for how long the drug will appear at the target organ. The pharmacodynamic concepts of maximum response and sensitivity determine the magnitude of the effect at a particular concentration (see E_{max} and C_{50}, Chapter 2; C_{50} is also known as EC_{50}).

Figure 3–1 illustrates a fundamental hypothesis of pharmacology, namely, that a relationship exists between a beneficial or toxic effect of a drug and the concentration of the drug. This hypothesis has been documented for many drugs, as indicated by the Target Concentrations and Toxic Concentrations columns in Table 3–1. The apparent lack of such a relationship for some drugs does not weaken the basic hypothesis but points to the need to consider the time course of concentration at the actual site of pharmacologic effect (see below).

Knowing the relationship between dose, drug concentration, and effects allows the clinician to take into account the various pathologic and physiologic features of a particular patient that make him or her different from the average individual in responding to a drug. The importance of pharmacokinetics and pharmacodynamics in patient care thus rests upon the improvement in therapeutic benefit and reduction in toxicity that can be achieved by application of these principles.

PHARMACOKINETICS

The "standard" dose of a drug is based on trials in healthy volunteers and patients with average ability to absorb, distribute, and eliminate

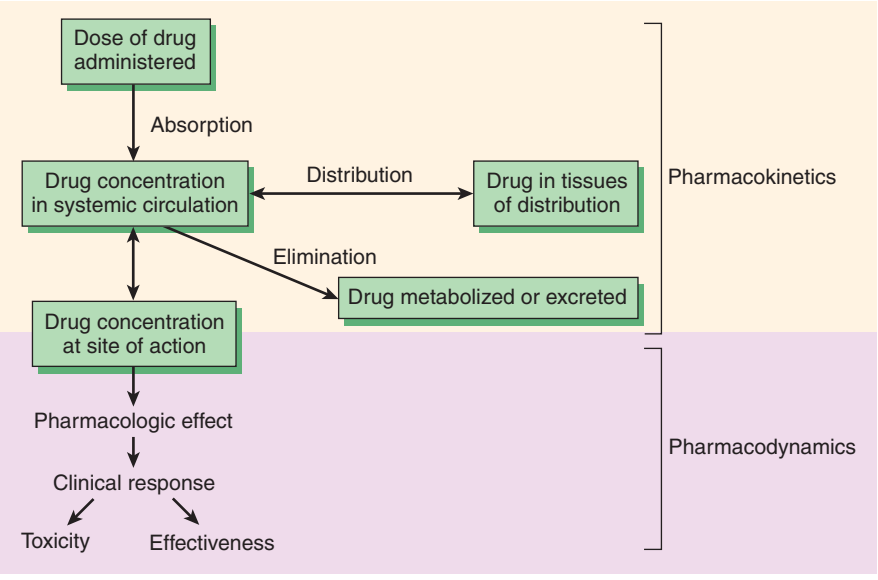


FIGURE 3–1 The relationship between dose and effect can be separated into pharmacokinetic (dose-concentration) and pharmacodynamic (concentration-effect) components. Concentration provides the link between pharmacokinetics and pharmacodynamics and is the focus of the target concentration approach to rational dosing. The three primary processes of pharmacokinetics are absorption, distribution, and elimination.

the drug (see Clinical Trials: The IND and NDA in Chapter 5). This dose will not be suitable for every patient. Several physiologic processes (eg, maturation of organ function in infants) and pathologic processes (eg, heart failure, renal failure) dictate dosage adjustment in individual patients. These processes modify specific pharmacokinetic parameters. The two basic parameters are **clearance,** the measure of the ability of the body to eliminate the drug; and **volume of distribution,** the measure of the apparent space in the body available to contain the drug. These parameters are illustrated schematically in Figure 3–2 where the volume of the beakers into which the drugs diffuse represents the volume of distribution and the size of the outflow "drain" in Figures 3–2B and 3–2D represents the clearance.

Volume of Distribution

Volume of distribution (V) relates the amount of drug in the body to the concentration of drug (C) in blood or plasma:

$$V = \frac{\text{Amount of drug in body}}{C} \qquad (1)$$

The volume of distribution may be defined with respect to blood, plasma, or water (unbound drug), depending on the concentration used in equation (1) ($C = C_b$, C_p, or C_u).

That the V calculated from equation (1) is an *apparent* volume may be appreciated by comparing the volumes of distribution of drugs such as digoxin or chloroquine (Table 3–1) with some of the physical volumes of the body (Table 3–2). Volume of distribution can vastly exceed any physical volume in the body because it is the volume *apparently* necessary to contain the amount of drug *homogeneously* at the concentration found in the blood, plasma, or water. Drugs with very high volumes of distribution have much higher concentrations in extravascular tissue than in the vascular compartment, ie, they are *not* homogeneously distributed. Drugs that are completely retained within the vascular compartment, on the other hand, have a minimum possible volume of distribution equal to the blood component in which they are distributed, eg, 0.04 L/kg body weight or 2.8 L/70 kg (Table 3–2) for a drug that is restricted to the plasma compartment.

Clearance

Drug clearance principles are similar to the clearance concepts of renal physiology. Clearance of a drug is the factor that predicts the rate of elimination in relation to the drug concentration:

$$CL = \frac{\text{Rate of elimination}}{C} \qquad (2)$$

Clearance, like volume of distribution, may be defined with respect to blood (CL_b), plasma (CL_p), or unbound in water (CL_u), depending on the concentration measured.

It is important to note the additive character of clearance. Elimination of drug from the body may involve processes occurring in the kidney, the lung, the liver, and other organs. Dividing the rate of elimination at each organ by the concentration of drug presented to it yields the respective clearance at that organ.

TABLE 3–1 Pharmacokinetic and pharmacodynamic parameters for selected drugs. (See Speight & Holford, 1997, for a more comprehensive listing.)

Drug	Oral Availability (F) (%)	Urinary Excretion (%)[1]	Bound in Plasma (%)	Clearance (L/h/70 kg)[2]	Volume of Distribution (L/70 kg)	Half-Life (h)	Target Concentration	Toxic Concentration
Acetaminophen	88	3	0	21	67	2	15 mg/L	> 300 mg/L
Acyclovir	23	75	15	19.8	48	2.4	…	…
Amikacin	…	98	4	5.46	19	2.3	10 mg/L[3] …	…
Amoxicillin	93	86	18	10.8	15	1.7	…	…
Amphotericin	…	4	90	1.92	53	18	…	…
Ampicillin	62	82	18	16.2	20	1.3	…	…
Aspirin	68	1	49	39	11	0.25	…	…
Atenolol	56	94	5	10.2	67	6.1	1 mg/L	…
Atropine	50	57	18	24.6	120	4.3	…	…
Captopril	65	38	30	50.4	57	2.2	50 ng/mL	…
Carbamazepine	70	1	74	5.34	98	15	6 mg/L	> 9 mg/L
Cephalexin	90	91	14	18	18	0.9	…	…
Cephalothin	…	52	71	28.2	18	0.57	…	…
Chloramphenicol	80	25	53	10.2	66	2.7	…	…
Chlordiazepoxide	100	1	97	2.28	21	10	1 mg/L	…
Chloroquine	89	61	61	45	13,000	214	20 ng/mL	250 ng/mL
Chlorpropamide	90	20	96	0.126	6.8	33	…	…
Cimetidine	62	62	19	32.4	70	1.9	0.8 mg/L	…
Ciprofloxacin	60	65	40	25.2	130	4.1	…	…
Clonidine	95	62	20	12.6	150	12	1 ng/mL	…
Cyclosporine	30	1	98	23.9	244	15	200 ng/mL	> 400 ng/mL
Diazepam	100	1	99	1.62	77	43	300 ng/mL	…
Digoxin	70	67	25	9	500	39	1 ng/mL	> 2 ng/mL
Diltiazem	44	4	78	50.4	220	3.7	…	…
Disopyramide	83	55	2	5.04	41	6	3 mg/mL	> 8 mg/mL
Enalapril	95	90	55	9	40	3	> 0.5 ng/mL	…
Erythromycin	35	12	84	38.4	55	1.6	…	…
Ethambutol	77	79	5	36	110	3.1	…	> 10 mg/L
Fluoxetine	60	3	94	40.2	2500	53	…	…
Furosemide	61	66	99	8.4	7.7	1.5	…	> 25 mg/L
Gentamicin	…	76	10	4.7	20	3	3 mg/L[3]	…
Hydralazine	40	10	87	234	105	1	100 ng/mL	…
Imipramine	40	2	90	63	1600	18	200 ng/mL	> 1 mg/L
Indomethacin	98	15	90	8.4	18	2.4	1 mg/L	> 5 mg/L
Labetalol	18	5	50	105	660	4.9	0.1 mg/L	…
Lidocaine	35	2	70	38.4	77	1.8	3 mg/L	> 6 mg/L
Lithium	100	95	0	1.5	55	22	0.7 mEq/L	> 2 mEq/L

(*continued*)

TABLE 3–1 Pharmacokinetic and pharmacodynamic parameters for selected drugs. (Continued)

Drug	Oral Availability (F) (%)	Urinary Excretion (%)[1]	Bound in Plasma (%)	Clearance (L/h/70 kg)[2]	Volume of Distribution (L/70 kg)	Half-Life (h)	Target Concentration	Toxic Concentration
Meperidine	52	12	58	72	310	3.2	0.5 mg/L	…
Methotrexate	70	48	34	9	39	7.2	750 μM-h[4,5]	> 950 μM-h
Metoprolol	38	10	11	63	290	3.2	25 ng/mL	…
Metronidazole	99	10	10	5.4	52	8.5	4 mg/L	…
Midazolam	44	56	95	27.6	77	1.9	…	…
Morphine	24	8	35	60	230	1.9	15 ng/mL	…
Nifedipine	50	0	96	29.4	55	1.8	50 ng/mL	…
Nortriptyline	51	2	92	30	1300	31	100 ng/mL	> 500 ng/mL
Phenobarbital	100	24	51	0.258	38	98	15 mg/L	> 30 mg/L
Phenytoin	90	2	89	Conc dependent[5]	45	Conc dependent[6]	10 mg/L	> 20 mg/L
Prazosin	68	1	95	12.6	42	2.9	…	…
Procainamide	83	67	16	36	130	3	5 mg/L	> 14 mg/L
Propranolol	26	1	87	50.4	270	3.9	20 ng/mL	…
Pyridostigmine	14	85	…	36	77	1.9	75 ng/mL	…
Quinidine	80	18	87	19.8	190	6.2	3 mg/L	> 8 mg/L
Ranitidine	52	69	15	43.8	91	2.1	100 ng/mL	…
Rifampin	?	7	89	14.4	68	3.5	…	…
Salicylic acid	100	15	85	0.84	12	13	200 mg/L	> 200 mg/L
Sulfamethoxazole	100	14	62	1.32	15	10	…	…
Terbutaline	14	56	20	14.4	125	14	2 ng/mL	…
Tetracycline	77	58	65	7.2	105	11	…	…
Theophylline	96	18	56	2.8	35	8.1	10 mg/L	> 20 mg/L
Tobramycin	…	90	10	4.62	18	2.2	…	…
Tocainide	89	38	10	10.8	210	14	10 mg/L	…
Tolbutamide	93	0	96	1.02	7	5.9	100 mg/L	…
Trimethoprim	100	69	44	9	130	11	…	…
Tubocurarine	…	63	50	8.1	27	2	0.6 mg/L	…
Valproic acid	100	2	93	0.462	9.1	14	75 mg/L	> 150 mg/L
Vancomycin	…	79	30	5.88	27	5.6	20 mg/L[3]	…
Verapamil	22	3	90	63	350	4	…	…
Warfarin	93	3	99	0.192	9.8	37	…	…
Zidovudine	63	18	25	61.8	98	1.1	…	…

[1] Assuming creatinine clearance 100 mL/min/70 kg.

[2] Convert to mL/min by multiplying the number given by 16.6.

[3] Average steady-state concentration.

[4] Target area under the concentration-time curve after a single dose.

[5] Can be estimated from measured C using $CL = V_{max}/(K_m + C)$; V_{max} = 415 mg/d, K_m = 5 mg/L. See text.

[6] Varies because of concentration-dependent clearance.

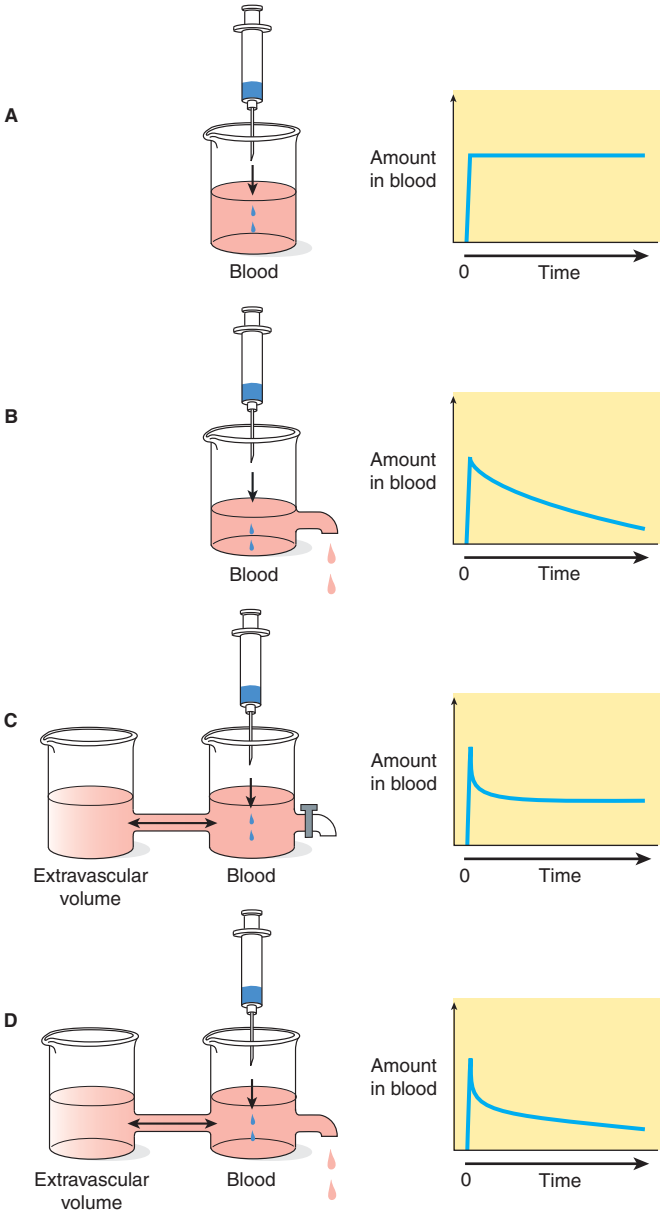


FIGURE 3–2 Models of drug distribution and elimination. The effect of adding drug to the blood by rapid intravenous injection is represented by expelling a known amount of the agent into a beaker. The time course of the amount of drug in the beaker is shown in the graphs at the right. In the first example (**A**), there is no movement of drug out of the beaker, so the graph shows only a steep rise to maximum followed by a plateau. In the second example (**B**), a route of elimination is present, and the graph shows a slow decay after a sharp rise to a maximum. Because the level of material in the beaker falls, the "pressure" driving the elimination process also falls, and the slope of the curve decreases. This is an exponential decay curve. In the third model (**C**), drug placed in the first compartment ("blood") equilibrates rapidly with the second compartment ("extravascular volume") and the amount of drug in "blood" declines exponentially to a new steady state. The fourth model (**D**) illustrates a more realistic combination of elimination mechanism and extravascular equilibration. The resulting graph shows an early distribution phase followed by the slower elimination phase.

TABLE 3–2 Physical volumes (in L/kg body weight) of some body compartments into which drugs may be distributed.

Compartment and Volume	Examples of Drugs
Water	
Total body water (0.6 L/kg[1])	Small water-soluble molecules: eg, ethanol
Extracellular water (0.2 L/kg)	Larger water-soluble molecules: eg, gentamicin
Blood (0.08 L/kg); plasma (0.04 L/kg)	Strongly plasma protein-bound molecules and very large molecules: eg, heparin
Fat (0.2-0.35 L/kg)	Highly lipid-soluble molecules: eg, DDT
Bone (0.07 L/kg)	Certain ions: eg, lead, fluoride

[1]An average figure. Total body water in a young lean man might be 0.7 L/kg; in an obese woman, 0.5 L/kg.

Added together, these separate clearances equal total systemic clearance:

$$CL_{kidney} = \frac{\text{Rate of elimination}_{kidney}}{C} \quad (3a)$$

$$CL_{liver} = \frac{\text{Rate of elimination}_{liver}}{C} \quad (3b)$$

$$CL_{other} = \frac{\text{Rate of elimination}_{other}}{C} \quad (3c)$$

$$CL_{systemic} = CL_{kidney} + CL_{liver} + CL_{other} \quad (3d)$$

"Other" tissues of elimination could include the lungs and additional sites of metabolism, eg, blood or muscle.

The two major sites of drug elimination are the kidneys and the liver. Clearance of unchanged drug in the urine represents renal clearance. Within the liver, drug elimination occurs via biotransformation of parent drug to one or more metabolites, or excretion of unchanged drug into the bile, or both. The pathways of biotransformation are discussed in Chapter 4. For most drugs, clearance is constant over the concentration range encountered in clinical settings, ie, elimination is not saturable, and the rate of drug elimination is directly proportional to concentration (rearranging equation [2]):

$$\text{Rate of elimination} = CL \times C \quad (4)$$

This is usually referred to as first-order elimination. When clearance is first-order, it can be estimated by calculating the **area under the curve (AUC)** of the time-concentration profile after a dose. Clearance is calculated from the dose divided by the AUC.

A. Capacity-Limited Elimination

For drugs that exhibit capacity-limited elimination (eg, phenytoin, ethanol), clearance will vary depending on the concentration

of drug that is achieved (Table 3–1). Capacity-limited elimination is also known as mixed-order, saturable, dose- or concentration-dependent, nonlinear, and Michaelis-Menten elimination.

Most drug elimination pathways will become saturated if the dose and therefore the concentration are high enough. When blood flow to an organ does not limit elimination (see below), the relation between elimination rate and concentration (C) is expressed mathematically in equation (5):

$$\text{Rate of elimination} = \frac{V_{max} \times C}{K_m + C} \qquad (5)$$

The maximum elimination capacity is V_{max}, and K_m is the drug concentration at which the rate of elimination is 50% of V_{max}. At concentrations that are high relative to the K_m, the elimination rate is almost independent of concentration—a state of "pseudo-zero order" elimination. If dosing rate exceeds elimination capacity, steady state cannot be achieved: The concentration will keep on rising as long as dosing continues. This pattern of capacity-limited elimination is important for three drugs in common use: ethanol, phenytoin, and aspirin. Clearance has no real meaning for drugs with capacity-limited elimination, and AUC should not be used to describe the elimination of such drugs.

B. Flow-Dependent Elimination

In contrast to capacity-limited drug elimination, some drugs are cleared very readily by the organ of elimination, so that at any clinically realistic concentration of the drug, most of the drug in the blood perfusing the organ is eliminated on the first pass of the drug through it. The elimination of these drugs will thus depend primarily on the rate of drug delivery to the organ of elimination. Such drugs (see Table 4–7) can be called "high-extraction" drugs since they are almost completely extracted from the blood by the organ. Blood flow to the organ is the main determinant of drug delivery, but plasma protein binding and blood cell partitioning may also be important for extensively bound drugs that are highly extracted.

Half-Life

Half-life ($t_{1/2}$) is the time required to change the amount of drug in the body by one-half during elimination (or during a constant infusion). In the simplest case—and the most useful in designing drug dosage regimens—the body may be considered as a single compartment (as illustrated in Figure 3–2B) of a size equal to the volume of distribution (V). The time course of drug in the body will depend on both the volume of distribution and the clearance:

$$t_{1/2} = \frac{0.7 \times V}{CL} \qquad (6)$$

Because drug elimination can be described by an exponential process, the time taken for a twofold decrease can be shown to be proportional to the natural logarithm of 2. The constant 0.7 in equation (6) is an approximation to the natural logarithm of 2.

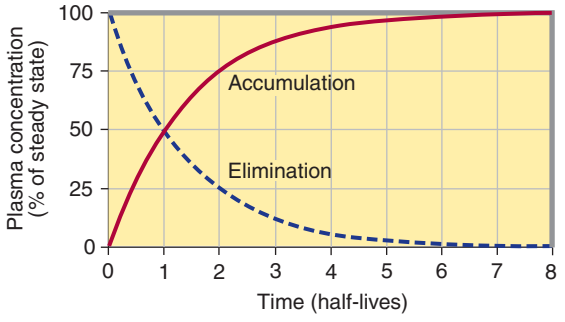


FIGURE 3–3 The time course of drug accumulation and elimination. **Solid line:** Plasma concentrations reflecting drug accumulation during a constant-rate infusion of a drug. Fifty percent of the steady-state concentration is reached after one half-life, 75% after two half-lives, and over 90% after four half-lives. **Dashed line:** Plasma concentrations reflecting drug elimination after a constant-rate infusion of a drug had reached steady state. Fifty percent of the drug is lost after one half-life, 75% after two half-lives, etc. The "rule of thumb" that four half-lives must elapse after starting a drug-dosing regimen before full effects will be seen is based on the approach of the accumulation curve to over 90% of the final steady-state concentration.

Half-life is useful because it indicates the time required to attain 50% of steady state—or to decay 50% from steady-state conditions—after a change in the rate of drug administration. Figure 3–3 shows the time course of drug accumulation during a constant-rate drug infusion and the time course of drug elimination after stopping an infusion that has reached steady state.

Disease states can affect both of the physiologically related primary pharmacokinetic parameters: volume of distribution and clearance. A change in half-life will not necessarily reflect a change in drug elimination. For example, patients with chronic renal failure have decreased renal clearance of digoxin but also a decreased volume of distribution; the increase in digoxin half-life is not as great as might be expected based on the change in renal function. The decrease in volume of distribution is due to the decreased renal and skeletal muscle mass and consequent decreased tissue binding of digoxin to Na^+/K^+-ATPase.

Many drugs will exhibit multicompartment pharmacokinetics (as illustrated in Figures 3–2C and 3–2D). Under these conditions, the "true" terminal half-life, as given in Table 3–1, will be greater than that calculated from equation (6).

Drug Accumulation

Whenever drug doses are repeated, the drug will accumulate in the body until dosing stops. This is because it takes an infinite time (in theory) to eliminate all of a given dose. In practical terms, this means that if the dosing interval is shorter than four half-lives, accumulation will be detectable.

Accumulation is inversely proportional to the fraction of the dose lost in each dosing interval. The fraction lost is 1 minus the fraction remaining just before the next dose. The fraction remaining can be predicted from the dosing interval and the

half-life. A convenient index of accumulation is the **accumulation factor:**

$$\text{Accumulation factor} = \frac{1}{\text{Fraction lost in one dosing interval}}$$

$$= \frac{1}{1 - \text{Fraction remaining}} \qquad (7)$$

For a drug given once every half-life, the accumulation factor is 1/0.5, or 2. The accumulation factor predicts the ratio of the steady-state concentration to that seen at the same time following the first dose. Thus, the peak concentrations after intermittent doses at steady state will be equal to the peak concentration after the first dose multiplied by the accumulation factor.

Bioavailability

Bioavailability is defined as the fraction of unchanged drug reaching the systemic circulation following administration by any route (Table 3–3). The area under the blood concentration-time curve (AUC) is proportional to the extent of bioavailability for a drug if its elimination is first-order (Figure 3–4). For an intravenous dose of the drug, bioavailability is assumed to be equal to unity. For a drug administered orally, bioavailability may be less than 100% for two main reasons—incomplete extent of absorption across the gut wall and first-pass elimination by the liver (see below).

A. Extent of Absorption

After oral administration, a drug may be incompletely absorbed, eg, only 70% of a dose of digoxin reaches the systemic circulation. This is mainly due to lack of absorption from the gut. Other drugs are either too hydrophilic (eg, atenolol) or too lipophilic (eg, acyclovir) to be absorbed easily, and their low bioavailability is also due to incomplete absorption. If too hydrophilic, the drug cannot cross the lipid cell membrane; if too lipophilic, the drug is not soluble enough to cross the water layer adjacent to the cell. Drugs may not be absorbed because of a reverse transporter associated with P-glycoprotein. This process actively pumps drug out of gut wall cells back into the gut lumen. Inhibition of P-glycoprotein and gut wall metabolism, eg, by grapefruit juice, may be associated with substantially increased drug absorption.

TABLE 3–3 Routes of administration, bioavailability, and general characteristics.

Route	Bioavailability (%)	Characteristics
Intravenous (IV)	100 (by definition)	Most rapid onset
Intramuscular (IM)	75 to ≤ 100	Large volumes often feasible; may be painful
Subcutaneous (SC)	75 to ≤100	Smaller volumes than IM; may be painful
Oral (PO)	5 to < 100	Most convenient; first-pass effect may be significant
Rectal (PR)	30 to < 100	Less first-pass effect than oral
Inhalation	5 to < 100	Often very rapid onset
Transdermal	80 to ≤ 100	Usually very slow absorption; used for lack of first-pass effect; prolonged duration of action

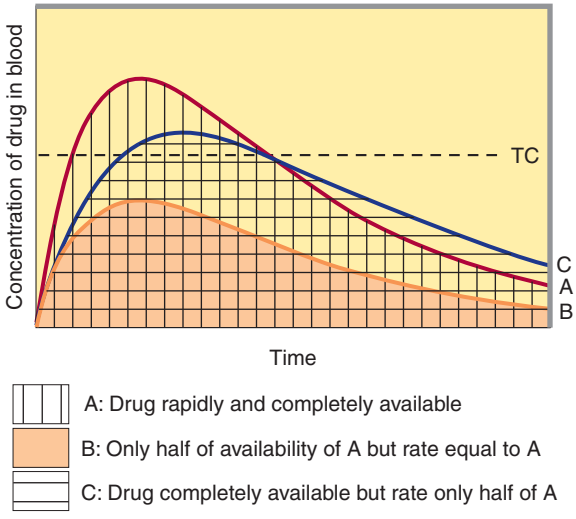


FIGURE 3-4 Blood concentration-time curves, illustrating how changes in the rate of absorption and extent of bioavailability can influence both the duration of action and the effectiveness of the same total dose of a drug administered in three different formulations. The dashed line indicates the target concentration (TC) of the drug in the blood.

B. First-Pass Elimination

Following absorption across the gut wall, the portal blood delivers the drug to the liver prior to entry into the systemic circulation. A drug can be metabolized in the gut wall (eg, by the CYP3A4 enzyme system) or even in the portal blood, but most commonly it is the liver that is responsible for metabolism before the drug reaches the systemic circulation. In addition, the liver can excrete the drug into the bile. Any of these sites can contribute to this reduction in bioavailability, and the overall process is known as first-pass elimination. The effect of first-pass hepatic elimination on bioavailability is expressed as the extraction ratio (ER):

$$ER = \frac{CL_{liver}}{Q} \qquad (8a)$$

where Q is hepatic blood flow, normally about 90 L/h in a person weighing 70 kg.

The systemic bioavailability of the drug (F) can be predicted from the extent of absorption (f) and the extraction ratio (ER):

$$F = f \times (1 - ER) \qquad (8b)$$

A drug such as morphine is almost completely absorbed (f = 1), so that loss in the gut is negligible. However, the hepatic extraction ratio for morphine is morphine clearance (60 L/h/70 kg) divided by hepatic blood flow (90 L/h/70 kg) or 0.67, so (1 – ER) is 0.33. The bioavailability of morphine is therefore expected to be about 33%, which is close to the observed value (Table 3–1).

C. Rate of Absorption

The distinction between rate and extent of absorption is shown in Figure 3–4. The rate of absorption is determined by the site of administration and the drug formulation. Both the rate of absorption and the extent of input can influence the clinical effectiveness of a drug. For the three different dosage forms depicted in Figure 3–4, there would be significant differences in the intensity of clinical effect. Dosage form B would require twice the dose to attain blood concentrations equivalent to those of dosage form A. Differences in rate of absorption may become important for drugs given as a single dose, such as a hypnotic used to induce sleep. In this case, drug from dosage form A would reach its target concentration earlier than drug from dosage form C; concentrations from A would also reach a higher level and remain above the target concentration for a longer period. In a multiple dosing regimen, dosage forms A and C would yield the same average blood level concentrations, although dosage form A would show somewhat greater maximum and lower minimum concentrations.

The mechanism of drug absorption is said to be zero-order when the rate is independent of the amount of drug remaining in the gut, eg, when it is determined by the rate of gastric emptying or by a controlled-release drug formulation. In contrast, when the dose is dissolved in gastrointestinal fluids, the rate of absorption is usually proportional to the gastrointestinal concentration and is said to be first-order.

Extraction Ratio & the First-Pass Effect

Systemic clearance is not affected by bioavailability. However, clearance can markedly affect the extent of availability because it determines the extraction ratio (equation [8a]). Of course, therapeutic blood concentrations may still be reached by the oral route of administration if larger doses are given. However, in this case, the concentrations of the drug *metabolites* will be increased significantly over those that would occur following intravenous administration. Lidocaine and verapamil are both used to treat cardiac arrhythmias and have bioavailability less than 40%, but lidocaine is never given orally because its metabolites are believed to contribute to central nervous system toxicity. Other drugs that are highly extracted by the liver include isoniazid, morphine, propranolol, and several tricyclic antidepressants (Table 3–1).

Drugs with high extraction ratios will show marked variations in bioavailability between subjects because of differences in hepatic function and blood flow. These differences can explain the marked variation in drug concentrations that occurs among individuals given similar doses of highly extracted drugs. For drugs that are highly extracted by the liver, bypassing hepatic sites of elimination (eg, in hepatic cirrhosis with portosystemic shunting) will result in substantial increases in drug availability, whereas for drugs that are poorly extracted by the liver (for which the difference between entering and exiting drug concentration is small), shunting of blood past the liver will cause little change in availability. Drugs in Table 3–1 that are poorly extracted by the liver include chlorpropamide, diazepam, phenytoin, theophylline, tolbutamide, and warfarin.

Alternative Routes of Administration & the First-Pass Effect

There are several reasons for different routes of administration used in clinical medicine (Table 3–3)—for convenience (eg, oral), to maximize concentration at the site of action and minimize it elsewhere (eg, topical), to prolong the duration of drug absorption (eg, transdermal), or to avoid the first-pass effect.

The hepatic first-pass effect can be avoided to a great extent by use of sublingual tablets and transdermal preparations and to a lesser extent by use of rectal suppositories. Sublingual absorption provides direct access to systemic—not portal—veins. The transdermal route offers the same advantage. Drugs absorbed from suppositories in the lower rectum enter vessels that drain into the inferior vena cava, thus bypassing the liver. However, suppositories tend to move upward in the rectum into a region where veins that lead to the liver predominate. Thus, only about 50% of a rectal dose can be assumed to bypass the liver.

Although drugs administered by inhalation bypass the hepatic first-pass effect, the lung may also serve as a site of first-pass loss by excretion and possibly metabolism for drugs administered by nongastrointestinal ("parenteral") routes.

THE TIME COURSE OF DRUG EFFECT

The principles of pharmacokinetics (discussed in this chapter) and those of pharmacodynamics (discussed in Chapter 2 and Holford & Sheiner, 1981) provide a framework for understanding the time course of drug effect.

Immediate Effects

In the simplest case, drug effects are directly related to plasma concentrations, but this does not necessarily mean that effects simply parallel the time course of concentrations. Because the relationship between drug concentration and effect is not linear (recall the E_{max} model described in Chapter 2), the effect will not usually be linearly proportional to the concentration.

Consider the effect of an angiotensin-converting enzyme (ACE) inhibitor, such as enalapril, on plasma ACE. The half-life of enalapril is about 3 hours. After an oral dose of 10 mg, the peak plasma concentration at 3 hours is about 64 ng/mL. Enalapril is usually given once a day, so seven half-lives will elapse from the time of peak concentration to the end of the dosing interval. The concentration of enalapril after each half-life and the corresponding extent of ACE inhibition are shown in Figure 3–5. The extent of inhibition of ACE is calculated using the E_{max} model, where E_{max}, the maximum extent of inhibition, is 100% and the C_{50}, the

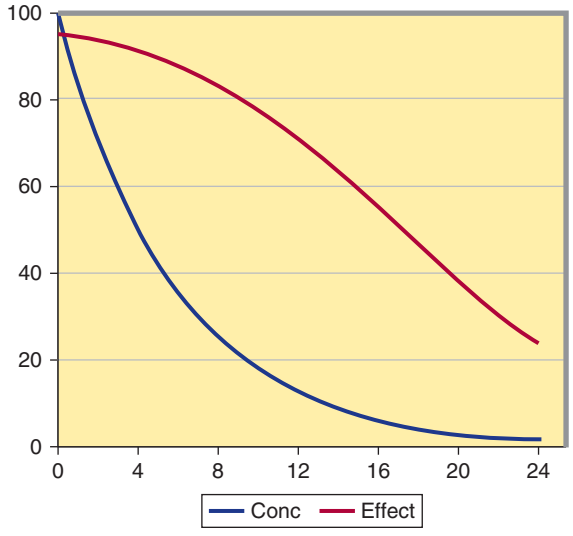


FIGURE 3–5 Time course of angiotensin-converting enzyme (ACE) inhibitor concentrations and effects. The blue line shows the plasma enalapril concentrations in nanograms per milliliter after a single oral dose. The red line indicates the percentage inhibition of its target, ACE. Note the different shapes of the concentration-time course (exponentially decreasing) and the effect-time course (linearly decreasing in its central portion).

concentration of the drug that produces 50% of maximum effect, is about 1 ng/mL.

Note that plasma concentrations of enalapril change by a factor of 16 over the first 12 hours (four half-lives) after the peak, but ACE inhibition has only decreased by 20%. Because the concentrations over this time are so high in relation to the C_{50}, the effect on ACE is almost constant. After 24 hours, ACE is still 33% inhibited. This explains why a drug with a short half-life can be given once a day and still maintain its effect throughout the day. The key factor is a high initial concentration in relation to the C_{50}. Even though the plasma concentration at 24 hours is less than 1% of its peak, this low concentration is still half the C_{50}. This is very common for drugs that act on enzymes (eg, ACE inhibitors) or compete at receptors (eg, propranolol).

When concentrations are in the range between four times and one fourth of the C_{50}, the time course of effect is essentially a linear function of time. It takes four half-lives for concentrations to drop from an effect of 80% to 20% of E_{max}—15% of the effect is lost every half-life over this concentration range. At concentrations below one fourth the C_{50}, the effect becomes almost directly proportional to concentration and the time course of drug effect will follow the exponential decline of concentration. It is only when the concentration is low in relation to the C_{50} that the concept of a "half-life of drug effect" has any meaning.

Delayed Effects

Changes in drug effects are often delayed in relation to changes in plasma concentration. This delay may reflect the time required for the drug to distribute from plasma to the site of action. This will be the case for almost all drugs. The delay due to distribution is a pharmacokinetic phenomenon that can account for delays of a few minutes. This distributional delay can account for the lag of effects after rapid intravenous injection of central nervous system (CNS)–active agents such as thiopental.

A common reason for more delayed drug effects—especially those that take many hours or even days to occur—is the slow turnover of a physiologic substance that is involved in the expression of the drug effect. For example, warfarin works as an anticoagulant by inhibiting vitamin K epoxidase in the liver. This action of warfarin occurs rapidly, and inhibition of the enzyme is closely related to plasma concentrations of warfarin. The *clinical effect* of warfarin, eg, on the International Normalized Ratio (INR), reflects a decrease in the concentration of the prothrombin complex of clotting factors. Inhibition of vitamin K epoxidase decreases the synthesis of these clotting factors, but the complex has a long half-life (about 14 hours), and it is this half-life that determines how long it takes for the concentration of clotting factors to reach a new steady state and for a drug effect to reflect the average warfarin plasma concentration.

Cumulative Effects

Some drug effects are more obviously related to a cumulative action than to a rapidly reversible one. The renal toxicity of aminoglycoside antibiotics (eg, gentamicin) is greater when administered as a constant infusion than with intermittent dosing. It is the accumulation of aminoglycoside in the renal cortex that is thought to cause renal damage. Even though both dosing schemes produce the same average steady-state concentration, the intermittent dosing scheme produces much higher peak concentrations, which saturate an uptake mechanism into the cortex; thus, total aminoglycoside accumulation is less. The difference in toxicity is a predictable consequence of the different patterns of concentration and the saturable uptake mechanism.

The effect of many drugs used to treat cancer also reflects a cumulative action—eg, the extent of binding of a drug to DNA is proportional to drug concentration and is usually irreversible. The effect on tumor growth is therefore a consequence of cumulative exposure to the drug. Measures of cumulative exposure, such as AUC, provide a means to individualize treatment.

THE TARGET CONCENTRATION APPROACH TO DESIGNING A RATIONAL DOSAGE REGIMEN

A rational dosage regimen is based on the assumption that there is a **target concentration** that will produce the desired therapeutic effect. By considering the pharmacokinetic factors that determine the dose-concentration relationship, it is possible to individualize the dose regimen to achieve the target concentration. The effective concentration ranges shown in Table 3–1 are a guide to the concentrations measured when patients are being effectively treated. The initial target concentration should usually be chosen from the lower end of this range. In some cases, the target concentration will also depend on the specific therapeutic objective—eg, the control of atrial fibrillation by digoxin often requires a target

concentration of 2 ng/mL, while heart failure is usually adequately managed with a target concentration of 1 ng/mL.

Maintenance Dose

In most clinical situations, drugs are administered in such a way as to maintain a steady state of drug in the body, ie, just enough drug is given in each dose to replace the drug eliminated since the preceding dose. Thus, calculation of the appropriate maintenance dose is a primary goal. Clearance is the most important pharmacokinetic term to be considered in defining a rational steady-state drug dosage regimen. At steady state, the dosing rate ("rate in") must equal the rate of elimination ("rate out"). Substitution of the target concentration (TC) for concentration (C) in equation (4) predicts the maintenance dosing rate:

$$\text{Dosing rate}_{ss} = \text{Rate of elimination}_{ss}$$
$$= CL \times TC \qquad (9)$$

Thus, if the desired target concentration is known, the clearance in that patient will determine the dosing rate. If the drug is given by a route that has a bioavailability less than 100%, then the dosing rate predicted by equation (9) must be modified. For oral dosing:

$$\text{Dosing rate}_{oral} = \frac{\text{Dosing rate}}{F_{oral}} \qquad (10)$$

If intermittent doses are given, the maintenance dose is calculated from:

$$\text{Maintenance dose} = \text{Dosing rate} \times \text{Dosing interval} \qquad (11)$$

(See Box: Example: Maintenance Dose Calculations.)

Note that the steady-state concentration achieved by continuous infusion or the average concentration following intermittent dosing depends only on clearance. The volume of distribution and the half-life need not be known in order to determine the average plasma concentration expected from a given dosing rate or to predict the dosing rate for a desired target concentration. Figure 3–6 shows that at different dosing intervals, the concentration-time curves will have different maximum and minimum values even though the average level will always be 10 mg/L.

Estimates of dosing rate and average steady-state concentrations, which may be calculated using clearance, are independent of any specific pharmacokinetic model. In contrast, the determination of maximum and minimum steady-state concentrations requires further assumptions about the pharmacokinetic model. The accumulation factor (equation [7]) assumes that the drug follows a one-compartment body model (Figure 3–2B), and the peak concentration prediction assumes that the absorption rate is much faster than the elimination rate. For the calculation of estimated maximum and minimum concentrations in a clinical situation, these assumptions are usually reasonable.

Loading Dose

When the time to reach steady state is appreciable, as it is for drugs with long half-lives, it may be desirable to administer a loading dose that promptly raises the concentration of drug in plasma to the target concentration. In theory, only the amount of the loading dose need be computed—not the rate of its administration—and, to a first approximation, this is so. The volume of distribution is the proportionality factor that relates the total amount of drug in the body to the concentration; if a loading dose is to achieve the target concentration, then from equation (1):

Example: Maintenance Dose Calculations

A target plasma theophylline concentration of 10 mg/L is desired to relieve acute bronchial asthma in a patient. If the patient is a nonsmoker and otherwise normal except for asthma, we may use the mean clearance given in Table 3–1, ie, 2.8 L/h/70 kg. Since the drug will be given as an intravenous infusion, F = 1.

$$\text{Dosing rate} = CL \times TC$$
$$= 2.8\ \text{L/h/70 kg} \times 10\ \text{mg/L}$$
$$= 28\ \text{mg/h/70 kg}$$

Therefore, in this patient, the proper infusion rate would be 28 mg/h/70 kg.

If the asthma attack is relieved, the clinician might want to maintain this plasma level using oral theophylline, which might be given every 12 hours using an extended-release formulation to approximate a continuous intravenous infusion. According to Table 3–1, F_{oral} is 0.96. When the dosing interval is 12 hours, the size of each maintenance dose would be:

$$\text{Maintenance dose} = \frac{\text{Dosing rate}}{F} \times \text{Dosing interval}$$
$$= \frac{28\ \text{mg/h}}{0.96} \times 12\ \text{hours}$$
$$= 350\ \text{mg}$$

A tablet or capsule size close to the ideal dose of 350 mg would then be prescribed at 12-hourly intervals. If an 8-hour dosing interval was used, the ideal dose would be 233 mg; and if the drug was given once a day, the dose would be 700 mg. In practice, F could be omitted from the calculation since it is so close to 1.

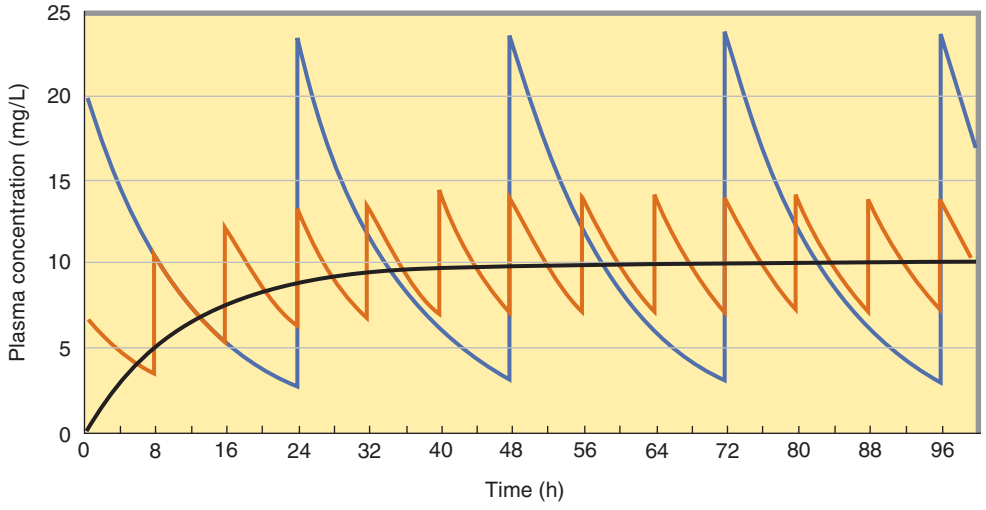


FIGURE 3–6 Relationship between frequency of dosing and maximum and minimum plasma concentrations when a steady-state theophylline plasma level of 10 mg/L is desired. The smoothly rising black line shows the plasma concentration achieved with an intravenous infusion of 28 mg/h. The doses for 8-hourly administration (orange line) are 224 mg; for 24-hourly administration (blue line), 672 mg. In each of the three cases, the mean steady-state plasma concentration is 10 mg/L.

$$\text{Loading dose} = \begin{matrix}\text{Amount in the body}\\ \text{immediately following}\\ \text{the loading dose}\end{matrix} = V \times TC \qquad (12)$$

For the theophylline example given in the Box, Example: Maintenance Dose Calculations, the loading dose would be 350 mg (35 L × 10 mg/L) for a 70-kg person. For most drugs, the loading dose can be given as a single dose by the chosen route of administration.

Up to this point, we have ignored the fact that some drugs follow more complex multicompartment pharmacokinetics, eg, the distribution process illustrated by the two-compartment model in Figure 3–2. This is justified in the great majority of cases. However, in some cases the distribution phase may not be ignored, particularly in connection with the calculation of loading doses. If the rate of absorption is rapid relative to distribution (this is always true for rapid intravenous administration), the concentration of drug in plasma that results from an appropriate loading dose—calculated using the apparent volume of distribution—can initially be considerably higher than desired. Severe toxicity may occur, albeit transiently. This may be particularly important, eg, in the administration of antiarrhythmic drugs such as lidocaine, where an almost immediate toxic response may occur. Thus, while the estimation of the *amount* of a loading dose may be quite correct, the *rate of administration* can sometimes be crucial in preventing excessive drug concentrations, and slow administration of an intravenous drug (over minutes rather than seconds) is almost always prudent practice.

When intermittent doses are given, the loading dose calculated from equation (12) will only reach the average steady-state concentration and will not match the peak steady-state concentration (Figure 3–6). To match the peak steady-state concentration, the loading dose can be calculated from equation (13):

$$\text{Loading dose} = \text{Maintenance dose} \times \text{Accumulation factor} \qquad (13)$$

TARGET CONCENTRATION INTERVENTION: APPLICATION OF PHARMACOKINETICS & PHARMACODYNAMICS TO DOSE INDIVIDUALIZATION

The basic principles outlined above can be applied to the interpretation of clinical drug concentration measurements on the basis of three major pharmacokinetic variables: absorption, clearance, and volume of distribution (and the derived variable, half-life). In addition, it may be necessary to consider two pharmacodynamic variables: maximum effect attainable in the target tissue and the sensitivity of the tissue to the drug. Diseases may modify all of these parameters, and the ability to predict the effect of disease states on pharmacokinetic parameters is important in properly adjusting dosage in such cases. (See Box: The Target Concentration Strategy.)

Pharmacokinetic Variables

A. Absorption

The amount of drug that enters the body depends on the patient's adherence to the prescribed regimen and on the rate and extent of transfer from the site of administration to the blood.

The Target Concentration Strategy

Recognition of the essential role of concentration in linking pharmacokinetics and pharmacodynamics leads naturally to the target concentration strategy. Pharmacodynamic principles can be used to predict the concentration required to achieve a particular degree of therapeutic effect. This target concentration can then be achieved by using pharmacokinetic principles to arrive at a suitable dosing regimen (Holford, 1999). The target concentration strategy is a process for optimizing the dose in an individual on the basis of a measured surrogate response such as drug concentration:

1. Choose the target concentration, TC.
2. Predict volume of distribution (V) and clearance (CL) based on standard population values (eg, Table 3–1) with adjustments for factors such as weight and renal function.
3. Give a loading dose or maintenance dose calculated from TC, V, and CL.
4. Measure the patient's response and drug concentration.
5. Revise V and/or CL based on the measured concentration.
6. Repeat steps 3–5, adjusting the predicted dose to achieve TC.

Overdosage and underdosage relative to the prescribed dosage—both aspects of failure of adherence—can frequently be detected by concentration measurements when gross deviations from expected values are obtained. If adherence is found to be adequate, absorption abnormalities in the small bowel may be the cause of abnormally low concentrations. Variations in the extent of bioavailability are rarely caused by irregularities in the manufacture of the particular drug formulation. More commonly, variations in bioavailability are due to metabolism during absorption.

B. Clearance

Abnormal clearance may be anticipated when there is major impairment of the function of the kidney, liver, or heart. Creatinine clearance is a useful quantitative indicator of renal function. Conversely, drug clearance may be a useful indicator of the functional consequences of heart, kidney, or liver failure, often with greater precision than clinical findings or other laboratory tests. For example, when renal function is changing rapidly, estimation of the clearance of aminoglycoside antibiotics may be a more accurate indicator of glomerular filtration than serum creatinine.

Hepatic disease has been shown to reduce the clearance and prolong the half-life of many drugs. However, for many other drugs known to be eliminated by hepatic processes, no changes in clearance or half-life have been noted with similar hepatic disease. This reflects the fact that hepatic disease does not always affect the hepatic intrinsic clearance. At present, there is no reliable marker of hepatic drug-metabolizing function that can be used to predict changes in liver clearance in a manner analogous to the use of creatinine clearance as a marker of renal drug clearance.

C. Volume of Distribution

The apparent volume of distribution reflects a balance between binding to tissues, which decreases plasma concentration and makes the apparent volume larger, and binding to plasma proteins, which increases plasma concentration and makes the apparent volume smaller. Changes in either tissue or plasma binding can change the apparent volume of distribution determined from plasma concentration measurements. Older people have a relative decrease in skeletal muscle mass and tend to have a smaller apparent volume of distribution of digoxin (which binds to muscle proteins). The volume of distribution may be overestimated in obese patients if based on body weight and the drug does not enter fatty tissues well, as is the case with digoxin. In contrast, theophylline has a volume of distribution similar to that of total body water. Adipose tissue has almost as much water in it as other tissues, so that the apparent total volume of distribution of theophylline is proportional to body weight even in obese patients.

Abnormal accumulation of fluid—edema, ascites, pleural effusion—can markedly increase the volume of distribution of drugs such as gentamicin that are hydrophilic and have small volumes of distribution.

D. Half-Life

The differences between clearance and half-life are important in defining the underlying mechanisms for the effect of a disease state on drug disposition. For example, the half-life of diazepam increases with patient age. When clearance is related to age, it is found that clearance of this drug does not change with age. The increasing half-life for diazepam actually results from changes in the volume of distribution with age; the metabolic processes responsible for eliminating the drug are fairly constant.

Pharmacodynamic Variables

A. Maximum Effect

All pharmacologic responses must have a maximum effect (E_{max}). No matter how high the drug concentration goes, a point will be reached beyond which no further increment in response is achieved.

If increasing the dose in a particular patient does not lead to a further clinical response, it is possible that the maximum effect has been reached. Recognition of maximum effect is helpful in avoiding ineffectual increases of dose with the attendant risk of toxicity.

B. Sensitivity

The sensitivity of the target organ to drug concentration is reflected by the concentration required to produce 50% of maximum effect, the C_{50}. Diminished sensitivity to the drug can be detected by measuring drug concentrations that are usually associated with therapeutic response in a patient who has not responded. This may be a result of abnormal physiology—eg, hyperkalemia diminishes responsiveness to digoxin—or drug

antagonism—eg, calcium channel blockers impair the inotropic response to digoxin.

Increased sensitivity to a drug is usually signaled by exaggerated responses to small or moderate doses. The pharmacodynamic nature of this sensitivity can be confirmed by measuring drug concentrations that are low in relation to the observed effect.

INTERPRETATION OF DRUG CONCENTRATION MEASUREMENTS

Clearance

Clearance is the single most important factor determining drug concentrations. The interpretation of measurements of drug concentrations depends on a clear understanding of three factors that may influence clearance: the dose, the organ blood flow, and the intrinsic function of the liver or kidneys. Each of these factors should be considered when interpreting clearance estimated from a drug concentration measurement. It must also be recognized that changes in protein binding may lead the unwary to believe there is a change in clearance when in fact drug elimination is not altered (see Box: Plasma Protein Binding: Is It Important?). Factors affecting protein binding include the following:

1. **Albumin concentration:** Drugs such as phenytoin, salicylates, and disopyramide are extensively bound to plasma albumin. Albumin levels are low in many disease states, resulting in lower total drug concentrations.
2. **Alpha$_1$-acid glycoprotein concentration:** α_1-Acid glycoprotein is an important binding protein with binding sites for drugs such as quinidine, lidocaine, and propranolol. It is increased in acute inflammatory disorders and causes major changes in total plasma concentration of these drugs even though drug elimination is unchanged.
3. **Capacity-limited protein binding:** The binding of drugs to plasma proteins is capacity-limited. Therapeutic concentrations of salicylates and prednisolone show concentration-dependent protein binding. Because unbound drug concentration is determined by dosing rate and clearance—which is not altered, in the case of these low-extraction-ratio drugs, by protein binding—increases in dosing rate will cause corresponding changes in the pharmacodynamically important unbound concentration. In contrast, total drug concentration will increase less rapidly than the dosing rate would suggest as protein binding approaches saturation at higher concentrations.

Dosing History

An accurate dosing history is essential if one is to obtain maximum value from a drug concentration measurement. In fact, if the dosing history is unknown or incomplete, a drug concentration measurement loses all predictive value.

Timing of Samples for Concentration Measurement

Information about the rate and extent of drug absorption in a particular patient is rarely of great clinical importance. However, absorption usually occurs during the first 2 hours after a drug dose and varies according to food intake, posture, and activity. Therefore, it is important to avoid drawing blood until absorption is complete (about 2 hours after an oral dose). Attempts to measure peak concentrations early after oral dosing are usually unsuccessful and compromise the validity of the measurement, because one cannot be certain that absorption is complete.

Some drugs such as digoxin and lithium take several hours to distribute to tissues. Digoxin samples should be taken at least 6 hours after the last dose and lithium just before the next dose (usually 24 hours after the last dose). Aminoglycosides distribute

Plasma Protein Binding: Is It Important?

Plasma protein binding is often mentioned as a factor playing a role in pharmacokinetics, pharmacodynamics, and drug interactions. However, there are no clinically relevant examples of changes in drug disposition or effects that can be clearly ascribed to changes in plasma protein binding (Benet & Hoener, 2002). The idea that if a drug is displaced from plasma proteins it would increase the unbound drug concentration and increase the drug effect and, perhaps, produce toxicity seems a simple and obvious mechanism. Unfortunately, this simple theory, which is appropriate for a test tube, does not work in the body, which is an open system capable of eliminating unbound drug.

First, a seemingly dramatic change in the unbound fraction from 1% to 10% releases less than 5% of the total amount of drug in the body into the unbound pool because less than one third of the drug in the body is bound to plasma proteins even in the most extreme cases, eg, warfarin. Drug displaced from plasma protein will of course distribute throughout the volume of distribution, so that a 5% increase in the amount of unbound drug in the body produces at most a 5% increase in pharmacologically active unbound drug at the site of action.

Second, when the amount of unbound drug in plasma increases, the rate of elimination will increase (if unbound clearance is unchanged), and after four half-lives the unbound concentration will return to its previous steady-state value. When drug interactions associated with protein binding displacement and clinically important effects have been studied, it has been found that the displacing drug is also an inhibitor of clearance, and it is the change in *clearance* of the *unbound* drug that is the relevant mechanism explaining the interaction.

The clinical importance of plasma protein binding is only to help interpretation of measured drug concentrations. When plasma proteins are lower than normal, then total drug concentrations will be lower but unbound concentrations will not be affected.

quite rapidly, but it is still prudent to wait 1 hour after giving the dose before taking a sample.

Clearance is readily estimated from the dosing rate and mean steady-state concentration. Blood samples should be appropriately timed to estimate steady-state concentration. Provided steady state has been approached (at least three half-lives of constant dosing), a sample obtained near the midpoint of the dosing interval will usually be close to the mean steady-state concentration.

Initial Predictions of Volume of Distribution & Clearance

A. Volume of Distribution

Volume of distribution is commonly calculated for a particular patient using body weight (70-kg body weight is assumed for the values in Table 3–1). If a patient is obese, drugs that do not readily penetrate fat (eg, gentamicin and digoxin) should have their volumes calculated from fat-free mass (FFM) as shown below. Total body weight (WT) is in kilograms and height (HTM) is in meters:

$$\text{For women: FFM (kg)} = \frac{37.99 \times \text{HTM}^2 \times \text{WT}}{35.98 \times \text{HTM}^2 + \text{WT}} \quad (14a)$$

$$\text{For men: FFM (kg)} = \frac{42.92 \times \text{HTM}^2 \times \text{WT}}{30.93 \times \text{HTM}^2 + \text{WT}} \quad (14b)$$

Patients with edema, ascites, or pleural effusions offer a larger volume of distribution to the aminoglycoside antibiotics (eg, gentamicin) than is predicted by body weight. In such patients, the weight should be corrected as follows: Subtract an estimate of the weight of the excess fluid accumulation from the measured weight. Use the resultant "normal" body weight to calculate the normal volume of distribution. Finally, this normal volume should be increased by 1 L for each estimated kilogram of excess fluid. This correction is important because of the relatively small volumes of distribution of these water-soluble drugs.

B. Clearance

Drugs cleared by the renal route often require adjustment of clearance in proportion to renal function. This can be conveniently estimated from the creatinine clearance, calculated from a single serum creatinine measurement and the predicted creatinine production rate.

The predicted creatinine production rate in women is 85% of the calculated value, because they have a smaller muscle mass per kilogram and it is muscle mass that determines creatinine production. Muscle mass as a fraction of body weight decreases with age, which is why age appears in the Cockcroft-Gault equation.*

The decrease of renal function with age is independent of the decrease in creatinine production. Because of the difficulty of obtaining complete urine collections, creatinine clearance calculated in this way is at least as reliable as estimates based on urine collections. The fat-free mass (equation [14]) should be considered rather than total body weight for obese patients, and correction should be made for muscle wasting in severely ill patients.

Revising Individual Estimates of Volume of Distribution & Clearance

The commonsense approach to the interpretation of drug concentrations compares predictions of pharmacokinetic parameters and expected concentrations to measured values. If measured concentrations differ by more than 20% from predicted values, revised estimates of V or CL for that patient should be calculated using equation (1) or equation (2). If the change calculated is more than a 100% increase or 50% decrease in either V or CL, the assumptions made about the timing of the sample and the dosing history should be critically examined.

For example, if a patient is taking 0.25 mg of digoxin a day, a clinician may expect the digoxin concentration to be about 1 ng/mL. This is based on typical values for bioavailability of 70% and total clearance of about 7 L/h (CL_{renal} 4 L/h, $CL_{nonrenal}$ 3 L/h). If the patient has heart failure, the nonrenal (hepatic) clearance might be halved because of hepatic congestion and hypoxia, so the expected clearance would become 5.5 L/h. The concentration is then expected to be about 1.3 ng/mL. Suppose that the concentration actually measured is 2 ng/mL. Common sense would suggest halving the daily dose to achieve a target concentration of 1 ng/mL. This approach implies a revised clearance of 3.5 L/h. The smaller clearance compared with the expected value of 5.5 L/h may reflect additional renal functional impairment due to heart failure.

This technique will often be misleading if steady state has not been reached. At least a week of regular dosing (three to four half-lives) must elapse before the implicit method will be reliable.

REFERENCES

Benet LZ, Hoener B: Changes in plasma protein binding have little clinical relevance. Clin Pharmacol Ther 2002;71:115.

Holford NHG: Pharmacokinetic and pharmacodynamic principles. 2011; http://www.fmhs.auckland.ac.nz/sms/pharmacology/holford/teaching/medsci722

Holford NHG: Target concentration intervention: Beyond Y2K. Br J Clin Pharmacol 1999:48:9.

Holford NHG, Sheiner LB: Understanding the dose-effect relationship. Clin Pharmacokinet 1981;6:429.

Speight T, Holford NHG: *Avery's Drug Treatment*, 4th ed. Adis International, 1997.

* The Cockcroft-Gault equation is given in Chapter 60.

CASE STUDY ANSWER

Sixty seven percent of total standard digoxin clearance is renal, so the standard renal clearance is 0.67 × 9 L/h = 6 L/h/70 kg with creatinine clearance of 100 mL/min and nonrenal clearance is (1 – 0.67) × 9 L/h = 3 L/h/70 kg (see Table 3–1 for standard pharmacokinetic parameters). Her predicted creatinine clearance is 22 mL/min (Cockcroft & Gault), so for digoxin, her renal clearance is 6 × 22/100 × 60/70 = 1.1 L/h, nonrenal clearance 2.7 × 60/70 = 2.6 L/h, and total clearance 3.7 L/h. The parenteral maintenance dose rate is 2 mcg/L × 3.7 L/h = 7.4 mcg/h. Once-a-day oral dosing with bioavailability of 0.7 would require a daily maintenance dose of 7.4/0.7 × 24 = 254 mcg/day. A practical dose would be one 250 mcg tablet per day.

This page intentionally left blank

CHAPTER 4

Drug Biotransformation

Maria Almira Correia, PhD

Humans are exposed daily to a wide variety of foreign compounds called **xenobiotics**—substances absorbed across the lungs or skin or, more commonly, ingested either unintentionally as compounds present in food and drink or deliberately as drugs for therapeutic or "recreational" purposes. Exposure to environmental xenobiotics may be inadvertent and accidental or—when they are present as components of air, water, and food—inescapable. Some xenobiotics are innocuous, but many can provoke biologic responses. Such biologic responses often depend on conversion of the absorbed substance into an active metabolite. The discussion that follows is applicable to xenobiotics in general (including drugs) and to some extent to endogenous compounds.

WHY IS DRUG BIOTRANSFORMATION NECESSARY?

Renal excretion plays a pivotal role in terminating the biologic activity of some drugs, particularly those that have small molecular volumes or possess polar characteristics, such as functional groups that are fully ionized at physiologic pH. However, many drugs do not possess such physicochemical properties. Pharmacologically active organic molecules tend to be lipophilic and remain unionized or only partially ionized at physiologic pH; these are readily reabsorbed from the glomerular filtrate in the nephron. Certain lipophilic compounds are often strongly bound to plasma proteins and may not be readily filtered at the glomerulus. Consequently, most drugs would have a prolonged duration of action if termination of their action depended solely on renal excretion.

An alternative process that can lead to the termination or alteration of biologic activity is metabolism. In general, lipophilic xenobiotics are transformed to more polar and hence more readily excreted products. The role that metabolism plays in the inactivation of lipid-soluble drugs can be quite dramatic. For example, lipophilic barbiturates such as thiopental and pentobarbital would have extremely long half-lives if it were not for their metabolic conversion to more water-soluble compounds.

Metabolic products are often less pharmacodynamically active than the parent drug and may even be inactive. However, some biotransformation products have *enhanced* activity or toxic properties. It is noteworthy that the synthesis of endogenous substrates such as steroid hormones, cholesterol, active vitamin D congeners, and bile acids involves many pathways catalyzed by enzymes associated with the metabolism of xenobiotics. Finally, drug-metabolizing enzymes have been exploited in the design of pharmacologically inactive prodrugs that are converted to active molecules in the body.

THE ROLE OF BIOTRANSFORMATION IN DRUG DISPOSITION

Most metabolic biotransformations occur at some point between absorption of the drug into the general circulation and its renal elimination. A few transformations occur in the intestinal lumen or intestinal wall. In general, all of these reactions can be assigned to one of two major categories called **phase I** and **phase II reactions** (Figure 4–1).

Phase I reactions usually convert the parent drug to a more polar metabolite by introducing or unmasking a functional group (–OH, $-NH_2$, –SH). Often these metabolites are inactive, although in some instances activity is only modified or even enhanced.

If phase I metabolites are sufficiently polar, they may be readily excreted. However, many phase I products are not eliminated rapidly and undergo a subsequent reaction in which an endogenous substrate such as glucuronic acid, sulfuric acid, acetic acid, or an amino acid combines with the newly incorporated functional group to form a highly polar conjugate. Such conjugation or synthetic reactions are the hallmarks of phase II metabolism. A great variety of drugs undergo these sequential biotransformation reactions, although in some instances the parent drug may already possess a functional group that may form a conjugate directly. For example, the hydrazide moiety of isoniazid is known to form an *N*-acetyl conjugate in a phase II reaction. This conjugate is then a

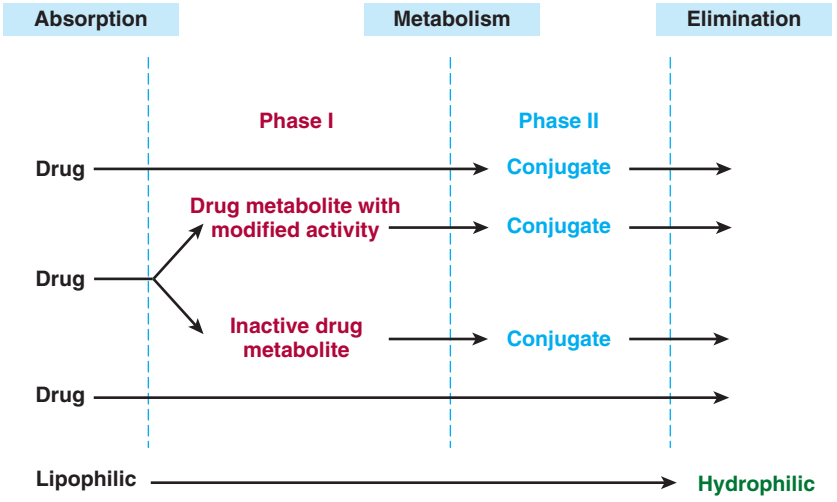


FIGURE 4–1 Phase I and phase II reactions, and direct elimination, in drug biodisposition. Phase II reactions may also precede phase I reactions.

substrate for a phase I type reaction, namely, hydrolysis to isonicotinic acid (Figure 4–2). Thus, phase II reactions may actually precede phase I reactions.

WHERE DO DRUG BIOTRANSFORMATIONS OCCUR?

Although every tissue has some ability to metabolize drugs, the liver is the principal organ of drug metabolism. Other tissues that display considerable activity include the gastrointestinal tract, the lungs, the skin, the kidneys, and the brain. After oral administration, many drugs (eg, isoproterenol, meperidine, pentazocine, morphine) are absorbed intact from the small intestine and transported first via the portal system to the liver, where they undergo extensive metabolism. This process is called the **first-pass effect** (see Chapter 3). Some orally administered drugs (eg, clonazepam, chlorpromazine, cyclosporine) are more extensively metabolized in the intestine than in the liver, while others (eg, midazolam) undergo significant (≈ 50%) intestinal metabolism. Thus, intestinal metabolism can contribute to the overall first-pass effect, and individuals with compromised liver function may rely increasingly on such intestinal metabolism for drug elimination. Compromise of intestinal metabolism of certain drugs (eg, felodipine, cyclosporine A) can also result in significant elevation of their plasma levels and clinically relevant drug-drug interactions (DDIs, see below). First-pass effects may so greatly limit the bioavailability of orally administered drugs (eg, lidocaine) that alternative routes of administration must be used to achieve therapeutically effective blood levels. Furthermore, the lower gut harbors intestinal microorganisms that are capable of many biotransformation reactions. In addition, drugs may be metabolized by gastric acid (eg, penicillin), by digestive enzymes (eg, polypeptides such as insulin), or by enzymes in the wall of the intestine (eg, sympathomimetic catecholamines).

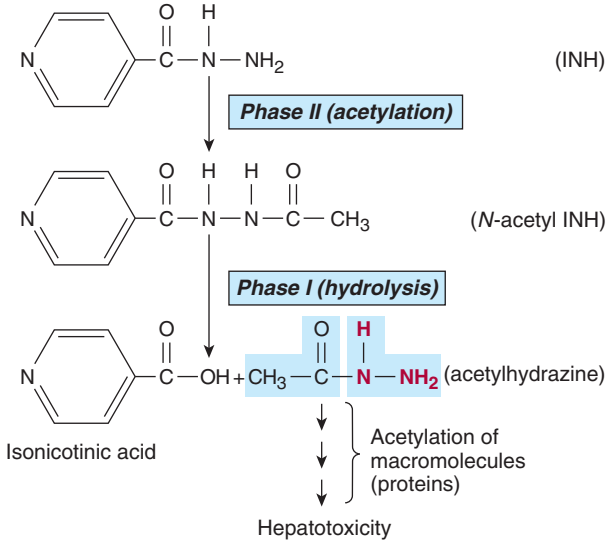


FIGURE 4–2 Phase II activation of isoniazid (INH) to a hepatotoxic metabolite.

Although drug biotransformation in vivo can occur by spontaneous, noncatalyzed chemical reactions, most transformations are catalyzed by specific cellular enzymes. At the subcellular level, these enzymes may be located in the endoplasmic reticulum (ER), mitochondria, cytosol, lysosomes, or even the nuclear envelope or plasma membrane.

MICROSOMAL MIXED FUNCTION OXIDASE SYSTEM & PHASE I REACTIONS

Many drug-metabolizing enzymes are located in the lipophilic endoplasmic reticulum membranes of the liver and other tissues.

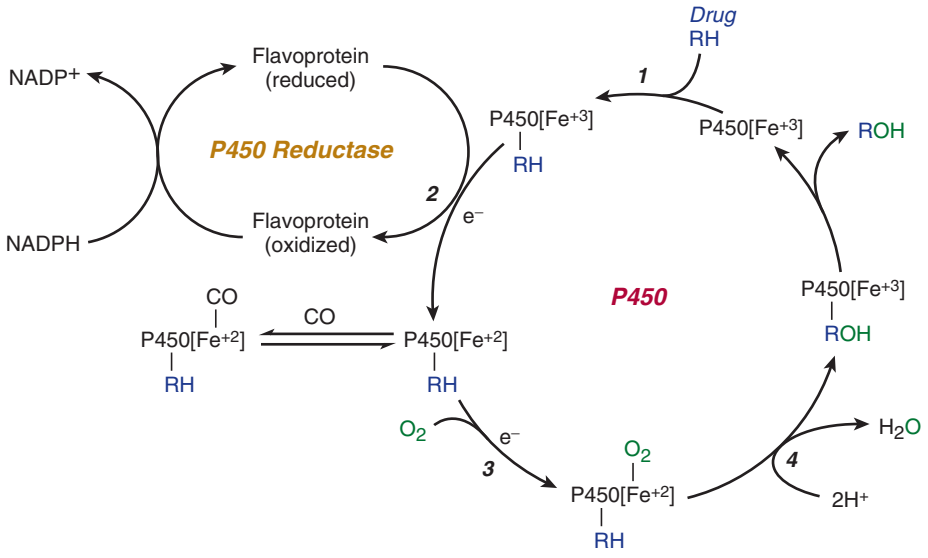


FIGURE 4–3 Cytochrome P450 cycle in drug oxidations. RH, parent drug; ROH, oxidized metabolite; e^-, electron.

When these lamellar membranes are isolated by homogenization and fractionation of the cell, they re-form into vesicles called **microsomes.** Microsomes retain most of the morphologic and functional characteristics of the intact membranes, including the rough and smooth surface features of the rough (ribosome-studded) and smooth (no ribosomes) endoplasmic reticulum. Whereas the rough microsomes tend to be dedicated to protein synthesis, the smooth microsomes are relatively rich in enzymes responsible for oxidative drug metabolism. In particular, they contain the important class of enzymes known as the **mixed function oxidases** (MFOs), or **monooxygenases.** The activity of these enzymes requires both a reducing agent (nicotinamide adenine dinucleotide phosphate [NADPH]) and molecular oxygen; in a typical reaction, one molecule of oxygen is consumed (reduced) per substrate molecule, with one oxygen atom appearing in the product and the other in the form of water.

In this oxidation-reduction process, two microsomal enzymes play a key role. The first of these is a flavoprotein, **NADPH-cytochrome P450 oxidoreductase** (POR). One mole of this enzyme contains 1 mol each of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The second microsomal enzyme is a hemoprotein called **cytochrome P450,** which serves as the terminal oxidase. In fact, the microsomal membrane harbors multiple forms of this hemoprotein, and this multiplicity is increased by repeated administration of or exposure to exogenous chemicals (see text that follows). The name cytochrome P450 (abbreviated as **P450** or **CYP**) is derived from the spectral properties of this hemoprotein. In its reduced (ferrous) form, it binds carbon monoxide to give a complex that absorbs light maximally at 450 nm. The relative abundance of P450s, compared with that of the reductase in the liver, contributes to making P450 heme reduction a rate-limiting step in hepatic drug oxidations.

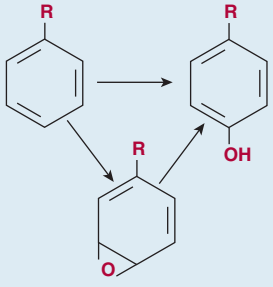
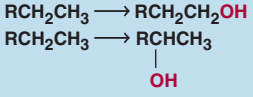
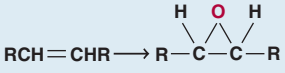
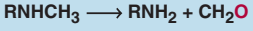
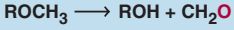
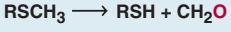
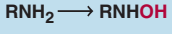
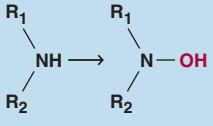
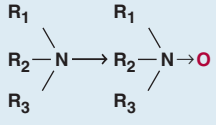
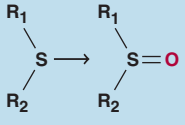
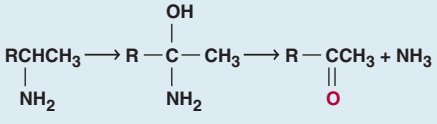
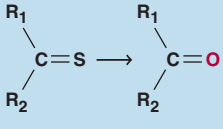
Microsomal drug oxidations require P450, P450 reductase, NADPH, and molecular oxygen. A simplified scheme of the oxidative cycle is presented in Figure 4–3. Briefly, oxidized (Fe^{3+}) P450 combines with a drug substrate to form a binary complex (step 1). NADPH donates an electron to the flavoprotein P450 reductase, which in turn reduces the oxidized P450-drug complex (step 2). A second electron is introduced from NADPH via the same P450 reductase, which serves to reduce molecular oxygen and to form an "activated oxygen"-P450-substrate complex (step 3). This complex in turn transfers activated oxygen to the drug substrate to form the oxidized product (step 4).

The potent oxidizing properties of this activated oxygen permit oxidation of a large number of substrates. Substrate specificity is very low for this enzyme complex. High lipid solubility is the only common structural feature of the wide variety of structurally unrelated drugs and chemicals that serve as substrates in this system (Table 4–1). However, compared with many other enzymes including phase II enzymes, P450s are remarkably sluggish catalysts, and their drug biotransformation reactions are slow.

HUMAN LIVER P450 ENZYMES

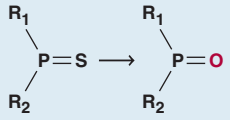
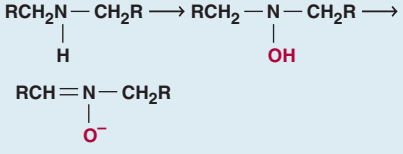
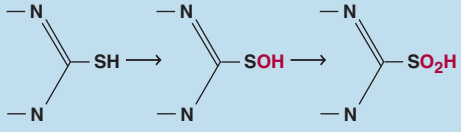
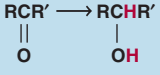
Gene arrays combined with immunoblotting analyses of microsomal preparations, as well as the use of relatively selective functional markers and selective P450 inhibitors, have identified numerous P450 isoforms (CYP: 1A2, 2A6, 2B6, 2C8, 2C9, 2C18, 2C19, 2D6, 2E1, 3A4, 3A5, 4A11, and 7) in the human liver. Of these, **CYP1A2, CYP2A6, CYP2B6, CYP2C9, CYP2D6, CYP2E1,** and **CYP3A4** appear to be the most important forms, accounting for approximately 15%, 4%, 1%, 20%, 5%, 10%, and 30%, respectively, of the total human liver P450 content. Together, they

TABLE 4–1 Phase I reactions.

Reaction Class	Structural Change	Drug Substrates
Oxidations		
Cytochrome P450-dependent oxidations:		
Aromatic hydroxylations		Acetanilide, propranolol, phenobarbital, phenytoin, phenylbutazone, amphetamine, warfarin, 17α-ethinyl estradiol, naphthalene, benzpyrene
Aliphatic hydroxylations		Amobarbital, pentobarbital, secobarbital, chlorpropamide, ibuprofen, meprobamate, glutethimide, phenylbutazone, digitoxin
Epoxidation		Aldrin
Oxidative dealkylation		
N-Dealkylation		Morphine, ethylmorphine, benzphetamine, aminopyrine, caffeine, theophylline
O-Dealkylation		Codeine, *p*-nitroanisole
S-Dealkylation		6-Methylthiopurine, methitural
***N*-Oxidation**		
Primary amines		Aniline, chlorphentermine
Secondary amines		2-Acetylaminofluorene, acetaminophen
Tertiary amines		Nicotine, methaqualone
***S*-Oxidation**		Thioridazine, cimetidine, chlorpromazine
Deamination		Amphetamine, diazepam
Desulfuration		Thiopental

(*continued*)

TABLE 4–1 Phase I reactions. (Continued)

Reaction Class	Structural Change	Drug Substrates
Cytochrome P450- dependent oxidations: (continued)		
		Parathion
Dechlorination	$CCl_4 \longrightarrow [CCl_3^{\bullet}] \longrightarrow CHCl_3$	Carbon tetrachloride
Cytochrome P450-independent oxidations:		
Flavin monooxygenase (Ziegler's enzyme)	$R_3N \longrightarrow R_3N^+ \rightarrow O^- \xrightarrow{H^+} R_3N^+OH$	Chlorpromazine, amitriptyline, benzphetamine
		Desipramine, nortriptyline
		Methimazole, propylthiouracil
Amine oxidases	$RCH_2NH_2 \longrightarrow RCHO + NH_3$	Phenylethylamine, epinephrine
Dehydrogenations	$RCH_2OH \longrightarrow RCHO$	Ethanol
Reductions		
Azo reductions	$RN{=}NR_1 \longrightarrow RNH{-}NHR_1 \longrightarrow RNH_2 + R_1NH_2$	Prontosil, tartrazine
Nitro reductions	$RNO_2 \longrightarrow RNO \longrightarrow RNHOH \longrightarrow RNH_2$	Nitrobenzene, chloramphenicol, clonazepam, dantrolene
Carbonyl reductions		Metyrapone, methadone, naloxone
Hydrolyses		
Esters	$R_1COOR_2 \longrightarrow R_1COOH + R_2OH$	Procaine, succinylcholine, aspirin, clofibrate, methylphenidate
Amides	$RCONHR_1 \longrightarrow RCOOH + R_1NH_2$	Procainamide, lidocaine, indomethacin

are responsible for catalyzing the bulk of the hepatic drug and xenobiotic metabolism (Table 4–2, Figure 4–4).

It is noteworthy that CYP3A4 alone is responsible for the metabolism of over 50% of the prescription drugs metabolized by the liver. The involvement of individual P450s in the metabolism of a given drug may be screened in vitro by means of selective functional markers, selective chemical P450 inhibitors, and P450 antibodies. In vivo, such screening may be accomplished by means of relatively selective noninvasive markers, which include breath tests or urinary analyses of specific metabolites after administration of a P450-selective substrate probe.

Enzyme Induction

Some of the chemically dissimilar P450 substrate drugs, on repeated administration, *induce* P450 expression by enhancing the

TABLE 4–2 Human liver P450s (CYPs), and some of the drugs metabolized (substrates), inducers, and selective inhibitors.

CYP	Substrates	Inducers	Inhibitors
1A2	Acetaminophen, antipyrine, caffeine, clomipramine, phenacetin, tacrine, tamoxifen, theophylline, warfarin	Smoking, charcoal-broiled foods, cruciferous vegetables, omeprazole	Galangin, furafylline, fluvoxamine
2A6	Coumarin, tobacco nitrosamines, nicotine (to cotinine and 2′-hydroxynicotine)	Rifampin, phenobarbital	Tranylcypromine, menthofuran, methoxsalen
2B6	Artemisinin, bupropion, cyclophosphamide, efavirenz, ifosfamide, ketamine, *S*-mephobarbital, *S*-mephenytoin (*N*-demethylation to nirvanol), methadone, nevirapine, propofol, selegiline, sertraline, ticlopidine	Phenobarbital, cyclophosphamide	Ticlopidine, clopidogrel
2C8	Taxol, all-*trans*-retinoic acid	Rifampin, barbiturates	Trimethoprim
2C9	Celecoxib, flurbiprofen, hexobarbital, ibuprofen, losartan, phenytoin, tolbutamide, trimethadione, sulfaphenazole, *S*-warfarin, ticrynafen	Barbiturates, rifampin	Tienilic acid, sulfaphenazole
2C18	Tolbutamide, phenytoin	Phenobarbital	
2C19	Diazepam, *S*-mephenytoin, naproxen, nirvanol, omeprazole, propranolol	Barbiturates, rifampin	*N*3-benzylnirvanol, *N*3-benzylphenobarbital, fluconazole
2D6	Bufuralol, bupranolol, clomipramine, clozapine, codeine, debrisoquin, dextromethorphan, encainide, flecainide, fluoxetine, guanoxan, haloperidol, hydrocodone, 4-methoxy-amphetamine, metoprolol, mexiletine, oxycodone, paroxetine, phenformin, propafenone, propoxyphene, risperidone, selegiline (deprenyl), sparteine, tamoxifen, thioridazine, timolol, tricyclic antidepressants	Unknown	Quinidine, paroxetine
2E1	Acetaminophen, chlorzoxazone, enflurane, halothane, ethanol (a minor pathway)	Ethanol, isoniazid	4-Methylpyrazole, disulfiram
3A4[1]	Acetaminophen, alfentanil, amiodarone, astemizole, cisapride, cocaine, cortisol, cyclosporine, dapsone, diazepam, dihydroergotamine, dihydropyridines, diltiazem, erythromycin, ethinyl estradiol, gestodene, indinavir, lidocaine, lovastatin, macrolides, methadone, miconazole, midazolam, mifepristone, nifedipine, paclitaxel, progesterone, quinidine, rapamycin, ritonavir, saquinavir, spironolactone, sulfamethoxazole, sufentanil, tacrolimus, tamoxifen, terfenadine, testosterone, tetrahydrocannabinol, triazolam, troleandomycin, verapamil	Barbiturates, carbamazepine, glucocorticoids, pioglitazone, phenytoin, rifampin, St. John's wort	Azamulin, clarithromycin, diltiazem, erythromycin, fluconazole, grapefruit juice (furanocoumarins), itraconazole, ketoconazole, ritonavir, troleandomycin

[1]CYP3A5 has similar substrate and inhibitor profiles, but except for a few drugs is generally less active than CYP3A4.

rate of its synthesis or reducing its rate of degradation (Table 4–2). Induction results in accelerated substrate metabolism and usually in a decrease in the pharmacologic action of the inducer and also of co-administered drugs. However, in the case of drugs metabolically transformed to reactive metabolites, enzyme induction may exacerbate metabolite-mediated toxicity.

Various substrates induce P450 isoforms having different molecular masses and exhibiting different substrate specificities and immunochemical and spectral characteristics.

Environmental chemicals and pollutants are also capable of inducing P450 enzymes. As previously noted, exposure to benzo[*a*]pyrene and other polycyclic aromatic hydrocarbons, which are present in tobacco smoke, charcoal-broiled meat, and other organic pyrolysis products, is known to induce CYP1A enzymes and to alter the rates of drug metabolism. Other environmental chemicals known to induce specific P450s include the polychlorinated biphenyls (PCBs), which were once used widely in industry as insulating materials and plasticizers, and 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (dioxin, TCDD), a trace byproduct of the chemical synthesis of the defoliant 2,4,5-T (see Chapter 56).

Increased P450 synthesis requires enhanced transcription and translation along with increased synthesis of heme, its prosthetic cofactor. A cytoplasmic receptor (termed AhR) for polycyclic aromatic hydrocarbons (eg, benzo[*a*]pyrene, dioxin) has been identified. The translocation of the inducer-receptor complex into the nucleus, followed by ligand-induced dimerization with Arnt, a closely related nuclear protein, leads to subsequent activation of regulatory elements of *CYP1A* genes, resulting in their induction. This is also the mechanism of CYP1A induction by cruciferous vegetables, and the proton pump inhibitor, omeprazole. A pregnane X receptor (PXR), a member of the steroid-retinoid-thyroid hormone receptor family, has recently been shown to mediate CYP3A induction by various chemicals (dexamethasone, rifampin, mifepristone, phenobarbital, atorvastatin, and hyperforin,

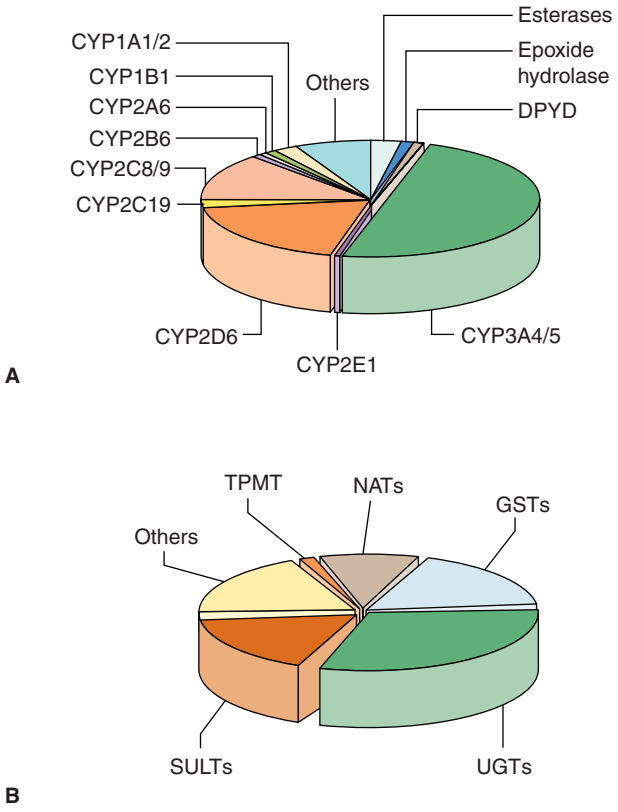


FIGURE 4–4 Relative contributions of various cytochrome P450 isoforms (**A**) and different phase II pathways (**B**) to metabolism of drugs in clinical use. Many drugs are metabolized by two or more of these pathways. Note that two pathways, CYP3A4/5 and UGT, are involved in the metabolism of more than 75% of drugs in use. DPYD, dihydropyrimidine dehydrogenase; GST, glutathione-*S*-transferase; NAT, *N*-acetyltransferase; SULT, sulfotransferase; TPMT, thiopurine methyltransferase; UGT, UDP-glucuronosyltransferase. (Reproduced, with permission, from Brunton LL, Lazo JS, Parker KL: *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 11th ed. McGraw-Hill Medical, 2006.)

a constituent of St. John's wort) in the liver and intestinal mucosa. A similar receptor, the constitutive androstane receptor (CAR), has been identified for the relatively large and structurally diverse phenobarbital class of inducers of CYP2B6, CYP2C9, and CYP3A4. Peroxisome proliferator receptor α (PPAR-α) is yet another nuclear receptor highly expressed in liver and kidneys, which uses lipid-lowering drugs (eg, fenofibrate and gemfibrozil) as ligands. Consistent with its major role in the regulation of fatty acid metabolism, PPAR-α mediates the induction of CYP4A enzymes, responsible for the metabolism of fatty acids such as arachidonic acid and its physiologically relevant derivatives. It is noteworthy that on binding of its particular ligand, PXR, CAR, and PPAR-α each form heterodimers with another nuclear receptor, the retinoid X-receptor (RXR). This heterodimer in turn binds to response elements within the promoter regions of specific *P450* genes to induce gene expression.

P450 enzymes may also be induced by **substrate stabilization,** eg, decreased degradation, as is the case with troleandomycin- or clotrimazole-mediated induction of CYP3A enzymes, the ethanol-mediated induction of CYP2E1, and the isosafrole-mediated induction of CYP1A2.

Enzyme Inhibition

Certain drug substrates inhibit cytochrome P450 enzyme activity (Table 4–2). Imidazole-containing drugs such as cimetidine and ketoconazole bind tightly to the P450 heme iron and effectively reduce the metabolism of endogenous substrates (eg, testosterone) or other co-administered drugs through competitive inhibition. Macrolide antibiotics such as troleandomycin, erythromycin, and erythromycin derivatives are metabolized, apparently by CYP3A, to metabolites that complex the cytochrome P450 heme iron and render it catalytically inactive. Another compound that acts through this mechanism is the inhibitor proadifen (SKF-525-A, used in research), which binds tightly to the heme iron and quasi-irreversibly inactivates the enzyme, thereby inhibiting the metabolism of potential substrates.

Some substrates irreversibly inhibit P450s via covalent interaction of a metabolically generated reactive intermediate that may react with the P450 apoprotein or heme moiety or even cause the heme to fragment and irreversibly modify the apoprotein. The antibiotic chloramphenicol is metabolized by CYP2B1 to a species that modifies the P450 protein and thus also inactivates the enzyme. A growing list of such **suicide inhibitors**—inactivators that attack the heme or the protein moiety—includes certain steroids (ethinyl estradiol, norethindrone, and spironolactone); fluroxene; allobarbital; the analgesic sedatives allylisopropylacetylurea, diethylpentenamide, and ethchlorvynol; carbon disulfide; grapefruit furanocoumarins; selegiline; phencyclidine; ticlopidine and clopidogrel; ritonavir; and propylthiouracil. On the other hand, the barbiturate secobarbital is found to inactivate CYP2B1 by modification of *both* its heme and protein moieties. Other metabolically activated drugs whose P450 inactivation mechanism is not fully elucidated are mifepristone, troglitazone, raloxifene, and tamoxifen.

PHASE II REACTIONS

Parent drugs or their phase I metabolites that contain suitable chemical groups often undergo coupling or conjugation reactions with an endogenous substance to yield **drug conjugates** (Table 4–3). In general, conjugates are polar molecules that are readily excreted and often inactive. Conjugate formation involves high-energy intermediates and specific transfer enzymes. Such enzymes (**transferases**) may be located in microsomes or in the cytosol. Of these, uridine 5′-diphosphate (UDP)-glucuronosyl transferases **(UGTs)** are the most dominant enzymes (Figure 4–4). These microsomal enzymes catalyze the coupling of an activated endogenous substance (such as the UDP derivative of glucuronic acid) with a drug (or endogenous compound such as bilirubin, the end product of heme metabolism). Nineteen *UGT* genes (*UGTA1*

TABLE 4–3 Phase II reactions.

Type of Conjugation	Endogenous Reactant	Transferase (Location)	Types of Substrates	Examples
Glucuronidation	UDP glucuronic acid	UDP glucuronosyltransferase (microsomes)	Phenols, alcohols, carboxylic acids, hydroxylamines, sulfonamides	Nitrophenol, morphine, acetaminophen, diazepam, *N*-hydroxydapsone, sulfathiazole, meprobamate, digitoxin, digoxin
Acetylation	Acetyl-CoA	*N*–Acetyltransferase (cytosol)	Amines	Sulfonamides, isoniazid, clonazepam, dapsone, mescaline
Glutathione conjugation	Glutathione (GSH)	GSH-*S*-transferase (cytosol, microsomes)	Epoxides, arene oxides, nitro groups, hydroxylamines	Acetaminophen, ethacrynic acid, bromobenzene
Glycine conjugation	Glycine	Acyl-CoA glycinetransferase (mitochondria)	Acyl-CoA derivatives of carboxylic acids	Salicylic acid, benzoic acid, nicotinic acid, cinnamic acid, cholic acid, deoxycholic acid
Sulfation	Phosphoadenosyl phosphosulfate	Sulfotransferase (cytosol)	Phenols, alcohols, aromatic amines	Estrone, aniline, phenol, 3-hydroxycoumarin, acetaminophen, methyldopa
Methylation	*S*-Adenosylmethionine	Transmethylases (cytosol)	Catecholamines, phenols, amines	Dopamine, epinephrine, pyridine, histamine, thiouracil
Water conjugation	Water	Epoxide hydrolase (microsomes)	Arene oxides, *cis*-disubstituted and monosubstituted oxiranes	Benzopyrene 7,8-epoxide, styrene 1,2-oxide, carbamazepine epoxide
		(cytosol)	Alkene oxides, fatty acid epoxides	Leukotriene A_4

and *UGT2*) encode UGT proteins involved in the metabolism of drugs and xenobiotics. Similarly, 11 human sulfotransferases (**SULTs**) catalyze the sulfation of substrates using 3′-phosphoadenosine 5′-phosphosulfate (**PAPS**) as the endogenous sulfate donor. Cytosolic and microsomal glutathione (**GSH**) transferases (**GSTs**) are also engaged in the metabolism of drugs and xenobiotics, and in that of leukotrienes and prostaglandins, respectively. Chemicals containing an aromatic amine or a hydrazine moiety (eg, isoniazid) are substrates of cytosolic *N*-acetyltransferases **(NATs**), encoded by *NAT1* and *NAT2* genes, which utilize **acetyl-CoA** as the endogenous cofactor.

S-Adenosyl-L-methionine (**SAMe**; AdoMet)-mediated *O*-, *N*-, and *S*-methylation of drugs and xenobiotics by methyltransferases (**MTs)** also occurs. Finally, endobiotic, drug, and xenobiotic epoxides generated via P450-catalyzed oxidations can also be hydrolyzed by microsomal or cytosolic epoxide hydrolases (**EHs**). Conjugation of an activated drug such as the *S*-CoA derivative of benzoic acid, with an endogenous substrate, such as glycine, also occurs. Because the endogenous substrates originate in the diet, nutrition plays a critical role in the regulation of drug conjugations.

Phase II reactions are relatively faster than P450-catalyzed reactions, thus effectively accelerating drug biotransformation.

Drug conjugations were once believed to represent terminal inactivation events and as such have been viewed as "true detoxification" reactions. However, this concept must be modified, because it is now known that certain conjugation reactions (acyl glucuronidation of nonsteroidal anti-inflammatory drugs, *O*-sulfation of *N*-hydroxyacetylaminofluorene, and *N*-acetylation of isoniazid) may lead to the formation of reactive species responsible for the toxicity of the drugs. Furthermore, sulfation is known to activate the orally active prodrug minoxidil into a very efficacious vasodilator, and morphine-6-glucuronide is more potent than morphine itself.

METABOLISM OF DRUGS TO TOXIC PRODUCTS

Metabolism of drugs and other foreign chemicals may not always be an innocuous biochemical event leading to detoxification and elimination of the compound. Indeed, as previously noted, several compounds have been shown to be metabolically transformed to reactive intermediates that are toxic to various organs. Such toxic reactions may not be apparent at low levels of exposure to parent compounds when alternative detoxification mechanisms are not yet overwhelmed or compromised and when the availability of endogenous detoxifying cosubstrates (GSH, glucuronic acid, sulfate) is not limited. However, when these resources are exhausted, the toxic pathway may prevail, resulting in overt organ toxicity or carcinogenesis. The number of specific examples of such drug-induced toxicity is expanding rapidly. An example is acetaminophen (paracetamol)-induced hepatotoxicity (Figure 4–5). Acetaminophen, an analgesic antipyretic drug, is quite safe in therapeutic doses (1.2 g/d for an adult). It normally undergoes glucuronidation and sulfation to the corresponding conjugates, which together make up 95% of the total excreted metabolites. The alternative P450-dependent GSH conjugation pathway

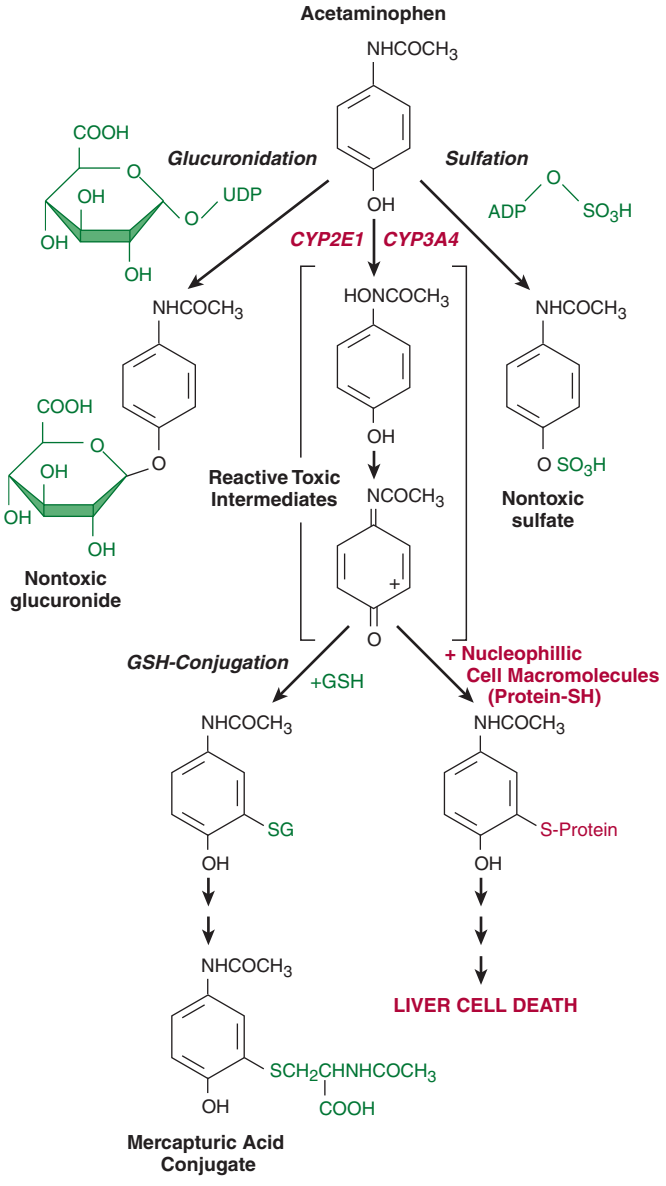


FIGURE 4–5 Metabolism of acetaminophen (top center) to hepatotoxic metabolites. GSH, glutathione; SG, glutathione moiety.

accounts for the remaining 5%. When acetaminophen intake far exceeds therapeutic doses, the glucuronidation and sulfation pathways are saturated, and the P450-dependent pathway becomes increasingly important. Little or no hepatotoxicity results as long as hepatic GSH is available for conjugation. However, with time, hepatic GSH is depleted faster than it can be regenerated, and a reactive, toxic metabolite accumulates. In the absence of intracellular nucleophiles such as GSH, this reactive metabolite (*N*-acetylbenzoiminoquinone) reacts with nucleophilic groups of cellular proteins, resulting in hepatotoxicity.

The chemical and toxicologic characterization of the electrophilic nature of the reactive acetaminophen metabolite has led to the development of effective antidotes—cysteamine and *N*-acetylcysteine. Administration of *N*-acetylcysteine (the safer of the two) within 8–16 hours after acetaminophen overdosage has been shown to protect victims from fulminant hepatotoxicity and death (see Chapter 58). Administration of GSH is not effective because it does not cross cell membranes readily.

CLINICAL RELEVANCE OF DRUG METABOLISM

The dose and frequency of administration required to achieve effective therapeutic blood and tissue levels vary in different patients because of individual differences in drug distribution and rates of drug metabolism and elimination. These differences are determined by genetic factors and nongenetic variables, such as age, sex, liver size, liver function, circadian rhythm, body temperature, and nutritional and environmental factors such as concomitant exposure to inducers or inhibitors of drug metabolism. The discussion that follows summarizes the most important of these variables.

Individual Differences

Individual differences in metabolic rate depend on the nature of the drug itself. Thus, within the same population, steady-state plasma levels may reflect a 30-fold variation in the metabolism of one drug and only a two-fold variation in the metabolism of another.

Genetic Factors

Genetic factors that influence enzyme levels account for some of these differences, giving rise to "genetic polymorphisms" in drug metabolism. The first examples of drugs found to be subject to genetic polymorphisms were the muscle relaxant succinylcholine, the anti-tuberculosis drug isoniazid, and the anticoagulant warfarin. A true genetic polymorphism is defined as the occurrence of a variant allele of a gene at a population frequency of ≥ 1%, resulting in altered expression or functional activity of the gene product, or both. Well-defined and clinically relevant genetic polymorphisms in both phase I and phase II drug-metabolizing enzymes exist that result in altered efficacy of drug therapy or adverse drug reactions (**ADRs**). The latter frequently necessitate dose adjustment (Table 4–4), a consideration particularly crucial for drugs with low therapeutic indices.

A. Phase I Enzyme Polymorphisms

Genetically determined defects in the phase I oxidative metabolism of several drugs have been reported (Table 4–4). These defects are often transmitted as autosomal recessive traits and may be expressed at any one of the multiple metabolic transformations that a chemical might undergo. Human liver P450s 3A4, 2C9, 2D6, 2C19, 1A2, and 2B6 are responsible for about 75% of all clinically relevant phase I drug metabolism (Figure 4–4), and thus for about 60% of all physiologic drug biotransformation and elimination. Thus, genetic polymorphisms of these enzymes, by significantly influencing phase I drug metabolism, can alter their

TABLE 4–4 Some examples of genetic polymorphisms in phase I and phase II drug metabolism.

Enzyme Involved	Defect	Genotype	Drug and Therapeutic Use	Clinical Consequences[1]
CYP1A2	*N*-Demethylation	**EM**	Caffeine (CNS stimulant)	Reduced CNS stimulation due to increased gene inducibility and thus increased metabolism/clearance in cigarette smokers and frequent ingesters of omeprazole.
	N-Demethylation	**PM**	Caffeine (CNS stimulant)	Enhanced CNS stimulation.
CYP2A6	Oxidation	**PM**	Nicotine (cholinoceptor stimulant)	Nicotine toxicity. Lesser craving for frequent cigarette smoking.
	Oxidation	**EM**	Nicotine (cholinoceptor stimulant)	Increased nicotine metabolism. Greater craving for frequent cigarette smoking.
	Oxidation	**PM**	Coumarin (anticoagulant)	Increased risk of bleeding.
	Oxidation	**EM**	Coumarin (anticoagulant)	Increased clearance. Greater risk of platelet aggregation and thrombosis.
CYP2B6	Oxidation, *N*-Dechloroethylation	**PM**	Cyclophosphamide, ifosfamide (anticancer)	Reduced clearance. Increased risk of ADRs.
	Oxidation	**PM**	Efavirenz (anti-HIV)	Reduced clearance. Increased risk of ADRs.
CYP2C8	Hydroxylation	**PM**	Repaglinide, rosiglitazone, pioglitazone (antidiabetic)	Reduced clearance. Increased risk of ADRs.
	Hydroxylation	**PM**	Paclitaxel (anticancer)	Reduced clearance. Increased risk of ADRs (myelosuppression).
	N-Deethylation/ *N*-Dealkylation	**PM**	Amodiaquine, chloroquine (antimalarial)	Reduced clearance. Increased risk of ADRs.
	N-Deethylation	**PM**	Amiodarone (antiarrhythmic)	Reduced clearance. Increased risk of ADRs.
CYP2C9	Hydroxylation	**PM**	Celecoxib, diclofenac, flurbiprofen, *S*-ibuprofen (NSAIDs)	Reduced clearance. Increased risk of ADRs.
	Hydroxylation	**PM**	*S*-Warfarin, *S*-acenocoumarol (anticoagulants)	Enhanced bleeding risk. Clinically highly relevant. Dose adjustment required.
	Hydroxylation	**PM**	Tolbutamide (antidiabetic)	Cardiotoxicity.
	Hydroxylation	**PM**	Phenytoin (antiepileptic)	Nystagmus, diplopia, and ataxia.
CYP2C19	*N*-Demethylation	**PM**	Amitriptyline, clomipramine (antidepressants)	Reduced clearance. Increased risk of ADRs. Dose adjustment required.
	Oxidation	**PM**	Moclobemide (MAOI)	
	N-Demethylation	**PM**	Citalopram (SSRI)	Increased risk of gastrointestinal side effects.
	O-Demethylation	**PM**	Omeprazole (PPI)	Increased therapeutic efficacy.
	Hydroxylation	**PM**	Mephenytoin (antiepileptic)	Overdose toxicity.
	N-Demethylation	**EM**	Escitalopram (antidepressants)	Increased gene transcription resulting in increased activity and thus reduced therapeutic efficacy.
	O-Demethylation	**EM**	Omeprazole (PPI)	Reduced therapeutic efficacy.
	Hydroxylation	**EM**	Tamoxifen (anticancer)	Increased metabolic activation, increased therapeutic efficacy; reduced risk of relapse. Dose adjustment required.
	Oxidative cyclization	**EM**	Chlorproguanil (antimalarial)	Increased metabolic activation, increased therapeutic efficacy. Dose adjustment required.
	Oxidation	**EM**	Clopidogrel (antiplatelet)	Increased metabolic activation, increased therapeutic efficacy. Dose adjustment required.
CYP2D6	Oxidation	**PM**	Bufuralol (β-adrenoceptor blocker)	Exacerbation of β blockade, nausea.
	O-Demethylation	**PM**	Codeine (analgesic)	Reduced metabolic activation to morphine and thus reduced analgesia.
	Oxidation	**PM**	Debrisoquin (antihypertensive)	Orthostatic hypotension.

(*continued*)

TABLE 4–4 Some examples of genetic polymorphisms in phase I and phase II drug metabolism. (Continued)

Enzyme Involved	Defect	Genotype	Drug and Therapeutic Use	Clinical Consequences[1]
	N-Demethylation	**PM**	Nortriptyline (antidepressant)	Reduced clearance. Increased risk of ADRs.
	Oxidation	**PM**	Sparteine	Oxytocic symptoms.
	O-Demethylation	**PM**	Dextromethorphan (antitussive)	Reduced clearance. Increased risk of ADRs.
	O-Demethylation	**PM**	Tramadol (analgesic)	Increased risk of seizures.
	Hydroxylation	**PM**	Tamoxifen (anticancer)	Reduced metabolic activation to the therapeutically active endoxifen and thus reduced therapeutic efficacy.
	O-Demethylation	**UM**	Codeine (analgesic)	Increased metabolic activation to morphine and thus increased risk of respiratory depression.
	N-Demethylation	**UM**	Nortriptyline (antidepressant)	Reduced therapeutic efficacy due to increased clearance.
	O-Demethylation	**UM**	Tramadol (analgesic)	Reduced therapeutic efficacy due to increased clearance.
CYP3A4		**PM?**	All drugs metabolized by this enzyme would be potentially affected	Reduced clearance. Dose adjustment may be required to avoid drug-drug interactions.
CYP3A5		**PM?**	Saquinavir, and other CYP3A substrates	Usually less catalytically active than CYP3A4. A higher frequency of a functional CYP3A5∗1 allele is seen in Africans than in Caucasians; the latter most often carry the defective CYP3A5∗3 allele. This may significantly affect therapeutics of CYP3A substrates in CYP3A5∗1 or CYP3A5∗3 homozygous individuals.
ALDH	Aldehyde dehydrogenation	**PM**	Ethanol (recreational drug)	Facial flushing, hypotension, tachycardia, nausea, vomiting.
BCHE	Ester hydrolysis	**PM**	Succinylcholine (muscle relaxant) Mivacurium (neuromuscular blocker) Cocaine (CNS stimulant)	Prolonged apnea. Prolonged muscle paralysis. Increased blood pressure, tachycardia, ventricular arrhythmias.
GST	GSH-conjugation	**PM**	Acetaminophen (analgesic), Busulfan (anticancer)	Impaired GSH conjugation due to gene deletion.
NAT2	*N*-Acetylation	**PM**	Hydralazine (antihypertensive)	Lupus erythematosus-like syndrome.
	N-Acetylation	**PM**	Isoniazid (antitubercular)	Peripheral neuropathy.
TPMT	*S*-Methylation	**PM**	6-Thiopurines (anticancer)	Myelotoxicity.
UGT1A1	Glucuronidation	**PM**	Bilirubin (heme metabolite) Irinotecan (anticancer)	Hyperbilirubinemia. Reduced clearance. Dose adjustment may be required to avoid toxicity (GI dysfunction, immunosuppression).

[1]Observed or predictable.

ADR, adverse drug reaction; EM, extensive metabolizer; PM, poor metabolizer; UM, ultrarapid metabolizer.

pharmacokinetics and the magnitude or the duration of drug response and associated events.

Three P450 genetic polymorphisms have been particularly well characterized, affording some insight into possible underlying molecular mechanisms, and are clinically noteworthy, as they require therapeutic dosage adjustment. The first is the **debrisoquin-sparteine oxidation** type of polymorphism, which apparently occurs in 3–10% of Caucasians and is inherited as an autosomal recessive trait. In affected individuals, the **CYP2D6**-dependent oxidations of debrisoquin and other drugs (Table 4–2; Figure 4–6) are impaired. These defects in oxidative drug metabolism are probably co-inherited. The precise molecular basis for the defect appears to be faulty expression of the P450 protein due to either defective mRNA splicing or protein folding, resulting in little or no isoform-catalyzed drug metabolism and thereby conferring a **poor metabolizer (PM)** phenotype. This PM phenotype correlates with a higher risk of relapse in patients with breast cancer treated with tamoxifen, an anti-cancer drug that relies on its CYP2D6-dependent metabolic activation to endoxifen for its efficacy. More recently, however, another polymorphic genotype has been reported that results in **ultrarapid metabolism** of relevant drugs due to the presence of CYP2D6 allelic variants with

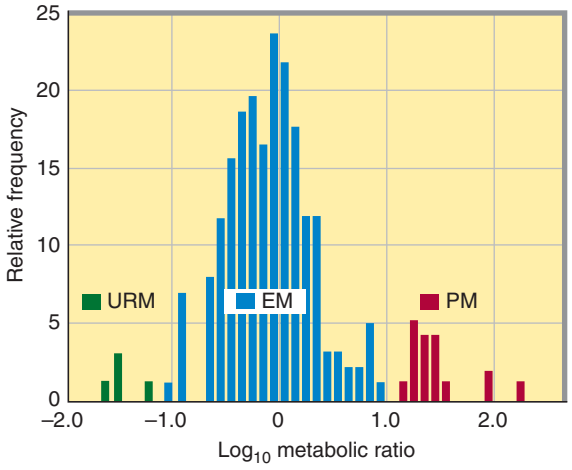


FIGURE 4–6 Genetic polymorphism in debrisoquin 4-hydroxylation by CYP2D6 in a Caucasian population. The semilog frequency distribution histogram of the metabolic ratio (MR; defined as percent of dose excreted as unchanged debrisoquin divided by the percent of dose excreted as 4-hydroxydebrisoquin metabolite) in the 8-hour urine collected after oral ingestion of 12.8 mg debrisoquin sulfate (equivalent to 10 mg free debrisoquin base). Individuals with MR values > 12.6 were phenotyped as poor metabolizers (PM, *red bars*), and those with MR values < 12.6 but > 0.2 were designated as extensive metabolizers (EM, *blue bars*). Those with MR values < 0.2 were designated as ultrarapid metabolizers (URM, *green bars*) based on the MR values (0.01–0.1) of individuals with documented multiple copies of CYP2D6 allelic variants resulting from inherited amplification of this gene. (Data from Woolhouse et al: Debrisoquin hydroxylation polymorphism among Ghanians and Caucasians. Clin Pharmacol Ther 1979;26:584.)

up to 13 gene copies in tandem. This ultrarapid metabolizer (**UM**) genotype is most common in Ethiopians and Saudi Arabians, populations that display it in up to one third of individuals. As a result, these subjects require two-fold to three-fold higher daily doses of nortriptyline (an antidepressant and a CYP2D6 substrate) to achieve therapeutic plasma levels. The poor responsiveness to antidepressant therapy of the UM phenotype also clinically correlates with a higher incidence of suicides relative to that of deaths due to natural causes in this patient population. Conversely, in these UM populations the prodrug codeine (another CYP2D6 substrate) is metabolized much faster to morphine, often resulting in undesirable adverse effects of morphine, such as abdominal pain. Indeed, intake of high doses of codeine by a mother of the ultrarapid metabolizer type was held responsible for the morphine-induced death of her breast-fed infant.

The second well-studied genetic drug polymorphism involves the stereoselective **aromatic (4)-hydroxylation** of the anticonvulsant mephenytoin, catalyzed by **CYP2C19.** This polymorphism, which is also inherited as an autosomal recessive trait, occurs in 3–5% of Caucasians and 18–23% of Japanese populations. It is genetically independent of the debrisoquin-sparteine polymorphism. In normal "**extensive metabolizers**" (EMs) (*S*)-mephenytoin is extensively hydroxylated by CYP2C19 at the 4 position of the phenyl ring before its glucuronidation and rapid excretion in the urine, whereas (*R*)-mephenytoin is slowly *N*-demethylated to nirvanol, an active metabolite. PMs however, appear to totally lack the stereospecific (*S*)-mephenytoin hydroxylase activity, so both (*S*)- and (*R*)-mephenytoin enantiomers are *N*-demethylated to nirvanol, which accumulates in much higher concentrations. Thus, PMs of mephenytoin show signs of profound sedation and ataxia after doses of the drug that are well tolerated by normal metabolizers. Two defective CYP2C19 variant alleles (CYP2C19*2 and CYP2C19*3), the latter predominant in Asians, are responsible for the PM genotype. The molecular bases include splicing defects resulting in a truncated, nonfunctional protein. CYP2C19 is responsible for the metabolism of various clinically relevant drugs (Table 4–4). Thus, it is clinically important to recognize that the safety of each of these drugs may be severely reduced in persons with the PM phenotype. On the other hand, the PM phenotype can notably increase the therapeutic efficacy of omeprazole, a proton-pump inhibitor, in gastric ulcer and gastroesophageal reflux diseases.

Another CYP2C19 variant allele (CYP2C19*17) exists that is associated with increased transcription and thus higher CYP2C19 expression and even higher functional activity than that of the wild type CYP2C19-carrying EMs. Individuals carrying this CYP2C19*17 allele exhibit higher metabolic activation of prodrugs such as the breast cancer drug tamoxifen, the antimalarial chlorproguanil, and the antiplatelet drug clopidogrel. The former event is associated with a lower risk of breast cancer relapse, and the latter event with an increased risk of bleeding. Carriers of the CYP2C19*17 allele are also known to enhance the metabolism and thus the elimination of drugs such as the antidepressants escitalopram and imipramine, as well as the antifungal voriconazole. This consequently impairs the therapeutic efficacy of these drugs, thus requiring clinical dosage adjustments.

The third relatively well-characterized genetic polymorphism is that of **CYP2C9.** Two well-characterized variants of this enzyme exist, each with amino acid mutations that result in altered metabolism. The CYP2C9*2 allele encodes an Arg144Cys mutation, exhibiting impaired functional interactions with **POR**. The other allelic variant, CYP2C9*3**,** encodes an enzyme with an Ile359Leu mutation that has lowered affinity for many substrates. For example, individuals displaying the CYP2C9*3 phenotype have greatly reduced tolerance for the anticoagulant warfarin. The warfarin clearance in CYP2C9*3-homozygous individuals is about 10% of normal values, and these people have a much lower tolerance for the drug than those who are homozygous for the normal wild type allele. These individuals also have a much higher risk of adverse effects with warfarin (eg, bleeding) and with other CYP2C9 substrates such as phenytoin, losartan, tolbutamide, and some nonsteroidal anti-inflammatory drugs (Table 4–4).

Allelic variants of CYP3A4 have also been reported, but their contribution to its well-known interindividual variability in drug metabolism apparently is limited. On the other hand, the expression of **CYP3A5,** another human liver isoform, is markedly polymorphic, ranging from 0% to 100% of the total hepatic CYP3A content. This CYP3A5 protein polymorphism is now known to result from a single nucleotide polymorphism (**SNP**) within intron 3, which enables normally spliced CYP3A5 transcripts in

5% of Caucasians, 29% of Japanese, 27% of Chinese, 30% of Koreans, and 73% of African Americans. Thus, it can significantly contribute to interindividual differences in the metabolism of preferential CYP3A5 substrates such as midazolam. Two other CYP3A5 allelic variants that result in a PM phenotype are also known.

Polymorphisms in the *CYP2A6* gene have also been recently characterized, and their prevalence is apparently racially linked. CYP2A6 is responsible for nicotine oxidation, and tobacco smokers with low CYP2A6 activity consume less and have a lower incidence of lung cancer. CYP2A6 1B allelic variants associated with faster rates of nicotine metabolism have been recently discovered. It remains to be determined whether patients with these faster variants will fall into the converse paradigm of increased smoking behavior and lung cancer incidence.

Additional genetic polymorphisms in drug metabolism (eg, **CYP2B6**) that are inherited independently from those already described are being discovered. For instance, a 20- to 250-fold variation in interindividual CYP2B6 expression partly due to genetic polymorphisms has been reported. This may have a significant impact on the metabolism of several clinically relevant drugs such as cyclophosphamide, methadone, efavirenz, selegiline, and propofol. Studies of theophylline metabolism in monozygotic and dizygotic twins that included pedigree analysis of various families have revealed that a distinct polymorphism may exist for this drug and may be inherited as a recessive genetic trait. Genetic drug metabolism polymorphisms also appear to occur for aminopyrine and carbocysteine oxidations. Regularly updated information on human P450 polymorphisms is available at http://www.imm.ki.se/CYPalleles/.

Although genetic polymorphisms in drug oxidations often involve specific P450 enzymes, such genetic variations can also occur in other enzymes. Recently, genetic polymorphisms in POR, the essential P450 electron donor, have been reported. In particular, an allelic variant (at a 28% frequency) encoding a POR A503V mutation has been reported to result in impaired CYP17-dependent sex steroid synthesis and impaired CYP3A4- and CYP2D6-dependent drug metabolism in vitro. Its involvement in clinically relevant drug metabolism, while predictable, remains to be established. Descriptions of a polymorphism in the oxidation of trimethylamine, believed to be metabolized largely by the **flavin monooxygenase (Ziegler's enzyme),** result in the "fish-odor syndrome" in slow metabolizers, thus suggesting that genetic variants of other non–P450-dependent oxidative enzymes may also contribute to such polymorphisms.

B. Phase II Enzyme Polymorphisms

Succinylcholine is metabolized only half as rapidly in persons with genetically determined deficiency in pseudocholinesterase (now generally referred to as butyrylcholinesterase [**BCHE**]) as in persons with normally functioning enzyme. Different mutations, inherited as autosomal recessive traits, account for the enzyme deficiency. Deficient individuals treated with succinylcholine as a surgical muscle relaxant may become susceptible to prolonged respiratory paralysis (succinylcholine apnea). Similar pharmacogenetic differences are seen in the acetylation of isoniazid. The defect in slow acetylators (of isoniazid and similar amines) appears to be caused by the synthesis of less of the NAT2 enzyme rather than of an abnormal form of it. Inherited as an autosomal recessive trait, the **slow acetylator phenotype** occurs in about 50% of blacks and whites in the USA, more frequently in Europeans living in high northern latitudes, and much less commonly in Asians and Inuits (Eskimos). The slow acetylator phenotype is also associated with a higher incidence of isoniazid-induced peripheral neuritis, drug-induced autoimmune disorders, and bicyclic aromatic amine-induced bladder cancer.

A clinically important polymorphism of the *TPMT* (thiopurine *S*-methyltransferase) gene is encountered in Europeans (frequency, 1:300), resulting in a rapidly degraded mutant enzyme and consequently deficient *S*-methylation of aromatic and heterocyclic sulfhydryl compounds including the anti-cancer thiopurine drugs 6-mercaptopurine, thioguanine, and azathioprine, required for their detoxification. Patients inheriting this polymorphism as an autosomal recessive trait are at high risk of thiopurine drug-induced fatal hematopoietic toxicity.

Genetic polymorphisms in the expression of other phase II enzymes (UGTs and GSTs) also occur. Thus, UGT polymorphisms (*UGT1A1*28*) are associated with hyperbilirubinemic diseases (Gilbert's syndrome) as well as toxic side effects due to impaired drug conjugation and/or elimination (eg, the anticancer drug irinotecan). Similarly, genetic polymorphisms (*GSTM1*) in GST (mu1 isoform) expression can lead to significant adverse effects and toxicities of drugs dependent on its GSH conjugation for elimination.

C. The Role of Pharmacogenetic Testing in Clinically Safe & Effective Drug Therapy

Despite our improved understanding of the molecular basis of pharmacogenetic defects in drug-metabolizing enzymes, their impact on drug therapy and ADRs, and the availability of validated pharmacogenetic biomarkers to identify patients at risk, this clinically relevant information has not been effectively translated to patient care. Thus, the much-heralded potential for personalized medicine, except in a few instances of drugs with a relatively low therapeutic index (eg, warfarin), has remained largely unrealized. This is so even though 98% of US physicians are apparently aware that such genetic information may significantly influence therapy. This is partly due to the lack of adequate training in translating this knowledge to medical practice, and partly due to the logistics of genetic testing and the issue of cost-effectiveness. Severe ADRs are known to contribute to 100,000 annual US deaths, about 7% of all hospital admissions, and an increased average length of hospital stay. Genotype information could greatly enhance safe and efficacious clinical therapy through dose adjustment or alternative drug therapy, thereby curbing much of the rising ADR incidence and its associated costs.

Diet & Environmental Factors

Diet and environmental factors contribute to individual variations in drug metabolism. Charcoal-broiled foods and cruciferous vegetables

are known to induce CYP1A enzymes, whereas grapefruit juice is known to inhibit the CYP3A metabolism of co-administered drug substrates (Table 4–2). Cigarette smokers metabolize some drugs more rapidly than nonsmokers because of enzyme induction (see previous section). Industrial workers exposed to some pesticides metabolize certain drugs more rapidly than unexposed individuals. Such differences make it difficult to determine effective and safe doses of drugs that have narrow therapeutic indices.

Age & Sex

Increased susceptibility to the pharmacologic or toxic activity of drugs has been reported in very young and very old patients compared with young adults (see Chapters 59 and 60). Although this may reflect differences in absorption, distribution, and elimination, differences in drug metabolism also play a role. Slower metabolism could be due to reduced activity of metabolic enzymes or reduced availability of essential endogenous cofactors.

Sex-dependent variations in drug metabolism have been well documented in rats but not in other rodents. Young adult male rats metabolize drugs much faster than mature female rats or prepubertal male rats. These differences in drug metabolism have been clearly associated with androgenic hormones. Clinical reports suggest that similar sex-dependent differences in drug metabolism also exist in humans for ethanol, propranolol, some benzodiazepines, estrogens, and salicylates.

Drug-Drug Interactions during Metabolism

Many substrates, by virtue of their relatively high lipophilicity, are not only retained at the active site of the enzyme but remain nonspecifically bound to the lipid endoplasmic reticulum membrane. In this state, they may induce microsomal enzymes, particularly after repeated use. Acutely, depending on the residual drug levels at the active site, they also may competitively inhibit metabolism of a simultaneously administered drug.

Enzyme-inducing drugs include various sedative-hypnotics, antipsychotics, anticonvulsants, the antitubercular drug rifampin, and insecticides (Table 4–5). Patients who routinely ingest barbiturates, other sedative-hypnotics, or certain antipsychotic drugs may require considerably higher doses of warfarin to maintain a therapeutic effect. On the other hand, discontinuance of the sedative inducer may result in reduced metabolism of the anticoagulant and bleeding—a toxic effect of the ensuing enhanced plasma levels of the anticoagulant. Similar interactions have been observed in individuals receiving various combinations of drug regimens such as rifampin, antipsychotics, or sedatives with contraceptive agents, sedatives with anticonvulsant drugs, and even alcohol with hypoglycemic drugs (tolbutamide).

It must also be noted that an inducer may enhance not only the metabolism of other drugs but also its own metabolism. Thus, continued use of some drugs may result in a pharmacokinetic type of **tolerance**—progressively reduced therapeutic effectiveness due to enhancement of their own metabolism.

TABLE 4–5 Partial list of drugs that enhance drug metabolism in humans.

Inducer	Drugs Whose Metabolism Is Enhanced
Benzo[*a*]pyrene	Theophylline
Carbamazepine	Carbamazepine, clonazepam, itraconazole
Chlorcyclizine	Steroid hormones
Ethchlorvynol	Warfarin
Glutethimide	Antipyrine, glutethimide, warfarin
Griseofulvin	Warfarin
Phenobarbital and other barbiturates[1]	Barbiturates, chloramphenicol, chlorpromazine, cortisol, coumarin anticoagulants, desmethylimipramine, digitoxin, doxorubicin, estradiol, itraconazole, phenylbutazone, phenytoin, quinine, testosterone
Phenylbutazone	Aminopyrine, cortisol, digitoxin
Phenytoin	Cortisol, dexamethasone, digitoxin, itraconazole, theophylline
Rifampin	Coumarin anticoagulants, digitoxin, glucocorticoids, itraconazole, methadone, metoprolol, oral contraceptives, prednisone, propranolol, quinidine, saquinavir
Ritonavir[2]	Midazolam
St. John's wort	Alprazolam, cyclosporine, digoxin, indinavir, oral contraceptives, ritonavir, simvastatin, tacrolimus, warfarin

[1]Secobarbital is an exception. See Table 4–6 and text.

[2]With chronic (repeated) administration; acutely, ritonavir is a potent CYP3A4 inhibitor/inactivator.

Conversely, simultaneous administration of two or more drugs may result in impaired elimination of the more slowly metabolized drug and prolongation or potentiation of its pharmacologic effects (Table 4–6). Both competitive substrate inhibition and irreversible substrate-mediated enzyme inactivation may augment plasma drug levels and lead to toxic effects from drugs with narrow therapeutic indices. Indeed, such acute interactions of terfenadine (a second-generation antihistamine) with a CYP3A4 substrate-inhibitor (ketoconazole, erythromycin, or grapefruit juice) resulted in fatal cardiac arrhythmias (torsades de pointe) requiring its withdrawal from the market. Similar drug-drug interactions with CYP3A4 substrate-inhibitors (such as the antibiotics erythromycin and clarithromycin, the antidepressant nefazodone, the antifungals itraconazole and ketoconazole, and the HIV protease inhibitors indinavir and ritonavir), and consequent cardiotoxicity led to withdrawal or restricted use of the 5-HT_4 agonist, cisapride. Similarly, allopurinol both prolongs the duration and enhances the chemotherapeutic and toxic actions of mercaptopurine by competitive inhibition of xanthine oxidase. Consequently, to avoid bone marrow toxicity, the dose of mercaptopurine must be reduced in patients receiving allopurinol. Cimetidine, a drug used in the treatment of peptic ulcer, has been shown to potentiate the pharmacologic actions of anticoagulants and sedatives. The metabolism of the sedative chlordiazepoxide has been shown to be

TABLE 4–6 Partial list of drugs that inhibit drug metabolism in humans.

Inhibitor[1]	Drug Whose Metabolism Is Inhibited
Allopurinol, chloramphenicol, isoniazid	Antipyrine, dicumarol, probenecid, tolbutamide
Chlorpromazine	Propranolol
Cimetidine	Chlordiazepoxide, diazepam, warfarin, others
Dicumarol	Phenytoin
Diethylpentenamide	Diethylpentenamide
Disulfiram	Antipyrine, ethanol, phenytoin, warfarin
Ethanol	Chlordiazepoxide (?), diazepam (?), methanol
Grapefruit juice[2]	Alprazolam, atorvastatin, cisapride, cyclosporine, midazolam, triazolam
Itraconazole	Alfentanil, alprazolam, astemizole, atorvastatin, buspirone, cisapride, cyclosporine, delavirdine, diazepam, digoxin, felodipine, indinavir, loratadine, lovastatin, midazolam, nisoldipine, phenytoin, quinidine, ritonavir, saquinavir, sildenafil, simvastatin, sirolimus, tacrolimus, triazolam, verapamil, warfarin
Ketoconazole	Astemizole, cyclosporine, terfenadine
Nortriptyline	Antipyrine
Oral contraceptives	Antipyrine
Phenylbutazone	Phenytoin, tolbutamide
Ritonavir	Amiodarone, cisapride, itraconazole, midazolam, triazolam
Saquinavir	Cisapride, ergot derivatives, midazolam, triazolam
Secobarbital	Secobarbital
Spironolactone	Digoxin
Troleandomycin	Theophylline, methylprednisolone

[1]While some inhibitors are selective for a given P450 enzyme, others are more general and can inhibit several P450s concurrently.

[2]Active components in grapefruit juice include furanocoumarins such as 6′, 7′-dihydroxybergamottin (which inactivates both intestinal and liver CYP3A4) as well as other unknown components that inhibit P-glycoprotein-mediated intestinal drug efflux and consequently further enhance the bioavailability of certain drugs such as cyclosporine.

inhibited by 63% after a single dose of cimetidine; such effects are reversed within 48 hours after withdrawal of cimetidine.

Impaired metabolism may also result if a simultaneously administered drug irreversibly inactivates a common metabolizing enzyme. These inhibitors, in the course of their metabolism by cytochrome P450, inactivate the enzyme and result in impairment of their own metabolism and that of other cosubstrates. This is indeed the case of the furanocoumarins in grapefruit juice that inactivate CYP3A4 in the intestinal mucosa and consequently enhance its proteolytic degradation. This impairment of their intestinal first-pass CYP3A4-dependent metabolism significantly enhances the bioavailability of drugs, such as felodipine, nifedipine, terfenadine, verapamil, ethinylestradiol, saquinavir, and cyclosporine A, and is associated with clinically relevant drug-drug and food-drug interactions.

Recovery from these interactions is dependent on CYP3A4 resynthesis and thus may be slow.

Interactions between Drugs & Endogenous Compounds

Some drugs require conjugation with endogenous substrates such as GSH, glucuronic acid, or sulfate for their inactivation. Consequently, different drugs may compete for the same endogenous substrates, and the faster-reacting drug may effectively deplete endogenous substrate levels and impair the metabolism of the slower-reacting drug. If the latter has a steep dose-response curve or a narrow margin of safety, potentiation of its therapeutic and toxic effects may result.

Diseases Affecting Drug Metabolism

Acute or chronic diseases that affect liver architecture or function markedly affect hepatic metabolism of some drugs. Such conditions include alcoholic hepatitis, active or inactive alcoholic cirrhosis, hemochromatosis, chronic active hepatitis, biliary cirrhosis, and acute viral or drug-induced hepatitis. Depending on their severity, these conditions may significantly impair hepatic drug-metabolizing enzymes, particularly microsomal oxidases, and thereby markedly affect drug elimination. For example, the half-lives of chlordiazepoxide and diazepam in patients with liver cirrhosis or acute viral hepatitis are greatly increased, with a corresponding increase in their effects. Consequently, these drugs may cause coma in patients with liver disease when given in ordinary doses.

Some drugs are metabolized so readily that even marked reduction in liver function does not significantly prolong their action. However, cardiac disease, by limiting blood flow to the liver, may impair disposition of those drugs whose metabolism is flow-limited (Table 4–7). These drugs are so readily metabolized by the liver that hepatic clearance is essentially equal to liver blood

TABLE 4–7 Rapidly metabolized drugs whose hepatic clearance is blood flow-limited.

Alprenolol	Lidocaine
Amitriptyline	Meperidine
Clomethiazole	Morphine
Desipramine	Pentazocine
Imipramine	Propoxyphene
Isoniazid	Propranolol
Labetalol	Verapamil

flow. Pulmonary disease may also affect drug metabolism, as indicated by the impaired hydrolysis of procainamide and procaine in patients with chronic respiratory insufficiency and the increased half-life of antipyrine (a P450 functional probe) in patients with lung cancer. The impaired enzyme activity or defective formation of enzymes associated with heavy metal poisoning or porphyria also results in reduced hepatic drug metabolism.

Although the effects of endocrine dysfunction on drug metabolism have been well explored in experimental animal models, corresponding data for humans with endocrine disorders are scanty. Thyroid dysfunction has been associated with altered metabolism of some drugs and of some endogenous compounds as well. Hypothyroidism increases the half-life of antipyrine, digoxin, methimazole, and some β blockers, whereas hyperthyroidism has the opposite effect. A few clinical studies in diabetic patients indicate no apparent impairment of drug metabolism, although impairment has been noted in diabetic rats. Malfunctions of the pituitary, adrenal cortex, and gonads markedly reduce hepatic drug metabolism in rats. On the basis of these findings, it may be supposed that such disorders could significantly affect drug metabolism in humans. However, until sufficient evidence is obtained from clinical studies in patients, such extrapolations must be considered tentative.

Finally, the release of inflammatory mediators, cytokines, and nitric oxide associated with bacterial or viral infections, cancer, or inflammation are known to impair drug metabolism by inactivating P450s and enhancing their degradation.

REFERENCES

Benowitz NL: Pharmacology of nicotine: Addiction, smoking-induced disease, and therapeutics. Annu Rev Pharmacol Toxicol 2009;49:57.

Correia MA: Human and rat liver cytochromes P450: Functional markers, diagnostic inhibitor probes and parameters frequently used in P450 studies. In: Ortiz de Montellano P (editor). *Cytochrome P450: Structure, Mechanism and Biochemistry,* 3rd ed. Kluwer-Academic/Plenum Press, 2005.

Correia MA, Ortiz de Montellano P: Inhibitors of cytochrome P-450 and possibilities for their therapeutic application. In: Ruckpaul K (editor): *Frontiers in Biotransformation*, vol 8. Taylor & Francis, 1993.

Correia MA, Ortiz de Montellano P: Inhibition of cytochrome P450 enzymes. In: Ortiz de Montellano P (editor). Cytochrome P450: Structure, *Mechanism and Biochemistry,* 3rd ed. KluwerAcademic/Plenum Press, 2005.

Daly AK: Pharmacogenetics and human genetic polymorphisms. Biochem J 2010;429:435.

Guengerich FP: Human cytochrome P450 enzymes. In: Ortiz de Montellano P (editor). *Cytochrome P450: Structure, Mechanism and Biochemistry*, 3rd ed. Kluwer-Academic/Plenum Press, 2005.

Guengerich FP: Role of cytochrome P450 enzymes in drug-drug interactions. Adv Pharmacol 1997;43:7.

Hustert E et al: The genetic determinants of the CYP3A5 polymorphism. Pharmacogenetics 2001;11:773.

Ingelman-Sundberg M et al: Influence of cytochrome P450 polymorphisms on drug therapies: Pharmacogenetic, pharmacoepigenetic and clinical aspects. Pharmacol Ther 2007;116:496.

Ingelman-Sundberg M, Sim SC: Pharmacogenetic biomarkers as tools for improved drug therapy; emphasis on the cytochrome P450 system. Biochem Biophys Res Commun 2010;396:90.

Ingelman-Sundberg M: Pharmacogenetics: An opportunity for a safer and more efficient pharmacotherapy. J Intern Med 2001;250:186.

Kroemer HK, Klotz U: Glucuronidation of drugs: A reevaluation of the pharmacological significance of the conjugates and modulating factors. Clin Pharmacokinet 1992;23:292.

Kuehl P et al: Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. Nat Genet 2001;27:383.

Lown KS et al: Grapefruit juice increases felodipine oral availability in humans by decreasing intestinal CYP3A protein expression. J Clin Invest 1997;99:2545.

Meyer UA: Pharmacogenetics—Five decades of therapeutic lessons from genetic diversity. Nat Rev Genet 2004;5:669.

Morgan ET et al: Regulation of drug-metabolizing enzymes and transporters in infection, inflammation, and cancer. Drug Metab Dispos 2008;36:205.

Nelson DR et al: The P450 superfamily: Update on new sequences, gene mapping, accession numbers, and nomenclature. Pharmacogenetics 1996;6:1.

Nelson DR et al: Updated human P450 sequences available at http://drnelson.utmem.edu/human.P450.seqs.html.

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

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

Wang L, McLeod HL, Weinshilboum RM: Genomics and drug response. N Engl J Med 2011;364:1144.

Williams SN et al: Induction of cytochrome P450 enzymes. In: Ortiz de Montellano PR (editor). *Cytochrome P450. Structure, Mechanism, and Biochemistry.* Plenum Press, 2005; *and references therein.*

Willson TM, Kliewer SA: PXR, CAR and drug metabolism. Nat Rev Drug Discov 2002;1:259.

Xu C et al: CYP2A6 genetic variation and potential consequences. Adv Drug Delivery Rev 2002;54:1245.

CHAPTER 5

Development & Regulation of Drugs

Bertram G. Katzung, MD, PhD*

A few useful drugs have been known since humans first began ingesting or injecting substances and recording the results (see The History of Pharmacology in Chapter 1), but the majority of agents in current use have been developed during the last 100 years using a variety of pharmacologic and toxicologic techniques. These new chemicals and the efforts to market them have in turn led to a variety of methods of legal regulation. This chapter describes the methods of new drug development and some aspects of drug regulation in the United States.

The most common first steps in the development of a new drug are the discovery or synthesis of a potential new drug compound or the elucidation of a new drug target. When a new drug molecule is synthesized or discovered, subsequent steps seek an understanding of the drug's interactions with its biologic targets. Repeated application of this approach leads to compounds with increased efficacy, potency, and selectivity (Figure 5–1). In the United States, the safety and efficacy of drugs must be defined before marketing can be legally carried out. In addition to in vitro studies, relevant biologic effects, drug metabolism, pharmacokinetic profiles, and particularly an assessment of the relative safety of the drug must be characterized in vivo in animals before human drug trials can be started. With regulatory approval, human testing may then go forward (usually in three phases) before the drug is considered for approval for general use. A fourth phase of data gathering and safety monitoring is becoming increasingly important and follows after approval for marketing. Once approved, the great majority of drugs become available for use by any appropriately licensed practitioner. Highly toxic drugs that are nevertheless considered valuable in lethal diseases may be approved for restricted use by practitioners who have undergone special training in their use and who maintain detailed records.

*The author thanks Barry A. Berkowitz, PhD, a previous author of this chapter, for his contributions.

THE PHARMACEUTICAL INDUSTRY

Careful analysis indicates that a majority of new drugs *originate* in research carried out in public sector institutions (universities, research institutes). However, because of the economic investment required and the need to efficiently access and integrate multiple technologies, most new drugs are then *developed* in pharmaceutical companies. Enormous and increasing costs, with estimates from $150 million to several billion, are involved in the development of a single new drug that reaches the marketplace. Only 2 of 10 marketed drugs return their research and development (R&D) investments, thus providing considerable motivation to develop "blockbuster drugs." Thousands of compounds may be synthesized and hundreds of thousands tested from libraries of compounds for each successful new drug lead, which then frequently needs to be further optimized for reasons of potency, selectivity, drug metabolism, and dosing convenience before each drug reaches the market. Increasing regulatory challenges and litigation resulting from real or suspected drug toxicity after approval further increase the cost of developing new drugs. Unfortunately, only 10–15% of the new drugs that achieve marketing approval represent significant advances in safety and effectiveness; the remainder are merely molecular variants ("me-too drugs") on true breakthrough drugs.

In spite of the cost of development, the financial rewards in drug development can be enormous. The global market for pharmaceuticals in 2007 was estimated to be $712 billion and the return on investment in the pharmaceutical industry is among the highest of all industries. This is ensured by setting the price of a new, important drug very high and lowering the price only when competition forces it down; for example, when me-too variants or generic versions of the original molecule become available. Even in Europe, where drug prices are lower than in the USA, industry profits are comparable. The 2007 worldwide sales of the top-selling drug (Lipitor) exceeded $12 billion. In the USA, approximately 10–12% of the health care dollar is presently spent on prescription drugs. At the same time, the investment in drugs can have

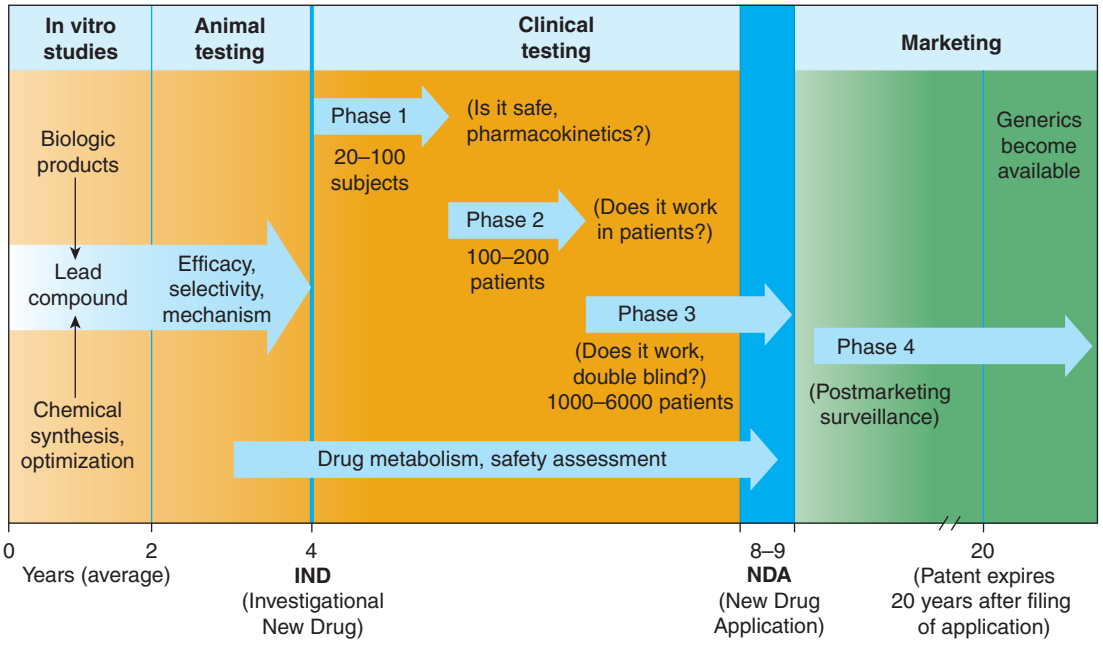


FIGURE 5–1 The development and testing process required to bring a drug to market in the USA. Some of the requirements may be different for drugs used in life-threatening diseases (see text).

enormous health benefits—new drugs can reduce suffering and save lives.

DRUG DISCOVERY

Most new drugs or drug products are discovered or developed through the following approaches: (1) identification or elucidation of a new drug target; (2) rational design of a new molecule based on an understanding of biologic mechanisms and drug receptor structure; (3) screening for biologic activity of large numbers of natural products, banks of previously discovered chemical entities, or large libraries of peptides, nucleic acids, and other organic molecules; and (4) chemical modification of a known active molecule, resulting in a me-too analog. Steps (1) and (2) are often carried out in academic research laboratories, but the costs of steps (3) and (4) usually ensure that industry carries them out.

Once a new drug target or promising molecule has been identified, the process of moving from the basic science laboratory to the clinic begins. This **translational research** involves the preclinical and clinical steps described next.

Drug Screening

Regardless of the source or the key idea leading to a drug candidate molecule, testing it involves a sequence of experimentation and characterization called drug screening. A variety of assays at the molecular, cellular, organ system, and whole animal levels are used to define the activity and selectivity of the drug. The type and number of initial screening tests depend on the pharmacologic and therapeutic goal. For example, anti-infective drugs may be tested against a variety of infectious organisms, some of which are resistant to standard agents; hypoglycemic drugs may be tested for their ability to lower blood sugar, etc.

The molecule will also be studied for a broad array of other actions to determine the mechanism of action and selectivity of the drug. This can reveal both expected and unexpected toxic effects. Occasionally, an unexpected therapeutic action is serendipitously discovered by a careful observer. The selection of compounds for development is most efficiently conducted in animal models of human disease. Where good predictive preclinical models exist (eg, antibacterials, hypertension, or thrombotic disease), we generally have good or excellent drugs. Good drugs or breakthrough improvements are conspicuously lacking and slow for diseases for which preclinical models are poor or not yet available, eg, autism and Alzheimer's disease.

Studies are performed during drug screening to define the **pharmacologic profile** of the drug at the molecular, cellular, organ, system, and organism levels. The value of these tests is highly dependent on the reproducibility and reliability of the assays. For example, a broad range of tests would be performed on a drug designed to act as an antagonist for a new vascular target for the treatment of hypertension.

At the molecular level, the compound would be screened for activity on the target, for example, receptor binding affinity to cell membranes containing the homologous animal receptors (or if possible, on the cloned human receptors). Early studies would be done to predict effects that might later cause undesired drug metabolism or toxicologic complications. For example, studies on liver cytochrome P450 enzymes would be performed to determine whether the molecule of interest is likely to be a substrate or inhibitor of these enzymes or to interfere with the metabolism of other drugs. Effects on cardiac ion channels such as the hERG

potassium channel, possibly predictive of life-threatening arrhythmias, are considered.

Effects on cell function determine whether the drug is an agonist, partial agonist, inverse agonist, or antagonist at the relevant receptors. Isolated tissues, especially vascular smooth muscle, would be used to characterize the pharmacologic activity and selectivity of the new compound in comparison with reference compounds. Comparison with other drugs would also be undertaken in other in vitro preparations such as gastrointestinal and bronchial smooth muscle. At each step in this process, the compound would have to meet specific performance and selectivity criteria to be carried further.

Whole animal studies are generally necessary to determine the effect of the drug on organ systems and disease models. Cardiovascular and renal function studies of new drugs are generally first performed in normal animals. Studies on disease models, if available, are then performed. For a candidate antihypertensive drug, animals with hypertension would be treated to see whether blood pressure was lowered in a dose-related manner and to characterize other effects of the compound. Evidence would be collected on duration of action and efficacy after oral and parenteral administration. If the agent possessed useful activity, it would be further studied for possible adverse effects on other major organs, including the respiratory, gastrointestinal, endocrine, and central nervous systems.

These studies might suggest the need for further chemical modification (compound optimization) to achieve more desirable pharmacokinetic or pharmacodynamic properties. For example, oral administration studies might show that the drug was poorly absorbed or rapidly metabolized in the liver; modification to improve bioavailability might be indicated. If the drug was to be administered long term, an assessment of tolerance development would be made. For drugs related to or having mechanisms of action similar to those known to cause physical or psychological dependence, abuse potential would also be studied. Drug interactions would be examined.

The desired result of this screening procedure (which may have to be repeated several times with analogs or congeners of the original molecule) is a **lead compound**, ie, a leading candidate for a successful new drug. A patent application would be filed for a novel compound (a composition of matter patent) that is efficacious, or for a new and nonobvious therapeutic use (a use patent) for a previously known chemical entity.

PRECLINICAL SAFETY & TOXICITY TESTING

All drugs are toxic in some individuals at some dose. Seeking to correctly define the limiting toxicities of drugs and the therapeutic index comparing benefits and risks of a new drug is an essential part of the new drug development process. Most drug candidates fail to reach the market, but the art of drug development consists of effective assessment and management of risk versus benefit and not total risk avoidance.

Candidate drugs that survive the initial screening procedures must be carefully evaluated for potential risks before and during clinical testing. Depending on the proposed use of the drug, preclinical toxicity testing includes most or all of the procedures shown in Table 5–1. Although no chemical can be certified as completely "safe" (free of risk), the objective is to estimate the risk associated with exposure to the drug candidate and to consider this in the context of therapeutic needs and likely duration of drug use.

The goals of preclinical toxicity studies include identifying potential human toxicities, designing tests to further define the toxic mechanisms, and predicting the most relevant toxicities to be monitored in clinical trials. In addition to the studies shown in Table 5–1, several quantitative estimates are desirable. These include the **no-effect dose**—the maximum dose at which a specified toxic effect is not seen; the **minimum lethal dose**—the smallest dose that is observed to kill any experimental animal; and, if necessary, the **median lethal dose (LD_{50})**—the dose that kills

TABLE 5–1 Safety tests.

Type of Test	Approach and Goals
Acute toxicity	Usually two species, two routes. Determine the no-effect dose and the maximum tolerated dose. In some cases, determine the acute dose that is lethal in approximately 50% of animals.
Subacute or subchronic toxicity	Three doses, two species. Two weeks to 3 months of testing may be required before clinical trials. The longer the duration of expected clinical use, the longer the subacute test. Determine biochemical, physiologic effects.
Chronic toxicity	Rodent and at least one nonrodent species for ≥ 6 months. Required when drug is intended to be used in humans for prolonged periods. Usually run concurrently with clinical trials. Determine same end points as subacute toxicity tests.
Effect on reproductive performance	Two species, usually one rodent and rabbits. Test effects on animal mating behavior, reproduction, parturition, progeny, birth defects, postnatal development.
Carcinogenic potential	Two years, two species. Required when drug is intended to be used in humans for prolonged periods. Determine gross and histologic pathology.
Mutagenic potential	Test effects on genetic stability and mutations in bacteria (Ames test) or mammalian cells in culture; dominant lethal test and clastogenicity in mice.

approximately 50% of the animals. Presently, the LD_{50} is estimated from the smallest number of animals possible. These doses are used to calculate the initial dose to be tried in humans, usually taken as one hundredth to one tenth of the no-effect dose in animals.

It is important to recognize the limitations of preclinical testing. These include the following:

1. Toxicity testing is time-consuming and expensive. Two to 6 years may be required to collect and analyze data on toxicity before the drug can be considered ready for testing in humans.
2. Large numbers of animals may be needed to obtain valid preclinical data. Scientists are properly concerned about this situation, and progress has been made toward reducing the numbers required while still obtaining valid data. Cell and tissue culture in vitro methods and computer modeling are increasingly being used, but their predictive value is still limited. Nevertheless, some segments of the public attempt to halt all animal testing in the unfounded belief that it has become unnecessary.
3. Extrapolations of therapeutic index and toxicity data from animals to humans are reasonably predictive for many but not for all toxicities. Seeking an improved process, a Predictive Safety Testing Consortium of five of America's largest pharmaceutical companies with an advisory role by the Food and Drug Administration (FDA) has been formed to share internally developed laboratory methods to predict the safety of new treatments before they are tested in humans. In 2007, this group presented to the FDA a list of biomarkers for early kidney damage.
4. For statistical reasons, rare adverse effects are unlikely to be detected in preclinical testing.

EVALUATION IN HUMANS

Less than one third of the drugs tested in clinical trials reach the marketplace. Federal law in the USA and ethical considerations require that the study of new drugs in humans be conducted in accordance with stringent guidelines. Scientifically valid results are not guaranteed simply by conforming to government regulations, however, and the design and execution of a good clinical trial require interdisciplinary personnel including basic scientists, clinical pharmacologists, clinician specialists, statisticians, and others. The need for careful design and execution is based on three major confounding factors inherent in the study of any drug in humans.

Confounding Factors in Clinical Trials

A. The Variable Natural History of Most Diseases

Many diseases tend to wax and wane in severity; some disappear spontaneously, even, on occasion, cancer. A good experimental design takes into account the natural history of the disease by evaluating a large enough population of subjects over a sufficient period of time. Further protection against errors of interpretation caused by disease fluctuations is sometimes provided by using a **crossover design,** which consists of alternating periods of administration of test drug, placebo preparation (the control), and the standard treatment (positive control), if any, in each subject. These sequences are systematically varied, so that different subsets of patients receive each of the possible sequences of treatment.

B. The Presence of Other Diseases and Risk Factors

Known and unknown diseases and risk factors (including lifestyles of subjects) may influence the results of a clinical study. For example, some diseases alter the pharmacokinetics of drugs (see Chapters 3 and 4). Other drugs and some foods alter the pharmacokinetics of many drugs. Concentrations of blood or tissue components being monitored as a measure of the effect of the new agent may be influenced by other diseases or other drugs. Attempts to avoid this hazard usually involve the crossover technique (when feasible) and proper selection and assignment of patients to each of the study groups. This requires obtaining accurate diagnostic tests, medical and pharmacologic histories (including use of recreational drugs), and the use of statistically valid methods of randomization in assigning subjects to particular study groups. There is growing interest in analyzing genetic variations as part of the trial that may influence whether a person responds to a particular drug. It has been shown that age, gender, and pregnancy influence the pharmacokinetics of some drugs, but these factors have not been adequately studied because of legal restrictions and reluctance to expose these populations to unknown risks.

C. Subject and Observer Bias and Other Factors

Most patients tend to respond in a positive way to any therapeutic intervention by interested, caring, and enthusiastic medical personnel. The manifestation of this phenomenon in the subject is the **placebo response** (Latin, "I shall please") and may involve objective physiologic and biochemical changes as well as changes in subjective complaints associated with the disease. The placebo response is usually quantitated by administration of an inert material with exactly the same physical appearance, odor, consistency, etc, as the active dosage form. The magnitude of the response varies considerably from patient to patient and may also be influenced by the duration of the study. In some conditions, a positive response may be noted in as many as 30–40% of subjects given placebo. Placebo adverse effects and "toxicity" also occur but usually involve subjective effects: stomach upset, insomnia, sedation, and so on.

Subject bias effects can be quantitated—and minimized relative to the response measured during active therapy—by the **single-blind** design. This involves use of a placebo as described above, administered to the same subjects in a crossover design, if possible, or to a separate control group of well-matched subjects. Observer bias can be taken into account by disguising the identity of the medication being used—placebo or active form—from both the subjects and the personnel evaluating the subjects' responses (**double-blind** design). In this design, a third party holds the code identifying each medication packet, and the code is not broken until all the clinical data have been collected.

Drug effects seen in clinical trials are obviously affected by the patient taking the drugs at the dose and frequency prescribed. In a recent phase 2 study, one third of the patients who said they were taking the drug were found by blood analysis to have not taken the

Drug Studies—The Types of Evidence*

As described in this chapter, drugs are studied in a variety of ways, from 30-minute test tube experiments with isolated enzymes and receptors to decades-long observations of populations of patients. The conclusions that can be drawn from such different types of studies can be summarized as follows.

Basic research is designed to answer specific, usually single, questions under tightly controlled laboratory conditions, eg, does drug *x* inhibit enzyme *y*? The basic question may then be extended, eg, if drug *x* inhibits enzyme *y*, what is the concentration-response relationship? Such experiments are usually reproducible and often lead to reliable insights into the mechanism of the drug's action.

First-in-human studies include phase 1–3 trials. Once a drug receives FDA approval for use in humans, *case reports* and *case series* consist of observations by clinicians of the effects of drug (or other) treatments in one or more patients. These results often reveal unpredictable benefits and toxicities but do not generally test a prespecified hypothesis and cannot prove cause and effect. *Analytic epidemiologic studies* consist of observations designed to test a specified hypothesis, eg, that thiazolidinedione antidiabetic drugs are associated with adverse cardiovascular events. *Cohort* epidemiologic studies utilize populations of patients that have (exposed group) and have not (control group) been exposed to the agents under study and ask whether the exposed groups show a higher or lower incidence of the effect. *Case control* epidemiologic studies utilize populations of patients that have displayed the end point under study and ask whether they have been exposed or not exposed to the drugs in question. Such epidemiologic studies add weight to conjectures but cannot control all confounding variables and therefore cannot conclusively prove cause and effect.

Meta-analyses utilize rigorous evaluation and grouping of similar studies to increase the number of subjects studied and hence the statistical power of results obtained in multiple published studies. While the numbers may be dramatically increased by meta-analysis, the individual studies still suffer from their varying methods and end points and a meta-analysis cannot prove cause and effect. *Large randomized controlled trials* are designed to answer specific questions about the effects of medications on clinical end points or important surrogate end points, using large enough samples of patients and allocating them to control and experimental treatments using rigorous randomization methods. Randomization is the best method for distributing all foreseen confounding factors, as well as unknown confounders, equally between the experimental and control groups. When properly carried out, such studies are rarely invalidated and can be very convincing.

*I thank Ralph Gonzales, MD, for helpful comments.

drug. Confirmation of **compliance** with protocols (also known as **adherence**) is a necessary element to consider.

The various types of studies and the conclusions that may be drawn from them are described in the accompanying text box. (See Box: Drug Studies—The Types of Evidence.)

The Food & Drug Administration

The FDA is the administrative body that oversees the drug evaluation process in the USA and grants approval for marketing of new drug products. To receive FDA approval for marketing, the originating institution or company (almost always the latter) must submit evidence of safety and effectiveness. Outside the USA, the regulatory and drug approval process is generally similar to that in the USA.

As its name suggests, the FDA is also responsible for certain aspects of food safety, a role it shares with the US Department of Agriculture (USDA). Shared responsibility results in complications when questions arise regarding the use of drugs, eg, antibiotics, in food animals. A different type of problem arises when so-called food supplements are found to contain active drugs, eg, sildenafil analogs in "energy food" supplements.

The FDA's authority to regulate drugs derives from specific legislation (Table 5–2). If a drug has not been shown through adequately controlled testing to be "safe and effective" for a specific use, it cannot be marketed in interstate commerce for this use.*

Unfortunately, "safe" can mean different things to the patient, the physician, and society. Complete absence of risk is impossible to demonstrate, but this fact may not be understood by the public, who frequently assume that any medication sold with the approval of the FDA should be free of serious "side effects." This confusion is a major factor in litigation and dissatisfaction with aspects of drugs and medical care.

The history of drug regulation (Table 5–2) reflects several health events that precipitated major shifts in public opinion. The Pure Food and Drug Act of 1906 became law in response to revelations of unsanitary and unethical practices in the meat-packing industry. The Federal Food, Drug, and Cosmetic Act of 1938 was largely a reaction to deaths associated with the use of a preparation of sulfanilamide marketed before it and its vehicle were adequately

*Although the FDA does not directly control drug commerce within states, a variety of state and federal laws control interstate production and marketing of drugs.

TABLE 5–2 Some major legislation pertaining to drugs in the United States.

Law	Purpose and Effect
Pure Food and Drug Act of 1906	Prohibited mislabeling and adulteration of drugs.
Opium Exclusion Act of 1909	Prohibited importation of opium.
Amendment (1912) to the Pure Food and Drug Act	Prohibited false or fraudulent advertising claims.
Harrison Narcotic Act of 1914	Established regulations for use of opium, opiates, and cocaine (marijuana added in 1937).
Food, Drug, and Cosmetic Act of 1938	Required that new drugs be safe as well as pure (but did not require proof of efficacy). Enforcement by FDA.
Durham-Humphrey Act of 1952	Vested in the FDA the power to determine which products could be sold without prescription.
Kefauver-Harris Amendments (1962) to the Food, Drug, and Cosmetic Act	Required proof of efficacy as well as safety for new drugs and for drugs released since 1938; established guidelines for reporting of information about adverse reactions, clinical testing, and advertising of new drugs.
Comprehensive Drug Abuse Prevention and Control Act (1970)	Outlined strict controls in the manufacture, distribution, and prescribing of habit-forming drugs; established drug schedules and programs to prevent and treat drug addiction.
Orphan Drug Amendments of 1983	Provided incentives for development of drugs that treat diseases with less than 200,000 patients in USA.
Drug Price Competition and Patent Restoration Act of 1984	Abbreviated new drug applications for generic drugs. Required bioequivalence data. Patent life extended by amount of time drug delayed by FDA review process. Cannot exceed 5 extra years or extend to more than 14 years post-NDA approval.
Prescription Drug User Fee Act (1992, reauthorized 2007)	Manufacturers pay user fees for certain new drug applications.
Dietary Supplement Health and Education Act (1994)	Established standards with respect to dietary supplements but prohibited full FDA review of supplements and botanicals as drugs. Required the establishment of specific ingredient and nutrition information labeling that defines dietary supplements and classifies them as part of the food supply but allows unregulated advertising.
Bioterrorism Act of 2002	Enhanced controls on dangerous biologic agents and toxins. Seeks to protect safety of food, water, and drug supply.
Food and Drug Administration Amendments Act of 2007	Granted FDA greater authority over drug marketing, labeling, and direct-to-consumer advertising; required post-approval studies, established active surveillance systems, made clinical trial operations and results more visible to the public.

tested. The Kefauver-Harris amendments of 1962 were, in part, the result of a teratogenic drug disaster involving thalidomide. This agent was introduced in Europe in 1957–1958 and, based on animal tests then commonly used, was marketed as a "nontoxic" hypnotic and promoted as being especially useful during pregnancy. In 1961, reports were published suggesting that thalidomide was responsible for a dramatic increase in the incidence of a rare birth defect called phocomelia, a condition involving shortening or complete absence of the arms and legs. Epidemiologic studies provided strong evidence for the association of this defect with thalidomide use by women during the first trimester of pregnancy, and the drug was withdrawn from sale worldwide. An estimated 10,000 children were born with birth defects because of maternal exposure to this one agent. The tragedy led to the requirement for more extensive testing of new drugs for teratogenic effects and stimulated passage of the Kefauver-Harris Amendments of 1962, even though the drug was not then approved for use in the USA. In spite of its disastrous fetal toxicity and effects in pregnancy, thalidomide is a relatively safe drug for humans other than the fetus. Even the most serious risk of toxicities may be avoided or managed if understood, and despite its toxicity, thalidomide is now approved by the FDA for limited use as a potent immunoregulatory agent and to treat certain forms of leprosy.

Clinical Trials: The IND & NDA

Once a new drug is judged ready to be studied in humans, a Notice of Claimed Investigational Exemption for a New Drug (IND) must be filed with the FDA (Figure 5–1). The IND includes (1) information on the composition and source of the drug, (2) chemical and manufacturing information, (3) all data from animal studies, (4) proposed plans for clinical trials, (5) the names and credentials of physicians who will conduct the clinical trials, and (6) a compilation of the key data relevant to study of the drug in humans that has been made available to investigators and their institutional review boards.

It often requires 4–6 years of clinical testing to accumulate and analyze all required data. Testing in humans is begun only after sufficient acute and subacute animal toxicity studies have been completed. Chronic safety testing in animals, including carcinogenicity studies, is usually done concurrently with clinical trials. In each of the three formal phases of clinical trials, volunteers or

patients must be informed of the investigational status of the drug as well as the possible risks and must be allowed to decline or to consent to participate and receive the drug. These regulations are based on the ethical principles set forth in the Declaration of Helsinki (1966). In addition to the approval of the sponsoring organization and the FDA, an interdisciplinary institutional review board (IRB) at the facility where the clinical drug trial will be conducted must review and approve the scientific and ethical plans for testing in humans.

In **phase 1**, the effects of the drug as a function of dosage are established in a small number (20–100) of healthy volunteers. Although a goal is to find the maximum tolerated dose, the study is designed to prevent severe toxicity. If the drug is expected to have significant toxicity, as may be the case in cancer and AIDS therapy, volunteer patients with the disease are used in phase 1 rather than normal volunteers. Phase 1 trials are done to determine the probable limits of the safe clinical dosage range. These trials may be nonblind or "open"; that is, both the investigators and the subjects know what is being given. Alternatively, they may be "blinded" and placebo controlled. The choice of design depends on the drug, disease, goals of investigators, and ethical considerations. Many predictable toxicities are detected in this phase. Pharmacokinetic measurements of absorption, half-life, and metabolism are often done. Phase 1 studies are usually performed in research centers by specially trained clinical pharmacologists.

In **phase 2**, the drug is studied in patients with the target disease to determine its efficacy ("proof of concept"), and the doses to be used in any follow-on trials. A modest number of patients (100–200) are studied in detail. A single-blind design may be used, with an inert placebo medication and an established active drug (positive control) in addition to the investigational agent. Phase 2 trials are usually done in special clinical centers (eg, university hospitals). A broader range of toxicities may be detected in this phase. Phase 2 trials have the highest rate of drug failures, and only 25% of innovative drugs move on to phase 3.

In **phase 3**, the drug is evaluated in much larger numbers of patients with the target disease—usually thousands—to further establish and confirm safety and efficacy. Using information gathered in phases 1 and 2, phase 3 trials are designed to minimize errors caused by placebo effects, variable course of the disease, etc. Therefore, double-blind and crossover techniques are often used. Phase 3 trials are usually performed in settings similar to those anticipated for the ultimate use of the drug. Phase 3 studies can be difficult to design and execute and are usually expensive because of the large numbers of patients involved and the masses of data that must be collected and analyzed. The drug is formulated as intended for the market. The investigators are usually specialists in the disease being treated. Certain toxic effects, especially those caused by immunologic processes, may first become apparent in phase 3.

If phase 3 results meet expectations, application is made for permission to market the new agent. Marketing approval requires submission of a New Drug Application (NDA)—or for biologicals, a Biological License Application—to the FDA. The application contains, often in hundreds of volumes, full reports of all preclinical and clinical data pertaining to the drug under review. The number of subjects studied in support of the new drug application has been increasing and currently averages more than 5000 patients for new drugs of novel structure (new molecular entities). The duration of the FDA review leading to approval (or denial) of the new drug application may vary from months to years. Priority approvals are designated for products that represent significant improvements compared with marketed products; in 2007, the median priority approval time was 6 months. Standard approvals, which take longer, are designated for products judged similar to those on the market—in 2007, the median standard approval time was 10.2 months. If problems arise, eg, unexpected but possibly serious toxicities, additional studies may be required and the approval process may extend to several years.

In cases in which an urgent need is perceived (eg, cancer chemotherapy), the process of preclinical and clinical testing and FDA review may be accelerated. For serious diseases, the FDA may permit extensive but controlled marketing of a new drug before phase 3 studies are completed; for life-threatening diseases, it may permit controlled marketing even before phase 2 studies have been completed. Roughly 50% of drugs in phase 3 trials involve early, controlled marketing. Such "accelerated approval" is usually granted with the requirement that careful monitoring of the effectiveness and toxicity of the drug be carried out and reported to the FDA. Unfortunately, FDA enforcement of this requirement has not always been adequate.

Once approval to market a drug has been obtained, **phase 4** begins. This constitutes monitoring the safety of the new drug under actual conditions of use in large numbers of patients. The importance of careful and complete reporting of toxicity by physicians after marketing begins can be appreciated by noting that many important drug-induced effects have an incidence of 1 in 10,000 or less and that some adverse effects may become apparent only after chronic dosing. The sample size required to disclose drug-induced events or toxicities is very large for such rare events. For example, several hundred thousand patients may have to be exposed before the first case is observed of a toxicity that occurs with an average incidence of 1 in 10,000. Therefore, low-incidence drug effects are not generally detected before phase 4 no matter how carefully the studies are executed. Phase 4 has no fixed duration. As with monitoring of drugs granted accelerated approval, phase 4 monitoring has often been lax.

The time from the filing of a patent application to approval for marketing of a new drug may be 5 years or considerably longer. Since the lifetime of a patent is 20 years in the USA, the owner of the patent (usually a pharmaceutical company) has exclusive rights for marketing the product for only a limited time after approval of the new drug application. Because the FDA review process can be lengthy, the time consumed by the review is sometimes added to the patent life. However, the extension (up to 5 years) cannot increase the total life of the patent to more than 14 years after approval of a new drug application. As of 2005, the average effective patent life for major pharmaceuticals was 11 years. After expiration of the patent, any company may produce the drug, file an abbreviated new drug application (ANDA), demonstrate required equivalence, and, with FDA approval, market the drug as a **generic** product without paying license fees to the original patent owner. Currently, 67% of prescriptions in the USA are for

generic drugs. Even biotechnology-based drugs such as antibodies and other proteins are now qualifying for generic designation, and this has fueled regulatory concerns.

A **trademark** is the drug's proprietary trade name and is usually registered; this registered name may be legally protected as long as it is used. A generically equivalent product, unless specially licensed, cannot be sold under the trademark name and is often designated by the official generic name. Generic prescribing is described in Chapter 65.

The FDA drug approval process is one of the rate-limiting factors in the time it takes for a drug to be marketed and to reach patients. The Prescription Drug User Fee Act (PDUFA) of 1992, reauthorized in 2007, attempts to make more FDA resources available to the drug approval process and increase efficiency through use of fees collected from the drug companies that produce certain human drugs and biologic products. In 2009, the FDA approved 19 new molecular entity drug applications for new nonbiologic entities and six biological license applications, one more than in 2008. The traditional sequential and linear drug development process previously described is being increasingly modified in an attempt to safely accelerate clinical trials that provide "proof of mechanism" of action and "proof of concept" that the drug does work in the target disease. In these newer approaches, certain development activities such as full dose-response studies, final drug formulation work, and long-term toxicology studies may be deferred. It is hoped that this approach will focus resources on drugs more likely to succeed and minimize later-stage failures. In one example, a phase 0 (phase zero) clinical trial is designed to study the pharmacodynamic, pharmacokinetic properties of a drug and its links to useful biomarkers and measures of mechanism. Unlike a phase 1 trial with dose-response studies, in a phase 0 trial, a limited number of low doses are administered. These trials are not designed to be therapeutic.

Conflicts of Interest

Several factors in the development and marketing of drugs result in conflicts of interest. Use of pharmaceutical industry funding to support FDA approval processes raises the possibility of conflicts of interest within the FDA. Supporters of this policy point out that chronic FDA underfunding by the government allows for few alternatives. Another important source of conflicts of interest is the dependence of the FDA on outside panels of experts that are recruited from the scientific and clinical community to advise the government agency on questions regarding drug approval or withdrawal. Such experts are often recipients of grants from the companies producing the drugs in question. The need for favorable data in the new drug application leads to phase 2 and 3 trials in which the new agent is compared only to placebo, not to older, effective drugs. As a result, data regarding the efficacy and toxicity of the new drug *relative to a known effective agent* may not be available when the new drug is first marketed.

Manufacturers promoting a new agent may pay physicians to use it in preference to older drugs with which they are more familiar. Manufacturers sponsor small and often poorly designed clinical studies after marketing approval and aid in the publication of favorable results but may retard publication of unfavorable results. The need for physicians to meet continuing medical education (CME) requirements in order to maintain their licenses encourages manufacturers to sponsor conferences and courses, often in highly attractive vacation sites, and new drugs are often featured in such courses. Recognition of the obvious conflicts of interest is leading some clinical specialty organizations to reject industry support of such conferences. Finally, the common practice of distributing free samples of new drugs to practicing physicians has both positive and negative effects. The samples allow physicians to try out new drugs without incurring any cost to the patient. On the other hand, new drugs are usually much more expensive than older agents and when the free samples run out, the patient (or insurance carrier) may be forced to pay much more for treatment than if the older, cheaper, and possibly equally effective drug were used. Finally, when the patent for a drug is nearing expiration, the patent-holding manufacturer may try to extend its exclusive marketing privilege by paying generic manufacturers to *not* introduce a generic version ("pay to delay").

Translational Research

Unfortunately, the rate of introduction of new drugs has fallen during the last two decades. This has raised concerns about our ability to deal with the increasing prevalence of resistant microorganisms, and the problems of degenerative diseases in an aging population. In an effort to facilitate this process, the National Institutes of Health are currently (2011) considering the establishment of a new institute specializing in translational research.

Adverse Drug Reactions

An adverse drug event (ADE) or reaction to a drug (ADR) is a harmful or unintended response. Adverse drug reactions are claimed to be the fourth leading cause of death, higher than pulmonary disease, AIDS, accidents, and automobile deaths. The FDA has further estimated that 300,000 preventable adverse events occur in hospitals, many as a result of confusing medical information or lack of information (for example, regarding drug incompatibilities). Some adverse reactions, such as overdose, excessive effects, and drug interactions, may occur in anyone. Adverse reactions occurring only in susceptible patients include intolerance, idiosyncrasy (frequently genetic in origin), and allergy (usually immunologically mediated). During IND studies and clinical trials before FDA approval, all adverse events (serious, life-threatening, disabling, reasonably drug related, or unexpected) must be reported. After FDA approval to market a drug, surveillance, evaluation, and reporting must continue for any adverse events that are related to use of the drug, including overdose, accident, failure of expected action, events occurring from drug withdrawal, and unexpected events not listed in labeling. Events that are both serious and unexpected must be reported to the FDA within 15 days. In 2008, the FDA began publishing quarterly a list of drugs being investigated for potential safety risks. The ability to predict and avoid adverse drug reactions and optimize a drug's therapeutic index is an increasing focus of pharmacogenetic and personalized medicine.

It is hoped that greater use of electronic health records will reduce some of these risks (see Chapter 65).

Orphan Drugs & Treatment of Rare Diseases

Drugs for rare diseases—so-called orphan drugs—can be difficult to research, develop, and market. Proof of drug safety and efficacy in small populations must be established, but doing so is a complex process. Furthermore, because basic research in the pathophysiology and mechanisms of rare diseases receives relatively little attention or funding in both academic and industrial settings, recognized rational targets for drug action may be few. In addition, the cost of developing a drug can greatly influence priorities when the target population is relatively small. Funding for development of drugs for rare diseases or ignored diseases that do not receive priority attention from the traditional industry has received increasing support via philanthropy or similar funding from not-for-profit foundations such as the Cystic Fibrosis Foundation, the Huntington's Disease Society of America, and the Gates Foundation.

The Orphan Drug Amendments of 1983 provides incentives for the development of drugs for treatment of a rare disease or condition defined as "any disease or condition which (a) affects less than 200,000 persons in the U.S. or (b) affects more than 200,000 persons in the U.S. but for which there is no reasonable expectation that the cost of developing and making available in the U.S. a drug for such disease or condition will be recovered from sales in the U.S. of such drug." Since 1983, the FDA has approved for marketing more than 300 orphan drugs to treat more than 82 rare diseases.

REFERENCES

Avorn J: Debate about funding comparative effectiveness research. N Engl J Med 2009;360:1927.

Avorn J: *Powerful Medicines: The Benefits and Risks and Costs of Prescription Drugs.* Alfred A. Knopf, 2004.

Boutron I et al: Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA 2010;303:2058.

Brown WA: The placebo effect. Sci Am 1998;1:91.

DiMasi JA: Rising research and development costs for new drugs in a cost containment environment. J Health Econ 2003;22:151.

Editor's Page: Code of Ethics of the World Medical Association: Declaration of Helsinki. Clin Res 1966;14:193.

Evans RP: *Drug and Biological Development: From Molecule to Product and Beyond.* Springer, 2007.

FDA website: http://www.fda.gov

Ferrara N et al: Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. Nat Rev Drug Discov 2004;3(5):391.

Grabowski H, Vernon J, DiMasi J: Returns on research and development for 1990s new drug introductions. Pharmacoeconomics 2002;20(Suppl 3):11.

Hennekens CH, DeMets D: Statistical association and causation. Contributions of different types of evidence. JAMA 2011;305:1134.

Hoffman GA, Harrington A, Fields, HK: Pain and the placebo response: What have we learned. Perspect Biol Med 2005;48:248.

Holaday JW, Berkowitz BA: Antiangiogenic drugs: Insights into drug development from endostatin, Avastin, and thalidomide. Mol Interv 2009;9:157.

Huang S-M, Temple R: Is this the drug or dose for you? Impact and consideration of ethnic factors in global drug development, regulatory review, and clinical practice. Clin Pharmacol Ther 2008;84:287; or http://www.fda.gov/cder/genomics/publications.htm

Hughes B: 2009 FDA drug approvals. Nat Rev Drug Discov 2010;9:89

Kesselheim AS et al: Whistle-blowers experiences in fraud litigation against pharmaceutical companies. N Engl J Med 2010;362:1832.

Lee C-J et al: *Clinical Trials of Drugs and Biopharmaceuticals.* CRC Publishing, 2005.

Lesko LJ: The critical path of warfarin dosing: Finding an optimal dosing strategy using pharmacogenetics. Clin Pharmacol Ther 2008;84:301; http://www.fda.gov/cder/genomics/publications.htm

Light DW: Global drug discovery: Europe is ahead. Health Affairs 2009;28:969.

Mogolian E, Mrydral P: What's the difference between brand-name and generic prescription drugs? Sci Am 2004;Dec 13; http://www.sciam.com/article.cfm?id=whats-the-difference-betw-2004-12-13

Ng R: *Drugs from Discovery to Approval.* Wiley-Blackwell, 2008.

Pharmaceutical Research and Manufacturers of America website: http://www.phrma.org

Rockey SJ, Collins FS: Managing financial conflict of interest in biomedical research. JAMA 2010;303:2400

Scheindlin S: Demystifying the new drug application. Molec Interventions 2004;4:188.

Sistare FD, DeGeorge JJ: Preclinical predictors of clinical safety: Opportunities for improvement. Clin Pharmacol Ther 2007;82(2):210.

Stevens AJ et al: The role of public sector research in the discovery of drugs and vaccines. N Engl J Med 2011;364:535.