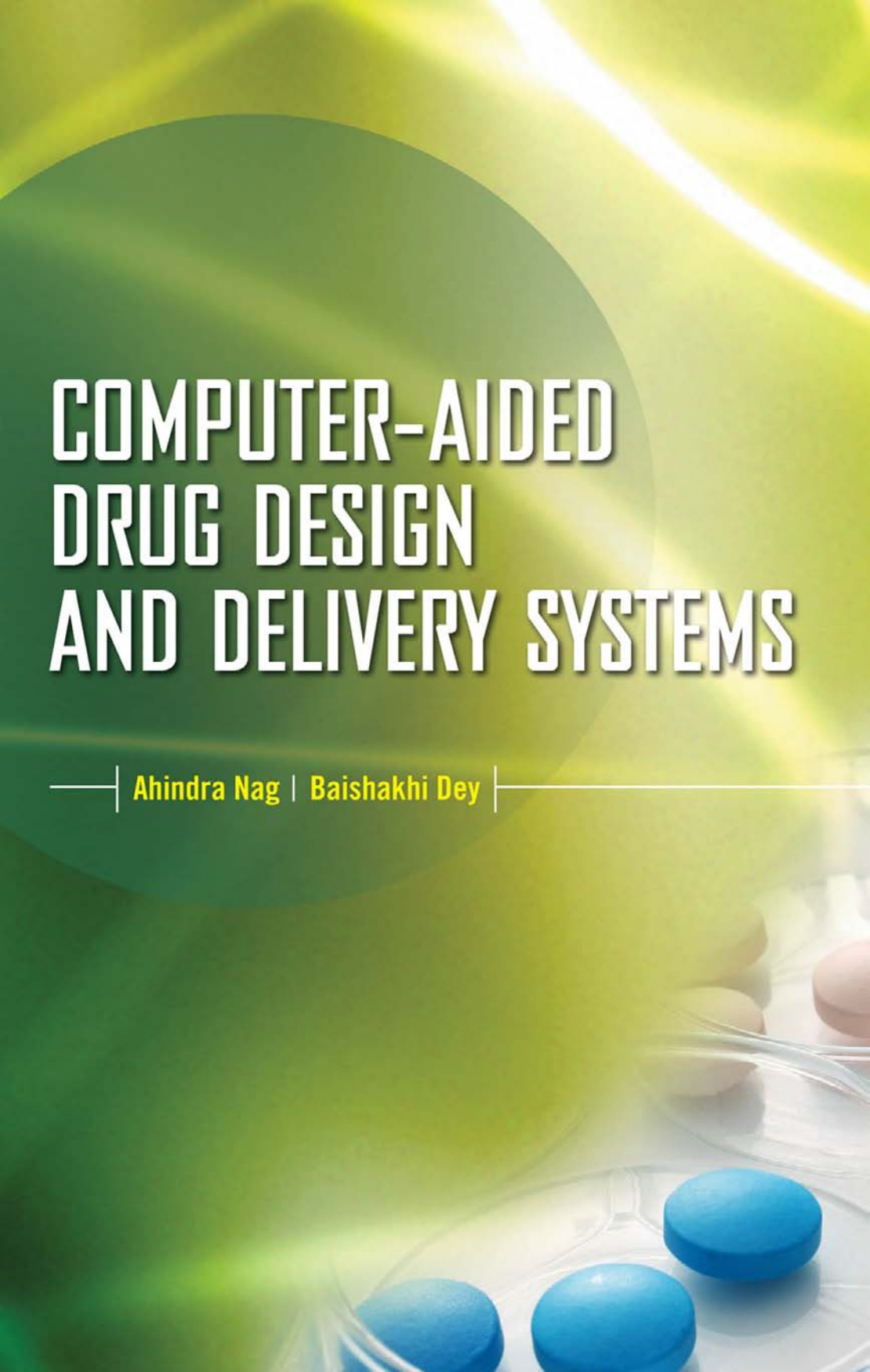
COMPUTER-AIDED DRUG DESIGN AND DELIVERY SYSTEMS

Ahindra Nag | Baishakhi Dey



Computer-Aided Drug Design and Delivery Systems

This page intentionally left blank

Computer-Aided Drug Design and Delivery Systems

Ahindra Nag
Baishakhi Dey

New York Chicago San Francisco
Lisbon London Madrid Mexico City
Milan New Delhi San Juan
Seoul Singapore Sydney Toronto

ISBN: 978-0-07-170125-9

MHID: 0-07-170125-7

The material in this eBook also appears in the print version of this title: ISBN: 978-0-07-170124-2, MHID: 0-07-170124-9.

About the Authors

Ahindra Nag is a professor in the Chemistry Department at the Indian Institute of Technology.

Baishakhi Dey is affiliated with A. M. Reddy Memorial College of Pharmacy.

This page intentionally left blank

Contents

This page intentionally left blank

Preface

One of the important aims of this book is to delineate essential information about drugs and their delivery in the human system and measurement of disorders in the human system by means of bioinstruments and biosensors. The information and views have been arranged in an orderly sequence of distinct statements. This book can serve as an important guide for undergraduates and postgraduate students of chemistry, pharmacy, biochemistry, medicine, and biotechnology. It can be a good handbook for professionals and academicians as well as researchers.

The book consists of seven chapters. Chapter 1 discusses computer-aided drug design. Computer-aided drug design can be used to create drugs based on knowledge about the receptors on which they act. Computers help not only in drug design but also in biologic screening of molecules. By means of computers, screening of tens of thousands of compounds can be carried out in a week. At the same time, the incentive to develop drugs by playing around with their structures is spurred by the need to develop molecules that are safer, more effective, and cheaper, and have better kinetics. Chapter 2 addresses drug delivery systems. In this chapter, the various methodologies and techniques of drug administration to the body, how drug delivery technique can be modified to prolong drug action, and the latest techniques of drug delivery are discussed. Chapter 3 deals with the bioinformatics of drug molecules and databases. Starting from the fundamental definitions, applications of bioinformatics in rational drug design, biotechnology, medicine, and engineering and details about bioinformatics databases are presented. Application of lipase and esterase in the preparation of drugs and drug intermediates is the topic of Chapter 4.

Objective and quantitative data have been provided on the use of drugs in humans, pharmacokinetic and pharmacodynamic considerations (i.e., the relationship between plasma concentration and the intensity of therapeutic or toxic actions), plasma half-lives, the relative efficacy of different medications, and the incidences of adverse effects. The study of pharmacokinetics shows what our body does to

a drug. Pharmacokinetics is actually the quantitative study of drug movement in, through, and out of the body—the intensity of response being related to drug concentration at the site of action, which, in turn, depends on pharmacokinetic properties. Pharmacokinetic considerations determine the route of administration, dose, latency of onset, time of peak action, duration of action, and frequency of administration of a drug.

Pharmacodynamics is the study of the action of drugs on our body. Here, discussions are based on the physiologic and biochemical effects of drugs and their modes of action at the macromolecular/subcellular and organ-system levels. Attempts have been made to elucidate the complete action-effect sequence and the dose-effect relationship. How the effect of one drug modifies the action of another drug is also a part of pharmacodynamics. Pharmacokinetics and pharmacodynamics of drug molecules are the subjects of Chapter 5.

Chapter 6 discusses biomarkers, biosensors, and robotics in medicine. A biomarker is a substance that can be used as an indicator of a normal biologic process, a pathogenic process, or a pharmacologic response. Biomarkers have widespread applications in medicine, technology, epidemiology, genetics, and environmental exposures. In this book, however, Chapter 6 deals mainly with the role of biomarkers in medicine, giving a special emphasis to cancer biomarkers, which have become a major focus of cancer research and cancer screening. Biomarkers are also useful for diagnosis, monitoring disease progression, predicting disease recurrence, and determining therapeutic efficacy. Robotic technologies, miniaturization, the development of microscience technologies, and the application of microrobots in surgery are innovative fields of research. Briefly, an attempt has been made to provide information on different robotic technologies being applied in biomedical science as well as in robot-assisted surgery.

The concluding chapter of this book addresses biomedical instrumentation. The field of medical diagnostic instruments has developed greatly over the past 30 years based on improvements in electronic engineering. It is now so important that most engineering graduate and science postgraduate courses include the subject in the curriculum. Although this field is vast in terms of applications, here we discuss the important instruments related to our body systems, as well as sense organs, from both medical and technical angles.

It is our hope that this book will meet the demand of all those who study this subject either in a course or for research applications in both medical and electronic technology.

Ahindra Nag

Baishakhi Dey

Computer-Aided Drug Design and Delivery Systems

This page intentionally left blank

CHAPTER 1

Computer-Aided Drug Design (CADD)

1.1 Introduction

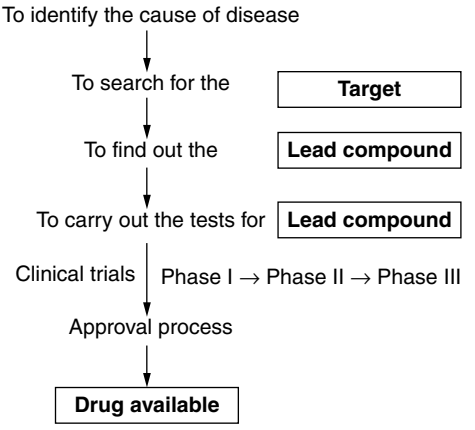
People in every civilization have used drugs of plant or animal origin to prevent and treat disease. The quest for substances to combat sickness and to alter mood and consciousness is nearly as basic as the search for food and shelter. Many drugs obtained from natural sources are highly valued, but most drugs used in modern medicine are the products of advances in synthetic organic chemistry and biotechnology. Thus a *drug* can be defined as a substance of either natural or synthetic origin that is used in the diagnosis, cure, relief, treatment, or prevention of disease or intended to affect the structure or function of the body. Thus a drug is a chemical that affects the body and its processes.[1,2]

A Little History of Computer-Aided Drug Design

1960s	Review the target-drug interaction
1980s	Automation: High-throughput target/drug selection
1980s	Databases (information technology): Combinatorial libraries
1980s	Fast computers: Docking
1990s	Fast computers: Genome assembly, genomic-based target selection
2000s	Vast information handling: Pharmacogenomics

1.2 Rational Drug Design

The development of a drug follows five classic steps:



However, the traditional approach to drug discovery is really time-consuming and cost-intensive. Thus the new approach to drug discovery exceeds the limitations of the traditional research. It evolved out of the following consideration: The target in the body and the potential active compound are directly related to each other.

For designing a drug, knowledge about the disease and previous infectious processes is a must. For *rational* drug design, the first step is to identify a molecular target that is critical to a disease process or an infectious pathogen. The next important step is to determine the molecular structure of the target, which makes sense of the word *rational*. The validity of *rational* or *structure-based* drug discovery rests largely on a high-resolution target structure of sufficient molecule detail to allow selectivity in the screening of compounds.

Modern drug research owes its rationality to German physician and immunologist Paul Ehrlich. He developed the intellectual tools of medical science, such as receptors for drugs. Domagk discovered "prontosil" as an antibacterial agent, which laid the foundation for the concepts of biochemical metabolites and steric analogues. This enabled medicinal chemists to engage chemical *drug design*. Drug design was based on modification of the structure of a *lead* compound when the *lead* compound suffered from some therapeutically undesirable side effects. Hansch (1964) through quantitative structural activity relationship (QSAR) has helped medicinal chemists to target the design of drugs using *Hammett's substituent constants*.[1]

After studying about 5,000 to 10,000 compounds, only one drug comes to the market. In the discovery phase, each drug costs about $156 million. Food and Drug Administration (FDA) processes I, II, and III cost another $75 million. This brings the total to about $231 million for each drug placed on the market for consumers. Then, for gaining FDA approval, a long and expensive procedure also needs to be followed.[2]

1.3 Factors Contributing to Drug Discovery

Besides the long and expensive drug discovery cycle, other factors contribute to the rapidly changing landscape of the drug discovery environment:

1. Advances in molecular biology and high-throughput screening
2. Demand fundamentals
   a. Aging population of the baby boomers
   b. Consumer demand for quality health care
   c. Expanded access and universal health care
   d. New breakthrough technologies
   e. Consumer awareness of the quality of nutrition and supplements
3. Supply fundamentals
   a. Hospital downsizing
   b. Insurers reluctance to pay high reimbursements
   c. Transition to outpatient procedures
   d. Disease management
   e. Global management

Owing to these factors—regulation, the cost-effectiveness of drug discovery, and the fundamentals of demand and supply—the process of drug discovery is undergoing a complete overhaul. Companies that have been making a fortune from the sale of drugs are expected to shift their focus to tap into information. Companies are now using intelligent software and are employing cheminformation to shorten the cycle of drug discovery and thus make the drug discovery process cost-effective.[2–4]

1.4 Drug Design with the Help of Software

Based on the fine structure of the target molecule, a whole new ligand is constructed (Fig. 1.1). This is just de novo discovery of a ligand. Ligbuild is a powerful tool to build a ligand.

1.5 Drug Discovery Process

The major steps in the drug discovery process following identification of the biologic target involve the following:

1. *Hit identification*. This involves screening large collections of compounds to identify those which interact with the biologic

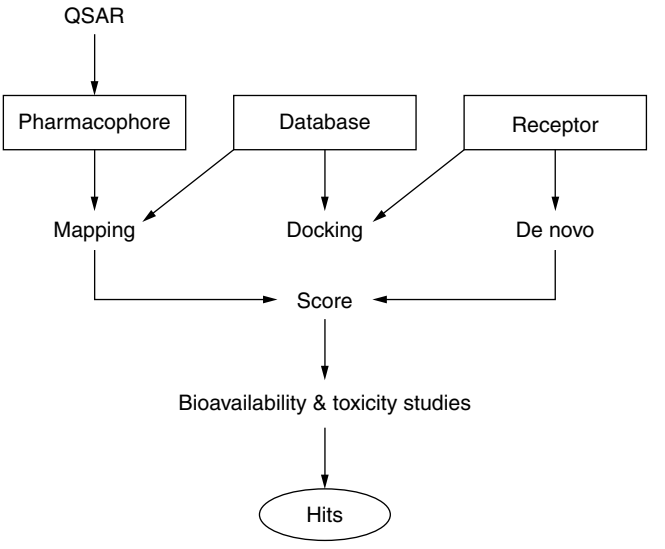


Figure 1.1 Flowchart for drug design.

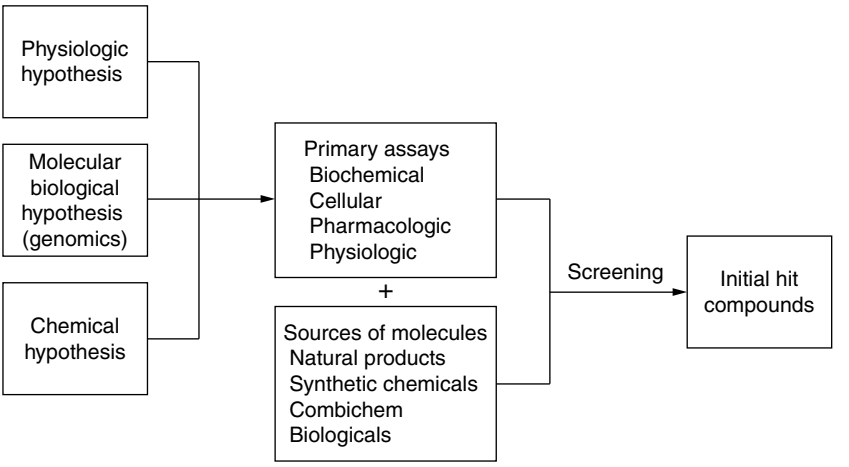
target. A compound that interacts with a target protein is referred to as a *hit.*

2. *Lead generation.* This involve the chemical modification of hits by repeated cycles of synthesis and testing of analogues to produce *leads,* which are compounds with improved chemical characteristics, thereby increasing their suitability as potential drugs.

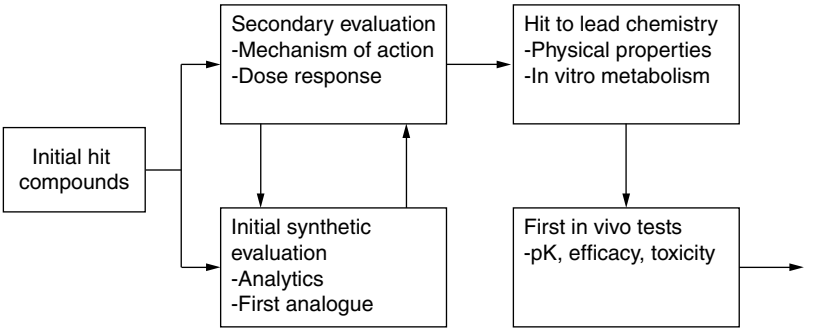
   Lead optimization involves the further optimization of leads by additional repeated modification to produce drug development candidates with optimized characteristics for further preclinical and clinical development. Then *target validation* seeks to establish the link between the target protein and the particular clinical disease. These tests usually involve correlating changes in the level of the target protein in cells or animals with changes in cell biology or animal physiology characteristic of the disease state (Fig. 1.2).[2–4]

1.6 How Computers Help in Drug Design

1. For storage and retrieval of information, for example, structures determined experimentally by x-ray crystallography for biologic targets (enzymes) and drug molecules
2. To develop information about toxicity and structural activity relationship (SAR)



Drug Discovery Processes - II



Drug Discovery Processes - III

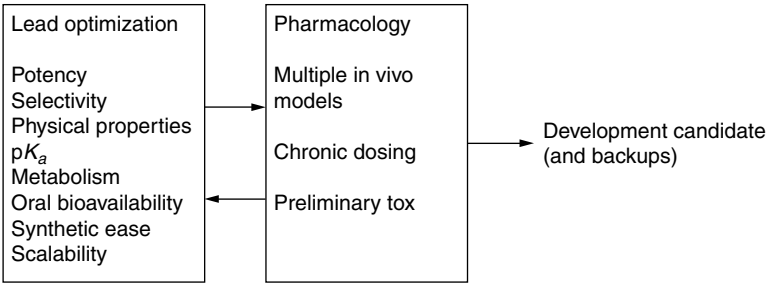


Figure 1.2 Drug discovery processes (today).

3. To visualize the similarities/differences between molecules (drugs) acting in the same way
4. To study drug-receptor interactions
5. To do calculations involving interaction strength and motion dynamics

1.7 Computer Simulation for Drug Design

When developing a new drug, the first step is to design the suitable candidate compound, the ligand, and then to recognize the target protein and how to bind it. However, proteins are very versatile, and often their shapes change as they fulfill their different functions. Hence different techniques are adopted to study protein dynamics. In all cases, though, the cost and time factor are matters of concern. This is why computer simulation of the dynamics of molecules is becoming increasingly important.

High-performance computing (HPC) is required to cross the threshold to where *MD simulation* becomes a valuable tool for industry. However, in most pharmaceutical companies, HPC is something very new, and supercomputers are not readily available to industrial researchers. With the arrival of affordable high-performance multiprocessors and corresponding development of parallel software, it now has become possible for industrial researchers to undertake more realistic calculations that previously were out of reach.

Scientists at NOVO Nordisk, a large Danish pharmaceutical company, are convinced that this new capability will change the acceptance of MD simulation dramatically as a tool in the design of new ligands. During "Europort-D,"* they could, for the first time, study the dynamics of the complex molecular interactions critical for recognition of ligands by their target proteins.[3,4]

1.8 Drug Design Theory

The basic concept behind drug design is to understand the method by which the active site of a receptor selectively restricts the binding of inappropriate structures. Any potential molecule that can bind to a receptor is called a *ligand*. For a ligand to bind and interact with a receptor, a specific combination of atoms presenting the

*Europort was a large cluster of projects with the goal of moving a large number of commercial application codes to parallel platforms. In Europort-D, the results are redesigned into demonstrators that can be used to show companies what parallel computing could do for them. Both the Europort and the Europort-D projects were funded by the European Commission's Esprit Programme.

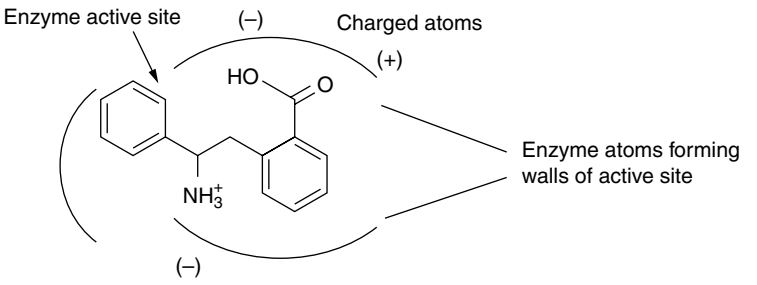


Figure 1.3 Enzyme-substrate complementary interactions.

correct size, shape, and charge composition of the atoms is necessary. A putative *ligand-receptor interaction* has complementary ligand-receptor size and shape, phenomenon called *steric complementarity* (Fig. 1.3).

In addition to steric complementarity, electrostatic interactions also influence ligand binding by restricting the binding of inappropriate molecules because the ligand must contain correctly placed complementary charged atoms for interaction to occur. However, the main driving force for receptor binding is hydrophobic interaction. Two-thirds of our bodies is water, and it is the hydrophobic nature of the ligand that provides the driving force to compel the ligand to leave the water and bind to a receptor.

Numerous potential interactions occur between ligands and receptors, and the specific interactions that are crucial for ligand recognition and binding by a receptor are termed the *pharmacophore.*

By using a *lock-and-key analogy,* we can imagine a lock having numerous tumblers. Now, there may be many keys that can sterically complement the lock and fit within the keyhole. All but the correct key will displace the wrong tumblers, however, leading to a suboptimal interaction with the lock. Only the correct key, which presents the pharmacophore to the receptor, contacts the appropriate tumblers and interacts properly with the lock to open it. This is crucial to the design of pharmaceuticals because any successful drug must incorporate the appropriate chemical structures and present the pharmacophore to the receptor (Fig. 1.4).

Now to face the challenge of designing a drug that addresses a specific target receptor, the major concerns are as follows:

1. To characterize the medical condition and determine receptor targets
2. To achieve active-site complementarity: steric, electrostatic, and hydrophobic
3. To consider biochemical mechanism for receptors
4. To adhere to the laws of chemistry

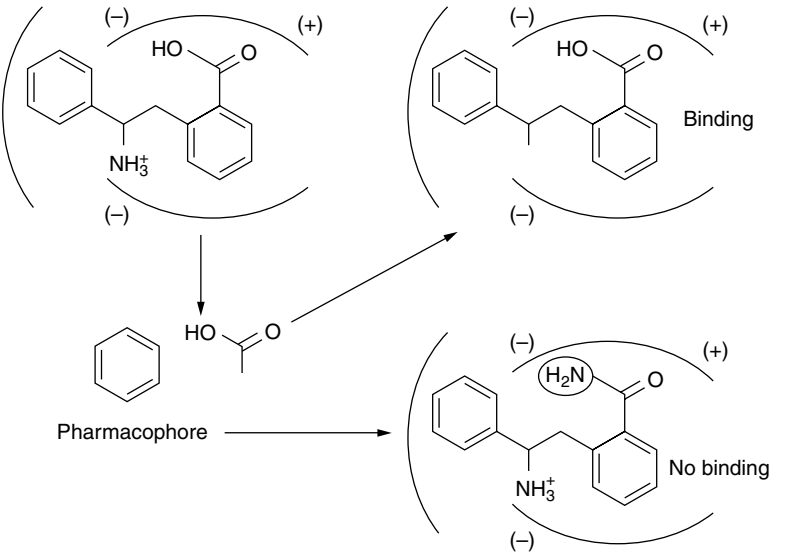


Figure 1.4 Pharmacophore and receptor binding.

5. To achieve synthetic feasibility
6. To address biologic considerations
7. To develop patent considerations

However, the modern drug discovery pipeline can be presented as shown in Fig. 1.5.

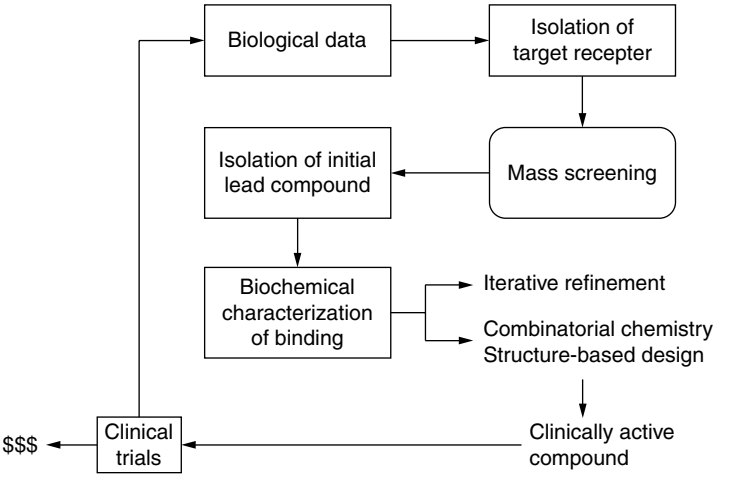


Figure 1.5 Modern discovery of drugs.

1.9 Role of Computers in Drug Design: Their Success and Failure

In the early 1990s, there was a great deal of optimism that CADD would revolutionize the way in which drugs are developed. The enduring exponential increase in computing power progressed to such a point that rudimentary estimations of ligand-receptor complementarities could be performed. Through *computer graphics technology*, scientists acquired the ability to generate vector models of chemical structures and manipulate them in real time. By using computers, computational chemists believed that they could circumvent much of the time and effort required for drug synthesis and testing simply by generating novel compounds with the help of computers. The concept of generating virtual lead compounds entirely through the computer simulation was termed *de novo design.*

The world's largest pharmaceutical firms spent millions of dollars on hardware and software to turn de novo design into a reality. Unfortunately, success is rare, and except for few cases, de novo design proved to be an utter failure. De novo design could not prove itself to be an effective method to discover lead compounds. The main reasons behind are limitations in computing power and a lack of useful software functions. In scientific computing accuracy and processing time are very important. Thus, to make calculations run in a finite period of time, assumptions, algorithms, approximations, and other shortcuts are necessary. This greatly diminished the calculated accuracy of any ligand-receptor interactions. Even though chemists postulated numerous chemical structures that potentially could complement the active site based on computer simulations, the calculated binding had no correlation with reality.

This remains the most significant challenge in de novo design. Although computers have become exponentially faster, the sheer number of calculations needed to accurately predict the binding of a de novo–generated ligand to its receptor in a useful time frame still requires significant approximations. In de novo design, attempts are being made to generate whole ligands from scratch and dock them within their receptors. However, the problem remains how the predicted structure behaves in real life.

The second significant problem in computer-aided de novo design is the generation of undesired chemical structures that are of no use. The technological advancement of RACHEL attempts to circumvent these problems.

In this situation, *mass screening* and *combinatorial chemistry* came to the fore and gained wide acceptance. By using these techniques, researchers were provided with an opportunity to discover lead compounds in a rapid and efficient manner. De novo design tools and their associated problems were no longer needed to generate lead structures. Now it became apparent that computational tools

were needed that could *optimize* these lead compounds into potent drugs.

The concept of drug optimization versus de novo design is an important fact. In de novo ligand generation, an entire structure is created from scratch. In drug optimization, we begin with a lead compound whose bound structure with the receptor has been characterized by x-ray crystallography.

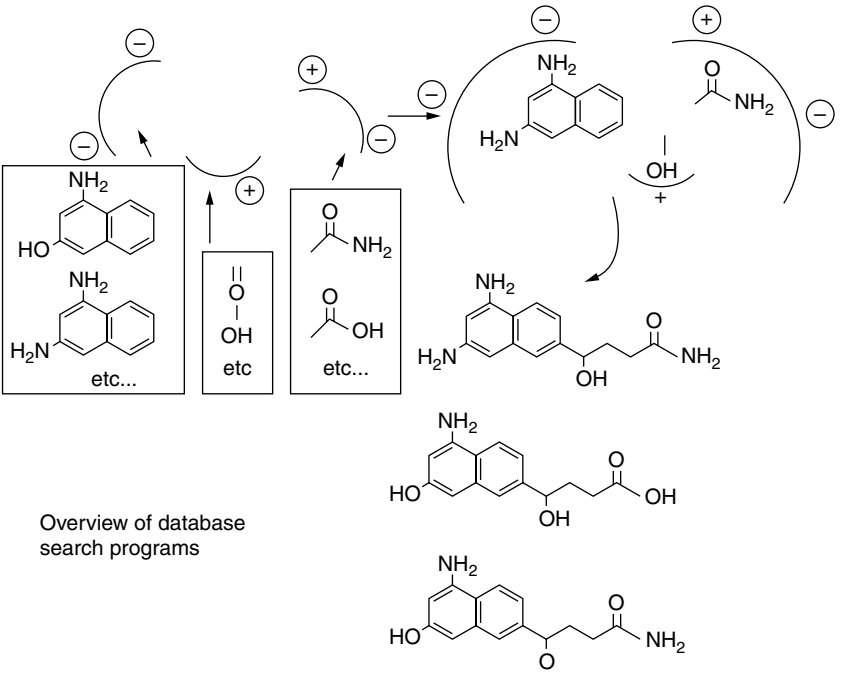
However, generation of chemical derivatives is highly amenable to computerized automation. Computers can rapidly generate and predict the binding of all potential derivatives, creating a list of the best potential candidates. Thus, using CADD software helps in the refinement of weakly binding lead compounds in the most effective manner.[5,6]

1.10 Rational Drug Design Software

Since 1990, dozens of rational drug design packages have been published. They fall in three major categories: scanners, builders, and hybrids.

1.10.1 Scanners

Generally, these types of programs are used for lead compound screening. All database search programs fall into this category.



Overview of database search programs

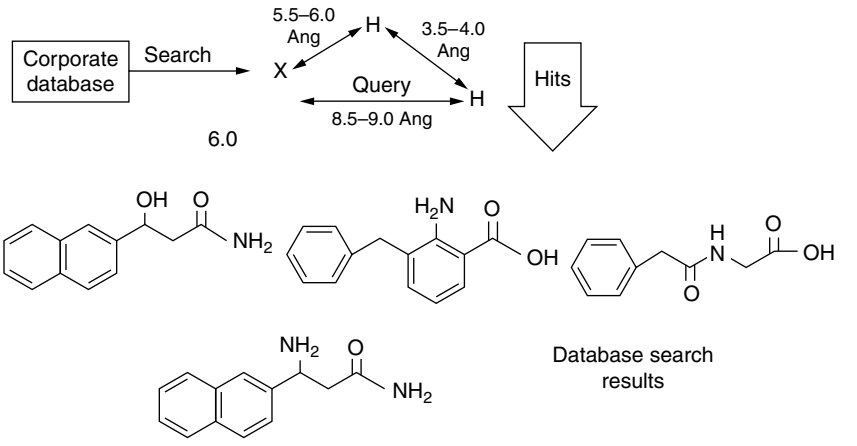


Figure 1.6 Overview of corporate databases.

Figure 1.6 shows how the database programs are used. In the first part of the figure, an active site with a lead compound whose binding structure is determined is found. After biochemical analysis of the ligand-receptor interaction we get three groups of ligands (hits) that make up the pharmacophore—a phenyl ring, an amide hydrogen, and a hydroxyl group.

The pharmacophore is transformed into a query that specifies the three-dimensional (3D) relationship between the various functional groups desired. A database of the compounds is used to develop the 3D structures. Then the query is used to search the database.

These database search programs help users to have complete control over the query specifications. We can retrieve the structures that meet the requirements of the pharmacophore and get an opportunity to complement the receptor.

The advantages of these programs are as follows:

1. Since the programs use the database of known compounds, synthetic feasibility is not an issue here.
2. These programs are highly optimized for speed and thus allow rapid determination of potential binding ligands.
3. The compounds that are retrieved mirror the query, so no scoring functions are needed.
4. The assumption that 3D structures stored in the database are representative of the biologic reality holds true for small molecules.

However, database search programs have certain limitations. First and foremost is the need is to maintain a wealth of corporate

databases. Unfortunately, most academic institutions lack such databases because of poor funding and lack of personnel. Second, all databases have a finite number of structures, and this limits and biases potential solutions. The diversity of potential hits is limited as well.

1.10.2 Builders

Builder-type programs are used in de novo design, but their chief use is in the optimization of lead compounds. These programs, instead of using complete compounds, contain fragments and chemical building blocks. Figure 1.7 shows a lead compound containing a stable, tight-binding region and a phenyl ring that should be replaced to improve receptor complementarity.

Builder-type programs require the attachment point of the weakly-binding portion as input. The software then removes the offending ligand region and uses the attachment point to create a population of derivatives by adding, deleting, and substituting fragments chosen from the component database to fill the active site. The binding energies of the resulting derivative ligands are then calculated. The structures that augment binding are retained, and those that do not are discarded. This process continues as the new population of structures is built on to generate the next round of derivatives.

The advantages of using these programs are as follows:

1. Anyone can use these programs, and no database of structures is required.
2. The component database that is used here is built into the software.
3. By the combinatorial addition of fragments, a vast number of potential derivative structures can be obtained.

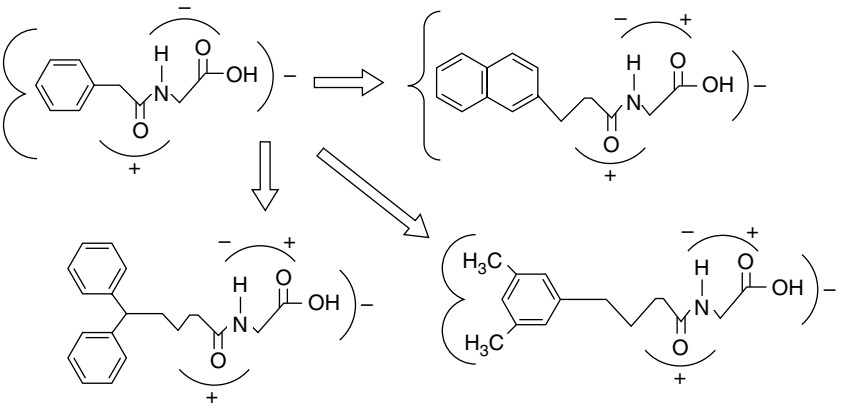


Figure 1.7 Builder-type program.

4. These programs automatically generate a diverse set of chemical solutions that contribute to the creation of truly novel ligands.
5. These programs lend themselves to the current drug development pipeline because they can be used to optimize the hits that result from mass screening.

However, the disadvantages of these programs are as follows:

1. Because of combinatorial attachment, a number of diverse chemical components arise that, in reality, may be synthetically unfeasible structures.
2. Many ligands are produced that are chemically unstable.
3. Although a diverse set of chemical building blocks is used, the manner of attachment depends entirely on the developer of the software.

1.10.3 Hybrids

Hybrid-type programs are used mostly for de novo generation of lead compounds. The strength of such programs is their ability to generate a large number of diverse potential hits. However, they suffer from all the shortcomings of the de novo design packages. The combinatorial nature of the algorithm leads to the generation of undesired chemical structures that may be unstable or synthetically unfeasible. Moreover, the developer of the software also may bias the generation of compounds.[6]

1.11 RACHEL Software Package

RACHEL stands for *real-time automated combinatorial heuristic enhancement of lead compounds* and is a drug optimization package written by Chris M. W. Ho of Drug Design Methodologies, LLC. This program is designed to optimize weakly binding lead compounds in an automated combinatorial fashion. This software comes under builder type of drug refinement program.

Through relevant discussions, we will show that RACHEL is superior to all the current computer-aided drug refinement software packages. Pharmaceutical firms are always competing with rival companies. Intellectual property in the form of patented database structures, synthetic know-how, and the biochemical data of characterized lead compounds provides a competitive edge. As such, using a program that contains both a standard database and a standard scoring function offers no such advantage over another company.

RACHEL avoids this shortcoming by allowing the firm or research laboratory to use its own intellectual property in the design of new drugs. RACHEL pulls building blocks directly from the user's corporate database and thereby offers the company an opportunity to save considerable time and money that it previously spent on the synthesis of new drugs. RACHEL software offers this tremendous advantage, and until now, no other design package employs this feature.

RACHEL's component extraction method creates a massive diversity index of the entire corporate database. Data such as the size of the component, atom composition, connectivity, ring structure, and electrostatic charge are included. Thus a means of rapidly cross-referencing chemical components on demand is available. RACHEL implements random sampling in the initial stages of lead compound optimization. RACHEL also incorporates a heuristic active-site mapping algorithm to determine the optimal chemical characteristics to complement a given region of the active site. In addition, RACHEL can determine the chemical characteristics to complement the receptor, can provide a list of candidate fragments and substitute them in a combinatorial fashion, and also can cross-reference other database components that exhibit similar characteristics.

User-directed structure generation is the most powerful and unique feature of RACHEL that truly allows the application of combinatorial chemistry. RACHEL has developed a component-specification language that contains a combination of keywords, target values, and Boolean operators. Moreover, RACHEL filters the master component database and generates individual databases for each subsite (Fig. 1.8).[7]

1.11.1 RACHEL Component Specification Language

Command	Function
$\mathrm{CMPNTS}_{\mathrm{min\ -\ max}}$	Number of total components to use
$\mathrm{ATOMS}_{\mathrm{min\ -\ max}}$	Number of atoms in a specific component
$\mathrm{R\text{-}ATOMS}_{\mathrm{min\ -\ max}}$	Number of ring atoms in a specific component
$\mathrm{MW}_{\mathrm{min\ -\ max}}$	Molecular weight
$\mathrm{BOND}_{\mathrm{(list)(<,\ >,\ =value)}}$	Specifies bonded atom types within a specific component
$\mathrm{PHARM}_{\mathrm{(atype)(x,\ y,\ z)}}$	Specifies a specific pharmacophoric group at coordinates (*x*, *y*, and *z*)

Another unique feature of RACHEL is that component specification language permits the removal of undesired structures. RACHEL is an automated method to generate diversity using templates.

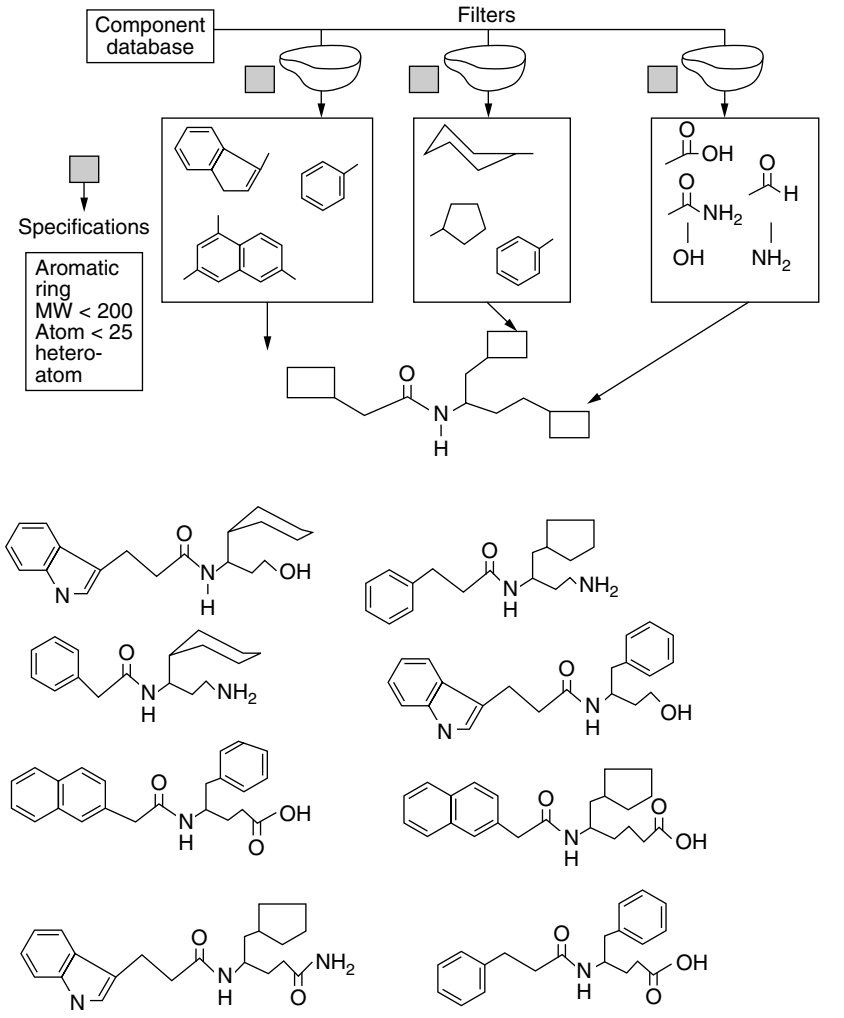


Figure 1.8 RACHEL: User-directed structure generation.

RACHEL has the unique ability to use structure-activity data and retain the competitive edge gained through research and development. By incorporating the necessary statistical and analytical tools, RACHEL allows the user to easily generate focused scoring functions to estimate ligand binding to a specific target receptor using proprietary structure-activity data. This allows companies that have characterized the receptor binding of a number of lead compound derivatives to use this knowledge in the design of future drugs.[6,7]

References

1. Fujita T, Iwasa J, and Hansch C. A new substituent constant, π, derived from partition coefficients. *J. Amer. Chem. Soc.* 86:5175–80, 1964.
2. Foye WO (ed.). *Principles of Medicinal Chemistry*, 3rd ed. Philadelphia: Lea & Febiger, 1989.
3. Leo A, Hansch C, Elkins D. Progress in medicinal chemistry. *Chem. Rev.* 71:525, 1971.
4. Wolff ME. *Burger's Medicinal Chemistry and Drug Discovery*, New York: Wiley, 1995.
5. Smith RN, Hansch C, Ames MM. Selection of a reference partitioning system for drug design work. *J. Pharm. Sci.* 64:599–606, 1975.
6. Hansch C. *Qualitative Structure-Activity Relationship in Drug Design*, Vol. 1. New York: Academic Press, 1971, pp. 271–342.
7. Cramer RD 3rd, Snader KM, Willis CR, et al. Application of quantitative structure-activity relationships in the development of the antiallergic pyranenamines. *J. Med. Chem.* 22:714–725, 1979.

CHAPTER 2

Drug Delivery Systems

2.1 Introduction

The goal of any drug delivery system is to provide a therapeutic amount of drug to the proper site in the body to achieve promptly and then maintain the desired drug concentration. Thus the drug delivery system should deliver the drug at a rate dictated by the need of the body over the period of treatment.[1]

Two aspects are very important to drug delivery, namely, spatial placement and temporal delivery of the drug. *Spatial placement* relates to targeting the drug to a specific organ or tissue, whereas *temporal delivery* refers to controlling the rate of drug delivery to the target tissue.

Drug delivery technologies are aimed at modulating the pharmacokinetics of a drug. The two important approaches are

1. To modify the molecule itself (either covalent, complexation, or physical state)
2. To modify the encapsulation of the active compound

Technologies also differ on the type of the molecules they deliver, the applicable routes of administration, and their overall effect on the pharmacokinetic profile. Manufacturing of a therapeutic agent includes scale-up production, purification, and packaging.

In this present era, science and technology have placed their maximum emphasis on the development of sustained-release pharmaceuticals, and this matter continues to be the focus of a great deal of attention in both industrial and academic laboratories. There currently exists numerous products on the market that are formulated for both oral and parenteral routes of administration that claim sustained or controlled drug delivery. Although current research is directed toward the oral dosage forms that satisfy the temporal aspect of drug delivery, newer approaches are under investigation that allow for spatial placement.[2]

Studies have shown that conventional drug therapy, given either by injection or through an extravascular route, does not maintain blood drug levels within the therapeutic range for extended periods of time. This short duration of action is due to the inability of conventional dosage forms to control temporal delivery. Hence, to maintain blood drug levels in the therapeutic range for longer periods, different techniques must be adopted. An attempt can be made to increase the initial dose, but this may result in early toxic levels. Obviously, such an approach is undesirable and unsuitable. An alternative approach is to administer the drug repetitively using a constant dosing interval, as in multiple-dose therapy.[3] However, several potential problems are inherent in multiple-dose therapy:

1. If the dosing interval is not appropriate for the biologic half-life of the drug, large peaks and valleys in the blood drug level may result.
2. The blood drug level may not be within the therapeutic range at sufficiently early times, an important consideration for certain disease states.

Moreover, patient noncompliance with the multiple-dosing regimen can result in failure of this approach. Thus there have been compelling motives to strive for the development of sustained-release drug delivery systems.

2.2 Development of Sustained-Release Drug Delivery Systems

For the sake of discussion, the conventional dosage forms (e.g., capsules, tablets, solutions, emulsions, aerosols, etc.) can be considered to release their active ingredients immediately into the absorption pool. For non-immediate-release dosage forms, though, release of drug from the dosage form is the rate-limiting step.[4]

Non-immediate-release delivery systems can be divided conveniently into four categories:

1. Delayed release
2. Sustained release
   a. Controlled release
   b. Prolonged release
3. Site-specific release
4. Receptor release

2.2.1 Delayed-Release Systems

These systems use repetitive intermittent dosing of a drug from one or more immediate-release units incorporated into a single dosage form (e.g., repeated action tablets and capsules or enteric-coated tablets). A delayed-release dosage form, though, does not produce uniform blood levels within the therapeutic range but is more effective for patient compliance than conventional dosage forms. *Site specific* and *receptor release* refer to targeting of a drug directly to a certain biologic location. In the case of site-specific release, the target is adjacent to or in the diseased organ or tissue; for receptor release, the target is the particular receptor for a drug within an organ or tissue. Both these systems satisfy the spatial aspect of drug delivery.

Sustained-release systems include any drug delivery system that achieves slow release of drug over an extended period of time. If the system can provide some control, whether it is of a temporal or a spatial nature or both, of drug release in the body or, in other words, the system is successful at maintaining constant drug levels in the target tissue or cells, it is considered a *controlled-release system*.[4,5] If it is unsuccessful at this but nevertheless prolongs therapeutic blood or tissue levels of the drug for an extended period of time, it is considered a *prolonged-release system*.

The objective in designing a sustained-release system is to deliver drug at a rate necessary to achieve and maintain a constant blood drug level. This rate should be analogous to that achieved by continuous infusion, where a drug is provided to the patient at a constant rate just equal to its rate of elimination. This implies that the rate of delivery must be independent of the amount of drug remaining in the dosage form and constant over time (Fig. 2.1). That is, release from the dosage form should follow zero-order kinetics, as given by

$$K_r^o = \text{rate in} = \text{rate out} = K_e C_d V_d$$

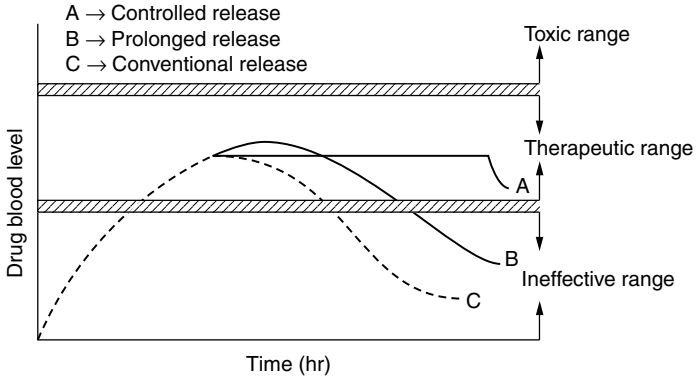


Figure 2.1 Sustained drug delivery system.

where K_r^o = zero-order rate constant for drug release
K_e = first-order rate constant for overall drug elimination
C_d = desired drug level in the body
V_d = volume space in which the drug is distributed

2.2.2 Advantages of Sustained Drug Delivery Systems

1. Improved efficiency in treatment
2. Improved bioavailability
3. Minimized local side effects and systemic side effects
4. Reduce fluctuation in drug level
5. Make use of special effects (e.g., sustained-release aspirin for morning relief of arthritis by dosing before bedtime)
6. Economical

For the design of any drug delivery system considerable importance is to be placed on several variables, the chief among which are

1. Route of drug delivery
2. Type of delivery system
3. Disease being treated
4. Nature of the patient
5. Length of therapy
6. Properties of the drug

These properties have the greatest effect on the behavior of the drug in the delivery system and body.[5]

2.3 Important Physicochemical Properties of the Drug

2.3.1 Aqueous Solubility and pK_a

Aqueous solubility of a drug influences its dissolution rate and hence the absorptive power of the drug. Since most drugs are either weak acids or weak bases, pK_a is also an important parameter in drug absorption. The pH of the gastric or intestinal fluid and the pH of the blood are indicative of the driving force for absorption based on pH gradient.

2.3.2 Partition Coefficient

After a drug is administered in the body and before it is eliminated, the drug molecules have to diffuse through a variety of biologic membranes, and the major criterion in evaluating the ability to

penetrate these lipid membranes is the oil/water partition coefficient defined as $K = C_o/C_w$, where C_o is the equilibrium concentration of all forms of the drug, ionized or unionized in an organic phase, and C_w is the equilibrium concentration of all forms in the aqueous phase.[5,6]

2.3.3 Drug Stability

Generally, a drug in the solid state undergoes degradation at a much slower rate than a drug in solution or suspension. The relative bioavailability of a drug that is unstable in the gastrointestinal (GI) tract can be improved by using a sustained-release drug delivery system. For a drug that is unstable in the stomach, an appropriate delivery system would deliver the contents to the intestine. The reverse is the case for drugs that are unstable in the intestine. However, sustained-release delivery systems are designed in such a way that they release their contents over the entire length of the GI tract.

2.3.4 Protein Binding

The distribution of a drug in the extravascular space is governed by the equilibrium process of dissociation of the drug from the protein. The drug-protein complex can serve therefore as a reservoir in the vascular space for sustained drug release to extravascular tissues, but only for drugs that exhibit a high degree of binding. Thus the protein-binding characteristics of a drug can play a significant role in its therapeutic effect regardless of the dosage form.[5,6]

2.4 Biologic Properties

2.4.1 Absorption

To design a definite drug delivery system, the rate, extent, and uniformity of absorption of a drug are important considerations. The rate-limiting step in drug delivery from a sustained-release system is not its release from a dosage form but rather its absorption. A rapid rate of absorption of the drug relative to its release is essential to make the system successful.

2.4.2 Distribution

The distribution of a drug into vascular and extravascular spaces in the body is an important factor in the drug's overall elimination kinetics. This fact influences the formulation of the drug delivery system primarily by restricting the magnitude of the release rate and the dosage size. Two parameters that are used to describe the distribution characteristics of a drug are its *apparent volume of distribution* and the *ratio of drug concentration in tissue to that in plasma at the steady state*,

the so-called T/P ratio. For drugs that obey a one-compartment model, the apparent volume of distribution is

$$V = \text{dose}/C_o$$

where C_o is the initial drug concentration.

2.4.3 Metabolism

The metabolic conversion of a drug into another chemical form must be considered in the design of drug delivery systems. For this reason, it is necessary to keep information on the location, rate, and extent of metabolism and the rate constant for the process. The two important factors associated with the metabolism of the drug are first the ability of the drug to induce or inhibit enzyme synthesis that results in a fluctuating blood drug level with chronic dosing. The other factor is the fluctuating blood drug level owing to intestinal metabolism or a hepatic first-pass effect.

2.4.4 Elimination and Biologic Half-Life

The rate of elimination of a drug is described quantitatively by its biologic half-life $t_{½}$. The half-life of a drug is related to its apparent volume of distribution V and its systemic clearance:

$$\begin{aligned} t_{½} &= 0.693V/Cl_s \\ &= 0.693V \cdot AUC/\text{dose} \end{aligned}$$

The systemic clearance Cl_s is equal to the ratio of the administered dose to the total area under the blood drug level versus time curve AUC.[3–6]

2.5 Safety Considerations

The most widely used measure of the margin of safety of a drug is its therapeutic index *TI:*

$$TI = TD_{50}/ED_{50}$$

where TD_{50} is the median toxic dose, and ED_{50} is the median effective dose. The value of TI varies from as little as unity, where the ED_{50} is also producing toxic symptoms, to several thousand. For very potent drug, whose therapeutic concentration range is narrow, the value of TI is small. The larger the value of TI, the safer is the drug.[6]

2.6 Transdermal Delivery Systems

Our skin serves as a barrier against the penetration of microorganisms, viruses, and toxic chemicals and as a restraint against loss of physiologically vital fluids. New approaches are now being adopted

to use this route for systemic drug delivery. Advanced research in the field of biomedical applications has led to the development of different types of *transdermal therapeutic systems*. Several technologies have been developed to provide rate control over the release and transdermal permeation of drugs.

2.6.1 Membrane-Moderated Systems

Membrane-moderated systems consist of a drug reservoir encapsulated in a shallow compartment molded from a drug-impermeable backing and a rate-controlling polymeric membrane. The drug molecules are released only through the rate-controlling polymeric membrane, a thin layer of compatible hypoallergenic adhesive polymer such as silicone or polyacrylate applied to achieve intimate contact of the transdermal system with the skin. Examples include nitroglycerine-releasing transdermal therapeutic systems such as Transderm-Nitro (Ciba) and scopolamine-releasing transdermal therapeutic systems such as Transderm Scop (Ciba) (Fig. 2.2).[6,7]

2.6.2 Adhesive Diffusion-Controlled Systems

In this system, the drug reservoir is formulated by directly dispersing the drug in an adhesive polymer and then spreading the medicated adhesive by solvent casting. On the top of the drug, as reservoir layer, layers of nonmedicated material, and a rate-controlling adhesive polymer of constant thickness are applied to produce an adhesive diffusion-controlled drug delivery system. Examples of this type include nitroglycerin-releasing transdermal therapeutic systems such as the Deponit system and isosorbide dinitrate–releasing transdermal therapeutic systems such as Frandol tape.

2.6.3 Matrix Dispersion-Type Systems

The drug reservoir in this system is formed by homogeneously dispersing the drug in a hydrophilic or lipophilic polymer matrix, and the medicated polymer then is molded into a medicated disk with a defined surface area and controlled thickness. The disk then is glued onto an occlusive base plate in a compartment fabricated from a drug-impermeable backing. An example of this type of system is the nitroglycerin-releasing Nitro-Dur transdermal therapeutic system.

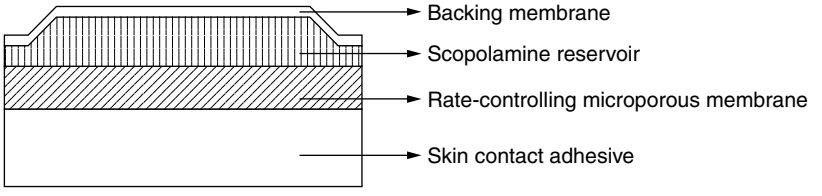


Figure 2.2 Transdermal device for the delivery of scopolamine.

2.6.4 Microreservoir Systems

In this system, the drug reservoir is formed by first suspending the drug particles in an aqueous solution of water-soluble polymer and then dispersing it homogeneously in a lipophilic polymer by high-shear mechanical force to form a large number of unleachable microscopic spheres of drug. This thermodynamically unstable dispersion is stabilized quickly by immediately cross-linking the polymer in situ, which produces a medicated polymer disk with a constant surface area and fixed thickness. An example of this type is the Nitro Disk nitroglycerin-releasing system.[7]

2.7 Targeted Delivery Systems

2.7.1 Nanoparticles

This is also known as the *colloidal drug delivery system*, and included in this group are microcapsules, nanocapsules, microspheres, liposomes, polymeric beads, etc. A nanoparticle is a particle that contains dispersed drug within a diameter of 200 to 500 nm, and because of such small sizes, nanoparticles can be easily administered through intravenous injection. The materials used for the preparation of nanoparticles should be sterile, nontoxic, and biodegradable. Nanoparticles generally are prepared by a process similar to the coacervation method of microencapsulation. The two main applications of nanoparticles are as carriers of medical diagnostic agents such as radioisotopic technetium-99m and fluorescein isothiocyanate and for the delivery of liver flukicides in veterinary medicine. The behavior of nanoparticles in vivo is the same as that exhibited by other colloidal systems of similar size and points to the possibility of using nanoparticles to target drugs to the liver and phagocytic cells. Experiments have shown the potential use of nanoparticles for the targeted delivery of anticancer agents to tumorous tissue.

2.7.2 Liposomes

When phospholipids are dispersed gently in an aqueous medium, they swell, hydrate, and spontaneously form multilamellar concentric bilayer vesicles with layers of aqueous medium separating the lipid bilayers. These systems are referred to as *multilamellar liposomes* or *multilamellar vesicles* (MLVs). Sonication or solvent dilution of MLVs results in the formation of small unilamellar vesicles (SUVs). Phospholipids can form a variety of structures other than liposomes when dispersed in water, depending on the molar ratio of lipid to water. At low ratios, the liposomes are the preferred structures.

The physical characteristics of liposomes depend on pH, ionic strength, and the presence of divalent cations. They show permeability to ionic and polar substances, including many drugs, but at elevated

temperatures they undergo a phase transition that markedly alters their permeability.

Liposomes can interact with cells by four different mechanisms:

1. Endocytosis by phagocytic cells of the reticuloendothelial system such as macrophages and neutrophils
2. Adsorption to the cell surface either by nonspecific weak hydrophobic or electrostatic forces or by specific interactions with cell surface components
3. Fusion with the plasma cell membrane by insertion of the lipid bilayer of the liposome into the plasma membrane, with simultaneous release of liposomal contents into the cytoplasm
4. Transfer of liposomal lipids to cellular or subcellular membranes without any association of the liposome contents

Liposomes can be injected, and they persist in tissues for hours or days depending on their composition and half-lives in the blood. They can exist only in places where large openings or pores exist in the capillary endothelium, such as the sinusoids of the liver or spleen. This behavior limits the potential targeting of liposomes.

Attempts to overcome the limitation on targeting of liposomes have centered around two approaches. One is the use of antibodies bound to the liposome surface to direct the antibody and its drug contents to specific antigenic receptors located on a particular cell type. The second approach is to use carbohydrate determinants as recognition sites.

Potential therapeutic applications of the liposomes include their use in the treatment of malignant tumors, lysosomal storage disease, intracellular parasites, metal toxicity, and diabetes.

2.7.3 Resealed Erythrocytes

When erythrocytes are suspended in a hypotonic medium, they swell to about one and a half times their normal size and the membrane ruptures, resulting in the formation of pores with diameters of 200 to 500 Å. The pores allow equilibration of the intracellular and extracellular solutions. If the ionic strength of the medium then is adjusted to its tonicity and the cells are incubated at 37°C, the pores will close and cause the erythrocytes to "reseal." Using this technique with a drug present in the extracellular solution, it is possible to entrap up to 40 percent of the drug inside the resealed erythrocyte and to use this system for targeted delivery via intravenous injection.[7,8]

The ability of resealed erythrocytes to deliver drugs to the liver or spleen can be viewed as a disadvantage in that other organs and tissues are inaccessible. Thus the application of this system to targeted delivery has been limited mainly to treatment of lysosomal storage diseases and metal toxicity, where the site of drug action is the reticuloendothelial system.

2.8 Transdermal Delivery Systems

The system of drug delivery that employs a skin portal to the systemic circulation at a predetermined rate and maintains clinically effective concentrations over a prolonged period of time is known as a *transdermal drug delivery system (TDDS)*. The functional course of a TDDS is determined mainly by its four central components, namely, the skin, the drug candidate, the penetration-enhancement method, and the patch design.[9]

An ideal TDDS has following characteristics:

1. *Agent independence*. The *ideal drug delivery system (IDDS)* is capable of delivering any drug, regardless of size or structure, at the specified rate of delivery.
2. *Selected delivery profile.* Drug is delivered by the IDDS as per the specified quantity-time profile.
3. *Multiple drugs.* The IDDS is capable of delivering more than one therapeutic agent at a time.
4. *Flexibility*. The IDDS has the capability of changing or adjusting the rate or timing of delivery or the quantity to be delivered.
5. *Sensoring/monitoring and decision making*. The IDDS will sense patient needs, determine the appropriate action, and deliver the necessary quantities in accordance with a calculated quantity-time delivery profile. Such a system is referred to as a smart delivery system.
6. *Targeting.* The IDDS has the capability of focusing drug transport toward target sites.
7. *Capacity*. The system is capable of making repeated deliveries before replacement.
8. *Absence of problems*. The IDDS reduces or causes no new problems or concerns.
9. *Reliability*.
10. *Marketplace value*. The IDDS offers high value by featuring maximum functionality at minimum complexity and cost.

Advantages

1. Avoidance of presystemic metabolism
2. Reduction of intra- and interpatient variability
3. Maintenance of drug level in the systemic circulation within the therapeutic window
4. Extension of drug duration of action by a single administration thereby reducing the frequency of dosing

5. Improved patient compliance and acceptability
6. Termination of drug input simply by removing the patch

Disadvantages

1. The barrier function of the skin severely constrains the absolute amount of a drug that is absorbed across a reasonable area of the skin during a closing period.
2. The system is limited to potent drug molecules, typically those requiring a daily dose on the order of 10 mg or less.
3. Even though a drug may be potent enough, it must have certain physicochemical properties such as molecular weight and lipophilicity that allow it to reach the dermal microcirculation so that it will be absorbed percutaneously. Absence of either oil or water solubility will preclude desirable permeation.
4. The pharmacokinetic and pharmacodynamic characteristics of the drug must be such that the relatively sustained and slow input provided by the TDDS produces the desirable therapeutic effect.
5. Tolerance-inducing compounds are not good choices for this mode of the administration.
6. Locally irritating or sensitizing orally administered drugs with a reproducible bioavailability need not be administered with the TDDS.

2.9 Basic Components of Transdermal Devices

Transdermal devices are designed to support the passage of drugs from the surface of the skin through its various layers and into the systemic circulation.[9–11] There are two basic types of transdermal systems:

1. Those which control the rate of drug delivery to the skin
2. Those which allow the skin to control the rate of drug absorption

The different components of transdermal systems include

- Polymer matrix or matrices that reserve and regulate drug release and drug absorption/permeation
- Enhancers
- Excipients, which are adhesives to register the preparation topically
- Polymer matrix

The polymers used in the preparation of various components of a TDDS should fulfill the following requirements:

1. The molecular weight, physical characteristics, and chemical functionality of the polymers must allow the diffusion of drug at the desired rate.
2. The polymer should be chemically nonreactive or be an inert drug carrier.
3. The polymer should not decompose on storage or during the life of the device.
4. The polymer and its decomposed product should be nontoxic.
5. The device should be easy to manufacture and fabricate into a desired product.
6. The cost of polymer should not be excessively high.

Widely used polymers include polypropylene, polyvinyl carbonate, cellulose acetate, polyacrylonitrile, hydroxy propyl cellulose (HPC) ethylene vinyl acetate, and polyethylene terephthalate.

The different components of a TDDS include

1. Foil cover strip
2. Drug matrix
3. Release liner
4. Foil base plate
5. Microporous tape
6. Adsorbent pad
7. Occlusive overlay

2.10 Choice of Drug

To prepare a successful transdermal product, a judicious choice of drug is essential. The important drug properties to be taken into consideration are molecular weight, chemical functionality, and physical properties.

Drugs with high oil/water partitioning result in selective localization and poor permeation owing to lipid affinity. Similarly, drugs with high hydrophilicity fail to permeate the hydrophobic skin barrier. Some drugs that suffer skin metabolism and lose their pharmacodynamic activity should be formulated as transdermal systems. However, drugs with a high molecular weight are poor candidates for transdermal systems because they fail to react at therapeutically desired systemic level. Again, the drugs for a transdermal system should be nonirritating and nonallergenic to human skin.[11]

2.11 Excipients

Skin permeability enhancers and excipients promote skin permeation and thus are regarded as an integral part of transdermal formulations. The penetration enhancers are broadly classified into three categories: lipophilic solvents, surface-active agents, and two-component systems. Dimethyl sulfoxide (DMSO) increases permeation of lipophilic drugs possibly by affecting their continuous lipophilic pathway of the skin.

Surfactants also enhance the permeation of hydrophilic drugs. Two-component systems are very effective permeation promoters. They are composed of oleic acid and propylene glycol. However, in commercially available transdermal systems, use of enhancers in not indicated; excipients can be incorporated additionally that facilitate and support the permeation.

2.12 Adhesive and Packaging

Adhesives are necessary in all-transdermal systems to attach them to the skin, and attachment is accomplished through the use of a pressure-sensitive polymeric adhesive. The adhesive system should have the following characteristics:

1. It should not cause irritation, sensitization, or imbalance in the normal skin flora.
2. It should strongly adhere to the skin and should be resistant to routine disturbances such as bathing, clothing, abrasion, and exercise.
3. It should be easily removable.
4. It should have intimate contact with the skin at both the macroscopic and the microscopic levels.

The pressure-sensitive adhesive should be positioned on the face of the delivery system or at the back of the device. A face adhesive–based transdermal system should possess some additional characteristics, including the following:

1. It should be physically and chemically compatible with the drug and other dosage-form excipients.
2. It should not interfere with the permeation characteristics of the drug.
3. It should allow the delivery of simple or blended percutaneous absorption enhancers.
4. The adhesive properties should not deteriorate with aging.

2.13 Pressure-Sensitive Adhesives

A *pressure-sensitive adhesive* generally is defined as a material that adheres to a substrate when a light pressure is applied and leaves no residue when removed. In transdermal systems, this type of adhesive is required to maintain intimate contact of the system with the skin surface.[10,11] Pressure-sensitive adhesives must be evaluated for general adhesive properties as well as for dermal toxicity and human wear.

Pressure-sensitive adhesive can be evaluated on the basis of their three basic properties: peel adhesive, tack, and shear strength.

2.13.1 Peel Adhesion Properties

The force required to remove an adhesive coating from a test substrate is referred to as its *peel adhesion*. The molecular weight of an adhesive polymer and the type and amount of additives (i.e., tacifiers and polymer compositions) are the variables that determine peel adhesion properties. Adequate adhesion to skin and the nontraumatic removal of the system depend on these properties.

A single tape is applied to a stainless steel plate or backing material of choice, then the tape is pulled from the substrate at an angle of 180 degrees, and the force required for tape removed is measured. This force is expressed in ounces per inch width of the tape, with higher values indicating greater bond strength. If the pulled tape does not leave any residue on the plate, this indicates *adhesive failure*. If some residue is left behind, this suggests *cohesive failure*, which often signifies a lack of cohesive strength.

2.13.2 Tack Properties

This is the ability of a polymer to adhere to a substrate with little contact pressure. Molecular weight, composition of the polymer, and tackifying resins affect the tack properties of transdermal systems. In the case of a transdermal system that is applied with finger pressure, tack is an important property. The four generally used tests for tack determination are the thumb tack, rolling-ball tack, quick-stick tack, and probe tack tests.

Thumb Tack Test

This is a qualitative test applied for tack property determination of adhesives. In this test, the thumb is simply pressed on the adhesive, and the relative tack property is measured. Based on experience, one can differentiate between relative degrees of tack.

Rolling Ball Tack Test

This test measures the softness of a polymer that relates to tack. In this test, a stainless steel ball of 7/16 inches in diameter is released on

an inclined track so that it rolls down and comes in contact with horizontal, upward-facing adhesive. The distance the ball travels along the adhesive is the measurement of tack, usually expressed in inches. The less tacky the adhesive, the further the ball will travel.

Quick-Stick (Peel) Tack Test

In this test, the tape is pulled away from the substrate at 90°C at a speed of 12 inches/minute. The peel force required to break the bond between the adhesive and the substrate is measured and recorded as tack value, expressed in ounces per inch of width. The higher values of force required indicate a higher degree of tack.

Probe Tack Test

In this test, a probe tack tester is used. The tip of a clean probe with a defined surface roughness is brought into contact with adhesive, where a bond is formed between the probe and the adhesive. The subsequent removal of the probe mechanically breaks the bond. The force required to pull the probe away from the adhesive at a fixed rate is recorded as tack (expressed in grams).

2.13.3 Shear Strength Properties

Shear strength is a measurement of the cohesive strength of an adhesive polymer. If a transdermal device has adequate cohesive strength, it will not slip after application and will leave no residue on removal. This can be influenced by molecular weight, degree of cross-linking, and composition of the polymer. In this particular test, adhesive-coated tape is applied onto a stainless steel plate. A specified weight is hung from the tape to pull it in a direction parallel to the plate. Shear strength is determined by measuring the time it takes to pull the tape off the plate. The longer the time taken for removal, the greater is the shear strength.[11]

2.14 Transdermal Devices

Based on components, delivery systems essentially can be classified as

- Membrane moderated
- Adhesive diffusion controlled
- Matrix dispersion type
- Microreservoir type

2.14.1 Membrane-Moderated Devices (Reservoir Devices)

In this system, the drug reservoir is encapsulated in a shallow compartment molded from a drug-impermeable metallic-plastic lamination, and the drug delivery side is covered by a controlling polymeric

membrane. The drug molecules are released only through the rate-controlling membrane. In the drug reservoir compartment, the solid-polymer drug matrix is suspended in an unleachable viscous fluid forming a pastelike suspension. On the external surface, a thin layer of drug-compatible adhesive polymer may be applied for an intimate delivery phase to skin contact.

The rate of drug release from this type of system can be tailored by varying the polymer composition, permeability coefficient, or thickness of the rate-controlling membrane. Several transdermal systems have been developed successfully using this technology. These include the Transderm-Scop for 3 days of protection from motion sickness and the Transderm Nitro for once-a-day medication for **angina pectoris.**[6–9]

Figure 2.3 shows that the rate of drug release can be controlled by varying the polymer composition, the permeability coefficient, and the thickness of the rate-controlling membrane and adhesive. The intrinsic rate of drug release from this type of system is defined by the relation

$$\frac{dQ}{dt} = \frac{C_R}{\frac{1}{P_m} + \frac{1}{P_a}}$$

where C_R is the drug concentration in the reservoir compartment, and P_a and P_m are the permeability coefficients of the adhesive layer and

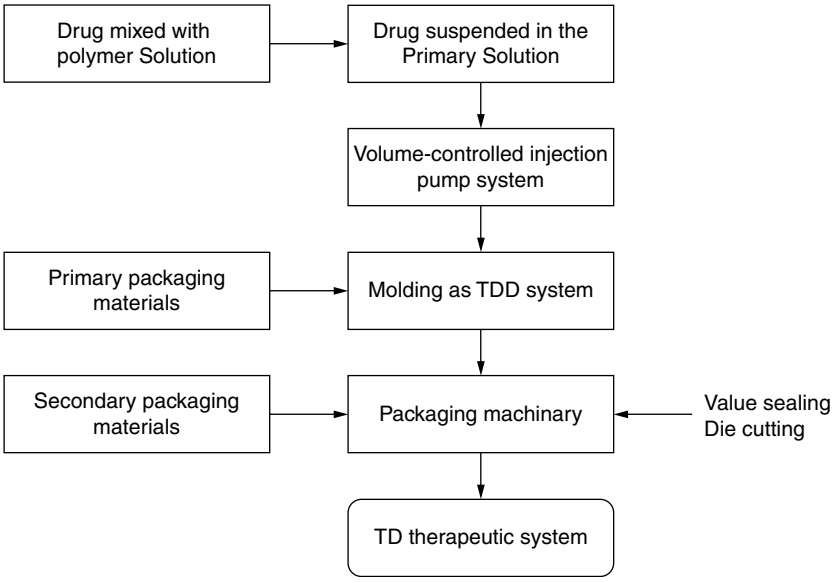


Figure 2.3 Representation of the process: Membrane-moderated TDDS.

rate-controlling membrane, respectively. For a microporous membrane, P_m is essentially the sum of permeability coefficients for simultaneous permeation across the pores and the polymeric material.

The coefficients P_m and P_a are defined as follows:

$$P_a = \frac{K_{a/m}D_a}{h_a} \qquad \text{and} \qquad P_m = \frac{K_{m/r}D_m}{h_m}$$

where $K_{m/r}$ and $K_{a/m}$ are, respectively, the partition coefficients for the interfacial partitioning of drug from the reservoir to the membrane and from the membrane to the adhesive. Similarly, D_m and D_a are the respective diffusion coefficients in the rate-controlling membrane and the adhesive layer, and h_m and h_a are the thicknesses of the rate-controlling membrane and adhesive layer, respectively.

2.14.2 Adhesive Diffusion-Type Devices (Mixed Monolithic Reservoir Devices)

This type of delivery system is a simplified form of the membrane-moderated drug delivery system. The drug reservoir is prepared by directly dispensing the drug in an adhesive polymer and then spreading the medicated adhesive by solvent film casting method over a flat sheet of drug-impermeable metallic/plastic backing membrane. This forms a thin drug reservoir layer.[6,7] The drug reservoir layer then is covered with a nonmedicated rate-controlling drug delivery system. This system is also applied for controlled administration of verapamil. The rate of drug release is defined by

$$\frac{dQ}{dt} = \left(\frac{K_{a/r}D_a}{h_a}\right)C_R$$

where $K_{a/r}$ is the partitioning coefficient.

2.14.3 Matrix Dispersion-Controlled Device (Monolithic Device)

These devices are formed by homogeneously dispersing the drug in a mixture of hydrophilic/lipophilic polymer (matrix), and the medicated polymer is molded on a medicated disk of defined surface area and thickness (Fig. 2.4). This drug reservoir–containing polymer disk is glued over an occlusive base plate consisting of a compartment fabricated using an impermeable plastic backing. Instead of applying the adhesive polymer on the surface of the medicated disk, it can be applied as an adhesive rim around the medicated disk. The rate of drug release of this system is given as

$$\frac{dQ}{dt} = \left(\frac{AC_pD_p}{2t}\right)^{1/2}$$

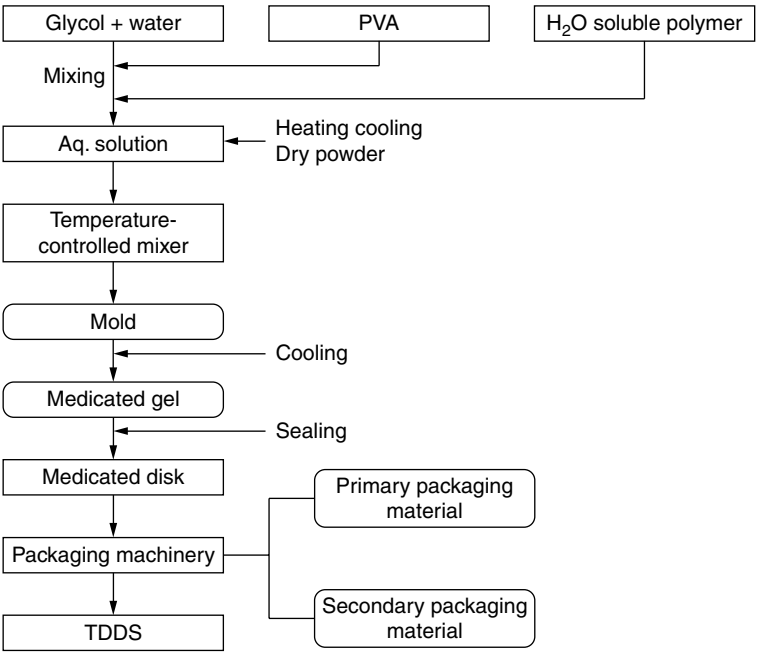


Figure 2.4 Matrix diffusion-controlled TDDS.

where A is the initial drug loading doses, and C_p and D_p are solubility and diffusibility of the drug in polymer, respectively.

2.14.4 Microreservoir-Type Devices

These devices have essentially the same features as both reservoir and matrix dispersion-type devices (Fig. 2.5). Here, the drug reservoir is formed by suspending the drug solid in an aqueous solution of water-soluble polymer. The drug suspension then is dispersed homogeneously in a lipophilic polymer by high-shear mechanical agitation to form thousands of unleachable microspheres of the drug. The rate of drug release from this type of TDD system can be defined quantitatively by following equation:

$$\frac{dQ}{dt} = \frac{D_p D_s \alpha^1 K_p}{D_p \delta d + D_s \delta s \alpha^1 / K_p} \times \left[\beta S_p - \frac{D_1 S_1 (1-\beta)}{\delta_1} (1/K_1 - 1/K_m) \right]$$

where $\alpha^1 = \delta^1/\beta^1$; δ^1 is the ratio of drug concentration in the bulk of the elution solution to drug solubility in the same medium; β^1 is the ratio of drug concentration at the outer edge of the polymer coating membrane to drug solubility in the same polymer composition; K_1, K_m, and K_p are the partition coefficients; D_1, D_p, and D_s are the drug diffusibilities; S_1 and S_p are the solubilities of the drug in the liquid

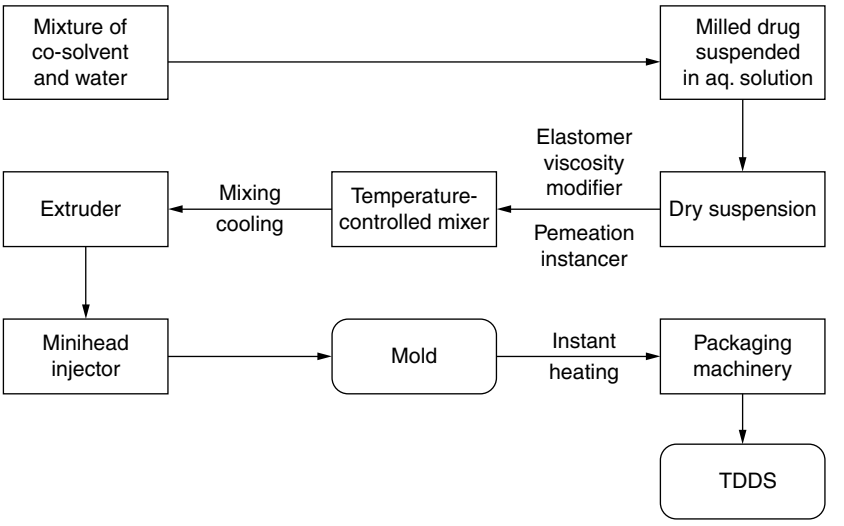


Figure 2.5 Microreservoir dissolution-controlled TDDS.

compartment and in the polymer matrix; δ_1, δ_p, and δ_d are the thicknesses of the liquid layers surrounding the drug particles, the polymer coating membrane around the polymer matrix, and the hydrodynamic diffusion layer surrounding the polymer coating membrane; and β is the ratio of the drug concentration at the inner edge of the interfacial barrier to the carry and solubility of drug in the polymer matrix.

2.14.5 Microreservoir Dissolution-Controlled Devices

Transdermal devices can be evaluated for their drug release and permeation kinetics by using a two-compartment diffusion cell assembly. The different types of diffusion cells can be summarized on the basis of their physical design and methods of sampling and measurement. Among the various diffusion cells used, Franz diffusion cell is the most commercialized.

2.15 In Vitro Release Kinetics

Controlled-release kinetics of drugs from technologically diffusion transdermal therapeutic systems can be evaluated and compared by using a Franz diffusion cell assembly. The release of nitroglycerin from the Transderm—Nitro is three times greater than that from the Deponit system. It was suggested that the rate-controlling adhesive layers in the Deposit system play a greater role in the release of nitroglycerin than does the rate-controlling membrane in the Transderm—Nitro system. Similarly, the release flux of nitroglycerin from the Nitro—Dur system is about two fold greater than the Nitro—Disk system.[3,7]

2.16 In Vitro Skin-Permeation Kinetics

In vitro permeation kinetics studies can be performed on hairless mouse skin or human cadaver skin by using either a Franz diffusion cell or a two-reservoir diffusion cell. In two-reservoir diffusion cells, sink conditions must be maintained. The permeation of nitroglycerin across human cadaver and hairless mouse skin from different transdermal therapeutic systems was compared for their kinetics. It was noted there the rate of skin permeation across the excised skins of hairless mouse matched the data obtained from human cadaver skin, suggesting that hairless mouse skin could be an acceptable animal model for human skin permeation kinetics study.

2.17 IV-IV Correlation

To examine the feasibility of using freshly excised hairless mouse skin as an animal model for studying the transdermal controlled permeation kinetics of the drug across human skin, the in vivo rate of skin permeation can be determined for comparison. It can be calculated from steady-state plasma level data $(C_p)_{ss}$ by using the relationship

$$(Q/t)_{\text{i.v.}} = (C_p)_{ss}{}^{K_e V_d}$$

where K_e is the first-order elimination rate constant.

2.18 Optimization of Transdermal Systems

To formulate a TDDS, the thing to be taken into consideration is the relationship between the rate of drug delivery R_d to the skin surface and the rate of skin absorption R_a of the drug.[5,6] Ideally, a TDDS should be designed to have a transdermal permeation rate determined based on rate of drug delivery from the TDDS and not by skin permeability to the drug to be delivered. In such a case, the transdermal bioavailability of the drug becomes less dependent on intra- and interpatient variability in skin permeability. The rate of transdermal permeation of a drug at steady state $(R_p)_{ss}$ is mathematically related to $(R_d)_a$—actual rate of drug delivery from the TDDS—and to maximum achievable rate of skin adsorption $(R_a)_m$ by the relation

$$\frac{1}{(R_p)_{ss}} = \frac{1}{(R_d)_a} + \frac{1}{(R_a)_m}$$

This equation also helps to determine the actual rate of drug delivery from a TDDS.

2.19 Skin Irritation Sensitization

These two factors must be assayed:

1. Severe skin irritation may affect the efficacy or safety of the product.
2. Dermatologic adverse effects must be evaluated.

Separate skin irritation and sensitization studies must be done for this purpose.

References

1. *Physician's Desk Reference*, 48th ed. Oradell, NJ: Medical Economics, 1994.
2. Chien YW. *Novel Drug Delivery Systems*. New York: Marcel Dekker, 1982.
3. Poznansky MJ, Juliano RL. Biological approaches to the controlled delivery of drugs: A Critical Review. 36:277–236, 1984.
4. Lee VHK, Robinson JR, Lee VHL. *Controlled Drug Delivery*, 2nd ed. New York: Marcel Dekker, 1987, p. 3.
5. Urquhart J. *Controlled Release Pharmaceuticals*. Washington, DC: American Pharmaceutical Association, 1981.
6. Higuchi, T. Application of the Higuchi model for drug release from dispersed matrices to particles of general shape. *J Pharm Sci*. 52:1145, 1963.
7. Chandrasekaran SK, Hillman R. Heterogeneous model of drug release from polymeric matrix. *J Pharm Sci.* 69:1311, 1980.
8. McGinity JW, Cameron CG, Cuff GW. A controlled release matrix using a mixture of hydrophilic and hydrophobic polymers. *Drug Dev Ind Pharm.* 9:57, 1983.
9. Beckett AH, Stanforth DH, Raisi A. Therapeutic applications of injectable poly(*ortho*)esters. *Drug Dev Ind Pharm*. 6:121, 1980.
10. Hsieh DST, Smith N, Chien YW. In Meyers WE, Dunn RL, eds., *Proceeding of 11th International Symposium on Controlled Release of Bioactive Materials*. Chicago: Controlled Release Society, 1984, p. 134.
11. Inlet G. In Gregoriadis E, ed., *Drug Carriers in Biology and Medicine*. London: Academic Press, 1979, p. 129.

This page intentionally left blank

CHAPTER 3

Bioinformatics of Drug Molecules and Databases

3.1 Introduction

The word *bioinformatics* is of recent invention, not appearing until around 1991. With the progress of time, however, computers have become an important part of the biologic research because of their unique capacity to handle large quantities of data and probe the complex dynamics observed in nature. Bioinformatics can be defined as the application of computational techniques to understand and organize the information associated with the biologic macromolecules. Bioinformatics is an interface between the biologic and computational sciences. It is an application of computational techniques that helps us to understand and organize the information associated with biologic macromolecules. It conceptualizes biology in terms of molecules and applies informatics techniques such as applied math, computer science, and statistics. Bioinformatics is an integration of computer, mathematical, and statistical methods to manage and analyze biologic information. Indeed, bioinformatics has become a unique tool for biologists.[1,2]

From a biologist's point of view, bioinformatics is a set of tools that allows the scientist to see the differences in DNA sequence between individuals that affect the relationship between disease and polymorphism. In today's world, bioinformatics allows us to collect data faster than they are interpreted. In bioinformatics, the computational techniques of information technology are used to interpret biology.[2] These techniques allow us to

- Assign functions to proteins
- Mine data

- Compare protein–protein interactions in the different protein families
- Transform gene data to protein structure and correlate gene and protein function

3.2 The Concepts, Goals, and Aims of Bioinformatics

Bioinformatics is a fast-growing discipline that deals with the storage, analysis, integration, and simulation of molecular biologic data. It can be regarded as a cutting-edge technology and a knowledge revolution.

The major goal of bioinformatics is to obtain the complete sequence of as many different genomes as possible. This genome sequence information is of great importance to multinational corporations as well as the research organizations so that they can play biology in the *in silico* environment. The information derived from this research helps scientists in designing and analyzing more advanced experiments.[2,3] Thus the effect of modern bioinformatics has been to cause a fundamental shift in the approach of the scientists to molecular biology.

The aims of bioinformatics are threefold. First, bioinformatics organizes data in such a way that it allows researchers easy access to existing information and creates new entries such as the Protein Data Bank (PDB). The PDB is designed for three-dimensional (3D) macromolecular structures. Bioinformatics helps scientists to study the data already stored and to develop the tools to analyze themselves.

Second, bioinformatics aims to develop the tools and resources that aid in the analysis of data. For example, in order to sequence a particular protein, it is often necessary to compare it with previously characterized sequences. In such a case, one needs a simple text-based search and the use of such programs as FASTA and the PSI-BLAST (basic local alignment tool). Now development of such type of programs requires the knowledge of both computational theory and biology.[4]

The third aim is to use this tool to analyze data and interpret the results in a biologically meaningful manner. With the help of bioinformatics global analyses can be carried out that can be applied across many systems and highlight novel features.

The information sources in bioinformatics can be divided into raw DNA sequences, protein sequences, macromolecular structures, genome sequences, and other whole-genome data.[4,5]

3.3 Fields of Bioinformatics: Role and Relevance in Biotechnology

The interest of the pharmaceutical industries in the genome sequencing projects has made bioinformatics increasingly important. Bioinformatics is increasingly important for medical diagnostic and

therapeutic uses and has wide industrial applications. This rapidly evolving field is now very challenging for biotechnology professionals to keep up with.[5,6]

The field has evolved to deal with four distinct problems, namely:

1. Handling and managing biologic data, including organization, control, linkages, and analysis.
2. Communicating with the people and institutions engaged in biologic research and applications. Such communications include e-mail, file transfers, remote logins, electronic bulletin boards, and so on.
3. Organizing, accessing, searching, and the retrieving biologic information, documents, and literature.
4. Analyzing and interpreting biologic data via computational approaches, including visualization, modeling, and simulation, and developing algorithms for highly parallel processing of complex biologic structures.[6–8]

The following thrust areas of biotechnology are fully dependent on bioinformatics:

Genomics. The genomics studies in bioinformatics have concentrated on model organisms and the analysis of regulatory systems.

Comparative genomics. A successful bioinformatics application implies functional correlation that assumes a similarity between two sequences, whether it is DNA, RNA, or protein.

Functional genomics. This deals with calculation of the 3D structure of a macromolecule based on its sequence, providing the theoretical basis of relationships between sequences and structure, one of the most fundamental problems in *in silico* biology. It also deals with folding problems, a demanding objective of computational biology.

Pharmacogenomics. In drug design, the basic concept to be kept in mind is that the therapeutic effect of a drug is to be maximized and the side effects of the drug minimized. It therefore would be very convenient to personalize drugs for each patient. The genetic variations among all human constitute only 0.1 percent of total DNA. These differences are mostly *point mutations.* These so-called single-nucleotide polymorphisms (SNPs) become good candidates for drug development. Investigations of SNPs using microarrays made examination of the entire population feasible. Databases of SNPs reveal the causes of cancer and so on.

Cellomics. For carrying out investigations on complex interactions, the availability of sufficient data and identification of the relevant components of life are very essential. Simulations of life

in silico constitute a future direction for bioinformatics that has just begun.

Nucleotide and genome sequencing. The main interest of bioinformatics lies in mapping complete genome sequences for different organisms. The GenBank, European Molecular Biology Laboratory (EMBL), and DNA database of Japan (DDBJ) contain DNA sequences for individual genes that encode protein and RNA products.

Like the composite protein sequence database, the Entrez Genome Database complies sequence data from these primary databases.[5,7,8] The Entrez Genome Database brings together all complete and partial genomes in a single location. In addition to providing information about raw nucleotide sequences, it also provides

- A list of complete genomes
- A list of all the chromosomes in an organism
- Detailed views of single chromosomes, marking coding and noncoding regions

The Gene Census provides an entry point for genome analysis. The database allows the building of phylogenetic trees based on different criteria, such as ribosomal RNA or protein fold occurrence.

The cluster of orthologous groups (COGs) database classifies proteins encoded in 21 completed genomes on the basis of sequence similarity. The COG database provides the 3D domain architecture and similar functions. One of the important applications of this database is to predict the function of uncharacterized proteins and to identify phylogenetic patterns of protein occurrence (www.ncbi.nlm.nih.gov/COG).

3.4 Proteomics: Protein Sequence Databases

Protein sequence databases are categorized as primary, composite, or secondary. *Primary databases* contain over 300,000 protein sequences and function as repositories for raw data. Some common repositories are the Swiss-PROT and protein information resource (PIR) International, which describe protein functions, domain structure, and post-translational modifications. *Composite database* such as OWL and NRDB compile and filter sequence data from different primary databases to produce combined nonredundant sets that are more complete than the individual databases and also include protein sequence data. *Secondary databases* contain information derived from protein sequences and help the user to determine whether a new sequence belongs to a new known protein family. One of the most popular is *PROSITE*, a database of short sequence patterns and profiles that characterize biologically significant sites in a protein.

Structural Databases

It is sometimes very difficult to extract the information provided in individual protein databank (PDB) entries, so PDB Sum provides a separate Web page for detailed display of structural analyses, schematic diagrams, and data on interactions between different molecules in a given entry. Three major databases classify proteins by structure in order to identify structural and evolutionary relationships: the classification by class, architecture, topology, and homology (CATH), structural classification of proteins (SCOP), and fold classification based on structure-structure alignment of protein (FSSP) databases.[9–11]

3.5 Data Retrieval Systems

There are three data retrieval systems with particular relevance to molecular biologists: Entrez, the Sequence Retrieval System, and DBGET (http://genome.ad.jp/dbget/dbget2.html). The World Wide Web (WWW) is really astonishing and provides good access to biologic information. Depending on the type of data at hand, there are two basic ways to search:

1. Using descriptive words to search text databases
2. Using a nucleotide or protein sequence to search sequence databases

The three database systems tools are Entrez, SRS, and DBGET.

Entrez

This is a molecular biology database and retrieval system that was developed by the National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.goc/Entrez). This Entrez system provides access to nucleotide and protein sequence databases, a molecular modeling of 3D structures database, a genomes and maps database, and a literature database. The literature database, PubMed, provides excellent and easy access to *MEDLINE* and *pre-MEDLINE* articles.[12,13] Of the three text-based database systems, Entrez is the easiest to use.

3.6 System Retrieval Systems

The types of databases in this category include the sequence and sequence-related, metabolic pathways, transcription factors, application results, protein 3D structure, genome mapping, mutations, and locus-specific mutations databases. Over 30 versions of system retrieval systems are currently running on the WWW.[14]

3.7 Applications of Bioinformatics

One of the most recent developments in bioinformatics is *DNA microarray,* the technology that promises to monitor the whole genome on a single chip so as to get an idea of the interactions among thousands of genes simultaneously. The genome chip used in this technology is also very important in *protein-chip technology*. Base pairing or h0079 hybridization is the underlying principle of this technology. An *array* is an orderly arrangement of samples that provides a medium for matching known and unknown DNA samples based on base pairing rules, thus automating the process of identifying the unknown. The two major applications of DNA microarrays involve (1) identification of sequences and (2) determination of the expression levels of genes.

Finding homologues is a search for similarities and is the main driving force between different biomolecules. Identification of protein homologues and organization of data are some of the practical uses. For example, for a poorly characterized protein, it would be possible to search for homologues and cautiously apply some of the knowledge of the latter to the former. Specifically with structural data, theoretical models of proteins are usually based on experimentally solved structures of close homologues. Similar techniques can be used in fold recognition for predicting structures at the tertiary level.[6–13]

3.8 Bioinformatics in Rational Drug Design and the Medical Sciences

Medical applications of bioinformatics have aided rational drug design. The nucleotide sequence of a probable amino acid can be determined using translation software. Sequence search techniques are used to find homologues in model organisms based on sequence similarity. The docking algorithms can be used to design molecules that bind to the model structure, leading the way for biochemic assays to test their biologic activity on the actual protein.[3,15,16]

Further advancements in bioinformatics will revolutionize the future of health care. Postnatal genotyping of patients provides information about the susceptibility to or immunity from specific diseases and pathogens. This type of information helps researchers to formulate a unique combination of vaccines so as to minimize the health care cost of unnecessary treatments. Through proper drug design, drugs can be tailored specifically to both the patient and the disease so as to provide effective treatment with minimal side effects.

Bioinformatics technology is emerging very fast, encompassing medical information, analysis, and interpretation and combining the tools and techniques of computer science and biology, with the aim of understanding the significance of a variety of data. The hope is to achieve greater success in biomedical science, particularly in drug

discovery and in finding and curing various dreaded diseases using these computational tools.

Bioinformatics will be useful for designing experiments, for developing strategies for the testing schemes and analyzing the results, and for projecting new options and possibilities. In the near future, much of the research work in the medical sciences will be left to computers in laboratories and hospitals for the advanced treatment and better cure of patients.

The advancements in information technology (IT) have accelerated the pace of information exchange and the interface of computers with medical and biologic data.[6–8] Hundreds of small steps have been taken by the scientists in drug firms, biotech companies, and university laboratories leading into the era of genomic medicine using powerful tools. DNA chips and microarrays have led scientists to see at a glance the entire genomic era, which shows that thousands of genes are active in a given tissue sample. A number of sophisticated software programs can organize gigabytes of genetic data, and huge database of genes, disease tissue samples, and mRNA delineate the molecules that initiate the actual construction of working proteins.[6,12,13]

Computer databases are an essential component of growing collections of DNA and protein data because there is a need to apply computer-based analytical methods to the data to shore the data among several geographically separated laboratories. This is the result of the need to synthesize information from isolated literature references and to provide a means for searching biologically significant patterns to compare new sequence in reasonable time limits.

The existence of a database provides for surveys, prevention of duplication of work, cross-verification of experiments and hypotheses and so on. Computers take science beyond the traditional scientific method based on theoretical analyses and laboratory research.[8–10]

Alterations in human genes are responsible for hereditary diseases such as Huntington's disease, cystic fibrosis, sickle cell anemia, and so on. The most important step at present is to uncover the functional genome of these life forms not only in humans but also in microbes and plants. This will help us to make new discoveries about various diseases and their treatments. With this approach, scientists now can get help in identifying genes along the DNA chains of the genome. This will facilitate understand of the function of the gene and also monitor the function of the genes in the genome. Such technological conveniences are of great help in the medical sciences. The DNA chip is becoming day by day useful in the diagnosis and pathologic analysis of diseases[7] (http://bioinformatics.oupjournals.org, Bioinformatics).

The process of drug design and development is expensive and time-consuming. The application of the tools and techniques of bioinformatics has resulted in a reduction in both the cost and the development time of drugs. This has had a tremendous impact on society. If a

newly lifesaving drug is discovered, the gains achieved include not only financial savings but also the saving of the lives of millions of people.

Drug design, more specifically rational drug design, is really an innovative process. Today, computer modeling techniques play a significant role in rational drug design. Such computer modeling of drugs is known as *computer-aided drug design* (CADD). Drugs may be small organic molecules or protein-based drugs popularly called as *biologics*. Several approaches are followed for drug design. Rational approaches to drug design and development are really the major route to lead discovery. First, it is necessary to identify the cause of the disease, whether it is due to microorganisms or aberrant cell growth or to hyper/hyposecretion of an essential chemical substance in the body. Diseases owing to hyper/hyposecretion of a chemical can be rectified by antagonism/agonism of a receptor. Diseases arising from microorganisms and aberrant cell growth can be treated with drugs that will inhibit a particular enzyme or interfere with DNA biosynthesis or function. Some drugs aim to inhibit the functioning of a pathway in the diseased state by causing a key molecule to stop functioning. Drugs designed with such an approach will bind to the active site and inhibit the key molecule. Again, some approaches aim to enhance the normal pathway by promoting specific molecules in the normal pathway that may have been affected by the disease state. In the case of direct drug design or structure-based drug design, first, a knowledge of the 3D structure of the biologic target is obtained from x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy, and then, based on the structure of the biologic target, candidate drugs that will bind with high affinity and selectivity are predicted and designed using interactive graphics and the intuition of a medicinal chemist.[10,11]

Indirect drug design relies on knowledge of the pharmacophore that defines the necessary structural characteristics that a molecule should possess in order to bind to the biologic target of interest. It is also known as *ligand-based drug design*. A *pharmacophore* can be defined as a relevant group on a molecule that interacts with the receptor and is responsible for the activity of a drug molecule. The other extraneous atoms in the lead molecule are called the *auxophore*. Some of these extra atoms help the pharmacophoric groups to remain in the appropriate position and thus maintain the integrity of the molecule. Again, some extraneous atoms that interfere with binding of pharmacophore need to be excised from the lead compound. Moreover, some groups neither bind to the receptor nor prevent the pharmacophore from binding. Thus there are groups that interfere with lead compound binding and some that are not detrimental to binding. Therefore, after identifying the pharmacophoric and auxophoric groups of the lead molecule, it must be decided which group to excise, which to retain,

and which to modify. If, on the removal of a group, there is a decrease in potency, then that group is pharmacophoric; an increase in potency on removal of a group suggests that the group is auxophoric, and no change in potency suggests that the group is auxophoric but not interfering with binding. Structure-based drug design uses the 3D structures of protein and a predocked ligand in the PDB format.[17,18] Rational drug design stresses the fact that modulation of a specific biologic target should have a therapeutic value. Disease linkage studies that show an association between mutations of the biologic target and disease states will give information in this regard.

The second criterion of a biomolecule to act as drug target is that the molecule should be capable of binding to another small molecule and have its activity modulated by that molecule. Screening libraries of potential drug compounds will give information about the molecules that will bind to drug targets.

Computer-assisted drug design uses computational chemistry to study in detail drugs and related biologically active molecules. In contrast, the traditional method of drug discovery relies on trial and error of the chemical substance in microbial culture or an animal study and then matching the effects to treatment. In the entire process of drug discovery, computers may assist in hit identification, hit to lead optimization of affinity and selectivity, lead optimization, and so on. The carbonic anhydrase inhibitor dorzolamide was designed by application of structure-based drug design. Another important example of rational drug design is the remarkable anticancer drug imatinib, a tyrosine kinase inhibitor (http://psb.stanford.edu/psb-online, Pacific Symposium on Biocomputing). Major pharmaceutical and biotechnology companies, including Smithkline Beecham, Merck, Affamix Corporation, Curagen Corporation, and Glaxo Welcome, have set up large research and development groups in bioinformatics. A virtual reality laboratory for molecular and drug design has been established by Glaxo Welcome with strong support from bioinformatics.

3.9 Bioinformatics and Chemical Engineering

Biologic systems are complex entities. From an engineering point of view, they consist a large number of physico-chemical and mechanical processes that operate in parallel and in series. Understanding the functions of individual processes and their interaction is critical for advances in drug discovery as well as in industrial and agricultural biotechnology. Chemical engineering has contributed to many aspects of biotechnology research and development.

Transcription and translation are no doubt complex processes, but in some ways they are similar to the *polymerization process* that is studied in chemical engineering. Multiple mRNA and protein species

are synthesized in parallel, and they compete for the same monomers, nucleic acids, amino acids, and so on.[12,18]

Now, transformations, transcriptions, and translations are complex chemical processes. Their complexity and the uncertain latter physicochemical properties resemble those of many problems in combustion, atmospheric chemistry, polymer chemistry, and the thermodynamics of complex mixtures. The large number of the components, the thermodynamic interactions among the components, the nonlinearity of the kinetic properties, and the spatial organization of the cellular environment are some of the qualitative similarities between biologic processes and processes whose uses are central to chemical engineering.[15]

3.10 Brief Introduction to Bioinformatics Databases

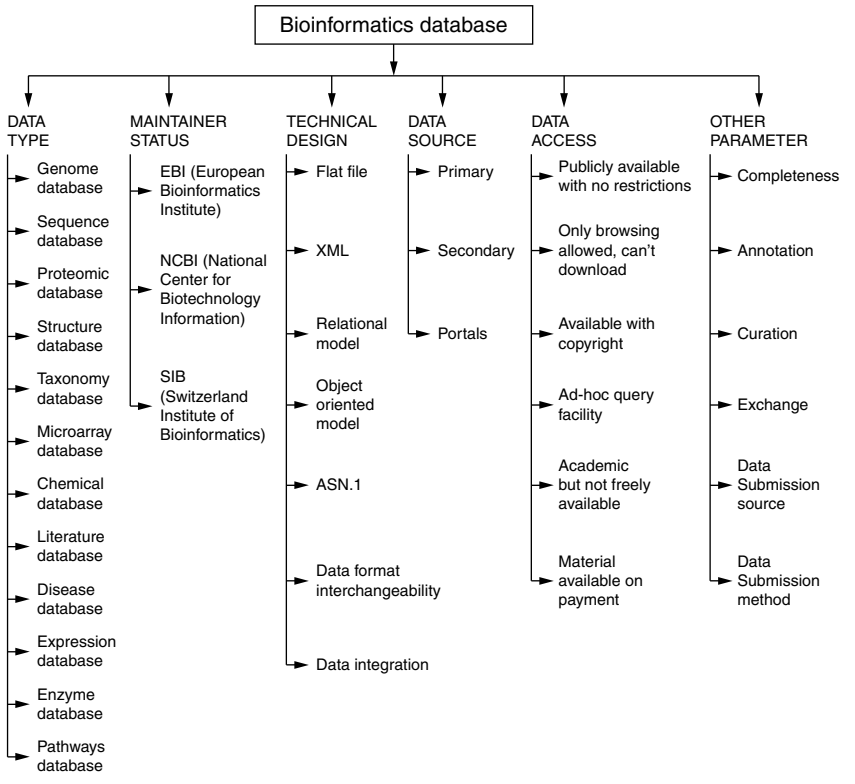
This section deals with the databases used in bioinformatics that are available today. After completion of the Human Genome Project (HGP), bioinformatics turned into a data-rich science. Databases used in bioinformatics, unravel a wealth of biologic information hidden in the jumble of DNA and RNA sequences, structures, and other sources of biologic data. When dealing with a bioinformatics database, the primary objective is to understand its criteria and its storage system and to be able to update and manipulate the database and to have knowledge about the whereabouts of the information retrieval system and the various categories of databases.[17,18]

Features of Bioinformatics Databases

1. Bioinformatics database are dynamic in nature with a constant change in the database scheme.
2. The databases hold data ranging from molecules, molecular pathways, cells, tissues, organisms, populations, etc., that is, they have a hierarchical organization.
3. The databases store 3D molecular structures, different images of cells and tissues and complex structural description of biochemical pathways.
4. Databases in bioinformatics are heterogeneous in nature.
5. These databases exclusively require Internet access as well as a search or browsing facility.
6. Because of the constant research work going on in this field, bioinformatics databases require day-to-day updating.

Types of Bioinformatics Databases

Among the different categories of bioinformatics databases, the proprietors and maintainers of databases are actually government-funded



agencies that enable free data access so as to support research work in these areas. Federal bodies such as the NCBI, EBI, and SIB maintain a pool of database on their own servers. NCBI is a federal agency whose main directive is to place in the public domain the increasing body of information in the field of molecular biology and genome research and thereby promoting scientific progress. The Entrez sites for PubMed, genome, nucleotide, protein, and structure databases are maintained in the ESA. The EBI, which is a nonprofit academic organization and a center for research and service in bioinformatics, manages huge amounts of biologic data, including nucleic acids, protein sequences, and macromolecular structure. It has already created 200 computational biology tools. The European Molecular Biology Laboratory (EMBL) represents over 185,000 organisms. EBI is funded by the United Kingdom. It is similar to the SIB, which is a quasi-academic, nonprofit foundation promoting research, teaching, and service facilities in bioinformatics either alone or with international collaborators.[17,18]

Developing a biologic database is really a daunting task for computer scientists because most of the fields are hard to define,

or there may be too many existing correlations that are likely to change owing to the fact that bioinformatics databases are developed using different database management systems and may be stored on different hardware platforms such as spreadsheet, flat files, simple text files, and so on. A *flat file* is the simplest database model, and it is easy to program and allows fast data retrieval. If new data categories need to be added, however, it requires reprogramming. Also, it is very difficult to process multiple values of a data item.

Extensible Mark Up Language (XML) is a hierarchical, semistructured model that allows easy addition or deletion of records with fast data retrieval and multiple associations with similar records in different files. Here, however, the computer path restricts access and also needs large amounts of computer storage.

The *relational model* is provided with rows and columns in a highly structured manner, and the database management system supporting the model is known as a *relational database management system.* It is flexible, easily accessible, and allows easy modification, and the physical storage of data can be changed without affecting relationships between records. However, new relations require considerable processing, and the sequential access is slow. Moreover, given the flexibility of relationships between records, logical mistakes are likely.

Object-oriented database management systems and *object-relational database management systems* have objects and classes with structured as well as abstract data types. They are very popular databases in molecular biology.

ASN.1, or *Abstract Syntax Notation 1,* was developed by the International Organization for Standardization (ISO). It is an application protocol to represent data format. It contains two basic types of syntax: *abstract syntax*, for defining information, and *transfer syntax*, for communicating information. The format of ASN.1 is simple data submission and thus is also easy.[17,18]

The two major categories of universal protein sequence databases are those which store simple archives of sequence data and those which store annotated data. The EBI/SIB Swiss-Prot and TrEMBL databases and the PIR protein sequence database (PIR-PSD) have separate protein databases. EBI, SIB, and PIR merged in 2002 to form the Uniprot (Universal Protein) Consortium. This consortium provides accurate, consistent, and rich sequence and functional annotations. This consortium facilitated biologic research by maintaining high quality database[15,16] (www.ncbi.nlm.nih.gov/BLAST).

To search a PSD, two basic methods are followed:

1. *By a particular keyword*. PIR allows a variety of keyword searches.
2. *By using search engines*.

UniProt has three components:

1. UniProt knowledge base (UniProtKB)
2. UniProt reference clusters (UniRef)
3. UniProt archive (UniParc)

UniProt KB. This is a central database of protein sequences providing accurate information on sequences and functional annotation. Its two sections are

Swiss-Prot. This contains manually annotated records with information extracted from literature and curator-evaluated computational analysis.

TrEMBL. This contains computationally analyzed records that are waiting for full manual annotation.

UniRef. This combines closely related sequences into a single record so as to speed up searches.

UniParc. This comprehensive repository consists of the unique identifiers and sequences, and the protein sequences are integrated from different publicly accessible sources such as the UniProt Consortium databases Swiss-Prot, PIR-PSD, TrEMBL, DDBJ, GenBank, PDB, and others. It also gives the history of a sequence.[17,18]

The PIR-PSD is a function-annotated database of PIR, MIPS, and JIPID at NBRF, Georgetown University. PIR—PSD is the world's first database of classified and functionally annotated protein sequences that grew out of the *Atlas of Protein Sequence and Structure,* edited by Margaret Day Hoff.

PSD data are collected from sequences in GenBank, EMBL, and DDBJ translations, published literature, and direct submissions to PIR International (http://www.prosci.uci.edu, Prosci.uci.edu, Protein science). Among the several PIR database, the noted ones are

PIR –PSD. This is the main protein sequence database. ASDB refers to Annotation and Similarity Database (www.ncbi.nlm.nih.gov/BLAST).

ALN. This is a database of protein sequences.

iProClass. This classifies proteins according to structure and function.

NRL_3D. This is a database of sequences and annotations of the proteins of known structure deposited in the PDB.

RESID. This is a database of covalent protein structure modifications.

3.11 Microarray Database

Microarrays can be specifically referred to as DNA microarrays, gene chips, DNA chips or biochips. Microarrays are actually a collection of microscopic DNA spots or probes in the form of an array attached to a

solid base that may be a glass, plastic, or a silicon chip. These DNA spots monitor thousands of gene expressions simultaneously in a single experiment. The purpose of a microarray is to measure gene expressions in diseased and normal cells so as to study and identify the disease. In May 2000, the Microarray Informatics Group at EBI was established for the purpose of managing, storing, and analyzing microarray data (http://www.fbsmres.leeds.ac.uk/users/bms8rjg/5130%20resource.html, Microarry informatics—A long road ahead Sivakumar, A).

Array Express is the public database for microarray gene expression that stores the minimum information about microarray experiments. The Array Express database contains three major components. The *Array* gives information about the design and manufacture of the array itself. The *Experiment* gives information on the experimental factors and actual data obtained. The *Protocol* describes the procedure used in the production of the array or execution of the experiment (www.Gene-chips.com) (http://www.fbsmres.leeds.ac.uk/users/bms8rjg/5130%20resource.html, Microarry informatics—A long road ahead Sivakumar, A).

3.12 Bioinformatics Scenario

This section discusses the present bioinformatics scenario in India. The roles of both government and private organizations are discussed here. In 1986, the Indian Department of Biotechnology (DBT) launched the Biotechnology Information System (BTIS) during the seventh five-year plan that is actually a nationwide network having 10 distributed information centers (DICs) and 48 distributed information subcenters (sub-DICs). The primary aim was to establish India as a leader in bioinformatics.

In the eighth five-year plan, BTIS Net established a distributed database and network infrastructure having six national facilities for high-end interactive graphics and molecular modeling in addition to those mentioned earlier. These centers are connected through satellites and terrestrial links through two major network service providers, NICNET and ERNET.

The DICs and sub-DICs established by the DBI provide information to all institutions regarding research in the field of bioinformatics. JNU New Delhi, IISC Bangalore, Bose Institute Calcutta, and Madurai Kamraj University, Madurai provide the subject genetic engineering. Plant tissue culture, photosynthesis, and plant molecular biology are provided by the Indian Agricultural Research Institute, New Delhi. The Institute of Microbial Technology, Chandigarh, deals with protein modeling and protein engineering. The University of Pune provides the virology and animal cell culture. The National Institute of Immunology, New Delhi, deals with immunology. Neuroinformatics is provided by the National Brain Research Center, Delhi. IIT-Delhi

provides computational biology and supercomputing facilities for bioinformatics. Department of biotechnology, government of India, New Delhi, is obviously specialized for biotechnology.

Side by side with government organizations, the private organizations have also come with their plans to harness Indian personnel, among which are international biotech companies such as Strand Genomics and Jubilant Biosys, who have established their offices in India. Indian companies such as Biocon India, Informatics Pvt. Ltd., and Advance Biochemical Lab are working as CROs both in India and abroad. TCS, Nicholas Piramal, and Satyam have formed collaborations with the Center for DNA Finger Printing and Diagnostic (CDFD), the Center for Biochemical Technology (CBT), the Centre for Cellular and Molecular Biology (CCMB), and Spectra Mind e-services. These professionals analyze the biologic literature and create databases. The GE call centers also work along similar lines. Companies such as Avestha Gengraine, Mahindra, BT, and DSQ software have developed into the bioinformatics service sector.

Pharmaceutical companies have greatly increased their annual incomes by setting up bioinformatics research and development (R&D) with new research horizons in the fields of proteomics, genomics, CROs for data management and remote sales and marketing. The leading pharmaceutical companies are the Ranbaxy, Dabur, Dr. Reddys, Astra Zeneca, Wockhardt, Pfizer, Zydus Cadila, Cipla, and East India Pharmaceuticals, which are rapidly setting up bioinformatics R&D. However, success in IT-related fields will be difficult and needs to focus on biologic knowledge and innovation, not just investment.

India also has geared up to improve the academic scenario in the field of bioinformatics. The different five-year plans undertaken by DBT have started a one-year advanced diploma in bioinformatics in five noted Indian Universities (JNU, CU, Pune University, Pondicherry University, and Madurai Kamraj University) through the All-India Entrance Examination. Many universities, such as Bharathiar University, have started two-year M.Sc courses in bioinformatics, and IIT Delhi is conducting certificate courses in supercomputing and has plans to offer M.Tech. degree in bioinformatics. The Bioinformatics Institute of India (BII) is the only institute that provides the scope for distance learning on bioinformatics in the country. Overall, there is a tremendous scope in the field of bioinformatics in India, and the low cost of Indian IT and biology work forces are the two key success factors.

Educational institutes in India are now offering bioinformatics as a primary degree course or as an elective subject. If computer science students opt for it, they should have basics in molecular biology, biochemistry, and evolutionary biology. On the other hand, biology students choosing it as their career need to cover programming in C, C^{++}, Java, Perl, Fortran, and Linux, and RDBMS such as Oracle/Sybase, statistical techniques, and calculus. Some of the new languages essential

for bioinformatics are BioCORBA, BioJava, BioPython, BioPerl, Cell Modeling, and so on.

Job prospects for both non-IT biology candidates as well as nonbiology science graduates are equally bright. Nonbiology bioinformatics science graduates can work as system analysts, system engineers, scientific programmers, database or network administrators, database developers, and so on. Non-IT bioinformatics science graduates can handle production and data submission and mining tolls, annotation tools, map integration tools, and so on.

India has expertise in the field of software and also maintains a good pool of qualified biologists. Infact, a joint venture of the two Indian groups can achieve a leading position in bioinformatics (http://bioinformatics.oupjournals.org, bioinformatics).

References

1. Gibas C, Beck PJ. *Developing Bioinformatics Computer Skill: An Introduction to Software Tools for Biological Application*. Ed. Mumbai, India: Shrotf, 2010.
2. Baxevanis AD, Ovellette BF. *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*. New York: Wiley, 1998.
3. Rashidi HH, Buehlar LK. *Bioinformatics Basics: Application in Biological Science and Medicine.* Boca Raton, FL: CRC Press, 2000.
4. Hatzimanikatis V. Bioinformatics and functional genomics: Challenges and opportunities. *AIChE J.* 12:2340–2343, 2000.
5. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics: An introduction. In *Yearbook of Medical Informatics*. Oxford, England: Oxford University Press, 2001, pp. 83–96.
6. Ekins R, Micro arrays: Their origins and applications. *Trends Biotech.* 17:217–218, 1999.
7. Chothia C, Gerstein M. Advances in structural genomics. *Curr Opin Stuct Biol.* 27:4658–4670, 1999.
8. Person WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA.* 85:2444–2448, 1998.
9. Reichhardt T. It's sink or swin as a tidal wave of data approaches. *Nature*. 399: 517–520, 1999.
10. Sivakumar A, Balaji R, Krishnan VG, Parish JH. *Trends in Molecular Bioinformatics*, 10th International Conference ISMB, 2002.
11. Ladunga I. PHYSEAN: Physical sequence analysis for the identification of protein domains on the basis of physical and chemical properties of amino acids. *Bioinformatics*. 15:1028–1038, 1999.
12. Brazma A, Vilo J. Gene expression data analysis. *FEBS Lett*. 480:17–24, 2000.
13. Brazma A, Robinson A, Cameron G, Ashburner M. One stop shop for microarray data. *Nature*. 403:699–700, 2000.
14. Etzold T, Ulyanoy A, Argos P. SRS: Information retrieval system for molecular biology databanks. *Methods Enzymol*. 266:114, 1996.
15. Bateman S, Birney E, Cerrut L, et al. The Pfam protein families database. *Nucleic Acids Res.* 30:276–280, 2002.
16. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol.* 215:403–410, 1990.
17. Altschul SF, Madden TL, Scholffer AA, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucluic Acids Res.* 25:3389–3402, 1997.
18. Attwood TK, Croning MD, Flower DR, et al. PRINTS-S: The database formerly known as PRINTS. *Nucleic Acids Res.* 28:225–7, 2000.

Internet Sites

http://bioinformatics.oupjournals.org, bioinformatics
http://www.prosci.uci.edu, Prosci.uci.edu, protein science
http://psb.stanford.edu/psb-online, Pacific Symposium on Biocomputing
http://www.ncbi.nlm.nih.gov, National Center for Biotechnology Information
http://www.fbsmres.leeds.ac.uk/users/bms8rjg/5130%20resource.html, Microarry informatics, A long road ahead, A. Sivakumar.
www.Gene-chips.com
www.ncbi.nlm.nih.gov/, BLAST
www.ncbi.nlm. nih.goc/, Entrez
http://genome.ad.jp/dbget/dbget2.html.
www.ncbi.nlm.nih.gov/COG

This page intentionally left blank

CHAPTER 4

Lipase- and Esterase-Mediated Drugs and Drug Intermediates

4.1 Introduction

Use of biocatalysts helps to carry out chemical processes in mild reaction conditions and with a high degree of selectivity. Synthesis of chiral drugs with the help of enzymes or preparation of chiral pharmaceuticals through biotransformations are already accepted methodologies. Lipases, esterases, oxidoreductases, and so on act as useful biocatalysts for the synthesis of chiral drugs.[1,2]

4.2 Lipase Used in Biotransformations

Lipases are a class of enzymes that catalyze the hydrolysis of long-chain triglycerides, and biotransformations are now accepted as a common methodology for the preparation of chiral pharmaceuticals. The usefulness of lipases is as follows:

1. Lipases or triacylglycerol acyl hydrolases are the most frequently used biocatalysts in organic synthesis. Lipases are able to discriminate between enantiotopic groups and between the enantiomers of a racemate. Use of such biocatalysts helps to perform enantioselective hydrolytic reactions and the formation of ester bonds. Biocatalysts efficiently catalyze various other reactions, such as amidation, aminolysis, thio transesterification, and oximolysis in organic solvents.[1]

2. Lipases can be used in asymmetric syntheses, such as the kinetic resolution of racemic alcohols, acids, esters, or amines and the desymmetrization of prochiral compounds.
3. Lipases find use in stereoselective biotransformations to carry out the kinetic resolution (KR) of mixtures and enzymatic desymmetrization (EED) of prochiral compounds.
4. Lipases in organic solvent can be used for the enantioselective preparation of intermediates for the synthesis of chiral drugs. Synthesis of chiral drugs (single isomers) occurs by aminolysis, transesterification, or enzymatic hydrolysis.[1]

4.2.1 Solvents for Biocatalysis

Biocatalysis is mostly favored in the nonaqueous medium over aqueous medium because water is a poor solvent for many organic compounds, and many compounds are also unstable in aqueous medium. Moreover, removal of water is both tedious and expensive. On the other hand, use of organic solvents offers the following advantages:

- Easy recovery of substrate and product with high yield
- Avoidance of side reactions
- The potential use of nonpolar substrates
- Shift of the thermodynamic equilibrium to favor synthesis over hydrolysis

4.2.2 Sources of Lipase

Lipases are distributed among higher animals, microorganisms, and plants where they fulfill a key role in the turnover of lipids. This is a digestive enzyme that facilitates deposition and mobilization of fat that is used as an energy reservoir within the organism and in metabolism of intracellular lipids.[3]

Of the plentiful enzymes used, over half are from fungi and yeast, and over a third are from bacteria, with the remainder divided between animal (8 percent) and plant (4 percent) sources.[4] Although pancreatic lipases have been used traditionally for various purposes, it is now well established that microbial lipases are preferred for commercial applications owing to their multifold properties, easy extraction procedures, and unlimited supply. Microbes are preferred to plants and animals as sources of enzymes for the following reasons:

1. They are generally cheaper to produce.
2. The raw materials used for enzyme production are easily available and their quality and composition can be easily identified and controlled.

3. Microbial lipases are produced by methods that can be scaled up easily to meet the current market demand.
4. The level of contaminants is much less in microbial enzymes than in plant and animal tissues.

Among the various lipase-producing microorganisms (bacteria, fungi, yeasts, and *Actinomyces*), *Candida, Pseudomonas, Rhizomucor*, and *Rhizopus* spp. stand out currently as sources of most of the commercially available enzyme preparations.[5]

4.2.3 Location of Action

Some lipases work within the interior spaces of living cells to degrade lipids. In the example of lysosomal lipase, the enzyme is confined within an organelle called a *lysosome*. Other lipases, such as pancreatic lipases, are found in the spaces outside cells and have roles in metabolism, absorption, and transport of lipids throughout the body. Since biologic membranes are integral to living cells and are largely composed of phospholipids, these enzymes play important roles in cell biology.[6] Furthermore, lipases are involved in diverse biologic processes ranging from routine metabolism of dietary triglycerides to cell signaling and inflammation.[7]

4.2.4 Structure of Lipase

Lipases belong to the class of serine hydrolases and therefore do not require any cofactor. The catalytic triad in this enzyme is composed of Ser-Asp-His, and in some lipases, Glu is present instead of Asp. Another pentapeptide consensus sequence found around the active site of serine is Gly-X_1-Ser-X_2-Gly (X_1 = histidine, X_2 = glutamic or aspartic acid).[8] In most lipases, a part of the enzyme molecule covers the active site with a short α-helix called *lid* or *flap*. The side of the α-helical lid facing the catalytic site, as well as the protein chains surrounding the catalytic site, is composed mainly of hydrophobic side chains.[8] The phenomenon of interfacial activation is often associated with the reorientation of this α-helical lid structure in the vicinity of the active site, making the active site accessible, which otherwise remains buried inside.[9]

4.2.5 Mechanism of Action

The mechanism[10] for hydrolysis or esterification is same for both lipases and consists of four steps:

Step 1. Substrate is bound to the active serine, resulting in an intermediate that is stabilized by the histidine and aspartic acid residues of the protein molecule.

Step 2. Release of alcohol and formation of an acyl enzyme complex.

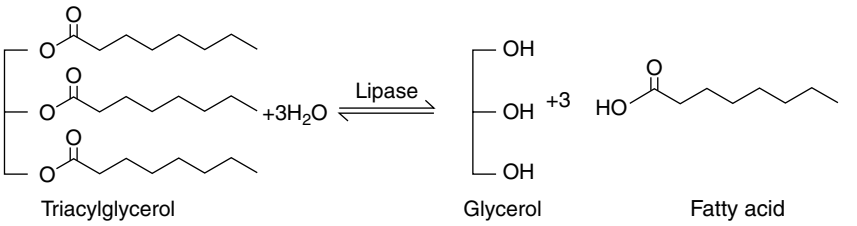


Figure 4.1 The catalytic action of lipase.

Step 3. Attack of nucleophile (water in hydrolysis, alcohol or ester in esterification and transesterification).

Step 4. Yield of product and free enzyme after resolution.

Most lipases act at a specific position on the glycerol backbone of a lipid substrate and convert triglyceride substrates to monoglycerides and free fatty acids, as shown in Fig. 4.1.

4.2.6 Production of Lipase

For production of lipases from microorganisms, two different modes of fermentation are common:

1. Submerged fermentation (SmF)
2. Solid-state fermentation (SSF)

Although SmF systems control culture conditions such as degree of aeration, pH, temperature of the medium, and so on, SSF systems have been gaining interest in recent years owing to several potential advantages such as those listed below:[11]

1. Owing to low moisture levels, bacterial growth is restricted, and hence serious contamination rarely occurs.
2. Extraction cost is less owing to low solvent requirements.
3. Microbial use of gaseous oxygen reduces the energy cost of aeration.
4. Treatment of fermented residue is simple by using the dried fermented mass as animal feed or fertilizer.

SSF has proven to be an efficient way to produce enzymes, especially by filamentous fungi, because it provides the microorganisms with environmental conditions similar to their natural habitat suggested by Holker and colleagues[12] where enzymes are produced at higher concentration and with higher temperature and pH stability under SSF conditions.

4.2.7 Statistical Design Approach for Optimization of Process Parameters and Nutritional Requirement for Lipase Production

There are many ways to approach the problem of optimization and design of a fermentation process. Traditionally, in case of biologic systems, optimization was achieved by the *one-factor-at-a-time* (OFAT) strategy. The rationale behind this strategy is to change one control variable at a time while holding the rest constant. However, the probable interactions among the factors may not be taken into account by the single-variable search technique.[13] Moreover, it is a time-consuming process and does not guarantee the achievement of optimal conditions when interactions among the factors are strong.[14]

To achieve the true optimum, a suitable alternative to OFAT is the technique of *evolutionary operation* (EVOP). The fundamental idea of the EVOP method is to incorporate a continuous or systemic scheme of changes in the control variables in place of the common static operation of a process. The effect of these changes is evaluated, and the process is shifted in the direction of improvement, if any. The EVOP methodology therefore can be considered as a multivariable search technique in which the effects of two or three factors are studied together, and the responses are analyzed statistically to arrive at the appropriate decision. The search is made sequentially, and thus it gets its name from the fact that the process slowly evolves toward the optimum.

Banerjee and colleagues[15] for the first time reported the application of EVOP methodology to design experiments for optimizing different conditions of a biologic process for the synthesis of protease. Because of its iterative nature, little needs to be known about the system before beginning of the process. In theory, the process improves at each step, making it an ideal tool for biologic process optimization. The EVOP methodology is constrained by its inability to represent a system having more than three independent variables at a time. However, n factors can be studied together by a factorial technique. Thus a hybridization of the advantages of the two methods, that is, EVOP and the factorial technique, is capable of providing a new and more efficient approach toward the optimization of an n-variable system.[16] This new methodology has the advantages of both EVOP and factorial design, where experiments are designed based on the factorial technique, and results are analyzed by EVOP procedure. The major strength of the EVOP factorial design technique is its relatively simple and clear-cut decision-making procedure that directs the change of variables to reach the maximum or minimum response.

4.2.8 Immobilization

Enzyme immobilization is a technique specifically designed to restrict the freedom of movement of the biocatalyst. Although enzymes are

becoming increasingly important in sustainable technology and green chemistry,[17] their application for a given reaction is often hampered by major limitations, such as high cost. Immobilization of an enzyme on a rigid support overcomes this limitation because the immobilized biocatalyst enables easy separation,[18] the possibility of reuse, and simple operation.

The advantages of immobilized enzymes over their free form are as follows:[19,20]

1. Immobilized lipase technology facilitates the development of continuous, large-scale commercial processes.
2. Use of immobilized lipase reduced contamination of products by residual lipase, thereby simplifying the downstream processing.
3. Immobilization permits multiple use of the enzyme and often enhances chemical and thermal stability.
4. It also enables easy separation of the enzyme from reaction mixture.
5. Use of immobilized lipase provides opportunities for better control of both the process and product quality and enables continuous operation.
6. In using immobilized lipase, the high costs of enzymatic reactions can be reduced by reuse and lower refining costs of products.

The various methods used for enzyme immobilization can be subdivided into three categories based on the interaction between the catalyst and the solid support:

Covalent binding. In this method, covalent bonds are formed between enzyme and support material using the functional groups (such as amino, carboxyl, hydroxyl, and sulfydryl) present in various amino acid residues on the enzyme surface.[21,22] Various organic and mineral supports are used for covalent immobilization, but prior to attachment, these supports usually require activation. This method has the advantage of good retention of activity and often renders excellent reusability and thermal stability. The drawback of covalent immobilization is loss of enzyme activity owing to exposure of the biocatalyst to toxic reagents and severe reaction conditions.[23]

Physical adsorption. This method is based on the reversible surface interaction between enzyme and support material. The forces involved are very weak and include van der Waal's forces, hydrogen bonding, hydrophobic bonding interactions, and their combinations.[22] Immobilization of biocatalyst through adsorption is simple and easy to perform. The significant drawback of this technique is leakage of biocatalyst through desorption.

Entrapment and encapsulation. Entrapment of enzymes involves physical confinement of enzymes in the lattice structure of gel or cross-linkable resins. Entrapment of enzyme in alginate is recognized as one of the rapid, nontoxic, inexpensive, and versatile methods for enzyme immobilization.

Another significant development in this area has been made by introduction of K-carrageenan polymers. The serious drawback of entrapped enzymes is diffusional limitation, particularly in case of high-molecular-weight reactants.[24]

Encapsulation is very similar to entrapment. In encapsulation, the catalyst is enclosed in the pore space, where the diameter of the pore opening is smaller than the pore space. Nylon and cellulose nitrate have proven to be popular materials for construction of microcapsules suitable for enzyme immobilization. A distinct advantage of this method is co-immobilization, whereas disadvantage includes difficulty in mass transfer and in certain cases may result in rupture of the membranes if product accumulation is rapid.[25]

4.2.9 Role of Lipases in the Preparation of Drugs and Drug Intermediates

Recently, there has been tremendous demand for chiral drug substances to focus on single-stereoisomer instead of racemic mixtures. For this reason, kinetic resolution processes and enzymatic desymmetrization of prochiral compounds have special relevance in pharmaceutical industries.

In order to prepare enantiometrically pure secondary alcohols, structures that contain many chiral drugs or intermediates for the synthesis of pharmaceuticals, their resolution can be carried out by acylation of alcohols or hydrolysis of esters using lipases in organic solvents. These compounds have been synthesized for the preparation of β-androgenic blocking agents with the structure of 1,2-aminoalcohols. By using *Pseudomonas cepacia* lipase (PSLs) resolution of azido alcohols can be carried for the synthesis of (*S*)-propranolol (Fig. 4.2).[26] The general strategy for generation of these alcohols by enzymatic acylation is shown in Fig. 4.3.[1]

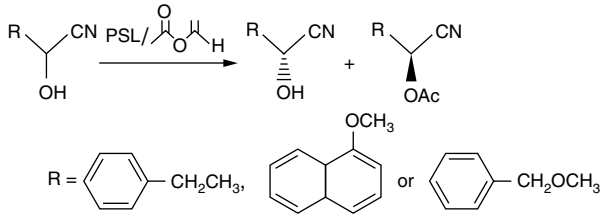


Figure 4.2 Lipase-catalyzed resolution of cyanohydrins and syntheses of ethyl-(*R*)-2-hydroxy-4-phenylbutyrate and (*S*)-propranolol.

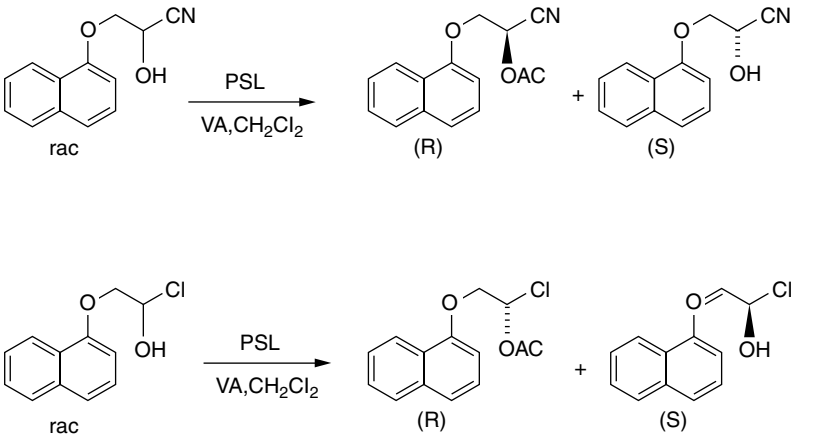


Figure 4.3 Resolution of alcohol intermediates in the preparation of (*S*)-propranolol.

Many 2-amino-1-arylethanols derived from (*R*)-adrenaline and (*R*)-noradrenaline as the parent compounds are important adrenergic drugs.[27] Studies on these drugs have shown that the biological activity resides mostly in their (*R*)-enantiomer, whereas the (*S*)-isomer is usually less active or even may cause undesired side effects. Resolution of 1-(4-amino-3-chloro-5-cyanophenyl)-2-bromo-1-ethanol has been achieved by ethanol enantioselective lipase-catalyzed alcoholysis, hydrolysis, and acylation. Although a good enantioselectivity was observed in the three reactions, the best results were obtained by hydrolysis (Fig. 4.4). This α-bromohydrin is an intermediate in the synthesis of a new adrenergic agent.

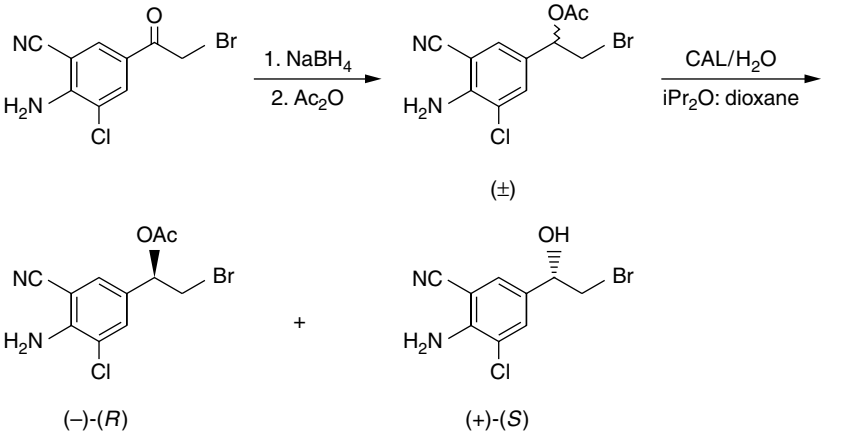


Figure 4.4 Enantioselective hydrolysis of the racemic ester (±).

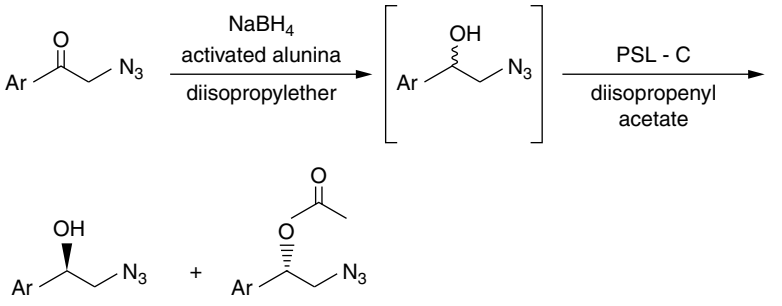


Figure 4.5 Chemoenzymatic synthesis of chiral β-azidoalcohols.

Enantiopure β-azidoalcohols are of great significance as potential precursors for optically active aziridines and β-amino alcohols. Chiral 1,2-amino alcohols have become of great significance because of their presence in biologically active natural products, such as ephedrine and pharmacologically active bronchodilators salmeterol and albuterol. The enantioselective synthesis of chiral β-azidoalcohols was done via reduction of the corresponding ketoazides with $NaBH_4$ in the presence of moist aluminium oxide followed by an in situ lipase-mediated resolution (Fig. 4.5).[28]

The chiral γ-azidoalcohols are immediate precursors of γ-aminoalcohols and are found to be useful in the ring expansion of symmetric cyclohexanones in the asymmetric Schmidt reaction. Synthesis of enantiomerically pure γ-azidoalcohols by lipase-catalyzed transesterification by Ahmed and colleagues.[29] *P. cepacia* immobilized on diatomaceous earth (PS-D) in *n*-hexane catalyzed the transesterification process in an efficient manner, providing γ-azidoalcohols in high enantiomeric excess. The enantiomerically pure γ-azidoalcohols obtained by kinetic resolution in high enantiopurity have been used to synthesize enantiomeric pairs of the antidepressant drugs fluoxetine and duloxetine (Fig. 4.6).

1, 4-Dihydropyridine derivatives are of great interest in the pharmaceutical industry as the calcium antagonists. Enantiometrically pure 2-methoxy-2-phenyl ethanol can be used for the preparation of

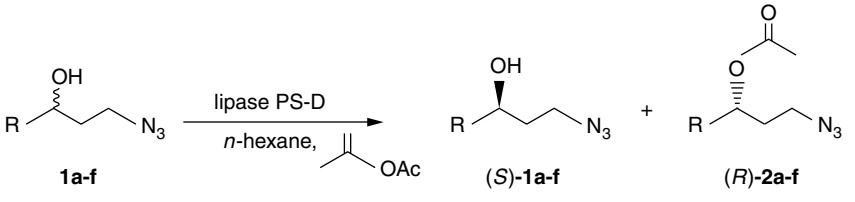


Figure 4.6 Synthesis of enantiomerically pure γ-azidoalcohols by lipase-catalyzed transesterification.

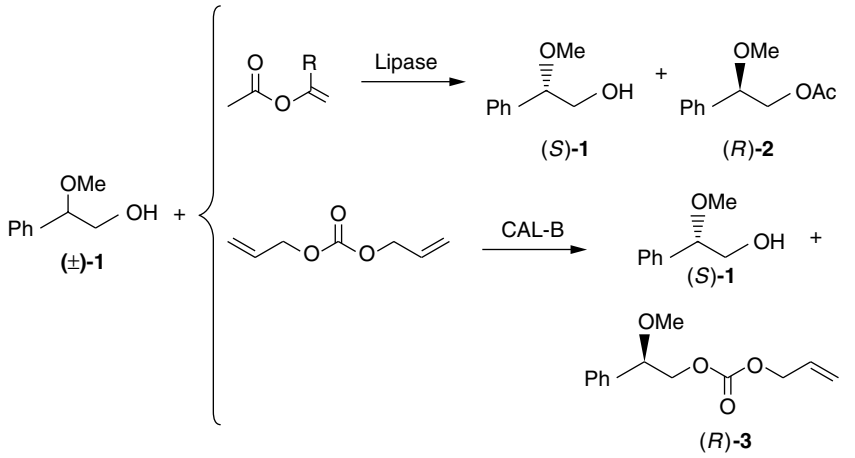


Figure 4.7 Lipase-catalyzed resolution of 2-methoxy-2-phenyl ethanol.

such derivatives, and the resolution of such primary alcohol can be carried out with different lipases, *Candida antarctica* lipase B (CAL-B) being the most efficient biocatalyst (Fig. 4.7).[30]

2-Arylpropionic acids, bearing a single stereogenic tertiary center, constitute an important class of nonsteroidal anti-inflammatory agents that relieve inflammation by inhibiting cyclooxygenase, thereby regulating the arachidonic acid cascade. Because the pharmacologic activity of the (*S*)-isomer of 2-arylpropionic acid is reported to be stronger than that of the (*R*)-isomer, development of an efficient enantioselective synthetic route to the (*S*)-isomer has received considerable attention. The lipase selectivity, however, was not sufficient for preparative scale resolution, and the water content strongly affected the reaction rate. López-Belmonte and colleagues[31] reported the *Rhizopus miehei* lipase-catalyzed resolution of the racemic mixture, but selectivity was extremely low (Fig. 4.8).

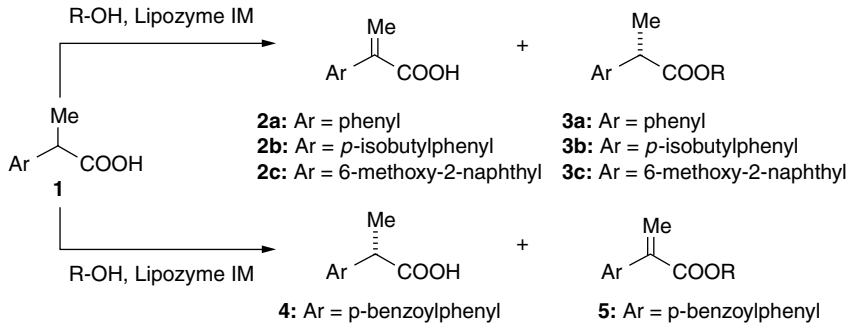


Figure 4.8 Enantioselective esterification of 2-arylpropionic acids catalyzed by immobilized *R. miehei* lipase.

Park and colleagues[32] suggested that enantioselectivity in lipase-catalyzed resolution of racemic ketoprofen depended mainly on the sources of lipase, alcohol moiety, organic solvent, and water content. Ethanol was used as the alkyl donor, and the optimal water content required for highly-efficient enzymatic resolution was determined to be 0.1 to 0.15 percent (v/v), which was maintained using salt hydrates such as $Na_2SO_4 \cdot 10H_2O$. (*S*)-Ketoprofen could be obtained with high enantioselectivity (E = 15) in *n*-hexane supplemented with ethylene dichloride (20 percent v/v) using commercially available *C. antarctica* lipase. Moreno and colleagues[33] suggested covalent immobilization of pure lipases A and B from *C. rugosa* on agarose and silica, and the percentage immobilization of lipases A and B is similar in both supports (33 to 40 percent). The remaining activity of the biocatalysts immobilized on agarose (70 to 75 percent) is greater than that of the enzymatic derivatives immobilized on silica (40 to 50 percent). The hydrophobic/hydrophilic properties of the support control the lipase activity of these derivatives. The thermal stability of the immobilized lipase A derivatives is greater than that of lipase B derivatives. The nature of the support influences the thermal deactivation profile of the immobilized derivatives. The immobilization in agarose (hydrophilic support) gives biocatalysts that show a greater initial specific reaction rate than the biocatalysts immobilized in silica (hydrophobic support) using the hydrolysis of the esters of (*R*)- or (*S*)-2-chloropropanoic and of (*R,S*)-2-phenylpropanoic acids as the reaction test. The enzymatic derivatives are active for at least 196 hours under hydrolysis conditions. The stereospecificity of the native and immobilized enzymes is the same. Kato and colleagues[34] reported an efficient synthetic procedure for obtaining optically active ketoprofen using *Mucor javanicus* lipase, one of nine commercially available hydrolytic enzymes, showed good enantioselectivity (E = 50) for racemic ketoprofen trifluoroethyl ester in phosphate buffer (pH 7.0) containing 30 percent acetone (Fig. 4.9). Lipase immobilized on Toyonite 200-A showed the best selectivity (E = 55) and reactivity. Moreover, the lipase could be recycled at least five times.

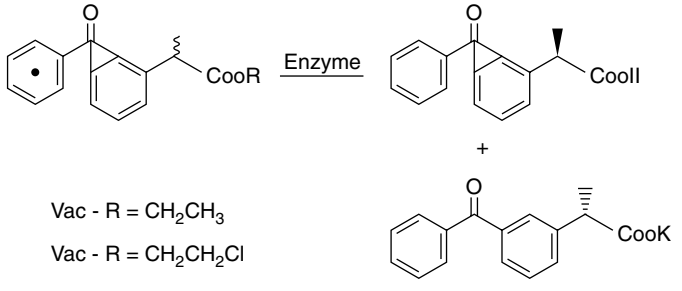


Figure 4.9 Enzymatic-catalyzed enantiosolective hydrolysis of ketoprofen monochloroethyl ester and trifluoroethyl ester.

Basak and Nag and colleagues[35] anticipated the following strategies to synthesize anti-inflammatory pure chiral drugs. The methyl ester of ibuprofen, naproxen, or flubiprofen was reduced with sodium borohydride in methanol. The mixture was allowed to stir for 4 hours, after which the reaction volume was reduced to 5 mL. Water was added, and the mixture was extracted with ethyl acetate. The combined organic layers were dried over anhydrous sodium sulfate and then evaporated. The product was purified by silica gel chromatography using hexane:ethyl acetate ratio 3:1. The alcohol then was acetylated with acetic anhydride, triethylamine, and a catalytic amount of DMAP using dichloromethane as solvent. The mixture was allowed to stir for 1 hour at room temperature and then was washed with water and extracted with dichloromethane. The organic layer was dried over Na_2SO_4 and then evaporated to leave a white solid from which the product was isolated by silica gal chromatography using hexane: ethyl acetate ratio 7:1. The acetyl derivative was hydrolyzed with the help of the enzyme (PPL) in phosphate buffer (pH 8.0) and acetone. After 50 percent conversion (estimated by TLC), the mixture was extracted with ethyl acetate and dried over Na_2SO_4, and the solvent was evaporated. The residue, on chromatography on silica gel, furnished the product, which was isolated from hexane: ethyl acetate ratio 1:1. The absolute configuration was determined by comparison with the literature values for the *S*-isomer. The oxidation of the (+) alcohols was carried out with PDC using dimethylformamide (DMF) as solvent. After 72 hours, the mixture was partitioned between ethyl acetate and aqueous $NaHCO_3$. The aqueous layer was adjusted to pH 2 and then reextracted with ethyl acetate. The organic layer was dried with Na_2SO_4, filtered, and evaporated to leave the acid as a white solid (yield ≈ 50 percent) having identical ^1H NMR with the authentic racemic sample (Fig. 4.10).

(–)-Paroxetine hydrochloride is a selective serotonin (5-HT) reuptake inhibitor used as an antidepressant. A (–)-paroxetine alcohol intermediate also can be prepared by the enzymatic hydrolysis of the corresponding ester derivative or enzymatic acylation of the primary alcohol in a reaction catalyzed by *C. antarctica* lipases CAL-A and CAL-B, as shown by Fernandez and colleagues.[1] Both enzymes gave very good results in terms of yield and enatioselectivity. In the reaction process, the two *C. antarctica* lipases showed opposite stereochemical preference. CAL-A catalyzes the acylation of the (3*S*, 4*R*)-alcohol, whereas CAL-B prefers the (3*R*, 4*S*)-enantiomer. In the last case, the remaining alcohol possesses the correct absolute (3*S*, 4*R*) for the synthesis of (–)-paroxetine.

Citalopram is a very selective inhibitor of serotonin and an efficient human antidepressant. It can be obtained by resolution of the corresponding cyanodiol with a primary and a tertiary hydroxyl group. Again, to get optically pure citalopram, on testing with

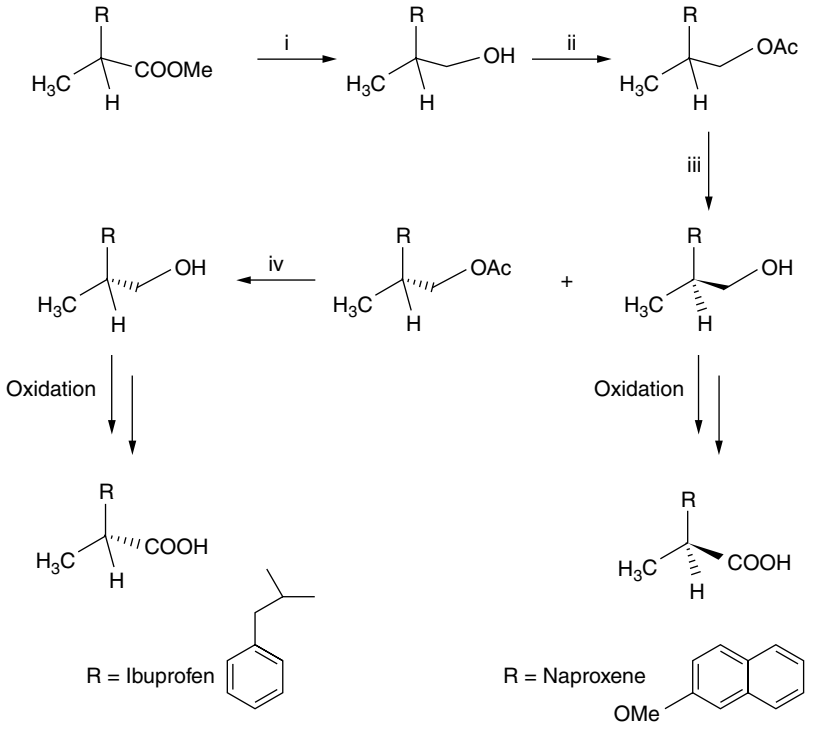


Figure 4.10 Enantioselective hydrolysis of ester group to produce anti-inflammatory drugs.

several lipases, CAL-B, vinyl acetate, and acetonitrile gave the best results (E = 70), as shown by Fernandez-Solares and colleagues.[36] *C. antarctica* lipase B (CAL-B) catalyzes the enzymatic acetylation of the primary benzylic alcohol with high enantioselectivity at the quaternary stereogenic center (Fig. 4.12). The enzymatic enantioselective hydrolysis of the 3-acetyloxymethyl derivative catalyzed by CAL-B is also possible.

Caffeic acid phenethyl ester (CAPE) has been used widely in oxidative stress-related research. CAPE has several biologic and pharmacologic properties such as anti-inflammatory and neuroprotective effects. The CAPE molecule is small and lipophilic and thus easily enters into cells by crossing the cell membrane. CAPE was synthesized successfully from caffeic acid and phenyl ethyl alcohol catalyzed by Novezym-435.[37] Reactions carried out in nonpolar solvents resulted in higher conversion than in polar solvents despite

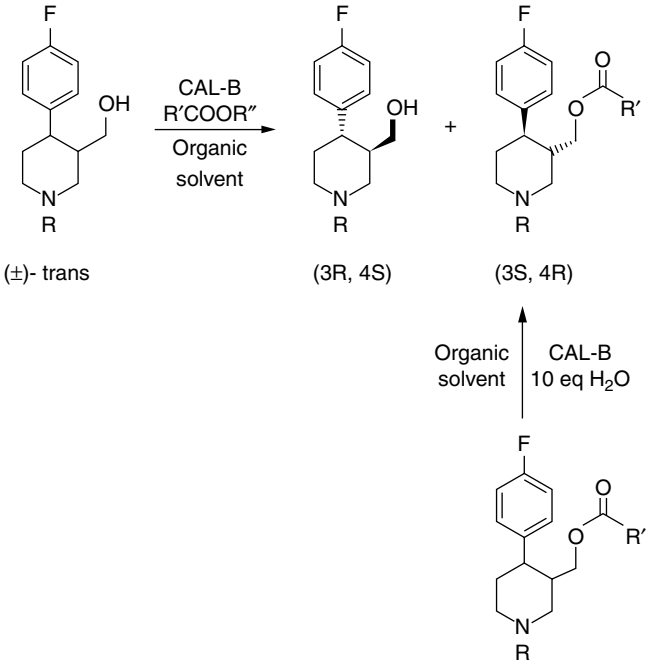


Figure 4.11 Lipase-catalyzed preparation of (−)-paroxetine precursor.

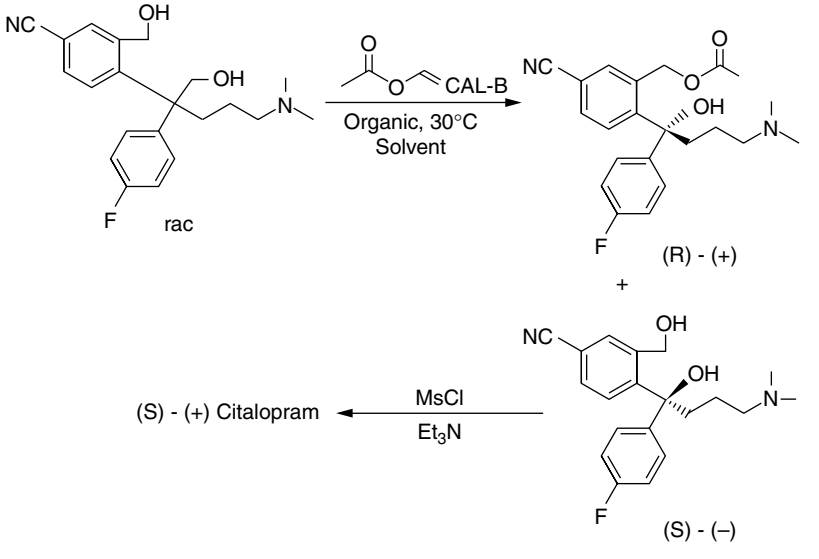


Figure 4.12 Chemoenzymatic synthesis of (S)-(+)-citalopram.

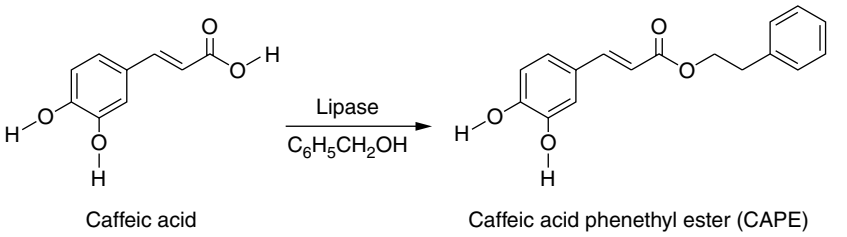


Figure 4.13 Enzymatic synthesis of caffeic acid phenethyl ester.

the low-solubility caffeic acid in nonpolar solvents. Isooctane gave the highest conversion (Fig. 4.13).

Trans-2-phenylcyclohexan –1-ol (TPCH) is an excellent intermediate with potential to be used widely in the manufacture of several chiral pharmaceuticals. TPCH can be used in the assembly of several chiral target molecules such as diltiazem, hydroxyl amino acids, etc. TPCH can be used in a wide variety of asymmetric reactions such as the Diels-Alder re-formation, Darzen's glycidic ester condensation, and so on.

Traditionally, TPCH is synthesized by copper catalyzed opening of cyclohexene oxide by a phenyl Grignard reagent. Synthesis of TPCH by resolution with *C. rugosa* lipase (CRL) was shown to be efficient by Luis del Río and colleagues (Fig. 4.14).[38]

Ferulic acid (FA), obtained from medicinal herbs such as *Angelica sinensis* and *Cimicifuga heracleifolia*, is a phenolic acid having strong antioxidant action and exhibiting a number of physiologic activities such as anticarcinogenic and anti-Alzheimer's disease effects. An enzymatic synthesis approach to prepare novel feruloylated lipids through the lipase-catalyzed transesterification reaction of ethyl ferulate (EF) with tributyrin (TB) in toluene was investigated by Zheng and colleagues.[39] The two major products of this enzymatic synthesis are 1(3)-feruloyl-monobutyryl glycerol (FMG) and 1(3)-feruloyl-dibutyryl-glycerol (FDG) (Fig. 4.15).

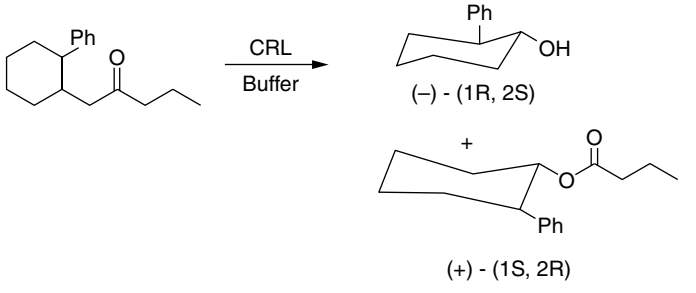


Figure 4.14 Lipase-mediated resolution of *trans*-phenylcyclohexanol.

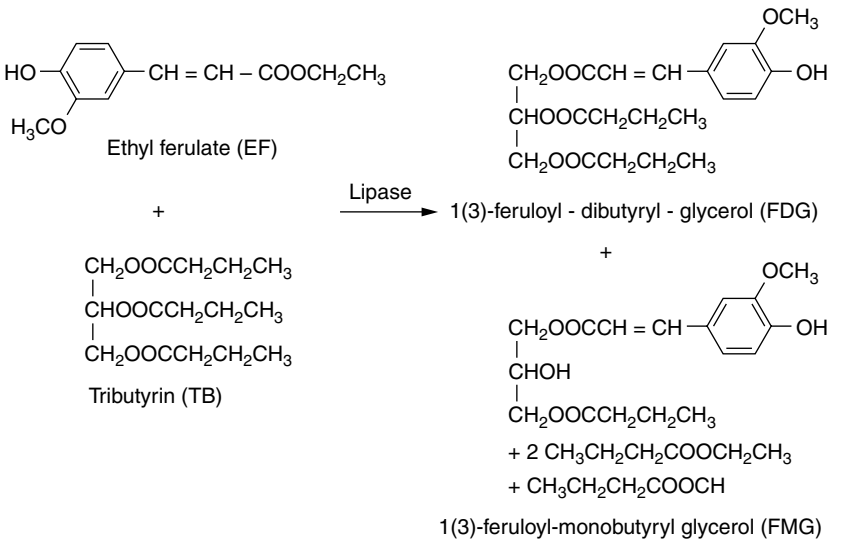
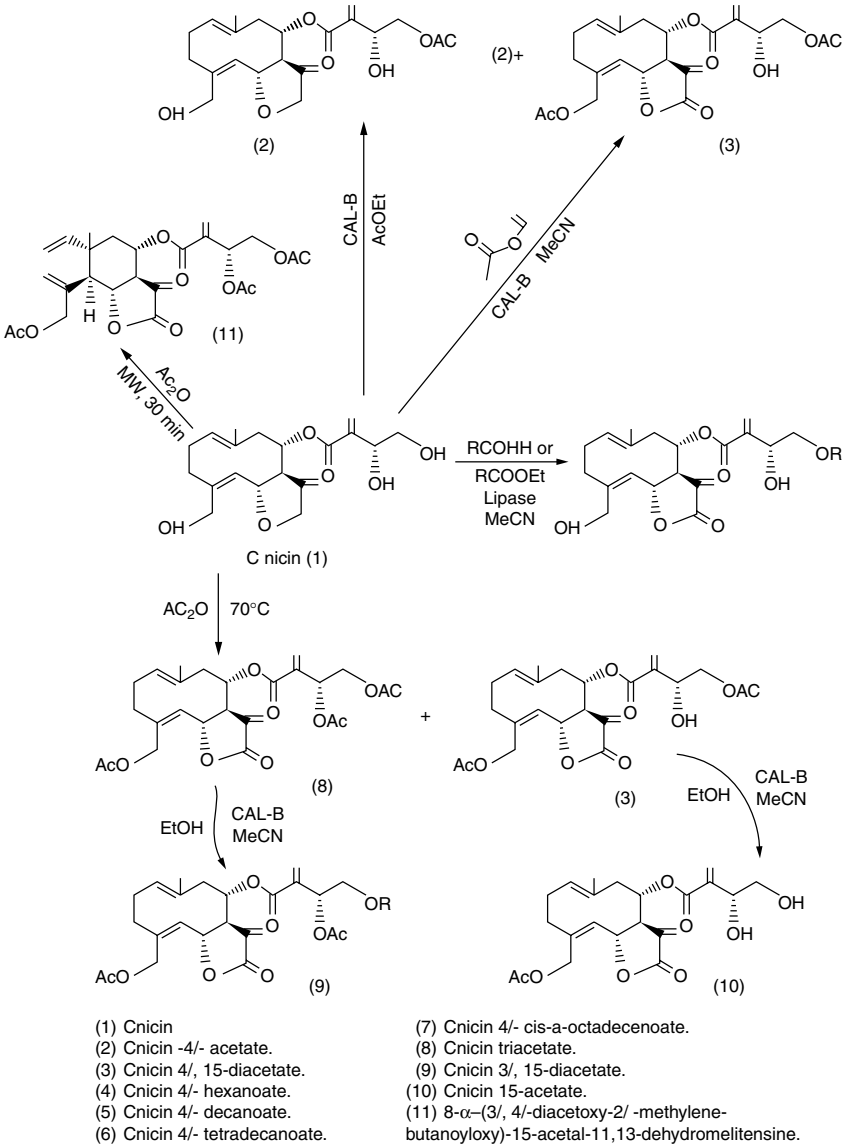


Figure 4.15 Lipase-catalyzed transesterification of ethyl ferulate with tributyrin.

The germacranolide cnicin, an important sequiterpene, exhibits prominent antibacterial, cytotoxic, and antifungal properties. The presence of an ester at C-8 of the germacranolide is important for its antibacterial activity. The esterified derivatives are more active against gram-positive bacteria. Owing to the presence of three free hydroxyl groups in the cnicin molecule, scientists made an attempt to synthesize various esters so that the position of the ester group could influence the numerous biologic activities displayed by cnicin. When the conventional synthetic methods did not give significant results, scientists took an enzymatic approach because lipases are efficient biocatalysts for chemoselective, region-selective, and stereoselective reactions under mild conditions.

Thus the behavior of the lipases from several sources in the acetylation of this germacranolide, cnicin, have been investigated by dissolving cnicin in ethyl acetate, which works as acylating agent and solvent. The transesterification was conducted using isopropyl acetate in acetonitrile with both enzymes *C. antarctica* lipase B and *P. cepacia* lipase by Monslave and colleagues.[40] Both the enzymes gave the same product cnicin-4-acetate in high yields, but *C. antarctica* lipase B was more efficient than *P. cepacia* lipase. The same enzymatic strategy can be applied for preparation of acyl derivatives of cnicin that have longer chain acyl moieties (Fig. 4.16).

Enzymatic acylation of amines can be used for the preparation of β-substituted isopropyl amines because of their pharmacologic significance. Owing to the increasing demand for optically active



(1) Cnicin
(2) Cnicin -4/- acetate.
(3) Cnicin 4/, 15-diacetate.
(4) Cnicin 4/- hexanoate.
(5) Cnicin 4/- decanoate.
(6) Cnicin 4/- tetradecanoate.
(7) Cnicin 4/- cis-a-octadecenoate.
(8) Cnicin triacetate.
(9) Cnicin 3/, 15-diacetate.
(10) Cnicin 15-acetate.
(11) 8-α–(3/, 4/-diacetoxy-2/ -methylene-butanoyloxy)-15-acetal-11,13-dehydromelitensine.

Figure 4.16 Lipase-catalyzed preparation of acyl derivatives of the germacranolide cnicin.

compounds in the pharmaceutical industry, as well as the synthetic applicability of enantiomerically pure amines (as chiral auxiliaries, bases, and ligands), the design of efficient methods for the preparation of optically active amines is of special interest. The presence of methyl groups on the nitrogen atom is known to increase the efficiency of drugs. Both enantiomers of the amphetamines can be prepared with

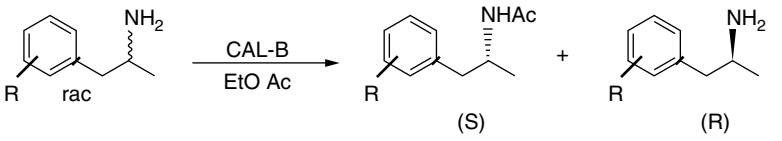


Figure 4.17 Enzymatic kinetic resolution of β-substituted isopropyl amines.

high efficiency by *C. antarctica* B–catalyzed resolution of the racemic amines using ethyl acetate as acyl donor and solvent, as shown by González and colleagues.[41] The unreacted amines and the converted acetamides are easily separated by selective extraction (Fig. 4.17).

Synthesis of nucleoside analogues is undergoing increasing demand because of their significant biologic activity, such as antitumor, antiviral, and activity against HIV. Chemoenzymatic synthesis using lipases provided many favored methods to prepare the enantiomerically pure nucleoside analogues. Some of the prochiral cyclopentane that acted as a starting material for the synthesis of some enatiomerically pure nucleoside analogues is given in Fig. 4.18.

Among these compounds, Roberts and colleagues[42] hydrolyzed the trisubstituted mesodiacetate (*A*) with PPL into monoacetate (*B*) with a yield of 92 percent. This monoacetate obtained in enantiomerically pure form is used to synthesize various carbocyclic nucleoside analogues such as adenine derivatives and the precursors for both enantiomers of aristeromycin (*C*) and neplanocin (*D*) (Fig. 4.19).

Several processes in the synthesis of different antibiotic structures took advantage of biocatalytic lipase-mediated enantioselective steps to introduce chirality into the precursors. Sih and colleagues[43] used compound *A*, which is a diastereoisomerically pure but racemic acetate, to synthesize *C*, a biosynthetic monensin *A*, a pentacyclic polyether ionophore produced by *Streptomyces cinnamonensis*. It has a

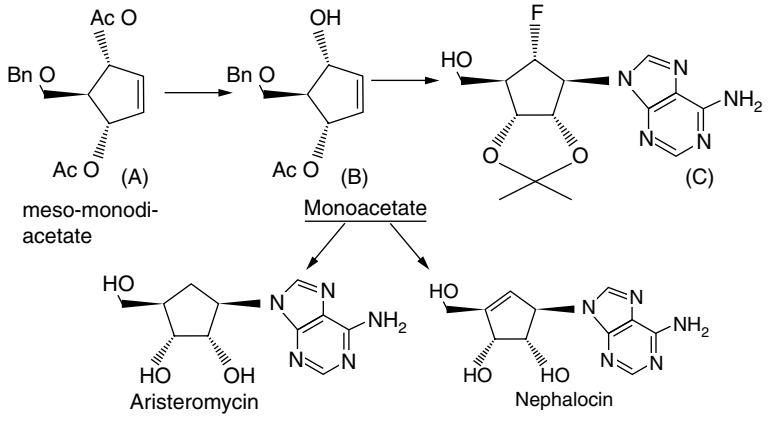


Figure 4.18 Prochiral cyclopentane starting materials used for synthesis of nucleoside analogues.

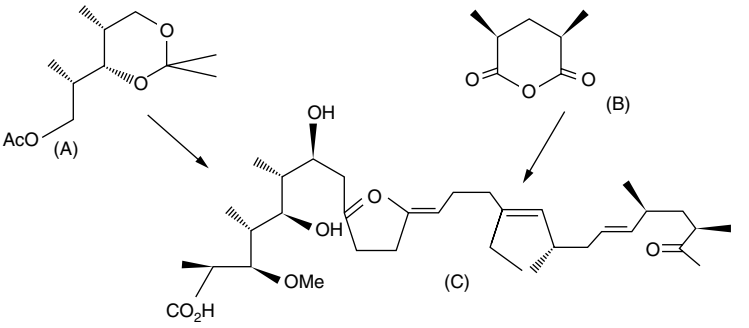


Figure 4.19 Lipase-catalyzed synthesis of monensin.

polyoxygenated branched-carbon skeleton derived from the building units acetate, propionate, and butyrate and possesses stereochemical features in common with many polyethers of its class. Compound *A* was resolved by hydrolysis in the presence of PPL.

Compound *B*, that is, meso-2,4-dimethyl glutaric anhydride, was asymmetrized by alcoholysis with 2-methyl propanol in the presence of lipase SP 382 to give a monoester that acted as the starting material for compound *C* and other biomolecules.

The synthesis of the β-lactone antibiotic 1233A was based on the lipase-catalyzed asymmetrization of prochiral diacetate (X) to afford a chiral monoacetate with 86 percent yield, as shown by Mori and colleagues (Fig. 4.20).[44]

Different racemic intermediates (1–3) shown in Fig. 4.21 were used by Akita and colleagues[45] as substrates in lipase-catalyzed hydrolyses or transesterification for the enantioselective synthesis of oudemansins (4), antibiotics with strong antifungal activities. In the case of the hydrolytic separation of compounds 1 and 2, the methyl ester function was unaffected by lipases.

Chloramphenicol is a natural antibiotic with a fairly wide spectrum of antimicrobial activity. It inhibits protein synthesis in bacteria and, to a lesser extent, in eukaryotic cells by binding to the 50S ribosomal subunit, thus preventing access of aminoacyl tRNA to the ribosome. Chloramphenicol is employed widely as medicine against

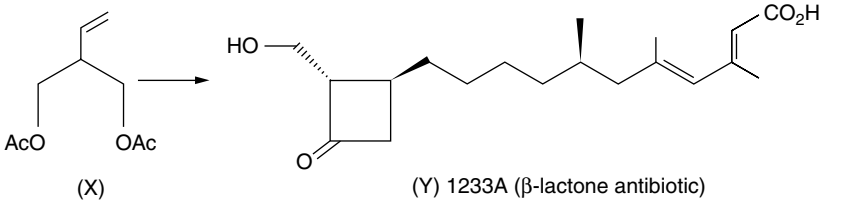


Figure 4.20 Synthesis of β-lactone.

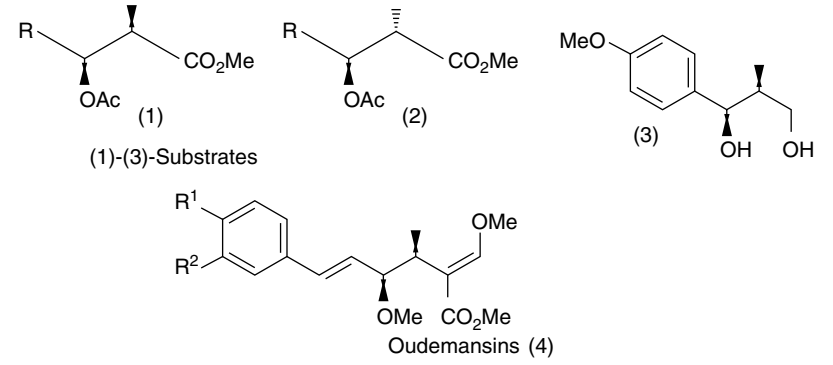


Figure 4.21 Synthesis of oudemansins.

infections of the urinary tract and other bacterial diseases. Ottolina and colleagues[46] achieved regioselective esterification of chloramphenicol by the action of lipase in acetone and several methyl carboxylates. Aliphatic and aromatic esters of different sizes and natures have been introduced selectively on the primary hydroxyl group of these molecules by modification of the reaction conditions (e.g., temperature, solvent, and lipase source) (Fig. 4.22).

Macromolecular drugs have attracted considerable interest because they can effectively control the rate of the drug release and increase therapeutic benefit. Synthesis of drug monomers is an important step for macromolecular drug preparation. Thiamphenicol [D(+)-*threo*-1-(4′-methyl sulfonylphenyl)-2-dichloroacetamido-propane-1,3-diol) is characterized by broad-spectrum antibiotic activity similar to that of chloramphenicol, with lower toxicity and higher activity than chloramphenicol. Yu Zhen and colleagues[47] synthesized three polymerizable vinyl thiamphenicol esters with different acyl donor carbon chain lengths (C4, C6, and C10) by Lipozyme (immobilized from *Mucor miehei*) in acetone at 50°C for 12 hours to give 73, 81, and 63 percent yields, respectively. Products were isolated by silica gel column chromatography, with the eluent consisting

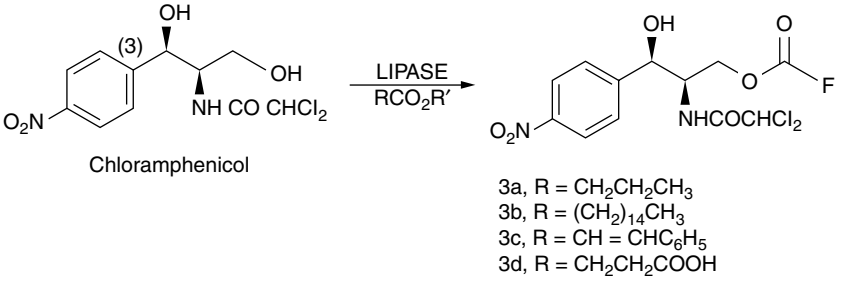


Figure 4.22 Ester derivatives of chloramphenicol by lipase in organic solvents.

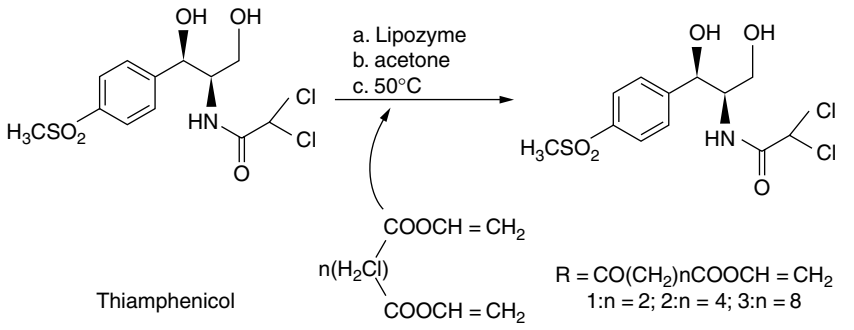


Figure 4.23 Lipase-catalyzed synthesis of thiamphenicol esters.

of ethyl acetate:petroleum ether (2:1 v/v). The products were valuable monomers for preparation of the macromolecular antibiotic (Fig. 4.23).

4.3 Esterase as Hydrolase Enzyme

An esterase is a hydrolase enzyme that splits esters into an acid and an alcohol in a chemical reaction with water called *hydrolysis*. It differs from lipase in that esterases preferentially break ester bonds of shorter-chain fatty acids, whereas lipases can break long-chain fatty acids, which typically are insoluble or at least poorly soluble in water. Fojan and colleagues[9] suggested that the lipases display high activity toward the aggregated state of their substrate, whereas the esterases typically show their highest activity toward the soluble state of their substrate. Lipases display a statistically significant enhanced occurrence of nonpolar residues close to the surface, clustering around the active site. Lid opening appears to strengthen this pattern further. The active site of lipases displays negative potential in the pH range associated with their maximum activity, typically at pH values above 8. The esterases show a very similar pattern, but at pH values around 6, correlated with their usually lower pH activity optimum.

4.3.1 Esterase Production

Different sources yield different types of esterases. Based on sources, generally varied nomenclature is reported in the literature, for example, carboxylesterase, cholinesterase, acetyl xylan esterase, aryl esterase, phosphotriesterase, phenolic esterase, pig liver esterase, acetylcholine esterase, cholesterol esterase, ferulic acid esterase, tannin esterase, and so on.

Generally, knowledge of specific microbes and the choice of an appropriate fermentation technique are necessary to produce a specific esterase. However, human and pig liver esterases cannot

be produced by fermentation techniques. In all these fermentations, though, other hydrolytic enzymes such as cellulases, hemicellulases, and so on are also produced, which makes for greater costs in downstream processing. Thus recombination production methods are used to produce a pure specific esterase. Here, the source and production of some esterases[48] are given in Table 4.1.

Animal esterase, PLE, is generally isolated as acetone powder from fresh pig liver.[49] The specific activity of esterase is determined by spectrophotometrically using *p*-nitrophenyl acetate. Microbial esterases can be used in the form of cells from both bacteria and fungi to catalyze the reactions. The problem, however, is to produce the microbial cells in a suitable medium. Sometimes, controlling the reaction becomes more complex, but the selectivity is more than the isolated enzyme.

SI Number	Form of Esterase	Source
01	Feruloyl esterase	*A. Niger, Streptomyces avermitilis*
02	Pig liver esterase, porcine liver esterase, and recombinant pig liver esterase	
03	Methyl jasmonate esterase	*Lycopersicon esculetum*
04	Esterases from human system	
05	Recombinant esterase (PF1-K)	*Pseudomonas* sp.
06	Acetylcholinesterase	Blood of *Schistosoma* spp., *Mytilus edulis, Drawida willsi*
07	Aryl esterase	*Saccharamyces cerevisiae*
08	Acetyl esterase, methyl esterase, acetylglucomannan esterase and acetyl xylan esterase	*Aspergillus, Trichoderma* spp.
09	Cholesterol esterase and pseudocholinesterase, cholinesterase	Rat liver and other sources
10	Carboxylesterases	*Streptomyces lividans*, livers of rat and guinea pig, *Lucilia cuprina, Pediculus capitis, Bacillus coagulans*

Table 4.1

4.3.2 Preparation of Drugs and Drug Intermediates by Esterases

Acetaminophen is one of the safest analgesic drug in the world. However, the following difficulties are countered during preparation of acetaminophen from *p*-aminophenol:

1. *p*-Aminophenol is an unstable compound that undergoes oxidation to *p*-quinone and related compounds, thus isolation in the pure form may be tricky.
2. Selective acetylation of the amino group in preference to the phenolic hydroxyl group is a difficult proposition.
3. Chemoselective hydrolysis of *O*-acetyl in preference to *N*-acetyl is difficult to achieve through chemical methods.

To overcome these difficulties, Nag and colleagues[50] have selected *p*-nitrophenol rather *p*-aminophenol to prepare *N*,*O*-diacetyl *p*-aminophenol, which is resistant to oxidation. Selective hydrolysis of the *O*-acetyl group in preference to the *N*-acetyl group was achieved successfully with pig liver esterase (PLE) (Fig. 4.24).

N,*O*-Diacetyl *p*-aminophenol was prepared from *p*-nitrophenol by the addition of equal volumes of acetic anhydride and acetic acid followed by zinc dust. *N*,*O*-Diacetyl *p*-aminophenol was hydrolysed by PLE in phosphate buffer and acetone. The solution was filtered on celite under vacuum. The filtrate was acidified (pH 2.4) with 1 N HCl and was extracted with ethyl acetate (2 × 40 mL). The organic layer was dried with anhydrous Na_2SO_4, filtered, and evaporated to leave the compound as a light colored solid. In this study, acetaminophen was prepared by selective acetylation of the amino group in preference to the phenolic hydroxyl group, which is a difficult proposition, and also avoided unstable *p*-aminophenol. The simplicity of the isolation procedure is another interesting feature of the reaction.

Thromboxane A_2 is an exceptionally potent proaggregatory and vasoconstrictor substance produced by the metabolism of arachidonic acid in blood platelets and other tissues.[49] Together with the potent antiaggregatory and vasodilator actions, it is thought to play a role in the maintenance of vascular homeostasis and to contribute to the pathogenesis of a variety of vascular disorders. Compound 1 in Fig. 4.25

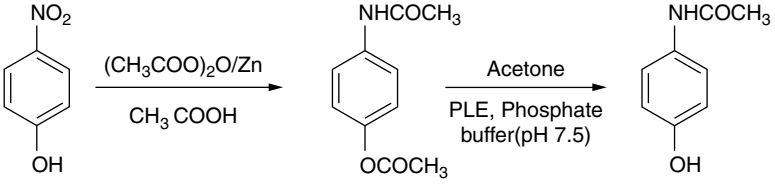


Figure 4.24 Enzymatic synthesis of acetaminophen.

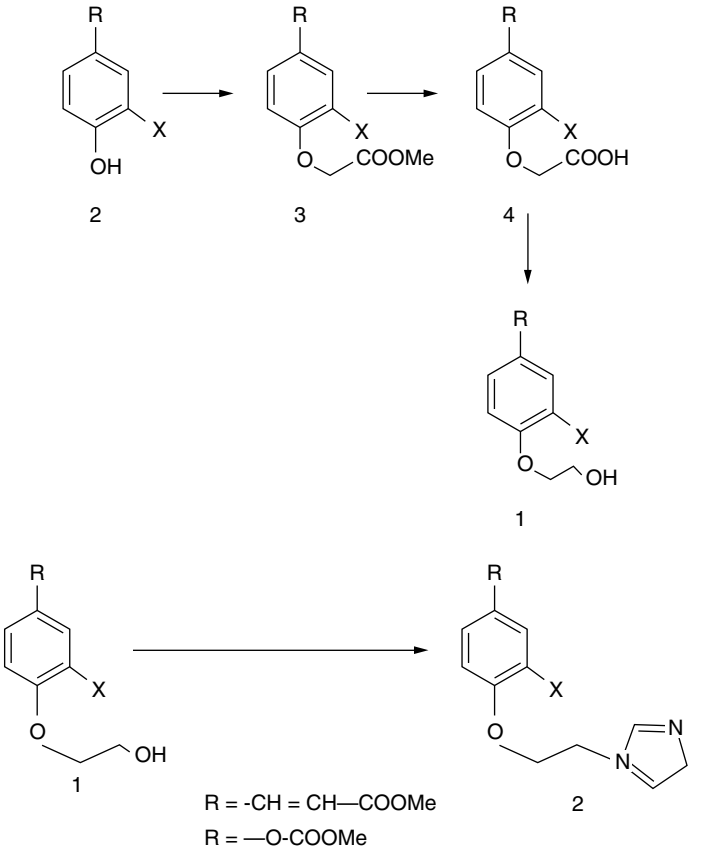


Figure 4.25 Esterase synthesis of thromboxane synthetase inhibitor.

is an intermediate for thromboxane synthetase inhibitor (compound 2). An attempt to synthesize compound 1 as given (scheme 2) is difficult because the conversion of compound 3 to compound 4 needs a selective transformation of saturated ester over α, β unsaturated or aromatic esters.

Ester groups attached with an α-unsaturated group or aromatic ring are much less likely to be hydrolyzed by normal hydrolytic conditions such as $NaOH/CH_3OH$ or LiOH/THF owing to the formation of an electronically unstable tetrahedral transition state. The selective transformation of saturated ester was not useful under these conditions owing to the formation of multiple products. When compounds having both saturated and unsaturated aromatic ester functionalities (acetone/buffer, pH 7.8; PLE), Basak and colleagues[49] found the saturated ester functionality to be smoothly hydrolyzed, leaving behind an unsaturated ester, the

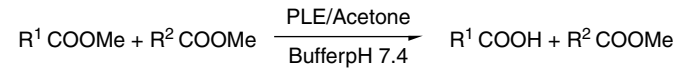


Figure 4.26 Selective transformation of saturated ester by esterase.

reaction time ranging from 4 to 20 hours with yield of saturated acid of 95 percent or higher (Fig. 4.26). The degree of hydrolysis was measured by ^{1}H NMR. The presence of unsaturated acid was never observed in NMR studies. This was a new finding based on the selectivity achieved by PLE.

Recently, β-lactams have been used extensively as synthons for various natural products, for example, side chain of taxol, the anticancer drug, and peptidase inhibitor. Basak and colleagues[52] have synthesized ethyl esters of several 4-substituted *N*-phenyl-β-lactam 3-carboxylic acids via the cycloaddition of disubstitued nitrons with methyl propionate in the presence of cuprous iodide.[51]

Basak and colleagues[52] adopted a strategy of enantioselective enzyme-catalyzed hydrolysis. Two classes of 3,4-disubstituted β-lactams 5–8 and 25–28 having trans-3-carboethoxy and *cis*-3-acetoxymethyl substituents, respectively, were prepared and then subjected to PLE- and PPL-catalyzed hydrolysis in phosphate buffer at pH 8.0. The reactions were stopped after about approximately 50 percent conversion. In the case of 3-carboethoxy β-lactams 5–8, the acids and the unconverted esters were separated by pH adjustments followed by solvent extraction. The enantiomeric excess was determined by converting the free acids to the diastereomeric amides 13–16 and 17–20, and their ratios were determined by 200-MHz NMR. The hydrogens at C-3 and C-4, as well as the methyl signals for the amides, appeared as separate doublets. In all cases, PPL failed to show any degree of enantioselectivity. However, with PLE, moderate enantioselection was observed for esters 6 and 7 substituted at C-4 with small heterocyclic rings such as furan or thiophene. Incorporation of conventional aromatic rings such as phenyl or *p*-methoxyphenyl at C-4 led to racemic acid during hydrolysis.

Hydrolysis of the acetates 25–28, on the other hand, showed complete reversal of enzyme selectivity. In these cases, practically no enzyme selectivity was observed with PLE, as revealed by negligible optical rotation and also by ^{1}H NMR in the presence of a chiral-shift reagent. However, with PPL, the enantioselectivity improved dramatically. In these cases, the alcohols 29–32 and the unhydrolyzed acetates 33–36 were separated by chromatography and then checked for enantiomeric purity. The absolute configurations of alcohols 29–32 and acetates 33–36 have been tentatively assigned by comparing their signs of rotation with those of similarly substituted analogues. Configurations 10 and 11 were assigned from the sign of rotation of their decarboxylated product 22–23 (Fig. 4.27).

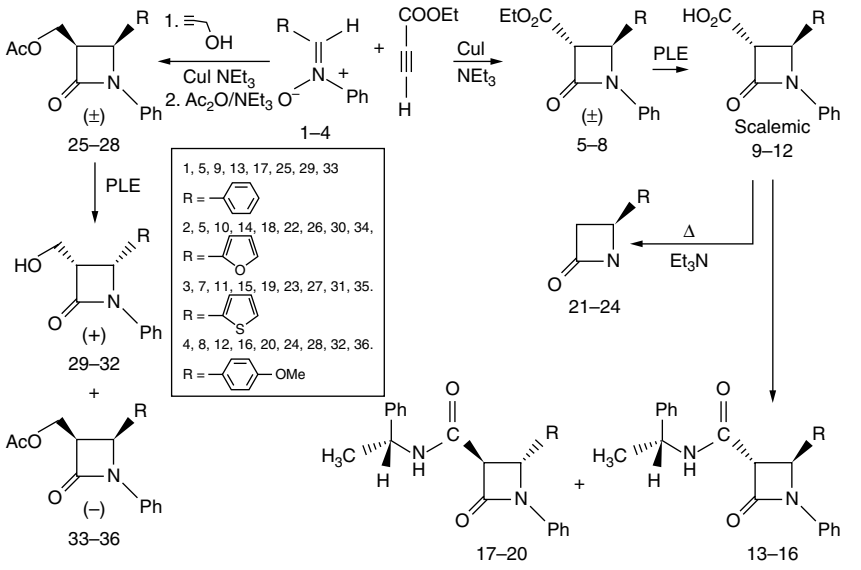


Figure 4.27 Enzymatic synthesis of different derivatives of β-lactams.

The highly enantioselective preparation of the *cis*-3-hydroxymethyl-β-lactams with a variety of substituents at C-4 was achieved via PPL-catalyzed hydrolysis of the corresponding acetates. Basak and colleagues[52] also have demonstrated an interesting dependence of ee (enantiomeric excess) on the nature of the substituent at C-4 during PLE-catalyzed hydrolysis of the *trans*-3-esters.

One interesting aspect of the hydrolysis is that the exomethylene-β-lactams obtained (Jones' model) from the hydrolyzed alcohols from the *cis*- and *trans*-acetates have the same sign of rotation and hence the same absolute conFiguration at C-4. This is possible only if the stereochemically preferable isomers from *cis*- and *trans*-β-lactam acetates undergoing hydrolysis have mirror-image configuration at C-4. The *cis*-β-lactam acetates have been shown to bind according to Jones' model. This necessitates the preferential binding mode of the *trans*-β-lactam acetates to be according to Seebachs' model,[6] which is enantiomeric to Jones' model. Thus adjacent chirality in the large hydrophobic group (in this case C-4 with the substituent) can dictate the mode of binding (Fig. 4.28).

Epoxides are widely distributed in nature and are of industrial, mechanical, and biochemical interest. Squalene 2,3-oxide is the biogenetic precursor of sterols. Leukotriene A (LTA) is the biogenetic precursor of leukotrienes C, D and E, which are important natural mediators of allergic asthma. Mamaghan and colleagues[53] used PLE efficiently in phosphate buffer for the separation of stereoisomeric

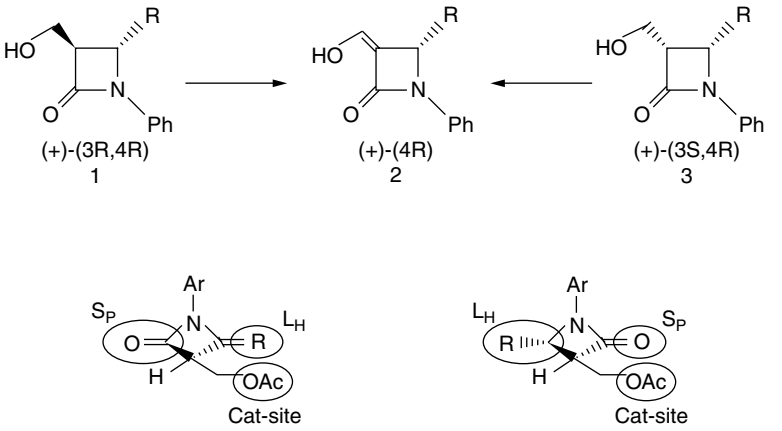


Figure 4.28 Jones' model in reference to β-lactam ester hydrolysis.

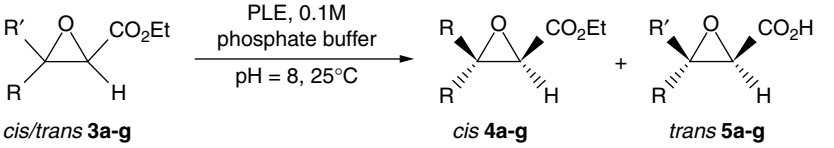


Figure 4.29 Estrase synthesis of *cis/trans*-ethyl glycidates.

mixtures of *cis/trans*-ethyl arylglycidates produced via Darzen's condensation reactions (Fig. 4.29 and Table 4.2).

Nitrate esters find widespread therapeutic use as drugs for treatment of heart and vascular diseases. Basavaiah and colleagues[54] developed a simple and convenient methodology for the synthesis of (1*R*,2*R*)- and (1*S*,2*S*)-2-nitroxycyclohexan-l-ols in enantiomerically pure form via pig liver acetone powder (PLAP)–mediated hydrolysis of racemic *trans*-1-acetoxy-2-nitroxycyclohexane (Fig. 4.30).

(–)-Physostigmine, initially isolated from the seeds of *Physostigma venenosum* in 1864, has been used clinically for the treatment of glaucoma, myasthenia gravis, and atropine and organophosphate intoxication, as well as for the relief of intoxication induced by overdoses of antidepressants, antihistamines, antipsychotics, and benzodiazepines. 1,4-6(–)-Physostigmine is a potent inhibitor of acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE), thereby showing wide biologic activities.

Asakawa and colleagues[55] described the preparation of a new chiral building block containing a benzylic quaternary stereogenic center by means of highly enantioselective PLE-mediated hydrolysis of dimethyl-2-(2-chloro-5-methoxyphenyl)-2-methylmalonate, as well as the absolute configuration of the new chiral building block,

Entry	*R*	*R′*	*cis/trans* Ratio	Reaction Time (h)	Conversion (%)	4a–g Yield (%)[a]	5a–g Yield (%)[a]
3a	C_6H_5	H	43/57	7.5	57	38	51
3b	C_6H_5	Me	39/61	9	61	36	57
3c	*p*-Me-C_6H_4	H	35/65	8.5	65	32	61
3d	*p*-Me-C_6H_4	Me	40/60	9.5	60	39	56
3e	*p*-MeO-C_6H_4	H	32/68	8	68	27	61
3f	*m*-NO_2-C_6H_4	H	30/70	10	70	27	67
3g	*p*-NO_2-C_6H_4	C_6H_5	45/55	13.5	55	40	49

[a]Isolated yields.

TABLE 4.2 Separation of *cis/trans*-Glycidates Using PLE

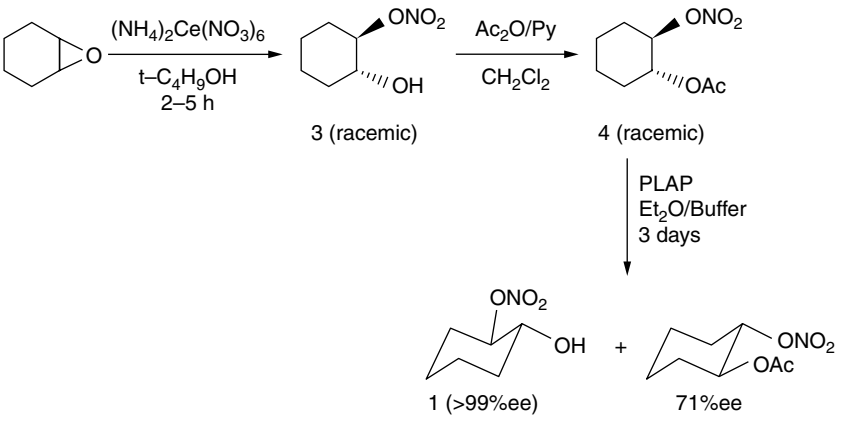


Figure 4.30 Esterase hydrolysis of racemic *trans*-1-acetoxy-2-nitroxycyclohexane.

which has been elucidated through the formal total synthesis of (–)-physostigmine (Fig. 4.31).

Naturally occurring enediynes attracted the attention of the scientific community and pharmaceutical companies for cancer treatment. These compounds show their antitumor activity via cleavage of DNA. The advantage of incorporating nitrogen into the enediyne framework is the scope of attaching different ligands onto the nitrogen for possible enhancement of the DNA-cleaving activity.

For the synthesis of monocyclic enediynes containing a nitrogen atom as a part of the ring, we required an access to the partially protected enediynyl acetates. Since selective acetylation of the diol or

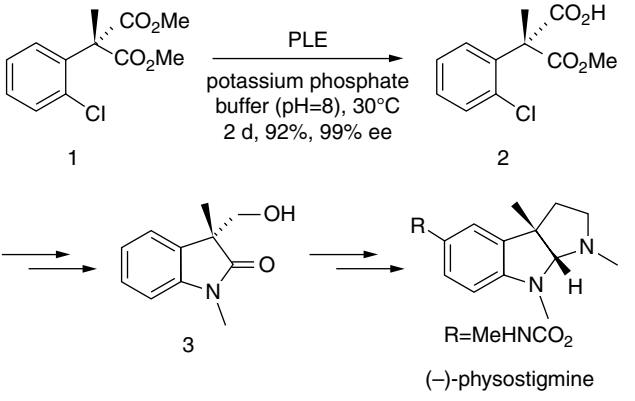
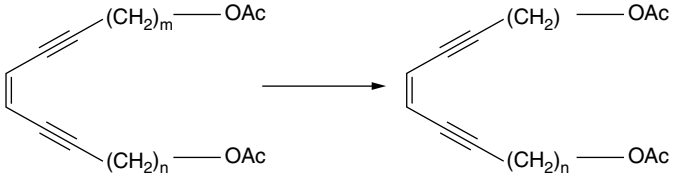


Figure 4.31 Enantioselective PLE-mediated hydrolysis of dimethyl-2-(2-chlorophenyl)-2-methylmalonate 1 and a formal total synthesis of (–)-physostigmine.

selective decetylation of the diacetate by chemical means is difficult, using the discriminatory catalytic power of enzymes to effect such a hydrolysis is possible. Basak and colleagues[56] succeeded in regio-selective hydrolysis of endiyne acetates, which provided access to important synthons for antitumor agents. Compounds 1a through 1f below were prepared by sequential coupling of *cis*-dicholoroethylene or 1,2-dibromobenzene with propargyl and homopropargyl alcohol and subsequent acetylation.

The diacetate dissolved in acetone was subjected to treatment with PLE and PPL and stirred at room temperature. The pH was kept at 7.8 by intermittent addition of 1 N NaOH. After about 45 percent conversion, the mixture was filtered through celite, and the filtrate was extracted into ethyl acetate, dried and evaporated. The product

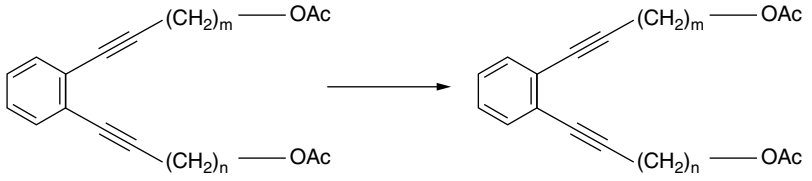


1a
1b
1c

m = n = 1
m = n = 2
m = 1 n = 2

2a
2b
2c

1 2

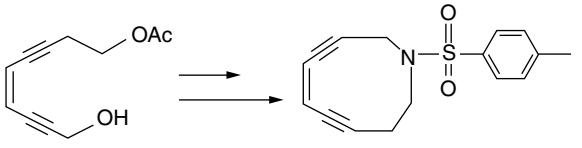


1d
1e
1f

m = n = 1
m = n = 2
m = 1 n = 2

2d
2e
2f

3 4



5 6

monoacetate was isolated by column chromatography (Si-gel). Both compounds 1e and 1f showed very good regioselectivity on being subjected to PPL–catalyzed hydrolysis. The acetate closer to the triple bond, that is, the propargylic one, underwent smooth hydrolysis, leaving the distinct ester intact. This was evident from a comparison of the ^{1}H NMR spectra of the diacetates with those of the hydrolysis products. Interestingly, the regioselectivity is the reverse of what was observed for PLE-catalyzed hydrolysis of substances containing an allylic and a β-alkoxy acetate. In that case, the rigidity imposed by a double bond slowed down the hydrolysis of the allylic acetate. The monoacetate and the unreacted diacetate were separated by column chromatography. The monoacetate compound (5) then was converted to the *N*-containing 10-membered cyclic enediynes compound (6), which has been tested for DNA cleavage.

References

1. Fernandez, Gotor-Vicente, Brieva R, Gotor V. Lipases: Useful biocatalysts for the preparation of pharmaceuticals. *J Mol Enzymatic Catalysts* 40:111–120, 2006.
2. Bailey J, Ollis DF, eds. *Biochemical Engineering Fundamentals*. New York: McGraw-Hill, 1994.
3. Villeneuve P, Pina M, Montet D, Graille J. Determination of lipase specificities through the use of chiral triglycerides and their racemics. *Chem Phys Lipids* 76: 109–113, 1995.
4. Das M, Ph.D thesis; Enzymatic polishing of germinated brown rice, 2006, Indian Institute of Technology, Kharagpur, West Bengal, India.
5. Costas M, Deive FJ, Longo MA. Lipolytic activity in submerged cultures of *Issatchenkia orientalis*. *Process Biochem*. 39:2109–2114, 2004.
6. Spiegel S, Foster D, Kolesnick R. Signal transduction through lipid second messengers. *Curr Opin Cell Biol*. 8:159–167, 1996.
7. Tjoelker LW, Eberhardt C, Unger J, et al. Plasma platelet-activating factor acetylhydrolase is a secreted phospholipase A_2 with a catalytic triad. *J Biotechnol*. 130: 47–56, 1995.
8. Saxena RK, Davidson WS, Sheoran A, et al. Purificationand characterization of an alkaline thermostable lipase from *Aspergillus carneus. Process Biochem* 39:239–247, 2003.
9. Fojan P, Jonson PH, Peterson MT, Peterson SB. What distinguishes an esterase from a lipase? Novel structural approach. *Biochimie*. 82:1033–1041, 2000.
10. Zaidi A, Gainer JL, Carta G, et al. Esterification of fatty acids using nylon-immobilized lipase in *n*-hexane: Kinetic parameters and chain length effects. *J Biotechnol*. 93:209–216, 2002.
11. Sato K, Sudo S. Small-scale solid state fermentations. In Demain AL, Davies JE, eds., *Manual of Industrial Microbiology and Biotechnology*, 2nd ed. Washington, DC: ASM Press, 1999, pp. 61–79.
12. Holker U, Hofer M, Lenz J. Biotechnological advances of laboratory solid state fermentation with fungi. *Appl Microbiol Biotechnol*. 64:175–186, 2004.
13. Kennedy M, Krouse D. Strategies for improving fermentation medium performance: A review. *J Ind Microbiol Biotechnol*. 23:456–475, 1999.
14. Strobel R, Sullivian G. Experimental design for improvement of fermentations. In Demain AL, Davies JE, eds., *Manual of Industrial Microbiology and Biotechnology*, 2nd ed. Washington DC: ASM Press, 1999, pp. 80–93.
15. Banerjee R, Bhattacharya BC. Evolutionary operation (EVOP) to optimize protease biosynthesis by *Rhizopus oryzae*. *Bioprocess Eng*. 8:151–155, 1992.

16. Tunga R, Bannerjee R, Bhattacharya BC. Optimization of *n*-variable biological experiments by evolutionary operations: Factorial design technique. *J Biosci*. 87:224–230, 1999.
17. Thomas JM, Raja R. Designing catalysts for clean technology, green chemistry and sustainable development. *Annu Rev Mater Res*. 35:315–350, 2005.
18. Petkar M, Lali A, Caimi M, et al. Immobilization of lipases for nonaqueous synthesis. *J. Mol Catal B Enzyme.* 30:83–90, 2006.
19. Messing RA. Adsorption and inorganic bridge formation. In Mosback K, ed., *Methods in Enzymology*. New York: Academic Press, 1976, pp. 44, 148–149.
20. Malcata FX, Reyes HR, Garcia HS, et al. Kinetics and mechanisms of reactions catalysed by immobilized lipases. *Enzyme Microb Technol.* 14:426–446, 1992.
21. Cabral JMS, Kennedy JF. Covalent and co-ordination immobilization of proteins. In Taylor RF, ed., *Protein Immobilization*, New York: Marcel Dekker, 1991, pp. 73–138.
22. Bickerstaff GF, ed. *Methods of Biotechnology*, Vol. 1: *Immobilization of Enzymes and Cells*. Totowa, NJ: Humana Press, 1997, pp. 1–11.
23. Tanaka A, Kawamoto T. Cell and enzyme immobilization. In Demain AL, Davies JE, eds., *Manual of Industrial Microbiology and Biotechnology*, 2nd ed. Washington DC: ASM Press, 1999, pp. 94–102.
24. Woodward J. Immobilized enzymes: Adsorption and covalent coupling. In Woodward J, ed., *Immobilized Cells and Enzymes: A Practical Approach*. Oxford, England: IRL, 1985, pp. 3–17.
25. Zhao XS, Bao XY, Guo W, Lee FY. Immobilizing catalysts on porous materials. *Materials Today*. 9:32–39, 2006.
26. Wang YF, Chen ST, Liu KC. Lipase-catalyzed irreversible transesterification using enol esters: Resolution of cyanohydrins and syntheses of ethyl (*R*)-2-hydroxy-4-phenylbutyrate and (*S*)-propranolol. *Tetrahedron Lett.* 30:1917–1920, 1989.
27. Conde S, Fierros M, Rodriguez-Franco M, Puig, C. Resolution of 1-(4-amino-3-chloro-5-cyanophenyl)-2-bromo-1-ethanol by lipase-mediated enantioselective alcoholysis, hydrolysis, and acytylation. *Tetrahedron Asym*. 9:229–2232, 1998.
28. Ahmed K, Ahmad AS, Sandbhor M, Malik SM. Synthesis of enantiopure β-azidoalcohols from their ketoazides by reduction with $NaBH_4$ in the presence of alumina and in situ lipase resolution. *Tetrahedron Asym*. 15:935–939, 2004.
29. Ahmed K, Ahmad AS, Sandbhor M, Malik SM. Synthesis of enantiomerically pure γ-azidoalcohols by lipase-catalyzed transesterification. *Tetrahedron Asym*. 19:1078–1083, 2008.
30. Monterde MI, Rosario B, Sánchez VM, Miguel B, Vicente G. Enzymatic resolution of the chiral inductor 2-methoxy-2-phenylethanol. *Tetrahedron Asym*. 13:1091–1096, 2002.
31. López-Belmonte MT, Alcántara Andrés R, Sinisterra José V. Enantioselective esterification of 2-arylpropionic acids catalyzed by immobilized *Rhizomucor miehei* lipase. *J Org Chem.* 62:1831–1840, 1997.
32. Park HJ, Chei WJ, Huh EC, et al. Production of optically active ketoprofen by direct enzymatic esterification. *J Biosci Bioeng*. 87:845–847, 1999.
33. Moreno JM, Hernaiz MJ, Sánchez-Montero JM,Sinisterra JV, Bustosb MT, Eva Sánchezb M, Bello JF. Covalent immobilization of pure lipases A and B from Candida rugosa. *J. Mol .Cat B: Enzymatic 2*. 177–184,1997.
34. Kato K, Tanaka S, Gong Y, et al. Preparation of optically active trifluoromethylated (3′-indolyl)thiacarboxylic acids, novel plant growth regulators, through lipase-catalyzed enantioselective hydrolysis. *J Biosci Bioeng*. 87:76–81,1999.
35. Basak A, Nag A, Bhattacharya G, et al. Chemoenzymatic synthesis of anti-inflammatory drugs in enantiomerically pure form. *Tetrahedron Asym.* 11:2403–2407, 2000.
36. Fernadez-Solares R, Brieva M, Quir´os I, et al. Enzymatic resolution of a quaternary stereogenic centre as the key step in the synthesis of (*S*)-(+)-citalopram. *Tetrahedron Asym.* 2:341–345, 2004.
37. Widjaja A, Yeh T-H, Yi-Hsu J. Enzymatic synthesis of caffeic acid phenethyl ester. *J Chinese Inst Chem Eng.* 39:413–418, 2008.

38. Rio J Luis del, Ignacio F. Resolution of (±)-*trans*-2-phenylcyclohexan-1-ol by lipases from *Candida rugosa*: Effect of catalyst source and reaction conditions. *Biotechnol. Lett.* 28:1021–1025, 1998.
39. Zheng Y, Wu Xiao-Mei, Branford-White B, et al. Enzymatic synthesis and characterization of novel feruloylated lipids in selected organic media. *J Mol Catalysis B: Enzymatic* 58:65–71, 2009.
40. Monslave NL, Rosseli S, Bruno M, Baldessari A. *J Mol Catalysis B: Enzymatic* 57:40–47, 2009.
41. Sabın J, Gonzalez GV, Rebolledo F. Kinetic resolution of (±)-*trans*- and (±)-*cis*-2-phenylcyclopentanamine by CALB-catalyzed aminolysis of esters: The key role of the leaving group. *Tetrahedron Asymm.* 12:1315–1320, 2002.
42. Roberts SM, Lee Grand DM. Enzyme-catalysed hydrolysis of 3,5-*cis*-diacetoxy-4-*trans*-benzyloxy-methylcyclopentene and the synthesis of aristeromycin precursors. *J Chem Soc Perkin Trans*. 1:1751–1752, 1992.
43. Sih C, Middlesworth FV, Patel DV, et al. Synthesis of the putative biosynthetic triene precursor of monensin. *J Am Chem Soc.* 107:2996–2997, 1985.
44. Mori K, Takhashi Y. Synthetic microbial chemistry: XXIV. Synthesis of antibiotic 1233A, an inhibitor of cholesterol biosynthesis *Liebigs Ann Chem.* 1057–1065, 1991.
45. Akita H, Umezawa I, Nozawa M, and Naeumo S. Total synthesis of (–)-oudemansin X based on enzymatic resolution using immobilized lipase. *Tetrahedron Asym.* 4:757–760, 1993.
46. Ottolina G, Carrea G, Riva S. Synthesis of ester derivatives of chloramphenicol by lipase-catalysed transesterification in organic solvents. *J Org Chem.* 55: 2366–2369, 1990.
47. Yu Zhen S, Zhi Chun C, Wang N, et al. Lipase-catalyzed regioselective synthesis of vinyl ester derivatives of thiamphenicol: Novel thiamphenicol monomers for preparation of macromolecular antibiotic. *Chinese Chem Lett.* 16:45–48, 2005.
48. Panda T, Gowrishankar BS. Production and applications of esterase. *Appl Microbiol Biotechnol*. 67:160–169, 2005.
49. Basak A, Bhattacharjee G, Nag A. Selectivity in enzyme catalyzed reaction: Preferential hydrolysis of saturated methyl ester over the corresponding unsaturated ester by pig liver esterase. *Biotech Lett.* 15:22–23, 1993.
50. Nag A, Nag S, Patil D, Basak A. Comparison of catalytic activities between esterase and lipase in the synthesis of drug, flavor and amide compounds. *J Pharmaceut. Chem.* 42:281–284, 2008.
51. Kinugasa M, Hashimoto S. The reactions of copper(I) phenylacetylide with nitrones. *J Chem Soc Chem Commun*. 466–467, 1972.
52. Basak A, Mahato A, Bhattacharya G, Mukherjee B. Differential effects of stereochemistry and C-4 substituents on the enantioselectivity of PLE- and PPL-catalyzed hydrolysis of 3,4-disubstituted β-lactams. *Tetrahedron Lett.* 38: 643–646, 1997.
53. Mamaghani M, Tabatabaeian K, Ghanadzadeh A, Fatemeh H. An efficient enzymatic method for the separation of stereoisomeric ***cis***- and ***trans-***glycidic esters synthesised via Darzen's condensation reactions. *Tetrahedron Lett*. 44:4775–4777, 2003.
54. Basavaiah D, Pandiaraju S, Muthukumaran K. Enantioselective synthesis of (1*R*,2*R*)- and (1*S*,2*S*)-2-nitroxycyclohexan-1-ols. *Tetrahedron Asym.* 7:13–16, 1996.
55. Asakawa K, Noguchi N, Takashima S, Nakada M. Preparation of a new chiral building block containing a benzylic quaternary stereogenic center and a formal total synthesis of (–)-physostigmine. *Tetrahedron Asym.* 19:2304–2309, 2008.
56. Basak A, Bdour Hussam MM. An easy access to partially protected enediyne diols, important intermediates for cyclic systems, via PPL catalysed hydrolysis. *Tetrahedron Lett*. 29:995–996, 1999.

This page intentionally left blank

CHAPTER 5

Pharmacokinetics and Pharmacodynamics in Drugs

5.1 Introduction

Pharmacokinetics is the quantitative study of drug movement in and through the body. It can be defined as the kinetics of drug absorption, distribution, metabolism, and excretion (KADME) and the interrelationship between drugs themselves and their pharmacologic, therapeutic, and toxicologic responses in humans.[1] Moreover, when this pharmacokinetics is applied in the safe and effective management of individual patients, the subject is known as *clinical pharmacokinetics*.

Now, before examining the KADME mechanism, let us see how the relationship between drug concentration at the site of action and rate of absorption works. The drug concentration-time curve clearly shows the pharmacodynamic and pharmacokinetic parameters.[2]

Let us discuss the pharmacokinetic and pharmacodynamic parameters briefly, which will assess the bioavailability of a drug.[3,4]

5.2 Pharmacokinetic Parameters

Peak plasma concentration (C_{max}). This is the point of maximum drug concentration in plasma. It is known as C_{max} and is expressed as microgram per milliliter. The intensity of pharmacologic response is often related to this factor.

Time of peak concentration (t_{max}). This is actually the time that a drug requires to attain its highest concentration in plasma. This factor can assess the efficacy of a drug used to treat acute conditions such as pain.

Area under the curve (*AUC*). This is actually the total integrated area under the plasma level–time profile and expresses the total

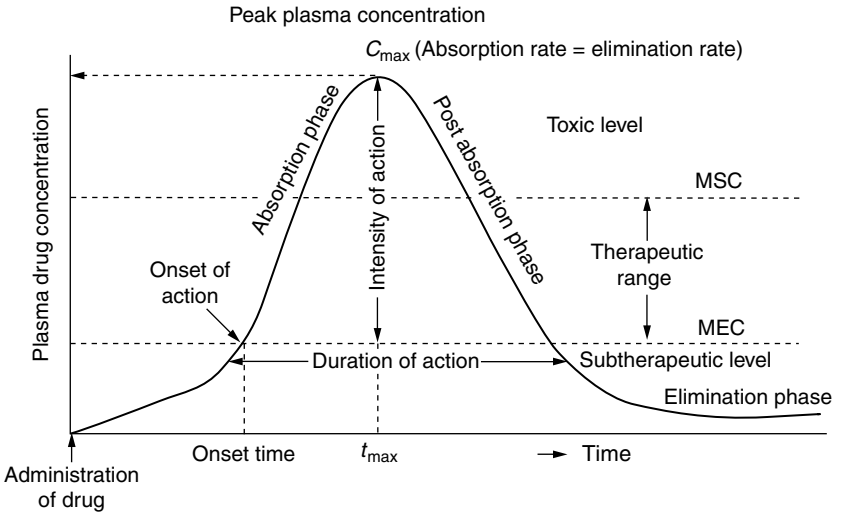


Figure 5.1 A typical plasma concentration–time profile.

amount of the drug that comes into the systemic circulation after administration (Fig. 5.1). It is expressed as micrograms per milliliter times hours. It is a very important parameter in assessing the bioavailability of a drug.[4]

5.3 Pharmacodynamic Parameters

Onset of action. This is the time needed for a drug to produce pharmacologic response.

Onset time. This is the time needed for a drug to produce pharmacologic response.

Minimum effective concentration (*MEC*). This is the minimum concentration of the drug in plasma needed to produce a therapeutic effect.

Maximum safe concentration (*MSC*). This is the concentration of a drug in plasma above which adverse effects are precipitated.

Therapeutic range (*T*). This is the drug concentration between minimum effective concentration and maximum safe concentration.

Duration of action. The duration of a drug action is the time period for which the plasma concentration of a drug remains above the MEC.

Intensity of action. This is the maximum pharmacologic response produced by a drug at its peak plasma concentration.

In a simple manner, pharmacokinetics is actually the kinetics of absorption, distribution, metabolism, and excretion (KADME). In short, it is the sum of all the processes inflicted by the body on the drug.

5.4 Absorption of Drugs

This can be defined as the movement of drug molecules from the site of administration to the systemic circulation. If a drug is absorbed completely but the rate of absorption is very slow, the drug may fail to show a therapeutic response. On the other hand, a rapidly absorbed drug attains therapeutic level immediately.[5]

5.4.1 Factors Affecting Drug Absorption

1. *Aqueous solubility.* A drug given in the form of a watery solution is obviously absorbed at a faster rate. In the case of poorly soluble drugs, dissolution governs the rate of absorption.
2. The larger the absorbing surface, the faster will be the rate of absorption.
3. A drug given in the form of a concentrated solution is absorbed faster than a dilute solution.
4. Whether a drug has been administered orally or parenterally affects the absorption rate to a great extent.
5. Increased blood flow at the site of absorption hastens drug absorption.

The rate and extent of drug absorption from a dosage form are referred to as *bioavailability*. Bioavailability of a drug is 100 percent if the drug is given in the form of an *intravenous injection* (i.v). For subcutaneous or intramuscular injection, bioavailability may be little lower owing to local binding.

However, for an orally administered drug, the bioavailability may be lower because the drug may be incompletely absorbed, or it may undergo *first-pass metabolism* or be excreted in the bile. The systemic absorption of a drug after topical application depends on the lipid solubility of the drugs.[6]

5.4.2 Mechanism of Drug Absorption

For a drug to be absorbed and distributed throughout the body, biotransformed in the body itself, and then eliminated out of the body, it has to cross a number of biologic membranes or barriers. Such a movement of drug is called *drug transport*. Certain principal mechanisms guide the transport of drug molecules across the membranes:

- Passive diffusion
- Pore transport
- Facilitated diffusion
- Active transport

- Ionic or electrochemical diffusion
- Ion-pair transport
- Endocytosis

Passive Diffusion

This is an energy-independent process—hence the name *passive diffusion*. The driving force here is the concentration gradient, which is actually the difference in drug concentration on either side of the membrane. The rate of drug transfer is directly proportional to the concentration gradient between gastrointestinal (GI) fluids and the blood compartment. The greater the membrane/water partition coefficient of the drug, the faster is the absorption. The rate of transfer of ionized drug species is three or four times the rate for ionized drugs.

Fick's first law of diffusion best expresses passive diffusion mathematically:

$$\frac{dQ}{dt} = \frac{DAK_{m/w}}{h}(C_{GIT} - C)$$

where dQ/dt = rate of drug diffusion
D = diffusion coefficient of the drug
$K_{m/w}$ = partition coefficient of the drug between the lipoidal membrane
A = surface area of the absorbing membrane
h = thickness of the membrane
$C_{GIT} - C$ = concentration gradient.

Now, D, A, $K_{m/w}$, and h are constants, and the term $DAK_{m/w/h}$ is replaced by the term P called *permeability coefficient*. Actually, permeability is the case in which a drug can penetrate or diffuse through a membrane.

Pore Transport

Also called *convective transport*, *bulk flow,* or *filtration*, this mechanism allows drug transport through aqueous filled channels or pores of mainly low-molecular weight or low-molecular-size drug molecules. Here, the main driving force is the hydrostatic pressure or osmotic difference. Drug transport promoted by such *water flux* is called *solvent drug*.

Carrier-Mediated Transport

In this case, a component of the membrane called a *carrier* binds reversibly or noncovalently with the solute molecules to be transported. The carrier-solute complex traverses the membrane to the other side, dissociates and discharges the solute molecule, and then returns back to the original site to again accept a fresh molecule.

This transport system is structure-specific, and because the number of carriers is limited, the system is subjected to competition among agents having similar structures. And since the number of carriers is limited, drug absorption will increase with concentration until the system becomes saturated. After saturation of the carriers, drug absorption will attain a constant value. Thus this transport system is capacity-limited.

This process follows a mixed-order kinetics because it is first order at subsaturation drug concentration and apparently zero order at and above saturation levels.[7]

Facilitated Diffusion

This is a type of the carrier-mediated transport system operating down the concentration gradient at a faster rate, the driving force being the concentration gradient itself. Here, no energy expenditure is involved. This transport mechanism is not very useful in the absorption of drugs, but it plays a classic role in the absorption of vitamin B_{12} in the GI tract. The gastric parietal cells secrete the intrinsic factor (IF) glycoprotein that complexes with vitamin B_{12}, and this complex is transported across the intestinal membrane by a carrier system (Fig. 5.2).

Active Transport

By this transport mechanism, a drug is transported from a lower to a higher concentration region. Here, movement occurs against concentration gradient. This energy is required for the work done by the carrier. Since this is an energy-dependent process, metabolic poisons that interfere with energy production, such as fluorides, cyanides, and so on, can inhibit the process. Many antineoplastic agents are absorbed by this mechanism. A good example of active transport is the impaired absorption of levodopa when ingested with meals rich in protein (Fig. 5.3).

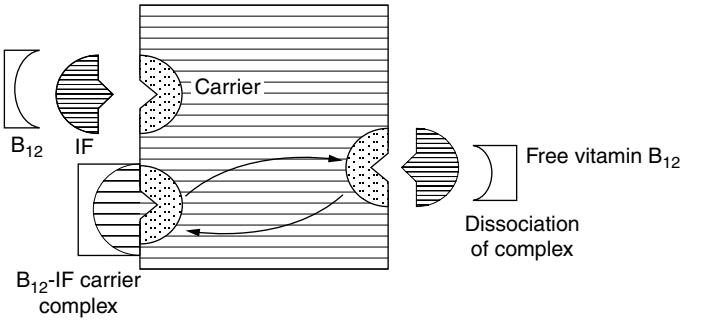


Figure 5.2 Facilitated diffusion of vitamin B_{12}.

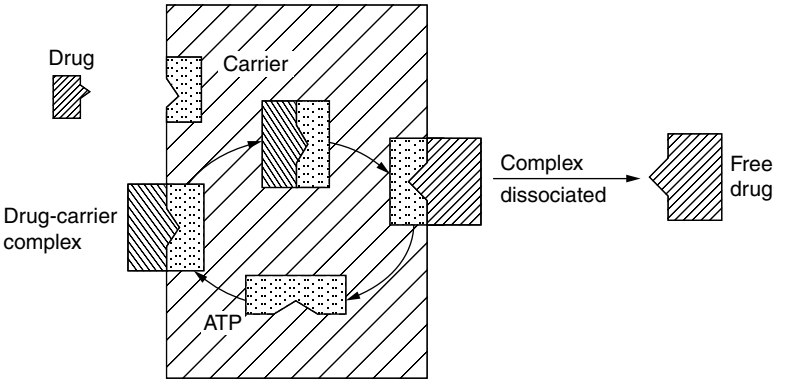


Figure 5.3 Active absorption of drug.

Ion-Pair Transport

This mechanism is used for the absorption of a number of drugs that ionize under all pH conditions. Examples include quaternary ammonium compounds and sulfonic acids. These agents form neutral complexes with endogenous ions of the gastrointestinal tract (GIT) such as mucin. These complexes have the required lipophilicity as well as aqueous solubility.

Electrochemical Diffusion

Mostly ionized drugs are absorbent by this process, where the driving force is the potential difference. A cationic drug is repelled owing to the positive charge outside the membrane. Only cations with a high kinetic energy penetrate the ionic barrier. After getting inside the membrane, the cations are attracted to the negatively charged intracellular membrane, creating an electrical gradient. Here, the drug moves down the electrical gradient (Fig. 5.4).[6,7]

Endocytosis

This involves the engulfing of extracellular materials. Also called *corpuscular* or *vesicular transport*, it involves two processes: (1) *phagocytosis*, or *cell eating*, which is the engulfing of solid materials, and (2) *pinocytosis*, or *cell drinking*, which is actual fluid uptake.

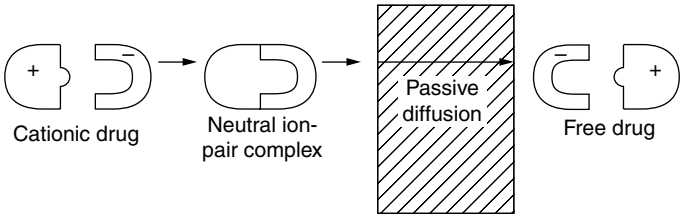


Figure 5.4 Ion-pair transport of a cationic drug.

5.4.3 Factors Influencing Drug Absorption

Both pharmaceutical factors and patient-related factors influence drug absorption. Among the pharmaceutical factors, the physicochemical properties of drug substances and the dosage-form characteristics are to be given important consideration. The anatomic, physiologic, and pathologic conditions of a patient also influence drug absorption.

The solubility of a drug, its dissolution rate, the pK_a and pH values of the drug, its particle size, its surface area, and a number of physicochemical properties of the drug substance influence drug absorption. Similarly, the dissolution time and disintegration time influence the pharmaceutical excipients, nature and type of dosage form, manufacturing variables, and storage conditions of the drug, all of which come under the dosage form characteristics, influence drug absorption to a significant extent.

The two rate-determining processes in the absorption of orally administered drugs are

- Rate of dissolution
- Rate of drug permeation through biomembrane

A necessary criterion for drug absorption by any mechanism is the *aqueous solubility* of the drug and its *dissolution rate*. Aqueous solubility of a drug is a *static process* that is actually the maximum amount of solute that a solvent will dissolve under the standard conditions of temperature, pH, and pressure. Dissolution rate, on the other hand, is a *dynamic process* that is actually the amount of solid substance that goes into solution per unit time under the standard conditions of pH, temperature, solvent composition, and constant solid surface area.

The rate of dissolution can be given by the *Noyes-Whitney equation*:

$$\frac{dc}{dt} = \frac{DAK_{w/o}(C_s - C_b)}{Vh}$$

where D = diffusion coefficient of the drug
A = surface area of the dissolving solid
V = volume of dissolution medium
h = thickness of the stagnant layer
$K_{w/o}$ = water/oil partition coefficient of the drug
$C_s - C_b$ = concentration gradient for diffusion of the drug

The greater the value of the diffusion coefficient of the drug, the faster will be the rate of the dissolution. Dissolution rate will be faster if the surface area of the solid drug is greater, and this can be achieved through micronization of the drug. Similarly, the greater the concentration gradient, the faster will be the diffusion of drug

molecules and hence drug dissolution. In addition, the greater the hydrophilicity of the drug molecules the faster will be its dissolution in aqueous fluids.

According to the Noyes-Whitney equation, the surface area of the drug particles should remain constant during dissolution. However, this is not true for dissolving particles. The decrease in particle size and change in surface area that accompany dissolution are given by the *Hixson and Crowell cubic root law of dissolution*:

$$W_o^{1/3} - W^{1/3} = K_t$$

where W_o = original mass of the drug
W = mass of the drug remaining to dissolve at time t
K = dissolution rate constant

The particle size and surface area of a solid drug are inversely related to each other. A decrease in particle size can be achieved by micronization, which will increase the surface area of the drug and enhance the dissolution rate. However, it is the effective surface area and not the absolute surface area that influences dissolution. The more effective the surface area, the more intimate will be the contact between solid surface and aqueous solvent, and the faster will be the dissolution rate. Drugs such as griseofulvin, chloramphenicol, and several salts of tetracycline have poor aqueous solubility. Micronizations of these drugs have shown excellent results in increasing their dissolution rates. Micronization has allowed decreases in the dosages of some drugs because of increased absorption efficiency. However, micronization of hydrophobic drugs such as aspirin, phenacetin, and phenobarbital has resulted in a fall in the dissolution rate.

Actually, hydrophobic drugs float on the dissolution medium by adsorbing air on their surfaces. Micronized particles have surface charges, and there may be electrically induced agglomeration of the drug particles. And reaggregation of the particles owing to high surface free energy will result in larger particles that will float on the surface of the dissolution medium. However, there are drugs such as penicillin G and erythromycin in which particle size reduction in order to increase surface area is not advisable because it produces a number of undesirable effects.

To improve the dissolution rate, particle size reduction through micronization can be done, although it has some adverse consequences. However, particle size can be reduced to the submicron level through the use of specialized techniques such as formation of molecular dispersion or solid solution by molecularly trapping the drug in the lattice of a hydrophilic agent such as cyclodextrin or by forming solid dispersions by dispersing the drug in polyethyl glycol (PEG), polyvinyl pyrrolidone (PVP), and so on.

Most drugs are either weak acids or weak bases. To enhance drug absorption, one of the easiest approaches is to improve the drug's solubility and dissolution rate, and one way to achieve this objective is to convert the drug to its salt form. At a given pH, the solubility of a drug, whether acidic, basic, or salt form, is a constant.

The pH-partition theory explains the process of drug absorption from the GIT and its distribution across biologic membranes. The theory states that for drug compounds of molecular weight greater than 100, which are primarily transported across biomembranes by passive diffusion, the process of absorption is governed by

1. The lipid solubility of the un-ionized drug
2. The pH at the absorption site
3. The dissociation constant pK_a of the drug

Drugs are weak electrolytes, and their ionization depends on the pH of the biologic fluid. The drug that exists in un-ionized form is a function of the dissociation constant pK_a of the drug and the pH of the fluid at the absorption site.

From the concept of pK_a of the drug and pH at the absorption site, the amount of ionized and un-ionized drug in solution can be determined by the *Henderson-Hasselbach equation*:

$$\text{For weak acids, } \text{pH} = pK_a + \log \frac{\text{conc. of ionized drug}}{\text{conc. of un-ionized drug}}$$

$$\% \text{ Drug ionized} = \frac{10^{\text{pH}-pK_a}}{1+10^{\text{pH}-pK_a}} \times 100$$

$$\text{For weak bases, } \text{pH} = pK_a + \log \frac{\text{conc. of un-ionized drug}}{\text{conc. of ionized drug}}$$

$$\% \text{ Drug ionized} = \frac{10^{pK_a-\text{pH}}}{1+10^{pK_a-\text{pH}}} \times 100$$

When the concentration of ionized and un-ionized drug becomes equal, then pH = pK_a. This pK_a is a characteristic of the drug. Weakly acidic drugs ionize more at an alkaline pH, and one scale change in pH causes a 10-fold change in ionization; weakly basic drugs ionize more at an acidic pH.

Un-ionized forms of acidic drugs will cross the membrane of the gastric mucosal cells and be converted to ionic form and then very slowly pass to the extracellular fluid. This is called as *ion trapping*. It is by this mechanism that aspirin causes gastric mucosal cell damage. Drugs such as phenytoin, several barbiturates, and ethosuximide are very weak acids and remain un-ionized at all pH values, and hence their absorption is rapid and independent of GIT pH.

Absorption of drugs such as aspirin, ibuprofen, phenylbutazone, and a number of penicillins whose pK_a values range from 2.5 to 7.5 is pH-dependent. Very strongly acidic drugs are poorly absorbed in the GIT.

On the other hand, many benzodiazepines, caffeine, and theophylline remain un-ionized at all pH values, and their absorption remains rapid and pH-independent. However, strongly basic drugs are ionized in the entire pH range of the GIT and are poorly absorbed. There are some basic drugs whose pK_a values lie in the range of 5 to 11, and their absorption is pH-dependent. Examples of such drugs include amitriptyline, imipramine, and chloroquine.

For a drug to be well absorbed and show good bioavailability, it should have sufficient aqueous solubility to dissolve in the fluids at the absorption site and high enough lipid solubility to affect the partitioning of the drug in the lipoidal membrane and thus into systemic circulation. Therefore, it is true that a perfect hydrophilic-lipophilic balance (HLB) should be inherent to the drug molecule itself for optimal bioavailability.

So far we have discussed how different physicochemical properties of a drug influence its absorption. Now we are going to see how some of the patient-related factors and even the anatomy/physiology of the GIT influence the absorption of drugs and thus their bioavailability.

Gastrointestinal pH

The pH of the GI fluid influences drug absorption in several ways. It has already been discussed that dissolution is an important rate-limiting step in drug absorption. Because drugs are either weak acids or weak bases, their solubility is influenced by pH. Thus the GI pH is of great significance in the oral bioavailability of the drugs. Weakly acidic drugs dissolve rapidly in the alkaline pH of the intestine, and basic drugs dissolve in the acidic pH of the stomach.[6,7] The disintegration of some pharmaceutical dosage forms is pH-sensitive, and special mention may be made of enteric coated formulations.

GI pH influences drug absorption by determining the amount of drug that is allowed to exist in the un-ionized form at the site of absorption. The chemical stability of some drugs is also influenced by GI pH. The acidic pH of the stomach may cause the degradation of some drugs such as penicillin, and the alkaline intestinal pH also can interfere with the action of basic drugs.

Gastric Emptying

The passage of drug from the stomach to the small intestine is called *gastric emptying*. Rapid gastric emptying is said to increase the bioavailability of a drug. A large number of factors influence this process.

The physical and mental states of the patient, any disease condition, the nature of the drugs being taken, GI pH, and the volume, composition, physical nature, and temperature of the meal all influence gastric emptying.

The method of gastric emptying of a drug can be studied by using radioopaque contrast medium such as barium sulfate or tagging the drug with a radioisotope and scanning the stomach at regular intervals. Rapid gastric emptying is recommended in case of rapid onset of drug action, for enteric coated formulations, and for drugs that are unstable in gastric pH. Gastric emptying can be promoted by taking the drug on an empty stomach. Delay in gastric emptying is recommended in the case where food promotes drug dissolution and absorption. If the drug dissolves slowly, the disintegration and dissolution of the drug are promoted by gastric fluids, as in case of absorption of vitamins C and B_{12}.

GT infections and diseases, gastrectomy, intestinal surgery, and hepatic diseases such as cirrhosis influence drug absorption to a considerable extent. Two intestinal disorders that cause malabsorption of drugs are *celiac disease* and *Crohn's disease.* Antineoplastics and alcohol also cause malabsorption of drugs. GI infections such as shigellosis, gastroenteritis, cholera, and food poisoning also result in the malabsorption of drugs.

Gastrointestinal Contests

The GI contents can influence drug absorption to a considerable extent. Within the GI tract, after a drug enters, there may be *food-drug interactions*, *drug-drug interactions*, or interaction between the drug and other GI constituents. All these interactions have an overall effect on the drug absorption procedure.

Food can either enhance or delay drug absorption or may not interfere with the process at all. The general concept says that drugs are better absorbed under fasting conditions, and the presence of food retards or prevents absorption. However, a drug taken half an hour before or after meal is not affected by food.

Increased drug absorption on interaction with food may occur because of enhanced solubility owing to GI secretions, increased time for dissolution of poorly soluble drugs, increased lymphatic absorption, and longer absorption-site contact of the drug, as well as the prolonged residence time. Again, there may be delayed drug absorption owing to the presence of food because of the formation of unabsorbable complexes. In addition, the presence of food may increase the viscosity of the medium and thus interfere with drug dissolution and disintegration. Moreover, because of food, there may be delayed gastric emptying that will affect drugs such as penicillins that are unstable in stomach or delayed transit of enteric-coated tablets into the intestine.

Although the absorption of methyldopa and propylthiouracil is unaffected by food, the absorption of aspirin, acetaminophen, and

diclofenac is delayed because of food. Absorption of the drugs such as griseofulvin, diazepam, and water-soluble vitamins is increased on interaction with food, but food decreases the absorption of drugs such as the penicillins, levodopa, erythromycin, and the tetracyclines.[8]

The bile salts, digestive enzymes, and mucin, a protective mucopolysaccharide lining the GI mucosa, influence drug absorption to a considerable extent. When two or more drugs are coadministered, there will be a drug-drug interaction in the GIT affecting the absorption of both, such as, for example, when antacid preparations containing heavy metal salts of Ca or Mg interfere with coadministered drugs such as Promazine and lincomycin.

Now we are going to discuss how the dosage-form characteristics, effect of pharmaceutical excipients, and different formulations influence drug absorption.

Pharmaceutical Excipients

Generally, it is not possible to administer a drug in its original form. To prepare a drug formulation, whether it is a tablet, capsule, or any other dosage form, a number of nondrug components are used other than the drug. These nondrug components are the *pharmaceutical excipients*. These excipients serve to increase consumer acceptance of the product, have a role in the physicochemical stability of the product during shelf life, and affect the optimal bioavailability and functional ability of the drug product. These pharmaceutical excipients, though inert in nature, have a role in the absorption of drugs.

Excipients include vehicles, diluents, binders, lubricants, disintegrants, surfactants, complexing and suspending agents, buffers, and colorants. If the dosage form is liquid, oral or parenteral (injections) vehicles or solvent systems can be used, and they may be aqueous vehicle, nonaqueous water-immiscible vehicle such as vegetable oils, or nonaqueous water-miscible vehicle such as glycerol, sorbitol, propylene glycol, and so on. The rate and extent to which a vehicle mixes with the biologic fluid affect the bioavailability of the drug. The more the vehicles mixes with the body fluids, the more quickly the drug will be absorbed. Viscosity of the vehicle also affects drug absorption.[9]

Binders are used in the tablet formulations to promote the cohesive compacts for directly compressible materials and to ensure that the tablet remains intact after compression. Some binders are starch, acacia, PVP, and so on. However, if the proportion of the binder becomes large, it will increase the hardness of the tablet and thus decrease the disintegration/dissolution rate of the tablet. Thus, with a greater amount of binders, drug absorption may be hampered.

Disintegrants, which are by nature hydrophilic, help a tablet to go into solution by overcoming the cohesive strength of the tablet, breaking the tablet so that it comes into contact with water. These agents are an important prerequisite for a tablet to dissolve. *Microcrystalline cellulose* is a very good disintegrant.

Diluents are commonly added to a tablet or capsule formulations when the required dose is inadequate to produce the necessary bulk. Diluents may be organic or inorganic. Hydrophilic powders may promote the dissolution of poorly water-soluble drugs by forming a coating over the drug particles, rendering them hydrophilic. Sometimes a drug-diluent interaction may lead to poor bioavailability of the drug.

Lubricants are agents that increase the flow properties of powders and granules in tablet formulations. They reduce the interparticle interaction. These substances, because they are hydrophobic, inhibit wettability and penetration of water into a tablet and its disintegration and dissolution.

Tablet coating, either sugar-coated or enteric-coated, may have a deleterious effect on the dissolution profile of a tablet.

Colorants, even in very low concentration, can have a negative impact on the dissolution rate of crystalline drugs.

Suspending agents actually impart viscosity. They stabilize the solid drug particles by reducing their rate of settling and by increasing the viscosity of the medium. They influence drug absorption in several ways. An increase in viscosity of these agents acts as a mechanical barrier to diffusion of the drug from the dosage form into the bulk of GI fluids and from GI fluids to the mucosal lining by forming a viscid layer on the GI mucosa. They also retard the GI transit of drugs.

Complexing agents influence the properties of drugs to a great extent. Sometimes such agents hinder drug absorption by forming poorly soluble or poorly absorbable complexes. However, complexation can enhance the bioavailability of a drug either by forming soluble complexes or by increasing membrane permeability and enhancing lipophilicity.

Surfactants influence drug absorption in a complex manner. They increase drug absorption by increasing the effective surface area, thereby promoting wetting. They also increase membrane contact with the drug and enhance membrane permeability. They also retard drug absorption by forming unabsorbable drug-micelle complexes, and if the surfactant concentration is large, it may show laxative action.

Buffers, though, create the real atmosphere for drug dissolution, the presence of K^+ cations may inhibit drug absorption, as seen in case of vitamin B_{12} and sulfanilamide. This is due to the fact that uptake of fluids by intestinal epithelial cells decreases the effective drug concentration in the tissues and thus reduces the absorption rate.

5.4.4 Manufacturing Variables

Drug dissolution is an obvious factor in drug absorption. Manufacturing processes influence drug dissolution from solid dosage form. The granulation technique, the amount of compression force

used in tabletting, and the intensity of packing of capsule contents have a direct impact on the dissolution profile and thus drug absorption.

A drug in the form of syrup or elixir undergoes rapid absorption in the body because the major rate-limiting step, dissolution, is absent. However, when a drug is administered in the form of a suspension, the major rate-limiting step in drug absorption is drug dissolution, which is generally rapid owing to the large surface area of the particles. Here, particle size, wetting agents, suspending agents, and viscosity of the medium all influence drug absorption.[6,7,9]

5.5 Distribution of Drugs

After a drug is absorbed from any site, by any one mechanism, it enters into the systemic circulation, where it is subjected to different processes such as *distribution* and *biotransformation*, and then is finally excreted out of the body.

Distribution is the reversible transfer of the drug throughout the blood, tissues, and extravascular fluids.[10–12]

If we assume the human body to be a homogeneous compartment, the volume of drug distribution is given by

$$V = \frac{\text{dose administered i.v.}}{\text{plasma concentration}}$$

5.5.1 Factors Affecting Drug Distribution

- Lipid solubility, that is, the lipid:water partition coefficient of the drug
- Ionization of the drug at physiologic pH, that is, the pK_a value of the drug
- Extent of plasma and tissue protein binding
- Differences in regional blood flow
- Effect of diseases such as cirrhosis and uremia

Drug distribution shows how extensively drugs are bound to plasma and tissue proteins and also sequestered in other tissues. However, a drug is not uniformly distributed throughout the body.

5.5.2 Factors Affecting Differences in the Distribution of Drugs

- Perfusion rate of the drug to different organs and tissues
- Drug binding to different blood and tissue components
- Physicochemical barriers to drug diffusion

- Physicochemical properties of the drug such as pK_a value and oil/water partition coefficient
- Other important factors such as age, diet, obesity, disease state, drug interactions, pregnancy, and so on

If a drug is ionic, polar, or water-soluble and has to diffuse through highly selective physiologic barriers, its distribution will be *permeability rate-limited*. If, on the other hand, the drug is lipophilic and has to diffuse through highly permeable capillaries and muscles, drug distribution is said to be *perfusion rate-limited*.

Perfusion rate is actually the volume of blood that flows per unit time per unit volume of the tissue. Lungs, liver, kidney, heart, brain, and adrenals are highly perfused organs; muscles and skin are moderately perfused, and bone and fat are said to be poorly perfused. Perfusion rate is given in milliliter square per minute.

The first-order distribution rate constant K_t is given by the equation

$$K_t = \frac{\text{perfusion rate}}{K_{t/b}}$$

where $K_{t/b}$ = tissue/blood partition coefficient of drug. The tissue distribution half-life is given by the equation

$$\text{Distribution half-life} = \frac{0.693}{K_t} = \frac{0.693K_{t/b}}{\text{perfusion rate}}$$

The extent to which a drug is distributed in a particular tissue or organ depends on the tissue volume and tissue/blood partition coefficient of the drug.

5.5.3 Plasma Protein Binding

A drug in the body binds to tissue and plasma proteins, blood components, and a number of extravascular components. Here, we are going to discuss *plasma protein binding* and how it affects drug distribution.[10]

Acidic drugs bind to plasma albumin, and basic drugs bind to α-amino acid glycoprotein. The extent of binding depends on the compound itself. The percentage of binding is actually the usual therapeutic plasma concentration of a drug. Drugs that bind to albumin include warfarin, tetracyclines, penicillins, sulfonamides, barbiturates, nonsteroidal anti-inflammatory drugs, phenytoin, and benzodiazepines. Drugs that bind to α_1-acid glycoprotein include β-blockers, lidocaine, methadone, verapamil, prazosin, imipramine, and quinidine.

A few implications of plasma protein binding are as follows:

1. Drugs that are highly bound to plasma proteins are restricted mostly to vascularized areas and have lower volumes of distribution.

2. The amount of drug that remains bound to the plasma protein is not available for action. The bound drug remains in equilibrium with the free drug in plasma, and the bound drug dissociates when the concentration of free drug in plasma is reduced owing to elimination.
3. When plasma concentrations of a drug are expressed, they refer to both bound and free drug.
4. High plasma protein binding of a drug makes a drug long-acting.
5. A competitive displacement interaction occurs among drugs binding to the same site. Obviously, a drug that binds with a higher affinity will displace a drug that bound with a lower affinity. However, drug binding may be reduced in the case of pathologic conditions such as hypoalbuminemia.

As with plasma protein binding, drugs also bind to specific tissue components. Drugs can get sequestered in various tissues and will have a large volume of distribution and longer duration of action, as in tetracycline on bone and teeth. It has already been discussed that drug distribution is either permeability rate-limited or perfusion rate-limited. Now we will discuss how specialized physiologic barriers affect drug distribution.

5.5.4 Physiologic Barriers

- Simple cell membrane barrier
- Simple capillary endothelial barrier
- Blood-brain barrier
- Placental barrier
- Cerebrospinal fluid barrier
- Blood-testis barrier

Drugs, whether ionized or un-ionized, if their molecular size is less than 600 Da, can diffuse easily through the capillary endothelium. Thus the endothelium is not really a barrier for drugs.[8,10] A simple cell membrane acts as a lipoidal barrier in the GI absorption of drugs.

Brain capillaries consist of the endothelial cells that are joined to one another by continuous tight intercellular junctions comprising of what is called the *blood brain-barrier* (*BBB*). This BBB is a lipoidal barrier that allows only drugs with a high oil/water partition coefficient to diffuse passively. It is much less permeable to water-soluble drugs and allows low penetration of moderately lipid-soluble and partially ionized drugs. Thus, in order to treat central nervous system (CNS) disorders, appropriate choice of a drug that will cross the BBB

is an important part of therapy. However, the trigger area and the median hypothalamic eminence area of the brain do not contain the BBB. Some drugs, if they can be administered intranasally, can diffuse directly to the CNS.

For drug targeting to the brain that crosses the BBB, some of the successful approaches include:

- To use permeation enhancers such as dimethyl sulfoxide (DMSO)
- To use drug carriers such as the dihydropyridine redox system
- To use the internal carotid artery; if it is infused with mannitol, there will be osmotic disruption of the BBB

The *blood-CSF barrier* is made up of the choroids plexuses of the lateral, third, and fourth ventricles, and the composition is similar to extracellular fluid of the brain. The permeability characteristics of the blood-CSF barrier is similar to the BBB. The blood-CSF barrier allows highly lipid-soluble drugs to cross with relative ease, but moderately lipid-soluble and partially ionized drugs permeate slowly.

The *blood-testis barrier* is located at the Sertoli–Sertoli cell junction, which restricts the passage of drugs to spermatocytes and spermatids.

The physicochemical parameters affecting drug distribution are molecular size, degree of ionization, and partition coefficient. Drugs are either weak acids or weak bases, and their degree of ionization in plasma depends on the pK_a value. Polar and hydrophilic drugs that ionize at plasma pH cannot penetrate the lipoidal cell membrane, and here, tissue permeability is the rate-limiting step in the distribution of the drugs. The extent to which a drug exists in un-ionized form governs the distribution pattern. Situations such as acidosis and alkalosis influence drug distribution. It is seen that acidosis favors extracellular distribution, whereas alkalosis favors intracellular distribution.

The effective partition coefficient of a drug is the driving force for drug distribution. It is calculated by the formula

$$\text{Effective } K_{o/w} = \text{fraction un-ionized at pH 7.4} \times K_{o/w} \text{ of the un-ionized drug}$$

Other miscellaneous factors such as age, diet, obesity, and pregnancy all influence drug distribution.

While discussing the distribution of drugs, another important parameter to be discussed is the *volume of distribution*. A constant relationship exists between the drug concentration in plasma C and the total amount of drug in the body X. It is given by $X = V_d C$. The proportionality constant V_d is the *apparent volume of distribution*, which is actually the hypothetical volume of body fluid in which the drug is

to be dissolved. The apparent volume of distribution and real volume of distribution are not essentially the same.

Real volume of distribution actually refers to the total-body water content, which is actually the sum total of intracellular, extracellular, and vascular fluid content.

The volume of such physiologic fluid compartments can be estimated with tracers or markers. The *plasma fluid content* can be determined using *Evans blue, indocyanine green*, *I-131*, and *albumin*. Total-body water (TBW) can be determined using heavy water (D_2O), tritiated water (HTO), and antipyrine. Extracellular fluid (ECF) can be determined using nonmetabolizable saccharides such as raffinose, insulin, and mannitol using some radioisotopes.

The apparent volume of distribution V_d is a characteristic of each drug under normal conditions, and it may be altered if any changes occur in the drug distribution pattern. The value of V_d ranges from 3 to 40,000 liters.

A generalization can be made about V_d. Drugs that bind selectively to plasma proteins have V_d values that are smaller than their true volume of distribution. Some of the drugs that bind selectively to extravascular tissues have V_d values that are larger than the real volume of distribution. If any factor causes an alteration in drug binding to blood components, it will cause an increase in the value of V_d, and if drug binding to extravascular tissues is altered, it will cause a decrease in V_d.

5.6 Biotransformation/Metabolism

The term *biotransformation* is synonymous with *metabolism*. It can be defined as the conversion of a drug from one chemical form to another. Drugs are actually chemical substances that are not nutrients to the body. They enter the body through inhalation, ingestion, or absorption by virtue of their lipophilicity and hence are termed as *xenobiotics*.[13]

Drugs are xenobiotics because they are exogenous compounds that are foreign to the body. Moreover, if such foreign substances or drugs continuously accumulate in the body, they can precipitate toxic reactions. Thus the human body has a metabolic system that transfers water-insoluble lipophilic nonpolar drugs to polar water-soluble compounds that are excreted easily. Thus the metabolism or biotransformation of drugs is really a *detoxification process*.[14] The liver is the principal site for drug metabolism, although other organs such as the lungs, skin, placenta, and kidneys have drug metabolizing ability.

Microsomal and nonmicrosomal enzymes catalyze a vast majority of biotransformation reactions. Lipid soluble substrates interact nonspecifically with microsomal enzymes, and these enzymes biotransforms the substrates to water-soluble metabolites that are excreted easily. Nonmicrosomal enzymes are present either in soluble

form in the cytoplasm or are attached to the mitochondria; mostly they are nonspecific in nature, and they catalyze a variety of biotransformation reactions.[15]

Before going to the types of biotransformation reactions, let us put a light on the various factors affecting biotransformation reactions:

1. A number of physicochemical properties of a drug, such as its molecular size/shape, pK_a values, lipophilicity, and other characteristics, influence its interactions during the biotransformation process.
2. Most metabolic reactions are influenced by enzymes, which are under genetic control. There is genetic variation in relation to drug-metabolizing ability—the subject at hand being called *pharmacogenetics.* Marked racial variation is observed in drug metabolism of isoniazid, phenytoin, dapsone, sulfadimidine, and succinyl choline etc. However, drugs such as antipyrine, phenylbutazone, and dicoumarol show very little genetic variation with regard to drug metabolism.
3. Rate of drug metabolism has sex-related differences because many such processes are influenced by the sex hormones. Again, there is a marked difference in drug metabolism in different age groups that is mainly attributed to variations in enzyme content, enzyme activity, and hemodynamics.
4. Drug metabolism is influenced in various stages of pregnancy, in hormonal imbalances, in women receiving oral contraceptive, or in those with changes in adrenocorticotropic hormone (ACTH) level because of stress.
5. An individual's diet influences drug metabolism to a great extent. The protein carbohydrate ratio, percentage of fat in the diet, any type of vitamin deficiency, the intake of alcohol, starvation, and malnutrition all have an influence on drug metabolism.
6. Disease states such as renal impairment, cirrhosis, hepatitis, obstructive jaundice, and hepatic carcinoma all influence biotransformation reactions.
7. A number of environmental chemicals such as DDT and cigarette smoke have enzyme-induction effects on drug metabolism. In addition, organophosphate insecticides and heavy metals have an inhibitory effect on the drug-metabolizing ability of enzymes.
8. There are certain agents that increase the drug-metabolizing ability of enzymes; such agents are called *inducers*. And agents that inhibit the drug-metabolizing ability of an enzyme are called *inhibitors*. Certain drugs stimulate their own metabolism, and the phenomenon is called *autoinduction* or *self-induction.*

Microsomal enzyme induction occurs as a result of an interaction between drugs, insecticides, and/or carcinogens with DNA that causes increased synthesis of microsomal enzyme protein, specially *cytochrome P-450* and *glucuronyl transferase*. As a result, the rate of metabolism of the inducing drug is increased. Owing to enzyme induction, the rate of drug metabolism in increased by two- to four-fold and can reach the maximum limit in 4 to 14 days and then returns to normal in 1 to 3 weeks.[16,17]

Drugs whose metabolism is affected by enzyme induction include chloramphenicol, theophylline, oral contraceptives, phenytoin, warfarin, griseofulvin, phenylbutazone, imipramine, and tolbutamide. As a result of enzyme induction, tolerance may develop in the case of drugs such as rifampin and carbamazepine because they are subjected to autoinduction.

Use of an inducer can interfere with the dose adjustment of drugs such as antihypertensives, and oral hypoglycaemics, which are taken on a regular basis. Drugs such as acetaminophen develop toxicity because they are activated by metabolism, resulting in an increase in the intensity of action.

Because of enzyme induction, some drugs such as oral contraceptives are inactivated by metabolism, and their duration of action may decrease. Thus there will be a failure of contraception.

Inhibition of drug-metabolizing enzymes can be due to changes in enzyme activity because of interaction at the enzymic site and because of competition between structurally similar compounds competing for the same site on the enzyme. If the metabolic product competes with the substrate for the same enzyme, a decrease in enzyme content occurs owing to a fall in enzyme synthesis, hormonal imbalance, or nutritional deficiency.[17]

Drugs that inhibit drug-metabolizing enzymes include allopurinol, cimetidine, disulfiram, isoniazid, diltiazem, monoamine oxidase (MAO) inhibitors, omeprazole, verapamil, sulfonamides, ketoconazole, and erythromycin.

Biotransformation of drugs leads to the following facts*:*

- There may be pharmacologic inactivation of the drugs.
- Active metabolite can be formed from an active drug. There is no change in pharmacologic action.
- Inactive drugs or prodrugs may get activated. This is the pharmacologic activation of the drug.
- Because of biotransformation, there may be a total change in pharmacologic action.
- In rare cases, there may be toxicologic activation of the drugs.

Biotransformations of amphetamines to phenylacetone, phenytoin to *p*-hydroxyphenytoin, and salicylic acid to salicyluric acid are

examples of pharmacologic inactivation of a drug. Here, active drugs are converted to inactive forms. The tissue acylating intermediate of isoniazid and conversions of acotaminophen to reactive metabolites causing hepatic necrosis are examples of toxicologic inactivation of drugs. When codeine is converted to morphine, digitoxin to digoxin, diazepam to temazepam, imipramine to desipramine, and amitriptyline to nortriptyline, no change in pharmacologic action is observed because metabolites with equal activity are being produced.[18–20]

Inactive prodrugs may be biotransformed to active metabolites, and there is a pharmacologic activation of the drugs. A few examples include the following:

Prodrug	Active Metabolite
Aspirin	Salicylic acid
Levodopa	Dopamine
Prednisone	Prednisolone
Cyclophosphamide	Aldophosphamide, acrolein, phosphoramide mustard
Bacampicillin	Ampicillin
Enalapril	Enalaprilat
Benorylate	Aspirin and acetaminophen
Proguanil	Proguanil triazine

Again, biotransformations can bring about a total change in pharmacologic activity. Iproniazid which is an antidepressant, may be biotransformed to Isoniazid, which is actually an antitubercular drug. In addition, a renowned tranquilizer such as diazepam, when converted to oxazepam, behaves as an anticonvulsant.

Drug biotransformation reactions are divided into the phase I and phase II reactions.[21] Phase I reactions include oxidation, reduction, and hydrolytic reactions. Phase II are mostly the conjugation reactions, which may be conjugations with glucuronic acid, α-amino acids, sulfate moieties, glutathione and mercapturic acid and acetylation and methylation reactions.

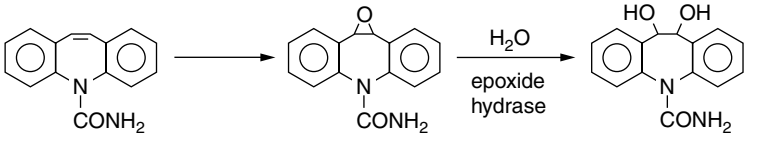
A wide variety of *oxidation* reactions takes place. The notable ones are the oxidation of olefins (C=C bonds), aromatic carbons, aliphatic and alicyclic carbons, benzylic and allylic carbon atoms and carbon atoms alpha to carbonyls and imines, alcohols, carbonyls, and acids. Oxidation of the carbon-heteroatom system also should be mentioned here and includes *N*-dealkylation, *N*-hydroxylation, *N*-oxide formation, oxidation, *S*-dealkylation, and desulfuration.

Reductive reactions mostly include the reduction of *N*-compounds, alcohols, carbonyls, C=C bonds, and so on. The *hydrolytic reactions* include the hydrolysis of esters, ethers, amides, hydrolytic cleavage of nonaromatic heterocycles, hydrolytic dehalogenation, and so on.[22–24] Now we will discuss the phase I reactions step by step with suitable examples.

Oxidative Reactions

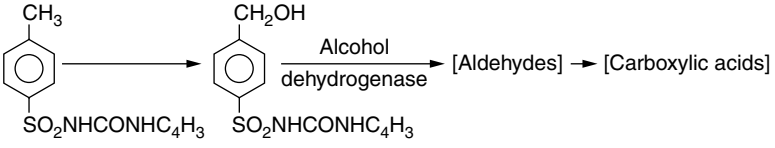
Oxidations are rather the most common and most important metabolic reactions because living beings derive energy mostly by the oxidative combustion of organic molecules containing carbon and hydrogen atoms. Oxidations increase the hydrophilicity of the xenobiotics by the introduction of polar functional groups.

1. Oxidation of nonaromatic carbon-carbon double bond proceeds via the formation of epoxides to yield the 1,2-dihydrodiols.[25–27]



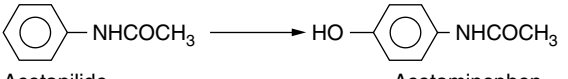
Carbamazepine — Carbamazepine-10,11-epoxide — Trans-10,11-dihydroxy carbamazepine

2. Tolbutamide is converted to corresponding carbinol, which is further oxidized to aldehydes and then to carboxylic acids. Here, oxidation of the benzylic carbon atoms actually takes place.

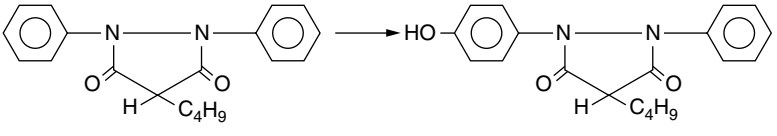


Tolbutamide — Primary carbinol

3. Oxidation of aromatic carbon atoms or aromatic hydroxylation.

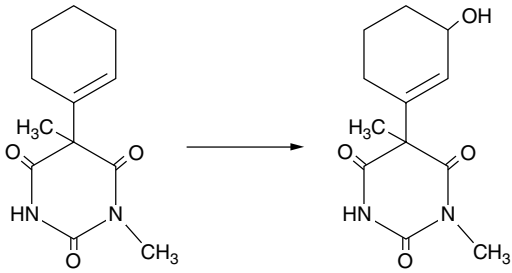


Acetanilide — Acetaminophen



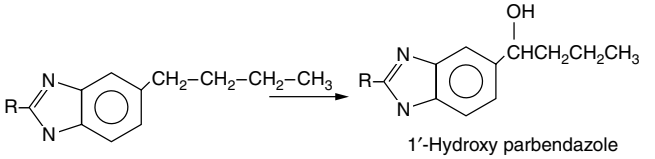
Phenyl butazone — Oxyphenbutazone

4. Carbon atoms adjacent to olefinic double bonds are allylic carbon atoms. An example of its oxidation is given below.

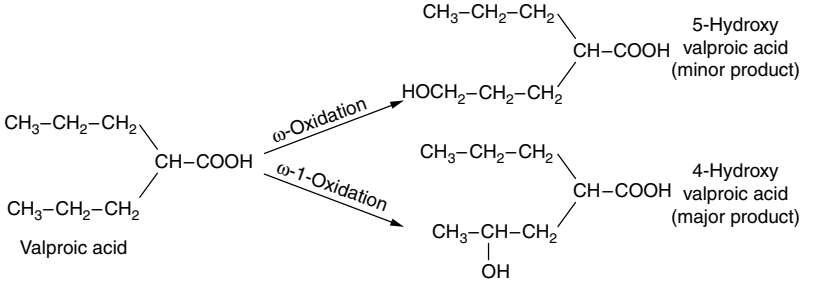


Hexobarbital 3′-Hydroxyhexobarbital

5. Alkyl or aliphatic carbon atoms may undergo ω-oxidation at the terminal methyl group or ω-1-oxidation at the penultimate carbon atom. Valproic acid through ω-oxidation produces 5-hydroxy valproic acid as the minor product and through ω-1-oxidation produces 4-hydroxy valprole acid as the major product. ω-1-Oxidation of secondary and tertiary penultimate carbons yield the corresponding alcohols. Parbendazole is converted to 1-hydroxyparbendazole; here, hydroxylation of aliphatic side chains attached to an aromatic ring occurs at benzylic methylene groups. This is an example of ω-1-oxidation.

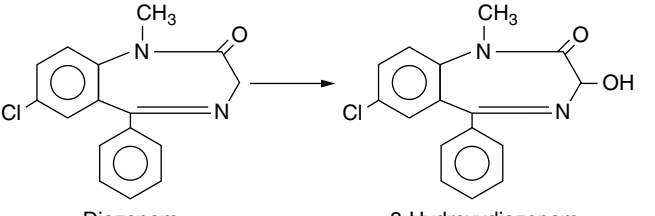


Parbendazole 1′-Hydroxy parbendazole



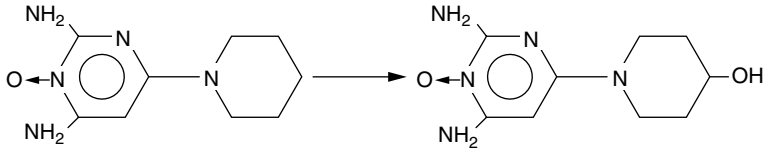
Valproic acid

6. Oxidation of carbon atoms that are alpha to carbonyl (C=O) and imino (C=N) functions.



Diazepam 3-Hydroxydiazepam (*N*-methyloxazepam)

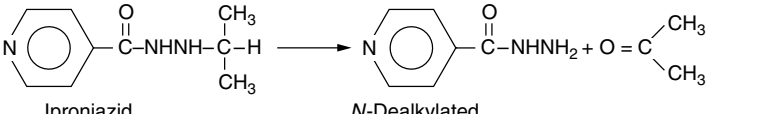
7. An example of oxidation of alicyclic carbon atoms is given below.



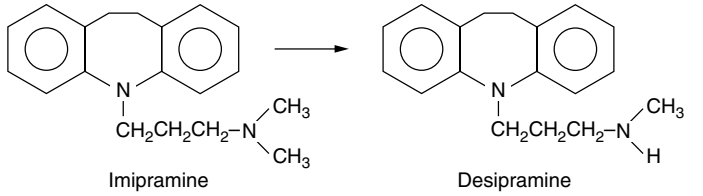
Minoxidil 4′-Hydroxy minoxidil

Now, when carbon-heteroatom systems undergo biotransformation reactions, either there will be hydroxylation of the carbon atom attached to the heteroatom and then cleavage at carbon-heteroatom bond, for example, *N*-, *O*-, and *S*-dealkylation, oxidative deamination, and desulfuration, or there may be oxidation of the heteroatom itself such as the *N*- and *S*-oxidation.[27,28]

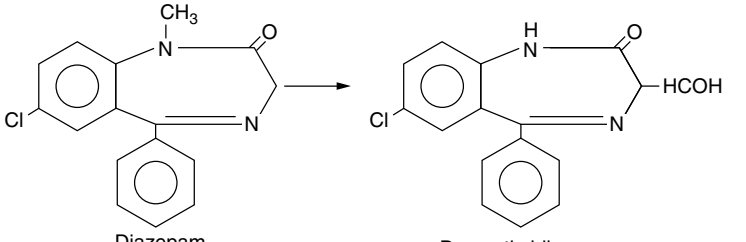
N-Dealkylation generally involves the oxidation of the alpha carbon to generate the carbinolamine intermediate that rearranges by the cleavage of C—N bond to yield the *N*-dealkylated product and the corresponding carbonyl of the alkyl group. Some of representative examples of compounds under going *N*-dealkylation are given below.



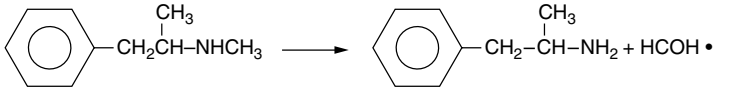
Iproniazid *N*-Dealkylated product Acetone



Imipramine Desipramine

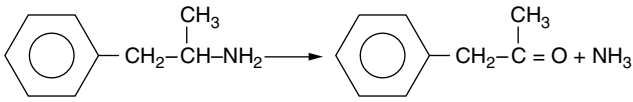


Diazepam Desmethyldiazepam



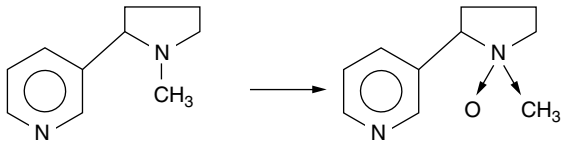
Methamphetamine Amphetamine

Oxidative deamination also proceeds via the carbinolamine pathway, and here, the C—N bond cleavage occurs at the bond that links the amino group to the larger portion of the drug molecule.

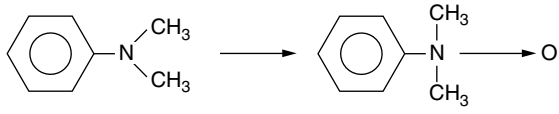


Amphetamine Phenylacetone

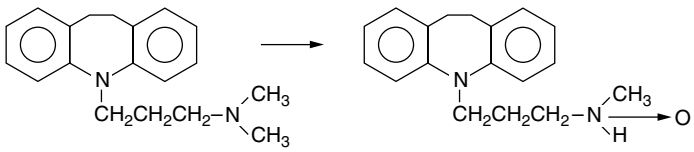
Aliphatic amines, alicyclic amines, aromatic heterocycles, and amines attached to aromatic rings generally yield the *N*-oxides.



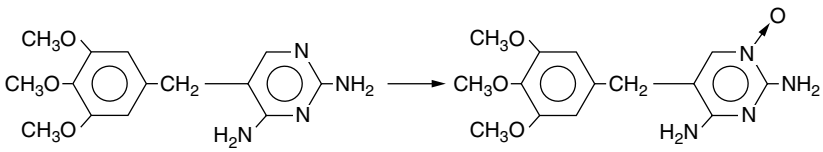
Nicotine Nicotine-1′-*N*-oxide



N, *N*-Dimethylaniline *N*-Oxide metabolite

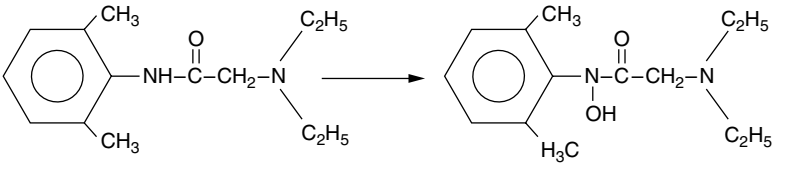


Imipramine Imipramine-*N*-oxide

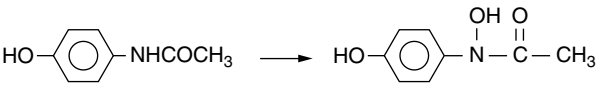


Trimethoprim Trimethoprim-1-*N*-oxide

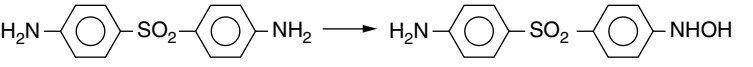
Mostly nonbasic nitrogen atoms undergo *N*-hydroxyl formation. Nonbasic aromatic amines also undergo this type of hydroxylation. *N*-Hydroxylation is also possible in the case of basic nitrogens such as primary and secondary amines. Some of important examples are cited below.



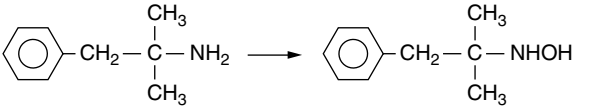
Lidocaine *N*-Hydroxylidocaine



Acetaminophen *N*-Hydroxy metabolite



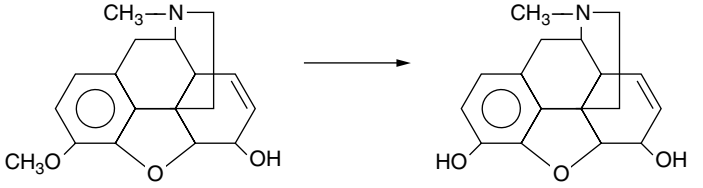
Dapsone *N*-Hydroxy dapsone



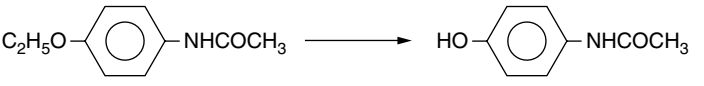
Phentermine *N*-Hydroxy phentermine

O-Dealkylation

Mostly ethers undergo *O*-dealkylation reactions. This reaction generally leads to the formation of active metabolites. It is similar to *N*-dealkylation, and proceeds via α-carbon hydroxylation to form an unstable hemiacetal or hemiketal intermediate that spontaneously undergoes C—O bond cleavage to form alcohol and a carbonyl moiety.



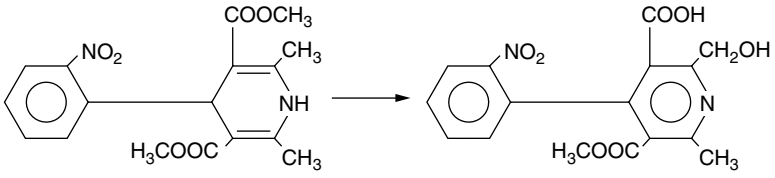
Codeine Morphine



Phenacetin Acetaminophen

Alcohols, carbonyls, and carboxylic acids also exhibit oxidation reactions, the reactions being catalyzed by nonmicrosomal enzymes such as dehydrogenases.

Dehydrogenation Reactions



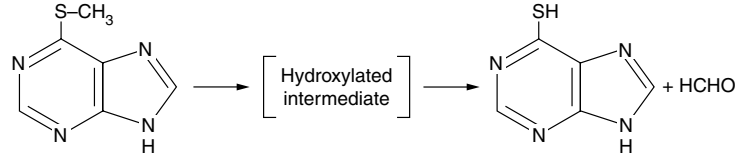
Nifedipine Pyridine metabolite

Oxidative Dehalogenation

These types of reactions are common in halogen atom–containing drug.

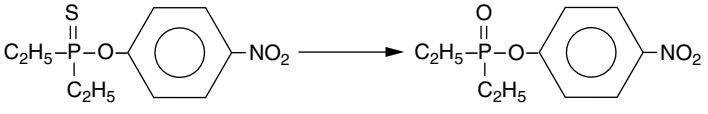
Now we will provide a few examples of the oxidation of the carbon-sulfur systems.

1. 6-Methyl mercaptopurine converted to 6-mercaptopurine is an example of *S*-dealkylation.



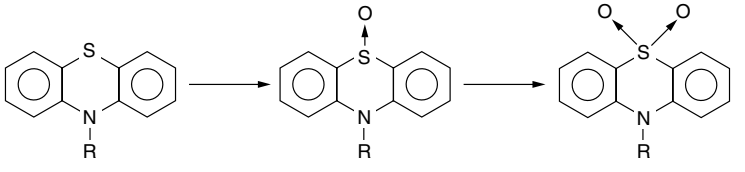
6-Methylmercaptopurine 6-Mercaptopurine

2. When parathion is converted to paraoxon, it is an example of desulfuration.



Parathion Paraoxon

3. Thioethers undergo *S*-oxidation to yield the sulfoxides, which may be further oxidized to the sulfones.

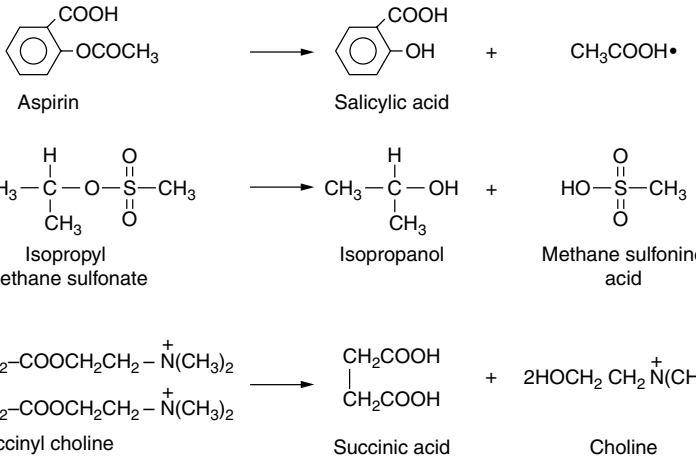


Phenothiazine Phenothiazine sulfoxide Phenothiazine sulfone

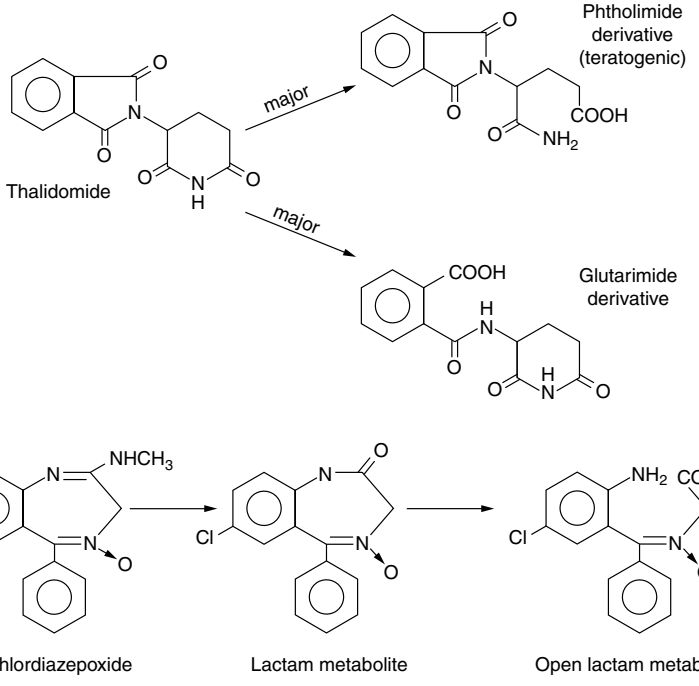
Hydrolytic Reactions

The important functional groups that undergo hydrolytic reactions are the esters, amides, ethers, and hydrazides. Because of hydrolysis, metabolites may result with complete loss of activity. In addition,

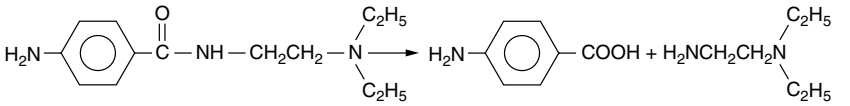
prodrugs on hydrolysis may lead to the formation of an active form, so there will be pharmacologic activation of the drug. After the hydrolytic cleavage of many drugs, if the hydrolyzed group contains the pharmacophore, then it may retain much of its activity. Examples of esters undergoing hydrolysis are listed below.

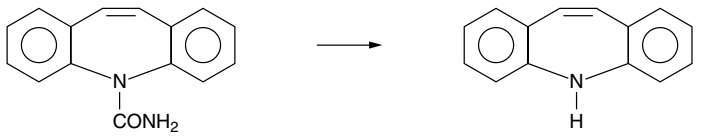


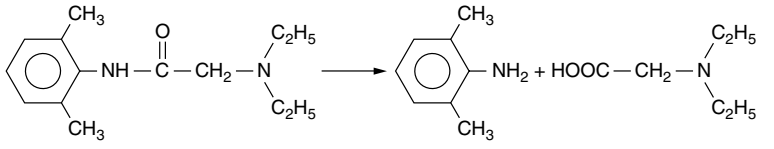
Hydrolysis of some lactams is given below.



Hydrolysis of amides is catalyzed by the amidases, resulting in C—N bond cleavage to yield carboxylic acid and amine.



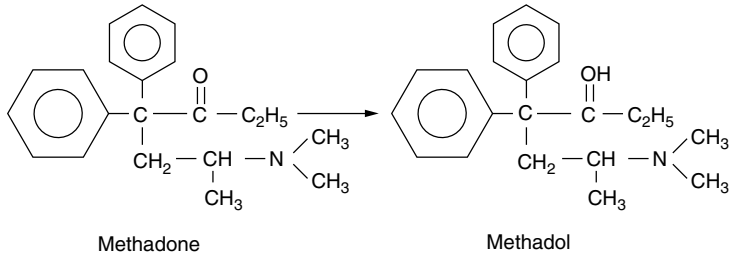




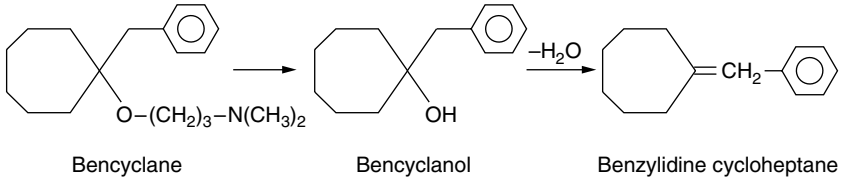
Reduction Reactions

Bioreductions are mostly reversible reactions that may result in a prolongation of action. In many cases, it is just the opposite of the oxidation process. There may be a reduction of the carbonyls, alcohols, N compounds, and C=C double bonds.[27,28] A few examples of bioreductions are cited below.

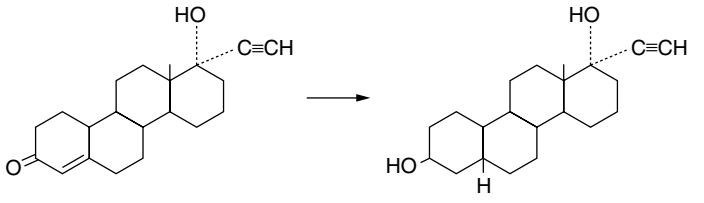
1. Methadone, which is an aliphatic ketone, is reduced to methadol.



2. The alicyclic ketone naltrexone is reduced to an isomorphine derivative.
3. Bencyclane, an antispasmodic before an alcohol is reduced, is dehydrated to a C=C bond.

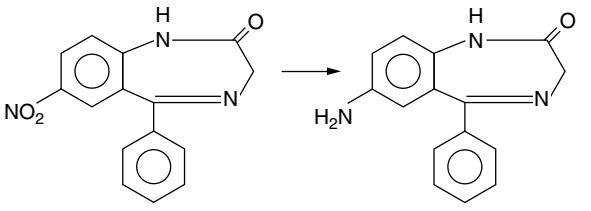


4. An unsaturated carbonyl compound on reduction results in the formation of alcohol, and the C=C double bond is reduced.

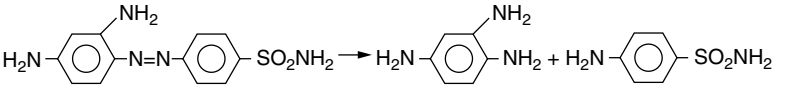


Norethindrone 3β,5β-Tetrahydronorethindrone

When the nitro group is reduced, hydroxylamine and nitroso groups are formed as intermediates that eventually yield the amines. In addition, the azo compounds via the formation of hydroso intermediates yield primary amines.



Nitrazepam 7-Aminometabolite



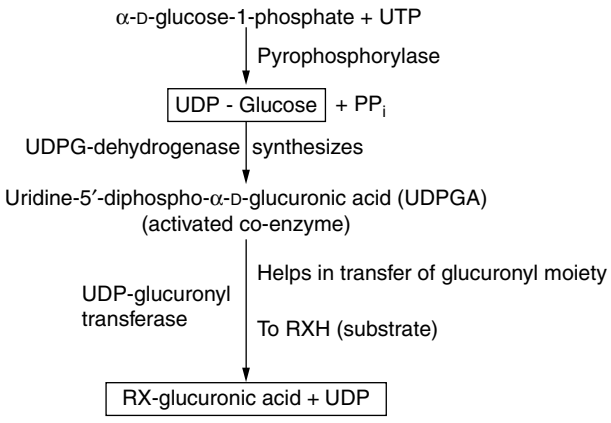
Prontosil 1,2,4-Triaminobenzene Sulfonilamide

So far we have discussed the various types of phase I reactions. Now we are going to discuss phase II reactions, which are actually *conjugation reactions*.[24–28] These reactions are actually drug detoxification reactions. The products of phase I reactions may conjugate with glucuronic acids, mercapturic acids, glutathione, amino acids, and sulfate moieties in the phase II reactions. Some of the important conjugation reactions are discussed below.

Conjugation with Glucuronic Acid

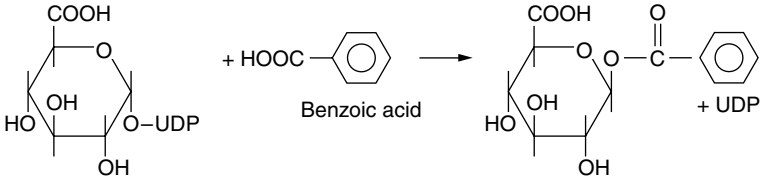
This is also known as *glucuronidation*.

- Conjugation with *D*-glucuronic acid also occurs to a high degree.
- There may be oxygen, nitrogen, sulfur, and carbon glucuronide formations.
- The steps of glucuronide formation are presented below:



The stable conjugate in the reaction is α-D-glucuronide, also called *glucosiduronic acid* or *glucopyranosiduronic acid conjugate*.

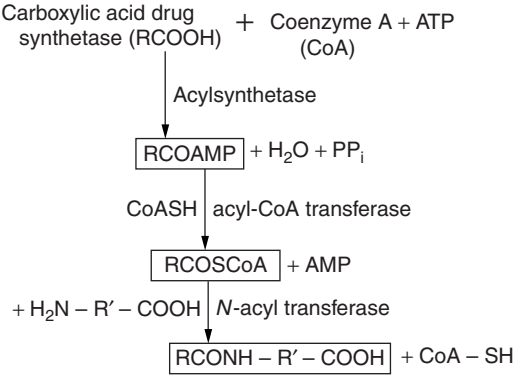
Benzoic acid combines with uridine-diphospho-glucuronic acid (UDPGA) to form the glucuronide of benzoic acid.



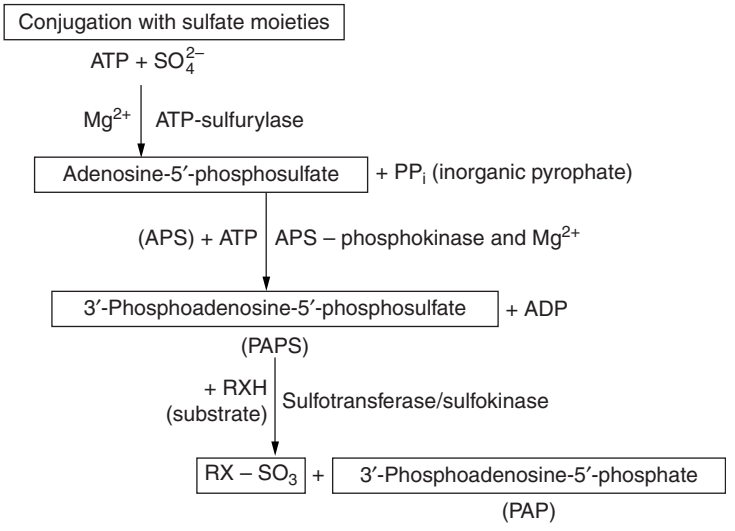
O-glucuronides, *N*-glucuronides, *C*-glucuronides, and *S*-glucuronides have been identified.

Conjugation with Amino Acids

The carboxylic acid drug substrate is activated by reacting with ATP and coenzyme A (CoA) to form the *acyl-CoA intermediate*. In the next step, the α-amino acid will be acylated in presence of acyl-CoA by action of the enzyme *N*-acyltransferase. However, these reactions occur to only a limited extent owing to the limited availability of amino acids.



Conjugation reactions with amino acids occur mostly with glycine and glutamine in liver mitochondria, and such reactions are used to estimate hepatic function. The diagnostic marker for the purpose is mostly benzoic acid. Benzoic acid conjugates with glycine to yield hippuric acid, which is rapidly excreted in urine; the extent and rate of hippuric acid excretion are an indicator of liver function.

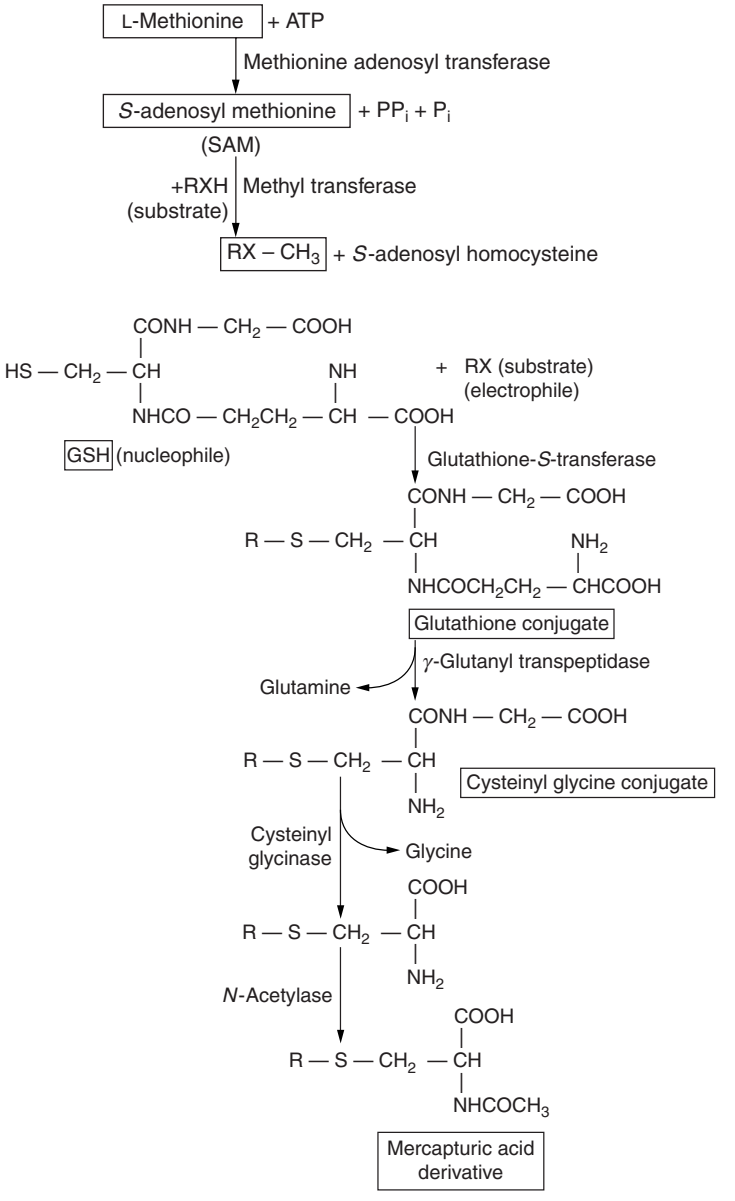


Glutathione Conjugation

This is a very important detoxification route from the perspective that highly electrophilic metabolites react with tissue nucleophilic groups such as –OH, $–NH_2$, and –SH and precipitate toxicities such as necrosis, carcinogenesis, teratogenesis, and so on. Glutathione or γ-glutamyl cysteinyl glycine (GSH), which itself is a strong nucleophile, is highly reactive to these electrophiles.

Acetylation and *methylation* are also important phase II detoxification reactions. Acetylation is an important metabolic pathway for drugs containing primary amino groups. It is basically an acylation reaction; the acylating agent is $CH_3COSCoA$, and enzyme involved is nonmicrosomal *N*-acetyl transferase. There is a pharmacogenetic difference in the rate of acetylation in humans called *acetylation polymorphism*. With respect to acetylation, there are slow acetylators and rapid acetylators.

Methylation is an intermediate of phase I and phase II reactions; it is a phase I reaction because it is just the opposite of the demethylation reaction, and because of its two-step mechanism involving synthesis of active coenzyme and transfer of the methyl group to a substrate, it also can be classed as a phase II reaction.



Sulfation	Glucuronidation
1. Catalyzed by nonmicrosomal Enzymes	1. Catalyzed by microsomal mixed-function oxidases
2. Occurs less frequently	2. Occurs more frequently
3. Occurs at low substrate concentration	3. Occurs at high substrate concentration

Examples of drugs undergoing glucuronidation: Chloramphenicol, acetaminophen, desipramine, fenoprofen, cyproheptadiene, meprobanate, *N*-hydroxy coumarin, and *N*-hydroxy dapsone.

Examples of drugs undergoing sulfation: Acetaminophen, salbutamol, and steroids.

Examples of drugs undergoing amino acid conjugation: Salicylic acid, phenylacetic acid, nicotinic acid, isopropoxyacetic acid, and cholic acid.

Examples of drugs undergoing acetylation: para-amino salicylic acid, *para*-amino benzoic acid, dapsone, histamine, mescaline, isoniazid, sulfanilamide, procainamide, and hydralazine.

Examples of drugs undergoing methylation: Morphine, methyldopa, L-DOPA, isoprenaline, nicotine, histamine, propylthiouracil, normorphine, and 6-mercaptopurine.

5.7 Excretion of Drugs

Drugs are absorbed into our body and are distributed throughout the body; then, under different types of biotransformation reactions, they are finally removed from the body by the process of *excretion*. Excretion of a drug in its intact form marks the termination of its pharmacologic action.[28–30]

Although kidneys are the main organ of excretion, other nonrenal router include the lungs, sweat glands, salivary glands, the biliary system, and the intestine. The rate of excretion is given by the equation

$$\text{Rate of excretion} = \text{rate of filtration} + \text{rate of secretion} - \text{rate of reabsorption}$$

The kidney is the primary organ for the excretion of drugs, and the urinary excretion of the drugs is governed by such processes as glomerular filtration, active tubular secretion, and active/passive tubular reabsorption.

Clearance

Clearance is used to describe renal excretion, and it is also a measure of kidney function. This is known as *total-body clearance* or *total systemic clearance*. It is actually the sum of both renal and nonrenal clearance. It also can be defined as the hypothetical volume of body fluids containing the drug that is to be removed completely in a specified time.

$$\text{Clearance (Cl)} = \frac{\text{elimination rate}}{\text{plasma drug concentration}}$$

$$\text{Renal clearance } (Cl_R) = \frac{\text{rate of urinary excretion}}{\text{plasma drug concentration}}$$

$$Cl_R = \frac{\text{rate of filtration} + \text{rate of secretion} - \text{rate of reabsorption}}{C}$$

From these equations, it appears that renal clearance is actually the ratio of the sum of the rate of glomerular filtration and active secretion minus the rate of reabsorption to plasma of drug concentration C''.

Renal excretion or renal clearance is affected by a number of factors, including the following:

1. The physicochemical factors affecting the renal excretion of a drug are pK_a, lipid solubility, and molecular weight. Drug molecules with a molecular weight below 300 Da, if they are water-soluble, are easily excreted by the kidneys. Those whose size ranges from 300 to 500 Da are excreted in both the bile and the urine. Drug molecules with size greater than 500 Da are excreted in the urine to a lesser extent.

   The pH-dependent excretion of any particular compound also depends on its pK_a and lipid solubility. Reabsorption is also influenced by the lipid solubility of the drug. Excretion of an acidic or basic drug that is polar in its un-ionized form is independent of the pH of urine and its flow rate. The rate of excretion of very weakly acidic and basic drugs that are mostly un-ionized throughout the entire range of urine pH is low and insensitive to urine pH. However, the rate of excretion of very strongly acidic or basic drugs is always high and insensitive to the pH of urine.

   Since a lipophilic drug is reabsorbed to a large extent, the urinary excretion of an unchanged drug is inversely related to its lipophilicity.
2. Renal blood flow is important in the case of drugs that are excreted only by glomerular filtration and for those which are actively secreted.
3. The plasma concentration of a drug and the distribution and binding characteristics of a drug also influence renal clearance. Renal clearance is inversely related to the apparent volume of distribution of a drug. A drug that has a large V_d value is poorly excreted in urine. For drugs that are actively secreted, the rate of excretion increases with increase in plasma concentration.
4. Age, sex, differences in genetic makeup, species and strain differences, and circadian rhythms also alter drug excretion.
5. Drug excretion is greatly impaired in the case of renal dysfunction. Disease states such as diabetes, hypertension, hypovolemia, pyelonephritis, and uremia all impair renal clearance of drugs.
6. *Drug interactions.* If a drug interaction causes any alteration in renal blood flow, binding characteristics, urine pH, intrinsic clearance, and/or forced dieresis, it would alter the renal clearance of the drug. Urinary excretion of digoxin is reduced by diazepam. All diuretics increase the elimination of drugs whose renal clearance is affected by the urine flow rate.

It is true that the kidney is the main route for drug excretion, but there are other nonrenal routes for drug excretion, including the following:

Pulmonary excretion. Mostly gases and inhalational anesthetics are absorbed and released through the lungs, and the mechanism involved is simple diffusion.

Salivary excretion. Drugs are excreted in saliva by passive diffusion, and it is predictable based on the pH partition hypothesis.

Biliary excretion. A drug whose biliary concentration is less than that in plasma has a small biliary clearance, and vice versa. Secretion of drugs in bile is influenced by several factors, including physicochemical properties of the drug itself, nature of the biotransformation process, age, sex, drug interactions, protein-drug binding, disease state, and even the route of drug administration. The ability of the liver to excrete a drug in bile is called *biliary clearance.*

$$\text{Biliary clearance} = \frac{\text{biliary excretion rate}}{\text{plasma drug concentration}}$$

$$= \frac{\text{bile flow} \times \text{biliary drug concentration}}{\text{plasma drug concentration}}$$

Enterohepatic circulation of drugs, which is actually drug cycling between the intestine and liver, is very important for conservation of important endogenous substances such as vitamin B_{12}, folic acid, vitamin D_3, steroid hormones, and bile salts.

Mammary excretion. This is a passive process and hence governed by the pH partition hypothesis. Drug excretion in milk is really an important issue from the point of view of a breast-feeding infant. Drugs that bind extensively to plasma protein, such as diazepam, are less secreted in milk. Drugs such as the barbiturates, morphine, and ergotamine may induce toxicity in infants. Nicotine is also secreted in the milk of mothers who smoke. Thus a breast-feeding mother is advised to avoid drugs and smoking.

5.7.1 Pharmacokinetic Equations

We have already discussed the KADME mechanism of pharmacokinetics. Now let's see how absorption, distribution, and excretion of intravenously and orally administration of drugs can be represented by different equations.[31–33]

Elimination rate constant, elimination half-life, and clearance are the three parameters that characterize the elimination phase.

$$t_{1/2} = \frac{0.693}{K_E}$$

This equation represents the elimination half-life. K_E is the elimination rate constant. It also can be represented as

$$\ln X = \ln X_0 - K_E t$$

where X_0 is the amount of drug at time $t = 0$, that is, the initial amount of drug injected. This equation in exponential form becomes

$$X = X_{0e}^{K_E t}$$

Now, converting it to common logarithms, we get

$$\log X = \log X_0 - \frac{K_E t}{2.303}$$

Clearance is another important parameter in clinical drug applications. Clearance is actually the theoretical volume of body fluid containing the drug from which the drug is completely removed in a given period of time. It is denoted by Cl_T. Cl_T, or *total systemic clearance*, is an additive property of individual organ clearances.

For drugs given by i.v. bolus,

$$\mathrm{Cl}_T = \frac{X_0}{AUC}$$

where *AUC* is the area under the curve. Cl_T also can be represented by the following equations:

$$\mathrm{Cl}_T = \frac{K_E X}{C}$$

$$\mathrm{Cl}_T = K_E X V_d$$

$$\mathrm{Cl}_T = \frac{0.693 V_d}{t_{1/2}}$$

V_d actually represents the *extent of distribution* of the drug and is expressed in Its.

$$V_d = \frac{\text{amount of drug in the body}}{\text{plasma drug concentration}} = \frac{X}{C}$$

$$V_d = \frac{X_0}{C_0} = \frac{\text{i.v. bolus dose}}{C_0}$$

For a drug given as an i.v. bolus,

$$V_d \text{ (area)} = \frac{X_0}{K_E AUC}$$

The *absorption rate constant* K_a is given by the equation

$$\frac{dx}{dt} = K_a X_a - K_E X$$

where, K_a is the first-order absorption rate constant, and X_a is the amount of drug at the absorption site remaining to be absorbed.[34–36] Transforming the equation in concentration terms, we get

$$C = \frac{K_a F X_0}{V_d (K_a - K_E)} (e_e - K_E t_{-e} - K_a t)$$

5.8 Pharmacodynamics

Pharmacodynamics deals with the actions of drugs in human body. It clearly elucidates the drug action-effect sequence, how a drug acts, and how the action of one drug is modified by the action of another drug. When a drug is administered, it does not impart any new function to any organ or system; rather, it alters the space of ongoing activity. A drug in the body may stimulate, depress, cause irritation or replacement, and also cause a cytotoxic effect. The stimulatory action of a drug enhances the activity of specialized cells. Examples include epinephrine and pilocarpine, which stimulate the heart and salivary glands, respectively. The opposite of stimulation is *depression*, which diminishes the activity of specialized cells, for example, barbiturates depress the CNS. *Replacement* is the use of natural metabolites, hormones, or congeners in deficiency states, for example, levodopa in parkinsonism and insulin in diabetes mellitus.

The actions of some drugs may be noxious for some specialized cells. Mild irritation, though, may stimulate function, for example, bitters that increase gastric secretion, but strong irritation can cause inflammation, necrosis, or even morphologic damage. Antimicrobials and antineoplastic drugs mostly exert cytotoxic action in the body.

If a person is administered two or more drugs simultaneously, the drugs either may be indifferent to each other or may show *synergism* or *antagonism.* The drug interaction may take place at the *pharmacokinetic* or *pharmacodynamics* level.[37] If the action of one drug is increased by the action of the other drug, it is a *synergistic* action. Synergism may be *additive* or *supra-additive*. If the effects of two drugs are in the same direction and simply add up, this is an *additive* effect. For example, in the combination of aspirin and acetaminophen, they both act as analgesics or antipyretics. Also, in the combination of sulfadiazine plus sulfamethazine plus sulfamerazine, all three act as antibacterials. Ephedrine and theophylline both act as bronchodilators. However, in a supra-additive effect, the effect of the combination is greater than the individual effects of the components. Thus the effect of drug A + B is greater than the effect of drug A + effect of drug B.

On the contrary, if one drug inhibits the action of another drug, this is known as *antagonism*. Depending on the mechanism involved, antagonism may be

Physical. For example, the adsorption of alkaloids by charcoal, which is used in alkaloids poisoning.

Chemical. Where two drugs react chemically to form an inactive product. For example, Tannins and alkaloids react to form insoluble alkaloidal tannate.

Physiological. Where two drugs act on different receptors or by different mechanisms but have opposite overt effects on the same physiologic function, for example, the action of insulin and glucagons on blood sugar level or the action of histamine and adrenaline on the bronchial muscles and blood pressure.

Receptor-mediated. Receptor antagonism is specific. The antagonist interferes with the binding of the agonist to its receptor or inhibits the generation of a response consequent to such binding. This type of antagonism may be competitive or noncompetitive.

For a drug to produce a certain degree of response in a patient, it should be given in an appropriate amount, referred to as the *dose* of a drug. The dose of a drug is governed by its inherent potency, that is the concentration in which it should be present at the target site and its pharmacokinetic characteristics. Some drugs have a wide safety margin, and a fixed dose is appropriate for most patients. This is known as the *standard dose.* It is applicable in the cases of some oral contraceptives and antimicrobials. In some cases, the drug may modify a finely regulated body function, and the dose is accurately adjusted by repeated measurement of the affected physiologic parameter, for example, antihypertensives, hypoglycemics, diuretics, anticoagulants, and general anesthetics. Drug dose in these cases is known as a *regulated dose.*

There are cases where an empirical dose is given at the beginning of treatment to obtain a target level, and adjustments are made later by actual monitoring of plasma concentrations. If proper drug level monitoring is not possible, adjustments can be made by observing the patient for a relatively long interval. This is *target-level dosing*. It is applicable for antidepressants and antiepileptics. The dose needed to produce maximal therapeutic effect cannot be given because of intolerable adverse effects. Thus the optimal dose is arrived at by titrating the drug against an acceptable level of adverse effects. This is known as a *titrated dose*, and it is used for anticancer drugs and corticosteroids.

However, a great advantage can be achieved by providing two or more drugs in a fixed-dose ratio. These fixed-dose combinations are convenient and have better patient compliance, and since most of the combinations are synergistic, there is a better therapeutic effect, and the side effects of one component may be counteracted by the other. However, such fixed-dose combinations suffer from such disadvantages as the drug dose needs to be adjusted and individualized, the time course of action of the components may be different, and the patient may not need one of the drugs present in the combination.

5.8.1 Factors Influencing Drug Action

The differences among individuals are responsible for the variations in drug response. Individuals differ in pharmacokinetic handling of drugs; there are variations in neurogenic or hormonal tone or concentrations of specific constituents. A few of the notable factors modifying drug action are discussed below.

1. *Age.* A drug dose for children generally is calculated by the use of some formulas:

$$\textit{Young's formula}: \text{Child dose} = \frac{\text{age}}{\text{age}+12} \times \text{adult dose}$$

$$\textit{Dilling's formula}: \text{Child dose} = \frac{\text{age}}{20} \times \text{adult dose}$$

   However, after first year of life, drug metabolism is faster than in adults. In aged people, renal function may decline progressively, and drug doses need to be reduced. The hepatic microsomal drug-metabolizing activity and liver blood flow also are reduced. This necessitates a reduction in drug doses in the elderly.

2. *Body size.* This influences the concentration of drug attained at the site of action:

$$\text{Individual dose} = \frac{\text{BW(kg)}}{70} \times \text{ average adult dose}$$

   where BW is body weight. However, body surface area (BSA) provides a more accurate basis for dose calculation:

$$\text{Individual dose} = \frac{\text{BSA } (\text{m}^2)}{1.7} \times \text{average adult dose}$$

   The BSA can be calculated from Dubois formula:

$$\text{BSA } (\text{m}^2) = \text{BW}^{0.425}\,(\text{kg}) \times \text{height}^{0.425}\,(\text{cm}) \times 0.007184$$

3. *Responsiveness to drugs* varies among different species. This is very apparent in the animal world. Even among human beings, some racial differences have been observed. For example, blacks require higher and Mongols require lower concentrations of atropines and ephedrine to dilate their pupil.

4. *Sex.* Generally, females have a smaller body size and thus require doses on the lower side of the normal range. Drugs given during pregnancy may affect the fetus. In addition, there are a number of antihypertensives that interfere with sexual function in males but not in females. Gynecomastia is a side effect that is observed only in men. Moreover, androgens are unacceptable in women and estrogens in men.

5. *Route of drug administration*. Obviously, this affects the intensity and speed of a drug response.
6. *Individual response.* An individual's response to a drug is also controlled to some extent genetically. Some specific genetic defects have lead to discontinuous variations in drug responses:

   G-6-PD deficiency → hemolysis with primaquine

   CYP2D6 abnormality → poor metoprolol metabolizer status

   Acetylator polymorphism → isoniazid neuropathy

   Acute intermittent porphyria → precipitated by barbiturates owing to a genetic defect in repression of porphyrin synthesis
7. The *psychology* of a patient and his or her beliefs, attitudes, and expectations affect the efficacy of a drug to a great extent.
8. *Environmental factors* such as exposure to insecticides, carcinogens, and tobacco smoke affect a drug's metabolism to a great extent. The time of drug administration also influences drug action. For example, hypnotics should be taken in night and in quiet surroundings.
9. *Pathologic states.* Although it is true that drugs modify disease processes, there are several disease that influence drug disposition and drug action. Persons who are suffering from liver diseases (especially cirrhosis) have altered drug dispositions in various ways. The drug action is also altered in liver disease. The bioavailability of drugs having high first-pass metabolism is increased owing to the loss of hepatocellular function. Hepatotoxic drugs should be strictly avoided in the liver disease. Any sort of renal disease markedly affects the pharmacokinetics of many drugs and also alters the effects of some drugs. Gastrointestinal (GI) disturbances also alter the absorption of orally administered drugs. It is really a complex matter, and drug absorption may be increased or decreased depending on the situation. In celiac disease, absorption of amoxicillin is decreased, but that of cephalexin and cotrimoxazole is increased. Achlorhydria decreases aspirin absorption by favoring its ionization. Congestive heart failure (CHF) and coronary artery disease (CAD) alter drug kinetics by decreasing drug absorption from GIT, altering the volume of distribution V_d and retarding drug elimination as a result of decreased perfusion and congestion of liver, reduced glomerular filtration rate, and increased tubular reabsorption. Thyroid patients show sensitivity to such drugs as digoxin, morphine and other CNS depressants.
10. If the rate of *drug administration* is faster than the rate of elimination, a drug will start to accumulate in the body. Such drug accumulation will give rise to *cumulative toxicity*.

11. *Actions of other drugs*. There are drugs that interact among themselves at the pharmacokinetic and pharmacodynamic levels to modify the action of each other. Some of the examples include the following:

    Phenothiazine and metoclopramide block the therapeutic action of levodopa.

    Indomethacin and aspirin blunt the diuretic action of furosemide.

    Imipramine and chlorpromazine abolish the antihypertensive action of guanethidine and clonidine.

12. *Tolerance.* In the presence of drug tolerance, higher doses of a drug are needed to produce a response. Tolerance may be natural or acquired. Tolerance is seen in the sedative action of the chlorpromazine and phenobarbitone, and it occurs in the analgesic and euphoric action of morphine. Cross-tolerance also may develop among pharmacologically related drugs. For example, alcoholics are relatively tolerant to barbiturates and general anesthetics. As to how tolerance to a drug develops, though, no definite answer is yet available, and only an incomplete mechanism has been presented. Owing to repeated use of a drug, the elimination kinetics may be greater, and hence the effective concentration of the drug at the site of action may be decreased. Moreover, overuse of the same drug can cause the cells of the target organ to become less responsive.

5.8.2 How Does a Drug Act?

Fundamentally, drug action can be divided into four categories:

1. Physical action
2. Chemical action
3. The action of enzymes
4. By means of receptors

Some of the physical properties of a drug, such as its mass, adsorptive properties, osmotic activity, radioactivity, and radiopacity influence the mode of drug action. Drugs such as mannitol exert their action through osmotic activity. The adsorptive property of charcoal and kaolin is used in their action. The mass of a drug such as dimethicone is exploited for its action. Barium sulfate and urografin are used as contrast media because of their radiopacity.

A drug reacts extracellularly based on some chemical reactions. Antacids neutralize gastric HCl. Chelating agents such as calcium disodium edentate and penicillamine sequester toxic metals. $KMnO_4$ and I_2 are germicidal because of their oxidative action.

Most biologic reactions are carried out under the catalytic influence of enzymes. Thus enzymes are a very important target of drug action. Drugs can increase or decrease the rate of enzymatically mediated reactions. Enzyme stimulation by the drugs is somewhat unusual; enzyme stimulation is applicable in the case of endogenous mediators and modulators. For example, epinephrine stimulates adenylyl cyclase. An increase in enzyme activity occurs by enzyme induction, that is, synthesis of more enzyme protein. However, this is not stimulation because the value of K_m does not change.

However, inhibition of enzymes is a common mode of drug action. Enzyme inhibition may be nonspecific or specific. Nonspecifically, many drugs are capable of denaturing proteins and altering and inhibiting the tertiary structure of any enzyme in which it comes into contact. This is a nonspecific action.[28,37] Specific enzyme inhibition by drugs can occur in a competitive or a noncompetitive way.

Competitive Inhibition

Here, drugs compete with the normal substrate or coenzyme, attaining a new equilibrium in the presence of the drug. V_{max} remains unchanged, but K_m is increased. In this case, a higher concentration of the substrate is required to achieve half-maximal reaction velocity. In addition, if the substrate concentration is sufficiently increased, it can displace the drug, and the same maximal reaction velocity can be attained. Some examples include the following:

- Sulfonamides compete with para-amino-benzoic acid (PABA) for bacterial folate synthetase.
- Carbidopa and methyldopa compete with levodopa for dopa decarboxylase.
- Warfarin competes with vitamin K.
- Allopurinol competes with hypoxanthine for xanthine oxidase.

The preceding discussion addresses competitive equilibrium-specific inhibition. In the nonequilibrium type of enzyme inhibition, the drug will react with the same catalytic site of the enzyme but either form strong covalent bonds or have such high affinity for the enzyme that the normal substrate is not able to displace the inhibitor. For example, organophosphates react covalently with the esteretic site of the enzyme cholinesterase.

Noncompetitive Inhibition

Here, the inhibitor reacts with the adjacent site and not with the catalytic site. The enzyme loses its catalytic property. K_m remains unchanged, and V_{max} is reduced. Examples include

Disulfiram and aldehyde dehydrogenase
Digoxin and Na^+, K^+-ATPase
Aspirin and cyclooxygenase
Acetazolomide and carbonic anhydrase
Theophylline and phosphodiesterase

Drug Action Through Receptors

There are a large number of drug molecules that exert their action through *receptors*. Receptors are actually specific regulatory macromolecules that are situated either on the surface or inside the effector cell, to which drug molecules bind and interact to initiate the characteristic response. Agonists will activate a receptor to produce an effect, and antagonists will prevent the action of an agonist on the receptor.

When discussing the action of receptors, one comes across two terms, *affinity* and *intrinsic activity*. Affinity is the ability of a drug to combine with the receptor. And intrinsic activity is the ability of the drug to activate the receptor consequent to receptor occupation.

Initially, the receptor combines with the drug to bring about a conformational change in the latter or prevent a conformational change through exclusion of the agonist. This is said to be *drug action*. As a result of drug action, an ultimate change in biologic function occurs through a series of intermediate steps. This is referred to as the *drug effect*.

Drugs exert an "all or none" action on the receptor. This means that either the receptor is fully activated or it is not activated at all. There is no question of partial activation. A drug and its receptor have complementary structural features and stand in rigid "lock and key" relationship. It is said that the intensity of response is proportional to the fraction of receptors occupied by the drug, and the maximal response occurs when all the receptors are occupied.

Many drugs exhibit structural specificity of action. A specific chemical configuration is necessary for a specific action. A three-carbon internitrogen separation in the side chain of phenothiazines results in antipsychotic compounds, whereas a two-carbon separation produces anticholinergic-antihistaminic compounds.

In addition, some drugs are stereospecific in action. For example, L-norepinephrine is 10 times more potent than D-norepinephrine, and D-Propranolol is 100 times less potent in blocking β receptors than the L-isomer.

Receptors can be divided into four groups:

1. G-protein-coupled receptors
2. Receptors with an intrinsic ion channel
3. Enzymatic receptors
4. Receptors regulating gene expression

The basic functions performed by the receptors are as follows:

1. Propagation of regulatory signals from outside to within the effector cell
2. Amplification of the signals
3. Integration of various extracellular and intracellular regulatory signals
4. Adaption to short- and long-term changes in the regulatory melieu
5. Maintenance of homeostasis

References

1. Gibaldi M. *Biopharmaceutics and Clinical Pharmacokinetics*, 4th ed. Philadelphia: Lea and Febiger, 1991.
2. Gibaldi M, Perrier, D. *Pharmacokinetics*, 2nd ed. New York: Marcel Dekker, 1982.
3. Niazi S. *Textbook of Biopharmaceutics and Clinical Pharmacokinetics.* New York: Appleton Century Crofts, 1979.
4. Notari RE. *Biopharmaceutics and Clinical Pharmacokinetics: An Introduction*, 4th ed. New York: Marcel Dekker, 1987.
5. Hoeneer BA, Benet LZ. Factors influencing drug absorption and drug availability. In Banker GS, Rhodes CT, eds., *Modern Pharmaceutics*, 2nd ed. New York: Marcel Dekker, 1990.
6. Loper, AE, Gardner CR. Gastrointestinal absorption of drugs. In Swarbrick J, Boylan JC, eds., *Encyclopedia of Pharmaceutical Technology*, Vol. 6, New York: Marcel Dekker, 1992.
7. Mayersohn M. Principles of drug absorption. In Banker GS, Rhodes CT, eds., *Modern Pharmaceutics*, 2nd ed. New York: Marcel Dekker, 1990.
8. Cohn VH. Transmembrane movement of drug molecules. In Mandel HG, Way EL, eds., *Fundamentals of Drug Metabolism and Drug Disposition*. Baltimore: Waverly Press, 1971, p. 3.
9. Welling PG. Absorption of drugs. In Swarbrick, J, and Boylan JC, eds., *Encyclopedia of Pharmaceutical Technology*, Vol. 1. New York: Marcel Dekker, 1988.
10. Reidenberg MM, Erill S. *Drug-Protein Binding*, New York: Praeger, 1986.
11. Ritschel WA. *Handbook of Clinical Pharmacokinetics*, 3rd ed. Chicago IL: Drug Intelligence Publications, 1986.
12. Rowland M, Tozer TN. *Clinical Pharmacokinetics: Concepts and Applications*, 2nd ed., Philadelphia: Lea and Febiger, 1989.
13. Benford D, et al. *Drug Metabolism: From Molecules to Man.* New York: Taylor and Francis, 1987.
14. Caldwell J, Jakoby WB. *Biological Basis of Detoxification*. New York: Academic Press, 1983.
15. Gibson GG, Skett Pl. *Introduction to Drug Metabolism.* London: Chapman & Hall, 1986.
16. Caldwell J, Marsh, MV. Metabolism of drugs by the gastrointestinal tract. In George CF, Shand DG, eds., *Clinical Pharmacology and Therapeutics 1: Presystemic Drug Elimination*. London: Butterworth, 1982, p. 29.
17. La Du BN, Mandel HG, Way EL. *Fundamentals of Drug metabolism and Drug Disposition*. Baltimore: Williams & Wilkins, 1972, p. 149.
18. Williams RT. *Detoxification Mechanisms*, 2nd ed. London: Chapman & Hall, 1959.
19. Chasseand LF, Hawkins DR. Biotransformation of drugs. In Swarbrick J, Boylan JC, eds., *Encyclopedia of Pharmaceutical Technology*, Vol. 2. New York: Marcel Dekker, 1990.

20. La Du BN, Mandel HG, Way EL. *Fundamentals of Drug Metabolism and Drug Disposition*. Baltimore: Williams & Wilkins, 1971.
21. Low LK, Castagnolim N Jr. Metabolic changes of drugs and related organic compounds. In Delgado JN, Remers WA, eds., *Wilson and Grisvold's Textbook of Organic, Medicinal and Pharmaceutical Chemistry*, 9th ed. Philadelphia: Lippincott, 1991.
22. Schmucker DL. Aging and drug disposition. *Pharmacol Rev*. 37:133, 1995.
23. Mitchell JR, Horming MG, eds. *Drug Metabolism and Drug Toxicity*, New York: Raven Press, 1984.
24. Mulder GJ, ed. *Conjugation Reactions in Drug Metabolism*. New York: Taylor and Francis, 1990.
25. Taylor JB. *Comprehensive Medicinal Chemistry*, Vol. 5: *Biopharmaceutics*, London: Pergamon Press, 1990.
26. Test B, Jenner P. *Drug Metabolism: Chemical and Biochemical Aspects*, New York: Marcel Dekker, 1979.
27. Wolff ME. *Burger's Medicinal Chemistry*, Part I: *The Basis of Medicinal Chemistry*, 4th ed. New York: Wiley, 1980.
28. Brahmankar DM, Jaiswal SB. *Biopharmaceutics and Pharmacokinetics*: *A Treatise* New Delhi: Vallabh Prakashan, 1995.
29. Sharjel L, Yu ABC. *Applied Biopharmaceutics and Pharmacokinetics*, 2nd ed. New York: Appleton Century Crofts, 1985.
30. Swarbrick J. *Current Concepts in Pharmaceutical Sciences: Biopharmaceutics*. Philadelphia: Lea and Febiger, 1970.
31. Wagner JG. *Biopharmaceutics and Relevant Pharmacokinetics*, Washington, DC: Drug Intelligence Publication, 1971.
32. Makrikanis A. *New Methods in Drug Research*. Barcelona: Prous, 1985.
33. Gregoriadis G, Senior J. *Targeting of Drugs with Synthetic Systems*. New York: Plenum Press, 1986.
34. Chein YW. *Novel Drug Delivery Systems*, 2nd ed. New York: Marcel Dekker, 1990.
35. Flerence AT, Attwood D. *Physicochemical Principles of Pharmacy*. London: Macmillan, 1981.
36. Welling PG, Tse FLS, Dighe SV. *Pharmaceutical Bioequivalence*, New York: Marcel Dekker, 1991.
37. Tripathi KD. *A textbook of Medical Pharmacology*, Jaypee Brothers, Medical Publishers, New Delhi.

CHAPTER 6

Biomarkers, Biosensors, and Robotics in Medicine

6.1 Introduction

A biomarker can be defined as an indicator of a biologic state. Just like an indicator, it is used to measure and evaluate normal biologic processes, pathogenic processes, or the pharmacologic response to a therapeutic intervention.[1]

Anatomic, physiologic, biochemical, and molecular parameters are associated with the presence and severity of specific disease states. Biomarkers are detectable and measurable by a variety of methods, including physical examination, laboratory assays, and medical imaging. A biologic process or biochemical indicator that precedes the development of disease and usually indicates the progression of disease may be used to measure the effects of treatment. Biomarkers may be any parameter of a patient that can be measured, for example, mRNA expression profiles, proteins, proteomic patterns, lipids, imaging methods, or electrical signals. The best biomarkers are the objects of accurate, relatively noninvasive, and easy-to-perform tests that can be done at the bedside or in the outpatient setting. These tests that involve a blood or spot urine specimen can be measured serially and have a fast turnaround. In the past, most efforts had focused on discovering tissue and urinary biomarkers. However, there has been a recent shift to finding serum biomarkers, with new methods and technologies making this approach more practical.

Biomarkers have uses and applications in various fields and are named differently in different fields. A biomarker is a molecule that

indicates the existence, past and present, of living organisms.[1,2] Hence biomarkers are useful for

- Diagnosis
- Monitoring disease progression
- Predicting disease recurrence
- Therapeutic treatment efficacy

6.2 Developmental History of Biomarkers

Biomarkers that were developed in the 1980s include CA-19-9 for colorectal cancer and CA-125 for the ovarian cancer. Prognostic and predictive biomarkers continue to be expected to play a significant role in cancer management. Prostate-specific antigen (PSA) is one of the best known cancer biomarkers to detect the disease in early stages. PSA determination, along with digital rectal examination (DRE), is the only screening biomarker approved by the Food and Drug Administration (FDA). However, recently, a few diagnostic bladder cancer biomarkers such as NMP22 and bladder tumor antigen (BTA) received FDA recommendations. Calcitonin, a thyroid cancer marker, also can be a potential screening biomarker.

There has been a remarkable advancement in the field of genomic and proteomic technologies, including gene-array technology, improved two-dimensional gel electrophoresis, and advanced mass spectroscopic (MS) techniques, coupled with bioinformatics tools. All these advancements show great promise of meeting the demand for discovering both sensitive and specific biomarkers. The notable techniques used in biomarker assays are enzyme-linked immunosorbent (ELISA), protein chip technology, and immunohistochemistry.[2–4]

The Early Detection Research Network (EDRN), established by National Cancer Institute of United States, led to improved coordination between biomarker research laboratories. The developmental phases of a biomarker have been outlined by the EDRN as follows:

- *Phase 1* is preclinical exploration in order to identify the promising directions of a biomarker development.
- *Phase 2* performs clinical assays to detect established disease and validation.
- A biomarker detects disease early before it becomes clinical, and a seven-positive rule is defined. This is the retrospective longitude done in *phase 3*.
- *Phase 4* tests the extent and a characteristic of the disease detected by the test and also identifies the false referral rate. This is *prospective screening*.

- *Phase 5* is the cancer control program that shows the impact of screening on reducing the burden of disease on the population.[5–7]

6.3 Different Types of Biomarkers

Biomarkers that validate genetic and molecular biology methods can be broadly divided into three types: *Type 0*, or natural history markers; *type 1*, or drug-activity markers; and *Type 2*, or surrogate markers. *Type 0 biomarkers* may be defined as markers of the natural history of a disease, and they correlate longitudinally with known clinical indices, such as symptoms over the full range of disease states. Type 0 markers can be characterized in phase 0 clinical studies, in which a reliable assay is used in a well-defined patient population for a specified period of time. Ideally, a linear (positive or negative) relationship is established with the "gold standard" clinical assessor.

Type 1 biomarkers are actually the capture effects of an intervention in accordance with the mechanism of action of a drug, even though the mechanism may not be associated with clinical outcome. A priori validation of type 1 biomarkers is impossible for truly novel targets without an effective positive control treatment. By definition, the more innovative the target, the less validated will be the associated biomarkers.

Type 2 biomarkers are considered *surrogate endpoints* because a change in the marker predicts clinical benefit. Type 2 biomarkers (or surrogate endpoints) must be relevant both to the mechanism of action of the drug and the pathophysiology of the disease. Changes in the biomarker should reflect treatment benefit, and therefore, effective therapy is necessary for this validation.[3,5–7]

Apart from the three types of biomarkers already mentioned, a few more biomarkers are discussed below:

Accessible biomarkers. This is a biomarker that can be obtained in a minimally invasive manner, typically from blood, plasma, serum, saliva, or urine.

Antecedent biomarkers. Most adult-onset degenerative diseases of the nervous system are likely to be a composite of related characteristics, heritable and environmental. The correlated combinations of these features constitute the trait or disease. Therefore, these types of antecedent biomarkers may or may not be directly involved in the etiology. In some instances, the genetic variant is neither necessary nor sufficient to cause the disease. However, genetic variants can be powerful antecedents at any stage of the disease pathway. By definition, these antecedent biomarkers exist before the disease or the outcome occurs and are independent of other exposures. They improve precision in the measurement of other associations because they may be synergistic or antagonistic.

Biochemical biomarkers. These have long contributed to the assessment of risk and benefit in cancer, and routine clinical assays are available.

Biologic markers. These are measurable and quantifiable biologic parameters (e.g., specific enzyme concentration, specific hormone concentration, specific gene phenotype distribution in a population, and presence of biologic substances) that serve as indices for health- and physiology-related assessments, such as disease risk, psychiatric disorders, environmental exposure and its effects, disease diagnosis, metabolic processes, substance abuse, pregnancy, cell line development, and epidemiologic studies. Biologic markers can reflect a variety of disease characteristics, including the level of exposure to an environmental or genetic trigger, an element of the disease process itself, an intermediate stage between exposure and disease onset, or an independent factor associated with the disease state but not causative of pathogenesis. Depending on the specific characteristic, biomarkers can be used to identify the risk of developing an illness (antecedent biomarkers), aid in identifying disease (diagnostic biomarkers), or predict future disease course, including response to therapy (prognostic biomarkers).

Cellular biomarkers. Fundamental to many tissue- engineered devices are problems of inflammation associated with how biologic cells respond to a given device when inserted into the body. To ensure that tissue-engineered materials are free of molecular changes that can occur during the in vitro development phase, cellular biomarkers are being identified that can be used during the manufacturing process.

Disease biomarkers. The goal of this initiative is to validate biomarkers for well-defined human diseases of the liver, kidney, urologic tract, digestive and hematologic systems, and endocrine and metabolic disorders such as diabetes and its complications and obesity, for which there are no or very few biomarkers or for which standard biomarkers are currently prohibitively invasive or expensive. New biomarkers will stimulate bench-to-bedside translation by providing measures of the biologic effects of potential new treatments. The ideal biomarker can be measured in a minimally invasive way, can be measured repeatedly over time, identifies early stages of disease, is indicative of disease prognosis, and correlates well with progression and response to therapy. Especially of interest would be studies designed to test the validity of candidate biomarkers or new technologies to monitor candidate biomarkers in patient tissue samples or small groups of well-characterized patients

Efficacy biomarkers. In oncology, a special class of extensively evaluated biomarkers of efficacy (surrogate endpoints) that generally correlate with desired clinical outcomes can be used as a basis for

corporate decisions as well as for gaining accelerated provisional regulatory approval of a drug.

Exploratory biomarkers. These are biomarkers based on general exploratory or research information, such as broad gene expression screening or collection of sera or tissue samples, that have not reached the status of a probable valid biomarker.

Genetic biomarkers. These are single genes (DNA) for which a mutation, deletion, single-nucleotide polymorphism (SNP), or some other feature provides predictive value.

Genomic biomarkers. These are a measurable DNA or RNA characteristics that are indicators of normal biologic processes, pathogenic processes, and/or response to therapeutic or other inventions. They could, for example, reflect the expression of a gene or the function of a gene. The regulation of a gene does not include the measurement and characterization of proteins or low-molecular-weight metabolites. DNA characteristics include, but are not limited, to SNPs, variability of short sequence repeats, DNA modification (e.g., methylation), insertions, deletions, copy number variation, and cytogenetic rearrangements (e.g., translations, duplications, deletions, or inversions). RNA characteristics include, but are not limited to, RNA sequence, RNA expression levels, RNA processing (e.g., splicing and editing), and micro-RNA levels. The use of genomic biomarkers in drug and diagnostic development and regulatory decision making is showing a lot of potential but still requires more qualification and validation. Two key application areas were identified. First, at the drug discovery stage, genomic biomarkers applied to compound profiling offer efficacy and toxicity data, allowing better decision making at earlier stages and reducing late-stage attrition. Second, in the clinical setting, genomic biomarkers hold promise in disease diagnosis/prognosis; treatment, patient, or dose selection; and clinical safety and efficacy assessment. This s gene expression pattern (RNAs) that is able to discriminate or predict.[8–10]

Imaging biomarkers. The term *biomarker* is often associated with the detection or measurement (in vitro) of expressed genes, proteins, or metabolites in biologic fluids. To drug developers, however, *biomarker* can refer equally well to morphologic, functional, or molecular measurements made in vivo using medical imaging equipment such as computed tomography (CT), magnetic resonance imaging (MRI), positron-emission tomography (PET), ultrasound and optical scanners, and molecular imaging in drug research and development (R&D) and medical practice.

Indirect biomarkers. Biomarkers are very useful tools when the metabolic fate of the compound or the etiology of a resulting disease is completely understood. They may contribute to confusion if it is not possible to distinguish between markers of exposure and markers of disease. Such is the case for biomarkers used in the

assessment of diisocyanate exposure. Biomarkers for diisocyanate exposure result from both direct and indirect effects. Molecules such as hemoglobin, albumin, tubulin, glutathione, and laminin have been implicated as having been directly modified as a result of exposure to toluene diisocyanate (TDI). In addition, indirect biomarkers have included profiles of molecules such as antibodies, cytokines, cell accumulation or proliferation, and markers of oxidative stress.

Kidney biomarkers. In May 2008, the FDA and European Medicine Evolution Agency (EMEA) confirmed their joint review and acceptance of seven new laboratory tests on urine that signal kidney injury. The protein signals, known as *biomarkers*, were confirmed in data from rat studies submitted to the FDA and EMEA by the Predictive Safety Testing Consortium (PSTC). The FDA and EMEA jointly came to the conclusion that the kidney biomarkers are acceptable in the context of nonclinical drug development for the detection of acute drug-induced kidney toxicity. The kidney biomarkers provide additional and complementary information to the currently available standards. The use of kidney biomarkers in clinical trials is to be considered on a case-by-case basis in order to gather further data to qualify their usefulness in monitoring drug-induced kidney toxicity in humans.

Molecular biomarkers. These are early signs of change in an organism's physiologic state, such as adaptation, stress, or injury, owing to environmental factors or disease. Changes in molecules such as metallothionein are sensitive and specific, making them useful sentinels of an organism's exposure to a specific environmental agent. Other molecular changes indicate progression of a disease process. The development of molecular biomarkers provides a link between ecologists or epidemiologists, who study health effects in populations, and molecular biologists, who study the underlying mechanisms of these health effects.[11–13]

Physiologic biomarkers. Various physiologic responses have been measured and used as biomarkers. These include studies of such basic physiologic functions as respiration, changes in growth rate, feeding, excretion, and so on. Physiologic responses are used to provide integrated measures of an organism's well-being based on a range of different functional attributes. An example of one such measure is growth. Growth is an important fitness component of individual organisms and may have an overall impact on the success of natural populations. It is worth noting that although changes in a single fitness component may not always have a direct influence on the overall fitness of an individual, growth tends to integrate and reflect most sublethal effects.

Prognostic biomarkers. These provide information regarding outcome irrespective of therapy. Candidate prognostic biomarkers for

breast cancer include elevated proliferation indices such as Ki-67 and proliferating cell nuclear antigen; estrogen receptor (ER) and progesterone receptor (PR) overexpression; markers of oncogene overexpression such as c-erbB-2, transforming growth factor α (TGF α), and EGFr; indicators of apoptotic imbalance, including overexpression of *bcl*-2 and an increased *bax*/*bcl*-2 ratio; markers of disordered cell signaling such as p53 nuclear protein accumulation; alteration of differentiation signals, such as overexpression of c-*myc* and related proteins; loss of differentiation markers, such as TGF-βII receptor and retinoic acid receptor; and alteration of angiogenesis proteins such as vascular endothelial growth factor (VEGF) overexpression.[13]

RNA biomarkers. A huge number of human transcripts has been found that do not code for proteins; these are named *non-protein-coding RNAs*. In most cases, small (miRNAs, snoRNAs) and long RNAs (antisense RNA, dsRNA, and long RNA species) have many roles, functioning as regulators of other mRNAs, at transcriptional and posttranscriptional levels, and controlling protein ubiquitination and degradation. Various species of non-protein-coding RNAs have been found differentially expressed in different types of cancer.[14,15]

Stratification biomarkers. These can be used to enrich a clinical trial population with one or more goals in mind, including increasing the responder population; reducing the size, time, and cost; and improving overall results of the clinical trial. Patient stratification involves the use of biomarkers to create subsets within a patient population that provide more detailed information about how the patient will respond to a given drug. Stratification can be a critical component to transforming a clinical trial from a negative or neutral outcome to one with a positive outcome by identifying the subset of the population most likely to respond to a novel therapy.

6.4 Use of Biomarkers in Cancer

The most important role of cancer biomarkers is to use them in widespread screening so that asymptomatic individuals can be detected with disease at a very early stage. An ideal tumor marker would be a protein or protein fragment that can be easily detected in the blood or urine of a patient but not detected in a healthy person.

For the first time, in 1965 Dr. Joseph Gold found a substance in the blood of patients with colon cancer that was normally found in fetal tissues and named it *carcinoembryonic antigen* (CEA). Additional biomarkers developed in 1980s were CA-19-9 for colorectal and pancreatic cancer, CA-15-3 for breast cancer, and CA-125 for ovarian cancer.[16–18,25]

However, the best known cancer biomarker used by physicians to detect early disease is the prostate-specific antigen (PSA). The upper limit of normal PSA level was considered to be 4 ng/mL. Thirty-three percent of tumors spread beyond the prostate in men with PSA values between 4 and 10 ng/mL. Alhough PSA screening provides a suspicion of prostate cancer, further clinical diagnosis relies on pathologic tissue examination. Any healthy individual with a PSA between 4.0 and 10 ng/mL is recommended for biopsy, although the lower limit for suspicion of prostate cancer has dropped recently to 2.5 ng/mL.

However PSA elevation is also associated with benign prostatic hyperplasia (BPH). Thus an elevated PSA level does not always indicate the presence of cancer.

For a biomarker to be called as a *perfect marker*, it is expected to have 100 percent sensitivity and 100 percent specificity. PSA, although a very sensitive marker, has low specificity. For now, the only tumor biomarker approved by the FDA for widespread screening purposes is PSA determination, along with digital rectal examination (DRE). Widespread PSA screening for the diagnosis of prostate cancer can be done in the patients with non-life-threatening small cancers in order to avoid complications of surgery or radiation therapy.

Change in PSA level with respect to time is the *PSA velocity*. A steep rise in PSA level increases the risk of malignant prostate cancer. Patients whose PSA level is increased more than 2 ng/mL prior to diagnosis of prostate cancer have the risk of this disease, even if they undergo radical prostatectomy. *PSA density* considers the relationship between PSA level and the size of the prostate.

Circulating PSA, found in the serum, may be either in free or bound form. The ratio of free to bound PSA decreases from benign to cancer. In benign conditions, free PSA is found, whereas the PSA in cancer is mostly in bound form.

Rapid advances in biomarker development have been made using gene arrays in addition to proteomic technologies, including two-dimensional (2D) electrophoresis (2DE) and mass spectrometry. The FDA has approved a small number of new urine-based biomarkers such as bladder tumor antigen (BTA) and nuclear matrix protein 22 as diagnostic markers for bladder cancer. Survivin, an inhibitor of apoptosis, has been identified as a urinary diagnostic biomarker for bladder cancer. Recently, calreticulin (CRT) was identified as a potential biomarker for bladder cancer.

Although PSA has been the major biomarker for the diagnosis of prostate cancer, in the last decade several potential predictive biomarkers have been identified, including thymosin β-15, antizyme, antizyme inhibitor, and collagen XIII, which help to distinguish metastatic prostate cancer.

Thymosin β-15 level is elevated in metastatic prostate cancer. Thymosin β-15 in combination with PSA predicts the recurrence of prostate cancer with more sensitivity and specificity than PSA alone. P53 tumor suppressor gene, bcl-2 proto oncogene, Ki-67 proliferation labeling index, osteopontin, osteocalcin, and metalloproteinases (MMPs) are also prognostic biomarkers.

Previously, mammography was used widely to detect breast cancer. Several research works have been successful in identifying proteins that contribute to the metastatic signature of breast cancer. Two of the breast cancer susceptibility genes are *BRCA1* and *BRCA2*. Persons with mutations of these two genes have an elevated risk of developing the breast cancer. *BRCA1* and *BRCA2* genetic analysis can best serve as a prognostic indicator but not as a decision maker. There are several factors that help to monitor prognosis in breast cancer. According to some, the apoptotic index can be a better predictor of 5-year disease-free survival than tumor volume, mitotic index, or status of organ confinement.

Vascular densities, a marker of blood vessels in a given volume of tumor, are important factors that influence tumor growth and dissemination. The microvessel density (MVD) increases with declining pathologic state of the tumor. Osteopontin (OPN) is a marker in several cancers, including breast cancer. Her-2 oncogene expression in tissues or serum is the most commonly used predictive biomarker for the breast cancer. Since it is a true predictive marker, *Her*-2 not only provides an indication of disease outcome but also can lead to selection of an appropriate cancer treatment, such as the use of herceptin to treat patients with *Her*-2 positive tumors. The other well-known indicators for breast cancer are urokinase-type plasminogen activator (uPA), plasminogen activator inhibitors (PAI) 1 and 2, cathepsin B, and cathepsin L.[19]

Ovarian cancer is also a target of intense biomarker research. Scientists have shown that overexpression of cyclin D (CD 1) is related to ovarian cancer. Tumor-associated trypsin inhibitor is a potentially important predictor of disease stage in ovarian cancer. Surface-enhanced laser desorption ionization (SELDI) associated with mass spectrometry has been used to compare serum protein profiles at different stages of ovarian cancer.

Extensive research is being carried out to identify genetic markers or signature genetic profiles for the diagnosis and prognosis of colorectal cancer. Microsatellite instability resulting from mutations in DNA mismatch repair genes such as *MLH1, MSH2*, and *MSH6* is associated with the prognosis of colorectal cancer.[16,18] P^{53} alterations are associated with non–small cell lung cancer, β-2 macroglobulin determines prognosis in multiple myeloma and lymphoma, and caspase-3 is a prognostic factor in gastric carcinoma.

Several biomarkers have been developed for tumor recurrence. Patients who have emerged from cancer surgery or chemotherapy

have a 30 to 70 percent statistical chance of a recurrent cancer owing to either local growth or distant metastasis. A study group has shown that two biomarkers, enhancer of zeste homologue (EZH) 2 and E-cadherin (ECAD), when found together in prostate tumor tissue, predict a threefold risk of cancer recurrence after surgery. Calcitonin, a biomarker for thyroid medullary carcinoma, is a widely accepted biomarker for thyroid cancer recurrence. Thyroglobulin is used to determine cancer recurrence after the thyroid is removed.[20,21]

6.5 Biomarkers in Medical Science

Biomarkers are introduced into the body of a patient to check organ function or other health aspects, for example, use of the radioisotope rubidium chloride to check perfusion of the heart muscle. Biomarkers indicate a particular disease state. They diagnose disease risk for a particular individual that may further necessitate altered drug treatment or administration regime. They indicates a change in expression or state of a protein that correlates with the risk or progression of a disease or with the susceptibility of a disease to a given treatment.

A biomarker serves as a *surrogate endpoint* for evaluating clinical benefit. This is true when a treatment alters a biomarker that has a direct connection with improved health. As a surrogate for a natural endpoint such as survival or irreversible morbidity, it is useful in evaluating potential drug therapies.

In psychiatric treatment for diseases such as schizophrenia, a special kind of biomarker is used for finding genetic causes for the disease and is called an *endophenotype*.

Previously, biomarkers were physiologic indicators of blood pressure or heart rate, but now the term has become a synonym for *molecular biomarker*. Molecular biomarkers indicate elevated PSA for prostate cancer and use enzyme assays as liver determinants of function. Molecular indicators cover the range of environmental exposure in epidemiologic studies, such as human papillomavirus or certain markers of tobacco exposure such as 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. Thus biomarkers are used in epidemiology and toxicology.

6.6 Recent Advances in Biomarkers

With the recent advances in biomarkers, they now involve imaging technology. *Imaging biomarkers* have the potential advantages of being

- Noninvasive
- Comfortable for patients

- Intuitive and multidimensional
- Qualitative and quantitative
- Very useful for doctors as a diagnostic aid

Use of imaging technology has revolutionized medical science, especially cardiac imaging, which is an active area of biomarker research. Today, physicians are trying to find noninvasive procedures that avoid invasive procedures such as catheterization. Cardiac computed tomography has great potential in this area, but researches are still attempting to overcome problems related to *calcium blooming*, a phenomenon in which calcium deposits interfere with image resolution. Other intravascular imaging techniques involve MRI. Optical coherence tomography (OCT) also is being investigated.

Another new imaging biomarker is positron-emission tomography (PET), which involves the use of radiolabeled glucose, and physicians have found it to be a promising technique because by tracking radiolabeled glucose, it directly measures a step known to be crucial to inflammation and tumor growth.[22–24]

6.7 Disadvantages of Biomarkers

1. If a biomarker is used to measure the success of a therapeutic intervention, the biomarker should reflect a direct effect of that intervention.
2. It is not advisable to always use biomarkers as surrogate endpoints to assess clinical outcomes.
3. It may be difficult to validate biomarkers, and they require different levels of validation depending on their intended use.

6.8 Biosensors

A *sensor* is a device that exhibits a characteristic of an electrical nature (e.g., charge, voltage, or current) when it is subjected to a phenomenon that is not electric. Sensors measure a physical quantity and convert it into a signal that can be read by an observer or by an instrument.[26]

Biosensors are sensors used to detect and measure the mass as well as structural aspects of biomolecular species. A biosensor also can be defined as any biologic component that will be used to sense the chemical or physical environment.

Biosensors are analogue devices that are based on the direct coupling of an immobilized biologically active compound with a

signal transducer and an electronic amplifier. Thus a biosensor is a self-contained integrated device that is capable of providing specific quantitative analytical information using a biologic recognition element, which is retained in direct spatial contact with a transduction element.[27]

6.9 Developmental History of Biosensors

Biosensors are the analytical tools for the analysis of biomaterial samples to gain an understanding of their biocomposition, structure, and function by converting a biologic response into an electrical signal. As regards the developmental status of a biosensor, it passes through three generations. In a first-generation biosensor, the normal product of the reaction diffuses to the transducer and causes the electrical response. In a second-generation biosensor, there is a specific mediator between the reaction and the transducer in order to generate an improved response. Among the third-generation biosensors, the reaction itself causes the response, and no product or mediator diffusion is directly involved.

In 1956, the oxygen electrode was invented by Leland C. Clark. The first potentiometric biosensor to detect urea was developed in 1969. In 1970, the ion-selective field effect transistor (ISFET) was invented. In 1975, the first commercial biosensor for the estimation of glucose—Yellow Springs Instruments glucose biosensor—was invented. In 1976, the first bedside artificial pancreas were developed. In 1980, the first fiberoptic pH sensor was developed for the in vivo blood-gas estimation, as well as a fiberoptic-based biosensor for glucose estimation. The first immunosensor, Surface Plasmon Resonance, was developed in 1983. In 1984, the first amperometric biosensor was developed in which ferrocene was used with glucose oxidase for the detection of glucose. In 1990, Pharmacias, BIA Core surface plasmon resonance (SPR)-based biosensor system was launched. In 1992, I-STAT launched a handheld blood analyzer; Glucocard was launched in 1996; and the Life Scan Fast Blood Glucose biosensor was launched in 1998.[28]

Often sensors are signal or stimuli receptors. The sciences of chemistry, physics, and biology, along with technological applications, have joined hands to assess stimuli and impulses. A genetically engineered organism that can emit light when exposed to some variable is often termed a biosensor.

The search for sensors has a long history. The fundamental senses are visual, auditory, smell, taste, and touch for which the receptors are eyes, ears, nose, tongue, and skin, respectively. Although sensed by the same human tongue, there are five distinct tastes, namely, bitter, salty, sour, sweet, and umami. The transducers for each are different, but signal outputs are relayed through nerves and interpreted (analyzed/monitored) at specific brain centers.

Luxmeters, audiometers, navigation radars, humidometers, pH meters, and thermometers are a few classic examples of sensors. Now, laborious laboratory exercise is not needed to test blood glucose or blood urea; a subcutaneous probe alone will instantaneously tell us the clinical data; even a droplet of blood assessed by color change of paper strip will indicate the approximate clinical concentration of glucose.[26,28] Very specific and exclusive sensors are available nowadays.

6.10 General Instrumentation and Operation of a Biosensor

In a biosensor, the phenomenon is recognized by a biologic system called a *bioreceptor* that is in direct contact with the sample and forms the sensitive component of the biosensor. The bioreceptor has a particularly selective site that identifies the analyte.

A biosensor consists of a biologic component for sensing the presence and concentration of substance and a transducer in which the sample is allowed to pass through a membrane so that selection may be exercised and the interfering molecules are retained outside the membrane. The sample then interacts with the biologic sensor and forms a product, which may be an electric current/charge, heat, gas, or a suitable chemical. The product then passes through another membrane and reaches the transducer, which converts the biochemical signal into an electrical signal that can be further amplified and can be read on a digital panel or recorded on a recorders.

The components of a biosensor (Fig. 6.1) are as follows:

- Converter converts the substrate to product
- Transducer
- Amplifier
- Processor
- Display unit

No single inventory or list will suffice to indicate the materials used for sensing or biosensing. Semiconductors are usually the midway

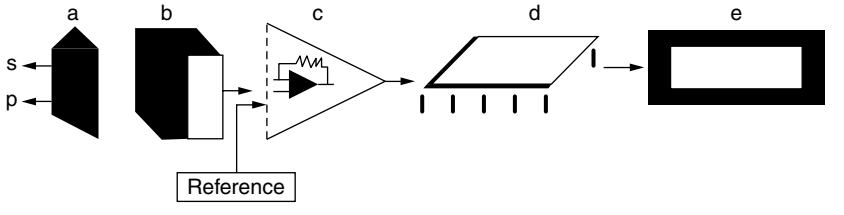


Figure 6.1 Biosensor components.

sensors. Selenium (Se) was used to sense and measure incident light, silicon was used in photovoltaic assembly, and SnO_2 also has some applications.

Nafion-I, a hydrophobic ionomer (polydifluorocarbon with SO_3H terminals), was used for sensing room temperature, dissolved oxygen in water, and so on. For humidity scaling, many polymers, namely, ethylene oxide, *O*-toluidine, *m*-phenylendiamine, vinyl alcohol, and others are already in use.

Solid-state sensors are made from various insulators, metals, and catalytic materials and also from organic membranes. For the detection of gases, mixtures of gases, or odors, a number of polymers are found to be useful. Polysiloxane, polypyrrole, and polyaniline with different polymers are used successfully for the detection of odor. PVA and PVC are very common polymers used as gas sensors. Using suitable doping material, polypyrrole, nanocomposite is a very sensitive sensor for the detection of CO_2, N_2, and CH_4.

Analyte	Sensor
H_2O, CH_4, CO_x, NO_x, NH_3, CFC	Organic semiconducting polymer of phenyl, acetylene, pyrrole, or amide
Ionic: Cationic or anionic oxygen, dissolved oxygen	Membranes Solid electrolytes ($Y_2O_3 + ZrO_2$)
N_2O, F, O_2, CD_x, SO_x, NO_x	Different sensors (LaF_3)
Different gases	SnO_2, ZnO, TiO_2, CoO, NiO, WO_3

If a chemical, biochemical, or physical change takes place either directly or indirectly in the target substrate, the change in color or polarization may be analyzed. Glucose is converted to gluconate by glucose oxidase enzyme in the presence of a leukodye that is converted to a colored dye that can be measured. Alternatively, the same reaction may be oxidized by H_2O_2 and the residual peroxide estimated by catalase and oxygen electrode directly.[29] L-Amino acid oxidase was covalently bound to an (electrodeposited) polytyramine electrode and could detect L-amino acids. A highly sensitive and rapid analysis of urea was possible with a composite polymerized membrane of polypyrrole and a polyion complex incorporating urease.

A multienzyme system of creatininase, creatinase, and sarcosine oxidase was immobilized on top of cross-linked proteins with gluteraldehyde and was used in sensors for the detection of a broad range of amino compounds, forming a fluorescent signal by

isoindole complex, formed by a polymer of allyl mercaptans and *O*-phthalic dialdehyde.[30]

$$O\text{—}[CF_2\text{—}CF\text{—}]_2\text{—}O\text{—}CF_2\text{—}CF_2\text{—}SO_3H$$

$$\text{—}\{CF_2\text{—}GF\text{—}(CF_2\text{—}CF_2)_X\text{—}\}_Y^-$$

Nafion - I

For the detection of total cholesterol, a two-enzyme electrode was constructed by incorporating cholesterol esterase and cholesterol oxidase in polypyrrole film with good storage stability. For the determination of organic peroxides in cosmetics and body fluids, poly-*N*-methyl pyrrole–horseradish peroxidase, a amperometric biosensor, was developed. The biosensor for hydrogen peroxide consists of a composite carbon electrode with ferrocene carboxylic acid sol—gel on which the horseradish peroxidase—polypyrrole membrane was electrodeposited. For the detection of a number of ions, such as Ag, Na, K, Sr, Pb, and Hg ions, selective sensors are available. It was reported to be a fast and reliable process controlled by the application of surface acoustic wave sensors.

Positional biosensors. These are fluorescent protein biosensors that monitor the target function in living and fixed cells. They give a robust automated readout in conjunction with cellomics, high-content screening (HCS), and bioapplications. They are used in life science research and in the HCS approach to drug discovery, and they provide valuable data regarding reaction kinetics, protein interactions, and posttranslational modifications.[31]

Conventional biosensors. Here, biosensors are used in combination with receptors for lead optimization, the most challenging area in drug research, in a fast, cheap, and effective manner. Preparation of the receptor surface plays a crucial role here and is achieved by immobilization of receptors in artificially prepared membranes on the sensor surface. The receptor sites, when occupied by the drug, will cause a measurable change in a physical parameter that is proportional to the mass change on the sensor surface. The important requirements of a biosensor are sensitivity, selectivity, a stable and reproducible reading, and reusability. The bound molecules from the receptor surfaces can be removed by aggressive media that reduce the activity of the receptors.

Optical biosensors. These sensors determine affinity and kinetics of wide variety of molecular interactions without the need of a molecular tag. They are used in drug discovery, target identification, lead selection, the KADME mechanism, and manufacturing quality control. These sensors exploit surface plasmon resonance, waveguides, and resonant mirrors to analyze bimolecular interactions.

6.11 Biosensors in Drug Design

Biosensors help to calculate reaction rate constants that govern drug interactions. This technique is very useful in kinetic experiments and measures bimolecular interaction in real time. Biosensing technology is really unmatched to perform kinetic experiments with ease and efficiency.

The pharmaceutical companies are spending billions of dollars on R&D to discover a blockbuster drug. About US$400 to 800 million and 10 years are spent by each of the companies to bring a drug to market, for which only one of five lead drug compounds makes it to final clinical use. These companies endure a tremendous pressure to ensure an improved success rate, reduced cycle time, and lower R&D costs. The latest technological advances are being used to ensure faster drug target discovery and drug development. High-throughput screening and massive parallelism increase the speed at which combinations of experiments can be used to find a blockbuster drug. Some of the biosensors used in drug discovery are discussed briefly below.[32,33]

So far, we have been discussing where biosensors contribute to drug discovery/design. Understanding interaction kinetics, especially molecular interactions, is very crucial for drug design. Biosensors are ideally poised for use in the devices and components employed routinely in drug discovery and development. Biosensors allow/help the pharmaceutical companies to lower drug discovery and development costs, achieve lower per-unit costs and achieve a high degree of manufacturing control. Nowadays, biosensors are becoming increasingly popular among industrial and academic scientists in the field of drug design/discovery.

6.12 Biosensors in Biomedical Research

Diagnostics is an area of increasing focus for the biosensors. There is increasing activity in the area of implantable MEMS for drug delivery and for in vivo biosensors. Biosensors can detect and differentiate among different chemical constituents of complex systems to provide unambiguous identification and accurate quantitation.

A new generation of biosensors is using DNA chips or probes. DNA biochips are used to diagnose genetic susceptibility and diseases, and DNA biosensors are used in the areas of nucleic acid identification. The latest biochip, using an antibody probe, is designed to detect the p53 protein system. The recognition performed by DNA biochips is based on the molecular hybridization process.

Biosensors are now being applied in cancer diagnosis. A minimally invasive method using laser-induced fluorescence (LIF) for in vivo cancer diagnosis has been developed by scientists at the Oak

Ridge National Laboratory (ORNL) and the Thompson Cancer Survival Center (TCSC). Here, the autofluorescence of normal and malignant tissues was measured directly using a fiberoptic probe inserted through an endoscope. This procedure does not require biopsies. The LIF procedure offers a rapid and cost-effective technique for cancer diagnosis.

Antibody-based fluoroimmunosensors (FISs) have been developed to detect the carcinogen benzo[a]pyrene (BaP) and benzopyrene tetrol (BPT). In both, the in vivo and in vitro fluorescence assay polyclonal or monoclonal antibodies are immobilized at the terminus of a fiberoptic probe or contained in a macro-sensing cavity within the FIS. The FIS device uses the backscattering of light emitted at the remote sensor probe. This device has excellent sensitivity to perform the trace analysis of chemical and biologic samples in complex matrices.[34]

Nanotechnology, biology, photonics, and advanced materials have joined hands to detect and manipulate atoms and molecules at the cellular level for a variety of medical purpose. A new class of nanoscale sensors called *nanosensors* allows measurements in the smallest of the environments. Using antibody-based nanoprobes in situ, intracellular measurements of single cells can be performed. With the aid of nanosensors, it is possible to probe individual chemical species and molecular signaling processes in specific locations within a cell. For the said purpose, if the nanoscale biosensor is inserted into a mammalian somatic cell, neither will it cause any damage to the cell membrane, nor will it affect normal cell function.

Because of the unique physical and chemical properties of nanoparticles, they are now being conjugated to biomolecules and have a widespread application in biologic and medical sciences. The human skull is a challenging obstacle to the coherent propagation of the ultraviolet rays that make imaging difficult. Ultrasonic tissue characterization has proven to provide sensitive and selective indicators of the state of vital tissues in more sonically accessible organs such as the heart. The primary goal of this work is to develop a rapid screening technology for monitoring brain injury and disease.[28,34]

Plasmonics is a field of research that refers to the enhanced electromagnetic properties of metallic nanostructures. *Plasmons* are actually the quanta associated with longitudinal waves propagating in matter through the collective motion of large numbers of electrons. Incident light irradiating the surfaces excites conduction electrons in the metal, exciting the surface plasmons and leading to enormous electromagnetic enhancement of a spectral signature such as surface-enhanced Raman scattering (SERS) and surface-enhanced fluorescence (SEF). Such techniques allow ultrasensitive detection of chemical and biologic species (Fig. 6.2).

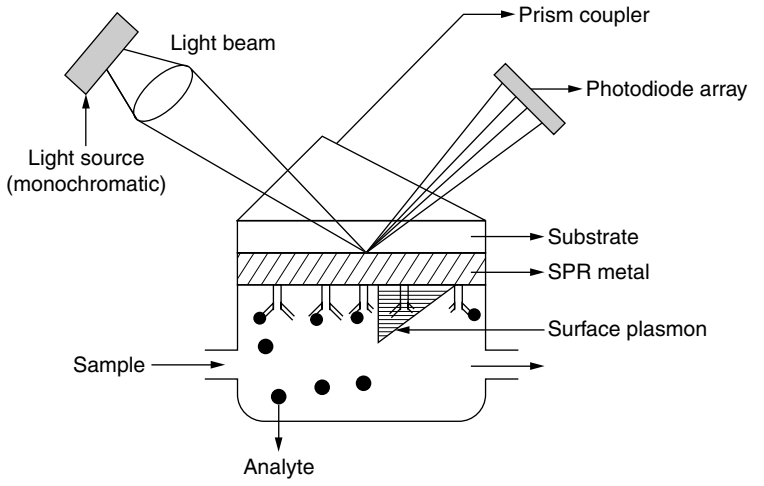


Figure 6.2 SPR (surface plasmon resonance) sensor.

In addition to SERS, confocal surface-enhanced Raman imaging (SERI) has been developed to monitor the intracellular distribution of molecular species associated with biologic abnormalities and localization of drugs and other cellular components within cells, thus offering a promising application for molecular signal monitoring for nanomedicine applications.

An emerging technique with excellent resolving power is the nearfield scanning optical microscopy (NSOM), a critical tool for visualization of cellular components labeled with fluorescent molecules. NSOM technology gives fundamental information about the cellular component's orientation. Along with confocal microscopy, NSOM technology is now used to study the localization of multidrug resistance (MDR) transport proteins and their effect on chemotherapeutic drugs. MDR is also used to indicate a variety of strategies that tumor cells develop owing to the cytotoxic effects of the anticancer drugs. MDR or brings three major changes in a cell:

1. Decreased accumulation of the cytotoxic drugs
2. Changes in cellular physiology
3. Changes in the activity or expression of some cellular proteins.

A class of unique optical nanobiosensors has been developed that are inserted into single living cells to monitor and measure biomolecules and biochemicals of biomedical interest without disrupting normal cellular processes (Fig. 6.3).

Optical nanobiosensors are integrated nanoscale devices consisting of a biologic recognition molecule coupled with an optical transducing element such as an optical nanofiber interfaced with a

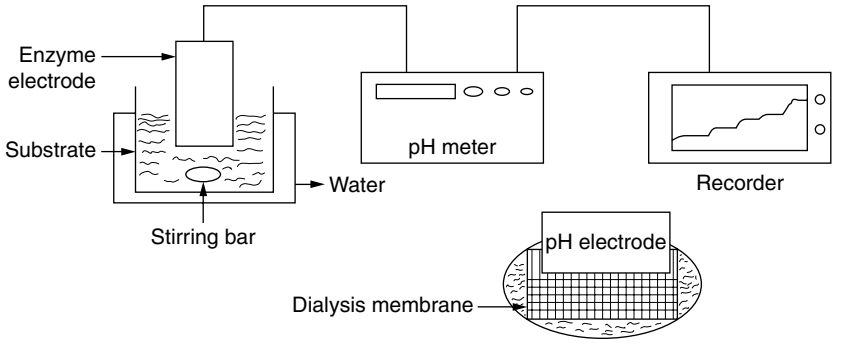


Figure 6.3 Optical biosensors. (*From Mulchandani P, Mulchandani A, Kanera I, Chen W, Biosens Bioelectron. 14:77, 1999, with permission.*)

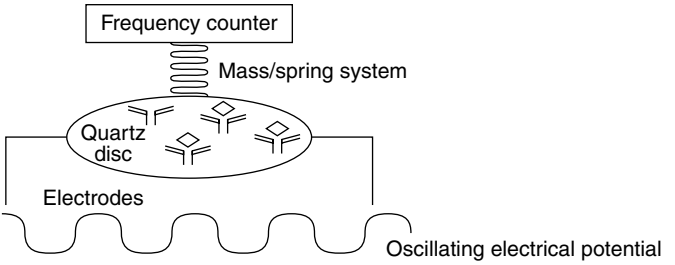


Figure 6.4 Piezoelectric biosensors.

photometric detection system. This nanoscale biotechnology-based device helps to monitor intact cell function and to study molecular functions such as apoptosis, DNA-protein interactions, and protein-protein interactions.

A new platform technology for biosensor and biochip development is magneto electronics (Fig. 6.4). Since the late 1990s magneto electronics is a new platform technology for the biosensor and biochip development. No doubt fluorescence-based microarrays are used successfully in the field of biomedical research (Fig. 6.5), but this technology, although in an infant stage, offers such advantages as

1. High-sensitivity detection to the level of a single molecular labeling system
2. A stable labeling system
3. Low magnetic background
4. Cheap device components

The biosensor research field is becoming vast day-by-day following the resurgence of interest in these devices in the 1990s; one of the contributing factors has been the continuous development of magreto-resistive materials, which have replaced many magnetic field sensors

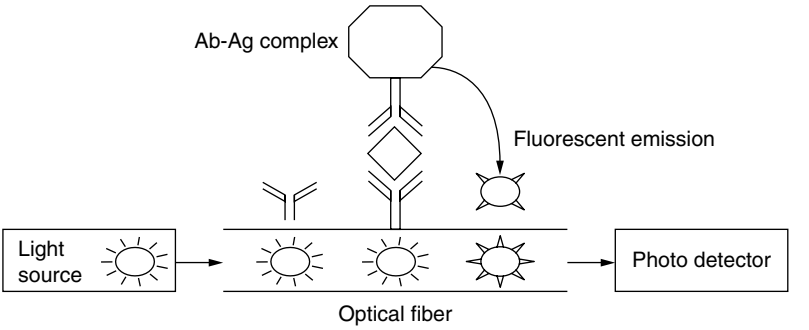


Figure 6.5 Fluorescence-labeled biosensors.

such as the super conducting quantum interference device (SQUID) and induction coils that offer, from their large size, low sensitivity and high power consumption.

Discovery of antiferromagnetic interlayer exchange coupling and the giant magnetoresistive effect (GMR) opened up several applications for high-sensitivity magnetic nanostructures, including magnetic recording media, read heads, and magnetic random access memory.

Microspheres, microbeads, and nanoparticles, also referred to as *magnetic labels* or *carriers*, have wide-ranging scientific and clinical applications in the field of biotechnology, drug delivery, molecular biology, and bioseparation and biomedical research. The important characteristics of these magnetic labels used in biosensors or biochipss are size and shape, chemical and magnetic composition, surface properties, stability, and ease of chemical functionalization for the immobilization of the biomolecules.

MR biosensing technique is versatile in that it potentially can be used to detect and analyze magnetically labeled nucleic acids, proteins, whole cells, or microorganisms.

The choice of DNA chip developed depends largely on the nature of the sensor signals obtained. Now, DNA hybridization and detection can be performed using an MR biochip in minutes, employing very low target DNA concentrations.[32,33] MR biochips have shown strong potential applications in high-sensitivity protein-based microassays. MR biochips with small-spin-valve sensors are now being applied to the development of diagnostic biochips for cystic fibrosis, a complex genetic disease.

Advancements are on the way to tailor DNA chip applications, more specifically toward the detection of biologic warfare agents such as *Bacillus anthracis, Yersinia pestis, Clostridium botulium, Vibrio cholerae,* and vaccinia virus. Today, MR biochips offer a rapid expanding research field with high sensitivity and high-quality quantitative molecular recognition detection data for a variety of biologic applications.

Thus it may be concluded that a decade ago, instrumental detection of smell and taste was almost a fiction. Now we are in a position to look forward to sensitive, specific, compact sensors. The situation demands the collaboration of biochemists, clinicians, biotechnologists, diagnostic experts, solid-state scientists, polymer experts, and instrument designers. Let us hope also to develop sensors for liee, hate, fetishes, dedication, and so on.

6.13 Robotics in Surgery

Robotics technologies have given rise to a variety of sophisticated industrial machines that assist industrial workers in carrying out their jobs with better quality by means of automation. Industrial robotics have developed different automated devices through computer vision programmed motion and movement by computer control of devices, control schemes, and algorithms to do a series of tasks involving interaction with positioning accuracies and forces. Previously, robotics was applied in autonomous vehicles, space exploration, land rovers, and so on. With the passage of time, tremendous advancement took place, miniaturization and microscience technologies were developed, and robotics became a good match for medical science. Several new inventions have been developed, and robotic technologies have revolutionized medical diagnostics and surgery. Extremely complex fully automated devices have been developed that will reliably execute a doctor's or surgeon's commands inside a patient's body.

Nowadays, doctors are using a robot probe to scan for seven parameters in the GI tract for any defects, avoiding surgical operations.[35] Surgery now uses robotic and image-processing systems to interactively assist the medical team, both in planning the surgical intervention and in execution. The objective of this new technique is to enhance the quality of surgical procedures by minimizing their side effects (smaller incisions, lesser trauma, more precision), thus increasing patient benefit while decreasing the surgical cost. These techniques are being introduced successfully in several areas of surgery: neurosurgery, orthopedics, microsurgery, cardiovascular and general surgery, and so on.

Three main steps can be pointed out in a general robotic surgical intervention: data acquisition and subsequent planning, intraoperative assistance, and postoperative patient control. In the *preoperative phase*, a patient-dependent model of the rigid (e.g., bones) and deformable (e.g., the heart) anatomic entities involved in the surgical act has to be built. For this, several medical imagery techniques (MRI, scanners, ultrasonics, etc.) are used, where the anatomic structures are detected, located, and modeled. At the same time, the mechanical model of the robotic system is fused in an overall geometric model. This will be used to describe and simulate the different potential problems that may occur during intervention.

The three major advances aided by surgical robots have been remote surgery, minimally invasive surgery, and unmanned surgery. Some major advances of robotic surgery include precision, miniaturization, smaller invasions, decreased blood loss, less pain, and quicker healing time. Further advantages are articulation beyond normal manipulation and three dimensional magnification resulting in improved ergonomics.[36]

6.14 Background and History of Surgical Robots

Since 1921 when Czech playwright Karel Capek introduced the notion and coined the term *robot* in his play, "Rossom's Universal Robots," robots have taken on increasing importance both in the imagination and in reality. *Robot*, taken from the Czech *robota*, meaning "forced labor," has evolved in meaning from dumb machines that perform menial, repetitive tasks to the highly intelligent anthropomorphic robots of popular culture. Although today's robots are still unintelligent machines, great strides have been made in expanding their utility.

Today, robots are used to perform highly specific, highly precise, and dangerous tasks in industry and research previously not possible with a human work force. Robots are used routinely to manufacture microprocessors used in computers, explore the deep sea, and work in hazardous environment, to name a few.

Robotics, however, has been slow to enter the field of medicine. The lack of crossover between industrial robotics and medicine, particularly surgery, is at an end. Surgical robots have entered the field in force. Robotic telesurgical machines have already been used to perform transcontinental cholecystectomy. Voice-activated robotic arms routinely maneuver endoscopic cameras, and complex-master slave robotic systems are currently FDA approved, marketed, and used for a variety of procedures. It remains to be seen, however, if history will look on the development of robotic surgery as a profound paradigm shift or as a bump in the road to something even more important.[36–38] Paradigm shift or not, the origin of surgical robotics is rooted in the strengths and weaknesses of its predecessors. Minimally invasive surgery began in 1987 with the first laparoscopic cholecystectomy. Since then, the list of procedures performed laparoscopically has grown at a pace consistent with improvements in technology and the technical skills of surgeons. The advantages of minimally invasive surgery are very popular among surgeons, patients, and insurance companies. Incisions are smaller, the risk of infection is less, hospital stays are shorter, if necessary at all, and convalescence is reduced significantly. Many studies have shown that laparoscopic procedures result in decreased hospital stays, a quicker return to the workforce, decreased pain, better cosmesis, and better postoperative immune function.

As attractive as minimally invasive surgery is, there are several limitations. Some of the more prominent limitations involve the technical and mechanical nature of the equipment. Inherent in current laparoscopic equipment is a loss of haptic feedback (force and tactile), natural hand-eye coordination, and dexterity.[39,40] Moving the laparoscopic instruments while watching a two-dimensional video monitor is somewhat counterintuitive. One must move the instrument in the opposite direction from the desired target on the monitor to interact with the site of interest. Hand-eye coordination therefore is compromised. Some refer to this as the *fulcrum effect*. Current instruments have restricted degrees of motion; most have four degrees of motion, whereas the human wrist and hand have seven degrees of motion. There is also a decreased sense of touch that makes tissue manipulation more heavily dependent on visualization. Finally, physiologic tremors in the surgeon are readily transmitted through the length of the rigid instruments. These limitations make more delicate dissections and anastomoses difficult, if not impossible. The motivation to develop surgical robots is rooted in the desire to overcome the limitations of current laparoscopic technologies and to expand the benefits of minimally invasive surgery.[41,42]

From their inception, surgical robots have been envisioned to extend the capabilities of human surgeons beyond the limits of conventional laparoscopy. The history of robotics in surgery begins with the Puma 560, a robot used in 1985 by Kwoh and colleagues to perform neurosurgical biopsies with greater precision. Three years later, Davies and colleagues performed a transurethral resection of the prostate using the Puma 560. This system eventually led to the development of PROBOT, a robot designed specifically for transurethral resection of the prostate. While PROBOT was being developed, Integrated Surgical Supplies, Inc., of Sacramento, CA, was developing ROBODOC, a robotic system designed to machine the femur with greater precision in hip-replacement surgeries. ROBODOC was the first surgical robot approved by the FDA.

Also in the middle to late 1980s, a group of researchers at the National Aeronautics and Space Administration (NASA) Ames Research Center, working on virtual reality, became interested in using this information to develop telepresence surgery.[43,44] This concept of telesurgery became one of the main driving forces behind the development of surgical robots. In the early 1990s, several of the scientists from the NASA Ames team joined the Stanford Research Institute (SRI). Working with SRI's other robotocists and virtual-reality experts, these scientists developed a dexterous telemanipulator for hand surgery. One of their main design goals was to give the surgeon the sense of operating directly on the patient rather than from across the room. While these robots were being developed, general surgeons and endoscopists joined the development team and realized the potential these systems had in ameliorating the limitations of conventional laparoscopic surgery.

The U.S. Army noticed the work at SRI, and it became interested in the possibility of decreasing wartime mortality by "bringing the surgeon to the wounded soldier—through telepresence." With funding from the U.S. Army, a system was devised whereby a wounded soldier could be loaded into a vehicle with robotic surgical equipment and be operated on remotely by a surgeon at a nearby Mobile Advanced Surgical Hospital (MASH) unit. This system, it was hoped, would decrease wartime mortality by preventing wounded soldiers from exsanguinating before they reached the hospital. This system has been demonstrated successfully on animal models but has not yet been tested or implemented in actual battlefield casualty care.

Several of the surgeons and engineers working on surgical robotic systems for the Army eventually formed commercial ventures that lead to the introduction of robotics to the civilian surgical community. Notably, Computer Motion, Inc., of Santa Barbara, CA, used seed money provided by the Army to develop the Automated Endoscopic System for Optimal Positioning (AESOP), a robotic arm controlled by surgeon voice commands to manipulate an endoscopic camera. Shortly after AESOP was marketed, Integrated Surgical Systems (now Intuitive Surgical), of Mountain View, CA, licensed the SRI Green Telepresence Surgery System. This system underwent extensive redesign and was reintroduced as the Da Vinci Surgical System. Within a year, Computer Motion put the Zeus System into production.

6.15 Robot-Assisted Surgery

Through the use of robotic technology, the researchers of Scuola Superiore Sant Anna, in Pisa, Italy, developed a pill that has a camera in it that resembles a normally prescribed capsules, but it is fitted with a sophisticated camera, electronics and computing devices, batteries, and a communication system. This radio-controlled crawling capsule has six legs with tiny hooks on the ends. Although it is tiny enough not to damage the soft tissues, it has the capability to adhere tightly to the gut wall, resisting muscular pulsations or mucosal slipperiness. The capsule can park at the site of interest by releasing a clamp with 5-mm-long jaws provided with teeth. This technique is much less painful and uncomfortable than colonoscopy or gastroscopy.

Another similar device is the Pillcam capsule, a patented single-use video color imaging capsule that, after ingestion, glides naturally through the digestive tract and is finally excreted. It has an associated data recorder, worn as a belt around the waist, that receives signals transmitted by the capsule through an array of sensors placed on the patient's body. After wearing the belt during the GI examination, patients can carry out their normal activities during an 8-hour test. During this time, the M2A capsule transmits approximately

50,000 images to the portable data recorder. The output is viewed with application software that permits physicians to view, edit, archive, e-mail, and save individual images and short video clips. The latest advancements enable physicians to localize images, detect suspected bleeding areas, and efficiently view and archive the content in the videos.[45–47]

Such microscopic robots have been developed that can be injected into the body of a patient, and physicians can analyze medical conditions, deliver drugs, or perform minimally invasive surgery. The biomedical microrobotic system developed by the group of scientists from the Swiss Federal Institute of Technology in Zurich, Switzerland, has been demonstrated to successfully maneuver through a watery maze using external energy from magnetic forces.

Robotics is using shape-memory alloys (SMAs) for developing prototype legs that will walk through animal stomach tissue without damaging it, movement of the legs being controlled automatically. Because the legs being made SMAs, they remember their original shape and revert back to it—much like a spring—each time a step is taken. Researchers have developed a prototype capsule that is 25 mm long and 10 mm in diameter—a little larger than the existing endoscope. Research is in progress to miniaturize even further.

These minimally invasive robots are very suitable to navigate through coronary blood vessels and ocular fluids for the purpose of diagnosis and treatment, resulting in less injury and faster recovery times for patients.

Applications of robots in surgery have helped surgeons to do microsurgery without making deep and long incisions. By means of robotics, surgical assistants are sporting three arms, a computerized brain, and a glowing track record that helps to repair heart valves, remove cancerous prostates, bypass blocked coronary arteries, and accomplish gastric by pass.

Through the use of sophisticated remote-controlled robotic instruments, physicians are now able to carry out surgery in fetuses if they are detected with any such disorder. This medical area requires considerable microscopic manipulation devices and skills. Robotic surgery poses minimal danger to both fetuses and their mothers. Robotic operations are more precise than conventional surgical methods; patients recover soon, return to normal activity, have shorter hospital stays, use less pain medications, and suffer fewer complications.

In robotic-assisted surgery, the most widely used machine is the Da Vinci Surgical System, manufactured by the Intuitive Surgical, a firm that specializes in robotic surgical equipment. The robots are used for prostate removal surgery, for hysterectomies cervical and endometrial cancers, pediatric gallbladder removal, and stomach surgery to prevent gastric reflux (Fig. 6.6).

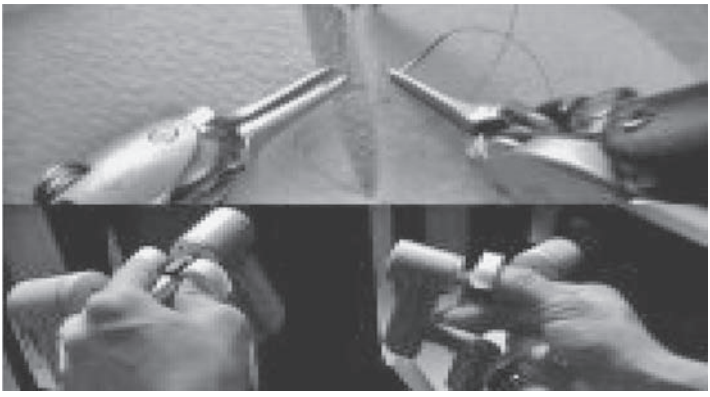


Figure 6.6 Robotic pyeloplasty of an obstructed kidney. (*youtube.com*.)

Laparoscopy is a minimally invasive surgery. As in laparoscopy, robotic surgery involves small incisions of one-fourth to three-fourths of an inch into which sleeves are inserted as ports for placement of specialized instruments and a video camera. Robotic surgery allows the surgeon to place his or her hands inside the patient without the need for large incisions (Fig. 6.7). Surgery with the Da Vinci system does not need close proximity of the surgeon to the patient. Unlike laparoscopy, the surgeon is seated across the room from the patient with his or her arms inserted into a nearby console, fingers

Figure 6.7 Robotic surgery with the Da Vinci system is an extension of the surgeon's hands and a big step beyond traditional laparoscopy. (*abcnews.go.com*.)

on stirrup-like holders and eyes fixed on lenses for sharp, magnified images of the surgical site, with the focus adjusted via foot pedals.

Laparoscopy allows manipulation of instruments up, down, and side to side, but surgery with the Da Vinci system allows more natural wrist movement.[45,47] Robot arms have wrists with the 8 degrees of freedom, allowing the surgeon to bend around corners and permitting full range of motion and the ability to rotate instruments 360 degrees through tiny incisions. The arrangement allows all-around vision and ability to zoom in and out. Another advantage of the Da Vinci system is the elimination of tremor.

The major benefit of this technology is in pediatrics, in fetal surgery in utero. However, use of robots in surgery is of great benefit to patients themselves, who recover quickly and are able to resume their normal life within days of major surgery instead of months.

6.16 Advantages and Disadvantages of Robot-Assisted Surgery

There are several disadvantages to these systems. First of all, robotic surgery is a new technology, and its uses and efficacy have not yet been well established. To date, mostly studies of feasibility have been conducted, and almost no long-term follow-up studies have been performed. Many procedures also will have to be redesigned to optimize the use of robotic arms and increase efficiency. However, time most likely will remedy these disadvantages.

Another disadvantage of these systems is their cost. With a price tag of a million dollars, their cost is nearly prohibitive. Whether the price of these systems will fall or rise is a matter of conjecture. Some people believe that with improvements in technology and as more experience is gained with robotic systems, the price will fall. Others believe that improvements in technology, such as haptics, increased processor speeds, and more complex and capable software, will increase the cost of these systems. Also at issue is the problem of upgrading systems. How much will hospitals and health care organizations have to spend on upgrades and how often? In any case, many people believe that to justify the purchase of these systems, they must gain widespread multidisciplinary use.

Another disadvantage is the size of these systems. Both systems have relatively large footprints and relatively cumbersome robotic arms. This is an important disadvantage in today's already crowded operating rooms. It may be difficult for both the surgical team and the robot to fit into the operating room. Some people suggest that miniaturizing the robotic arms and instruments will address the problems associated with their current size. Others believe that larger operating suites with multiple booms and wall mountings will be needed to accommodate the extra space requirements of robotic surgical

systems. The cost of making room for these robots and the cost of the robots themselves make them an especially expensive technology.

One of the potential disadvantages identified is a lack of compatible instruments and equipment. Lack of certain instruments increases reliance on tableside assistants to perform part of the surgery. This, however, is a transient disadvantage because new technologies have and will develop to address these shortcomings.

Most of the disadvantages identified will be remedied with time and improvements in technology. Only time will tell if the use of these systems justifies their cost. If the cost of these systems remains high and they do not reduce the cost of routine procedures, it is unlikely that there will be a robot in every operating room and thus unlikely that they will be used for routine surgeries.

While discussing the advantages of robotic surgery, it is important to mention that the conventional surgical system is flexible, adaptable, dexterous, intensive, and easy to instruct and debrief, but it suffers from the factors of limited geometric accuracy and human tremor and fatigue offers limited sterility and exposure to radiation and infection. In the case of robotic-assisted surgery, a surgeon finds good scale motion and good geometric accuracy, can use diverse sensors in control, the methods are resistant to radiation and infection, they offer 3D visualisation, in cases of laparoscopic surgery surgeons can get 7 degrees of freedom, and the procedures are nontiring and stable for surgeons.[48,49]

6.17 Applications of Robotic Surgery

Many general surgical procedures can be carried out by the robotic surgical system. The world's first-ever robotic pancreatectomy was performed by a team of the University of Illinois in Chicago, the medical team lead by Professor Pier Christoforo Giulianotti. In April 2008, the world's first minimally invasive liver resection was done for living-donar transplantation; 60 percent of the liver was removed with only four holes, and no surgical scars were left. The patient was released within a few days and had very much less surgical pain.

An endoscopic coronary artery bypass and mitral valve repair and replacement magnetic navigation system has been developed for safety in ablation procedures for arrhythmias and atrial fibrillation. It also has been used in interventional cardiology. The first robotic cardiac procedure was performed in United States in 1999 at Ohio State University, Columbus.[50] Other robotic-assisted surgeries include lung resections, tumor resections, and esophagectomy.

Robotic surgery has been used successfully in the field of gynecology. The Da Vinci system was used in benign gynecology, gynecologic oncology, hysterectomies, and gynecologic cancer staging. Use of robotic surgery has virtually eliminated the need for the large abdominal incisions previously needed in gynecology.[42,46,48]

In the field of neurosurgery, MD Robotic's Neuroarm is the world's first MRI-compatible surgical robot. Robotics also has gained a significant position in the field of orthopedics for total hip or knee replacement.

The Cyberknife robotic radiosurgery system uses computer-controlled robotics to treat tumors throughout the body. In the field of urology, the Da Vinci, system has been used to remove the prostate gland in cancers and to treat bladder abnormalities and kidney obstructions.

Surgical robots have been used in pediatric surgical procedures, including tracheoesophagal fistula repair, cholecystectomy, congenital diaphragmatic hernia, Morgagni's hernia, and so on.

6.18 Practical Uses of Surgical Robots Today

In today's competitive health care market, many organizations are interested in making themselves "cutting edge" institutions with the most advanced technological equipment and the very newest treatment and testing modalities. Doing so allows them to capture more of the health care market. Acquiring a surgical robot is, in essence, the entry fee into marketing an institution's surgical specialties as the "most advanced." It is not uncommon, for example, to see a photo of a surgical robot on the cover of a hospital's marketing brochure and yet see no mention of robotic surgery inside.

As far as ideas and science go, surgical robotics is deep, fertile soil. It may come to pass that robotic systems are used very little, but the technology they are generating and the advances in ancillary products will continue. Already the development of robotics is spurring interest in new tissue anastomosis techniques, improved laparoscopic instruments, and digital integration of already existing technologies.

As mentioned previously, applications of robotic surgery are expanding rapidly into many different surgical disciplines. The cost of procuring one of these systems remains high, however, making it unlikely that an institution will acquire more than one or two. This low number of machines and the low number of surgeons trained to use them make incorporation of robotics into routine surgeries unlikely. Whether this changes with the passing of time remains to be seen.

6.19 Future of Robotic Surgery

Robotic surgery is in its infancy. Many obstacles and disadvantages will be resolved in time, and no doubt many other questions will arise. Many questions have yet to be asked, questions such as malpractice liability, credentialing, training requirements, and interstate licensing for telesurgeons, to name just a few.

Many of current advantages of robotic-assisted surgery ensure its continued development and expansion. For example, the sophistication

of the controls and the multiple degrees of freedom afforded by the Zeus and Da Vinci systems allow increased mobility and no tremor without comprising the visual field to make microanastomoses possible. Many people have observed that robotic systems are information systems, and as such, they have the ability to interface and integrate with many of the technologies being developed for and currently used in the operating room. One exciting possibility is to expand the use of preoperative (computed tomography or magnetic resonance imaging) and intraoperative video image fusion to better guide the surgeon in dissection and identifying pathology. These data also may be used to rehearse complex procedures before they are undertaken. The nature of robotic systems also makes the possibility of long-distance intraoperative consultation or guidance possible, and it may provide new opportunities for teaching and assessment of new surgeons through mentoring and simulation. Computer Motion, the makers of the Zeus Robotic Surgical System, is already marketing a device called SOCRATES that allows surgeons at remote sites to connect to an operating room and share video and audio, to use a "telestrator" to highlight anatomy, and to control the AESOP endoscopic camera.

Technically, much remains to be done before robotic surgery's full potential can be realized. Although these systems have greatly improved dexterity, they have yet to develop the full potential in instrumentation or to incorporate the full range of sensory input. More standard mechanical tools and more energy-directed tools need to be developed. Some authors also believe that robotic surgery can be extended into the realm of advanced diagnostic testing with the development and use of ultrasonography and near-infrared and confocal microscopy equipment. Much like the robots in popular culture, the future of robotics in surgery is limited only by the imagination. Many future "advancements" are already being researched. Some laboratories, including our laboratory, are currently working on systems to relay touch sensation from robotic instruments back to the surgeon. Other laboratories are working on improving current methods and developing new devices for suture-less anastomoses. When most people think about robotics, they think about automation. The possibility of automating some tasks is both exciting and controversial. Future systems may include the ability for a surgeon to program the surgery and merely supervise as the robot performs most of the tasks. The possibilities for improvement and advancement are limited only by the imagination and cost.[47,49]

6.20 Conclusion

Although still in its infancy, robotic surgery has already proven itself to be of great value, particularly in areas inaccessible to conventional laparoscopic procedures. It remains to be seen, however, if robotic systems will replace conventional laparoscopic instruments in less

technically demanding procedures. In any case, robotic technology is set to revolutionize surgery by improving and expanding laparoscopic procedures, advancing surgical technology, and bringing surgery into the digital age. Furthermore, it has the potential to expand surgical treatment modalities beyond the limits of human ability. Whether or not the benefit of its use overcomes the cost to implement it remains to be seen, and much remains to be worked out. Although feasibility largely has been shown, more prospective randomized trials evaluating efficacy and safety must be undertaken. Further research must evaluate cost-effectiveness or the true benefit over conventional therapy for robotic surgery to take full root.

References

1. Hammond ME, Taube SE. Issues and barriers to development of clinically useful tumor markers: a development pathway proposal. *Semin Oncol*. 29: 213–221, 2002.
2. Negm RS, Verma M, Srivastava S. The promise of biomarkers in cancer screening and detection. *Trends Mol Med*. 8:288–293, 2002.
3. U.S. Department of Health and Human Services, U.S. Food and Drug Administration. *Challenge and Opportunity on the Critical Path of New Medical Products*. Available at www.fda.gov/oc/initiatives/criticalpath/whitepaper.tml (2004).
4. Anderson NL, Anderson NG. The human plasma proteome: History, character, and diagnostic prospects. *Mol Cell Proteomics*. 1:845–867, 2002.
5. Abulafia O, Sherer DM. Automated cervical cytology: Meta-analyses of the performance of the PAPNET system. *Obstet Gynecol Surv.* 54:253–264, 1999.
6. Erickson BJ, Bartholmai B. Computer-aided detection and diagnosis at the start of the third millennium. *J Digit Imaging*. 15:59–68, 2002.
7. American Joint Committee on Cancer and American Cancer Society. Green FL, et al., eds., *AJCC Cancer Staging Handbook: From the AJCC Cancer Staging Manual*. New York: Springer, 2002.
8. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. *Nature Genet.* 33:238–244, 2003.
9. Relling M, Dervieux T. Pharmacogenetics and cancer therapy. *Nature Rev Cancer*. 1:99–108, 2001.
10. Lenz HJ. The use and development of germline polymorphisms in clinical oncology. *J Clin Oncol.* 22:2519–2521, 2004.
11. Leon SA, Shapiro B, Sklaroff DM, and Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res*. 37:646–650, 1977.
12. Sidransky D. Emerging molecular markers of cancer. *Nature Rev Cancer*. 2:210–219, 2002.
13. Levine AJ. p53, the cellular gatekeeper for growth and division. *Cell*. 88: 323–331, 1997.
14. Jenuwein T, Allis C. Translating the histone code. *Science.* 293:1074–1080, 2001.
15. Esteller M, Corn PG, Baylin SB, Herman JG. A gene hypermethylation profile of human cancer. *Cancer Res*. 61:3225–3229, 2001.
16. Yamaguchi S, Asao T, Nakamura JI, et al. High frequency of DAP-kinase gene promoter methylation in colorectal cancer specimens and its identification in serum. *Cancer Lett.* 194:99–105, 2003.
17. Weinstein JN. Pharmacogenomics: Teaching old drugs new tricks. *N Engl J Med*. 343:1408–1409, 2000.
18. Esteller M. DNA methylation and cancer therapy: New developments and expectations. *Curr Opin Oncol*. 17:55–60, 2005.

19. Sorlie T, Perou C, Brown P, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA.* 98:10869–10874, 2001.
20. Verma M, Wright GL Jr, Hanash SM, et al. Proteomic approaches within the NCI early detection research network for the discovery and identification of cancer biomarkers. *Ann NY Acad Sci.* 945:103–115, 2001.
21. Wulfkuhle JD, Liotta LA, Petricoin EF. Proteomic applications for the early detection of cancer. *Nature Rev Cancer.* 3:267–275, 2003.
22. Barker PE. Cancer biomarker validation: Standards and process. Roles for the National Institute of Standards and Technology (NIST). *Ann NY Acad Sci*. 983:142–150, 2003.
23. Ransohoff DF. Rules of evidence for cancer molecular marker discovery and validation. *Nature Rev Cancer.* 4:309–314, 2004.
24. Schlieman MG, Ho HS, Bold RJ. Utility of tumor markers in determining resectability of pancreatic cancer. *Arch. Surg*. 138:951–956, 2003.
25. Duncan JL, Price A, Rogers K. The use of CA15.3 as a serum tumor marker in breast carcinoma. *Eur J Surg Oncol*. 17:16–19, 1991.
26. Madon MJ, Morrison SR. *Chemical Sensing with Solid State Devices*. New York: Academic Press, 1989.
27. Gangopadhay, R, De A. Conducting polymer composites, noble materials for gas sensing. *Sensor Actuators B*. 77:326–329, 2001.
28. *Introduction to Bioanalytical Sensors*. New York: Wiley, 1998.
29. Sukeerthi S, Contractor AQ. Application of conducting polymers as sensors. *Indian J Chem*. 33:565–571, 1994.
30. Turner AP. Tech. sight, biochemistry. Biosensors: Sense and sensitivity. *Science* 290:1315–1317, 2000.
31. Giuliano KA, Chen, YT, Haskins JR. Positional biosensors: A new tool for high content screening. *Mod Drug discovery*. 6:33–37, 2003.
32. Giuliano KA, Taylor Dl. Fluorescent protein biosensors: New tools for drug discovery. *Trends Biotechnol.* 16:135–140, 1998.
33. Giuliano KA, Post PL, Hahn Km, Taylor DL. Fluorescent protein biosensors: Measurement of molecular dynamics in living cells. *Annu Rev Biophys Biomol Struct*. 24:405–434, 1995.
34. Eggins BR. *Chemical Sensors and Biosensors*. New York: Wiley, 2002.
35. www.caimr.georgetown.edu/Medical%20Robotics%20Workshop/Presentations/Sakuma%20Presentation%20CARS%202005pdf.
36. http://ranier.hq.nasa.gov/telerobotics_page/Technologies?0903.html.
37. www.me.cmu.edu/facultyl/sitti/nano/links/rogotics.html.
38. http://staff.aist.go.jp/tamio.tanikawa/project/micro-e.html.
39. http://dsc.Discovery.Com/news/briefs/20050822/microbot.html.
40. www.wtec.org/robotics/workshop/PDF/06-BiologicalApplications-Zheng.pdf.
41. Satava RM. Surgical robotics: The early chronicles. A personal historical perspective. *Surg Laparosc Endosc Percutan Tech*. 12:6–16, 2002.
42. Cheah WK, Lee B, Lenzi JE, et al. Telesurgical laparoscopic cholecystectomy between two countries. *Surg Endosc*. 14:1085, 2000.
43. Marescaux J, Leroy J, Rubino F, et al. Transcontinental robot-assisted remote telesurgery: Feasibility and potential applications. *Ann Surg*. 235:487–492, 2002.
44. Felger JE, Nifong L. The evolution of and early experience with robot assisted mitral valve surgery. *Surg Laparosc Endosc Percutan Tech*. 12:58–63, 2002.
45. Jones SB, Jones DB. Surgical aspects and future developments in laparoscopy. *Anesthiol Clin North Am*. 19:107–124, 2001.
46. Fuchs KH. Minimally invasive surgery. *Endoscopy*. 34:154–159, 2002.
47. Davies B. A review of robotics in surgery. *Proc Inst Mech Eng*. 214:129–140, 2000.
48. Dario P, Corrozza MC, Peitrabissa A. Development and in vitro testing of a miniature robotic system for computer-assisted colonoscopy. *Comput Aided Surg*. 4:1–14, 1999.
49. Kennedy CW, Desai JP. Force feedback using vision. 11th International Conference on Advanced Robotics, June 30–July 3, 2003, University of Coimbra, Portugal.

CHAPTER 7
Biomedical Instrumentation

7.1 Introduction

Bioinstrumentation generally implies measurement of the biologic variables, although measurements in the field of biology are referred to as *biometrics*. The branch of science that includes the measurement of physiologic variables and parameters is known as *biometrics*. Biometrics deals with the measurement of physiologic variables and parameters, and it is biomedical instrumentation that provides the tools by which these measurements can be achieved. But biometrics also includes some mathematical and statistical methods. Biomedical instrumentation is really unique in the field of medicine.[1,2]

Bioengineering is defined as the study of both engineering and biologic sciences so that both can be fully used for the benefit of humankind. A person who works in research and development (R&D) in the interface area of medicine and engineering is called a *biomedical engineer*. A *clinical engineer* is a professional who brings to health care facilities a level of education, experience, and accomplishment that enables him or her to responsibly, effectively, and safely manage and interface with medical devices, instruments, and systems. Clinical engineers come into this profession through the engineering degree route, but some also may start out as physicists or physiologists.

A *biomedical equipment technician* is an individual who is knowledgeable about the theory of operation, underlying physiologic principles, and the practical and safe clinical application of biomedical equipment. Among his or her duties and responsibilities are included installation, calibration, inspection, preventive maintenance, and repair of the technical equipment and its operation, as well as ensuring the safety of the equipment.[3,4]

The field of medical instrumentation is developing day by day. Research in this field began as early as the nineteenth century when the electrocardiogram (ECG) was first used by Einthoven. However, progress in this field has been rather slow because of the high

development costs of medical instruments and sometimes because of the suspicious and uncooperative nature of the medical profession and hospital staff in terms of the use of new equipment.

Many developments with excellent potential seemed to have become lost causes. It was during this period that some progressive companies decided that rather than modifying existing hardware, they would design instrumentation especially for medical use. A large measure of help was provided by U.S. government, particularly the National Aeronautics and Space Agency (NASA). Aerospace medicine programs expanded both within NASA facilities and through grants to universities and hospital research units. Along with medical research programs at universities, a need to develop courses and curricula in biomedical engineering was appreciated.

Biomedical instrumentation can be viewed as information-gathering instrumentation, although it sometimes includes certain monitoring and control devices. Among biomedical instrumentation, some aid physicians in the diagnosis of diseases and disorders of the human body, and some special instrumentation systems are used for monitoring patients undergoing surgery or under intensive care.

Biomedical instrumentation may be classified into two major types:

1. *Clinical instrumentation* that is devoted mainly to the diagnosis, care, and treatment of the patients
2. *Research instrumentation* that is used primarily in the search for new knowledge pertaining to the various systems that compose the human organism

Clinical instruments can be used by the physician/nurse, and research instruments are used by skilled technologists.[4]

7.2 Biomedical Instrumentation Measurements

Measurements in which biomedical instrumentation is employed are divided into two categories: in vivo and in vitro. An *in vivo measurement* is one that is made on or within a living organism itself. An *in vitro measurement* is one performed outside the body, even though it relates to the functions of the body. However, in order to obtain valid measurements from a living human being, it is also necessary to have a sound knowledge of the subject on which the measurements are made. Within the human body can be found electrical, mechanical, thermal, hydraulic, pneumatic, and various other types of systems, each of which communicates with the external environment and internally with other systems of the body. By means of a multilevel control system and communication network, these individual systems are organized to perform many complex functions.

Next to the whole being in the hierarchy of organization comes the major functional systems of the body, including the nervous systems, cardiovascular system, pulmonary system, and so on. The functional systems can be broken down into subsystems and organs, which are further subdivided into smaller and smaller units. This process can continue down to cellular level and perhaps even to the molecular level. The major goal of biomedical instrumentation is to make possible the measurement of information communicated by these various elements.

The human body can be regarded as a conglomerate of chemical systems involving messenger agents for communication, materials for body repair and growth, and so on. For a biomedical engineer, the cardiovascular system can be viewed as a complex, closed hydraulic system with a four-chamber pump (the heart) that is connected to flexible and elastic tubing, the blood vessels. These tubes change their diameter to satisfy control requirements and are also provided with variable hydraulic resistances (the vasoconstrictors and vasodilators). The four-chamber pump acts as two synchronized but functionally isolated two-stage pumps. The first stage of each pump (the atrium) collects fluid (blood) from the system and then pumps it into second stage (the ventricle).[5] One of the two stage pumps (the right side of the heart) collects fluid from the main hydraulic system (the systemic circulation) and pumps it through an oxygenation system (the lungs). The other pump (the left side of the heart) receives fluid (blood) from the oxygenation system and pumps it into the main hydraulic system; the speed of the pump (heart rate) and its efficiency (stroke volume) are constantly changed to meet the overall requirements of the system.

Just as the cardiovascular system is the major hydraulic system in the body, the respiratory system is the pneumatic system. Here, an air pump (the diaphragm) alternately creates negative and positive pressures in the sealed chamber (the thoracic cavity) and causes the air to be sucked into and forced out of a pair of the elastic bags (the lungs). An automatic control center (the respiratory center of the brain) maintains pump operation at a speed that is adequate to supply oxygen and carry off carbon dioxide as required by the system.

The nervous system is the communication network for the body. The center of this network is a self-adapting central information processor or computer (the brain) with memory, computational power, decision-making capability, and a myriad of input-output channels.

The brain, or the central computer, is a self-adapting one. The central computer (the brain) is provided with millions of communication lines (the afferent and efferent nerves) that bring sensory information into and transmit control information out of the brain. Information is usually coded into the system by means of electrochemical pulses (nerve action potentials) that travel along signal lines

(the nerves). In addition to the central computer, a large number of simple decision making devices (spinal reflexes) are present to control directly certain motor devices from certain sensory inputs.

Now, the different systems of our body convey useful information about the functions they perform. These signals are the bioelectric potentials associated with nerve conduction, brain activity, heartbeat, muscle activity, and so on. Bioelectric potentials occur as a result of electrochemical activity of certain special types of cell. The biopotentials generated by the muscles of heart result in an *electrocardiogram* (ECG). The recorded representation of bioelectric potentials generated by the neuronal activity of the brain is called an *electroencephalogram* (EEG). Bioelectric potentials associated with muscle activity constitute an *electromyogram* (EMG). A record of the complex pattern of bioelectric potentials obtained from the retina of the eye constitutes an *electroretinogram* (ERG). A measure of the variations in the corneal-retinal potential as given by the position and movement of the eye is referred to as an *electrooculogram* (EOG). An *electrogastrogram* (EGG) is associated with the peristaltic movements of the gastrointestinal (GI) tract.

A number of electrodes are used to measure bioelectric events. The electrodes used to measure bioelectric potentials near or within a single cell are known as *microelectrodes*. Electrodes used to measure ECG, EEG, and EMG potentials from the surface of the skin are known as *skin-surface electrodes.* Electrodes used to penetrate the skin to record EEG potentials from a local region of the brain or EMG potentials from a specific group of muscles are called *needle electrodes. Bloodgas electrodes* are very important for the physiologic measurements of the partial pressures of oxygen and carbon dioxide in the blood. The partial pressure of oxygen (i.e., the oxygen tension) can be measured both in vivo and in vitro.[4,5]

For this purpose, a platinum cathode and the reference electrode are both integrated into a single unit—the *Clark electrode.* This electrode is placed at the tip of a catheter for insertion into various parts of the heart or vascular system for direct in vivo measurements.

Measurement of the partial pressure of carbon dioxide (P_{CO_2}) makes use of the fact that there is a linear relationship between the logarithm of the P_{CO_2} and the pH of the solution. The most improved type of P_{CO_2} electrode is called the *Severinghaus electrode.* In certain cases, measurements of P_{O_2} are combined into a single electrode that constitutes a *common reference half-cell.*

7.2.1 Instrumentation for Cardiovascular Measurements

This instrumentation deals with the physiologic measurements relating to the cardiovascular system. Here, stress will be placed on the measurement of biopotentials resulting in ECG and both direct and indirect methods of measuring blood pressure, blood flow, cardiac output, and blood volume.

Electrocardiography

The discovery of the fact that muscle contractions involve electrical processes dates back to the eighteenth century. In 1887, the first ECG was recorded by Waller, who used the capillary electrometer. The string galvanometer, which was introduced to electrocardiography by Einthoven in 1903, was a considerable improvement.[6]

1. The ECG is used to diagnose diseases and conditions relating to the heart.
2. The ECG is the graphic recording or display of the time-variant voltages produced by the myocardium during the cardiac cycle.
3. The P, QRS, and T waves reflect the rhythmic electrical depolarization and repolarization of the myocardium associated with contractions of the atria and ventricles.

This instrument consists of an extremely thin platinum wire or a gold-plated quartz fiber suspended in the air gap of a strong electromagnet. An electric current flowing through the wire or fiber causes movement of the wire or fiber perpendicular to the direction of the magnetic field. The magnitude of the movement is small, but it can be magnified several hundred times by an optical projection system for recording on a moving film or paper. The small mass of the moving wire or fiber results in a frequency response sufficiently high for faithful recording of the ECG. The previously used string galvanometers are now replaced by electronic amplifiers. However, electronic amplifiers suffer from interference problems, so differential amplifiers are now used to record other bioelectric signals, including the ECG.

To record an ECG, a number of electrodes (usually five) are affixed to the body of the patient, and the electrodes are connected to the ECG machine by the same number of electrical wires. These wires, more specifically the electrodes, are usually called *leads*. Actually, *leads* refer to a particular group of electrodes and the way in which they are connected to the amplifier.

While working with ECGs Einthoven postulated that at any given instant of the cardiac cycle, the frontal-plane representation of the electrical axis of the heart is a two-dimensional (2D) vector. Einthoven assumed that the heart (origin of the vector) is at the center of an equilateral triangle, the apexes of which are the right and left shoulders and the crotch. This triangle is known as the *Einthoven triangle*.[7,8]

Some normal values for amplitude and duration of important ECG parameters are as follows (Fig. 7.1):

Amplitude:	P wave	0.25 mV
	R wave	1.60 mV
	Q wave	25% of R wave
	T wave	0.1–0.5 mV

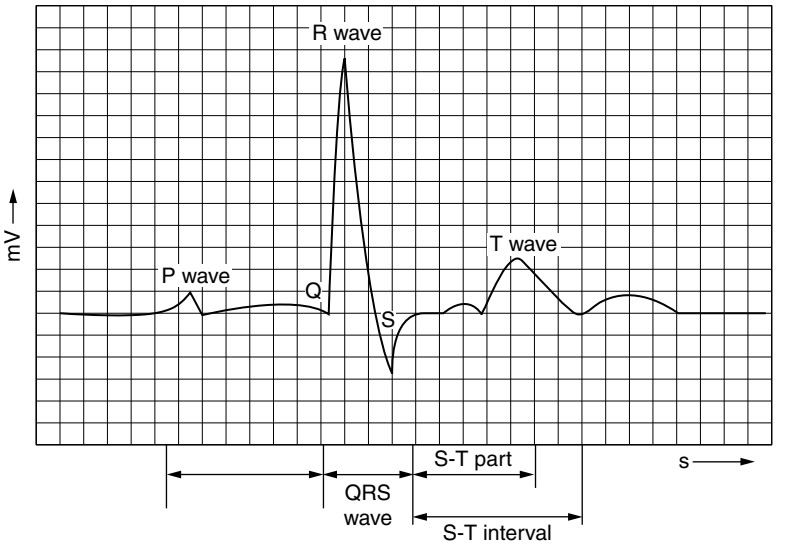


Figure 7.1 Normal pattern of ECG.

Duration:	PR interval	0.12–0.2 s
	QT interval	0.35–0.44 s
	ST segment	0.05–0.15 s
	P-wave interval	0.11 s
	QRS interval	0.09 s

For the purpose of diagnosis, when a cardiologist takes the ECG of a patient, he or she first observes whether heart rate lies within the normal range of 60 to 100 beats per minute. A rate that is slower than normal rate is *bradycardia*, and a rate that is faster than normal rate is *tachycardia*. The next object of observation is whether the cycles are evenly spaced, and if they are not, an *arrhythmia* is indicated. If the PR interval is greater than 0.2 s, a blockage of the atrioventricular (AV) node is suggested. If more basic features of the ECG are missing, a heart block may be suggested.

In healthy individuals, the ECG is reasonably normal. However, several changes in the ECG are seen in pathologic conditions. These include (1) altered path of excitation in the heart, (2) changed origin of waves (ectopic beats), (3) altered relationships of the features, (4) changed magnitudes of one or more features, and (5) differing durations of the waves or intervals. Although the ECG provides invaluable diagnostic information, especially about arrhythmias and myocardial infarction, certain disorders cannot be discerned from the ECG. Then other diagnostic techniques, such as angiography and echocardiography, are applied.

Measurement of Blood Pressure

Blood pressure measurement at the right time has saved the lives of many people indicating the need for treatment at the right time in the face of dangerously high blood pressure (i.e., hypertension). In routine clinical tests, blood pressure is usually measured by an indirect method using a *sphygmomanometer*. However, direct blood pressure measurements provide a continuous readout or recording of the blood pressure waveform and give more accurate readings than the indirect method.

Indirect Measurement This method involves measuring blood pressure with a sphygmomanometer and a stethoscope. The sphygmomanometer consists of an inflatable pressure cuff and a mercury or aneroid manometer to measure the pressure in the cuff. The cuff consists of a rubber bladder inside of an inelastic fabric covering that can be wrapped around the patient's upper arm and then fastened with either hooks or a Velcro fastener.

The sphygmomanometer works on the principle that when the cuff is placed on the upper arm and inflated, arterial blood can flow past the cuff only when the arterial pressure exceeds the pressure in the cuff. When the cuff is inflated to a pressure that only partially occludes the brachial artery, turbulence is generated in the blood as it spurts through the tiny arterial opening during each systole. The sound of this turbulence, known as *Korotkoff sounds*, can be heard through a stethoscope placed over the artery downstream of the cuff (Fig. 7.2).[9]

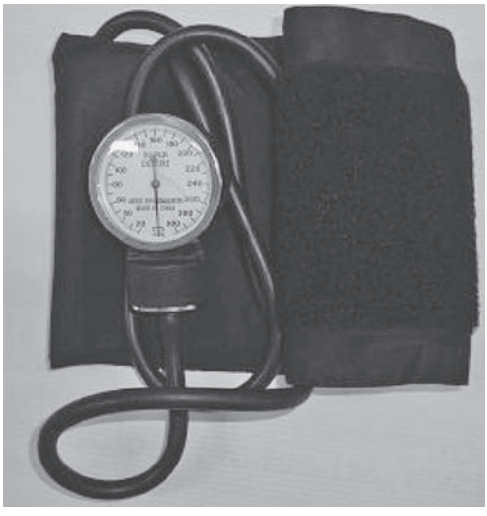


Figure 7.2 Sphygmomanometer.

To obtain a blood pressure measurement with a sphygmomanometer and a stethoscope, the pressure cuff on the upper arm is first inflated to a pressure well above the systolic pressure. At this point, no sounds can be heard through the stethoscope, which is placed over the brachial artery, because that artery has been collapsed by the pressure of the cuff. The pressure in the cuff is gradually reduced. As soon as cuff pressure falls below the systolic pressure, a small amount of blood spurts past the cuff, and Korotkoff sounds begin to be heard through stethoscope. The pressure of the cuff that is indicated on the manometer when the first Korotkoff sound is heard is recorded as the *systolic blood pressure*.

As the pressure in the cuff continues to drop, the Korotkoff sounds continue until the cuff pressure is no longer sufficient to occlude the vessel during any part of the cycle. Below this pressure, the Korotkoff sounds disappear, marking the value of the *diastolic pressure*.

The method of locating the systolic and diastolic pressures by listening to the Korotkoff sounds is called the *auscultatory method of sphygmomanometry*. An alternative method is the *palpatory method,* where the physician identifies the flow of blood in the artery by feeling the pulse of the patient downstream from the cuff instead of listening for the Korotkoff sounds.

Automated Indirect Methods For automation of the indirect procedure, a number of automatic and semiautomatic systems have been devised for recording blood pressure by the indirect procedure. Such devices use a pressure transducer connected to the sphygmomanometer cuff, a microphone placed beneath the cuff, and a standard physiologic recording system on which cuff pressure and the Korotkoff sounds are recorded. In this instrument, the pressure cuff is automatically inflated to about 220 mmHg and allowed to deflate slowly. The microphone picks up the Korotkoff sounds from the artery near the surface, just below the compression cuff. An example of an automatic blood pressure meter is the programmed electrosphygmomanometer PE-300. This instrument is designed for use in conjunction with an occluding cuff, microphone or pulse transducer, and a recorder for automatic measurement of indirect systolic and diastolic blood pressures.

Direct Measurements For direct measurement of blood pressure, one of the three methods can be adopted as follows:

1. Percutaneous insertion
2. Catheterization (vessel cutdown)
3. Implantation of a transducer in a vessel or in the heart

For the purpose of percutaneous insertion, a local anaesthetic is injected near the site of insertion. After occluding the vessel, a hollow

needle is inserted at a slight angle toward the vessel. When the needle is in place, a catheter is fed through the hollow needle with some sort of guide. When the catheter is securely in place in the vessel, the needle and guide are withdrawn. In some cases, a type of needle attached to an airtight tube is used so that the needle can be left in the vessel and the blood pressure measured directly by attaching a transducer to the tube.

The catheterization technique was first developed in 1940. This diagnostic technique, apart from obtaining blood pressures in the heart chambers and great vessels, is also used to obtain blood samples from the heart for oxygen-content analysis as well as to detect the location of abnormal blood flow pathways.

Catheters are also used for investigations with injection of radiopaque dyes for x-ray studies and colored dyes for indicator dilution studies and to study the actions of vasoactive drugs. A catheter is a long tube introduced into the heart or major vessel through a superficial artery or vein. A sterile catheter is designed in such a way that it travels easily through the vessel.

Measurement of blood pressure with a catheter can be done in two ways. In the first method, a sterile saline solution is introduced to the catheter so that the fluid pressure is transmitted to a transducer outside the body. A complete fluid pressure system is set up with provisions for checking against atmospheric pressure and for establishing a reference point. The frequency response of this system is a combination of the frequency response of the transducer and the fluid column in the catheter.

In the second method, pressure measurements can be obtained at the source, where a transducer is introduced into the catheter and pushed to the point at which the pressure is to be measured, or the transducer is mounted at the tip of the catheter. This device is a catheter-tip blood pressure transducer.

Implantation techniques involve major surgery and are applied for the purpose of research. Here, the transducer is kept fixed in place in the appropriate vessels for long periods of time.

Specific Direct Measurement Techniques Methods of direct blood pressure determination can be classified in two ways: first, by the clinical method by which the measuring device is coupled to the patient and, second by the electrical principle involved. Four different categories result. The first is a catheterization method involving the sensing of blood pressure through a liquid column. In this method, the transducer is external to the body, and the blood pressure is transmitted through a saline-solution column in a catheter to this transducer. The second category is a catheterization method involving placement of the transducer through a catheter at the actual site of measurement in the bloodstream or by mounting the transducer on the tip of the catheter. Third is a percutaneous method in which the blood pressure is

sensed in the vessel just under the skin by the use of a needle or catheter. Finally, there are implantation techniques in which the transducer is placed more permanently in the blood vessel or the heart by the surgical methods.

Floatation Catheters These have been designed so that after insertion, continuous monitoring of the pulmonary artery pressure becomes feasible in the clinical setting. The catheter may be inserted percutaneously or via a venous cutdown. By means of the use of continuous pressure and ECG monitoring, the catheter can be threaded into the subclavian vein with the balloon deflated. At this point, the balloon is partially inflated to half capacity and carried downstream to the right atrium by the flow of blood. The balloon then is inflated fully and advanced again so that the blood flow propels it through the tricuspid value into the right ventricle. From there it travels via pulmonary artery and wedges in a distal artery branch. The position of the catheter is verified by the pressure tracing, which shifts from a pulmonary artery pressure indicator to the "wedged" pressure waveform position.

Percutaneous Transducers Here, the transducer is connected to a hypodermic needle that has been placed in a vessel of the arm. The three-way stopcock dome permits flushing of the needle, administeration of drugs, and withdrawal of blood samples. This transducer can measure arterial and venous pressures or the pressures of other physiologic fluids by direct attachment to a needle at the point of attachment.

Implantable Transducers This type of transducer is made of titanium, which has excellent corrosion resistance, a low thermal coefficient of expansion, and a low modulus of elasticity, resulting in greater strain per unit stress. These types of transducers are implanted in the wall of the heart itself and are particularly useful for long-term investigations in animals.

7.3 Measurement of Blood Flow and Cardiac Output

Almost all blood flow meters, whether used clinically or for research applications, are based on the following physical principles:

1. Electromagnetic induction
2. Ultrasound transmission or reflection
3. Thermal convection
4. Radiographic principles
5. Indicator (dye or thermal) dilution

Magnetic and ultrasonic blood flow meters measure the velocity of the bloodstream. Ultrasound can be used transcutaneously to detect obstruction of blood vessels where a quantitative blood flow measurement is not required.[10]

Magnetic Blood Flow Meters or Electromagnetic Induction

Magnetic blood flow meters are based on the principle of magnetic induction. An electrical conductor, if moved through a magnetic field, will induce a voltage in the conductor that is proportional to the velocity of its motion. A permanent magnet or electromagnet is positioned around the blood vessels to generate a magnetic field perpendicular to the direction of blood flow. The voltage induced in the moving blood column is measured with stationary electrodes located on opposite sides of the blood vessel and perpendicular to the direction of the magnetic field. The commonly used types of implantable blood flow probes are slip-on or C-type transducers, cannula-type transducers, and catheter-tip transducers.

Ultrasonic Blood Flow Meters

In an ultrasonic blood flow meter, a beam of the ultrasonic energy is used to measure the velocity of flowing blood. In the transit-time ultrasonic flow meter, a pulsed beam is directed through a blood vessel at a shallow angle, and its transit time then is measured. When the blood flows in the direction of energy transmission, the transit time is shortened. If it flows in the opposite direction, the transit time is lengthened.

Most ultrasonic blood flow meters are based on the *Doppler principle*. An oscillator operating at a frequency of several megahertz excites a piezoelectric transducer that is coupled to the wall of an exposed blood vessel and sends an ultrasonic beam with a frequency F into the flowing blood. A small part of the transmitted energy is scattered back and is received by a second transducer arranged opposite the first one.

The scattering occurs mainly as a result of the moving blood cells; hence the reflected signal has a different frequency owing to the *Doppler effect*. Its frequency is either $F + F_D$ or $F - F_D$ depending on the direction of flow. The Doppler component F_D is directly proportional to the flowing blood.

The Doppler signal is typically in the low-frequency audio range. Because of the velocity profile of the flowing blood, the Doppler signal is not a pure sine wave but in the form of narrowband noise. Thus, from a loudspeaker or earphone, the Doppler signal of the pulsating blood flow can be heard as a characteristic "swish–swish."

Transducers for ultrasonic blood flow meters can be implanted for chronic use. Some of the commercially available blood flow meters of this type also incorporate a telemetry system.

Blood Flow Measurement by Thermal Convection

Blood velocity can be measured by this technique. A thermistor is positioned in the bloodstream at a constant temperature by a servo system. The electrical energy required to maintain this constant temperature is a measure of the flow rate.

In another method, an electrical heater is placed between two thermocouples or thermistors located some distance apart along the axis of a vessel. The temperature difference between the upstream and downstream sensors is a measure of blood velocity. A device of this type is called a *thermostromister*.

Blood Flow Determination by Radiographic Methods

Generally, blood is not visible on x-ray images, but after injecting a contrast medium into a blood vessel, which may be an iodinated organic compound, the blood circulation pattern can be made locally visible. After taking the x-ray image, the progress of the contrast medium can be followed, obstructions can be detected, and blood flow in certain vessels can be estimated. This technique is known as *cine* or *video angiography*, and it can be used to assess the extent of damage after a stroke or heart attack.

The second method involves the injection of a radioactive isotope into the blood circulation that allows the detection of vascular obstructions with an imaging device for nuclear radiation such as a scanner or gamma camera.

Measurement by Indicator Dilution Methods

The indicator or dye-dilution method is a method that measures real blood flow, not blood velocity. The substance used as an indicator should mix readily with blood, and its concentration in the blood should be easily determined after mixing. The substance must be stable, should not be retained by the body, and should have no toxic side effects.

An indocyanine dye, cardio green, in isotonic solution was used previously as an indicator. Its concentration was determined by measuring its light absorption with a densitometer or colorimeter. Radioactive isotopes also were used for the same purpose.

Nowadays, isotonic saline is used, and it is injected at a temperature lower than body temperature. The concentration of the saline after mixing with the blood is determined with a sensitive thermistor thermometer.

Plethysmography

These types of instruments are used to measure volume changes of blood in any part of the body and are based on the pulsations of blood occurring with each heartbeat. These measurements are very useful for the diagnosis of arterial obstructions, as well as for pulse-wave velocity measurements.

Plethysmographs are designed to measure constant pressure or constant volume within a chamber. A true plethysmograph actually responds to changes in volume. This type of instrument consists of a rigid cup or chamber placed over a limb or digit in which volume changes are to be measured. The cup is tightly sealed to the chamber to be measured so that any changes in volume in the limb or digit reflect pressure changes inside the chamber. Either fluid or air can be used to fill the chamber.

In plethysmographs, some form of pressure or displacement transducer is included to respond to pressure changes within the chamber and to provide a signal that can be calibrated to represent the volume of the limb or digit.

7.3.1 Measurement of Heart Sounds

For this purpose, the widely used instrument is the *stethoscope*, which is a device that carries sound energy from the chest of the patient to the ear of the physician via a column of air. However, only a small portion of the energy in heart sounds is in the audible frequency range. Moreover, with the dawning of the age of electronics, electronic stethoscopes are now available commercially, although they have never found favor with physicians.

A graphic record of heart sounds is a *phonocardiogram*, and the instrument used for recording it is *phonocardiograph.* The basic transducer for a phonocardiogram is a microphone having the necessary frequency response, generally ranging from below 5 Hz to above 1000 Hz. The readout of a phonocardiograph is either a high-frequency chart recorder or an oscilloscope.

Because most pen galvanometer recorders have an upper frequency limitation of around 100 or 200 Hz, photographic or light galvanometer recorders are required for faithful recording of heart sounds. Although normal heart sounds fall well within the frequency range of pen recorders, the high-frequency murmurs that are often important in diagnosis require the greater response of the photographic device.

Microphones for phonocardiograms are designed to be placed on the chest, over the heart. However, heart sounds are sometimes measured from other vantage points. For this purpose, special microphone transducers are placed at the tips of catheters to pick up heart sounds from within the chambers of the heart or from the major blood vessels near the heart. Moreover, there are *vibrocardiographs* and *apex cardiographs* that measure the vibrocardiogram and apex cardiogram, respectively, via the use of microphones.

7.3.2 Instrumentation for Measurements of the Nervous System

EEG measurements quantify the electrical activity of the brain and are obtained from electrodes placed on the surface of the scalp; the

waveforms represent a gross type of summation of potentials originating from an extremely large number of neurones in the vicinity of the electrodes.

The EEG potentials are actually the electrical patterns obtained from the scalp and are actually the result of graded potentials on the dendrites of neurons in the cerebral cortex and other parts of the brain as they are influenced by the firing of other neurons that impinge on those dendrites. EEG potentials are random-appearing waveforms with peak-to-peak amplitudes ranging from less than 10 μV to over 100 μV. Generally, surface or subdermal needle electrodes are used for the EEG measurements (Fig. 7.3).[11]

The ground reference electrode is a metal clip on the earlobe. A suitable electrolyte paste or jelly is used in conjunction with the electrodes to enhance coupling of the ionic potentials to the input of the measuring device. To reduce interference and minimize the effect of electrode movement, the resistance of the path through the scalp between electrodes must be kept as low as possible.

Placement of electrodes on the scalp is dictated by the requirements of the measurement to be made. In clinical practice, a standard pattern, called the *10–20 electrode placement system,* is used. This system was devised by the committees of the International Federation of Societies for Electroencephalography.

Figure 7.3 Electroencephalogram (EEG) measurements.

In addition to the electrodes, EEG measurement requires a readout or recording device. Most clinical EEGs have the capability of simultaneously recording EEG signals from several regions of the brain. For each signal, a complete channel of instrumentation is required, and EEGs having as many as 16 channels are available.

The readout in a clinical EEG is a multichannel pen recorder with a pen for each channel. The standard chart speed is 30 mm/s, and some devices provide a speed of 60 mm/s for higher-frequency signals. Some have a third speed of 15 mm/s to conserve paper during setup time.

For the purposes of research, EEG signals are separated into their conventional frequency bands by means of bandpass filters, and the output signals of the individual filters are recorded separately. A special form of EEG is the recording of *evoked potentials* from various parts of the nervous system. This EEG technique measures some form of sensory stimulus, such as a flash of a light or an audible click.

7.3.3 Electromyographic Measurements

Just like neurons, skeletal muscle fibers also generate action potentials when excited by motor neurons via the motor end plates, and measurement of these action potentials can be done either directly from the muscle or indirectly from the surface of the body, constituting an EMG. Although action potentials of individual muscle fibers are recorded, the electrical activity of the muscle is of primary importance.

The signal is the summation of all the action potentials within the range of electrodes, each weighted by its distance from the electrodes. A correlation exists between the overall amount of EMG activity for the whole muscle and the strength of muscular contraction. A linear relationship also exists between the voltage-time integral of the EMG signal and the isometric voluntary tension in a muscle. Special conditions such as tremor and fatigue are associated with characteristic EMG patterns.

Surface, needle, and fine-wire electrodes are used for different types of EMG measurements. The amplifiers for EMG measurements have high gain, high input impedance, and a differential input with good common-mode reflection.

A typical electromyograph has an oscilloscope readout instead of a graphic pen recorder. Most electromyographs include an audio amplifier and a loudspeaker in addition to the oscilloscope display to permit the operator to hear the "crackling" sounds of the EMG. This audio presentation is helpful for the placement of needle or wire electrodes into a muscle.

An important feature found in modern electromyographs is a built-in stimulator for nerve conduction time or nerve velocity measurements.

EMG signals can be quantified in several ways. A simple method is the measurement of amplitude alone. Another method is to count the number of spikes or, in some cases, zero crossing that occurs over a given time interval. A modification of this method is to count the number of times a given amplitude threshold is exceeded. The most useful method of quantifying the EMG uses the time integral of the EMG waveform. For the purpose of research, the EMG signal is rectified and filtered to produce a voltage that follows the envelope or contour of the EMG.

7.4 Instrumentation for Psychophysiologic Measurements

In order to study the behavior of a person, the electrical signals of the brain and nervous system that control behavior are measured. And these voltages can be easily recorded on the EEG. However, EEG recording can only identify events that involve larger areas of the brain such as epileptic seizures. Mental disorders generally cannot be diagnosed via EEG.

Many functions of the body, such as the blood pressure, heart rate, perspiration, and salivation, are controlled by the autonomic nervous system; changes in body language can be easily observed and recorded, practical application of which is the polygraph, commonly known as the *lie detector*, a device that records the body functions of a person that are likely to show changes when questions asked by the interrogator cause anxiety in the tested person.[12,13]

In psychophysiologic studies, instruments have been devised for the measurement of blood pressure, heart rate, respiration rate, and so on. For measuring the variations in perspiration, a technique has been devised in which if a sharp point is touched, the resistance of the skin shows a characteristic decrease called the *galvanic skin response* (GSR). The baseline value of the skin resistance is called the *basal skin resistance* (BSR). The GSR is believed to be caused by the activity of sweat glands. The GSR is measured readily at the palms of the hand, where there is the highest concentration of sweat glands. There are certain devices that allow the simultaneous measurement of both BSR and GSR.

Motor responses are under voluntary control, and nowadays, numerous devices are available commercially to measure motor responses and to study the influence of factors such as fatigue and stress or the effects of drugs. Some of the notable devices are the manual dexterity tests that consist of a number of small objects that the subject is required to assemble in a certain way while the time required for completion of the task is measured. In steadiness testers, a metal stylus is moved through touching the metal walls. An error closes the contact between the wall and stylus and advances an electromechanical counter.

In a pursuit rotor, a light spot moves with adjustable speed along a circular or star-shaped pattern on the top surface of the tester. An indicator and timer automatically measure the percentage of time during which the subject is "in target" during a certain test interval. The performance of certain muscles or muscle groups can be measured with various dynamometers that measure the force exerted either mechanically or with an electric transducer.

Human senses provide information inputs required by humans to orient themselves in their environment and to protect themselves from danger. Methods and instruments have been developed to measure the performance of sense organs, study their functioning, and detect impairment. For example, temperature sense can be measured with metal objects. To study optical perception, a device with a spot of controllable brightness and size is viewed against a background whose brightness also can be varied. Variations in the size and brightness of the spot and the brightness of the background are controlled independently. Another device for studying visual perception is the *tachistoscope,* in which the display of an illuminated card is presented to the viewer by means of a semitransparent mirror or a slide projector.[14,15]

Acuity of hearing can be measured with the help of an instrument called an *audiometer.* Here, the sound intensity in an earphone is gradually increased until the sound is perceived by the subject. The hearing in the other ear during this measurement is often masked by presenting a neutral stimulus to that ear. The threshold of hearing is determined at a number of frequencies. This process is automated in the *Bekesy audiometer.*

7.5 Instrumentation for the Experimental Analysis of Behavior

In order to describe and analyze behavior, special testing devices have been developed to obtain a numerical rating (e.g., a pursuit rotor). Many basic behavioral experiments are performed with animal subjects in a neutral environment in a soundproof enclosure called *Skinner box* so as to isolate the animal from uncontrolled environmental stimuli. The Skinner box is provided with a response bar and a stimulus light.

Behavior evidence by organisms to interact with and modify their environment is called *instrumental* or *operant behavior*. Insight into the behavior mechanism obtained in animal experiments can be extrapolated to human behavior. In a form of treatment called *behavior therapy*, behavioral and emotional problems are treated with electronic equipment. In bed wetting, the *Mower sheet,* which is actually a moisture sensor placed beneath the bed sheet, activates an acoustical alarm and turns on a light to awaken the subject when the presence of moisture is first detected.[14,15]

7.6 Respiratory Therapy Equipment

Important respiratory therapy equipment includes inhalators, respirators, ventilators, humidifiers, and nebulizers.

Inhalators

This is a device used to supply oxygen or other therapeutic gases to a patient who can breathe spontaneously without assistance. Devices for administering oxygen include nasal cannulas mouthpieces and face masks. The oxygen concentration presented to the patient is controlled by adjusting the flow of gas into the mask.

Ventilators and Respirators

These are used continuously or intermittently to improve ventilation of the lungs and to supply humidity or aerosol medications to the pulmonary tree. Commonly used respirators are classified as assist-control devices and can be operated in three different modes: (1) the assist mode, in which the inspiration is triggered by the patient, (2) the control mode, in which breathing is controlled by the timer set to provide the desired respiration rate, and (3) the assist-control mode, in which the apparatus is normally triggered by the patient's attempts to breathe, as in the assist mode.

Ventilators in clinical use are categorized into three basic types:

1. If the delivered gas reaches a predetermined pressure in the proximal or upper airways, the ventilator is said to be *pressure cycled.*
2. *Volume-cycled* ventilators operate so as to deliver a predetermined volume of gas to the patient.
3. If the air or oxygen has been applied for a predetermined period of time, the ventilator is said to be *time cycled*.[16–18]

Humidifiers, Nebulizers, and Aspirators

In order to prevent damage to a patient's lungs, the air or oxygen applied during respiratory therapy must be humidified by means of a *humidifier.* When the therapy requires that water or some type of medication is suspended in the inspired air as an aerosol, a device called a *nebulizer* is used. *Aspirators* or other types of suction apparatuses are used to remove mucous and other fluids from the airways.

7.7 Pacemakers and Defibrillators

A *pacemaker* is a device for generating artificial pacing impulses and then delivering them to the heart. Pacemakers may be internal or external.

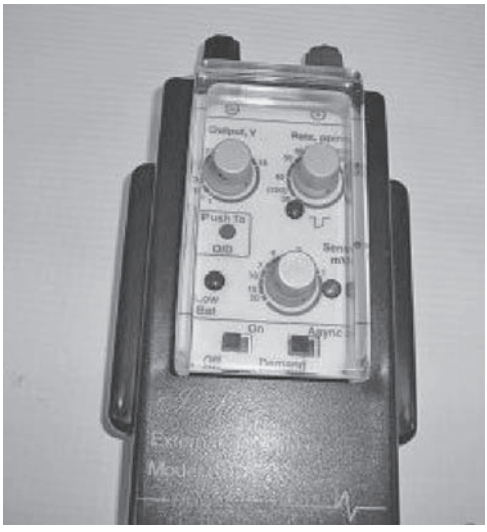


Figure 7.4 External pacemaker.

Internal pacemakers are permanently implanted in patients whose sinoatrial (SA) nodes have failed to function properly or in patients who suffer from permanent heart block because of a heart attack. They are implanted with the pulse generator placed in a surgically formed pocket below the right or left clavicle in the left subcostal area and for women beneath the left or right major pectoral muscle. Internal leads connect to electrodes that directly contact the inside of the right ventricle or the surface of the myocardium.

External pacemakers are used for patients with temporary heart irregularities, including heart block. They are also used for temporary management of certain arrhythmias. They include pulse generators located outside the body and normally connected through wires introduced into the right ventricle via a cardiac catheter. The pulse generator may be strapped to the lower arm of a patient who is confined to bed or worn at the midsection of an ambulatory patient (Fig. 7.4).[19]

Defibrillators

The heart performs its important pumping function through the synchronized action of heart muscle fibers. A condition in which this necessary synchronism is lost is known as *fibrillation*. Fibrillation of atrial muscles is called *atrial fibrillation*, and fibrillation of the ventricles is called *ventricular fibrillation.*

The most successful method of defibrillation is the application of electric shock to the area of the heart. Sufficient current is used to stimulate all the musculature of the heart simultaneously for a brief

period and then is released; all the heart muscle fibers enter their refractory period, after which normal heart action may resume. This application of an electric shock to resynchronize the heart is sometimes called *countershock*. If the patient does not respond, the burst is repeated until defibrillation occurs. This method of countershock is known as the *ac defibrillation*.

In a new method known as *dc defibrillation*, a capacitor is charged to a high dc voltage and then rapidly discharged through electrodes across the chest of the patient. It has been found that *dc defibrillation* is more successful than the ac method in correcting ventricular fibrillation, and it successfully corrects atrial fibrillation and other types of arrhythmias.

7.8 Instrumentation for the Medical Use of Radioisotopes

For diagnostic purpose, radioisotopes are introduced into the body. X-ray examinations are of immense importance in the field of medical science and technology. *Scintillation detectors* are used for determining the concentration of gamma-emitting radioisotopes in medical applications. For activity determinations inside the body, a collimated detector is used. A radioisotope scanner is a detector that is moved over the affected area to be examined in a zigzag fashion.

The ionizing effect of the x-rays is used for the treatment of certain diseases, especially tumors. Very hard x-rays generated from very high voltages are used in the therapy of deep-seated tumors. In dermatology, very soft x-rays, called *Grenz rays*, are used for skin treatments. For concentrating the radiation at the site of a tumor, very high voltage linear accelerators or betatrons can be used.[19,20]

7.9 Ophthalmic and ENT Instruments

It is through the sense organs of our body that we communicate with each other and also maintain our relationship with surrounding environments. Hence an attempt has been made to develop instruments for the study of the eye, ear, nose, and tongue.

Instrumentation in Ophthalmology

The three major instruments needed for ophthalmologic purposes are as follows:

1. *Opthalmoscope*. Used to study the interior of the eye.
2. *Retinoscope.* Used to measure the focusing power of the eye.
3. *Keratometer*. Used to measure the curvature of the cornea.

There are other instruments, such as the *tonometer*, which measures the pressure in the eye, and the *lensometer*, which is used to determine the prescription of an unknown lens.

Ophthalmoscope This instrument was invented in 1851 by Helmholtz. The basic principle of the ophthalmoscope is based on the fact that when a bright light is projected into a subject's eye and the returning light from the subject's retina is positioned so that it can be focused by the examiner, the lens system of the patient's eye acts as a built-in magnifier (Fig. 7.5). In addition to eye problems, an expert can detect a brain tumor with the help of an ophthalmoscope. A brain tumor cause a noticeable change in the interior of the eye owing to increased pressure inside the skull.

Retinoscope This instrument is used to determine the prescription of a corrective lens without the patient's active participation, although the eye has to be open and in a position suitable for examination. It also helps to check the prescription to determine which was better, the former or the latter. When a streak of light from a retinoscope dilutes eye, the light beam is reflected from the retina, acting as a light source for the operator. The retina's function in retinoscopy is the reverse of eye's far point. If focused at the retina of a relaxed eye, light from the retina of the relaxed eye will produce a focused image at the far point. By viewing the patient's eye through the retinoscope, the operator adds a lens in front of the patient's eye so as to cause the image from the patient's retina to be focused at the operator's own eye.

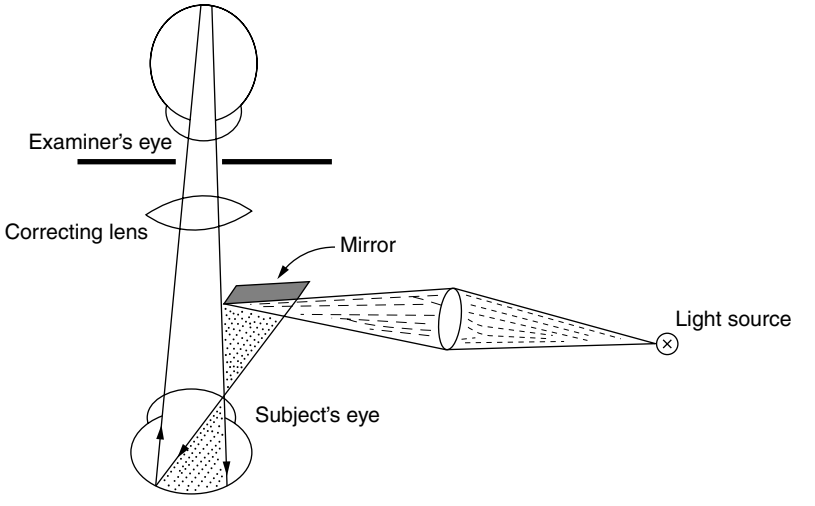


Figure 7.5 Ophthalmoscope.

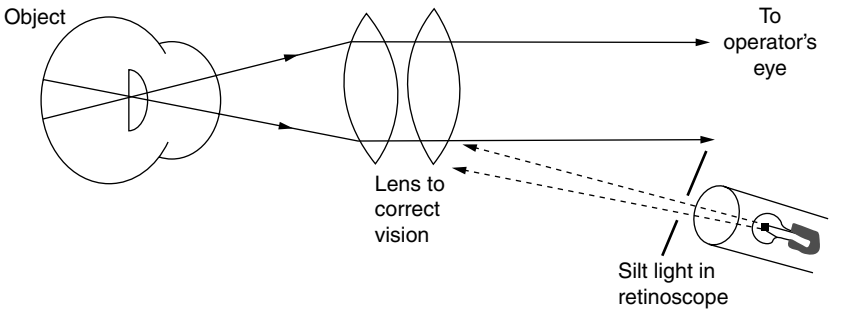


Figure 7.6 Retinoscopy.

To determine the prescription needed to correct the patient's eye, the operator must change the lens power of these added lenses by the dioptic power needed to focus the same eye at infinity (Fig. 7.6).

Keratometer This instrument is needed to fit contact lenses. It measures the curvature of the cornea. Keeping the patient's head fixed, the keratometer produces a lighted object that is reflected from the cornea. In keratometry, the cornea acts as a convex mirror. The reflected image is located at the focal plane at a distance $r/2$ behind the surface of the cornea. The keratometer produces a lighted object that is reflected from the cornea, keeping the patient's head in a fixed position. The instrument is placed at a known distance from the cornea, and the operator adjusts the focus control. Part of the reflected image passes through a prism that causes a second image to be seen by the operator. By adjusting the size of the reflected image so as to produce a coincidence of marker lines in the two images, the position of the prism after this adjustment is indicated in diopters of focusing power of the cornea (Fig. 7.7). The average value is 44 D, which corresponds to a cornea with a radius of curvature of 7.7 mm.

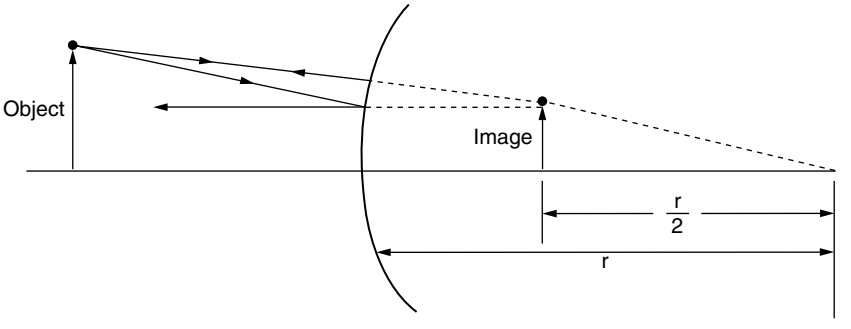


Figure 7.7 Keratometer.

Instruments to measure intraocular pressure are needed in treatment of the disease condition known as *glaucoma*, in which intraocular pressure increases abnormally and may lead to blindness if not treated quickly. The instrument used for this purpose is a *Schiotz tonometer*. This technique allows the tonometer to rest on the anesthetized cornea, the central plunger causing a slight depression in the cornea. The position of the plunger indicates the internal pressure of the eye on the scale. The force on the plunger can be varied by adding a variety of weights.

The pressure measured by a tonometer is the original pressure plus the increase resulting from the instrument. In order to remove the effect of the rigidity of the eye, another measurement is taken with a heavier weight or with a *Goldman tonometer.* The two readings, as well as table values, help to determine the original pressure and the rigidity of the eye.

Nowadays, ultrasound diagnostic techniques are also being used in ophthalmologic examinations. Tumors, foreign bodies, and detachment of the retina can be diagnosed by ultrasound. Applications of ultrasound make it possible to measure the distances to the lens, the distance to the retina, and the thickness of the vitreous humor curvature of the cornea; the prescription of corrective glasses is the other important quantity that also can be measured.

Instrumentation for Testing the Ear

The basic instrument for this purpose is the *audiometer*, which is used to determine hearing loss in an individual. Audiometers cause no discomfort to the patient and are capable of quick, precise measurements. The instrument is based on the principle of presenting specific pure-tone signals to the subject and determining the intensity at which the subject can barely hear these signals. The instrument has an oscillator and a pair of headphones and usually is calibrated in terms of frequency and acoustic output. The two main types of audiometers are pure-tone audiometers and speech audiometers, which are grouped based on the stimuli they provide to elicit audio responses.

Audiometers with a single pure-tone generator are provided with two channels. The first channel is for pure-tone or speech output, and the second channel has nominal masking. Pure-tone audiometers generate test tones in octave steps from 125 to 8000 Hz, with signal intensity ranging from –10 to +100 dB.

The *Bekesy audiometer system* represents the absolute threshold values at all frequencies in the range tested. The instrument generates a pure-tone signal that is made available to the subject through an air-conduction earphone. The subject presses a switch when the tone is heard and releases the switch when it is not heard. The pen connected to the attenuator produces a continuous record of the patient's intensity adjustments on the audiogram chart to produce a graphic presentation.[21]

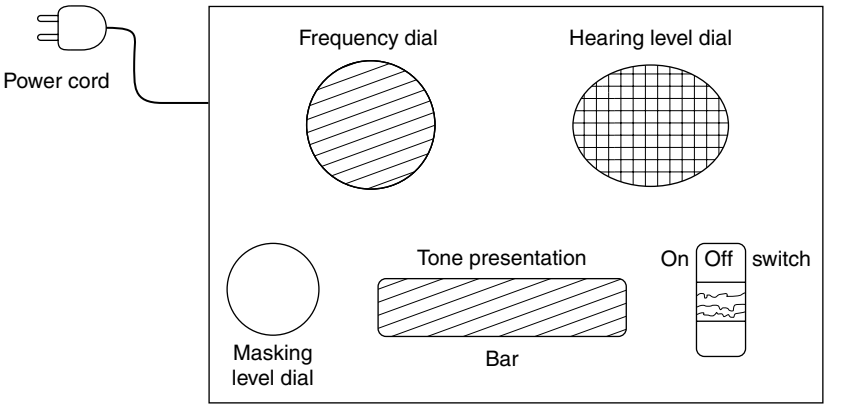


Figure 7.8 A simple audiometer.

Audiometers must be calibrated accurately to ensure that the instrument produces a pure tone and speech. The parameters commonly checked are frequency and intensity (Fig. 7.8).

7.10 Ultrasonography

This is a technique that uses ultrasonic energy to study the internal organs of the body (Fig. 7.9).

Advantages of Ultrasonography

1. The technique is not harmful to human tissue.
2. It is safe for pregnant women even with frequent use.
3. It can detect materials that are nonradiopaque, thus making angiographic dyes unnecessary.

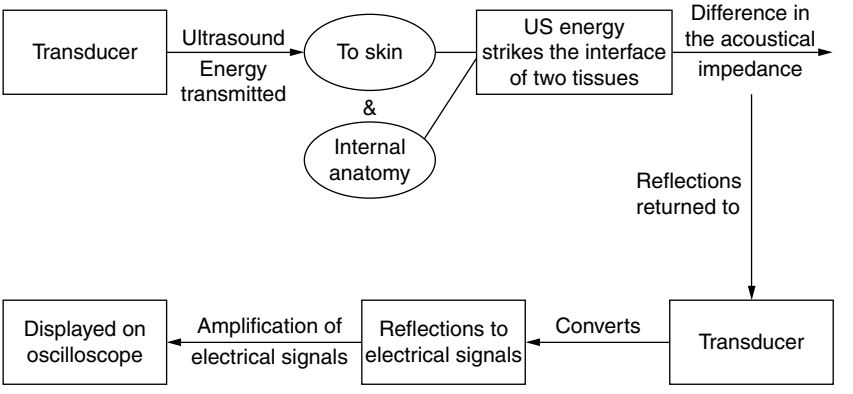


Figure 7.9 Block diagram of the ultrasonographic technique.

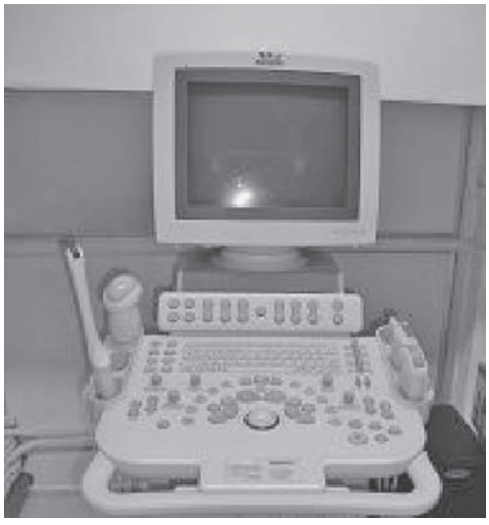


Figure 7.10 Ultrasonography.

4. It is easy to operate and is becoming a popular diagnostic technique (Fig. 7.10).

In ultrasonography, the main function is to measure the distances between interfaces that separate body structures by timing the echoes produced by these interfaces. This technique is called the *pulse-echo method*.

The commonly used display modes are (1) A-scan mode, (2) B-scan mode, and (3) T-M mode.

A-Scan Mode

A-scan does not produce an image in the usual sense of word. In this mode, pulses are transmitted into an object on a periodic basis, and then an oscilloscope is synchronized to the transmitted bursts. The echoes will appear on the screen. Since an A-scan trace gives only one-dimensional spatial information, it is not traditionally thought of as an image.

The A-scan ultrasonography instrument displays the amplified echo signal on the vertical channel of the oscilloscope, with the horizontal channel being deflected by a conventional sweep generator. The sweep generator is triggered from the impulse signal, and the time delay between the beginning of the sweep and the echo appearing on the CRT screen is proportional to tissue depth. A practical application of A-scan ultrasonography is echoencephalography, which detects brain midline position and the possible displacement of the midline owing to an abnormal space-occupying mass within one side of the skull, such as a tumor.

B-Scan Mode

Recently, this has become a versatile diagnostic tool because of developments during the last two decades. The picture that results from B-scanning is a two-dimensional cross-sectional presentation of part of the subject. The scanning system provides signals that are proportional to the position and direction of the probe. The principle of B-scan image formation is based on the fact that a transducer placed in each of the six positions gives rise to a series of intensified areas on the CRT scan view, where only six brightness dots would appear, but by slowly turning the transducer over the object, repetitive echo patterns at a rate of 500 Hz are observed, and there is a continuous display instead of several dots.

The transducer is pulsed repeatedly at 500 Hz while collecting the echo signals simultaneously. With movement, the image appears as a steady pattern, giving details of reflective objects inside the region of observation.

Time–Motion (T-M) Mode Ultrasonography

If a slight modification of the A-scan system is done by removing the echo signal from the vertical channel (Y) and connecting it to the identifying channel (Z), an echo response appears as an intensified spot on the screen. Any motion in the echo response would appear as horizontal motion of the spot between successive sweeps of the horizontal sweep generator; no vertical signal is involved. In time-motion mode ultrasonography, the hypothetical A-scan described earlier is used with a slow sweep generator, moving the display vertically. Any motion in the echo response is displayed in real time by the slow vertical sweep.[22–24]

In a T-M-mode ultrasonogram, the reflections are shown as a time course indicating the movements of cardiac chamber walls and so on. The machine has a cursor called an *M-cursor.* By positioning this cursor at the required point, an M-mode tracing of the heart is obtained at that point. This mode is useful for measuring chamber dimensions during systole and diastole.

Ultrasonography, especially the B-scan machines, by using a blend of technology and the latest microprocessors, have become powerful diagnostic aids and thus facilitate everyday work and present views very clearly and visibly. Probes have been specially designed for gynecology or obstetrics, urology, and cardiology, and thus previous errors in images have been eliminated considerably. These probes give clear internal details at variable distances. In orthopediatrics and pediatrics, probes with high frequencies and a wide sector scan are used. For diagnosis within the pelvic cavity, specially designed transvaginal probes are used.

Development of ultrasonography provides potential advantages in cases where conventional x-ray techniques are not feasible, for example, in study of a fetus in vitro because fetal development is

impaired by x-rays. X-ray resolution techniques are not suitable for eye scans because the x-ray plate cannot be located directly behind the region of concern and it is here that ultrasonography finds important uses.

7.11 Computed Tomographic Scanning (CT Scanning)

The *CT scan technique* was invented by Sir Godfrey Hounsfield, for which he was awarded the Nobel Prize in 1979. CT scanning uses x-rays but employs a computer simultaneously to reconstruct the image instead of recording it directly on photographic film. CT scanning is used in the diagnosis of diseases of the spinal cord, brain, chest, and abdomen.

This technique is highly useful in detecting tumors and monitoring the extent of their spread to neighboring tissues and organs. It helps to determine the feasibility of operative treatment, and it also assesses the results of treatment.

CT scanning is done mainly by scanning only thin, well-defined volumes of interest, which serves to minimize the superimposition effects and the used linear detectors with computer-based window functions, and it also minimizes scatter by collimating down to relatively thin volumes.

For the purpose of CT scanning the *fan angle* used lies between 30 and 50 degrees, with a provision for tilting the patient's table up to 15 degrees, and there is a height-adjustment provision for centering by which 84 to 103 cm is available for the head and 93 to 104 cm for body parts.

The attenuation coefficient μ is the ratio of x-ray intensity obtained on the detector with the body inserted and without it. Here, resolution of the image depends on the signal-to-noise ratio, which also depends on the radiation and the number of steps. A greater number of projections is used for resolution of the high-contrast areas, whereas low-contrast tissue matter is resolved by adjusting the x-irradiation to optimum.

The image intensity can be computed from the integrated value of the absorption by x-ray on any path.

$$\int_{p}^{q} \exp[-f(x,y)]dl$$

This is the amount of attenuation when the density at a point inside is denoted by $f(x, y)$. The integral is denoted as $g(r, O)$. This is the observed signal at an angle O. From the values of g at any r and O, it is necessary to find the function $f(x, y)$.

For the purpose of CT scanning, different techniques are adopted, and the most important ones are as follows: backprojection technique, the algebraic reconstruction technique, and the Fourier transform method.

Backprojection Technique

This is one of the most frequently used techniques. The image profile at any line is resolved by the [cosθ, sinθ] transformation to give the positions of the pixels w.r.t. the basic x, y axes. The measured data are projected back over the area at the same angle from which they were taken. A filter function is run before the backprojection.

Algebraic Reconstruction Technique

This algorithm is used to determine the values of the density coefficients of the pixels by treating the pixels of a slice as a matrix of values and by using the integral equations of the absorptions on several angles to invert the matrix.

Fourier Transform Method

In this technique, in order to provide negative value to each point in the projection (instead of zero), the shadow function with the filter is convolved and then transformed into the Fourier space of a 2D Fourier transform. Next, after summation, inverse transform provides the picture. The criteria of the picture are decided on the convolution function or kernel.

From the images obtained on CT scan, the following analysis is done: sagittal and coronal reconstruction from the axial images, histogram and image profile, region-of-interest (ROI) calculation, and 3D reconstruction. This reconstructed slice image is either printed or viewed on a CT monitor. With the help of present-day computers, these images can be stored and distributed on a CD for recording and scanning purposes.

7.12 Positron-Emission Tomographic (PET) Scanning

PET scanning is also a computerized imaging method and differs from CT scanning by providing quantitative information on various physiologic processes of the body in contrast to CT scanning, which provides static anatomic images. Positron-emitting radioisotopes such as ^{11}C, ^{13}N, ^{15}O, and ^{18}F are generated by using a cyclotron, and these radioactive atoms are incorporated by chemical methods into biomolecules such as glucose, amino acids, CO_2, and NH_3. The patient under test is either injected with or asked to inhale a small amount of a positron-emitting compound. The substance is distributed in the body. The 3D distribution of labeled radioisotopic atoms is recorded by powerful PET cameras, and the images are reconstructed by computer. Quantitative interpretation of image also can be accomplished by mathematical models.

Through the use of PET technique, color-processing centers in the visual cortex of human beings are being studied by U.K. scientists. The technique also effectively measures regional cerebral volume, blood flow, metabolic rates for glucose and oxygen, and so on.

7.13 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) actually represents the spatial distribution of NMR signal intensity or other NMR parameters in different parts of the human body when they are placed in a deliberately nonuniform magnetic field. Because of the nonuniformity in the magnetic field, the gradient in the magnetic field labels different parts of the specimen with different field strengths, thereby enabling the structure and internal processes of the specimen to be derived and displayed.

MRI systems are characterized by the strength of the magnetic field, and the imaging procedures are performed in field strengths ranging from 0.3 to 1.5 T. However, imaging outside this range is also possible.

It is the strength of the magnetic field that determines the tissue resonant frequency. This is the frequency that is receptive to the radiofrequency pulses applied to the tissue and is the frequency of the radiofrequency signals emitted during the imaging process. In an MRI scan, a number of pulse sequences are used for obtaining the image, such as partial saturation, inversion recovery, spin echo, fast low angle shot (FLASH), gradient-recalled acquisition in the steady state (GRASS), and others. The total scan time can be decreased by multislice and multiecho techniques.

By using pulse sequences, the NMR signal generated cannot be translated directly into the image. Here, frequency representation is converted to location representation through use of a digital computer with sufficient memory and storage space to perform the mathematics of this conversion. NMR signals yield one-dimensional information in the presence of the magnetic field gradient, and MRI employs different schemes for using this one-dimensional measurement to sample the two-dimensional image plane. Of the two techniques, projection reconstruction imaging and 2D Fourier transforms (FT) imaging (2D FT), the second one is the method of choice because of fast computational facility. After the entire frequency representation has been sampled by repeated cycles of the 2D FT technique using different phase-encoding gradient strengths, the sample frequency representation is converted to the image using the 2D FT method.

In most imaging modalities, such as x-ray and CT scanning, injectable contrast material or dyes are used, but the contrast used in MRI is fundamentally different. MRI contrast works by altering the local magnetic field in the tissue being examined. Normal and abnormal tissues respond differently to the slight alteration by giving different signals.

During MRI scanning, the machine produces a tremendous continual rapid hammering noise, and patients are provided with earplugs or stereo headphones to cope with the noise. The tremendous

noise is due to the rising electric current in the wires of the gradient magnets being opposed by the main magnetic field. The stronger the field, louder is the noise.

While undergoing MRI scans, patients are required to hold very still for extended time periods. The examination time ranges from 20 to 90 minutes or more. Just a slight movement of the part being scanned may give rise to very distorted images.

The MRI suite can become a very dangerous place if strict restrictions are not followed. Metal objects can become dangerous projectiles if they are taken into the scan room. Paperclips, pens, stethoscopes, and so on can be pulled out of pockets, and they fly at tremendous speed toward the opening of the magnet, where the patient is placed, posing a threat to everyone in the room. Credit cards and bank cards are destroyed by MRI machines, and machines have pulled mop buckets, vacuum cleaners, patient stretchers, heart monitors, and oxygen tanks. The largest known object ever pulled was a fully loaded pallet jack.

It is very risky to scan a patient with an implant or metallic object inside the body. MRI scanning of a pregnant woman is also not advised, and in the case of necessity, it is done by consultation between the MRI radiologist and the patient's obstetrician. However, pregnant MRI technologists can work in the department. In most cases, however, they are kept out of the actual scan room during pregnancy.[24]

The benefits of MRI are almost limitless. This scanning technique is ideal for diagnosing strokes at their earliest stages, multiple sclerosis, tumors in the pituitary gland and brain, tendonitis, and infections in the brain, spine, or joints, and for visualizing ligaments of the body, shoulder injuries, bone tumors, cysts, and herniated disks. Another advantage of MRI is the fact that no ionizing radiation is used; hence there is a low incidence of side effects. Moreover, MRI has the ability to image in any plane.

However, some persons cannot be scanned safely by MRI, especially those who are provided with pacemakers or are too big to be scanned. In addition, many claustrophobic people find it impossible to enter an MRI machine. Moreover, MRI machines are extremely expensive, and the test is also very costly. Nevertheless, the limitless benefits of MRI and its extensive applications in the field of medical science and research far outweigh its few drawbacks.

References

1. Best CM, Taylor NB. *The Living Body: A Text in Human Physiology*. New York: Holt, Rinehart and Winston, 1971.
2. Bronzini JD. *Technology for Patient Care*. St. Louis: Mosby.
3. Caceres CA, ed. *The Practice of Clinical Engineering*. New York: Academic Press, 1977.
4. Brown JHU, Jacobs, Stark L. *Biomedical Engineering*. Philadelphia: Davis, 1971.

5. Strong P. *Biophysical Measurements*. Beaverton, OR: Tektronix, 1970.
6. Burch GE, Winsor T. *A Primer of Electrocardiography*. Philadelphia: Lea & Febiger, 1960.
7. Gyton AC. *Textbook of Medical Physiology*. Philadelphia: Saunders, 1971.
8. *ECG Measurements, Application Note AN711*. Waltham, MA: Hewlett-Packard, Medical Electronics Division, 1970.
9. *Pressure Measurements, Application Note A 710.* Waltham, MA: Hewlett-Packard, Medical Electronics Division, 1969.
10. Franklin DL. "Techniques for measurement of blood flow through intact vessels," *Med Electron Bio Eng*. 3:27–37, 1965.
11. Brazier MAB. *The Electrical Activity of the Nervous System*, 3rd ed. Baltimore: Williams & Wilkins, 1968.
12. Thompson RF. *Foundations of Physiological Psychology*. New York: Harper & Row, 1967.
13. Grings WW. *Laboratory Instrumentation in Psychology*. Palo Alto, CA: National Press, 1954.
14. Stebbins WC. "Behavioral techniques." In Rushmer BF, ed., *Methods in Medical Research*, Vol. 11. Chicago: Year Book Medical Publishers, 1966.
15. Venables PH, Martin I. *A Manual of Psychophysiological Methods.* New York: Wiley, 1967.
16. Belinkoff S. *Introduction to Inhalation Therapy*. Boston: Little, Brown, 1969.
17. Cherniak RM, Cherniak L. *Respiration in Health and Disease*. Philadelphia: Saunders, 1965.
18. Filley GF. *Pulmonary Insufficiency and Respiratory Failure*. Philadelphia: Lea & Febiger, 1968.
19. *Cardiac Pace Making*. Somerville, NJ: Hoechst Pharmaceuticals, 1974.
20. Selman J. *The Fundamentals of X-Ray and Radium Physics*. Springfield, IL: Charles C Thomas Publisher, 1965.
21. Bekesy GV. "A new audiometer." *Acta Otolaryngol.* 34:411–422, 1947.
22. Haga E, ed. *Computer Techniques in Biomedicine and Medicine*. Philadelphia: Querbach Publishers, 1973.
23. Stacy RW, Waxman BD, eds. *Computers in Biomedical Research*, Vols. I, II, III, and IV. New York: Academic Press, 1965–1974.
24. Jurgen RK. "Electronics in medicine." *IEEE Spectrum*, January 1978, 68–72.

This page intentionally left blank

Index

Note: Page numbers referencing figures are followed by an "f"; page numbers referencing tables are followed by a "t".

B

C

D

E

F

G

H

I

M

N

O

P

Q

R

S

T