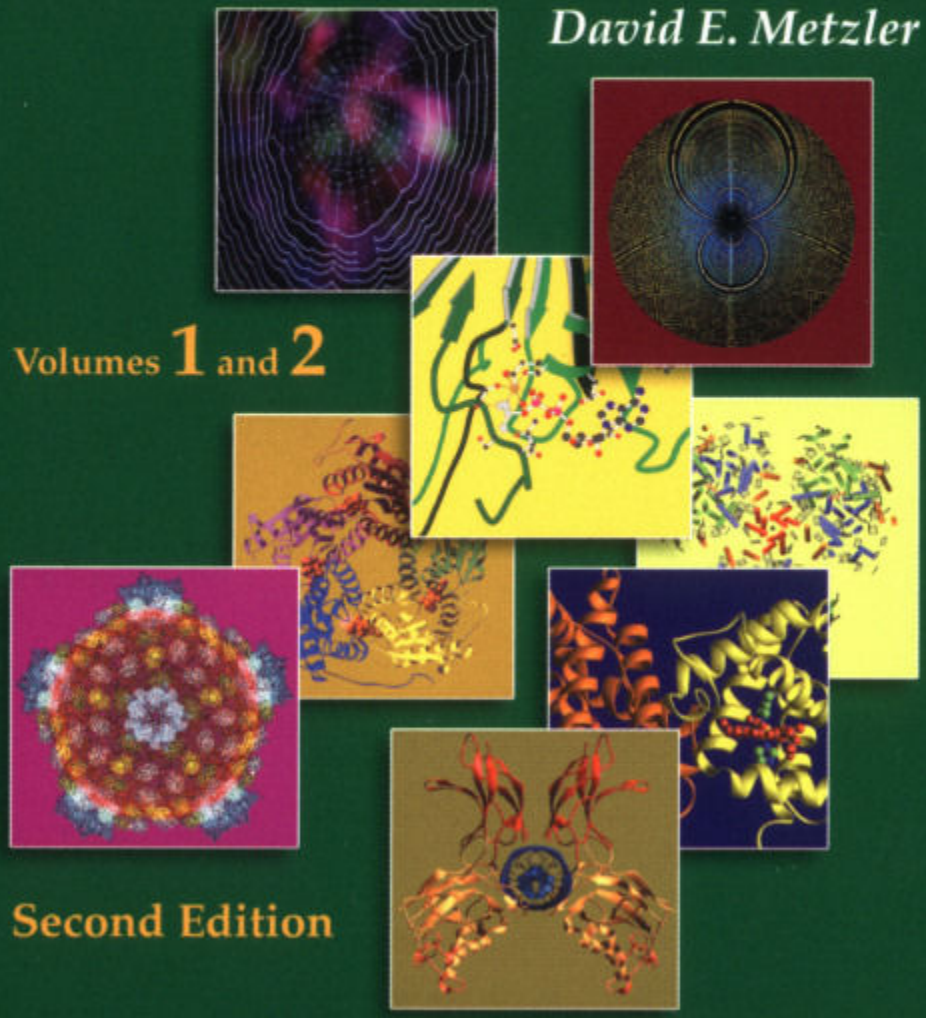
ELSEVIER
ACADEMIC
PRESS

# BIOCHEMISTRY

## The Chemical Reactions of Living Cells

*David E. Metzler*



Volumes 1 and 2

Second Edition

# Table of Contents
## *Volume 1*


Chapter 1. The Scene of Action *1*

Chapter 2. Amino Acids, Peptides, and Proteins *39*

Chapter 3. Determining Structures and Analyzing Cells *95*

Chapter 4. Sugars, Polysaccharides, and Glycoproteins *161*

Chapter 5. The Nucleic Acids *199*

Chapter 6. Thermodynamics and Biochemical Equilibria *281*

Chapter 7. How Macromolecules Associate *325*

Chapter 8. Lipids, Membranes, and Cell Coats *379*

Chapter 9. Enzymes: The Catalysts of Cells *455*

Chapter 10. An Introduction to Metabolism *505*

Chapter 11. The Regulation of Enzymatic Activity and Metabolism *535*

Chapter 12. Transferring Groups by Displacement Reactions *589*

Chapter 13. Enzymatic Addition, Elimination, Condensation, and Isomerization: Roles for Enolate and Carbocation Intermediates *677*

Chapter 14. Coenzymes: Nature's Special Reagents *719*

Chapter 15. Coenzymes of Oxidation – Reduction Reactions *765*

Chapter 16. Transition Metals in Catalysis and Electron Transport *837*

# Table of Contents
## *Volume 2*


Chapter 17. The Organization of Metabolism *938*

Chapter 18. Electron Transport, Oxidative Phosphorylation, and Hydroxylation *1012*

Chapter 19. The Chemistry of Movement *1088*

Chapter 20. Some Pathways of Carbohydrate Metabolism *1128*

Chapter 21. Specific Aspects of Lipid Metabolism *1180*

Chapter 22. Polyprenyl (Isoprenoid) Compounds *1226*

Chapter 23. Light and Life *1272*

Chapter 24. The Metabolism of Nitrogen and Amino Acids *1358*

Chapter 25. Metabolism of Aromatic Compounds and Nucleic Acid Bases *1420*

Chapter 26. Biochemical Genetics *1472*

Chapter 27. Organization, Replication, Transposition, and Repair of DNA *1528*

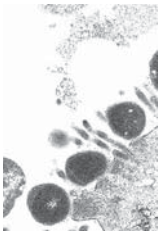
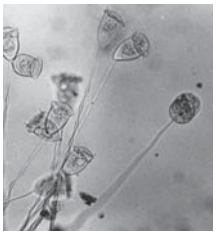
Chapter 28. The Transcription of Genes *1602*

Chapter 29. Ribosomes and the Synthesis of Proteins *1668*

Chapter 30. Chemical Communication Between Cells *1740*

Chapter 31. Biochemical Defense Mechanisms *1830*

Chapter 32. Growth and Development *1878*

Left: Cells of the pathogenic O157:H7 strain of *Escherichia coli* attached to the surface epithelium of the cecum of a neonatal piglet. Electron-dense filaments (presumably polymerized actin) in the host cytoplasm can be seen subjacent to attached bacteria. The bacteria have effaced most micro-villi but some remain between the bacterial cells. Courtesy of Evelyn A. Dean-Nystrom, National Animal Disease Center, USDA, Agricultural Research Service, Ames, IA. Center: Many unicellular organisms such as these *Vorticella* inhabit wet and moist environments throughout the biosphere. Invertebrates have evolved as long as humans and have complex specializations such as the contractile stem of these protozoa. Courtesy of Ralph Buchsbaum. Right: Although 97% of animals are invertebrates, ~ 3% of the several million known species have backbones. Giraffe: © M. P. Kahl, Photo Researchers

# Contents


2 ..... **A. The Simplest Living Things**
3 ..... 1. *Escherichia coli*
3 ..... 2. The Bacterial Genome
5 ..... 3. Ribonucleic Acids (RNA) and the Transcription and Translation of Genetic Information
5 ..... 4. Membranes and Cell Walls
6 ..... 5. Flagella and Pili
6 ..... 6. Classification and Evolution of Bacteria
8 ..... 7. Nutrition and Growth of Bacteria
9 ..... 8. Photosynthetic and Nitrogen-Fixing Prokaryotes
11 ..... **B. Eukaryotic Cells**
11 ..... 1. The Nucleus
11 ..... 2. The Plasma Membrane
12 ..... 3. Vacuoles, Endocytosis, and Lysosomes
13 ..... 4. The Endoplasmic Reticulum and Golgi Membranes
14 ..... 5. Mitochondria, Plastids, and Peroxisomes
15 ..... 6. Centrioles, Cilia, Flagella, and Microtubules
15 ..... 7. Cell Coats, Walls, and Shells
15 ..... **C. Inheritance, Metabolic Variation, and Evolution of Eukaryotes**
17 ..... 1. A Changing Genome
17 ..... 2. Genetic Recombination, Sex, and Chromosomes
18 ..... 3. Haploid and Diploid Phases
18 ..... **D. Survey of the Protists**
18 ..... 1. Protozoa
20 ..... 2. Fungi
20 ..... 3. Algae
23 ..... **E. The Variety of Animal Forms**
23 ..... 1. Major Groups of Multicellular Animals
25 ..... 2. Cell Types and Tissues
26 ..... *Tissues*
26 ..... *Blood cells*
26 ..... *Cell culture*
26 ..... 3. Communication
28 ..... *Cell contacts and junctions*
29 ..... *Cell recognition*
29 ..... **F. Higher Plants and Plant Tissues**
30 ..... **G. The Chemical Composition of Cells**

34 ..... **References**
36 ..... **Study Questions**


**Boxes**

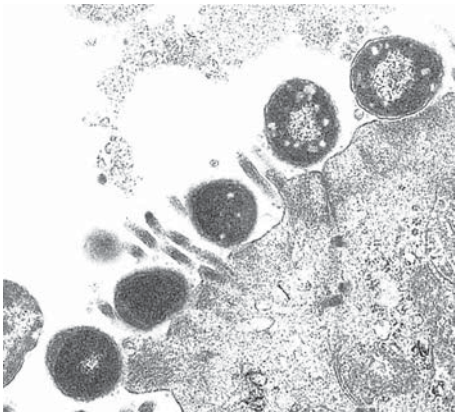
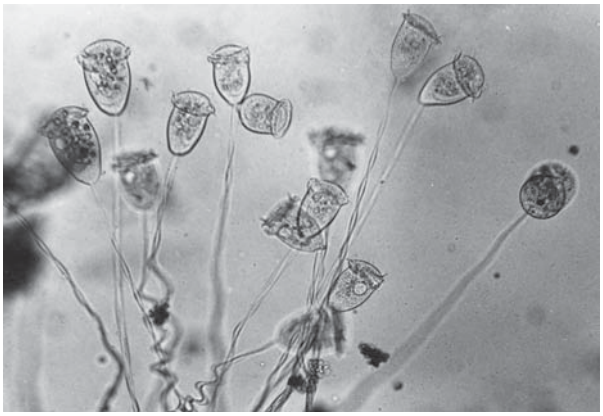

2 ..... Box 1-A About Measurements
4 ..... Box 1-B Relative Molecular Mass, $M_r$, and Daltons
9 ..... Box 1-C In the Beginning
16 ..... Box 1-D Inherited Metabolic Diseases
25 ..... Box 1-E Errors, Misconceptions, and Speculation
32 ..... Box 1-F About the References


**Tables**


7 ..... Table 1-1 A Systematic Classification Scheme for Bacteria
11 ..... Table 1-2 Approximate Sizes of Some Cells
12 ..... Table 1-3 Haploid Genome Sizes for Several Organisms
31 ..... Table 1-4 Approximate Composition of Metabolically Active Cells and Tissues

# The Scene of Action

# 1

This book is about the chemistry of living cells with special emphasis on the trillions of cells that make up your own body. Every aspect of life depends upon the chemical makeup of cells and on the chemical properties of the remarkable molecules found within the cells. The information presented here will give the reader a solid foundation for understanding not only the chemical basis of life but also the revolutionary developments in molecular biology, biochemical genetics, medicine, and agriculture which dominate today's scientific news and which will play an increasingly important role in our lives.

The first theme of the book is **biomolecular structure**. We'll look carefully at the complex structures of proteins, carbohydrates, RNA, DNA, and many other substances. We'll not only examine in-depth their molecular architecture but also study the chemical properties that make life possible.

A second theme is **metabolism**, the unceasing, complex network of thousands of chemical reactions by which cells grow and reproduce, take up foods and excrete wastes, move, and communicate with each other. Within cells we have a **steady state**, a condition in which the complex chemical constituents of cells are continuously being synthesized in one series of reactions and degraded in another. The result is a marvelous system of self-renewal or "turnover" of tissues. We'll examine the chemical reactions involved in these processes as well as the ways in which they are controlled. We will consider both the reaction sequences and the techniques such as cloning of genes, isotopic labeling, X-ray diffraction, and nuclear magnetic resonance spectroscopy, which are used today to study metabolism.

Human beings are surrounded by many other living creatures whose activities are important to us. Photosynthetic organisms obtain energy from sunlight and synthesize compounds that the human body requires but cannot make. Microorganisms cause decay of organic matter and convert it into forms usable by plants. This book deals with the chemical reactions occurring in all of these organisms. We'll look at strange and unusual reactions, along with those metabolic sequences common to most living things.

Each one of the thousands of chemical reactions of metabolism is catalyzed by an **enzyme**. Most of these enzymes are proteins, but others are made from **RNA** (ribonucleic acid). In both cases enzymes are very large molecules with precise three-dimensional structures. The study of the properties of enzymes and of **enzymatic catalysis** is a third theme of the book. Not only are the chemical mechanisms by which enzymes act of interest but also enzymes are often targets for useful drugs. Incorrectly formed enzymes can result in serious diseases.

The sequences of the amino acids in the chains from which proteins are constructed are encoded in the nucleotide sequences of **DNA** (deoxyribonucleic acid). The coding sequence for a protein in the DNA is found in the **structural gene** for that protein. The RNA enzymes are also encoded by DNA genes. A fourth major theme of the book deals with the nature of the **genetic code** used in DNA and with the processes by which cells read and interpret the code. It also includes study of the methods by which thousands of genes have been mapped to specific positions in chromosomes, isolated, cloned, and sequenced.

A large number of proteins present in the outer surfaces of cells serve as **receptors** that receive chemical messages and other signals from outside the cell. The receptors, which are sometimes enzymes, respond by generating internal signals that control metabolism and cell growth. Such **molecular signaling** is another major area of contemporary biochemistry.

Biologists have described over a million species, and several millions of others probably exist.[1] Many of these organisms have very specialized ways of life. However, they all have much chemistry in common. The same 20 amino acids can be isolated from proteins of plants, animals, and microorganisms. Formation of lactic acid in both bacteria and human muscle requires the same enzymes. Except for some small variations, the genetic code is universal—the same for all organisms. Thus, there is a unity of life and we can study metabolism as the entirety of chemical transformations going on in all living things. However, the differences among species are also impressive. Each species has its own gene for almost every protein.

When the enzyme that catalyzes a particular metabolic reaction is isolated from a number of different organisms, it is usually found to have similar properties and a similar mechanism of catalysis, regardless of the source. However, the *exact* sequence of amino acids in the enzyme will be almost unique to the organism that produced it. When the three-dimensional structures are compared it is found that differences between species often affect only the peripheral parts of an enzyme molecule. The interior structure of the protein, including the catalytic machinery, is highly conserved. However, the surface regions, which often interact with other macromolecules, vary greatly. Such interactions help to control metabolism and may account for many differences in the metabolism among living beings.

Variations in protein structures are not limited to differences between species. Individuals differ from one another. Serious genetic diseases sometimes result from the replacement of a single amino acid unit in a protein by a different amino acid. Genetic deviations from the "normal" structure of a protein result from **mutations**. Many mutations, whether they occurred initially in our own cells or in those of our ancestors, are detrimental.

However, such mutations also account for variation among individuals of a species and allow for evolution. The chemical nature and consequences of mutations and their significance to health, medicine, and agriculture are dealt with throughout the book. We now have reliable methods for inducing in the laboratory mutations at any specific place in a protein sequence and also for synthesizing new DNA sequences. These techniques of **genetic engineering** have given biochemists the ability to modify protein structures freely, to create entirely new proteins, and to provide a basis for the rapidly developing field of **genetic therapy**.

It should be clear from this introduction that **biochemistry** deals with virtually every aspect of life. The distinguishing feature of the science is that it approaches biological questions in terms of the underlying chemistry. The term **molecular biology** is often regarded as synonymous with biochemistry.

However, some scientists use it to imply a more biological approach. These molecular biologists also emphasize structure and function but may have a goal of understanding biological relationships more than chemical details. **Biophysics**, a closely related science, encompasses the application of physical and mathematical tools to the study of life.

## A. The Simplest Living Things

The simplest organisms are the **bacteria.**[2–5] Their cells are called **prokaryotic** (or procaryotic) because no membrane-enclosed nucleus is present. Cells of all other organisms contain nuclei separated from the cytoplasm by membranes. They are called **eukaryotic**. While viruses (Chapter 5) are sometimes regarded as living beings, these amazing parasitic objects are not complete organisms and have little or no metabolism of their own. The smallest bacteria are the **mycoplasmas**.[6–8] They do not have the rigid cell wall characteristic of most bacteria. For this reason they are easily deformed and often pass through filters designed to stop bacteria. They are nutritionally fussy and are usually, if not always, parasitic. Some live harmlessly in mucous membranes of people, but others cause diseases.

### BOX 1-A ABOUT MEASUREMENTS

In 1960 the International General Conference on Weights and Measures adopted an improved form of the metric system, **The International System of Units** (SI). The units of mass, length, and time are the kilogram (kg), meter (m), and second (s). The following prefixes are used for fractions and multiples:

| | | |
|---|---|---|
| $10^{-18}$, *atto* (a) | $10^{-6}$, *micro* (μ) | $10^{9}$, *giga* (G) |
| $10^{-15}$, *femto* (f) | $10^{-3}$, *milli* (m) | $10^{12}$, *tera* (T) |
| $10^{-12}$, *pico* (p) | $10^{3}$, *kilo* (k) | $10^{15}$, *peta* (P) |
| $10^{-9}$, *nano* (n) | $10^{6}$, *mega* (M) | $10^{18}$, *exa* (E) |

There is an inconsistency in that the prefixes are applied to the gram (g) rather than to the basic unit, the kilogram.

SI units have been used throughout the book whenever possible. There are no feet, microns, miles, or tons. Molecular dimensions are given uniformly in nanometers rather than in angstrom units (Å; 1Å = 0.1 nm). Likewise the calorie and kilocalorie have been replaced by the SI unit of energy, the **joule** (J; 1 calorie = 4.184 J).

Throughout the book frequent use is made of the following symbols:

, "appr oximately equal to"
~, "approximately" or "about"

For example, M*ycoplasma pneumoniae* is responsible for primary atypical pneumonia.

Cells of mycoplasmas sometimes grow as filaments but are often spherical and as small as 0.3 micrometer (μm) in diameter. Their outer surface consists of a thin **cell membrane** about 8 nanometers (nm) thick. This membrane encloses the **cytoplasm**, a fluid material containing many dissolved substances as well as submicroscopic particles. At the center of each cell is a single, highly folded molecule of DNA, which constitutes the bacterial chromosome. Besides the DNA there may be, in a small spherical mycoplasma, about 1000 particles ~20 nm in diameter, the **ribosomes**. These ribosomes are the centers of protein synthesis. Included in the cytoplasm are many different kinds of proteins, but there is room for a total of only about 50,000 protein molecules. Several types of RNA as well as many smaller molecules are also present. Although we don't know what minimum quantities of proteins, DNA, and other materials are needed to make a living cell, it is clear that they must all fit into the tiny cell of the mycoplasma.

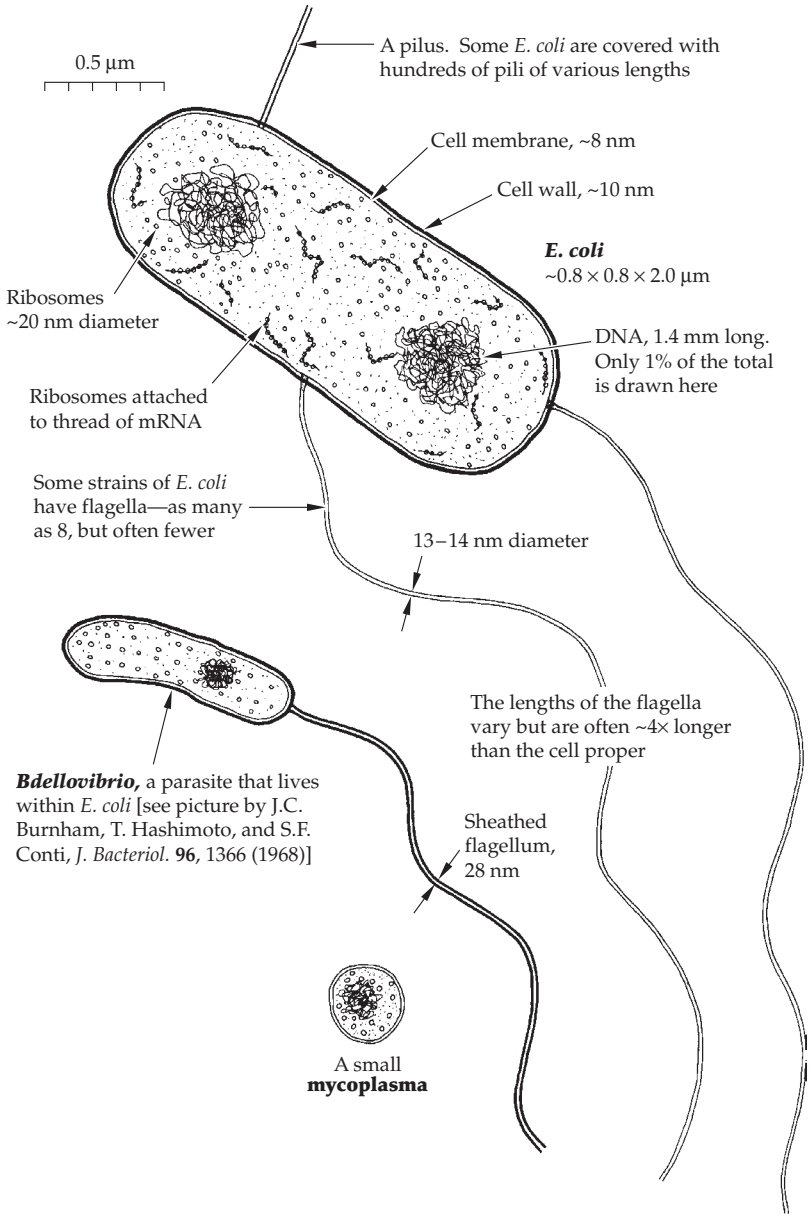



**Figure 1-1** *Escherichia coli* and some smaller bacteria.

## 1. *Escherichia coli*

The biochemist's best friend is *Escherichia coli*, an ordinarily harmless inhabitant of our intestinal tract. This bacterium is easy to grow in the laboratory and has become the best understood organism at the molecular level.[4,9] It may be regarded as a typical true bacterium or **eubacterium**. The cell of *E. coli* (Figs. 1-1, 1-2) is a rod ~2 μm long and 0.8 μm in diameter with a volume of ~1 $\mu m^3$ and a density of ~1.1 $g/cm^3$. The mass is ~1 x $10^{-12}$ g, i.e., 1 picogram (pg) or ~0.7 x $10^{12}$ daltons (Da) (see Box 1-B).[4] It is about 100 times bigger than the smallest mycoplasma but the internal structure, as revealed by the electron microscope, resembles that of a mycoplasma.

Each cell of *E. coli* contains from one to four identical DNA molecules, depending upon how fast the cell is growing, and ~15,000–30,000 ribosomes. Other particles that are sometimes seen within bacteria include food stores such as fat droplets and granules (Fig. 1-3). The granules often consist of **poly-β-hydroxybutyric acid**[10] accounting for up to 25% of the weight of *Bacillus megaterium*. **Polymetaphosphate**, a highly polymerized phosphoric acid, is sometimes stored in "metachromatic granules." In addition, there may be droplets of a separate aqueous phase, known as **vacuoles**.

## 2. The Bacterial Genome

The genetic instructions for a cell are found in the **DNA molecules**. All DNA is derived from four different kinds of monomers, which we call **nucleotides**. DNA molecules are double-stranded: two polymer chains are coiled together, their nucleotide units being associated as **nucleotide pairs** (see Fig. 5-7). The genetic messages in the DNA are in the form of

sequences of nucleotides. These sequences usually consist of a series of code "words" or **codons**. Each codon is composed of three successive nucleotides and specifies which one of the 20 different kinds of amino acids will be used at a particular location in a protein. The sequence of codons in the DNA tells a cell how to order the amino acids for construction of its many different proteins.

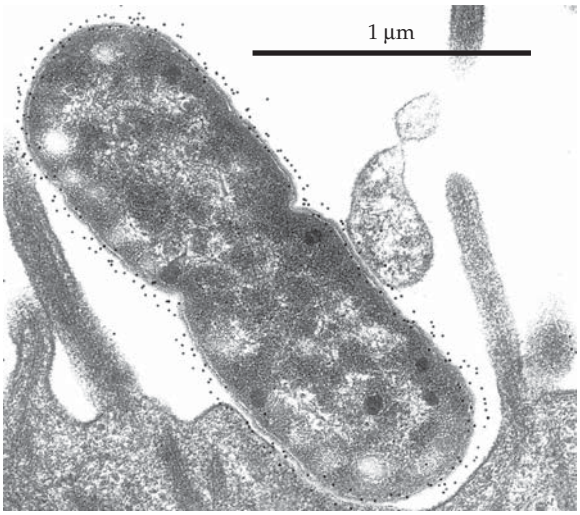



**Figure 1-2** Transmission electron micrograph of a dividing cell of *Escherichia coli* O157:H7 attached to the intestinal epithelium of a neonatal calf. These bacteria, which are able to efface the intestinal microvilli, form characteristic attachments, and secrete shiga toxins, are responsible for ~73,000 illnesses and 60 deaths per year in the U. S.[10a] After embedding, the glutaraldehyde-fixed tissue section was immunostained with goat anti-O157 IgG followed by protein A conjugated to 10-nm gold particles. These are seen around the periphery of the cell bound to the O-antigen (see Fig. 8-28). Notice the two microvilli of the epithelium. Courtesy of Evelyn A. Dean-Nystrom, National Animal Disease Center, USDA, Agricultural Research Service, Ames, IA.

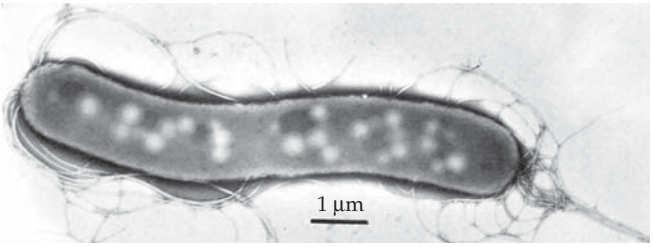



**Figure 1-3** A cell of a *Spirillum* negatively stained with phosphotungstic acid. Note the tufts of flagella at the ends, the rough appearance of the outer surface, the dark granules of poly-β-hydroxybutyric acid and the light-colored granules of unknown nature. Courtesy of F. D. Williams, Gail E. VanderMolen, and C. F. Amstein.

Assume that a typical protein molecule consists of a folded chain of 400 amino acids. Its structural gene will therefore be a sequence of 1200 nucleotide pairs. Allowing a few more nucleotides to form spacer regions between genes we can take ~1300 as the number of nucleotide pairs in a typical bacterial gene. However, some genes may be longer and some may be much shorter. The **genome** is the quantity of DNA that carries a complete set of genetic instructions for an organism. In bacteria, the genome is a single chromosome consisting of one double-stranded DNA molecule. *Mycoplasma genitalium* is the smallest organism for which the DNA sequence is known.[11] Its genome is a double-helical DNA circle of 580,070 nucleotide pairs and appears to contain about 480 genes (an average of ~1200 nucleotides per gene).

The average mass of a nucleotide pair (as the disodium salt) is 664 Da. It follows that the DNA of *M. genitalium* has a mass of ~385 x $10^6$ Da. The relative molecular mass ($M_r$) is 0.385 x $10^9$ (See Box 1-B for definitions of dalton and $M_r$). The DNA of *E. coli* is about seven times larger with a mass of ~2.7 x $10^9$ Da. It contains ~4.2 x $10^6$ nucleotide pairs and encodes over 4000 different proteins (see Table 1-3).

Each nucleotide pair contributes 0.34 nm to the length of the DNA molecule; thus, the total length of DNA of an *E. coli* chromosome is 1.4 mm. This is about 700 times the length of the cell which contains it. Clearly, the molecules of DNA are highly folded, a fact that accounts for their appearance in the electron microscope as dense aggregates called **nucleoids**, which occupy about one-fifth of the cell volume (Fig. 1-4).

## BOX 1-B RELATIVE MOLECULAR MASS, $M_r$, AND DALTONS

Atomic and molecular masses are assigned relative to the mass of the carbon isotope, $^{12}C$, whose atomic weight is defined as exactly 12. The actual mass of a single atom of $^{12}C$ is defined as 12 **daltons**, one dalton being 1.661 x $10^{-24}$ g. The mass of a molecule can be given in daltons **(Da)** or kilodaltons **(kDa)**. This molecular mass in daltons is numerically equivalent to the relative molecular mass **($M_r$)** or molecular weight **(MW)**[a] and also to the molar mass **(g/mol)**. However, it is not correct to use the dalton for the unitless quantity $M_r$. Masses of structures such as chromosomes, ribosomes, mitochondria, viruses, and whole cells as well as macromolecules can be given in daltons.[b]

[a] The Union of Pure and Applied Chemistry renamed molecular weight as **relative molecular mass** with the symbol $M_r$; $M_r$ = MW.
[b] J. T. Edsall (1970) *Nature* (*London*) **228**, 888.

Each bacterial nucleoid contains a complete set of genetic "blueprints" and functions independently. Each nucleoid is **haploid**, meaning that it contains only a single complete set of genes. In addition to their chromosome, bacteria often contain smaller DNA molecules known as **plasmids**. These plasmids also carry genetic information that may be useful to bacteria. For example, they often encode proteins that confer resistance to antibiotics. The ability to acquire new genes from plasmids is one mechanism that allows bacteria to adapt readily to new environments.[12] Plasmids are also used in the laboratory in the cloning of genes and in genetic engineering (Chapter 26).

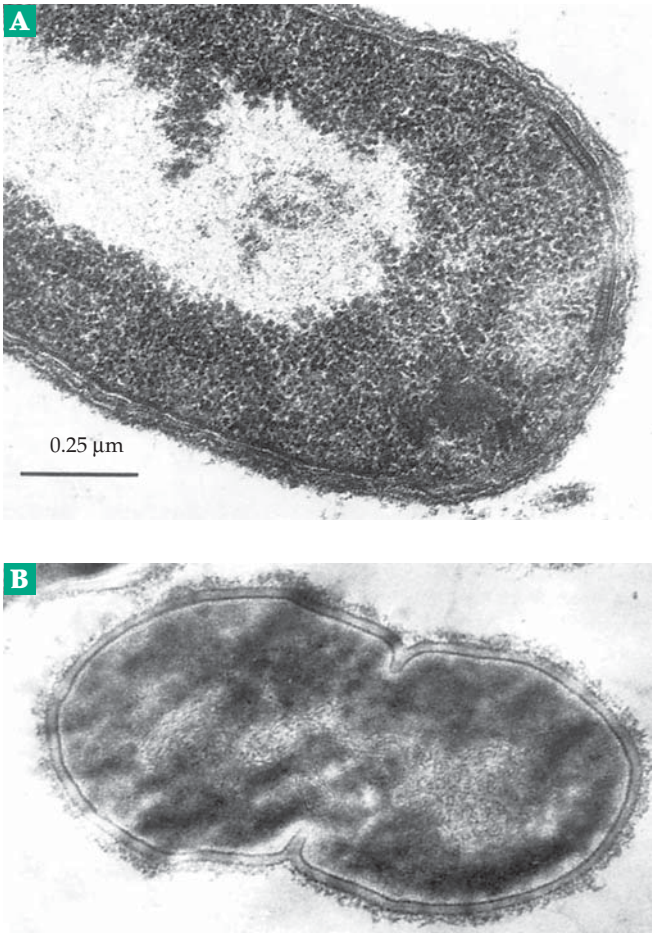



**Figure 1-4** (A) Thin (~60 nm) section of an aquatic gram-negative bacterium, *Aquaspirillum fasciculus*. Note the light-colored DNA, the dark ribosomes, the double membrane characteristics of gram-negative bacteria (Chapter 8, Section E), and the cell wall. In addition, an internal "polar membrane" is seen at the end. It may be involved in some way in the action of the flagella. (B) A thin section of dividing cell of *Streptococcus*, a gram-positive organism. Note the DNA (light-stranded material). A portion of a mesosome is seen in the center and septum can be seen forming between the cells. Micrographs courtesy of F. D. Williams, Gail E. VanderMolen, and C. F. Amstein.

## 3. Ribonucleic Acids (RNA) and the Transcription and Translation of Genetic Information

The genetic information in the DNA is not utilized directly by the protein-synthesizing machinery of cells. Instead, molecules of ribonucleic acid (RNA) are synthesized according to the instructions encoded in the DNA, a process called **transcription**. Although they differ from DNA significantly in their structure, these RNA molecules carry the same coded information as is found in a length of DNA that contains one or a few genes. If DNA is regarded as the "master blueprint" of the cell, molecules of RNA are "secondary blueprints." This concept is embodied in the name **messenger RNA** (mRNA) which is applied to a small, short-lived fraction of RNA that carries information specifying amino acid sequences of proteins. Each molecule of mRNA carries the genetic message from one or more genes to the ribosomes where the proteins are made.

Ribosomes are extraordinarily complex little protein-synthesizing machines. Each ribosome of *E. coli* has a mass of $2.7 \times 10^6$ Da and contains 65% of a stable **ribosomal RNA** and ~35% protein. About 50 different kinds of protein molecules are present as parts of the ribosomal structure. Working together with a variety of **transfer RNA** molecules and enzymes, the ribosomes are able to read the genetic messages from mRNA and to accurately assemble any kind of protein molecule that a gene may specify. This process is called **translation** of the genetic message.

## 4. Membranes and Cell Walls

Like the mycoplasma, the *E. coli* cell is bounded by an 8-nm membrane which consists of ~50% protein and 50% lipid. When "stained" (e.g., with permanganate) for electron microscopy, this single membrane appears as two very thin (2.0 nm) dark lines separated by an unstained center band (~3.5 nm) (Fig. 1-4; see also Fig. 8-4). Single membranes of approximately the same thickness and staining behavior occur in all cells, both of bacteria and of eukaryotes.

A cell membrane is much more than just a sack. It serves to control the passage of small molecules into and out of the cell. Its outer surface carries receptors for recognition of various materials. The inside surface of bacterial membranes contains enzymes that catalyze most of the oxidative metabolism of the cells. Bacterial cell membranes are sometimes folded inward to form internal structures involved in photosynthesis or other specialized reactions of metabolism such as oxidation of ammonia to nitrate.[2] In *E. coli* replication of DNA seems to occur on certain parts of the membrane surface, probably under the control of membrane-bound enzymes. The formation of the new membrane which

divides multiplying cells proceeds synchronously with the synthesis of DNA.

A characteristic of true bacteria (**eubacteria**) is a rigid **cell wall** which surrounds the cell membrane. The 40-nm-thick wall of *E. coli* is a complex, layered structure five times thicker than the cell membrane. Its chemical makeup is considered in Chapter 8. One of the layers is often referred to as the **outer membrane**. In some bacteria the wall may be as much as 80 nm thick and may be further surrounded by a thick **capsule** or **glycocalyx** (slime layer).[13] The main function of the wall seems to be to prevent osmotic swelling and bursting of the bacterial cell when the surrounding medium is hypotonic.

If the osmotic pressure of the medium is not too low, bacterial cell walls can sometimes be dissolved, leaving living cells bounded only by their membranes. Such **protoplasts** can be produced by action of the enzyme lysozyme on gram-positive bacteria such as *Bacillus megaterium*. Treatment of cells of gram-negative bacteria with penicillin (Box 20-G) produces **spheroplasts**, cells with partially disrupted walls. Spheroplasts and protoplasts are useful in biochemical studies because substances enter cells more readily when the cell wall is absent. Strains of bacteria lacking rigid walls are known as **L forms**.

## 5. Flagella and Pili

Many bacteria swim at speeds of 20–60 μm/s, ten or more body lengths per second! Very thin thread-like **flagella** of diameter 13–20 nm coiled into a helical form are rotated by the world's smallest "electric motors" to provide the motion.[14] While some bacteria have a single flagellum, the corkscrew-like *Spirillum* (Fig. 1-3) synchronously moves tufts of flagella at both ends. Some strains of *E. coli* have no flagella, but others contain as many as eight flagella per cell distributed over the surface. The flagella stream out behind in a bundle when the bacterium swims. The flagella of the helical **spirochetes** are located inside the outer membrane.[15,16]

In addition to flagella, extremely thin, long, straight filaments known as **pili** or **fimbriae** (Fig. 1-2) project from the surfaces of many bacteria.[14] The "sex pili" (F pili and I pili) of *E. coli* have a specific role in sexual conjugation. The similar but more numerous common pili or fimbriae range in thickness from 3 to 25 nm and in length from 0.2 to 2 μm. Pili are involved in adhesion of bacteria to surrounding materials or to other bacteria and facilitate bacterial infections.[17–19] A typical *E. coli* cell has 100–300 pili.[5]

## 6. Classification and Evolution of Bacteria

Bacteria vary greatly in their chemistry and metabolism, and it is difficult to classify them in a rational way. In higher organisms species are often defined as forms that cannot interbreed and produce fertile offspring, but such a criterion is meaningless for bacteria whose reproduction is largely asexual and which are able readily to accept "visiting genes" from other bacteria.[12] The classification into species and genera is therefore somewhat arbitrary. A currently used scheme (Table 1-1)[20] classifies the prokaryotes into 35 groups on the basis of many characteristics including shape, staining behavior, and chemical activities. Table 1-1 also includes genus names of most of the bacteria discussed in this book.

Bacteria may have the shape of spheres or straight or curved rods. Some, such as the **actinomycetes**, grow in a branching filamentous form. Words used to describe bacteria often refer to these shapes: a **coccus** is a sphere, a **bacillus** a rod, and a **vibrio** a curved rod with a flagellum at one end. A **spirillum** is screw-shaped. These same words are frequently used to name particular genera or families. Other names are derived from some chemical activity of the bacterium being described.

The **gram stain** provides an important criterion of classification that depends upon differences in the structure of the cell wall (see Chapter 20). Bacterial cells are described as **gram-positive** or **gram-negative** according to their ability to retain the basic dye crystal violet as an iodine complex. This difference distinguishes two of four large categories of bacteria.[20] Most actinomycetes, the spore-forming bacilli, and most cocci are gram-positive, while *E. coli*, other enterobacteria, and pseudomonads are gram-negative. A third category consists of eubacteria that lack cell walls, e.g. the mycoplasma.

Comparisons of amino acid sequences of proteins and the nucleotide sequences of DNA and RNA have provided a new approach to classification of bacteria.[21–28] Although the origins of life are obscure, we can easily observe that the genome changes with time through mutation and through the enzyme-catalyzed process of **genetic recombination**. The latter gives rise to the deletion of some nucleotides and the insertion of others into a DNA chain. When we examine sequences of closely related species, such as *E. coli* and *Salmonella typhimurium*, we find that the sequences are very similar. However, they differ greatly from those of many other bacteria. Consider the 23S ribosomal RNA, a molecule found in the ribosomes of all bacteria. It contains ~3300 nucleotides in a single highly folded chain. The basic structure is highly conserved but between any two species of bacteria there are many nucleotide substitutions caused by mutations as well as deletions and insertions. By asking what is the minimum number of

mutations that could have converted one 23S RNA into another and by assuming a more or less constant rate of mutation over millions of years it is possible to construct a **phylogenetic tree** such as that shown in Fig. 1-5.

One conclusion from these comparisons is that the methane-producing bacteria, the **methanogens**,[24] are only distantly related to most other bacteria. Methanogens together with the cell wall-less *Thermoplasma*,[28] some salt-loving **halobacteria**, and some **thermophilic** (heat-loving) sulfur bacteria form a fourth major category. They are often regarded as a separate kingdom, the **archaeobacteria**,[25] which together with the kingdom of the eubacteria form the superkingdom prokaryota. Certain archaeobacteria have biochemical characteristics resembling those of eukaryotes and

**TABLE 1-1**
**A Systematic Classification Scheme for Bacteria[a,b]**

**Kingdom Procaryotae**

The bacteria are classified according to the following 35 groups. Within these groups many genera are classified into subgroups or families. A few genera, most of which are mentioned elsewhere in this book, are listed here by name. Members of a single subgroup are placed together and are separated by semicolons from members of other subgroups.

1. The spirochetes (long bacteria, up to 500 µm, that are propelled by the action of filaments wrapped around the cell between the membrane and wall). *Borrelia* (*B. burgdorferi*, Lyme disease), *Leptospira*, *Treponema* (*T. pallidum*, syphilis)
2. Aerobic spiral and curved motile gram-negative bacteria. *Bdellovibrio*, *Campylobacter* (*C. jejuni*, diarrhea), *Helicobacter* (*H. pylori*, gastric ulcers), *Spirillum*
3. Nonmotile gram-negative curved bacteria
4. Gram-negative, aerobic rods and cocci. *Acetobacter, Agrobacterium, Azotobacter, Brucella* (*B. abortus*, brucellosis), *Flavobacterium, Gluconobacter, Legionella* (*L. pneumophila*, Legionnaire's disease), *Methylomonas*, *Neisseria* (*N. gonorrhea*, gonorrhea), *Pseudomonas*, *Rhizobium*, *Thermus, Xanthomonas, Rochalimaea* (*R. henselae*, cat scratch disease)
5. Gram-negative, facultatively anaerobic rods. *Enterobacter,*[c] *Proteus*, *Yersinia* (*Y. pestis*, plague), *Escherichia, Klebsiella, Salmonella* (*S. typhi*, typhoid fever), *Serratia, Shigella* (*S. dysenteriae*, bacterial dysentery), *Haemophilus; Vibrio* (*V. cholerae*, Asiatic cholera); *Zymomonas*
6. Gram-negative, anaerobic bacteria. *Butyrivibrio*
7. Dissimilatory sulfate- or sulfur-reducing bacteria. *Desulfovibrio*
8. Anaerobic gram-negative cocci. *Veillonella*
9. The rickettsias (parasitic bacteria with exacting nutritional requirements and small genome sizes) and chlamydias. *Chlamydia* (*C. trachomatis*, trachoma), *Rickettsia* (*R. rickettsii*, Rocky Mountain spotted fever)
10. Anoxygenic photosynthetic bacteria. Green sulfur bacteria. *Chlorobium, Prosthecochloris;* purple nonsulfur bacteria: *Rhodopseudomonas, Rhodospirillum;* purple sulfur bacteria: *Chromatium, Thiospirillum*
11. Oxygenic photosynthetic bacteria. Cyanobacteria (blue-green algae): *Synechocystis; Anabaena, Nostoc; Oscillatoria*
12. Aerobic, chemolithotrophic bacteria. Colorless sulfur bacteria: *Thiobacillus;* iron or manganese-oxidizing bacteria, magnetotactic bacteria; nitrifying bacteria: *Nitrobacter, Nitrosomonas*
13. Budding and/or appendaged bacteria. *Caulobacter*
14. Sheathed bacteria
15. Nonphotosynthetic, nonfruiting gliding bacteria. *Beggiatoa* (a filamentous bacterium containing sulfur granules)
16. Fruiting, gliding bacteria. *Myxococcus*
17. Gram-positive cocci. *Leuconostoc, Micrococcus, Peptococcus, Staphylococcus* (*S. aureus,* boils, infections), *Streptococcus* (*S. pyogenes,* scarlet fever, throat infections, *S. pneumoniae,* pneumonia)
18. Endospore-forming gram-positive rods and cocci. Aerobic: *Bacillus* (*B. anthracis,* anthrax), anaerobic: *Clostridium* (*C. tetani,* tetanus; *C. botulinum,* botulism)
19. Regular nonsporing gram-positive rods. *Lactobacillus*
20. Irregular nonsporing gram-positive rods. *Actinomyces, Bifidobacterium, Corynebacterium* (*C. diphtheriae,* diphtheria), *Propionibacterium*
21. Mycobacteria. *Mycobacterium* (*M. tuberculosis,* tuberculosis; *M. leprae,* leprosy)

22-29. Actinomycetes

30. Mycoplasmas. *Acholeplasma, Mycoplasma*
31. Methanogens. *Methanobacterium; Methanosarcina; Methanospirillum*
32. Archaeal sulfate reducers
33. Halobacteria. *Halobacterium*
34. Cell wall-less archaeobacteria. *Thermoplasma*
35. Very thermophilic $S^0$-Metabolizers. *Sulfolobus; Thermococcus*

[a] From Bergey's Manual of Systematic Bacteriology, 9th ed. J. G. Holt, N. R. Krieg, P. H. A. Sneath, J. T. Staley and S. T. Williams, Eds. (1994) Williams and Wilkins, Baltimore, Maryland. For another recent list see http://www.ncbi.nlm.nih.gov/

[b] The human diseases caused by some species are also listed.

[c] Formerly *Aerobacter*.

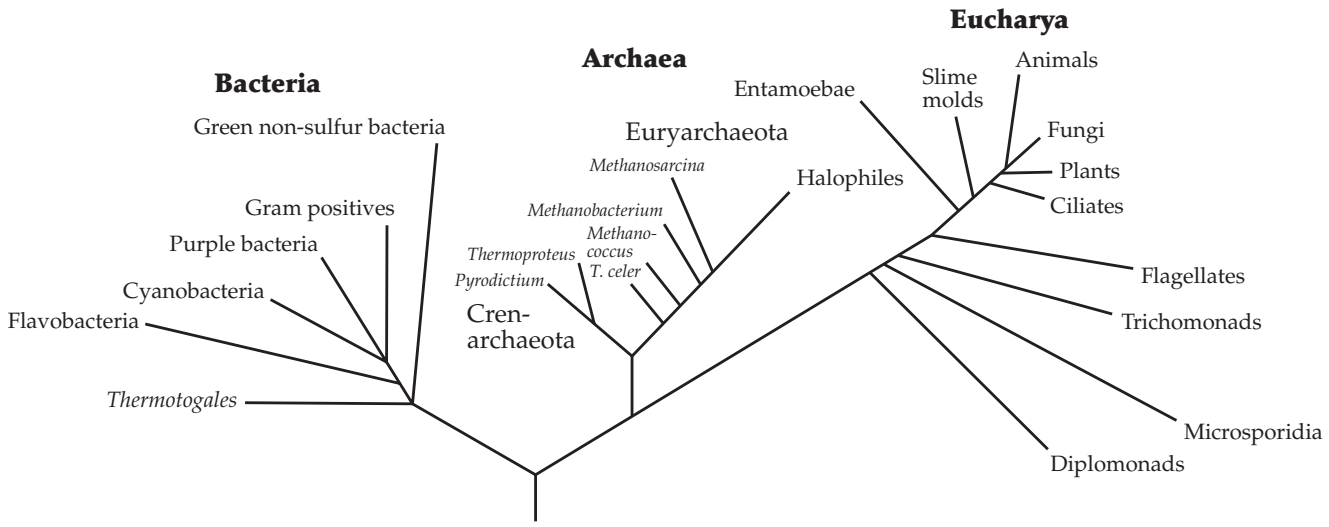



**Figure 1-5** Universal phylogenetic tree. From Wheelis *et al.*[29]

some biologists therefore classify them as **archaea** and rank their kingdom as equal to that of the bacteria and the eukaryotes (Fig. 1-5).[27,29,30,30a,30b] Others disagree.[31] In Table 1-1, the archaeobacteria are found in groups 31–35. Most bacteria are very small in size but there are species large enough to be confused with eukaryotic protozoa. The record for bacteria seems to be held by *Epulopiscium fishelsoni,* a parasite of the surgeonfish intestinal tract. A single cell measured > 600 μm by 80 μm diameter, over $10^6$ times larger in volume than a cell of *E. coli*.[32] The organism is a gram-positive bacterium as judged by analysis of its cloned ribosomal RNA genes.

## 7. Nutrition and Growth of Bacteria

**Autotrophic** (self-nourishing) bacteria can synthesize all of their organic cell constituents from carbon dioxide, water, and inorganic forms of nitrogen and sulfur. The **photoautotrophs** extract their energy from sunlight, while the **chemoautotrophs** obtain energy from inorganic chemical reactions. For example, the hydrogen bacteria oxidize $H_2$ to $H_2O$ and sulfur bacteria oxidize $H_2S$ to $H_2SO_4$. Like the fungi and animals, most bacteria are **chemoheterotrophic**; they obtain energy from the breakdown of organic compounds. Some of these heterotrophic bacteria are **anaerobes** which live without $O_2$. Many of them metabolize complex organic substances such as sugars in the absence of oxygen, a process called **fermentation**. Others oxidize organic compounds with an inorganic oxidant such as nitrite or sulfate. Members of the genus *Clostridium* are poisoned by oxygen and are known as **obligate anaerobes**. Others, including *E. coli*, are **facultative anaerobes**, able to grow either in the presence or in the absence of oxygen. **Obligate aerobes** depend for energy upon combustion of organic compounds with oxygen.

One of the largest groups of strictly aerobic heterotrophic bacteria, the pseudomonads (*Pseudomonas* and related genera), are of interest to biochemists because of their ability to oxidize organic compounds, such as alkanes, aromatic hydrocarbons, and steroids, which are not attacked by most other bacteria. Often, the number of oxidative reactions used by any one species of bacteria is limited. For example, the acetic acid bacteria that live in wine and beer obtain all of their energy by oxidation of ethanol to acetic acid:

$$CH_3CH_2OH + O_2 \rightarrow CH_3COOH + H_2O$$

Bacteria can grow incredibly fast. Under some conditions, it takes a bacterial cell only 10–20 min to double its size and to divide to form two cells.[4] An animal cell may take 24 h for the same process. Equally impressive are the rates at which bacteria transform their foods into other materials. One factor contributing to the high rate of bacterial metabolism may be the large surface to volume ratio. For a small spherical bacterium (coccus) of diameter 0.5 μm, the ratio of the surface area to the volume is $12 \times 10^6$ $m^{-1}$, while for an ameba of diameter 150 μm the ratio is only $4 \times 10^4$ $m^{-1}$ (the ameba can increase this by sticking out some pseudopods). Thimann[33] estimated that for a 90-kg human, the ratio is only 30 $m^{-1}$.

When food is limited, some bacteria such as the *Bacillus* form **spores**. These are compact little cells that form inside the vegetative cell and are therefore called **endospores**. They sometimes have only 1/10 the volume of the parent cell. Their water content is very low, their metabolic rate is near zero, and they are

extremely resistant to heat and further desiccation. Under suitable conditions, the spores can "germinate" and renew their vegetative growth. Spore formation is one of several examples of the development of specialized cells or **differentiation** among prokaryotes.

## 8. Photosynthetic and Nitrogen-Fixing Prokaryotes

It is likely that the earth was once a completely anaerobic place containing water, ammonia, methane, formaldehyde, and more complicated organic compounds. Perhaps the first forms of life, which may have originated about $3.5 \times 10^9$ years ago, resembled present-day anaerobic bacteria. The purple and green photosynthetic bacteria may be related to organisms that developed at a second stage of evolution: those able to capture energy from sunlight. Most of these gram-negative photosynthetic bacteria are strict anaerobes. None can make oxygen as do higher plants. Rather, the hydrogen needed to carry out the reduction of carbon dioxide in the photosynthetic process is obtained by the splitting of inorganic compounds, such as $H_2S$, thiosulfate, or $H_2$, or is taken from organic compounds. Today, photosynthetic bacteria are found principally in sulfur springs and in deep lakes, but at one time they were probably far more abundant and the only photosynthetic organisms on earth.

Before organisms could produce oxygen a second complete photosynthetic system, which could cleave $H_2O$ to $O_2$, had to be developed. The simplest oxygen-

### BOX 1-C IN THE BEGINNING

No one knows how life began. Theories ranging from the biblical accounts to recent ideas about the role of RNA are plentiful but largely unsatisfying. In the 1800s the great physical chemist Arrhenius was among scientists that preferred the idea held by some scientists today that a "seed" came from outer space. Until recently the only concrete data came from fossils. Making use of a variety of isotopic dating methods it can be concluded that cyanobacteria were present $2.2 \times 10^9$ years ago and eukaryotes $1.4 \times 10^9$ years ago. About $0.5 \times 10^9$ years ago the "Cambrian explosion" led to the appearance of virtually all known animal phyla. Many of these then became extinct about $0.2 \times 10^9$ years ago.

New insights published in 1859[a] were provided by Charles Darwin. However, his ideas were only put into a context of biochemical data after 1950 when sequencing of proteins and later nucleic acids began. From an astonishingly large library of sequence data available now we can draw one firm conclusion: *Evolution can be observed;*[b] *it does involve mutation of DNA*. Comparisons of sequences among many species allow evolutionary relationships to be proposed.[c-e] In general these are very similar to those deduced from the fossil record. They support the idea that evolution occurs by natural selection and that duplication of genes and movements of large pieces of DNA within the genome have occurred often. As many as 900 "ancient conserved regions" of DNA in the *E. coli* genome corresponding to those in human, nematode, and yeast DNA are thought to date back perhaps $3.5 \times 10^9$ years.[f] However, nobody has explained how life evolved before there was DNA.

One of the first scientists to devote his career to biochemical evolution was I. V. Oparin,[g] who published a book on the "origin of life" in 1924. Oparin and J. B. S. Haldane, independently, proposed that early life was anaerobic and that energy was provided by fermentation. In 1951 Stanley Miller built an apparatus that circulated $CH_4$, $NH_3$, $H_2O$, and $H_2$, compounds thought to be present in a primitive atmosphere, past an electric discharge. He found glycine, alanine, β-alanine, and other amino acids among the products formed.[h] Schrödinger pointed out that a flux of energy through a system will tend to organize the system. The solar energy passing through the biosphere induces atmospheric circulation and patterns of weather and ocean currents.[i,j] Perhaps in the primordeal oceans organic compounds arose from the action of light and lightning discharges. These compounds became catalysts for other reactions which eventually evolved into a rudimentary cell-less metabolism. It is a large jump from this to a cell! Among other problems is the lack of any explanation for the development of individual cells or of their genomes. However, because it helps to correlate much information *we will always take an evolutionary approach in this book* and will discuss the "beginnings" a little more in later chapters.

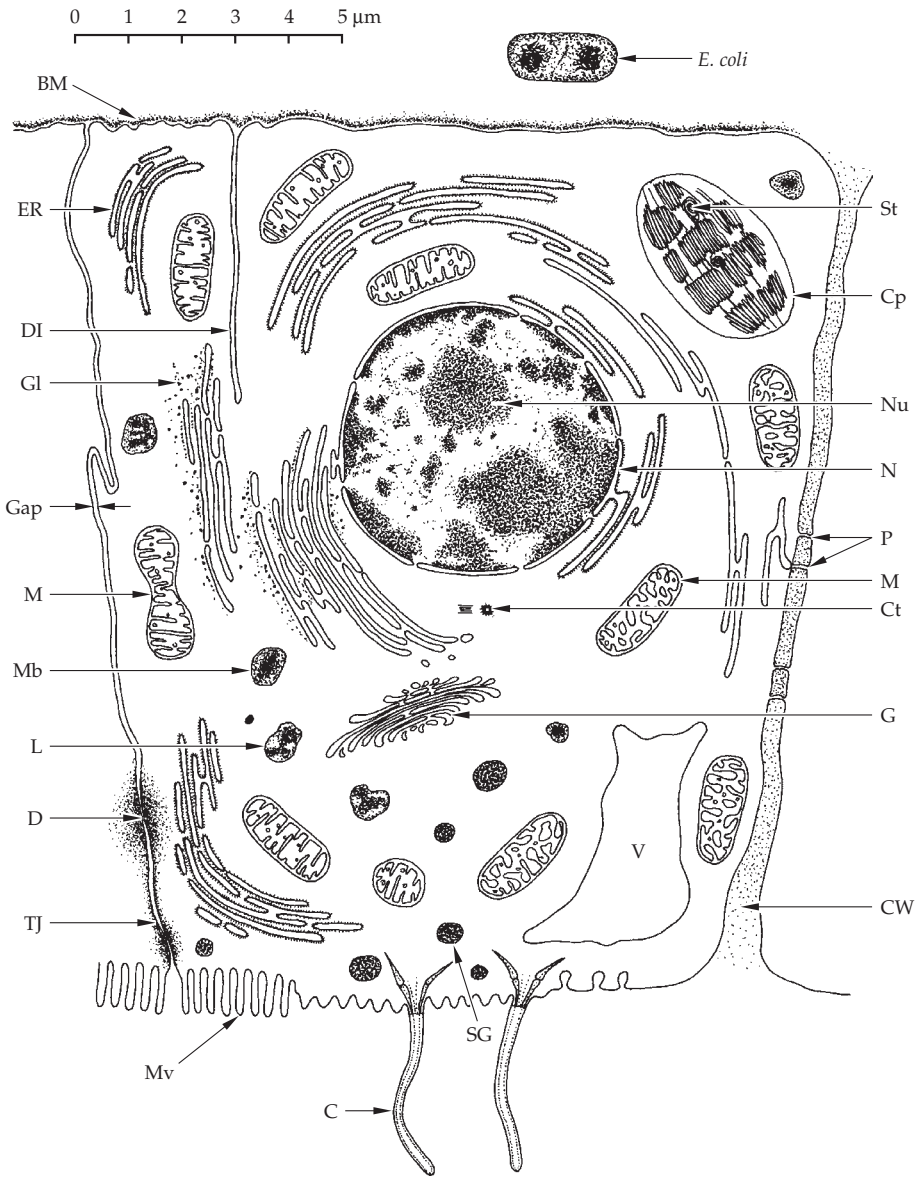
---

[a] Maynard-Smith, J. (1982) *Nature (London)* **296**, 599–601
[b] Lenski, R. E., and Travisano, M. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6808–6814
[c] Wilson, A. C. (1985) *Sci. Am.* **253**(Oct), 164–173
[d] Eigen, M., Gardiner, W., Schuster, P., and Winkler-Oswatitsch, R. (1981) *Sci. Am.* **244**(Apr), 88–118
[e] Doolittle, R. F. (1992) *Protein Sci.* **1**, 191–200
[f] Green, P., Lipman, D., Hillier, L., Waterston, R., States, D., and Claverie, J.-M. (1993) *Science* **259**, 1711–1716
[g] Broda, E. (1980) *Trends Biochem. Sci.* **5**, IV–V
[h] Miller, S. L. (1953) *Science* **117**, 528–529
[i] Mason, S. (1993) *Trends Biochem. Sci.* **18**, 230–231
[j] Welch, G. R. (1996) *Trends Biochem. Sci.* **21**, 452

producing creatures existing today are the **cyanobacteria**,[34] also known as blue-green algae. Many cyanobacteria are unicellular, but others such as *Oscillatoria*, a slimy "plant" that often coats the inside walls of household aquaria, consist of long filaments about 6 μm in diameter (see Fig. 1-11). All cyanobacteria contain two groups of pigments not found in other prokaryotes: **chlorophyll a** and **β-carotene**, pigments that are also found in the chloroplasts of true algae and in higher plants. A recently discovered group of bacteria, the **prochlorophytes**, are even closer to chloroplasts in their pigment composition.[35]

In addition to pigmented cells, some cyanobacteria contain paler cells known as **heterocysts**. They have a specialized function of fixing molecular nitrogen. The development of the ability to convert $N_2$ into organic nitrogen compounds represents another important evolutionary step. Because they can both fix nitrogen and carry out photosynthesis, the blue-green algae have the simplest nutritional requirements of any organisms. They need only $N_2$, $CO_2$, water, light, and minerals for growth.

Evolution of the photosynthetic cleavage of water to oxygen was doubtless a major event with far-reaching consequences. Biologists generally believe that as oxygen accumulated in the earth's atmosphere, the obligate anaerobes, which are poisoned by oxygen, became limited to strictly anaerobic environments. Meanwhile, a new group of bacteria, the **aerobes**, appeared with mechanisms for detoxifying oxygen




Abbreviations:

**BM**, basement membrane
**ER**, rough endoplasmic reticulum (with ribosomes attached; smooth ER is depicted nearer the nucleus and on the right side of the cell.)
**DI**, deep indentation of plasma membrane
**GI**, glycogen granules
**Gap**, space ~10-20 nm thick between adjacent cells
**M**, mitochondrion
**Mb**, microbody
**L**, lysosome
**D**, desmosome
**TJ**, tight junction
**Mv**, microvilli
**C**, cillium
**SG**, secretion granule
**V**, vacuole
**Nu**, nucleolus
**G**, Golgi apparatus
**CW**, cell wall (of a plant)
**Ct**, centrioles
**P**, plasmodesmata
**N**, nucleus
**Cp**, chloroplast
**St**, starch granule

**Figure 1-6** The "average" eukaryotic cell. This composite drawing shows the principal organelles of both animal and plant cells approximately to the correct scale. (Adapted from a drawing by Michael Metzler.)

and for using oxygen to oxidize complex organic compounds to obtain energy.

## B. Eukaryotic Cells

Cells of the **eukaryotes** contain true nuclei and are much larger and more complex internally than are those of prokaryotes. The nucleus of a cell contains most of its DNA and is separated from the cytoplasm by membranes. Within the cytoplasm are various **organelles** with characteristic structures. These include **mitochondria**, **lysosomes**, **peroxisomes,** and **centrioles**. Eukaryotic cells come in so many sizes and shapes and with so many specialized features that it is impossible to say what is typical. Nevertheless, Fig. 1-6 is an attempt to portray some sort of "average" cell, partly plant and partly animal.

As can be seen from Table 1-2, which lists the diameters and volumes of several roughly spherical cells, there is a great variation in size. However, a diameter of 10–20 µm may be regarded as typical for both plants and animals. For growth of a large cell such as the ovum, many adjacent cells assist in synthesis of foodstuffs which are transferred to the developing egg cell. Plant cells are often large but usually 90% or more of the cell is filled with a **vacuole** or **tonoplast**,[36] which is drawn unrealistically small in Fig. 1-6. The metabolically active protoplasm of plant cells often lies in a thin layer at their peripheries.

Many cells are far from spherical; for example, human red blood cells are discs 8 x 8 x 1 to 2 µm with a volume of 80 $\mu m^3$. Plant fiber cells may be several millimeters in length. Nerve cells of animals have long extensions, the **axons**, which in the human sometimes attain a length of a meter. Muscle cells fuse to give very long multinucleate fibers.

**TABLE 1-2**
**Approximate Sizes of Some Cells**

| Cell | Diameter (µm) | Approximate volume ($\mu m^3$) |
|---|---|---|
| *E. coli* | 1 | 1.0 |
| Small thymus cell | 6 | 120 |
| Liver cell | 20 | 4,000 |
| Human ovum (mature) | 120 | 500,000 |
| Hen's egg (white excluded) | 20,000 | $4 \times 10^{12}$ |
| Yeast cell | 10 | 500 |
| Onion root (meristematic cell) | 17 | 2,600 |
| Parenchyma cell of a fruit | 1,000 | $1 \times 10^8$ |

### 1. The Nucleus

In a typical animal cell the nucleus has a diameter of ~5 µm and a volume of 65 $\mu m^3$. Except at the time of cell division, it is densely and almost uniformly packed with DNA. The amount of DNA present is larger than that in bacteria as is indicated in Table 1-3. Yeast contains about three times as much genetic matter as *E. coli* and a human being or a mouse about 700 times as much. However, genes are sometimes duplicated in higher organisms and large amounts of **repetitive DNA** of uncertain significance are often present. Some amphibians have 25 times *more* DNA per cell than do humans. The fruit fly *Drosophila* contains about 13,600 functioning genes and a human being perhaps 50,000.[37]

Because of its acidic character, DNA is stained by basic dyes. Long before the days of modern biochemistry, the name **chromatin** was given to the material in the nucleus that was colored by basic dyes. At the time of cell division, the chromatin is consolidated into distinct **chromosomes** which contain, in addition to 15% DNA, about 10% RNA and 75% protein.

Nearly all of the RNA of the cell is synthesized (transcribed) in the nucleus, according to the instructions encoded in the DNA. Some of the RNA then moves out of the nucleus into the cytoplasm where it functions in protein synthesis and in some other ways. Many eukaryotic genes consist of several sequences that may be separated in the DNA of a chromosome by **intervening sequences** of hundreds or thousands of base pairs. The long RNA transcripts made from these **split genes** must be cut and **spliced** in the nucleus to form the correct messenger RNA molecules which are then sent out to the ribosomes in the cytoplasm.

Each cell nucleus contains one or more dense **nucleoli**, regions that are rich in RNA and may contain 10–20% of the total RNA of cells. Nucleoli are sites of synthesis and of temporary storage of ribosomal RNA, which is needed for assembly of ribosomes. The **nuclear envelope** is a pair of membranes, usually a few tens of nanometers apart, that surround the nucleus. The two membranes of the pair separate off a thin **perinuclear space** (Fig. 1-7). The membranes contain "pores" ~130 nm in diameter with a complex structure (see Fig. 27-8).[38,39] There is a central channel ~42 nm in diameter, which provides a route for controlled passage of RNA and other large molecules from the nucleus into the cytoplasm and also from the cytoplasm to the nucleus. Smaller ~10 nm channels allow passive diffusion of ions and small molecules.

**TABLE 1-3**
**Haploid Genome Sizes for Several Organisms**

| Organism; see footnotes for sequence information | Millions of nucleotide base pairs (Mb) | Number of chromosomes (haploid) | Estimated number of genes |
|---|---|---|---|
| *Mycoplasma genitalium*[a–c] | 0.580 | 1 | 482 |
| *Rickettsia prowazekii*[d] | 1.11 | 1 | 834 |
| *Haemophilus influenzae*[c,e–g] | 1.83 | 1 | 1,709 |
| *Methanococcus janaschii* (an archaeon)[h] | 1.66 | 1 | 1,738 |
| *Bacillus subtilis*[i] | 4.16 | 1 | |
| *Escherichia coli*[j,k] | 4.64 | 1 | 4,288 |
| *Myxococcus xanthis*[f] | 9.2 | 1 | |
| *Synechocystis* sp. (a cyanobacterium)[l] | 3.57 | 1 | 3,169 |
| *Saccharomyces cerevisiae* (a yeast)[k,m-p,t] | 13.5 | 17 | 6,241 |
| *Giardia lamblia* (a protozoan)[q] | 12 | | |
| *Plasmodium falciparum* (malaria parasite)[v] | 25-30 | 14 | |
| *Dictyostelium discoideum* (a slime mold)[r] | 34 | | |
| *Caenorhabditis elegans* (a nematode)[s,t] | 97 | 6 | 18,424 |
| Sea Urchin | 900 | | |
| *Drosophila melanogaster* (fruit fly)[u,t] | 180 | 4 | 13,601 |
| *Danio rerio* (zebrafish)[w] | 1,700 | 2.5 | |
| *Fugu rubripes* (pufferfish)[x] | 400 | | 60,000 |
| S. African lungfish[y] | 102,000 | 19 | |
| *Mus musculus* (mouse)[z] | ~3,000 | 20 | 80,000 |
| *Bos* (cow)[aa] | ~3,000 | 30 | 80,000 |
| *Homo sapiens* (human)[bb,cc] | ~3,000 | 23 | 50,000–150,000 |
| *Arabidopsis thaliana* (green plant)[dd,ee] | 115.4 | 5 | 25,498 |
| Rice[ff-hh] | 450 | 12 | |
| Maize[ii] or Wheat[ff] | ~2,700 | | |
| Lily[jj] | >100,000 | | |

## 2. The Plasma Membrane

The thin (8 nm) outer cell membrane or "plasma-lemma" (Fig. 1-7) controls the flow of materials into and out of cells, conducts impulses in nerve cells and along muscle fibrils, and participates in chemical communication with other cells. Deep infoldings of the outer membrane sometimes run into the cytoplasm. An example, is the "T system" of tubules which functions in excitation of muscle contraction (Figs. 19-7, 19-21). Surfaces of cells designated to secrete materials or to absorb substances from the surrounding fluid, such as the cells lining kidney tubules and pancreatic secretory cells, are often covered with very fine projections or **microvilli** which greatly increase the surface area. In other cases projections from one cell interdigitate with those of an adjacent cell to give more intimate contact.

## 3. Vacuoles, Endocytosis, and Lysosomes

Cells often contain vacuoles or smaller vesicles that are separated from the cytosol by a *single* membrane. Their content is often quite

[a] Fraser, C. M., and 28 other authors (1995) *Science* **270**, 397–403
[b] Goffeau, A. (1995) *Science* **270**, 445–446
[c] Brosius, J., Robison, K., Gilbert, W., Church, G. M., and Venter, J. C. (1996) *Science* **271**, 1302–1304
[d] Andersson, S. G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Pontén, T., Alsmark, U. C. M., Podowski, R. M., Näslund, A. K., Eriksson, A.-S., Winkler, H. H., and Kurland, C. G. (1998) *Nature (London)* **396**, 133–140
[e] Fleischmann, R. D., and 39 other authors (1995) *Science* **269**, 496–512
[f] He, Q., Chen, H., Kupsa, A., Cheng, Y., Kaiser, D., and Shimkets, L. J. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9584–9587
[g] Mrázek, J., and Karlin, S. (1996) *Trends Biochem. Sci.* **21**, 201–202
[h] Bult, C. J., and 39 other authors; corresponding author Venter, J. C. (1996) *Science* **273**, 1058–1073
[i] Azevedo, V., Alvarez, E., Zumstein, E., Damiani, G., Sgaramella, V., Ehrlich, S. D., and Serror, P. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6047–6051
[j] Blattner, F. R. and 16 other authors (1997) *Science* **277**, 1453–1462
[k] Winzeler, E. A., and 51 other authors (1999) *Science* **285**, 901–906
[l] Kaneko, T., and 23 other authors (1996) *DNA Res.* **3**, 109–136. See also http://www.kazusa.or.jp/cyano
[m] Dujon, B., and 107 other authors (1994) *Nature (London)* **369**, 371–378
[n] Taguchi, T., Seko, A., Kitajima, K., Muto, Y., Inoue, S., Khoo, K.-H., Morris, H. R., Dell, A., and Inoue, Y. (1994) *J. Biol. Chem.* **269**, 8762–8771
[o] Johnston, M., and 34 other authors. (1994) *Science* **265**, 2077–2082
[p] Williams, N. (1995) *Science* **268**, 1560–1561
[q] Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A., and Peattie, D. A. (1989) *Science* **243**, 75–77
[r] Loomis, W. F., and Insall, R. H. (1999) *Nature (London)* **401**, 440–441
[s] The *C. elegans* Sequencing Consortium (1998) *Science* **282**, 2012–2018 (See this article for list of authors.)
[t] Rubin, G. M., and 54 other authors. (2000) *Science* **287**, 2204–2215
[u] Adams, M. D., and 194 other authors. (2000) *Science* **287**, 2185–2195
[v] Su, X.-z, Ferdig, M. T., Huang, Y., Huynh, C. Q., Liu, A., You, J., Wootton, J. C., and Wellems, T. E. (1999) *Science* **286**, 1351–1353
[w] Postlethwait, J. H., Johnson, S. L., Midson, C. N., Talbot, W. S., Gates, M., Ballinger, E. W., Africa, D., Andrews, R., Carl, T., Eisen, J. S., Horne, S., Kimmel, C. B., Hutchinson, M., Johnson, M., and Rodriguez, A. (1994) *Science* **264**, 699–703
[x] Fishman, M. C. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10554–10556
[y] Kornberg, A., and Baker, T. A. (1992) *DNA Replication*, 2nd ed., pp. 19–21, Freeman, New York
[z] Dietrich, W. F., Copeland, N. G., Gilbert, D. J., Miller, J. C., Jenkins, N. A., and Lander, E. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10849–10853
[aa] Anonymous (1994) *Nature (London)* **368**, 167
[bb] Schuler, G. D. and 102 other authors (1996) *Science* **274**, 540–546
[cc] Koonin, S. E. (1998) *Science* **279**, 36–37
[dd] Olson, M. V. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4338–4344
[ee] The Arabidopsis Genome Initiative, (2000) *Nature* **408**, 796–815
[ff] Stevens, J. E. (1994) *Science* **266**, 1186–1187
[gg] Shimamoto, K. (1995) *Science* **270**, 1772–1773
[hh] Singh, K., Ishii, T., Parco, A., Huang, N., Brar, D. S., and Khush, G. S. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6163–6168
[ii] Carels, N., Barakat, A., and Gernardi, G. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11057–11060
[jj] Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. (1994) *Molecular Biology of the Cell*, 3rd ed., Garland, New York

acidic.[40] Small vesicles sometimes bud inward from the plasma membrane in a process called **endocytosis**. In this manner the cell may engulf particles (**phagocytosis**) or droplets of the external medium (**pinocytosis**). The resulting endocytotic vesicles or **endosomes** often fuse with **lysosomes**, which are small acidified vesicles containing a battery of enzymes powerful enough to digest almost anything in the cell. In cells that engulf bits of food (e.g., ameba) lysosomes provide the digestive enzymes. Lysosomes also take up and digest denatured or damaged proteins and may digest "worn out" or excess cell parts including mitochondria. Lysosomes are vital components of cells,[41] and several serious human diseases result from a lack of specific lysosomal enzymes.

## 4. The Endoplasmic Reticulum and Golgi Membranes

Although cytoplasm is fluid and in some organisms can undergo rapid streaming, the electron microscope

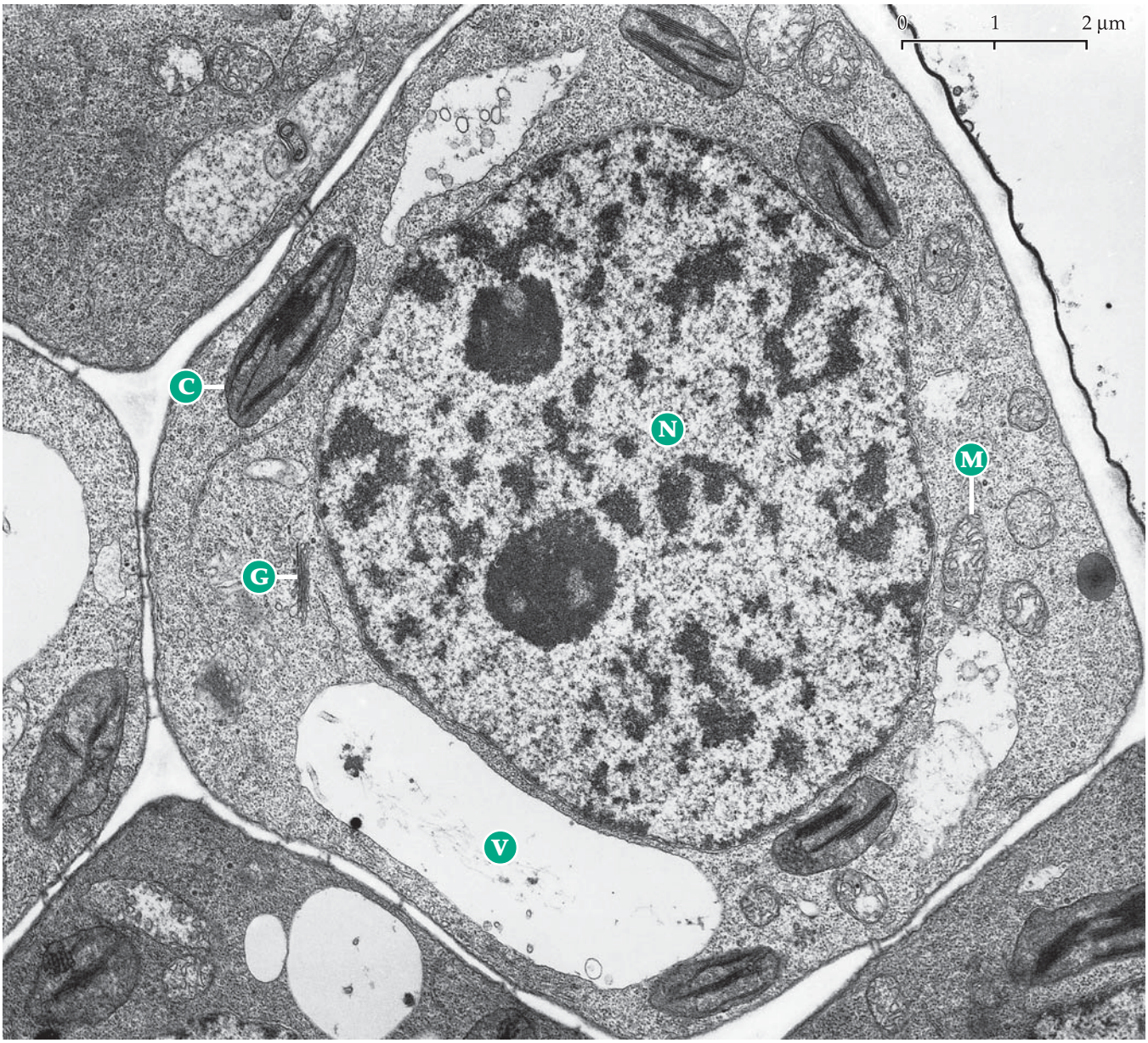
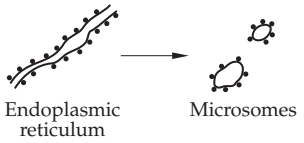



**Figure 1-7** Electron micrograph of a thin section of a young epidermal cell of a sunflower. The tissue was fixed and stained with uranyl acetate and lead citrate. Clearly visible are the nucleus (N), mitochondria (M), chloroplasts (C), a Golgi body dictyosome (G), endoplasmic reticulum, vacuole (V), cell wall, plasmodesmata, and cuticle (upper right, thin dark layer). Micrograph courtesy of H. T. Horner.

has revealed that within the liquid portion, the **cytosol**, there is a complex network of membranes known as the **endoplasmic reticulum** (ER). The membranes of the ER form tubes, vesicles, and flattened sacs called **cisternae**. The intracisternal spaces appear to connect with the perinuclear space and to a series of 3–12 flattened, slightly curved disk-shaped membranes known as the **Golgi apparatus** (Figs. 1-7, 20-8).[42,43] This organelle was first reported by Camillo Golgi in 1898.[44] Its existence was long doubted, but it is known now to play a vital role in metabolism.

The ER, the Golgi membranes, and secretion granules apparently represent an organized system for synthesis of secreted protein and formation of new membranes. Parts of the ER, the **rough endoplasmic reticulum** are lined with many ribosomes of 21–25 nm diameter. While resembling those of bacteria, these eukaryotic ribosomes are about 50% heavier ($4 \times 10^6$ Da). The **smooth endoplasmic reticulum** lacks ribosomes but proteins made in the rough ER may be modified in the smooth ER, e.g., by addition of carbohydrate chains. Small membrane vesicles break off from the smooth ER and pass to the Golgi membranes which lie close to the smooth ER on the side toward the center of the cell. Here additional modification reactions occur (Chapter 20). At the outer edges the membranes of the Golgi apparatus pinch off to form vacuoles which are often densely packed with enzymes or other proteins. These **secretion granules** move to the surface and are released from the cell. In this process of **exocytosis** the membranes surrounding the granules fuse with the outer cell membrane. The rough ER appears to contribute membrane material to the smooth ER and Golgi apparatus, while material from Golgi membranes can become incorporated into the outer cell membrane and into lysosomes. Outer mitochondrial membranes and membranes around vacuoles in plant cells may also be derived directly from the ER. Outer membrane materials are probably "recycled" by endocytosis.

The term **microsome**, frequently met in the biochemical literature, refers to small particles of 50–150 nm diameter which are mostly fragments of the ER together with some material from the plasma membrane. Microsomes are formed when cells are ground or homogenized. Upon centrifugation of the disrupted cells, nuclei and other large fragments sediment first, then the mitochondria. At very high speeds (e.g., at 100,000 times the force of gravity) the microsomes, whose masses are $10^8$–$10^9$ Da, settle. With the electron microscope we see that in the microsomes the membrane fragments have closed to give small sacs to the outside of which the ribosomes still cling:




## 5. Mitochondria, Plastids, and Peroxisomes

**Mitochondria**, complex bodies about the size of bacteria and bounded on the outside by a double membrane (Figs. 1-6 and 1-7), are present in all eukaryotic cells that use oxygen for respiration. The numbers per cell appear to vary from the *one* for certain tiny trypanosomes to as many as $3 \times 10^5$ in some oocytes. Liver cells often contain more than 1000 mitochondria apiece.[45] Study of ultrathin serial sections of a single yeast cell by electron microscopy has shown that under some growth conditions all of the yeast mitochondria are interconnected.[46] More recent evidence from new imaging procedures, e.g. using the green fluorescent protein (Box 23-A)[46a, 46b] also supports the idea that mitochondria are interconnected in a reticulum that can become fragmented under some conditions. The inner membrane of a mitochondrion is often highly folded to form the **cristae** (crests). The outer membrane is porous to small molecules but the passage of substances into and out of the inner space of the mitochondrion, known as the **matrix,** is tightly controlled by the inner membrane. Although some of the oxidative chemical activites of the cells are located in the ER and in peroxisomes, the major energy-yielding reactions for aerobic organisms are found in the mitochondria, which are also the principal site of utilization of oxygen. Within each mitochondrion is a small circular molecule of DNA whose genes encode only a few of the many proteins needed in this organelle. Also present within mitochondria are ribosomes of a size similar to those of bacteria and smaller than those lining the rough ER.

**Plastids** are organelles of plant cells that serve a variety of purposes.[45] Most important are the **chloroplasts**, the chlorophyll-containing sites of photosynthesis. Like mitochondria they contain folded internal membranes (see Fig. 23-19) and several small molecules of DNA.

Fragile organelles, the **peroxisomes** or **microbodies,** occur in many cells.[47–50] In green leaves they may occur in numbers up to one-third those of mitochondria. Peroxisomes are often about the size of mitochondria but have only a single membrane and do not contain DNA. They often contain an apparently crystalline "core." The single membrane of peroxisomes is porous to small molecules such as sucrose. This permits these organelles to be separated from mitochondria by centrifugation in a sucrose gradient where the microbodies assume a density of about 1.25 g/cm$^3$ compared to 1.19 for the impervious mitochondria.

Peroxisomes are rich in enzymes that produce and decompose hydrogen peroxide. They often make a major contribution to the oxidative metabolism of cells. In germinating oilseeds **glyoxysomes**, a type of peroxisome, contain enzymes that catalyze reactions of the biosynthetic "glyoxylate pathway" of metabolism.[51] Organelles that resemble peroxisomes in appearance

but which are functionally more closely related to mitochondria are the **hydrogenosomes** of anaerobic protozoa.[52] As the name suggests, these organelles are the site of formation of molecular hydrogen, a common product of anaerobic metabolism.

## 6. Centrioles, Cilia, Flagella, and Microtubules

Many cells contain **centrioles**,[53] little cylinders about 0.15 µm in diameter and 0.5 µm long, which are *not* enclosed by membranes. Each centriole contains a series of fine **microtubules** of 25 nm diameter. A pair of centrioles are present near the nucleus in most animal cells and play an important role in cell division. Together with surrounding materials they form the **centrosome**. However, centrioles have never been observed in plant cells.

Related in structure to centrioles are the long **flagella** and shorter **cilia** (the two words are virtually synonymous) which are commonly present as organelles of locomotion in eukaryotic cells. Stationary cells of our own bodies also often have cilia. For example, there are $10^9$ cilia/cm$^2$ in bronchial epithelium.[54] Modified flagella form the receptors of light in our eyes and of taste in our tongues. Flagella and cilia have a diameter of about 0.2 µm and a characteristic internal structure. Eleven hollow microtubules of ~24 nm diameter are usually arranged in a "9 + 2" pattern with nine pairs of fused tubules surrounding a pair of single tubules (Figs. 1-8 and 19-23). Each microtubule resembles a bacterial flagellum in appearance, but there are distinct and significant chemical differences. The **basal body** of the flagellum, the **kinetosome** (Fig. 1-8), resembles a centriole in structure, dimensions, and mode of replication. Recently a small 6–9 megabase pair DNA has been found in basal bodies of the protozoan *Chlamydomonas*.[55,56]

Microtubules similar to those found in flagella are also present in the cytoplasm. Together with thinner **microfilaments** of several kinds they form an internal **cytoskeleton** that provides rigidity to cells.[58,59] Microtubules also form the "spindle" of dividing cells. In nerve axons (Chapter 30) the microtubules run parallel to the length of the axons and are part of a mechanical transport system for cell constituents.

## 7. Cell Coats, Walls, and Shells

Like bacteria, most cells of higher plants and animals are surrounded by extracellular materials. Plants have rigid walls rich in cellulose and other carbohydrate polymers. Outside surfaces of plant cells are covered with a **cuticle** containing layers of a polyester called **cutin** and of wax (Fig. 1-6). Surfaces of animal cells are usually lined with carbohydrate molecules which are attached to specific surface proteins to form **glycoproteins**. Spaces between cells are filled with such "cementing substances" as **pectins** in plants and **hyaluronic acid** in animals. Insoluble proteins such as **collagen** and **elastin** surround connective tissue cells. Cells that lie on a surface (epithelial and endothelial cells) are often lined on one side with a thin, collagen-containing **basement membrane** (Figs. 1-6 and 8-31). Inorganic deposits such as calcium phosphate (in bone), calcium carbonate (eggshells and spicules of sponges), and silicon dioxide (shells of diatoms) are laid down, often by cooperative action of several or many cells.

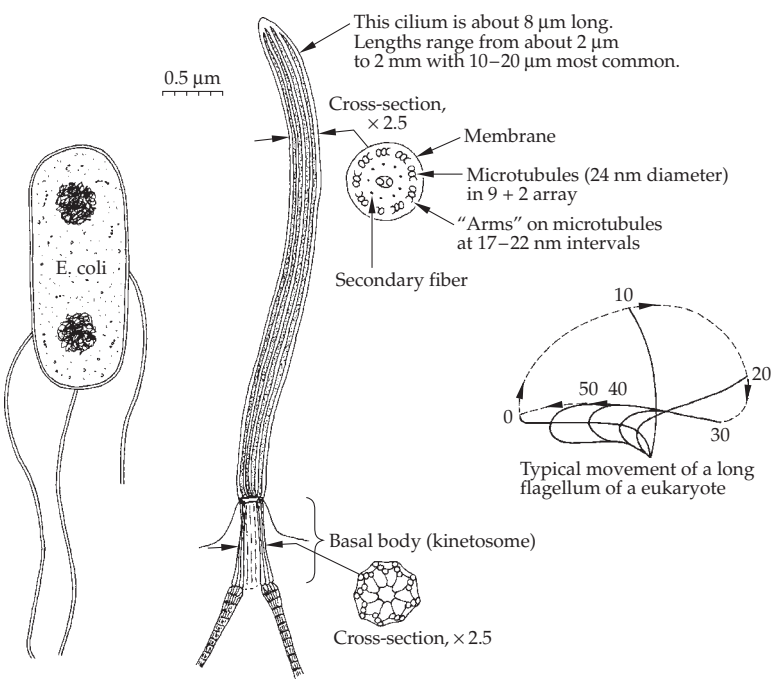



**Figure 1-8** Structure of cilia and flagella of eukaryotes. After P. Satir.[57]

## C. Inheritance, Metabolic Variation, and Evolution of Eukaryotes

The striking differences between eukaryotic and prokaryotic cells have led to many speculations about the evolutionary relationship of these two great classes of living organisms. A popular theory is that mitochondria, which are characteristic of most eukaryotes, arose from aerobic bacteria. After cyanobacteria had developed and oxygen

## BOX 1-D INHERITED METABOLIC DISEASES

In 1908 Archibald Garrod[a,b] proposed that **cystinuria** (Chapter 8) and several other defects in amino acid and sugar metabolism were "inborn errors of metabolism", i.e. inherited diseases. Since that time the number of recognized genetic defects of human metabolism has increased at an accelerating rate to ~4000.[c–e] Hundreds of other genetic problems have also been identified. For over 800 of these the defective gene has been mapped to a specific chromosome.[f] An example is **sickle cell anemia** (Box 7-B) in which a defective hemoglobin differs from the normal protein at one position in one of its constituent polypeptide chains. Many other defects involve loss of activity of some important enzyme.

Most genetic diseases are rare, affecting about one person in 10,000. However, **cystic fibrosis** affects one in 2500. There are so many metabolic diseases that over 0.5% of all persons born may develop one. Many die at an early age. A much greater number (>5%) develop such conditions as diabetes and mental illness which are, in part, of genetic origin. Since new mutations are always arising, genetic diseases present a problem of continuing significance.

At what rate do new mutations appear? From the haploid DNA content (Table 1-2) we can estimate that the total coding capacity of the DNA in a human cell exceeds two million genes (actually two million *pairs* of genes in diploid cells). However, only a fraction of the DNA codes for proteins. There are perhaps 50,000 pairs of structural genes in human DNA. The *easily detectable* rate of mutation in bacteria is about $10^{-6}$ per gene, or $10^{-9}$ per base per replication.[g] As a result of sophisticated "proofreading" and repair systems, it may be as low as $10^{-10}$ per base in humans.[h] Thus, in the replication of the $3 \times 10^9$ base pairs in diploid human chromosomes we might anticipate about one mistake per cell division. Only about 1/50 of these would be in structural genes and potentially harmful. Thus, if there are $10^{16}$ division cycles in a normal life span[h] each parent may pass on to future generations about 2 mutations in protein sequences. The ~$10^{14}$ body cells (somatic cells) also undergo mutations which may lead to cancer and to other problems of aging. Most mutations may be harmless or nearly so and a few may be beneficial.

*Photomicrograph of human male metaphase chromosomes. © Photo Researchers*

However, many are damaging and some are *lethal*. If a mutation is lethal, a homozygote will not survive and will be lost in an early (and usually undetected) spontaneous abortion. Healthy individuals carry as many as ten lethal recessive mutations as well as at least 3–5 autosomal recessive mutations of a seriously harmful type. Harmful dominant mutations are also frequent in the population. These include an elevated lipoprotein content of the blood and an elevated cholesterol level which are linked to early heart disease.

Biochemical disorders are also important because of the light they shed on metabolic processes. No other species is observed as carefully as *Homo sapiens*. As a consequence frequent reference will be made to genetic diseases throughout the book. A goal is to find ways to prevent or ameliorate the effects of these disorders. For example, in the treatment of **phenylketonuria** (Chapter 25) or of **galactosemia** (Chapter 20), a change in the diet can prevent irreversible damage to the brain, the organ most frequently affected by many of these diseases. Injection of a missing enzyme is giving life to victims of Gaucher's disease (Chapter 20). In many other cases no satisfactory therapy is presently available, but the possibilities of finding some way to supply missing enzymes or to carry out "genetic surgery" are among the most exciting developments of contemporary medical biochemistry (Chapter 26).

---

[a] Garrod, A. E. (1909) *Inborn Errors of Metabolism*, Oxford, London
[b] Bearn, A. G. (1993) *Archibald Garrod and the Individuality of Man*, Oxford, New York
[c] Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds. (1995) *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., McGraw-Hill, New York
[d] Davies, K. E. (1992) *Molecular Basis of Inherited Disease*, Oxford, New York
[e] McKusick, V. A. (1994) *Mendelian Inheritance in Man*, 11th ed., Johns Hopkins Univ. Press, Baltimore, Maryland
[f] McKusick, V. A., Amberger, J., and Steinberg, J. (1994) *J. NIH Res.* **6**, 115–134
[g] Watson, J. D. (1976) *Molecular Biology of the Gene*, 3rd ed., Benjamin, Menlo Park, California (p. 254)
[h] Koshland, D. E., Jr. (1994) *Science* **266**, 1925

had become abundant, a **symbiotic** relationship could have arisen in which small aerobic bacteria lived within cells of larger bacteria that had previously been obligate anaerobes. Sequence similarities of proteins suggest that these symbionts may have been related to present-day methanogens[60] and thermophilic sulfur bacteria.[61] The aerobes presumably used up any oxygen present, protecting the surrounding anaerobic organisms from its toxicity. The relationship became permanent and led eventually to the mitochondria-containing eukaryotic cell.[62–65] Further symbiosis with cyanobacteria or prochlorophytes could have led to the chloroplasts of the eukaryotic plants.

A fact that supports such ideas is the existence among present-day organisms of many endosymbiotic relationships. For example, the green paramecium (*Paramecium bursaria*) contains, within its cytoplasm, an alga (*Chlorella*), a common green plant that is quite capable of living on its own. Perhaps by accident it took up residence within the paramecium.[62] Some dinoflagellates (Fig. 1-9) contain endosymbiotic cyanobacteria[66] and recently a ciliate that contains endosymbiotic purple photosynthetic bacteria has been discovered.[67] These bacteria do not produce $O_2$ but utilize products of the host ciliates' metabolism such as acetate, lactate, and $H_2$ as electron donors for photosyntheses. They also utilize $O_2$ for respiration and may protect their hosts from the toxicity of $O_2$, just as may have happened in the distant past. According to this theory the symbionts would eventually have lost their photosynthetic ability and have become mitochondria. The relationship of mitochondria to bacteria is also supported by many biochemical similarities.

Fossils of bacteria and blue-green algae have been obtained from rocks whose age, as determined by geochemical dating, is more than $3 \times 10^9$ years.[68,69] However the first eukaryotic cells may have appeared about $1 \times 10^9$ years ago[70] and started to evolve into the more than one million species that now exist.[1,71,72]

## 1. A Changing Genome

How is it possible for the genome of an organism to increase in size as it evolved from a lower form to a higher one? Simple mutations that cause alterations in protein sequences could lead to changes in form and behavior of the organisms but could not, by themselves, account for the increase in genetic material that accompanied evolution. As a result of new techniques of genetic mapping and determining the sequence of nucleotides in DNA we are rapidly acquiring a detailed knowledge of the organization of the genome. It has been found that genes are often present as duplicate but not entirely identical copies. This suggests that there are mechanisms by which cells can acquire extra copies of one or more genes. Indeed it seems probable that at some time in the past the entire genome of bacteria was doubled and that it was later doubled again.[73] Evidence for this is that the masses of bacterial chromosomes group around values of 0.5, 1.4, and $2.7 \times 10^9$ Da. Genes can also be duplicated during the process of genetic recombination, which is discussed in Chapter 27. In addition, the size of the genome may have increased by incorporation of genetic material from extrachromosomal plasmids.

A possible advantage to a cell possessing an extra copy of a gene is that the cell would survive even when mutations rendered unusable the protein encoded by one of the copies. As long as one of the genes remained "good," the organism could grow and reproduce. The extra, mutated gene could be carried for many generations. As long as it produced only harmless, nonfunctioning proteins there might be little selection pressure to eliminate it and it might undergo repeated mutations. After many mutations and many generations later, the protein for which it coded could prove useful to the cell in some new way.

An example of evolution via gene duplication is provided by the oxygen-carrying proteins of blood. It appears that about a billion years ago, the gene for an ancestral **globin**, the protein of hemoglobin, was doubled. One gene evolved into that of present-day **globins** and the other into the gene of the muscle protein **myoglobin**. Still later, the globin gene again doubled leading to the present-day α and β chains of hemoglobin (Chapter 7). These are two distinctly different but related protein subunits whose genes are not even on the same chromosome. To complicate the picture further, most human beings have two or more copies of their α chain gene[74] as well as genes for fetal and embryonic forms of hemoglobin. However, some populations have lost one or more α chain genes. Thus, the genome changes in many details, even today.

## 2. Genetic Recombination, Sex, and Chromosomes

Bacteria usually reproduce by simple fission. The single DNA molecule of the chromosome is duplicated and the bacterium divides, each daughter cell receiving an identical chromosome. However, genetic recombination, which is accomplished in several ways by bacteria (Chapter 27), provides a deliberate process for mixing of genes. This process has been most fully developed in eukaryotic organisms that undergo sexual reproduction. The growth of a multicelled individual begins with the fusion of two haploid **gametes**, an egg and a spermatozoon. Each gamete carries a complete set of genetic instructions, and after the nuclei fuse the fertilized egg or **zygote** is **diploid**. Each diploid cell contains *two* complete sets of genetic blueprints of quite different origin. Even if a gene from one parent

is defective, the chances are that the gene from the other parent will be good. Sexual reproduction and the associated genetic recombination also provide a means for mixing of genes.

When eukaryotic cells prepare to divide in the process called **mitosis** (Fig. 26-11), the DNA molecules of the nucleus, which become spread out through a large volume, coil and fold. Together with proteins and other molecules they form the compact bodies known as chromosomes. Some organisms, such as *Ascaris* (a roundworm), have only two chromosomes, a **homologous pair**, one inheritied from the father and one from the mother. Both chromosomes divide in every mitotic cell division so that every cell of the organism has the homologous pair. Higher organisms usually have a larger number of chromosomes. Thus, humans have 23 homologous pairs. The mouse has 20, the toad 11, onions 8, mosquitos 3, and *Drosophila* 4. Human chromosomes vary in size but are usually 4–6 μm long and ~1 μm in diameter.

By the successive divisions of mitosis, a single fertilized eukaryotic egg cell can grow to an adult. Less than 50 successive mitotic divisions will produce the ~$10^{14}$ cells of a human. However, formation of gametes, which are haploid, requires the special process of **meiosis** (Fig. 26-12), by which the number of chromosomes is divided in half. During meiosis one chromosome of each of the homologous pairs of the diploid cell is passed to each of the gametes that are formed. In an organism such as *Ascaris*, which contains only a single pair of chromosomes, a gamete receives either the chromosome of maternal origin or that of paternal origin but not both. In organisms that have several pairs of chromosomes, one chromosome of each pair is passed to the gamete in a random fashion during meiosis. Most gametes receive some chromosomes of maternal and some of paternal origin. An important feature of meiosis is the genetic recombination that occurs during **crossing-over**. In this process, the strands of DNA are cut and genetic material is exchanged between the chromosomes of maternal and paternal origin. Thus, crossing-over breaks the **linkage** between genes and provides for greater variability in the offspring than would otherwise be possible. Each of us receives half of our genes from our mother and half from our father, but some of these genes have been inherited from each grandparent on both sides of the family, some from each great-grandparent, etc.

Many genes are passed down through many generations without substantial change, but others are evidently designed to be scrambled readily within somatic cells. Cell surface proteins[75] and antibody molecules are among the proteins whose genes undergo alteration during growth and differentiation of the tissues of the body (Chapter 32).

### 3. Haploid and Diploid Phases

In human beings and other higher animals, meiosis leads directly to formation of the gametes, the egg and sperm cells. These fuse to form a diploid nucleus and the adult develops by repeated mitosis of the diploid cells. While meiosis also occurs in the life cycle of all eukaryotic creatures, it is not always at a point corresponding to that in the human life cycle. Thus, the cells of many protozoa and of fungi are ordinarily haploid. When two haploid nuclei fuse to form a diploid cell, meiosis quickly occurs to produce haploid individuals again. Among lower plants and animals there is often an alternation of haploid and diploid phases of the life cycle. For example, gametes of ferns fall to the ground and germinate to form a low-growing green mosslike haploid or **gametophyte** form. The latter produces motile haploid gametes which fuse to a diploid zygote that grows into the larger and more obvious **sporophyte** form of the fern.

It is presumably the ability to survive as a heterozygote, even with one or more highly deleterious mutations, that has led to the dominance of the diploid phase in higher plants and animals.[76] However, to the biochemical geneticist organisms with a haploid phase offer experimental advantages because recessive mutants can be detected readily.

## D. Survey of the Protists

Unicellular eukaryotes have traditionally been grouped together with multicellular organisms in which all cells have similar functions, with little or no differentiation into tissues, as the kingdom **Protista**.[77,78] The fungi may also be included or may be regarded as a separate kingdom.[79] With present-day emphasis on DNA sequence comparisons the traditional classification is changing, however.[26]

### 1. Protozoa

Among the best known of the animal-like protista is the **ameba** (subphylum Sarcodina or Rhizopoda). The most striking feature of the ameba (Fig. 1-9) is its method of locomotion, which involves the transformation of cytoplasm from a liquid state to a semisolid gel. As the ameba moves, the cytoplasm at the rear liquifies and flows to the front and into the extending pseudopodia where it solidifies along the edges. The ameba poses several important biochemical questions: What chemistry underlies the reversible change from liquid to solid cytoplasm? How can the cell membranes break and reform so quickly when an ameba engulfs food particles?[80]

Relatives of the ameba include the **Radiolaria**,

marine organisms of remarkable symmetry with complex internal skeletons containing the carbohydrate polymer chitin together with silica ($SiO_2$) or strontium sulfate. The **Foraminifera** deposit external shells of calcium carbonate or silicon dioxide. Over 20,000 species are known and now as in the distant past their minute shells fall to the bottom of the ocean and form limestone deposits.

Tiny ameboid parasites of the subphylum **Sporozoa** attack members of all other animal phyla. Several genera of **Coccidia** parasitize rabbits and poultry causing enormous damage. Humans are often the victims of species of the genus *Plasmodium* (Fig. 1-9) which invade red blood cells and other tissues to cause **malaria**, one of our most serious ailments on a worldwide basis.[81–84] Throughout history malaria has probably killed more persons than any other disease. *Toxoplasma gondii* is another parasite which, in its haploid phase, is found throughout the world in wild animals and in humans. Although its presence usually elicits no symptoms, it sometimes causes blindness and mental retardation in children and can be fatal to persons with AIDS. Its sexual cycle occurs exclusively in cats.[85,86]

Another subphylum of protozoa, the Mastigophora, are propelled by a small number of flagella and are intermediate between animals and the algae. One of these is *Euglena viridis*, a small freshwater organism with a long flagellum in front, a flexible tapered body, green chloroplasts, and a light-sensitive "eyespot" which it apparently uses to keep itself in the sunshine (Fig. 1-9). Euglena is also able to live as a typical animal if there is no light. Treatment with streptomycin (Box 20-B) causes *Euglena* to lose its chloroplasts and to become an animal permanently. The **dinoflagellates** (Fig. 1-9), some colorless and some green, occur in great numbers among the plankton of the sea. *Giardia lamblia* is a troublesome intestinal parasite.

The **hemoflagellates** are responsible for some of our most terrible diseases. Trypanosomes (genus *Trypanosoma*) invade the cells of the nervous system causing African sleeping sickness. Mutating their surface proteins frequently by gene-scrambling mechanisms, these and other parasites are able to evade the immune response of the host.[87,88] For the same reason it is difficult to prepare vaccines against them. Other flagellates live in a symbiotic relationship within the alimentary canals of termites (Fig. 1-9) and roaches. Termites depend upon bacteria that live within the cells of these symbiotic protozoans to provide the essential enzymes needed to digest the cellulose in wood.

Members of the subphylum Ciliophora, structurally the most complex of the protozoa, are covered with a large number of cilia which beat together in an organized pattern.[89] The following question immediately comes to mind: How

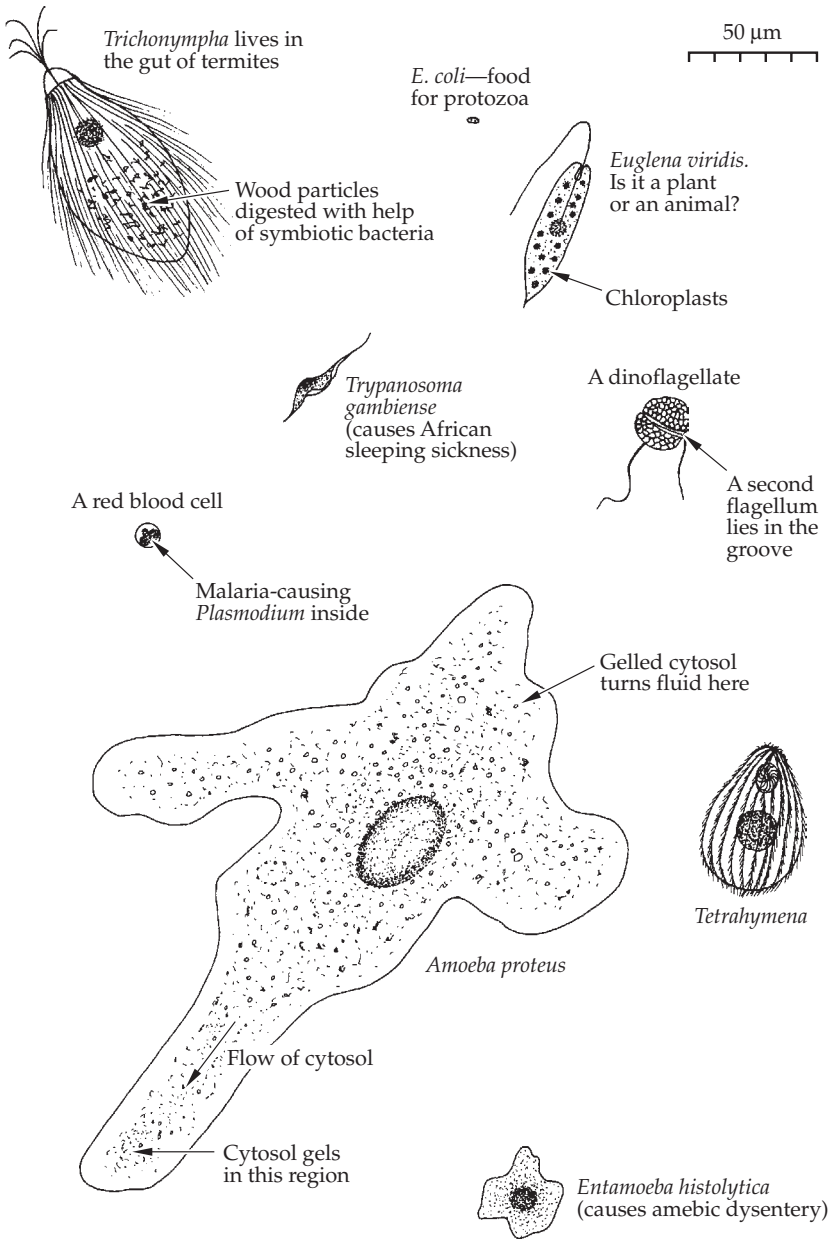



**Figure 1-9** A few well-known protists.

are the cilia able to communicate with each other to provide this organized pattern? Two ciliates that are often studied by biochemists are *Tetrahymena* (Fig. 1-9), one of the simplest, and *Paramecium*, one of the more complex.

The **Myxomycetes** or "slime molds" are more closely related to protozoa than to fungi.[90] Members of the family Acrasieae, the best studied member of which is *Dictyostelium discoideum*, start life as small amebas. After a time, when the food supply runs low, some of the amebas begin to secrete pulses of a chemical attractant **cyclic AMP**. Neighboring amebas respond to the pulses of cyclic AMP by emitting their own pulses about 15 s later, then moving toward the original source.[91,92] The ultimate effect is to cause the amebas to stream to centers where they aggregate and form fungus-like fruiting bodies. Asexual spores are formed and the life cycle begins again. Other Myxomycetes grow as a multinucleate (diploid) **plasmodium** containing millions of nuclei but no individual cell membranes. *Physarum polycephalum*, a species whose plasmodium may spread to a diameter of 30 cm, has become popular with biochemists. The 800,000 nuclei per square millimeter all divide synchronously.

## 2. Fungi

Lacking photosynthetic ability, living most often in soil but sometimes in water, the fungi are represented by almost half as many species (~$10^5$) as are the vascular plants.[93] The distinguishing characteristics of fungi are the lack of chlorophyll and growth as a series of many branched tubules (usually 6–8 μm diameter), the **hyphae**, which constitute the **mycelium**. The hyphae are not made up of separate cells but contain a mass of protoplasm with many nuclei. Only occasional septa divide the tubules. Most fungi are saprophytic, living on decaying plants or animal tissues. However, others are parasites that produce serious and difficult-to-treat infections in humans. An important medical problem is the lack of adequate antibiotics for treating fungal infections (mycoses).[94–96] On the other hand, fungi produce important antibiotics such as **penicillin**. Still others form some of the most powerful toxins known!

The lower fungi or **Phycomycetes** include simple aquatic molds and mildew organisms. Higher fungi are classified as **Ascomycetes** or **Basidiomycetes** according to the manner in which the sexual spores are born. In the Ascomycetes these spores are produced in a small sac called an **ascus** (Fig. 1-10). Each ascus contains four or eight spores in a row, a set of four representing the results of a single pair of meiotic divisions. A subsequent mitotic division will give eight spores. This is one of the features that has made *Neurospora crassa* (Fig. 1-10) a favorite subject for genetic studies.[97] The ascospores can be dissected out in order from the ascus and cultivated separately to observe the results of crossing-over during meiosis.

*Neurospora* also reproduces via haploid spores called **conidia**. The haploid mycelia exist as two mating types and conidia or mycelia from one type can fertilize cells in a special body (the protoperithecium) of the other type to form zygotes. The latter immediately undergo meiosis and mitosis to form the eight ascospores. Among other Ascomycetes are the highly prized edible truffles and morels. However, most mushrooms and puffballs are fruiting bodies of Basidiomycetes. Other Basidiomycetes include the **rusts**, which cause enormous damage to wheat and other grain crops.

Yeasts are fungi adapted to life in an environment of high sugar content and which usually remain unicellular and reproduce by budding (Fig. 1-10). Occasionally the haploid cells fuse in pairs to form diploid cells and sexual spores. Some yeasts are related to the Ascomycetes, others to Basidiomycetes. *Saccharomyces cerevisiae*, the organism of both baker's and brewer's yeast, is an Ascomycete. It can grow indefinitely in either the haploid or diploid phase. The genetics and biochemistry of this yeast have been studied extensively.[98–102] The genome is relatively small with $13.5 \times 10^6$ base pairs in 17 chromosomes. The sequence of the 315,000 base pairs of chromosome III was determined in 1992[101,102] and the sequence of the entire genome is now known.[103]

Fungi often grow in symbiotic association with other organisms. Of special importance are the **mycorrhizae** (fungus roots) formed by colonization of fine roots by beneficial soil fungi. Almost all plants of economic importance form mycorrihizae.[104]

## 3. Algae

Algae are chlorophyll-containing eukaryotic organisms which may be either unicellular or colonial.[105] The colonial forms are usually organized as long filaments, either straight or branched, but in some cases as blades resembling leaves. However, there is little differentiation among cells. The gold-brown, brown, and red algae contain special pigments in addition to the chlorophylls.

The euglenids (**Euglenophyta**) and dinoflagellates (**Pyrrophyta**), discussed in the protozoa section, can equally well be regarded as algae. The bright green **Chlorophyta**, unicellular or filamentous algae, are definitely plants, however. Of biochemical interest is *Chlamydomonas*, a rather animal-like creature with two flagella and a carotenoid-containing eyespot or **stigma** (Fig. 1-11). *Chlamydomonas* contains a single chloroplast. The "pyrenoid", a center for the synthesis of starch, lies, along with the eyespot, within the chloroplast. The organism is haploid with "plus" and "minus" strains

and motile gametes. Zygotes immediately undergo meiosis to form haploid spores. With a well-established genetic map, *Chlamydomonas* is another important organism for studies of biochemical genetics.[106]

The filamentous *Ulothrix* shows its relationship to the animals through formation of asexual spores with four flagella and biflagellate gametes. Only the zygote is diploid. On the other hand, the incomparably beautiful *Spirogyra* (Fig. 1-11) has no motile cells. The ameboid male gamete flows through a tube formed between the two mating cells, a behavior suggesting a relationship to higher green plants.

Some unicellular algae grow to a remarkable size. One of these is *Acetabularia* (Fig. 1-11), which lives in the warm waters of the Mediterranean and other tropical seas. The cell contains a single nucleus which lies in the base or rhizoid portion. In the mature alga, whose life cycle in the laboratory is 6 months, a cap of characteristic form develops. When cap development is complete, the nucleus divides into about $10^4$ secondary nuclei which migrate up the stalk and out into the rays of the cap where they form cysts. After the cap decays and the cysts are released, meiosis occurs and the flagellated gametes fuse in pairs to form zygotes which again grow into diploid algae. Because of its large size and the location of the nucleus in the base, the cells can be cut and grafted. Nuclei can be removed or transplanted and growth and development can be studied in the presence or absence of a nucleus.[107–110] The green algae **Volvox** live in wheel-like colonies of up to several thousand cells and are useful for biochemical studies of differentiation.[111]

Look through the microscope at almost any sample of algae from a pond or aquarium and you will see little boatlike **diatoms** slowly gliding through the water. The most prominent members of the division Chrysophyta, diatoms are characterized by their external "shells" of silicon dioxide. Large and ancient deposits of diatomaceous earth contain these durable silica skeletons which are finely marked, often with beautiful patterns (Fig. 1-11). The slow motion of diatoms is accomplished by streaming of protoplasm through a groove on the surface of the cell. Diatoms are an important part of marine plankton, and it is estimated that three-fourths of the organic material of the world is produced by diatoms and dinoflagellates. Like the brown algae, Chrysophyta contain the pigment **fucoxanthin**.

Other groups of algae are the brown and red marine algae or seaweed. The former (**Phaeophyta**) include the giant kelps from which the polysaccharide **algin** is obtained. The **Rhodophyta** are delicately branched plants containing the red pigment **phycoerythrin**. The polysaccharides, **agar** and **carrageenin**,

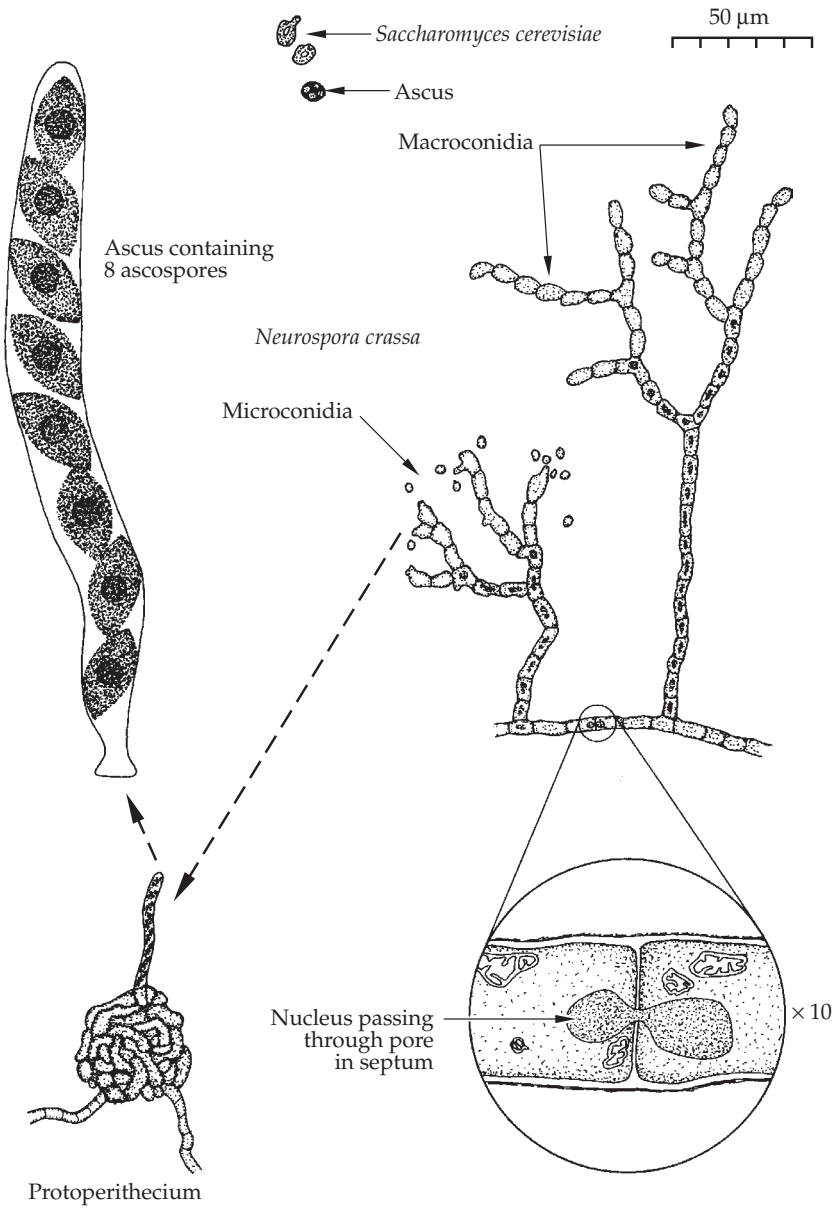



**Figure 1-10** Two frequently studied fungi. Top (including ascus): the yeast *Saccharomyces cerevisiae*. Below: *Neurospora crassa* showing various stages. After J. Webster.[93]

a popular additive to chocolate drinks and other foods, come from red algae.

Symbiotic associations of fungi with either true algae or with cyanobacteria are known as **lichens**. Over 15,000 varieties of lichens grow on rocks and in other dry and often cold places. While the algae appear to benefit little from the association, the fungi penetrate the algae cells and derive nutrients from them.[112] Although either of the two partners in a lichen can be cultured separately, the combination of the two is capable of producing special pigments and phenolic substances known as **depsides** which are not formed by either partner alone.

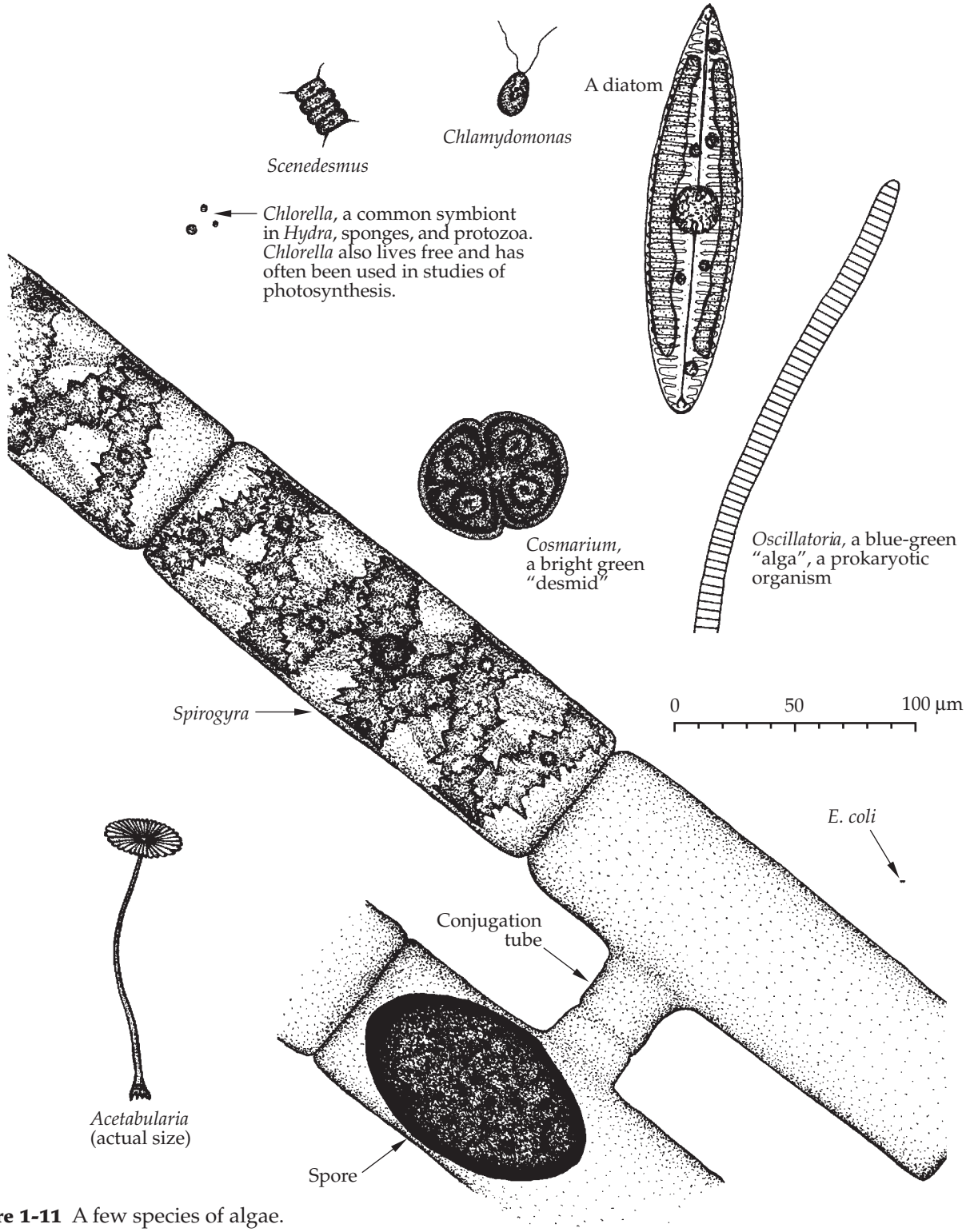



**Figure 1-11** A few species of algae.

## E. The Variety of Animal Forms

In this section, we will consider only a few biochemical and other aspects of multicellular animals or **Metazoa**. The sudden appearance of a large number of Metazoans about 0.5 x $10^9$ years ago[113,114] may have been an outcome of the appearance of split genes (see Section B, 1). As a result of gene duplication the coding pieces of split genes, the **exons**, could be moved to new locations in a chromosome where they could have become fused with other pieces of DNA to form entirely new genes.[115]

### 1. Major Groups of Multicellular Animals

The simplest metazoa are tiny symbiotic worms of the phylum (or subkingdom) **Mesozoa**, which live in the kidneys of deep sea-dwelling cephalopods (octopi and squid). Each worm is made up of only 25 cells in a single layer enclosing one or a small number of elongated axial cells (Fig. 1-12). Mesozoa have been regarded as parasitic, but they appear to facilitate excretion of $NH_3$ by the host through acidification of the urine.[118,119]

**Porifera** or sponges are the most primitive of multicelled animals.[120] They lack distinct tissues but contain several specialized types of cells. The body is formed by stationary cells that pump water through the pores to bring food to the sponge. Within the body **amebocytes** work in groups to form the **spicules** of calcium carbonate, silicon dioxide, or the protein **spongin** (Fig. 1-12). Sponges appear to lack a nervous system.

Individuals of the next most complex major phylum, **Cnidaria** (formerly Coelenterata), are radially symmetric with two distinct cell layers, the **endoderm** and **ectoderm**. Many species exist both as a polyp or **hydra** form (Fig. 1-13) and as a **medusa** or jellyfish. The jellyfish apparently has no brain but the ways in which its neurons interconnect in a primitive radial net are of interest. The Cnidaria have a very simple body form with remarkable regenerative powers. The freshwater hydra, a creature about 1 cm long (Fig. 1-13), contains a total of ~$10^5$ cells. A complete hydra can be regenerated from a small piece of tissue if the latter contains some of both the inner and the outer cell layers.[121,122]

The body of flatworms (phylum **Platyhelminthes**) consists of two external cell layers (endoderm and ectoderm) with a third layer between. A distinct excretory system is present. In addition to a nerve net resembling that of the Cnidaria, there are a cerebral ganglion and distinct eyes. One large group of flatworms, the **planarians** (typically about 15 mm in length, Fig. 1-14), inhabit freshwater streams. They are said to be the simplest creatures in which *behavior* can be studied.

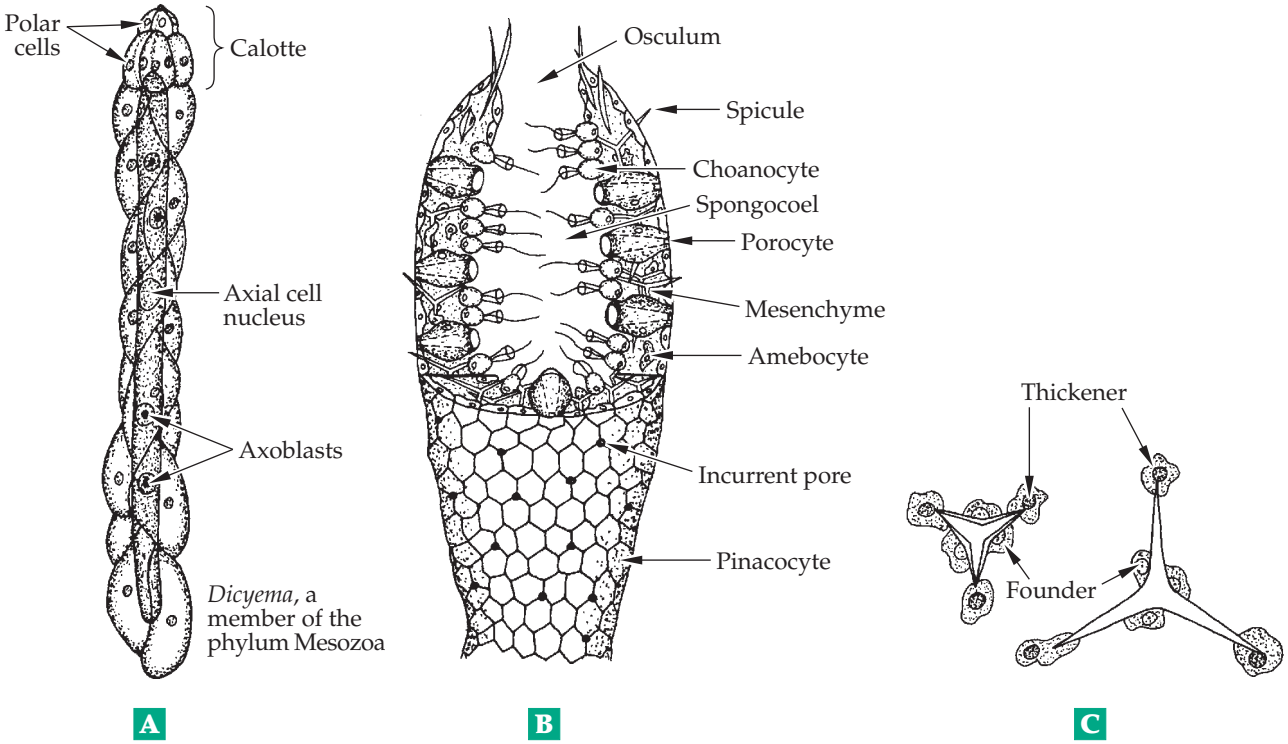



**Figure 1-12** Some lower forms of Metazoa. (A) Mesozoa (25 cells). After C. P. Hickman.[116] (B) A small asconoid sponge. After C. A. Villee, W. F. Walker, Jr., and R. D. Barnes.[117] (C) Ameboid cells of a sponge forming spicules. After Hickman.

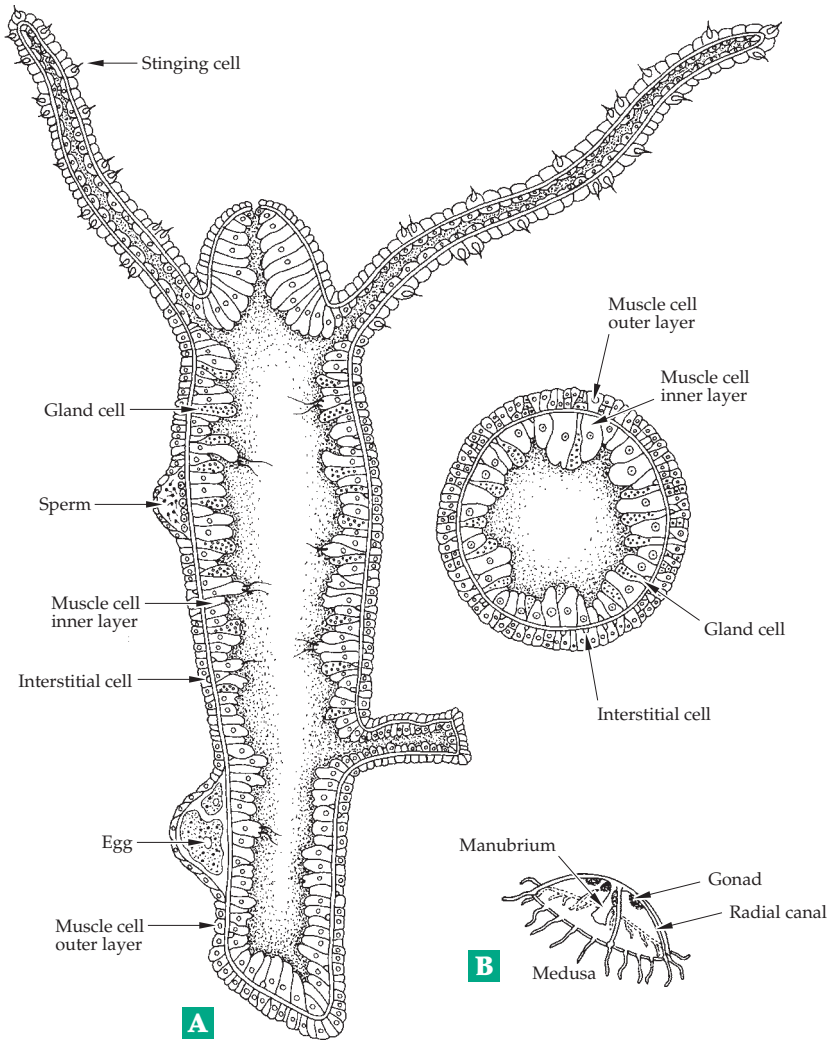



**Figure 1-13** (A) Hydra. After Loomis.[123] (B) The medusa stage of *Obelia*, a hydroid coelenterate.[123a]

Many parasitic flatworms (tapeworms and flukes) attack higher organisms.[124] Among them are the **Schistosoma**, tiny worms that are transmitted to humans through snails and which attack the blood vessels. The resulting **schistosomiasis** is one of the most widespread debilitating diseases on earth today, affecting 200 million people or more.[125,126]

The roundworms (**Nematoda**)[127–129] have, in addition to the **enteron** (alimentary tract), a separate body cavity. Free-living nematodes abound in water and soil but many species are parasitic. They do enormous damage to plants and to some animal species. Trichina, hookworms[129a], and filaria worms attack humans. However, in the laboratory the 1-mm-long, 810-cell nematode, *Caenorhabditis elegans* (Fig. 1-14) has become an important animal. In 1963 Sydney Brenner launched what has become a worldwide effort to make this tiny worm the equivalent in the animal kingdom of *E. coli* in the bacterial world.[129] The $10^8$ nucleotides in the worm's six chromosomes contain ~13,600 genes. *C. elegans* has become an important animal in which to study differentiation. Already the exact lineage of every cell has been traced, as has every connection among the 302 neurons in the animal's nervous system. The related **rotifers**,[130] with whirling "wheels" of cilia on their heads (Fig. 1-14) and transparent bodies, are a delight to the microscopist. Like nematodes, they are "cell constant" organisms. The total number of cells in the body is constant as is that in almost every part of every organ. Part of the developmental plan of such organisms is a "programmed cell death" (Chapter 32).

The **Annelida** (segmented worms)[131] are believed to be evolutionary antecedents of the arthropods. Present-day members include earthworms, leeches, and ~$10^5$ species of marine **polychaetes**. Annelids have a true body cavity separate from the alimentary canal and lined by a peritoneum. They have a well-developed circulatory system and their blood usually contains a type of hemoglobin.

About $10^6$ species of **arthropods** (80% of all known animals) have been described. Most are very small.[72] These creatures, which have a segmented exoskeleton of **chitin** and other materials, include the horseshoe crabs, the Arachnida (scorpions, spiders, and mites), the Crustacea, Myriopoda (centipedes and millipedes), and the Insecta. Important biochemical problems are associated with the development and use of insecticides and with our understanding of the metamorphosis that occurs during the growth of arthropods.[132] The fruit fly *Drosophila melanogaster* has provided much of our basic knowledge of genetics and continues to be the major species in which development is studied.[133–134a]

Among the molluscs (phylum **Mollusca**) the squids and octopuses have generated the most interest among biochemists. The neurons of squid contain giant axons, the study of which has led to much of our knowledge of nerve conduction. Octopuses show signs of intelligence not observed in other invertebrates whose nervous reactions seem to be entirely "preprogrammed." The brains of some snails contain only $10^4$ neurons, some of which are unusually large. The

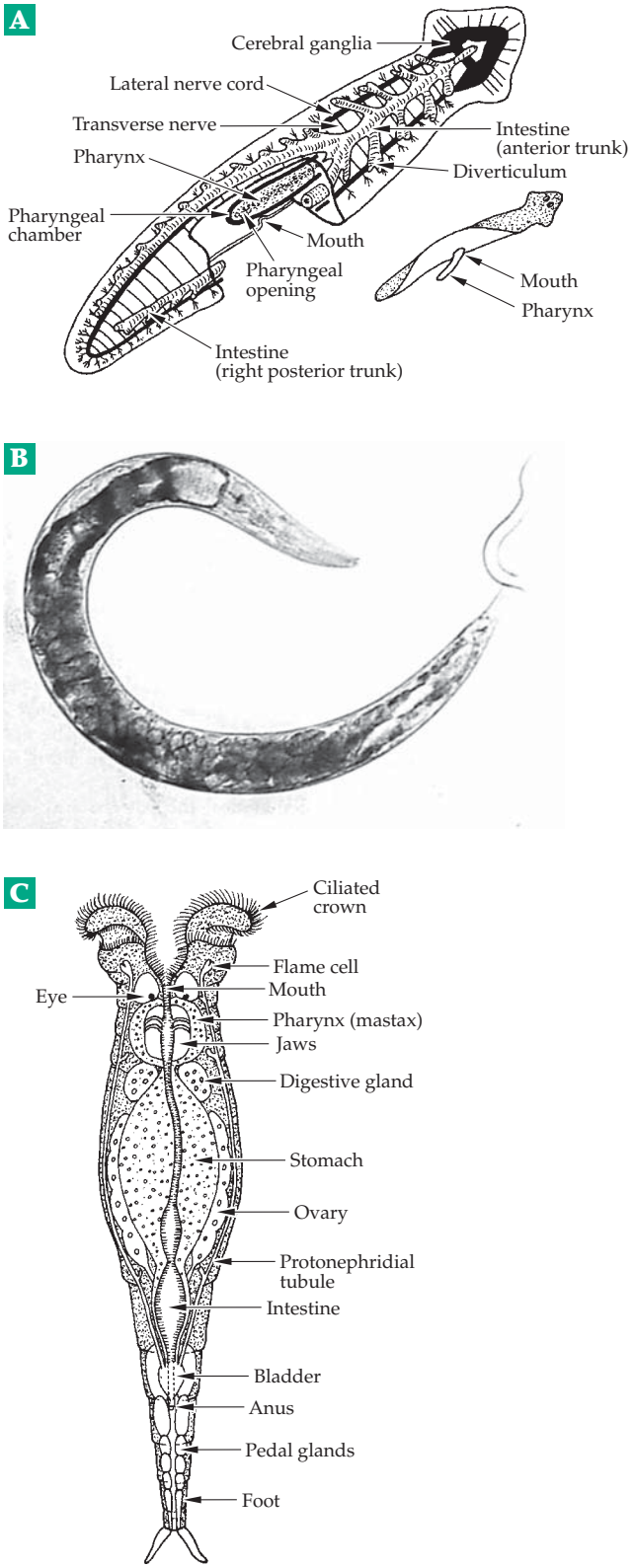



**Figure 1-14** (A) A planarian, length 15 mm. After Hickman.[116] Diagram of digestive and nervous sytems; cutaway section shows ventral mouth. Small drawing shows pharynx extended through ventral mouth. (B) The nematode *Caenorhabditis elegans*. *Ascaris* is very similar in appearance. From Buchsbaum.[77] (C) A rotifer, *Philodina* (~$10^3$ cells). After C. A. Villee *et al.*[117]

**Echinodermata** or spiny-skinned animals (starfish, sea urchins, and sea cucumbers) are regarded as a highly advanced phylum. Their embryological development has been studied intensively.

The phylum **Chordata**, to which we ourselves belong, includes not only the vertebrates but also more primitive marine animals that have a spinal cord. Among these primitive species, which may be related to early ancestral forms, are the **tunicates** or sea squirts. They have a very high concentration of vanadium in their blood.

## 2. Cell Types and Tissues

Isolated animal cells in tissue culture, no matter how highly differentiated, tend to revert quickly to one of three basic types known as **epitheliocytes**, **mechanocytes**, and **amebocytes**. Epitheliocytes are closely adherent cells derived from epithelial tissues and thought to be related in their origins to the two surface layers of the embryonic blastula. Mechanocytes, often called **fibroblasts** or **fibrocytes**, are derived from muscle, supporting, or connective tissue. Like the amebocytes, they arise from embryonic mesenchymal tissue cells that have migrated inward from the lower side of the blastula (Chapter 32). **Neurons**, **neuroglia**, and **lymphocytes** are additional distinct cell types.

### BOX 1-E ERRORS, MISCONCEPTIONS, AND SPECULATION

Warning: Not everything in this book is true. Despite all efforts to get it right, there are unintentional errors and misinterpretations of experimental results. Indeed, the history of biochemistry is replete with accounts of experimental findings that were interpreted incorrectly. Yet, the ideas expressed often stimulated others to develop a more correct picture later. The same is true today. Students should be critical, should look at experimental details, and *consult original literature as much as possible.*

Progress in science depends both upon careful observations and measurements and upon imaginative interpretations of unexpected findings. Speculative ideas, a number of which are mentioned in this book, provide an important stimulus in science. They should neither be ignored nor accepted as facts. I have tried to write in such a way that established facts will not often be confused with speculation.

***Tissues.*** Cells aggregate to form four major kinds of tissue. Epithelial tissues line the primary surfaces of the body: the skin, the digestive tract, urogenital tract, and glands. External skin is composed of flat platelike squamous epithelial cells whereas internal surfaces are often formed by columnar epithelial cells. Glands (sweat, oil, mammary, and internal secretory) as well as the sensory organs of the tongue, nose, and ear are all composed of epithelial cells. Epithelial cells are among the most highly polarized of cells. One side of each cell faces the outside, either air or water, while the other side is often directly against a basement membrane.

Supporting and connective tissues include the fatty **adipose tissue** as well as **cartilage** and **bone**. Both of the latter contain large amounts of intercellular material or **ground substance** consisting largely of complex polymers. Embryonic fibroblasts differentiate into white fibers, which produce collagen, and yellow fibers, which form elastin. The fibrils of both of these proteins are assembled in the intercellular space where they are embedded in the ground substance. **Osteoblasts** form bone by deposition of calcium phosphate in 3–7 µm thick layers within a ground substance that contains special proteins.

A third tissue is **muscle**, which is classified into three types: **striated** (voluntary skeletal muscle), **cardiac** (involuntary striated muscle), and **smooth** (involuntary) muscle. There are two major groups of cells in **nervous tissue**, the fourth tissue type. **Neurons** are the actual conducting cells whose cell membranes carry nerve impulses. Several kinds of **glial cells** lie between and around the neurons.

***Blood cells.*** Blood and the linings of blood vessels may be regarded as a fifth tissue type.[135,135a] The human body contains $5 \times 10^9$ **erythrocytes** or red blood cells per ml, a total of $2.5 \times 10^{13}$ cells in the five liters of blood present in the body. Erythrocytes are rapidly synthesized in the bone marrow. The nucleus is destroyed, leaving a cell almost completely filled with hemoglobin. With an average lifetime of 125 days, human red blood cells are destroyed by leukocytes in the spleen and liver.

The white blood cells or **leukocytes** are nearly a thousandfold less numerous than red cells. About $7 \times 10^6$ cells are present per ml of blood. There are three types of leukocytes: **lymphocytes** (~26% of the total), **monocytes** (~7% of the total), and **polymorphonuclear leukocytes** or **granulocytes** (~70% of the total). Lymphocytes are about the same size as erythrocytes and are made in lymphatic tissue. Individual lymphocytes may survive for as long as ten years. They function in antibody formation and are responsible for maintenance of long-term immunity.

Monocytes, two times larger, are active in ingesting bacteria. These cells stay in the blood only a short time before they migrate into the tissues where they become **macrophages**,[136] relatively fixed phagocytic cells. Macrophages not only phagocytize and kill invading bacteria, protozoa, and fungi but also destroy cancer cells. They also destroy damaged cells and cellular debris as part of the normal turnover of tissues. They play an essential role in the immune system by "processing" antigens and in releasing stimulatory proteins.

Granulocytes of diameter 9–12 µm are formed in the red bone marrow. Three types are distinguished by staining: **neutrophils**, **eosinophils**, and **basophils**. Neutrophils are the most numerous phagocytic cells of our blood and provide the first line of defense against bacterial infections. The functions of eosinophils and basophils are less well understood. The number of eosinophils rises during attacks of hay fever and asthma and under the influence of some parasites, while the basophil count is increased greatly in leukemia and also by inflammatory diseases. Granules containing histamine, heparin, and leukotrienes are present in the basophils. Blood **platelets** or **thrombocytes** are tiny (2–3 µm diameter) cell-like bodies essential for rapid coagulation of blood. They are formed by fragmentation of the cytoplasm of bone marrow **megakaryocytes**. One mature megakaryocyte may contribute 3000 platelets to the $1–3 \times 10^8$ per ml present in whole blood.

***Cell culture.*** Laboratory growth of isolated animal cells has become very important in biochemistry.[137] Sometimes it is necessary to have many cells with as nearly as possible identical genetic makeup. Such bacterial cells are obtained by plating out the bacteria and selecting a small colony that has grown from a single cell to propagate a "pure strain." Similarly, single eukaryotic cells may be selected for tissue culture and give rise to a **clone** of cells which remains genetically identical until altered by mutations.

The culture of embryonic fibroblasts is used to obtain enough cells to perform prenatal diagnosis of inherited metabolic diseases (Box 1-D). Tissue culture is easiest with embryonic or cancer cells, but many other tissues can be propagated. However, the cells that grow best and which can be propogated indefinitely are not entirely normal; the well-known **HeLa** strain of human cancer cells which was widely grown for many years throughout the world contains 70–80 chromosomes per cell compared with the normal 46.

## 3. Communication

Plants are able to maintain their form because the cells are surrounded by thick walls that cement the cells together. However, animal cells lack rigid walls and must be held together by specialized contacts.[138,139] Contacts between cells of both plants and animals are

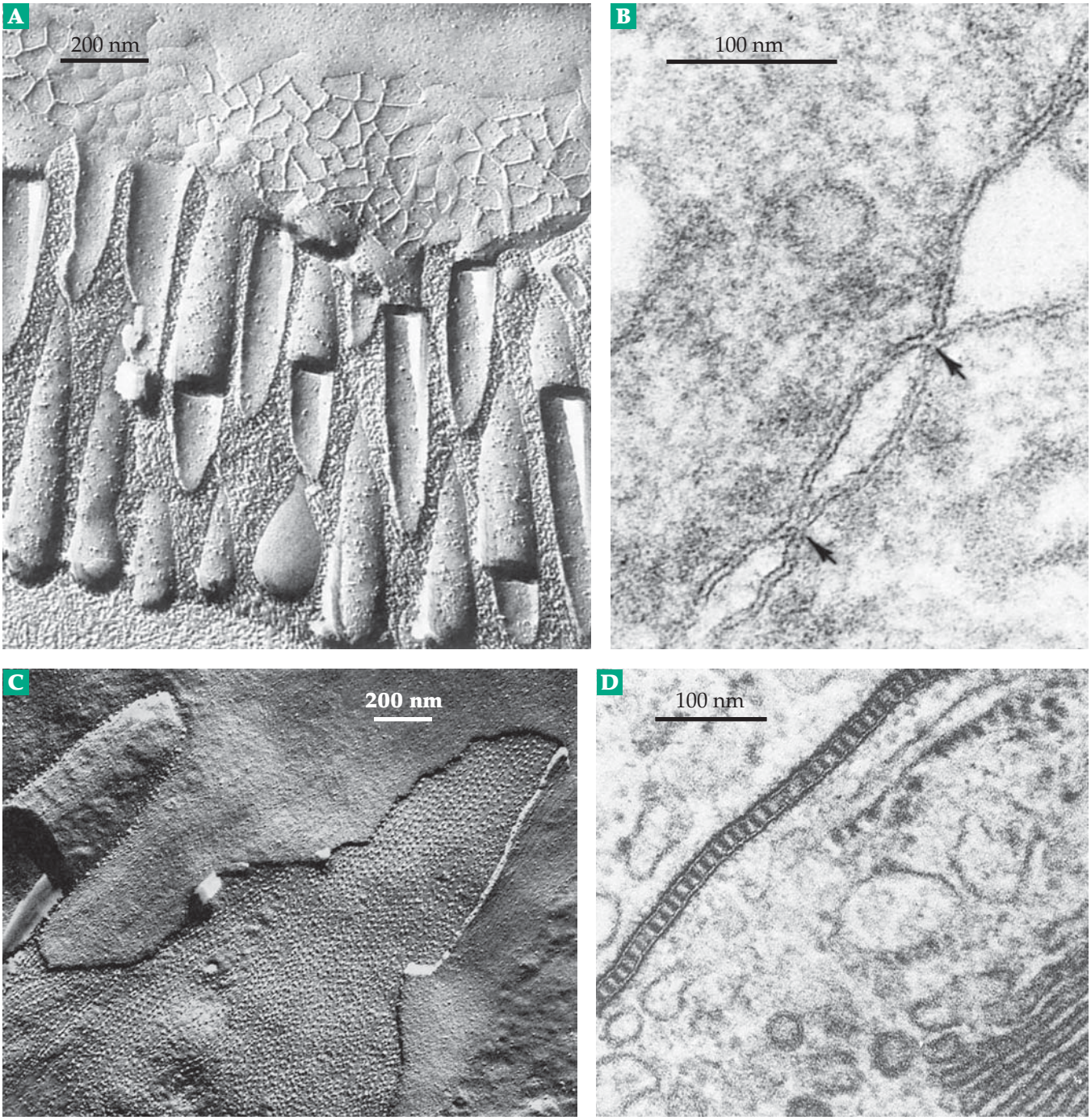
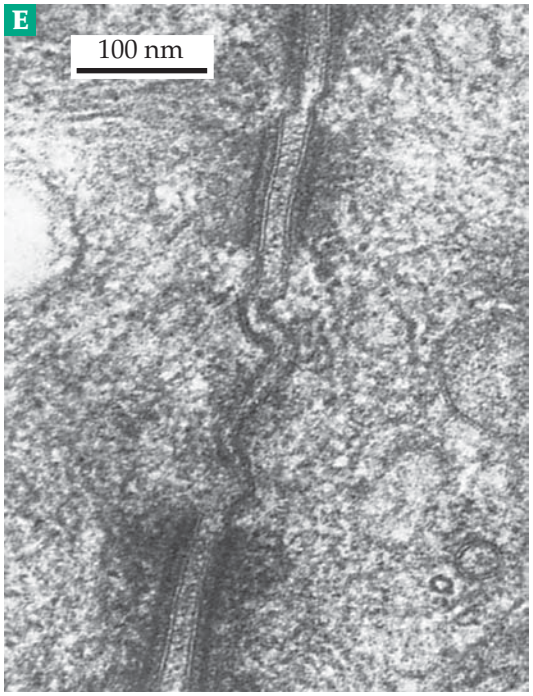
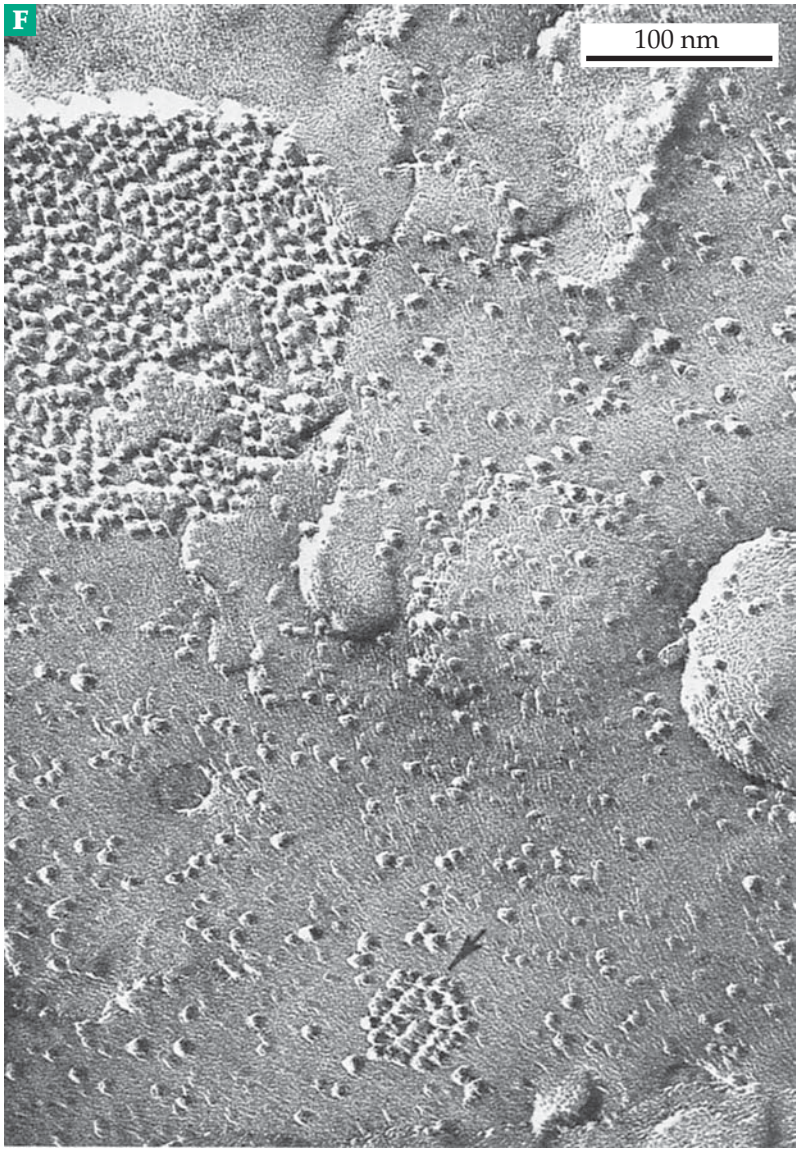



**Figure 1-15** Electron micrographs of cell junctions of three types. (A) Freeze-fractured zona occludens (occlusion zone) between epithelial cells of the rat small intestine. The tight junctions are repesented as a meshwork of ridges (in the P or protoplasmic fracture face) or grooves (in the E fracture face which looks toward the extracellular space). These represent the actual sites of membrane fusion. Microvilli are seen in the lower part of the photograph. From D. S. Friend and N. B. Gilula.[141] (B) Thin cross section of tight junction between mouse hepatocytes. The arrows indicate points of membrane fusion. From Gilula.[142] Copyright 1975 by The Williams & Wilkins Co., Baltimore. (C) A freeze-fractured septate junction from ciliated epithelium of a mollusc. This type of junction forms a belt around the cells. Fracture face P (central depressed area) contains parallel rows of membrane particles that correspond to the arrangmenet of the intracellular septa seen in thin sections. The surrounding fracture face E contains a complementary set of grooves. Particles in nonjunctional membrane regions (upper right corner) are randomly arranged. (D) Thin section of a septate junction of the type shown in (C). The plasma membranes of the two cells are joined by a periodic arrangement of electron-dense bars or septa, which are present within the intercellular space. Note the Golgi membranes in the lower right part of the photograph. (C) and (D) are from N. B. Gilula[143]



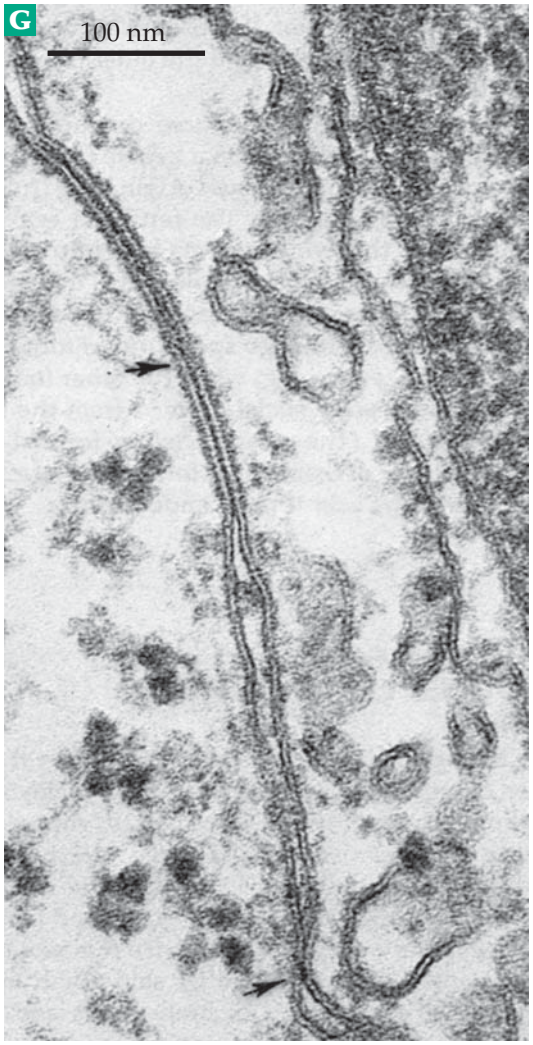




**Figure 1-15** (continued) (E) Desmosomes (macula adherens) in rat intestinal epithelium. Features include a wide (25–35 nm) intercellular space containing dense material, two parallel cell membranes, a dense plaque associated with the cytoplasmic surface and cytoplasmic tonofilaments that converge on the dense plaque. From Gilula.[143] (F) Freeze-fractured surface through gap junctions between communicating cells in culture. Both a large junction and a smaller one below (arrow) can be seen. (G) Gap junctions in thin section. (F) and (G) are from N. B. Gilula.[144]

important for a second reason: Cells must communicate, one with another.

***Cell contacts and junctions.*** Many epithelial cells, e.g., those lining the border of kidney tubules and secretory glands, form **tight junctions** with adjacent cells. Electron microscopy shows that in these junctions the outer portions of the membranes actually fuse in some places (Fig. 1-15). One way in which this has been

demonstrated is to freeze a tissue rapidly and to fracture it in the frozen state within a vacuum chamber. The fractured tissue is kept at about –100°C in a vacuum for a short time while water molecules evaporate from the fractured surfaces. A thin plastic replica is then made of the etched surfaces, which sometimes pass through tight junctions revealing details of their structure (Fig. 1-15, A, E). Study of electron micrographs of such surfaces shows that some cells are completely surrounded by belts of tight junctions, sometimes referred to as **occlusion zones** or **terminal bars**. Tight junctions between endothelial cells of blood capillaries in the brain prevent free diffusion of compounds from the blood stream into brain cells and form the **blood–brain barrier**.[140] Tight junctions between neurons and adjacent cells surround the **nodes of Ranvier** (Chapter 30).

Contacts of another type, known as **septate desmosomes** or **adhesion discs**, form a belt around the cells of invertebrate epithelia. In these contacts a space of ~18 nm between adjacent cell membranes is bridged in a number of places by thin walls. Behind the desmosomes the membrane is often backed up at these points by an electron-dense region to which are attached many fine microfilaments of ~6–10 nm diameter (Fig. 1-15, D).

One method of communication between cells is by passage of chemical substances through special junctions which, because of their appearance in electron micrographs of thin sections (Fig. 1-15, G) are known as **gap junctions**.[139,145,146] Gap junctions may cover substantial areas of the cell interface. In cross section, a thin 3–4 nm gap between the adjacent cell membranes is bridged by a lattice-like structure, which may appear in freeze-fractured surfaces as a hexagonal array of particles (Fig. 1-15, F, lower junction). These particles or **connexons** are each thought to be composed of six protein subunits. A central channel in the connexon is able to pass molecules of molecular mass up to about 500 Da.[147,148] Small molecules may be able to pass freely from one cell to another through the gap junction. Because of their low electrical resistance, gap junctions allow "electrical coupling" of cells. Such junctions form the **electrotonic synapses** that link some neurons to other excitable cells. Heart cells are all electrically coupled through gap junctions.[149]

Another type of communicating junction is also found in **synapses** of the nervous system. At these specialized contacts a nerve impulse transmitted along the membrane of one neuron triggers the release of a **neurotransmitter**, a chemical substance that passes across the gap between cells of the synapse and initiates a nerve impulse in the second neuron (Chapter 30).

***Cell recognition.*** Cells of higher organisms are able to recognize other cells as identical, as belonging to another tissue, or as being "foreign." This ability is developed most highly by cells of the immune system but is possessed to some extent by others. For example, cells of sponges can be separated by partial digestion of the protein "cement" that holds them together. When dissociated cells from orange sponges were mixed with those from yellow sponges, the cells clumped together to reform small sponges.[150,151] Furthermore, orange cells stuck to orange cells and yellow to yellow cells. Similar results have been obtained using a mixture of cultivated liver, kidney, and embryonic brain cells. When a wound heals, epithelial cells grow and move across the wound surface but they stop when they meet. Cells in tissue culture and growing on a glass surface experience this same **contact inhibition**[152] and spread to form a unicellular layer. Cancer cells in culture do not stop but climb one on top of the other, apparently lacking proper recognition and communication. Many chemical signals appear to pass between cells. An important goal of contemporary biochemistry is to understand how cells recognize each other and respond to signals that they receive.

## F. Higher Plants and Plant Tissues

Botanists recognize two divisions of higher plants. The **Bryophyta** or moss plants consist of the Musci (mosses) and Hepaticae (liverworts). These plants grow predominantly on land and are characterized by swimming sperm cells and a dominant gametophyte (haploid) phase. **Tracheophyta**, or vascular plants, contain conducting tissues. About $2 \times 10^5$ species are known. The ferns (class Filicineae, formerly Pteridophyta) are characterized by a dominant diploid plant and alternation with a haploid phase. Seed plants are represented by two classes: **Gymnosperms** (cone-bearing trees) and **Angiosperms**, the true flowering plants.

Genetically the simplest of the angiosperms is the little weed *Arabidopsis thaliana*, whose generation time is as short as five weeks. Its five chromosomes contain only $10^8$ base pairs in all, the smallest known genome among angiosperms[153] and one whose complete nucleotide sequence is being determined. Its biochemistry, physiology, and developmental biology are under intensive study. It may become the "fruit fly" of the plant kingdom.

There are several kinds of plant tissues. Undifferentiated, embryonic cells found in rapidly growing regions of shoots and roots form the **meristematic tissue**. By differentiation, the latter yields the simple tissues, the parenchyma, collenchyma, and sclerenchyma. **Parenchyma** cells are among the most abundant and least specialized in plants. They give rise through further differentiation to the **cambium layer**, the growing layer of roots and stems. They also

make up the pith or pulp in the center of stems and roots, where they serve as food storage cells.

The **collenchyma**, present in herbs, is composed of elongated supporting cells and the **sclerenchyma** of woody plants is made up of supporting cells with hard lignified cell walls and a low water content. This tissue includes **fiber cells**, which may be extremely long; e.g., pine stems contain fiber cells of 40 µm diameter and 4 mm long.

Two complex tissues, the **xylem** and **phloem**, provide the conducting network or "circulatory system" of plants. In the xylem or woody tissue, most of the cells are dead and the thick-walled tubes (**tracheids**) serve to transport water and dissolved minerals from the roots to the stems and leaves. The phloem cells provide the principal means of downward conduction of foods from the leaves. Phloem cells are joined end to end by **sieve plates**, so-called because they are perforated by numerous minute pores through which cytoplasm of adjoining sieve cells appears to be connected by strands 5–9 µm in diameter.[154] Mature sieve cells have no nuclei, but each sieve cell is paired with a nucleated "companion" cell.

**Epidermal tissue** of plants consists of flat cells, usually containing no chloroplasts, with a thick outer wall covered by a heavy waxy **cuticle** about 2 µm thick. Only a few specialized cells are found in the epidermis. Among them are the paired **guard cells** that surround the small openings known as **stomata** on the undersurfaces of leaves and control transpiration of water. Specialized cells in the root epidermis form **root hairs**, long extensions (~1 mm) of diameter 5–17 µm. Each hair is a single cell with the nucleus located near the tip.

Figure 1-16 shows a section from a stem of a typical angiosperm. Note the thin cambium layer between the phloem and the xylem. Its cells continuously undergo differentiation to form new layers of xylem increasing the woody part of the stem. New phloem cells are also formed, and as the stem expands all of the tissues external to the cambium are renewed and the older cells are converted into bark.

Plant **seeds** consist of three distinct portions. The **embryo** develops from a zygote formed by fusion of a sperm nucleus originating from the pollen and an egg cell. The fertilized egg is surrounded in the gymnosperms by a nutritive layer or **endosperm** which is **haploid** and is derived from the same gametophyte tissue that produced the egg. In angiosperms *two* sperm nuclei form; one of these fertilizes the egg, while the other fuses with *two* haploid **polar nuclei** derived from the female gametophyte. (The polar nuclei are formed by the same mitotic divisions that formed the egg.) From this develops a *3n* **triploid** endosperm.

## G. The Chemical Composition of Cells

**Water** is the major component of living cells, but the amount varies greatly. Thus, the pig embryo is 97% water; at birth a new-born pig is only 89% water. A lean 45-kg pig may contain 67% water but a very fat 135-kg animal only 40% water. Similar variations are encountered with other constitutents.

The water content of a tissue is often determined by thoroughly drying a weighed sample of tissue at low temperature in vacuum and then weighing it a second time. The solid material can then be extracted with a solvent that will dissolve out the fatty compounds. These are referred to collectively as **lipids**. After evaporation of the solvent the lipid residue may be weighed. By this procedure a young leafy vegetable might be found to contain 2–5% lipid on a dry weight

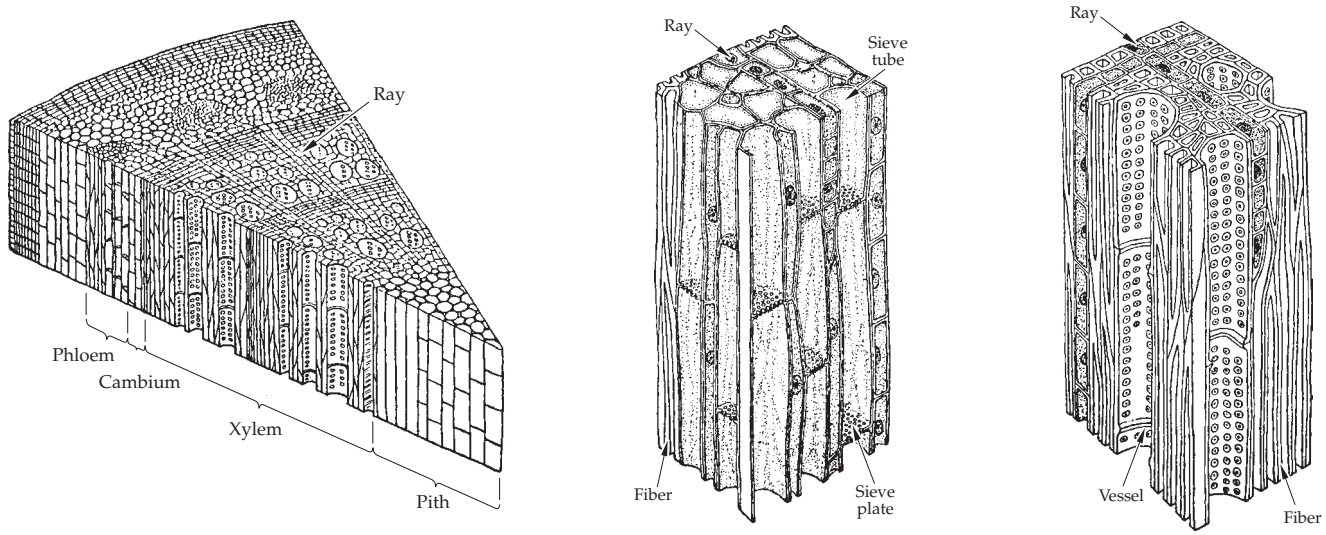



**Figure 1-16** Section of the stem of an angiosperm. Enlarged sections showing tubes of the phloem (left) and xylem (right). From S. Biddulph and O. Biddulph.[155] Drawn by Bunji Tagawa.

basis. Even very lean meats contain 10–30% lipid.

The residue remaining after removal of the lipid consists predominately of three groups of compounds: **proteins**, **nucleic acids**, and **carbohydrates**. Most of the nitrogen present in tissues is found in the proteins and the protein content is sometimes estimated by determining the percentage of nitrogen and multiplying by 6.25. In a young green plant, 20–30% of the dry matter may be protein, while in very lean meat it may reach 50–70%.

**TABLE 1-4**
**Approximate Composition of Metabolically Active Cells and Tissues[a]**

| Component | *E. coli*[b] (%) | Green plant (spinach, *Spinacia oleracea*)[c] | Rat liver[d] (%) |
|---|---|---|---|
| $H_2O$ | 70 | 93 | 69 |
| Protein | 15 | 2.3 | 21 |
| Amino acids | 0.4 | | |
| DNA | 1 | | 0.2 |
| RNA | 6 | | 1.0 |
| Nucleotides | 0.4 | | |
| Carbohydrates | 3 | 3.2 | |
| Cellulose | | 0.6 | |
| Glycogen | | | 3.8 |
| Lipids | 2 | 0.3 | 6 |
| Phospholipids | | | 3.1 |
| Neutral lipids | | | 1.6 |
| Sterols | | | 0.3 |
| Other small molecules | 0.2 | | |
| Inorganic ions | 1 | 1.5 | |
| $K^+$ | | | 0.4 |

| | Equivalents per liter in rat liver |
|---|---|
| Amino acid residues | 2.1 |
| Nucleotide units | 0.03 |
| Glycogen (glucose units) | 0.22 |
| $K^+$ | 0.1 |

[a] Data were not readily available for spaces left blank
[b] From J. D. Watson (1976) *Molecular Biology of the Gene*, 3rd ed., p. 69, Benjamin, New York The amounts of amino acids, nucleotides, carbohydrates, and lipids include precursors present in the cell.
[c] From B. T. Burton (1976) *Human Nutrition*, 3rd ed., McGraw-Hill, New York (p. 505)
[d] From C. Long, ed., (1961) *Biochemists' Handbook*, pp. 677–679, Van Nostrand-Reinhold, Princeton, New Jersey

A dried tissue sample may be burned at a high temperature to an **ash**, which commonly amounts to 3–10% and is higher in specialized tissues such as bone. It is a measure of the inorganic constituents of tissues.

The carbohydrate content can be estimated by the difference of the sum of lipid, protein, and ash from 100%. It amounts to 50–60% in young green plants and only 2–10% in typical animal tissues. In exceptional cases the carbohydrate content of animal tissues may be higher; the glycogen content of oysters is 28%.

The amount of nucleic acid in tissues varies from 0.1% in yeast and 0.5–1% in muscle and in bacteria to 15–40% in thymus gland and sperm cells. In these latter materials of high nucleic acid content it is clear that multiplication of % N by 6.25 is not a valid measure of protein content. For diploid cells of the body the DNA content per cell is nearly constant.

Table 1-4 compares the composition of a bacterium, of a green plant, and of an active animal tissue (rat liver). Although the solid matter of cells consists principally of C, H, O, N, S, and P, many other chemical elements are also present. Among the cations, $Na^+$, $K^+$, $Ca^{2+}$, and $Mg^{2+}$ are found in relatively large amounts. Thus, the body of a 70 kg person contains 1050 g Ca (mostly in the bones), 245 g K, 105 g Na, and 35 g Mg. Iron (3 g), zinc (2.3 g), and rubidium (1.2 g) are the next most abundant. Of these iron and zinc are essential to life but rubidium is probably not. It is evidently taken up by the body together with potassium.

The other metallic elements in the human body amount to less than 1 g each, but at least seven of them play essential roles. They include copper (100 mg), manganese (20 mg), and cobalt (~5 mg). Others, such as chromium (<6 mg), tin, and vanadium, have only recently been shown essential for higher animals.[156,157] Nickel, lead, and others may perhaps be needed.

Nonmetallic elements predominating in the ash are phosphorus (700 g in the human body), sulfur (175 g), and chlorine (105 g). Not only are these three elements essential to all living cells but also selenium, fluorine, silicon (Box 4-B), iodine, and boron are needed by higher animals and boron by plants (Fig. 1-17). Iodine deficiency may affect one billion human beings and may cause 20 million cases per year of **cretinism**, or less severe brain damage.[158]

What is the likelihood that other elements will be found essential? Consider a human red blood cell, an object of volume ~80 $\mu m^3$ and containing about $3 \times 10^8$ protein molecules (mostly hemoglobin). About $7 \times 10^5$ atoms of the "trace metal" copper and $10^5$ atoms of the nutritionally essential tin are present in a single red cell. Also present are $2 \times 10^4$ atoms of silver, a toxic metal. Its concentration, over $10^{-7}$ M, is sufficient that it could have an essential catalytic function. However, we know of none and it may simply have gotten into our bodies from handling money, jewelry, and other

## BOX 1-F ABOUT THE REFERENCES

The lists of references at the ends of chapters are provided to encourage readers to look at original research articles. The lists are neither complete nor critically selected, but they do increase the information given in this book many-fold. I apologize for the important papers omitted. However, the references that are here will help a student to get started in reading the literature. Each reference contains other references and names of persons active in the field. By searching recent journal indices or a computer database it is easy to find additional articles by the same authors or on the same subject.

Look at the various types of scientific articles including reviews, preliminary reports and full research papers. Be sure to examine those in the *primary source journals* which publish detailed research results. These articles have always been sent to referees, active scientists, who check to see that the experiments are described accurately, that the authors have cited relevant literature, and that the conclusions are logical. Some journals, e.g., *Biochem. Biophys. Res. Comm.* and *FEBS Letts.* are dedicated to rapid publication of short reports but are also refereed. Other journals provide mostly reviews or a mixture of reviews. Periodical review series, such as *Advances in Nucleic Acid Chemistry and Related Topics*, often appear annually. Every student who intends to become a professional biochemist should consider purchasing the *Annual Review of Biochemistry* each year. This indispensible source of current information on most aspects of biochemistry is available to students at a very low price.

Many journal papers are difficult to read. To start, pick papers that have an understandable introduction. Choose reviews that are short, such as those in *Trends in Biochemical Sciences*. Then go on to the more comprehensive ones. Never sit back and hope that your computer will automatically fetch just what you need! Many journals carry papers of biochemical importance. Those specializing in biochemistry include the following:

| Full Title | Abbreviation |
|---|---|
| Advances in Carbohydrate Chemistry and Biochemistry[a] | Adv. Carbohydr. Chem. Biochem. |
| Advances in Protein Chemistry[a] | Adv. Protein Chem. |
| Analytical Biochemistry | Anal. Biochem. |
| Annual Review of Biophysics and Biomolecular Structure[a] | Ann. Rev. Biophys. Biomolec. Struct. |
| Annual Review of Biochemistry[a] | Ann. Rev. Biochem. |
| Archives of Biochemistry and Biophysics | Arch. Biochem. Biophys. |
| Biochemical and Biophysical Research Communications | Biochem. Biophys. Res. Commun. |
| Biochemical Journal | Biochem. J. |
| Biochemistry | |
| Biochimica et Biophysica Acta | Biochim. Biophys. Acta |
| Bioorganic Chemistry | Bioorg. Chem. |
| Carbohydrate Research | Carbohydr. Res. |
| EMBO Journal[b] | EMBO J. |
| European Journal of Biochemistry | Eur. J. Biochem. |
| FASEB Journal[c] | FASEB J. |
| Journal of Bacteriology | J. Bacteriol. |
| Journal of Biochemistry | J. Biochem. |
| Journal of Biological Chemistry | J. Biol. Chem. |
| Journal of Lipid Research | J. Lipid Res. |
| Journal of Molecular Biology | J. Mol. Biol. |
| Journal of the American Chemical Society | J. Am. Chem. Soc. |
| Journal of Theoretical Biology | J. Theor. Biol. |
| Methods in Enzymology[a] | Methods Enzymol. |
| Nature | |
| Nucleic Acids Research | Nucleic Acids Res. |
| Proceedings of the National Academy of Sciences, USA | Proc. Natl. Acad. Sci. U.S.A. |
| Science | |
| Structure | |
| Trends in Biochemical Sciences | Trends Biochem. Sci. or TIBS |

[a] These are not journals but series of review and reference books. There are many other series of "Advances in..." and "Annual Reviews of ..." that are not listed here.
[b] EMBO–European Molecular Biology Organization
[c] FASEB–Federation of American Societies for Experimental Biology

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| H | | | | | | | | | | | | | | | | | He |
| Li | Be | | | | | | | | | | | B | C | N | O | F | Ne |
| Na | Mg | | | | | | | | | | | Al | Si | P | S | Cl | Ar |
| K | Ca | Sc | Ti | V | Cr | Mn | Fe | Co | Ni | Cu | Zn | Ga | Ge | As | Se | Br | Kr |
| R | Sr | Y | Zr | Nb | Mo | Tc | Ru | Rh | Pd | Ag | Cd | In | Sn | S | Te | I | Xe |
| Cs | Ba | La | Hf | Ta | W | Re | Os | Ir | Pt | Au | Hg | Tl | Pb | Bi | Po | At | Rn |

Essential to all animals and plants

Essential to several classes of animals and plants

Believed essential to a variety of species

Possible essential trace elements for some species

**Figure 1-17** Elements known to be essential to living things (after da Silva and Williams[157]). Essential elements are enclosed within shaded boxes. The 11 elements–C, H, O, N, S, P, Na, K, Mg, Ca, and Cl–make up 99.9% of the mass of a human being. An additional 13 are known to be essential for higher animals in trace amounts. Boron is essential to higher plants but apparently not to animals, microorganisms, or algae.

silver objects. The red blood cell also contains boron and aluminum ($3 \times 10^5$ atoms each), arsenic ($7 \times 10^5$ atoms), lead ($7 \times 10^4$ atoms), and nickel ($2 \times 10^4$ atoms). Of the elements (uranium and below) in the periodic table, only four (Ac, Po, Pa, and Ra) are present, on the average, in quantities less than one atom per cell.[156]

Of the apparently nonessential elements, several, e.g., Cs, Rb, Sr, and Ni (possibly essential) are not toxic at low concentrations. Others, such as Sb, As, Ba, Be, Cd, Pb, Hg, Ag, Tl, and Th, are highly toxic.

The ionic compositions of tissues and of body fluids vary substantially. Blood of marine organisms is similar to that of seawater in its content of $Na^+$, $Cl^-$, $Ca^{2+}$, and $Mg^{2+}$. Blood of freshwater and terrestrial organisms contains about ten times less $Na^+$ and $Cl^-$ and several times less $Ca^{2+}$ and $Mg^{2+}$ than is present in seawater, but it is nevertheless relatively rich in these ions.

In general, cells are rich in $K^+$ and $Mg^{2+}$, the $K^+$ predominating by far, and are poor in $Na^+$ and $Ca^{2+}$. Chloride is the principal inorganic anion, but organic carboxylate and phosphate groups contribute most of the negative charges (Table 1-4), many of which are fixed to proteins or other macromolecules. Ling estimated that cells typically contain about 1.66 M of amino acid residues in their proteins. Of these residues, 10% have negatively charged side chains and 8% positively charged. The difference is a net negative charge amounting to 33 mM within cells.[159.]

## References


1. May, R. M. (1992) *Sci. Am.* **267**(Oct), 42–48
2. Stanier, R. Y., Ingraham, J. L., Wheelis, M. L., and Painter, P. R. (1986) *The Microbial World*, 5th ed., Prentice-Hall, Englewood Cliffs, New Jersey
3. Brock, T. D., Smith, D. W., and Madigan, M. T. (1988) *Biology of Microorganisms*, 5th ed., Prentice-Hall, Englewood Cliffs, New Jersey
4. Neidhardt, F. C., ed. (1987) *Escherichia coli and Salmonella typhimurium*, Am. Soc. for Microbiology, Washington, D.C.
5. Neidhardt, F. C., Ingraham, J. L., and Schaechter, M. (1990) *Physiology of the Bacterial Cell*, Sinauer, Sunderland, Mass.
6. Maniloff, J., and Morowitz, H. J. (1972) *Bacteriol. Rev.* **36**, 263–290
7. Razin, S. (1978) *Microbiol. Rev.* **42**, 414–470
8. Barile, M. F., Razin, S., Tully, J. G., and Whitcomb, R. F., eds. (1979) *The Mycoplasmas*, Vol. I, Academic Press, New York
9. Watson, J. D. (1976) *Molecular Biology of the Gene*, 3rd ed., Benjamin, Menlo Park, California (p. 61)
10. Horowitz, D. M., and Sanders, J. K. M. (1994) *J. Am. Chem. Soc.* **116**, 2695–2702

10a. Gansheroff, L. J., and O'Brien, A. D. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2959–2961

11. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J.-F., Dougherty, B. A., Bott, K. F., Hu, P.-C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison, C. A., III, and Venter, J. C. (1995) *Science* **270**, 397–403
12. Sonea, S. (1988) *Nature (London)* **331**, 216
13. Costerton, J. W., Geesey, G. G., and Cheng, K.-J. (1978) *Sci. Am.* **238**(Jan), 86–95
14. Macnab, R. M. (1987) in *Escherichia coli and Salmonella typhimurium* (Niedhardt, F. C., ed), pp. 70–83, Am. Soc. for Microbiology, Washington, D.C.
15. Margulis, L., Ashen, J. B., Sonea, S., Solé, M., and Guerrero, R. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6966–6970
16. Goldstein, S. F., Charon, N. W., and Kreiling, J. A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3433–3437
17. Levine, M. M. (1985) *N. Engl. J. Med.* **313**, 445–447
18. Kuehn, M. J., Heuser, J., Normark, S., and Hultgren, S. J. (1992) *Nature (London)* **356**, 252–255
19. Alper, J. (1993) *Science* **262**, 1817
20. Holt, J. G., Krieg, N. R., Sneath, P. H. A., Staley, J. T., and Williams, S. T., eds. (1994) *Bergey's Manual of Determinative Bacteriology*, 9th ed., Williams & Wilkins, Baltimore, Maryland
21. Doolittle, R. F. (1992) *Protein Sci.* **1**, 191–200
22. Schwartz, R. M., and Dayhoff, M. O. (1978) *Science* **199**, 395–403
23. Cavalier-Smith, T. (1986) *Nature (London)* **324**, 416–417
24. Jones, W. J., Nagle, D. P., Jr., and Whitman, W. B. (1987) *Microbiol. Rev.* **51**, 135–177
25. Woese, C. R. (1987) *Microbiol. Rev.* **51**, 221–271
26. Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A., and Peattie, D. A. (1989) *Science* **243**, 75–77
27. Garrett, R. A., Dalgaard, J., Larson, N., Kjems, J., and Mankin, A. S. (1991) *Trends Biochem. Sci.* **16**, 22–26
28. Searcy, D. G. (1982) *Trends Biochem. Sci.* **7**, 183–185
29. Wheelis, M. L., Kandler, O., and Woese, C. R. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2930–2934
30. Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576–4579

30a. Doolittle, W. F. (1999) *Science* **284**, 2124–2128

30b. Pennisi, E. (1999) *Science* **284**, 1305–1307

31. Cavalier-Smith, T. (1992) *Nature (London)* **356**, 570
32. Angert, E. R., Clements, K. D., and Pace, N. R. (1993) *Nature (London)* **362**, 239–241
33. Thimann, K. V. (1963) *The Life of Bacteria*, 2nd ed., Macmillan, New York
34. Stanier, R. Y., and Cohen-Bazire, G. (1977) *Ann. Rev. Microbiol.* **31**, 225–274
35. Morden, C. W., and Golden, S. S. (1989) *Nature (London)* **337**, 382–385
36. Marme, D., Marre, E., and Hertel, R., eds. (1982) *Plasmalemma and Tonoplast: Their Functions in the Plant Cell*, Elsevier, Amsterdam
37. Collins, F. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10821–10823
38. Hinshaw, J. E., Carraghen, B. O., and Milligan, R. A. (1992) *Cell* **69**, 1133–1141
39. Dingwall, C., and Lasky, R. (1992) *Science* **258**, 942–947
40. Nolta, K. V., Padh, H., and Steck, T. L. (1991) *J. Biol. Chem.* **266**, 18318–18323
41. de Duve, C. (1975) *Science* **189**, 186–194
42. Griffiths, G., and Simons, K. (1986) *Science* **234**, 438–443
43. Hartley, S. M. (1992) *Trends Biochem. Sci.* **17**, 325–327
44. Golgi, C. (1898) *Arch. Ital. Biol.* **30**, 60 and 278
45. Kirk, J. T. O., and Tilney-Bassett, R. A. E. (1978) *The Plastids: Their Chemistry, Structure, Growth, and Inheritance*, 2nd ed., Elsevier-North Holland Biomedical Press, Amsterdam
46. Hoffmann, H. P., and Avers, C. J. (1973) *Science* **181**, 749–751

46a. Capaldi, R. A. (2000) *Trends Biochem. Sci.* **25**, 212–214

46b. Rutter, G. A., and Rizzuto, R. (2000) *Trends Biochem. Sci.* **25**, 215–221

47. Fukui, S., and Tanaka, A. (1979) *Trends Biochem. Sci.* **4**, 246–249
48. Kindl, H., and Lazarow, P. B., eds. (1982) *Peroxisomes and Glyoxysomes*, Vol. 386, Annals of the New York Academy of Sciences, New York
49. de Duve, C. (1983) *Sci. Am.* **248**(May), 74–84
50. Masters, C., and Crane, D. (1995) *The Peroxisome: A Vital Organelle*, Cambridge Univ. Press, London
51. Lord, J. M., and Roberts, L. M. (1980) *Trends Biochem. Sci.* **5**, 271–274
52. Cerkasov, J., Cerkasovova, A., Kulda, J., and Vilhelmova, D. (1978) *J. Biol. Chem.* **253**, 1207–1214
53. Glover, D. M., Gonzalez, C., and Raff, J. W. (1993) *Sci. Am.* **268**(Jun), 62–68
54. Olmsted, J. B., and Borisy, G. G. (1973) *Ann. Rev. Biochem.* **42**, 507–540
55. Hall, J. L., Ramanis, Z., and Luck, D. J. L. (1989) *Cell* **59**, 121–132
56. Hyams, J. S. (1989) *Nature (London)* **341**, 485–486
57. Satir, P. (1961) *Sci. Am.* **204**(Feb), 108–116
58. Lazarides, E. (1980) *Nature (London)* **283**, 244–256
59. Amos, L. A., and Amos, W. B. (1991) *Molecules of the Cytoskeleton*, Guilford, New York
60. Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9355–9359
61. Lake, J. A. (1991) *Trends Biochem. Sci.* **16**, 46–50
62. Margulis, L. (1981) *Symbiosis in Cell Evolution*, Freeman, San Francisco, California
63. Hurst, L. D. (1994) *Nature (London)* **369**, 451
64. Gupta, R. S., and Golding, G. B. (1996) *Trends Biochem. Sci.* **21**, 166–171
65. de Duve, C. (1996) *Sci. Am.* **274** (Apr), 50–57
66. Markert, C. L., Shaklee, J. B., and Whitt, G. S. (1975) *Science* **189**, 102–114
67. Fenchel, T., and Bernard, C. (1993) *Nature (London)* **362**, 300
68. Schopf, J. W. (1993) *Science* **260**, 640–646
69. Horgan, J. (1993) *Sci. Am.* **269**(Aug), 24
70. Doolittle, W. F. (1980) *Trends Biochem. Sci.* **5**, 146–149
71. May, R. M. (1986) *Nature (London)* **324**, 514–515
72. Seger, J. (1989) *Nature (London)* **337**, 305–306
73. Riley, M., and Anilionis, A. (1978) *Ann. Rev. Microbiol.* **32**, 519–560
74. Goossens, M., Dozy, A. M., Embury, S. H., Zachariades, Z., Hadjiminas, M. G., Stamatoyannopoulos, G., and Kan, Y. W. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 518–521
75. Williams, R. O., Young, J. R., and Majiwa, P. A. O. (1979) *Nature (London)* **282**, 847–849
76. Kondrashov, A. S. (1988) *Nature (London)* **336**, 435–440
77. Buchsbaum, R., Buchsbaum, M., Pearse, J., and Pearse, V. (1987) *Animals Without Backbones*, 3rd ed., Univ. Chicago Press, Chicago, Illinois
78. Margulis, L., and Schwartz, K. V. (1999) *Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth*, 3rd ed., Freeman, New York
79. Laver, W. G., Air, G. M., Dopheide, T. A., and Ward, C. W. (1980) *Nature (London)* **283**, 454–457
80. McKanna, J. A. (1973) *Science* **179**, 88–90
81. Miller, L. H. (1992) *Science* **257**, 36–37
82. Wyler, D. J. (1993) *N. Engl. J. Med.* **329**, 31–37
83. Miller, L. H., Good, M. F., and Milon, G. (1994) *Science* **264**, 1878–1883
84. Nussenzweig, R. S., and Long, C. A. (1994) *Science* **265**, 1381–1382
85. Sibley, L. D., and Boothroyd, J. C. (1992) *Nature (London)* **359**, 82–85
86. Soldati, D., and Boothroyd, J. C. (1993) *Science* **260**, 349–352
87. Donelson, J. E., and Turner, M. J. (1985) *Sci. Am.* **252**(Feb), 44–51
88. Blum, M. L., Down, J. A., Gurnett, A. M., Carrington, M., Turner, M. J., and Wiley, D. C. (1993) *Nature (London)* **362**, 603–609
89. Gall, J. G., ed. (1986) *The Molecular Biology of Ciliated Protozoa*, Academic Press, Orlando, Florida
90. Olive, L. S. (1975) *The Mycetozoans*, Academic Press, New York
91. Kimmel, A. R., ed. (1989) *Molecular Biology of Dictyostelium Development*, Liss, New York
92. Traynor, D., Kessin, R. H., and Williams, J. G. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8303–8307
93. Webster, J. (1980) *Introduction to Fungi*, 2nd ed., Cambridge Univ. Press, New York
94. Davies, D. A. L., and Pope, A. M. S. (1978) *Nature (London)* **273**, 235–236
95. Georgopapadakou, N. H., and Walsh, T. J. (1994) *Science* **264**, 371–373
96. Sternberg, S. (1994) *Science* **266**, 1632–1634
97. Davis, R. H., and DeSerres, F. J. (1970) *Meth. Enzymol.* **17A**, 79–148
98. Botstein, D., and Fink, G. R. (1988) *Science* **240**, 1439–1443
99. Pringle, J. R., Broach, J. R., and Jones, E. W., eds. (1991–1994) *The Molecular and Cellular Biology of the Yeast Saccharomyces*, Cold Springs Harbor Laboratory Press, Cold Spring Harbor, New York (3 Vols.)

## References


100. Wheals, A. E., Rose, A. H., and Harrison, S. J., eds. (1995) *The Yeasts*, 2nd ed., Vol. 6, Academic Press, San Diego, California
101. Gimeno, C. J., Ljungdahl, P. O., Styles, C. A., and Fink, G. R. (1992) *Cell* **68**, 1077–1090
102. Oliver, S. G., and 146 other authors (1992) *Nature (London)* **357**, 38–46
103. Williams, N. (1995) *Science* **268**, 1560–1561
104. Ruehle, J. L., and Marx, D. H. (1979) *Science* **206**, 419–422
105. Canter-Lund, H., and Lund, J. W. G. (1995) *Freshwater Algae: Their Microscopic World Explored*, Biopress, Bristol, England
106. Harris, E. H. (1989) *The Chlamydomonas Sourcebook*, Academic Press, San Diego, California
107. Brachet, J. L. A. (1965) *Endeavour* **24**, 155–161
108. Bonotto, S., Kefeli, V., and Puiseux-Dao, S. (1979) *Developmental Biology of Acetabularia*, Elsevier-North Holland, Amsterdam
109. Berger, S., and Kaever, M. J. (1992) *Dasycladales: An Illustrated Monograph of a Fascinating Algal Order*, Oxford Univ. Press, New York
110. Menzel, D., and Elsner-Menzel, C. (1990) *Protoplasm* **157**, 52–63
111. Jaenicke, L. (1982) *Trends Biochem. Sci.* **7**, 61–64
112. Ahmadjian, V. (1962) in *Physiology and Biochemistry of Algae*, Vol. 2 (Lewin, R. A., ed), pp. 817–822, Academic Press, New York
113. Morris, S. C. (1993) *Nature (London)* **361**, 219–221
114. Levinton, J. S. (1992) *Sci. Am.* **267**(Nov), 84–91
115. Keese, P. K., and Gibbs, A. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9489–9493
116. Hickman, C. P. (1973) *Biology of the Invertebrates*, Mosby, St. Louis, Missouri
117. Villee, C. A., Walker, W. F., Jr., and Barnes, R. D. (1973) *General Zoology*, Saunders, Philadelphia, Pennsylvania
118. Lapan, E. A., and Morowitz, H. (1972) *Sci. Am.* **227**(Dec), 94–101
119. Lapan, E. A. (1975) *Comp. Biochem. Physiol.* **52A**, 651–657
120. Bergquist, P. R. (1978) *Sponges*, Hutchinson, London
121. Grimmelikhuijzen, C. J. P., and Schaller, H. C. (1979) *Trends Biochem. Sci.* **4**, 265–267
122. Lenhoff, H. M., and Lenhoff, S. G. (1988) *Sci. Am.* **258**(Apr), 108–113
123. Loomis, W. F. (1959) *Sci. Am.* **200**(Apr), 150
123a. Hickman, C. P. *Integrated Principles of Zoology* (1966) Mosby, St. Louis
124. Trager, W. (1986) *Living Together: The Biology of Animal Parasitism*, Plenum, New York
125. Cherfas, J. (1991) *Science* **251**, 630–631
126. Kolberg, R. (1994) *Science* **264**, 1859–1861
127. Wood, W. B., ed. (1988) *The Nematode Caenorhabditis elegans*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
128. Kenyon, C. (1988) *Science* **240**, 1448–1453
129. Roberts, L. (1990) *Science* **248**, 1310–1313
129a. Hotez, P. J., and Pritchard, D. I. (1995) *Sci. Am.* **272**(Jun), 68–74
130. Dumont, H. J., and Green, J., eds. (1980) *Rotatoria*, p. 42, Junk, The Hague
131. Mill, P. J., ed. (1978) *Physiology of Annelids*, Academic Press, New York
132. Florkin, M., and Scheer, B. T., eds. (1970, 1971) *Chemical Zoology*, Vol. 5 and 6, Academic Press, New York
133. Goldstein, L. S. B., and Fyrberg, E. A., eds. (1994) *Drosophila melanogaster: Practical Uses in Cell and Molecular Biology*, Academic Press, San Diego, California
134. Lawrence, P. (1992) *The Making of a Fly: The Genetics of Animal Design*, Blackwell, Oxford
134a. Rubin, G. M., and Lewis, E. B. (2000) *Science* **287**, 2216–2218
135. Maddy, A. H. (1992) *Trends Biochem. Sci.* **17**, 125–126
135a. Harris, J. R., ed. (1991) *Blood Cell Biochemistry*, Plenum, New York (Three volumes)
136. Johnston, R. B., Jr. (1988) *N. Engl. J. Med.* **318**, 747–754
137. Freshney, R. I. (1994) *Culture of Animal Cells*, 3rd ed., Wiley-Liss, New York
138. Feldman, J., Gilula, N. B., and Pitts, J. D., eds. (1978) *Intercellular Junctions and Synapses*, Halsted Press, New York
139. Staehelin, L. A., and Hull, B. E. (1978) *Sci. Am.* **238**(May), 141–153
140. Bradbury, M. (1979) *The Concept of a Blood-Brain Barrier*, Wiley, Chichester, New York
141. Friend, D. S., and Gilula, N. B. (1972) *J. Cell Biol.* **53**, 771
142. Gilula, N. B. (1975) *Cellular Membranes and Tumor Cells*, Williams & Wilkins Co., Baltimore, Maryland (p. 221)
143. Gilula, N. B. (1974) in *Cell Communication* (Cox, R. P., ed), pp. 1–29, Wiley, New York
144. Gilula, N. B. (1972) *Nature (London)* **235**, 262–265
145. Zimmer, D. B., Green, C. R., Evans, W. H., and Gilula, N. B. (1987) *J. Biol. Chem.* **262**, 7751–7763
146. Hertzberg, E. L., and Johnson, R. G., eds. (1988) *Gap Junctions*, Liss, New York
147. Robinson, S. R., Hampson, E. C. G. M., Munro, M. N., and Vaney, D. I. (1993) *Science* **262**, 1072–1074
148. Stauffer, P. L., Zhao, H., Luby-Phelps, K., Moss, R. L., Star, R. A., and Muallem, S. (1993) *J. Biol. Chem.* **268**, 19769–19775
149. Veenstra, R. D., and DeHaan, R. L. (1986) *Science* **233**, 272–274
150. Moscona, A. A. (1961) *Sci. Am.* **205**(Sep), 143–162
151. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. (1983) *Molecular Biology of the Cell*, pp. 679–680, Garland Publ., New York
152. Heaysman, J. E. M. (1978) *Intl. Rev. Cytol.* **55**, 49–66
153. Palca, J. (1989) *Science* **245**, 131
154. Lloyd, A. M., Barnason, A. R., Rogers, S. G., Byrne, M. C., Fraley, R. T., and Horsch, R. B. (1986) *Science* **234**, 464–466
155. Biddulph, S., and Biddulph, O. (1959) *Sci. Am.* **200**(Feb), 44–49
156. Bowen, H. J. M. (1966) *Trace Elements in Biochemistry*, Academic Press, New York
157. Fraústo da Silva, J. J. R., and Williams, R. J. P. (1991) *The Biological Chemistry of the Elements: The Inorganic Chemistry of Life*, Clarendon Press, Oxford
158. Hetzel, B. S. (1994) *N. Engl. J. Med.* **331**, 1770–1771
159. Ling, G. N. (1962) *A Physical Theory of the Living State*, Ginn (Blaisdell), Boston, Massachusetts

## Study Questions

1. Describe the principal structural or organizational differences between prokaryotic and eukaryotic cells.

2. Describe two or more principal functions of proteins within cells, one function of DNA, two or more functions of RNA, and one function of lipids.

3. Compare the chemical makeup of ribosomes, of cell membranes, and of bacterial flagella.

4. Assume the following dimensions: Mycoplasma, sphere, 0.33 µm diameter; *E. coli*, cylinder, 0.8 µm diameter x 2 µm; liver cell, sphere 20 µm; root hair, cylinder, 10 µm diameter x 1 mm.

   a. Calculate for each cell the total volume, the mass in grams and in daltons (assume a specific gravity of 1.0).

   b. Assume that bacterial ribosomes are approximately spherical with a diameter of 23 nm. What is their volume? If the mass of a bacterial ribosome is $2.7 \times 10^6$ daltons, what is its apparent density (divide mass by volume)? Experimentally the buoyant density of bacterial ribosomes in a cesium chloride gradient (Chapter 5) is about 1.6 g/cm$^3$. How can this difference be explained? If eukaryotic ribosomes are 1.17 times larger than bacterial ribosomes in linear dimensions, what is the volume of a eukaryotic ribosome?

   c. What fraction of volume of *E. coli* consists of cell wall, of plasma membrane, of ribosomes (assume 15,000 are present)? If a cell of *E. coli* is 80% water, what fraction by weight of the total solids consists of ribosomes? Of DNA (assuming 2 chromosomes per cell)?

   d. What fraction by volume of a liver cell is composed of ribosomes, of plasma membrane, of mitochondria (assume 1000 mitochondria)? What fraction is accounted for by the nucleus?

5. a. What is the molar concentration of an enzyme of which only one molecule is present in an *E. coli* cell?

   b. Assume that the concentration of $K^+$ within an *E. coli* cell is 150 mM. Calculate the number of $K^+$ ions in a single cell.

   c. If the pH inside the cell is 7.0, how many $H^+$ ions are present?

6. If chromosomes (and chromatin) are 15% DNA, what will be the mass of 23 pairs of chromosomes in a human diploid cell? If the nucleus has a diameter of 5 mm and a density of 1.1 g/cm$^3$, what fraction by weight of the nucleus is chromatin?

7. Compare the surface to volume ratios for an *E. coli* cell, a liver cell, the nucleus of a eukaryotic cell, a root hair. If a cell of 20 µm diameter is 20% covered with microvilli of 0.1 µm diameter and 1 µm length centered on a 0.2 µm spacing, how much will the surface/volume ratio be increased?

8. It has been shown that the code for specifying a particular amino acid in a protein is determined by a sequence of three nucleotides (a codon) in a DNA chain. There are four different kinds of nucleotide units in DNA. How many different codons exist? Note that this is larger than the number of different amino acids (20) that are incorporated into proteins plus the three stop (termination) codons (see Tables 5-5 and 5-6 for a list of codons).

## Study Questions

9. State two similarities and two differences between cyanobacteria (blue-green algae) and green algae.

10. Compare the sizes and structures of bacterial and eukaryotic flagella.

11. How much larger in volume is a typical eukaryotic cell compared to a bacterium?

12. Compare the structure and properties of mitochondria, chloroplasts, and peroxisomes.

13. What are the possible origins of mitochondria and chloroplasts? What evidence can you cite to support your answer?

14. How many different kinds of polymers, e.g. proteins, RNA, that are present in or around living cells, can you name? Can you name some subgroups in any of your categories?

15. Compare the composition of these three, especially with respect to C, H, O, N, S, P, Fe, Cu, Al, and Si:

    a. The earth's crust

    b. Ocean water

    c. Cytoplasm

© John Gerlach/DRK PHOTO

A spider's orb-web is formed by extrusion of a concentrated protein solution and stretching of the resulting fiber. The cross-strands, which are stronger than steel, resemble silkworm silk. The molecules contain microcrystalline β sheet domains that are rich in Gly-Ala repeats as well as polyalanine segments. The capture spiral is formed from much more elastic molecules that contain many β-turn-forming sequences. These assume a spring-like β spiral. See Box 2-B.

# Contents


39 **A. Structural Principles for Small Molecules**
39 1. Bond Angles
40 2. Bond Lengths
40 3. Contact Distances
41 4. Asymmetry: Right-Handed and Left-Handed Molecules
42 *The D- and L- families of amino acids*
42 *The RS notation for configuration*
43 *Diastereoisomers*
43 *Geometrical isomers*
43 5. Conformations: The Shapes that Molecules can Assume
45 6. Tautomerism and Resonance
46 **B. Forces between Molecules and between Chemical Groups**
46 1. Van der Waals Forces
47 2. Attraction between Charged Groups (Salt Linkages)
47 3. Hydrogen Bonds
49 4. The Structure and Properties of Water
50 5. Hydration of Polar Molecules and Ions
50 6. Hydrophobic Interactions
51 **C. Amino Acids and Peptides**
51 1. Properties of α-Amino Acids
55 2. Acidic and Basic Side Chains
55 3. The Peptide Unit
56 4. Polypeptides
59 **D. The Architecture of Folded Proteins**
59 1. Conformations of Polypeptide Chains
61 2. The Extended Chain β Structures
62 *Pleated sheets*
63 *Twisted sheets*
64 *Properties of β sheets*
65 *Cylinders and barrels*
67 *Beta propellers*
68 3. Helices
69 *Properties of helices*
70 *Stacking of helices in proteins*
70 *Coiled coils*
71 *Helix bundles*
72 4. Polyglycine II and Collagen
72 5. Turns and Bends
74 6. Domains, Subunits, and Interfaces
75 7. Packing of Side Chains
75 8. The Network of Internal Hydrogen Bonds
76 **E. Folding Patterns and Protein Families**
76 1. Complex Folding Patterns
77 2. Symmetry
78 3. Effects of Sequence on Folding
79 **F. Chemical Modification and Crosslinking**
80 1. Disulfide Crosslinkages
80 2. Other Crosslinkages
81 **G. Dynamic Properties of Proteins**
81 1. Motion of Backbone and Side Chains
81 2. Conformational Changes
82 3. Denaturation and Refolding
83 4. Effects of pH and Solvent
84 5. Irreversible Damage to Proteins
85 **H. Design and Engineering of Proteins**

86 **References**
91 **Study Questions**

**Boxes**
58 Box 2-A Proteins of Blood Plasma
67 Box 2-B Silks
83 Box 2-C The Nobel Prizes

**Tables**
41 Table 2-1 The Sizes of Some Atoms
52 Table 2-2 Structure and Chemical Properties of Side Chain Groups of Amino Acids
61 Table 2-3 Approximate Torsion Angles for Some Regular Peptide Structures
78 Table 2-4 Classification of Protein Residues According to Their Tendencies to Form α Helix, β Structure, and β Turns

# Amino Acids, Peptides, and Proteins

# 2

Thousands of different proteins make up a very large fraction of the "machinery" of a cell. Protein molecules catalyze chemical reactions, carry smaller molecules through membranes, sense the presence of hormones, and cause muscle fibers to move. Proteins serve as structural materials within cells and between cells. Proteins of blood transport oxygen to the tissues, carry hormones between cells, attack invading bacteria, and serve in many other ways. No matter what biological process we consider, we find that a group of special proteins is required.

The amino acid units that make up a protein molecule are joined together in a precise sequence when the protein is made on a ribosome. The chain is then folded, often into a very compact form. Sometimes the chain is then cut in specific places. Pieces may be discarded and parts may be added. A metal ion, a coenzyme derived from a vitamin, or even a single methyl group may be attached to form the biologically active protein. The final product is a complex and sophisticated machine, often with moving parts, that is exquisitely designed for its particular role.

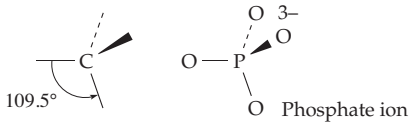
The biological functioning of a protein is determined both by the properties of the chemical groups in the amino acids that are joined to form the protein chain and by the way the chain is folded. The ways in which the different parts of the protein interact with each other and with other molecules are equally important. These interactions play a major role in determining the folding pattern and also provide much of the basis for the biological functioning of proteins. Similar considerations apply also to carbohydrates, nucleic acids, and other biopolymers. For these reasons it is appropriate to review some fundamentals of molecular structure and geometry.

## A. Structural Principles for Small Molecules

Stable organic molecules are held together by covalent bonds which are usually very strong. The standard Gibbs energies of formation ($\Delta G_f°$)[†] of many covalent single bonds are of the order of −400 kJ/mol (96 kcal/mol). The bonds have definite directions, which are measured by **bond angles** and definite **bond lengths**.

### 1. Bond Angles

Because of the tetrahedral arrangement of the four bonds around single-bonded carbon atoms and most phosphorus atoms, all six of the bond angles about the central atom have nearly the same tetrahedral angle of 109.5°.




Bond angles within chains of carbon atoms in organic compounds vary only slightly from this, and even atoms that are attached to fewer than four groups usually have similar angles; for example, the H–O–H angle in a water molecule is 105°, and the H–N–H angles of ammonia are 107°. In ethers the C–O–C angle is 111°. However, bond angles of only 101° are present in $H_2O_2$ and of 92° in $H_2S$ and $PH_3$.

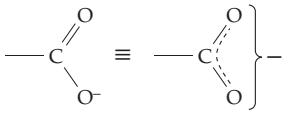
[†]See Chapter 6 for a review of thermodynamics

The presence of **double bonds** leads to **planarity** and to compounds with bond angles of 120°, the internal angle in a hexagon. The planar geometry imposed upon an atom by a double bond is often transmitted to an adjacent nitrogen or oxygen atom as a result of **resonance** (Section 6). For example, the amide groups that form the peptide linkage in proteins (see Fig. 2-5) are nearly planar and the angles all fall within four degrees of 120°.

## 2. Bond Lengths

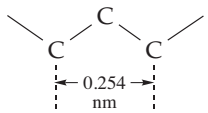
Chemists describe bond lengths as the distances between the nuclei of bonded atoms. The C–C single bond has a length of 0.154 nm (1.54 Å). The C–O bond is ~0.01 nm shorter (0.143 nm), and the typical C–H bond has a length of ~0.109 nm. The C–N bond distance is halfway between that for C–C and C–O (0.149 nm). Other lengths, such as that of O–H, can be estimated from the covalent radii given in Table 2-1.

The length of a double bond between any two atoms (e.g., C=C) is almost exactly 0.020 nm less than that for a single bond between the same atoms. If there is resonance, hence only partial double bond character, the shortening is less. For example, the length of the C–C bond in benzene is 0.140 nm; the C–O distances in the carboxylate anion are 0.126 nm.

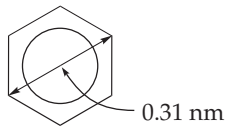


Carboxylate ion

Using simple geometry, it is easy to calculate overall lengths of molecules; here are two distances worth remembering:

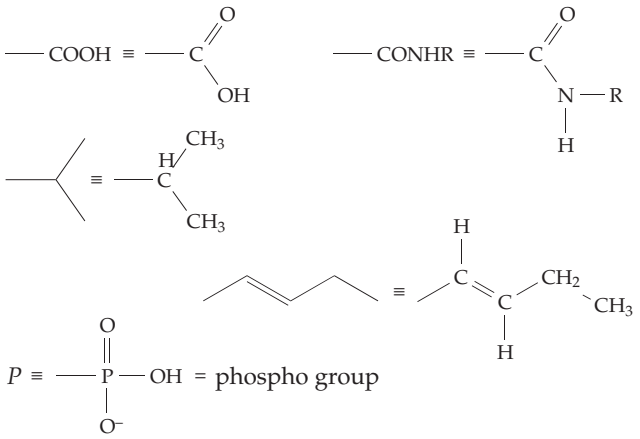



Distance between alternate C atoms in fully extended hydrocarbon chain




Distance across a benzene ring

In the preceding simplified structural formula for benzene the six hydrogen atoms have been omitted. Resonance between the two possible arrangements of the three double bonds[1] is indicated by the circle. Chemical shorthand of the following type is used throughout the book. Carbon atoms may be represented by an angle or the end of a line, but other atoms will always be shown.




## 3. Contact Distances

Covalent bond distances and angles tell us how the atomic nuclei are arranged in space but they do not tell us anything about the outside surfaces of molecules. The distance from the center of an atom to the point at which it contacts an adjacent atom in a packed structure such as a crystal (Fig. 2-1) is known as the **van der Waals radius**. The ways in which biological molecules fit together are determined largely by the van der Waals contact radii. These, too, are listed in Table 2-1. In every case they are approximately equal to the *covalent radius plus 0.08 nm*. Van der Waals radii

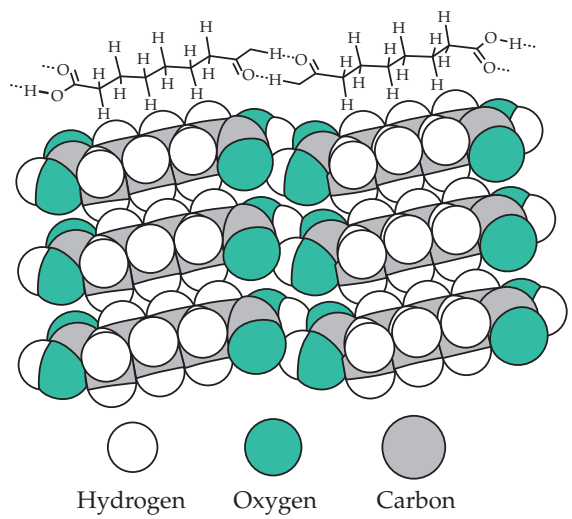


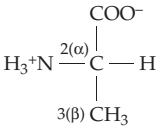

**Figure 2-1** Packing of molecules of suberic acid HOOC–$(CH_2)_6$–COOH in a crystal lattice as determined by neutron diffraction.[2] Notice the pairs of hydrogen bonds that join the carboxyl groups at the ends of the molecules and also the close contact of hydrogen atoms between the chains. Only the positions of the hydrogen nuclei were determined; the van der Waals radii have been drawn around them. However, the radii were originally determined from X-ray and neutron diffraction data obtained from many different crystalline compounds.

are not as constant as covalent radii because atoms can be "squeezed" a little, but only enough to decrease the contact radii by 0.005–0.01 nm. The radii of space-filling molecular models are usually made a little smaller than the actual scaled van der Waals radii to permit easier assembly.

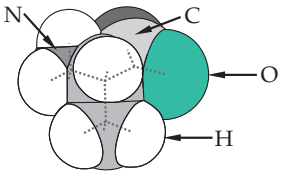
## 4. Asymmetry: Right-Handed and Left-Handed Molecules

The left hand looks much like the right hand, but they are different. One is the mirror image of the other. A practical difference is that your right hand will not fit into a left-handed glove. Despite our daily acquaintance with "handedness" it may seem difficult to explain in words how a right and a left hand differ. However, since most biochemical compounds are asymmetric,[3] it is important to be able to visualize these molecules in three dimensions and to draw their structures on paper. One of the best ways of learning to do this is to study molecular models. You may learn the most by making your own models (see Appendix).

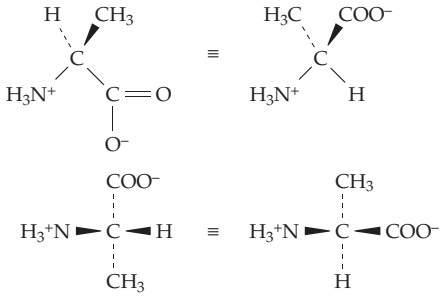
Whenever four different groups are bonded to a central carbon atom, the molecule is asymmetric and the four groups can be arranged in two different **configurations**. Consider alanine, one of the alpha (α)-amino acids from which proteins are built.




α–Alanine According to the Fisher projection formula it is L-alanine.




It is called an α-amino acid because the amino group is attached to the α (or number 2) carbon atom. To indicate its three-dimensional structure on a flat piece of paper, the bonds that project out of the plane of the paper and up toward the reader are often drawn as elongated triangles, while bonds that lie behind the plane of the paper are shown as dashed lines. The isomer of alanine having the configuration about the α-carbon atom shown in the following structural formulas is called *S*-alanine or L-alanine. The isomer which is a mirror image of *S*-alanine is *R*-alanine or D-alanine. Pairs of *R* and *S* compounds (see Section B for definitions) are known as **enantiomorphic** forms or **enantiomers**.




*S*-Alanine (L-Alanine)

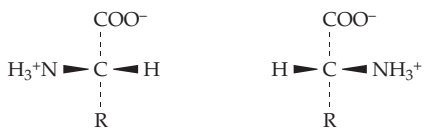
Notice that in the foregoing drawings, the carboxyl group ( —C<sup>O</sup>/OH ), abbreviated –COOH, is shown as having lost a proton to form the carboxylate ion $-COO^-$. Likewise, the amino group ($-NH_2$) has gained a proton to form the $-NH_3^+$ ion. The resulting **dipolar ionic** or **zwitterionic** structure is the one that actually exists for amino acids both in solution and in crystals.

**TABLE 2-1**
**The Sizes of Some Atoms[a–c]**

| Element | Covalent radius (nm)[a,b,d] | van der Waals radii (nm)[d] ***a*** Sideways contact[c] | ***b*** Polar contact[c] |
|---|---|---|---|
| H | 0.030 | 0.12 | |
| F | 0.064 | 0.138 | 0.130 |
| C | 0.077 | 0.16 | |
| N | 0.070 | 0.160 | 0.160 |
| O | 0.066 | 0.154 | 0.154 |
| Cl | 0.099 | 0.178 | 0.158 |
| Si | 0.117 | | |
| P | 0.110 | 0.19 | |
| S | 0.104 | 0.203 | 0.160 |
| Br | 0.114 | 0.184 | 0.154 |
| I | 0.133 | 0.213 | 0.176 |
| Se | | 0.215 | 0.170 |
| "Radius" of methyl group | | 0.20 | |
| Half-thickness of aromatic molecules | | 0.170 | |

[a] From Pauling, L. (1960) *The Nature of the Chemical Bond*, 3rd ed., Cornell Univ. Press, Ithaca, New York (pp. 224–227 and 260).
[b] Covalent radii for two atoms can be summed to give the inter-atomic distance. The van der Waals radii determine how closely molecules can pack. The closest observed contacts between atoms in macromolecules are approximately 0.02 nm less than the sum of the van der Waals radii. From Sasisekharian, V., Lakshminarayanan, A. V., and Ramachandran, G. N. (1967) in *Conformation of Biopolymers*, Vol. 2 (Ramachandran, G. N., ed), Academic Press, New York, (p. 641)
[c] Nyburg, S. C and Faerman, C. H. (1985) *Acta Crystal*. **B41**, 274–279 Shapes of many atomic surfaces are elliptical. The major radius ***a*** applies to sideways contacts and the minor radius ***b*** to "polar" contacts along a covalent bond axis. Distances are for atoms singly bonded to C and may differ slightly if bonds are to other atoms.
[d] For distances in Å multiply by 10.

***The D- and L- families of amino acids.*** The amino acids of which proteins are composed are related to L-alanine but have various side chains (R groups) in place of the methyl group of alanine. In the preceding section the structure of L-alanine was given in four different ways. To recognize them all as the same structure, we can turn them in space to an orientation in which the carboxyl group is at the top, the side chain ($-CH_3$) is down, and both project behind the paper. The amino group and hydrogen atom will then project upward from the paper at the sides as shown below. According to a convention introduced at the beginning of this century by Emil Fischer, *an amino acid is L if, when oriented in this manner, the amino group lies to the left and D if it lies to the right.*



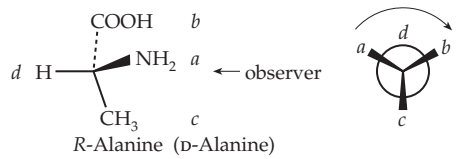

L-Amino acid D-Amino acid

Fischer further proposed that the amino acid in this orientation could be projected onto the paper and drawn with ordinary lines for all the bonds. This gives the previously shown **Fischer projection formula** of L-alanine.

Although the D and L system of designating configuration is old it is still widely used. Remember that D and L refer to the absolute configurations about a selected reference atom in the molecule; for an amino acid this is the number 2 or α-carbon. A quantity that is related to the asymmetry of molecules is the experimentally measurable **optical rotation** (Chapter 23). The sign of the optical rotation (+ or –) is sometimes given together with the name of a compound, e.g., D(+)-glucose. The older designations *d* (dextro) and *l* (levo) indicated + and –, respectively. However, compounds with the D configuration may have either + or – optical rotation.

In older literature optical isomerism of the type represented by D and L pairs was usually discussed in terms of "asymmetric carbon atoms" or "asymmetric centers." Now the terms **chiral** (pronounced *ki-ral*) **molecules**, chiral centers, and **chirality** (Greek: "handedness") are preferred.

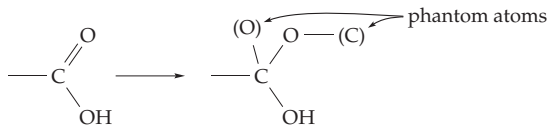
***The RS notation for configuration.*** This notation, devised by Cahn, Ingold, and Prelog, provides an unambiguous way of specifying configuration at any chiral center.[4,5] It is especially useful for classes of compounds for which no well-established DL system is available. The groups or atoms surrounding the central carbon atom, or other central atom, are ranked according to a **priority sequence**. The priority of a group is determined by a number of sequence rules, the first of which is (1) *Higher atomic number precedes lower*. In the following illustration, the priorities of the groups in D-alanine are indicated by the letters $a > b > c > d$. The highest priority (*a*) is assigned to the $NH_2$ groups which contain nitrogen bonded to the central atom. To establish the configuration, the observer views the molecule down the axis connecting the central atom to the group having the lowest priority, i.e., to group *d*. Viewed in this way, the sequence of groups *a*, *b*, and *c* can either be that of a right-handed turn (clockwise) as shown in the drawing or that of a left-handed turn (counterclockwise).




*R*-Alanine (D-Alanine)

The view down the axis and toward the group of lowest priority (d), which lies behind the page. The right-handed turn indicates the configuration R (rectus = right); the opposite configuration is S (sinister = left).

To establish the priority sequence of groups first look at the atoms that are bonded directly to the central atom, arranging them in order of decreasing atomic number. Then if necessary, move outward to the next set of atoms, again comparing atomic numbers. In the case of alanine, groups *b* and *c* must be ordered in this way because they both contain carbon directly bonded to the central atom. When double bonds are present at one of the atoms being examined, e.g., the carboxyl group in alanine, imagine that **phantom atoms** that replicate the real ones are present at the ends of the bonds:




These phantom atoms fill out the valences of the atoms involved in the multiple bonds and are considered to have zero atomic number and zero mass. They are not considered in establishing priorities.

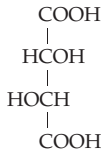
If the first rule and the expansion of multiple bonds are not sufficient to establish the priority, use these additional rules: (2) *Higher atomic mass precedes lower.* (3) When a double bond is present *Z precedes E* (see Geometrical isomers). For ring systems a *cis* arrangement of the highest priority substituents precedes *trans*. (4) When a pair of chiral centers is present *R,R* or *S,S* precedes *R,S* or *S,R*. (5) *An R chiral center precedes S.* For further details see Eliel *et al.*[5] and Bentley.[6] The following groups are ordered in terms of *decreasing priority*[6]: SH > OR > OH > NH–$COCH_3$ > $NH_2$ > COOR > COOH > CHO > $CH_2OH$ > $C_6H_5$ > $CH_3$ > $^3H$ > $^2H$ > H.

Although the *RS* system is unambiguous, closely

related compounds that belong to the same configurational family in the DL system may have opposite configurations in the *RS* system. Thus, L-cysteine (side chain $-CH_2SH$) has the *R* configuration. This is one of the reasons that the DL system is still used for amino acids and sugars.

***Diastereoisomers.*** Whereas compounds with one chiral center exist as an enantiomorphic pair, molecules with two or more chiral centers also exist as diastereoisomers (diastereomers). These are pairs of isomers with an opposite configuration at one or more of the chiral centers, but which are not complete mirror images of each other. An example is L-threonine which has the 2*S*, 3*R* configuration. The diastereoisomer with the 2*S*, 3*S* configuration is known as L-*allo*-threonine. L-isoleucine, whose side chain is $-CH(CH_3)\ CH_2CH_3$, has the 2*S*, 3*R* configuration. It can be called 2(*S*)-amino-3(*R*)-methyl-valeric acid but the simpler name L-isoleucine implies the correct configuration at both chiral centers.

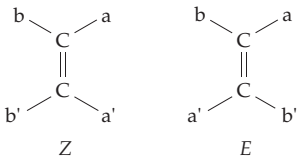
Sometimes the subscript s or g is added to a D or L prefix to indicate whether the chirality of a compound is being related to that of serine, the traditional configurational standard for amino acids, or to that of glyceraldehyde. In the latter case the sugar convention (Chapter 4) is followed. In this convention the configurations of the chiral centers furthest from C1 are compared. Ordinary threonine is $L_s$- or $D_g$-threonine. The configuration of dextrorotatory (+)-tartaric acid can be described as 2*R*, 3*R*, or as $D_s$, or as $L_g$.



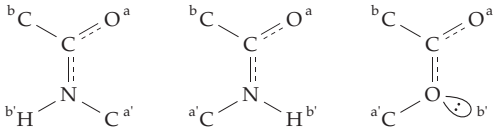

2*R*, 3*R*-tartaric acid

Biochemical reactions are usually stereospecific and a given enzyme will catalyze reactions of molecules of only a single configuration. A related fact is that proteins ordinarily consist entirely of amino acids of the L series.

***Geometrical isomers.*** The *RS* system also gives an unambiguous designation of geometrical isomers containing a double bond.[5,7] At each end of the bond, select the group of highest priority. If these two groups lie on the same side of the double bond the configuration is ***Z*** (from the German **zusammen**, "together"); if on opposite sides ***E*** (**entgegen**, "opposite").




Configurations of amide or ester linkages may also be specified in this manner. This is possible because the C–N bond of an amide has partial double-bond character, as to a lesser extent does the C–O bond to the bridge oxygen in an ester. In this case, assign the lowest priority to the unshared electron pair on the ester bridge oxygen.

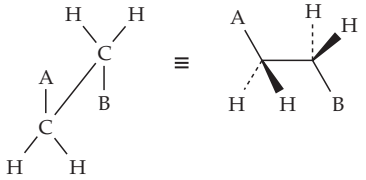



An amide of the *Z* configuration is ordinarily referred to as *trans* in protein chemistry because the main chain atoms are trans.

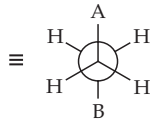
## 5. Conformations: The Shapes That Molecules Can Assume

As important to biochemists as configurations, the stable arrangements of bonded atoms, are **conformations**, the various orientations of groups that are caused by rotation about single bonds.[5,8] In many molecules such rotation occurs rapidly and freely at ordinary temperatures. We can think of a $-CH_3$ group as a kind of erratic windmill, turning in one direction, then another. However, even the simplest molecules have *preferred conformations*, and in more complex structures rotation is usually very restricted.

Consider a molecule in which groups A and B are joined by two $CH_2$ (methylene) groups. If A and B are pulled as far apart as possible, the molecule is in its fully extended **anti** or **staggered** conformation:



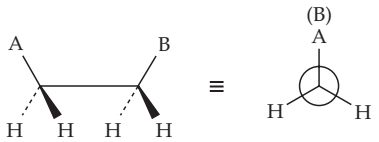
The antiperiplanar conformation



View down the axis joining the C atoms (Newton projection)

Groups A and B are said to be **antiperiplanar** (*ap*) in this conformation. Not only are A and B as far apart as possible but also all of the hydrogen atoms are at their maximum distances one from the other. This can be seen by viewing the molecule down the axis joining

the carbon atoms (Newman projection). Rotation of the second carbon atom 180° around the single bond yields the **eclipsed** conformation in which groups A and B are synperiplanar.

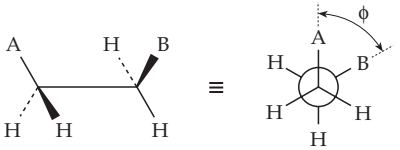



Eclipsed or synperiplanar conformation

If A and B are large bulky groups they will bump together, attainment of the eclipsed conformation will be almost impossible, and rotation will be severely restricted. Even if A and B are hydrogen atoms (ethane), there will be a rotational barrier in the eclipsed conformation which amounts to ~12 kJ (3 kcal) per mole because of the crowding of the hydrogen atoms as they pass each other.[5,9] This can be appreciated readily by examination of space-filling molecular models.

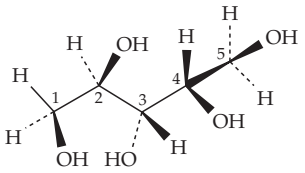
If groups A and B are methyl groups (butane), the steric hindrance between A and B leads to a rotational barrier of ~25 kJ ( 6 kcal) per mole. The consequence of this simple fact is that in fatty acids and related substances and in polyethylene the chains of $CH_2$ groups tend to assume fully extended zigzag conformations.

In addition to this extended conformation there are two **gauche** (**skewed** or synclinal) conformations which are only slightly less stable than the staggered conformation and in which A and B interfere only if they are very bulky. In one of the two gauche conformations B lies to the right of A and in the other to the left of A when viewed down the axis.




One of the two gauche or synclinal conformations

These two conformations are related to right-handed and left-handed screws, respectively. The threads on an ordinary right-handed household screw, when viewed down the axis from either end, move backward from left to right in the same fashion as do the groups A and B in the illustration. The angle ϕ is the **torsion angle** and is positive for right-handed conformations. Gauche conformations are important in many biological molecules; for example, the sugar alcohol **ribitol** stacks in crystals in a "sickle" conformation,[6] in which the chain starts out (at the left) in the zigzag arrangement but shifts to a gauche conformation around the fourth carbon atom, thereby minimizing steric interference between the OH groups on the second and fourth carbons.




Ribitol in sickle conformation

The complete series of possible conformations is shown in Fig. 2-2.

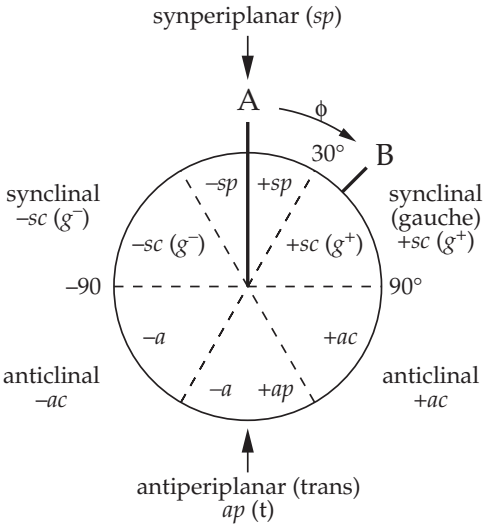
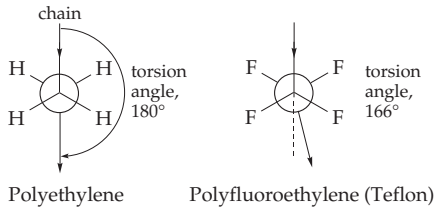



**Figure 2-2** Description of conformations about a single bond in the terminology of Klyne and Prelog[10,11] using the Newman projection. Group A is on the front atom at the top: the conformation is given for each possible position of group B on the other atom.

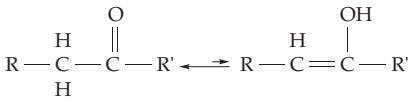
In the chain of methylene units, the hydrogen atoms on alternate carbon atoms of the fully extended chain barely touch (Fig. 2-1) but larger atoms cannot be accommodated. Thus, when fluorine atoms of van der Waals radius 0.135 nm replace the hydrogen atoms of radius 0.12 nm, a fully extended chain is no longer possible. For this reason the torsion angle in polyfluoroethylene is changed from the 180° of polyethylene to 166°, enough to relieve the congestion but not enough to cause severe eclipsing of the fluorines on adjacent carbons. The resulting **helical structure** is reminiscent of those occurring in proteins and other biopolymers. We see that helix formation can be a natural result of steric hindrance between groups of atoms.




Conformations of ring-containing molecules are dealt with in Chapter 4.

## 6. Tautomerism and Resonance

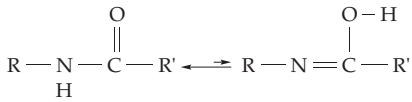
Many simple organic compounds exist as mixtures of two or more rapidly interconvertible isomers or **tautomeric forms**. Tautomers can sometimes be separated one from the other at low temperatures where the rate of interconversion is low. The classic example is the **oxo-enol** (or keto-enol) equilibrium (Eq. 2-1).



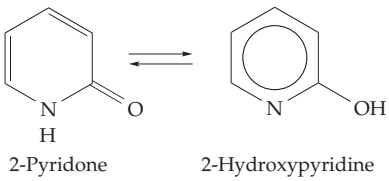

(2-1)

Although usually less stable than the oxo (keto) form, the enol is present in a small amount. It is formed readily from the oxo tautomer by virtue of the fact that hydrogen atoms attached to carbon atoms that are immediately adjacent to carbonyl (C=O) groups are remarkably acidic. Easy dissociation of a proton is a prerequisite for tautomerism. Since most hydrogen atoms bound to carbon atoms do not dissociate readily, tautomerism is unusual unless a carbonyl or other "activating group" is present.

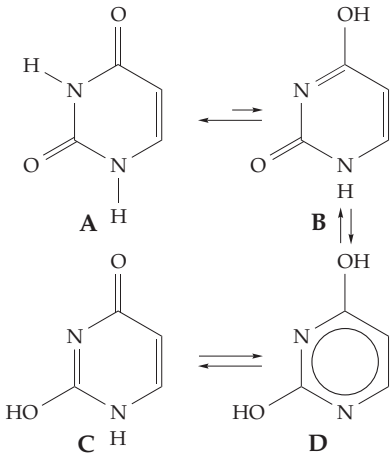
Since protons bound to oxygen and nitrogen atoms usually *do* dissociate readily, tautomerism also exists in amides and in ring systems containing O and N (Eqs. 2-2 to 2-5).



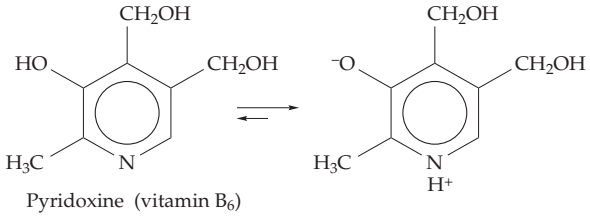

(2-2)




(2-3)




(2-4)



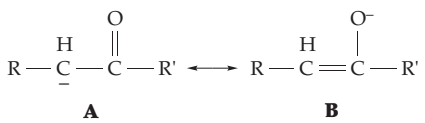

(2-5)

The tautomerism in Eq. 2-2 is the counterpart of that in the oxo-enol transformation. However, the equilibrium constant for aqueous conditions favors form A very strongly. 2-Pyridone is tautomerized to 2-hydroxypyridine (Eq. 2-3) to a greater extent. Pyrimidines (Eq. 2-4) and purines can form a variety of tautomers. The existence of form D of Eq. 2-4 is the basis for referring to uracil as dihydroxypyrimidine. However, the di-oxo tautomer A predominates. Pyridoxine (vitamin $B_6$) exists in water largely as the dipolar ionic tautomer B (Eq. 2-5) but in methanol as the uncharged tautomer A. In a pair of tautomers, a hydrogen atom always moves from one position to another and the lengths and bond character of these bonds also change.

The equilibrium constant for a tautomeric interconversion is simply the ratio of the mole fractions of the two forms; for example, the ratio of enol to oxo forms of acetone[12] in water at 25°C is $6.0 \times 10^{-9}$, while that for isobutyraldehyde is $1.3 \times 10^{-4}$. The ratio of 2-hydroxypyridine to 2-pyridone is about $10^{-3}$ in water but increases to 0.6 in a hydrocarbon solvent and to 2.5 in the vapor phase.[13,14] The ratio of dipolar ion to uncharged pyridoxine (Eq. 2-5) is ~4 at 25°C in water.[15] The ratios of tautomers B, C, and D to the tautomer A of uracil (Eq. 2-4) are small, but it is difficult to measure them quantitatively.[16] These tautomeric ratios are defined for given overall states of protonation (see Eq. 6-82). The constants are independent of pH but will change if the overall state of protonation of the molecule is changed. They may also be altered by

changes in temperature or solvent or by binding to a protein or other molecule.

It is important to distinguish tautomerism from **resonance**, a term used to indicate that the properties of a given molecule cannot be represented by a *single* valence structure but can be represented as a hybrid of two or more structures in which all the nuclei remain in the same places. Only bonding electrons move to convert one resonance form into another. Examples are the **enolate anion**, which can be thought of as a hybrid of structures A and B, and the amide linkage, which can be represented by a similar pair of resonance forms.

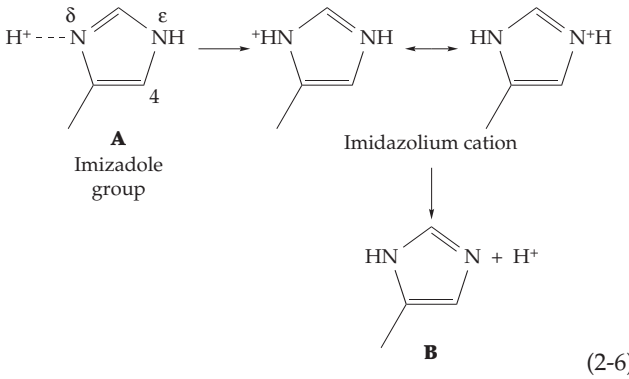



**A** **B**

Enolate anion

A double-headed arrow is often used to indicate that two structures drawn are resonance structures rather than tautomers or other separable isomers.

Although they are distinctly different, tautomerism and resonance are related. Thus, the acidity of carbon-bound hydrogens in ketones, which allows formation of enol tautomers, results from the fact that the enolate anion produced by dissociation of one of these hydrogens is stabilized by resonance. Similarly, tautomerism in the imidazole group of the amino acid histidine is related to resonance in the imidazolium cation. Because of this resonance, if a proton approaches structure A of Eq. 2-6 and becomes attached to the left-hand nitrogen atom ($N^{\delta}$), the positive charge in the resulting intermediate is distributed over both nitrogen atoms. This makes the proton on $N^{\varepsilon}$ acidic, permitting it to dissociate to tautomer B.




**A** Imizadole group

Imidazolium cation

**B**

(2-6)

Note: The nitrogen atom designated $N^{\delta}$ (or ND1)[11] in Eq. 2-6 may also be called N1 or $N^{\pi}$ (*pros*-N). Likewise, $N^{\varepsilon}$ (NE2) may be designated N3 or $N^{\tau}$ (*tele*- N). Since $N^{\sigma}$ has sometimes also been called N3, it is best not to use the numerical designations for the nitrogen atoms.

The tautomeric ratio of B to A for histidine in water (Eq. 2-6) has been estimated, using $^{15}$N- and $^{13}$C-NMR, as 5.0 when the α-amino group is protonated and as 2.5 when at high pH it is unprotonated.[17] This tautomerism of the imidazole group is probably important to the function of many enzymes and other proteins; for example, if $N^{\varepsilon}$ of structure A (Eq. 2-6) is embedded in a protein, a proton approaching from the outside can induce the tautomerism shown with the release of a proton in the interior of the protein, perhaps at the active site of an enzyme. The form protonated on $N^{\delta}$ (B of Eq. 2-6), which is the minor form in solution, predominates in some positions within proteins.[18]

## B. Forces between Molecules and between Chemical Groups

The structure of living cells depends very much on the covalent bonds within individual molecules and on covalent crosslinks that sometimes form between molecules. However, weaker forces acting between molecules and between different parts of the same molecule are responsible for many of the most important properties of biochemical substances. These are described as **van der Waals forces**, **electrostatic forces**, **hydrogen bonds**, and **hydrophobic interactions**. In the discussion that follows the thermodynamic quantities $\Delta H$, $\Delta S$, and $\Delta G$ will be used. If necessary, please see Chapter 6 for definitions and a brief review.

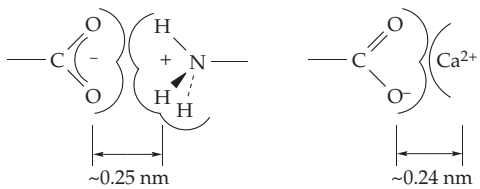
### 1. Van der Waals Forces

All atoms have a weak tendency to stick together, and because of this even helium liquifies at a low enough temperature. This is a result of the van der Waals or "London dispersion forces" which act strongly only at a very short distance. These forces arise from electrostatic attraction between the positively charged nucleus of one atom and the negatively charged electrons of the other.[19–21] Because nuclei are screened by the electron clouds surrounding them, the force is weak. The energy (enthalpy) of binding of one methylene ($-CH_2-$) unit into a monomolecular layer of a fatty acid is about $-\Delta H° = 1.7$ kJ/mol.[22] Although this is a small quantity, when summed over the 16 or more carbon atoms of a typical fatty acid the binding energy is substantial. When a methylene group is completely surrounded in a crystalline hydrocarbon, its van der Waals energy, as estimated from the heat of sublimation, is 8.4 kJ/mol; that of $H_2O$ in liquid water at the melting point of ice is 15 kJ/mol.[22]

While van der Waals forces between individual atoms act over very short distances, they can be felt at surprisingly great distances when exerted by large molecules or molecular aggregates.[23] Forces between very smooth surfaces have been measured experimentally at distances as great as 10 nm and even to 300 nm.[23a] However, these "long-range van der Waals forces" probably depend upon layers of oriented water molecules on the plates[23] (see also Section 5).

## 2. Attraction between Charged Groups (Salt Linkages)

Fixed positive and negative charges attract each other strongly. Consider a carboxylate ion in contact with $-NH_3^+$ or with an ion of calcium:




From the van der Waals radii of Table 2-1 and the ionic crystal radius of $Ca^{2+}$ of 0.10 nm, we can estimate an approximate distance between the centers of positive and negative charge of 0.25 nm in both cases. It is of interest to apply Coulomb's law to compute the force $F$ between two charged particles which are almost in contact. Let us choose a distance of 0.40 nm (4.0 Å) and apply Eq. 2-7.

$$F = 8.9875 \times 10^9 \times \frac{qq'}{r^2\varepsilon} \text{ newtons} \tag{2-7}$$

In this equation $r$ is the distance in meters, $q$ and $q'$ are the charges in coulombs (one electronic charge = $1.6021 \times 10^{-19}$ coulombs), ε is the dielectric constant, and $F$ is the force in newtons (N). The force per mole is $NF$ where $N$ is Avogadro's number.

An uncertainty in this kind of calculation is in the dielectric constant ε, which is 1.0 for a vacuum, about 2 for hydrocarbons, and 78.5 for water at 25°C. If ε is taken as 2, the force for $r$ = 0.40 nm is $4.3 \times 10^{14}$ N/mol. The force would be twice as great for the $Ca^{2+} - COO^-$ case. To move two single charges further apart by just 0.01 nm would require 4.3 kJ/mol, a substantial amount of energy. However, if the dielectric constant were that of water, this would be reduced almost 40-fold and the electrostatic force would not be highly significant in binding. It is extremely difficult to assign a dielectric constant for use in the interior of proteins.[23b] For charges spaced far apart within proteins the effective dielectric constant is usually as high as 30–60.[24] For closely spaced charges in hydrophobic niches it may be as low as 2–4.[25–27]

A calculation that is often made is the work required to remove completely two charges from a given distance apart (e.g., 0.40 nm) to an infinite distance (Eq. 2-8).

$$W \text{ (kJ mol}^{-1}) = 8.9875 \times 10^6 \times \frac{qq'}{r\varepsilon} \text{ N} = \frac{138.9}{\varepsilon r \text{ (in nm)}} \tag{2-8}$$

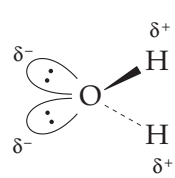
If ε = 2, this amounts to 174 kJ/mol for single charges at a distance of 0.40 nm; 69 kJ/mol at 1 nm; and only 6.9 kJ/mol at 10 nm, the distance across a cell membrane. We see that very large forces exist between closely spaced charges.

Electrostatic forces are of great significance in interactions between molecules and in the induction of changes in conformations of molecules. For example, attraction between $-COO^-$ and $-NH_3^+$ groups occurs in interactions between proteins. Calcium ions often interact with carboxylate groups, the doubly charged $Ca^{2+}$ bridging between two carboxylate or other polar groups. This occurs in carbohydrates such as agarose, converting solutions of these molecules into rigid gels (Chapter 4).

Individual macromolecules as well as cell surfaces usually carry a net negative charge at neutral pH. This causes the surfaces to repel each other. However, at a certain distance of separation the van der Waals attractive forces will balance the electrostatic repulsion.[21] Protruding hydrophobic groups may then interact and may "tether" bacteria or other particles at a fixed distance, often ~5 nm, from a cell surface.[28]

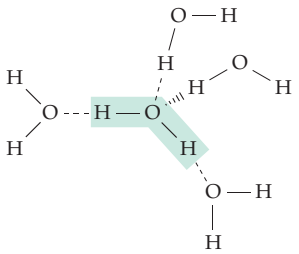
## 3. Hydrogen Bonds

One of the most important weak interactions between biologically important molecules is the hydrogen bond (H-bond).[29,30,30a] These "bonds" are the result of electrostatic attraction caused by the uneven distribution of electrons within covalent bonds. For example, the bonding electron pairs of the H–O bonds of water molecules are attracted more tightly to the oxygen atoms than to the hydrogen atoms. A small net positive charge is left on the hydrogen and a small net negative charge on the oxygen. Such **polarization** of the water molecules can be indicated in the following way:



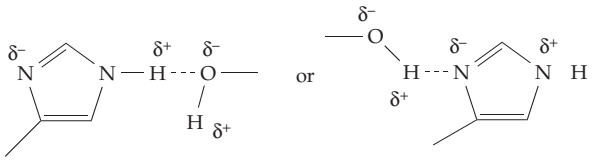
Here the $\delta^+$ and $\delta^-$ indicate a fraction of a full charge present on the hydrogen atoms and on the nonbonded electron pairs of the oxygen atom, respectively. Molecules such as $H_2O$, with strongly polarized bonds, are referred to as **polar molecules** and functional groups with such bonds as **polar groups**. They are to be contrasted with such nonpolar groups as the $-CH_3$ group in which the electrons in the bonds are nearly equally shared by carbon and hydrogen.

A hydrogen bond is formed when the positively charged end of one of the **dipoles** (polarized bonds) is attracted to the negative end of another dipole. Water molecules tend to hydrogen bond strongly one to another; each oxygen atom can be hydrogen-bonded to two other molecules and each hydrogen to another water molecule. Thus, every water molecule can have up to four hydrogen-bonded neighbors.



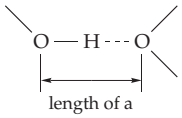
A water molecule hydrogen bonded to four other water molecules; note the tetrahedral arrangement of bonds around the central oxygen.

Many groups in proteins, carbohydrates, and nucleic acids form hydrogen bonds to one another and to surrounding water molecules. For example, an imidazole group of a protein can bond to an OH group of an amino acid side chain or of water in the following ways:



Remember that hydrogen bonds are always formed between pairs of groups, with one of them, often C=O or C=N-, containing the negative end of a dipole and the other providing the proton. The proton acceptor group, often OH or NH and occasionally SH, and even CH in certain structures,[31,31a,31b] donates an unshared pair of electrons. Dashed arrows are sometimes drawn from the hydrogen atom to the electron donor atom to indicate the direction of a hydrogen bond. Do not confuse these arrows with the curved arrows that indicate flow of electrons in organic reactions.

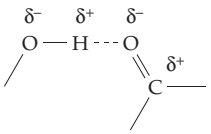
The strength of hydrogen bonds, as measured by the bond energy, varies over the range 10–40 kJ/mol. The stronger the hydrogen bond the shorter its length. Because hydrogen atoms can usually not be seen in X-ray structures of macromolecules, the lengths of hydrogen bonds are often measured between the surrounding heavy atoms:



length of a hydrogen bond

A typical —OH---O hydrogen bond will have a length of about 0.31 nm; a very strong hydrogen bond may be less than 0.28 nm in length, while weak hydrogen bonds will approach 0.36 nm, which is the sum of the van der Waals contact distances plus the O–H bond length. Beyond this distance a hydrogen bond cannot be distinguished easily from a van der Waals contact.

Hydrogen bonds are strongest when the hydrogen atom and the two heavy atoms to which it is bonded are in a straight line. For this reason hydrogen bonds tend to be linear. However, *the dipoles forming the hydrogen bond do not have to be colinear for strong hydrogen bonding:*[32,32a] There is some preference for hydrogen bonding to occur in the direction of an unshared electron pair on the oxygen or nitrogen atom.[33–35]



A linear O–H- - -O hydrogen bond with dipoles at an angle one to another.

Both ammonia, $NH_3$, and the $-NH_2$ groups of proteins are good electron donors for hydrogen bond formation. However, the hydrogen atoms of uncharged $-NH_2$ groups tend to be poor proton donors for H-bonds.[36] Do hydrogen bonds have some covalent character? The answer is controversial.[36a,36b,36c]

Hydrogen bonding is important both to the internal structure of biological macromolecules and in interactions between molecules. Hydrogen bonding often provides the specificity necessary to bring surfaces together in a complementary way. Thus, the location of hydrogen-bond forming groups in surfaces between molecules is important in ensuring an exact alignment of the surfaces.[37] The hydrogen bonds do not always have to be strong. For example, Fersht and coworkers, who compared a variety of mutants of an enzyme of known three-dimensional structure, found that deletion of a side chain that formed a good hydrogen bond to the substrate weakened the binding energy by only 2–6 kJ/mol. However, loss of a hydrogen bond to a charged group in the substrate caused a loss of 15–20 kJ/mol of binding energy.[37] Study of mutant

proteins created by genetic engineering is now an important tool for experimentally investigating the biological roles of hydrogen bonding.[37–39]

## 4. The Structure and Properties of Water

Water is the major constituent of cells and a remarkable solvent whose chemical and physical properties affect almost every aspect of life. Many of these properties are a direct reflection of the fact that most water molecules are in contact with their neighbors entirely through hydrogen bonds.[40–48] Water is the only known substance for which this is true.

In ordinary ice all of the water molecules are connected by hydrogen bonds, six molecules forming a hexagonal ring resembling that of cyclohexane. The structure is extended in all directions by the formation of additional hydrogen bonds to adjacent molecules (Fig. 2-3). As can be seen in this drawing, the molecules in ice assume various orientations in the hexagonal array, and frequently rotate to form their hydrogen bonds in different ways. This randomness remains as the temperature is lowered, and ice is one of few substances with a residual entropy at absolute zero.[49,50] Ice is unusual also in that the molecules do not assume closest packing in the crystal but form an open structure. The hole through the middle of the hexagon and on through the hexagons lying below it is ~0.06 nm in diameter.

The short hydrogen-bond length (averaging 0.276 nm) in ice indicates of strong bonding. The heat of sublimation ($\Delta H°$) of ice is −48.6 kJ/mol. If the van der Waals dispersion energy of −15 kJ/mol is subtracted from this, the difference of −34 kJ/mol can be attributed entirely to the hydrogen bonds—two for each molecule. Their average energy is 17 kJ/mol apiece. However, some of the hydrogen bonds are stronger and others weaker than the average.[51]

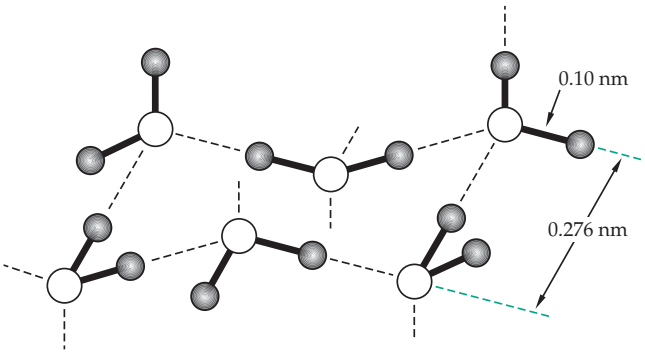
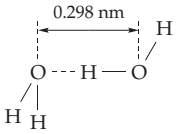
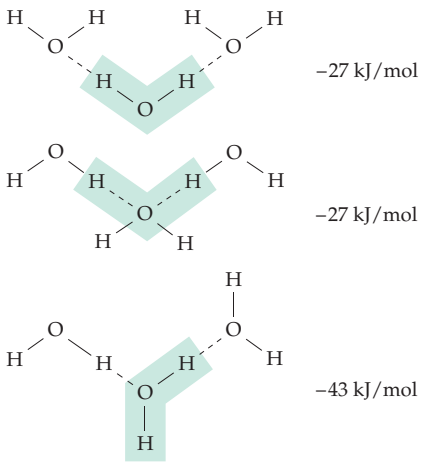



**Figure 2-3** Six water molecules in the lattice of an ice crystal. The hydrogen bonds, which connect protons with electron pairs of adjacent molecules, are shown as dashed lines.

In a gaseous water dimer the hydrogen bond is linear, a fact that suggests some covalent character.[52]




Its length is distinctly greater than that in ice. This is one of a number of pieces of evidence suggesting cooperativity in formation of chains of hydrogen bonds.[40,53,54] Consider the following three trimers for which theoretical calculations have predicted the indicated hydrogen bond energies.[40] In the first case the central water molecule *donates* two protons for hydrogen-bond formation; in the second it *accepts* the protons. In the third case it is both an electron acceptor and a donor. The OH dipoles are oriented "head to tail" and the hydrogen bonds are stronger than in the other




cases. Long chains of similarly oriented hydrogen bonds exist in ice and this may account for the short hydrogen bond lengths. Closed rings of hydrogen bonds oriented to give a maximum cooperative effect also exist in liquid water clusters[55] and within proteins, carbohydrates, and nucleic acids.[53,54]

The nature of liquid water is still incompletely understood,[46-48,56] but we know that water contains ice-like clusters of molecules that are continually breaking up and reforming in what has been called a "flickering cluster" structure. Judging by the infrared spectrum of water, about 10% of the hydrogen bonds are broken when ice melts.[41] A similar conclusion can be drawn from the fact that the heat of melting of ice is −5.9 kJ/mol. It has been estimated that at 0°C the average cluster contains about 500 water molecules.[41] At 50°C there are over 100 and at the boiling point about 40. Although most molecules in liquid water are present in these clusters, the hydrogen bonds are rapidly broken and reformed in new ways, with the average lifetime of a given hydrogen bond being ~$10^{-12}$ s.

## 5. Hydration of Polar Molecules and Ions

Water molecules are able to hydrogen bond not only to each other but also to polar groups of dissolved compounds. Thus, every group that is capable of forming a hydrogen bond to another organic group is also able to form hydrogen bonds of a somewhat similar strength with water. For this reason, hydrogen bonding is usually not a significant force in holding small molecules together in aqueous solutions. Polar molecules that stick together through hydrogen bonding when dissolved in a nonpolar solvent often do not associate in water. How then can biochemists assert that hydrogen bonding is so important in biochemistry? Part of the answer is that proteins and nucleic acids can be either properly folded with hydrogen bonds formed internally or denatured with hydrogen bonds from those same groups to water. The Gibbs energy change between these two states is small.

Every ion in an aqueous solution is surrounded by a shell of oriented water molecules held by the attraction of the water dipoles to the charged ion. The hydration of ions has a strong influence on all aspects of electrostatic interactions and plays a dominant role in determining such matters as the strength of acids and bases, the Gibbs energy of hydrolysis of ATP, and the strength of bonding of metal ions to negatively charged groups. For example, the previously considered interaction between carboxylate and calcium ions would be much weaker if both ions retained their hydration shells.

Consider the following example. $\Delta G°$ for dissociation of acetic acid in water is +27.2 kJ/mol (Table 6-5). The enthalpy change $\Delta H°$ for this process is almost zero (–0.1 kJ/mol) and $\Delta S°$ is consequently –91.6 J $K^{-1}$. This large entropy decrease reflects the increased amount of water that is immobilized in the hydration spheres of the $H^+$ and acetate$^-$ ions formed in the dissociation reaction. In contrast, dissociation of $NH_4^+$ to $NH_3$ and $H^+$ converts one positive ion to another. $\Delta H°$ is large (+52.5 kJ/mol) but the entropy change $\Delta S°$ is small (–2.0 J $K^{-1}$, Table 6-5).

Although effects of hydration are important in almost all biochemical equilibria, they are difficult to assess quantitatively. It is hard to know how many molecules of water are freed or immobilized in a given reaction. Charged groups in proteins are often hydrated. However, if they are buried in the interior of the protein, they may be solvated by polarizable protein side chain groups such as –OH or by backbone or side chain amide groups.[57,58]

## 6. Hydrophobic Interactions

Fats, hydrocarbons, and other materials whose molecules consist largely of nonpolar groups have a low solubility in water and a high solubility in nonpolar solvents. Similarly, the long alkyl groups of fatty acid esters aggregate within membranes and nonpolar side chains of proteins are often packed together in the centers of protein molecules. Because it is as if the nonpolar groups "fear" water, this is known as the **hydrophobic** effect.[59–70] The terms hydrophobic forces, hydrophobic interactions, and hydrophobic bonding have also been used. However, the latter term can be misleading because the hydrophobic effect arises not out of any special attraction between nonpolar groups but primarily from the strong internal cohesion of the hydrogen-bonded water structure.

How strongly do nonpolar groups "attract" each other in water? A partial answer can be obtained by measuring the standard Gibbs energy $\Delta G$ of transfer of a hydrocarbon molecule from a dilute aqueous solution into a dilute solution in another hydrocarbon. By studying a series of alkanes, Abraham[62] calculated the Gibbs energy change per $CH_2$ unit (Eq. 2-9) as: $\Delta G° = -3.8$ kJ/mol.

$$-CH_2-(\text{water}) \longrightarrow -CH_2-(\text{hexane}) \qquad \Delta G° = -3.8 \text{ kJ/mol} \tag{2-9}$$

This equation is a quantitative statement of the fact that the $CH_2$ group prefers to be in a nonpolar environment than to be surrounded by water. A similar Gibbs energy change would be expected to accompany the bringing together of a methylene unit from a small molecule and a hydrophobic surface on a protein molecule. However, in the latter case the accompanying decrease in entropy would make $\Delta G°$ less negative.

What causes the decrease in Gibbs energy when nonpolar groups associate in water? Jencks[60] suggested that we think of the transfer of a nonpolar molecule from a nonpolar solvent into water in two steps: (1) Create a cavity in the water of about the right size to accommodate the molecule. Since many hydrogen bonds will be broken, the Gibbs energy of cavity formation will be high. It will be principally an enthalpy ($\Delta H$) effect. (2) Allow the water molecules in the solvent to make changes in their orientations to accommodate the nonpolar molecule that has been placed in the cavity. The water molecules can move to give good van der Waals contacts and also reorient themselves to give the maximum number of hydrogen bonds. Since hydrogen bonds can be formed in many different ways in water, there may be as many or even more hydrogen bonds after the reorientation than before. This will be true especially at low temperature where most water exists as large icelike clusters. For dissolved hydrocarbons, the enthalpy of formation of the new hydrogen bonds often almost exactly balances the enthalpy of creation of the cavity initially so that $\Delta H$ for the overall process (transfer from inert solvent into water) is small. For the opposite transfer (from water to hydrocarbon hexane; Eq. 2-9) $\Delta H°$ is usually a small

positive number for aliphatic hydrocarbons and is nearly zero for aromatic hydrocarbons.

Since $\Delta G° = \Delta H° - T\Delta S°$ (see Chapter 6), it follows that the negative value of $\Delta G°$ for hydrophobic interactions must result from a positive entropy change, which may arise from the restricted mobility of water molecules that surround dissolved hydrophobic groups. When two hydrophobic groups come together to form a "hydrophobic bond," water molecules are freed from the structured region around the hydrophobic surfaces and the entropy increases. The $\Delta S°$ for Eq. 2-9 is about 12 J $deg^{-1}$ $mol^{-1}$. Attempts have been made to relate this value directly to the increased number of orientations possible for a water molecule when it is freed from the structured region.[64] However, interpretation of the hydrophobic effect is complex and controversial.[65–71a]

The formation constant $K_f$ for hydrophobic associations often increases with increasing temperature. This is in contrast to the behavior of $K_f$ for many association reactions that involve polar molecules and for which $\Delta H°$ is often strongly negative (heat is released). An example of the latter is the protonation of ammonia in an aqueous solution (Eq. 2-10).

$$NH_3 + H^+ \rightarrow NH_4^+$$
$$\Delta H° = -52.5 \text{ kJ/mol} \qquad (2\text{-}10)$$

Since $R1nK_f = -\Delta G°/T = -\Delta H°/T + \Delta S°$, $K_f$ decreases with increasing temperature if $\Delta H°$ is negative. Because for a hydrophobic interaction with a positive value of $\Delta H°$ $K_f$ *increases* with increasing temperature, an increase in stability at higher temperatures is sometimes used as a criterion for hydrophobic bonding. However, this criterion does not always hold. For example, base stacking interactions in polynucleotides (Chapter 5), whose strength does not increase with increasing temperature, are still thought to be hydrophobic.

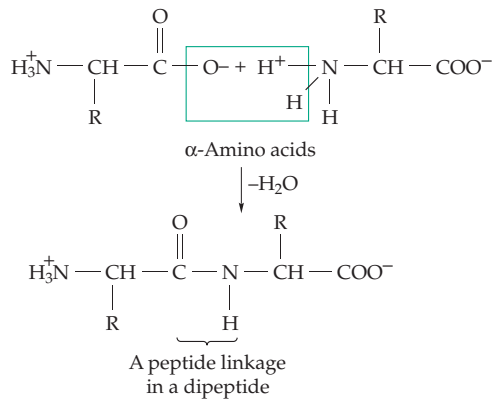
The water molecules that are in immediate contact with dissolved nonpolar groups are partially oriented. They form a cagelike structure around each hydrophobic group. When particles surrounded by such hydration layers are 1–2 nm apart, they sometimes experience either a fairly strong repulsion or an enhanced attraction caused by these hydration layers.[21,64–66,72] Direct experimental measurements have shown that these effects extend to distances of 10 nm[21,63] and can account for the previously mentioned long-range van der Waals forces.

Various efforts have been made to develop scales of **hydrophobicity** that can be used to predict the probability of finding a given amino acid side chain buried within a protein or in a surface facing water.[59,73] A new approach has been provided by the study of mutant proteins. For example, deletion of a single $-CH_2-$ group from an interior hydrophobic region of a protein was observed to decrease the stability of the protein by 4.6 kJ/mol.[74]

## C. Amino Acids and Peptides

Twenty α-amino acids are the monomers from which proteins are made. These amino acids share with other biochemical monomers a property essential to their role in polymer formation: *They contain at least two different chemical groups able to react with each other to form a covalent linkage*. In the amino acids these are the protonated amino ($NH_3^+$) and carboxylate ($COO^-$) groups. The characteristic linkage in the protein polymer is the **peptide** (amide) linkage whose formation can be imagined to occur by the splitting out of water between the carboxyl of one amino acid and the amino group of another (Eq. 2-11).

This equation is not intended to imply a mechanism for peptide synthesis. The equilibrium position for this reaction in an aqueous solution favors the free amino acids rather than the peptide. Therefore, both biological and laboratory syntheses of peptides usually do not involve a simple splitting out of water. Since the dipeptide of Eq. 2-11 still contains reactive carboxyl and amino groups, other amino acid units can be joined by additional peptide linkages to form **polypeptides**. These range from short-chain **oligomers** to polymers of from ~50 to several thousand amino acid units, the proteins.[75–77]




(2-11)

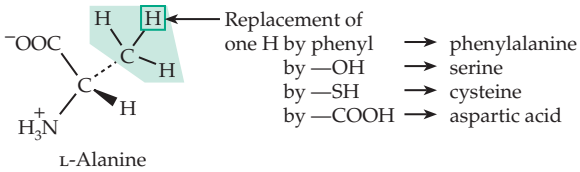
### 1. Properties of α-Amino Acids

The amino acids have in common a dipolar ionic structure and a chiral center. They are differentiated, one from another, by the structures of their **side chain groups**, designated R in the foregoing formulas. These groups are of varying size and chemical structure. The side chain groups fill much of the space in the interior of a protein molecule and also protrude from the external surfaces of the protein where they determine many of the chemical and physical properties of the molecule.

Table 2-2 shows the structures of the side chains

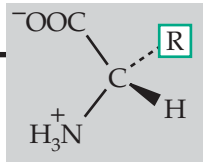
of the amino acids commonly found in proteins. The complete structure is given for proline. Both the three-letter abbreviations and one-letter abbreviations used in describing sequences of amino acids in proteins are also given in this table. Amino acids of groups **a**–**c** of Table 2-2 plus phenylalanine and methionine are sometimes grouped together as *nonpolar*. They tend to be found in a hydrophobic environment on the inside of a protein molecule. Groups **f** and **i** contain *polar*, *charged* side chains which usually protrude into the water surrounding the protein. The rest are classified as *polar* but *noncharged*.

To get acquainted with amino acid structures, learn first those of **glycine**, **alanine**, **serine**, **aspartic acid**, and **glutamic acid**. The structures of many other amino acids can be related to that of alanine ($R=CH_3$) by replacement of a β hydrogen by another group. Metabolic interrelationships will make it easier to learn structures of the rest of the amino acids later.




Since the –COOH groups of glutamic and aspartic acids are completely dissociated to $-COO^-$ at neutral pH, it is customary in the biochemical literature to refer to these amino acids as **glutamate** and **aspartate** without reference to the nature of the cation or cations present as counter ions. Such "-ate" endings are also used for most other acids (e.g., malate, oxaloacetate,

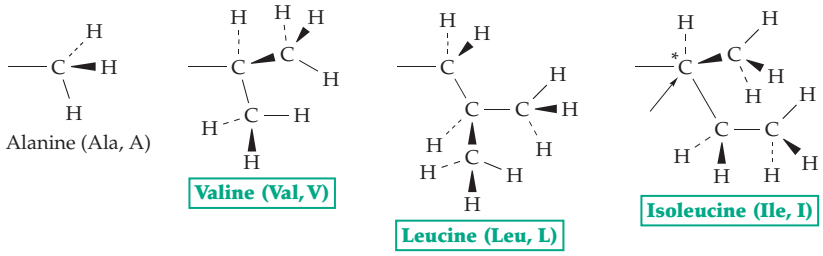
## TABLE 2-2
## Structure and Chemical Properties of Side Chain Groups (R) of Amino Acids



**a.** Glycine "side chain" = —H (Gly, G)[a]

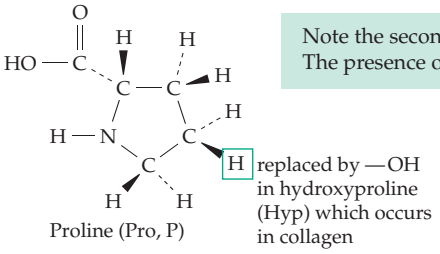
Strictly a link in the peptide chain, glycine provides a minimum of steric hindrance to rotation and to placement of adjacent groups.

**b.** Amino acids with alkyl groups as side chains



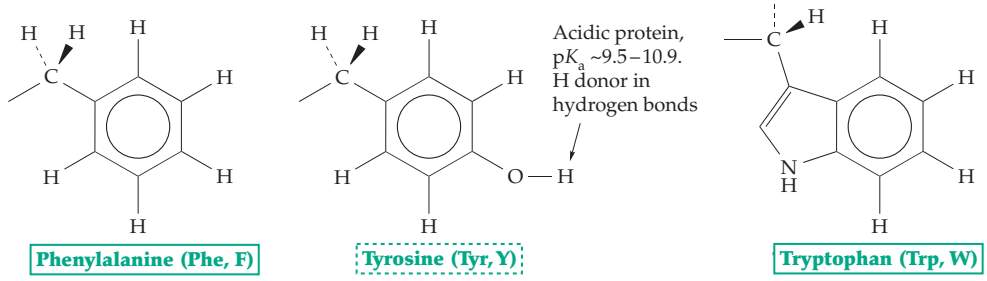

These bulky groups of distinctive shapes participate in hydrophobic interactions in protein interiors and in forming binding sites of specific shapes.

**c.** The imino acid proline. Because the side chain is fused to the α-amino group, the entire structure, not just the side chain, is shown.




Note the secondary amino group and the relatively rigid conformation. The presence of proline strongly influences the folding of protein chains.

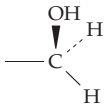
**d.** Aromatic amino acids



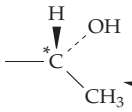

Tyrosine, phenylalanine, and tryptophan can form hydrophobic bonds and may be especially effective in bonding to other flat molecules.

**TABLE 2-2**
*(continued)*

**e.** Amino acid alcohols



Serine (Ser, S)

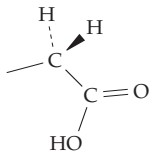


**Threonine (Thr, T)**

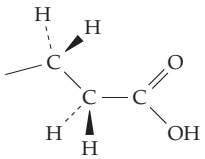
Note the second chiral center. The L-amino acid with the opposite configuration in the side chain is L-*allo*-threonine.

The —OH group is very weakly acidic ($pK_a$ ~13.6). It can form esters with phosphoric acid or organic acids and is a site of attachment of sugar rings in glycoproteins. Hydroxyl groups of serine are found at the active centers of some enzymes.

**f.** Acidic amino acids



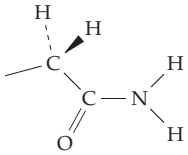
Aspartic acid (Asp, D)



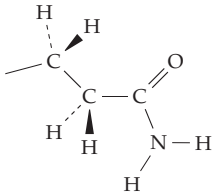
Glutamic acid (Glu, E)

Carboxyl groups of these side chains are dissociated at neutral pH ($pK_a$ values are 4.3–4.7) and provide anionic (–) groups on the surfaces of proteins.

**g.** Amides of aspartic acid and glutamic acid



Asparagine (Asn, N)

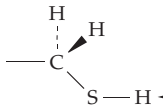


Glutamine (Gln, Q)

The amide group is not acidic but is polar and participates in hydrogen bonding.

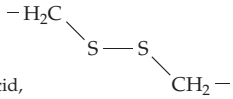
If it is uncertain whether a position in a protein is occupied by aspartic acid or asparagine, it may be designated Asx or B. If glutamic or glutamine, it may be designated as Glx or Z.

**h.** Sulfur-containing amino acids

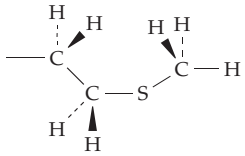


Cysteine (Cys, C)

A weak acid, $pK_a$ ~8.5



Two cysteine SH groups can be oxidatively joined to form a disulfide bridge in the "double-headed" amino acid cystine.



**Methionine (Met, M)**

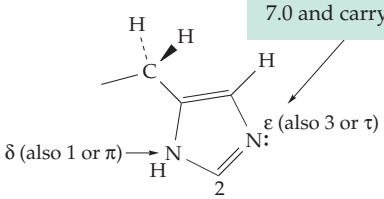
**Key**

- **Essential in the human diet**
- **Essential if phenylalanine or cysteine is inadequate**
- **Essential for rapid growth**

[a] The three-letter and one-letter abbreviations used for the amino acid residues in peptides and proteins are given in parentheses. B, J, U, X, and Z can be used to indicate modified or unusual amino acids.

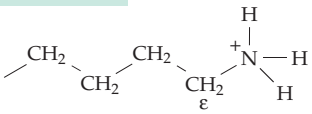
**i.** Basic amino acids

This basic site accepts a proton to form a conjugate acid of $pK_a$ ~6.4–7.0 and carrying a positive charge.



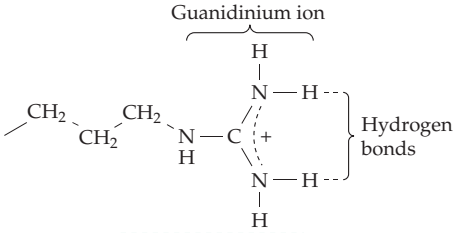
**Histidine (His, H)**

The **imidazole** groups in histidine side chains are parts of the active sites of many enzymes. Like other basic groups in proteins they also may bind metal ions.



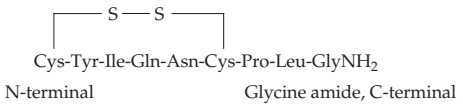
**Lysine (Lys, K)**

A flexible side arm with a potentially reactive amino group at the end. The high $pK_a$ of ~10.5 means that lysine side chains are ordinarily protonated in neutral solutions.

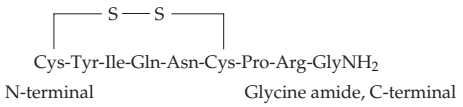



**Arginine (Arg, R)**

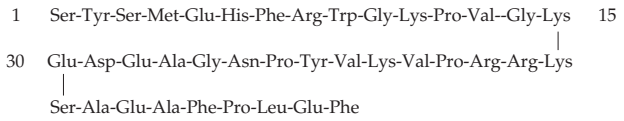
The guanidinium group has a high $pK_a$ of over 12 and remains protonated under most circumstances. It is stabilized by resonance as indicated by the dashed line. Guanidinium groups are often sites for binding of phosphate or carboxylate groups by pairs of hydrogen bonds.



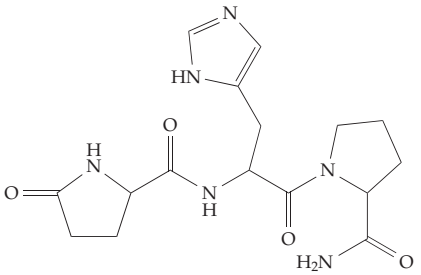

Human oxytocin



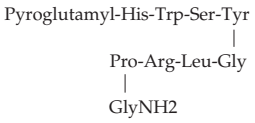

Human vasopressin



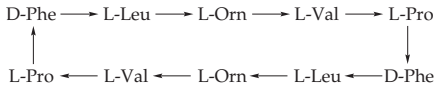

Human adrenocorticotropin (ACTH)



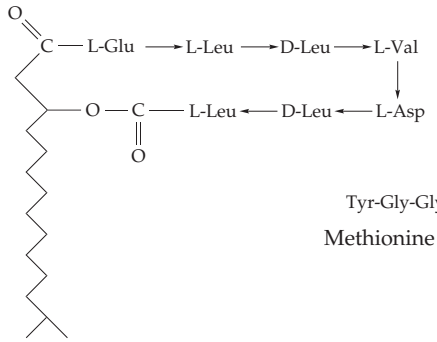
Hypothalamic thyrotropic hormone releasing factor (TRF); pyroglutamyl-histidyl-prolinamide




Luteinizing hormone releasing factor (LRF)




Gramicidin S




Surfactin

Tyr-Gly-Gly-Phe-Met

Methionine enkephalin

**Figure 2-4** Structures of some naturally occurring peptides. Oxytocin and vasopressin are hormones of the neurohypophysis (posterior lobe of the pituitary gland). Adrenocorticotropin is a hormone of the adenohypophysis (anterior pituitary). Hormones of the adenohypophysis are released under the influence of releasing factors (regulatory factors) produced in the neighboring hypothalamus (a portion of the brain) in response to neural stimulation. Structures of two releasing factors are shown. Note that the γ-carboxyl groups of the N-terminal glutamine residues have reacted, with loss of $NH_3$, with the neighboring terminal $-NH_2$ groups to form cyclic amide (pyroglutamyl) groups.[78] Gramicidin S is an antibiotic made by *Bacillus brevis*, and surfactin is a depsipeptide (containing an ester linkage), a surface active antibiotic of *Bacillus subtilis*. Methionine enkephalin is a brain peptide with opiate-like activity.[79]

phosphate, and adenylate) and in names of enzymes (e.g., lactate dehydrogenase).

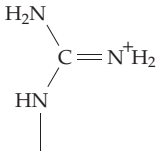
During the formation of polypeptides, the α-amino and carboxyl groups of the amino acids are converted into the relatively unreactive and uncharged amide (peptide) groups except at the two chain termini. In many cases the terminal amino and carboxyl groups are also converted within cells into uncharged groups (Chapter 10). Immediately after the protein is synthesized its terminal carboxyl group is often converted into an amide. The N terminus may be acetylated or cyclized to a pyroglutamyl group. Sometimes a cyclic peptide is formed (Fig. 2-4).

The properties of polypeptides and proteins are determined to a large extent by the chemistry of the side chain groups, which may be summarized briefly as follows. Glycine in a peptide permits a maximum of conformational mobility. The nine relatively nonpolar amino acids–alanine, valine, leucine, isoleucine, proline, methionine, phenylalanine, tyrosine, and tryptophan–serve as building blocks of characteristic shape. Tyrosine and tryptophan also participate in hydrogen bonding and in aromatic: aromatic interactions within proteins.

Much of the chemistry of proteins involves the side chain functional groups –OH, –SH, $-COO^-$, $-NH_3^+$, and imidazole (Eq. 2-6) and the guanidinium group of arginine. The side chains of asparagine and glutamine both contain the amide group $CONH_2$, which is relatively inert chemically but which can undergo hydrogen-bonding interactions. The amide linkages of the polypeptide backbone must also be regarded as important functional groups. Most polar groups are found on the outside surfaces of proteins where they can react chemically in various ways. When inside proteins they form H-bonds to the peptide backbone and to other polar groups.

## 2. Acidic and Basic Side Chains

The side chains of aspartic and glutamic acids carry negatively charged carboxylate groups at pH 7 while those of lysine and arginine carry the positively charged $-NH_3^+$ and guanidinium ions, respectively.




Guanidinium group

At pH 7 the weakly basic imidazole group of histidine may be partially protonated. Both the –SH group of cysteine and the phenolic –OH of tyrosine are weakly acidic and will dissociate and thereby acquire negative charges at a sufficiently high pH.

The number of positive and negative charges on a protein at any pH can be estimated approximately from the acid dissociation constants (usually given as $pK_a$ values) for the amino acid side chains. These are given in Table 2-2. However, $pK_a$ values of buried groups are often greatly shifted from these, especially if they associate as **ion pairs**. In addition, many proteins have free amino and carboxyl-terminal groups at the opposite ends of the peptide chain. These also participate in acid–base reactions with approximately the following $pK_a$ values.

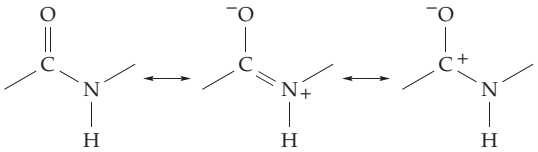
terminal —COOH, $pK_a = 3.6–3.7$
terminal —$NH_3^+$, $pK_a = 7.5–7.9$

The acid–base properties of an amino acid or of a protein are described by titration curves of the type shown in Figs. 3-1 and 3-2. In these curves the number of equivalents of acid or base that have reacted with an amino acid or protein that was initially at neutral pH are plotted against pH. The net negative or positive electrical charge on the molecule can be read directly from the curves. Both the net electrical charges and the distribution of positively and negatively charged groups are often of crucial importance to the functioning of a protein.

Additional aspects of the acid–base chemistry of amino acids and proteins are considered in Chapter 3, Section A and Chapter 6, Section E. The student may find it appropriate to study these sections at this time and to work the associated study problems.

## 3. The Peptide Unit

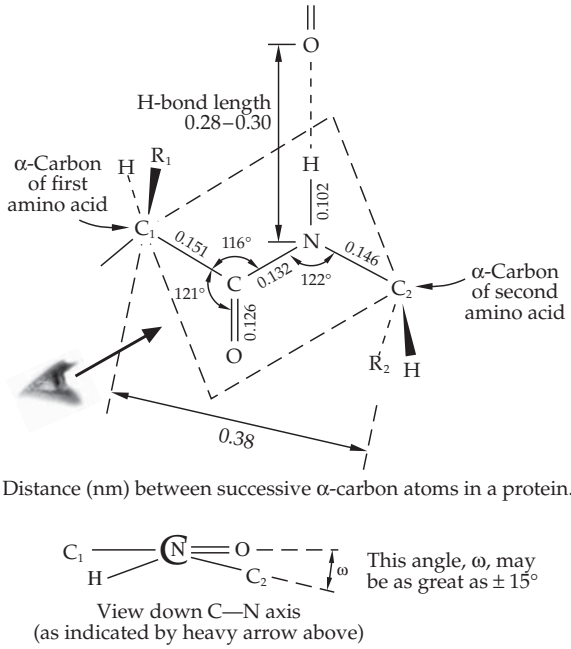
The very ability of a protein to exist as a complex three-dimensional structure depends upon the properties of the amide linkages between the amino acid units. Many of these properties follow from the fact that an amide can be viewed as a resonance hybrid of the following structures. Because of the partial double-bond character, the C–N bond is shorter than that of a normal single bond and the C=O bond is lengthened.




The observed lengths in nanometers determined by X-ray diffraction measurements are given in Fig. 2-5 (top). The partial double-bond character of the C–N bond has important consequences. The peptide unit is nearly planar as is indicated by the dashed parallelogram in Fig. 2-5.

However, the bonds around the nitrogen retain some pyramidal character (Fig. 2-5, bottom). Even more important is the fact that there is flexibility. As a result, the torsion angle ω may vary over a range of ± 15° or even more from that in the planar state.[80,81,81a]

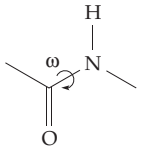
The resonance stabilization of the amide linkage is thought to be about 85 kJ/mol. Rotation around the C–N bond through 90° would be expected to require



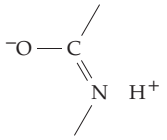

Distance (nm) between successive α-carbon atoms in a protein.

**Figure 2-5** Dimensions of the peptide linkage. Interatomic distances in nm, including the hydrogen bond length to an adjacent peptide linkage, are indicated. The atoms enclosed by the dotted lines all lie *approximately* in a plane. However, as indicated in the lower drawing, the nitrogen atom tends to retain some pyramidal character.

about this much energy. This fact immediately suggests a way in which proteins may sometimes be able to store energy—by having one or more peptide units twisted out of complete planarity.[69]



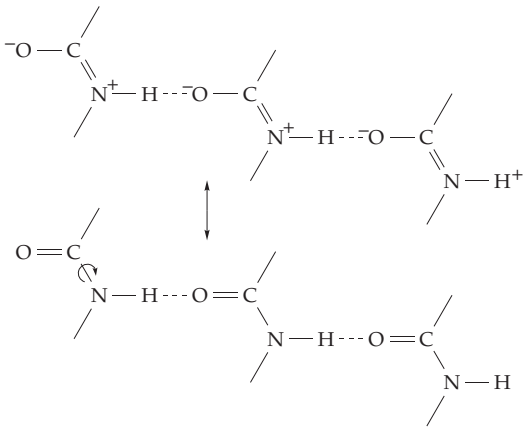

An important effect of the resonance of the amide linkage is that the oxygen atom acquires some negative charge and the NH group some positive charge. Some of the positive charge is usually depicted as residing on the nitrogen, but some is found on the hydrogen atom. The latter can be pictured as arising from a contribution of a fourth resonance form that contains no bond to hydrogen.




Fourth resonance form

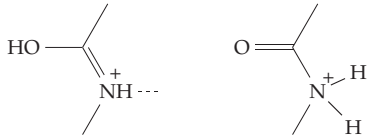
Nevertheless, this picture is inadequate. Various evidence indicates that the nitrogen actually carries a net negative charge.[81a]

The positive and negative ends of the dipoles in the amide group tend to associate to form strong hydrogen bonds. These hydrogen bonds together with the connecting amide linkages can form chains that may run for considerable distances through proteins. The tendency for cooperativity in hydrogen bond formation may impart unusual stability to these chains. As with individual amide linkages, these chains of hydrogen-bonded amides can also be thought of as resonance hybrids:



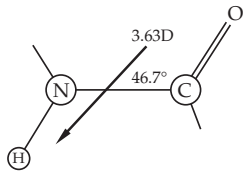
The two structures pictured are extreme forms, the true structure being something in between. In the lower form, rotation about the C–N bond would be permitted but then the charge separation present in the upper structure would no longer exist. Thus, the hydrogen bonds would be weakened. We can conclude that if an amide linkage in such a chain becomes twisted, the hydrogen bonds that it forms will be weakened. If there is cooperativity, the hydrogen bonds will all be strongest when there is good planarity in all of the amides in the chain.

Amides have very weak basic properties and protonation is possible either on the oxygen (A) or on the nitrogen (B).[82,83]



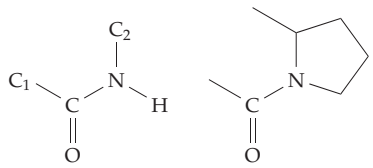

**A** **B**

The p$K_a$ values for such protonation are usually less than zero, but it is possible that a correctly placed acidic group in a protein could protonate either oxygen or nitrogen transiently during the action of a protein. Protonation on oxygen would strengthen hydrogen bonds from the nitrogen whereas protonation on nitrogen would weaken hydrogen bonds to oxygen and might permit rotation. The amide group has a permanent dipole moment of 3.63 debyes oriented as follows:




Here the arrow points toward the *positive* end of the dipole.

While the *trans* peptide linkage shown in Fig. 2-4 is usual, the following *cis* peptide linkage, which is ~8 kJ/mol less stable than the *trans* linkage, also occurs in proteins quite often. The nitrogen atom is usually but not always from proline.[81,84]

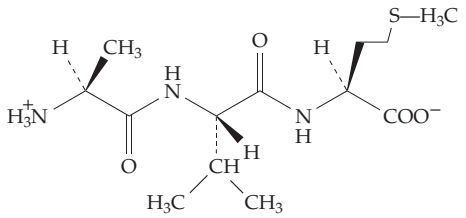



*cis* Peptide linkages

## 4. Polypeptides

The chain formed by polymerization of amino acid molecules provides the **primary structure** of a protein. Together with any covalent crosslinkages and other modifications, this may also be called the

**covalent structure** of the protein. Each monomer unit in the chain is known as an amino acid **residue**. This term acknowledges the fact that each amino acid has lost one molecule of $H_2O$ during polymerization. To be more precise, the number of water molecules lost is *one less* than the number of residues. Peptides are named according to the amino acid residues present and beginning with the one bearing the terminal amino group. Thus, L-alanyl-L-valyl-L-methionine has the following structure:




Like amino acids, this tripeptide is a dipolar ion. The same structure can be abbreviated Ala-Val-Met or, using one-letter abbreviations, AVM. It is customary in describing amino acid sequences to place the amino-terminal (N-terminal) residue at the left end and the carboxyl-terminal (C-terminal) residue at the right end. Residues are numbered sequentially with the N-terminal residue as 1. An example is shown in Fig. 2-6.

The sequence of amino acid units in a protein is always specified by a gene. The sequence determines how the polypeptide chain folds and how the folded protein functions. For this reason much effort has gone into "sequencing," the determination of the precise order of amino acid residues in a protein. Sequences of several hundreds of thousands of proteins and smaller peptides have been established and the number doubles each year.[75,87,88,88a] Most of these

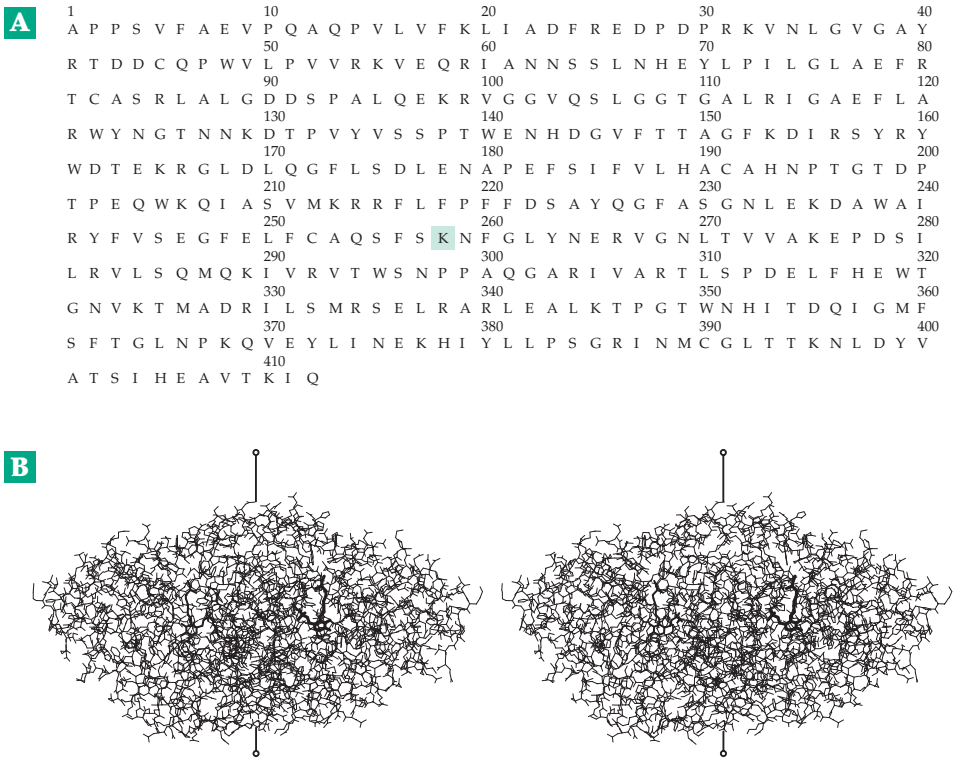



**Figure 2-6** (A) The complete amino acid sequence of the cytoplasmic enzyme aspartate aminotransferase from pig heart. The peptide has the composition $Lys_{19}$, $His_8$, $Arg_{26}$, $(Asp + Asn)_{42}$, $Ser_{26}$, $Thr_{26}$, $(Glu + Gln)_{41}$, $Pro_{24}$, $Gly_{28}$, $Ala_{32}$, $Cys_5$, $Val_{29}$, $Met_6$, $Ile_{19}$, $Leu_{38}$, $Tyr_{12}$, $Phe_{23}$, $Trp_9$. The molecular mass is 46.344 kDa and the complete enzyme is a 93.147-kDa dimer containing two molecules of the bound coenzyme pyridoxal phosphate attached to lysine-258 (enclosed in box).[85,86] (B) A stereoscopic view of a complete enzyme molecule which contains two identical subunits with the foregoing sequence. Coordinates from Arthur Arnone. In this "wire model" all the positions of all of the nearly 7000 atoms that are heavier than hydrogen are shown. The > 8000 hydrogen atoms have been omitted. The view is into the active site of the subunit on the right. The pyridoxal phosphate and the lysine residue to which it is attached are shown with heavy lines. The active site of the subunit to the left opens to the back side as viewed here. The drawing may be observed best with a magnifying viewer available from Abrams Instrument Corp., Lansing, Michigan or Luminos Photo Corp., Yonkers, New York. However, with a little practice, it is possible to obtain a stereoscopic view unaided. Hold the book with good illumination about 20–30 cm from your eyes. Allow your eyes to relax as if viewing a distant object. Of the four images that are visible, the two in the center can be fused to form the stereoscopic picture. Drawings by program MolScript (Kraulis, 1991).

## BOX 2-A PROTEINS OF BLOOD PLASMA

Among the most studied of all proteins are those present in blood plasma.[a–f] Their ready availability and the clinical significance of their study led to the early development of electrophoretic separations. Electrophoresis at a pH of 8.6 (in barbital buffer) indicated six main components. The major and one of the fastest moving proteins is **serum albumin**. Trailing behind it are the **$\alpha_1$-**, **$\alpha_2$-**, and **β-globulins**, **fibrinogen**, and **γ-globulins**. Each of these bands consists of several proteins and two-dimensional separation by electrophoresis and isoelectric focusing (Chapter 3) reveals over 30 different proteins.[e] Many of these contain varying numbers of attached carbohydrate units and appear as families of spots.

Fractionation of large quantities of plasma together with immunochemical assays has led to identification of over 200 different proteins. Sixty or more are enzymes, some in very small quantitites which may have leaked from body cells. Normally plasma contains 5.7 – 8.0 g of total protein per 100 ml (~1 mM). Albumin accounts for 3.5 – 4.5 g/100 ml. An individual's liver synthesizes about 12 g each day. Next most abundant are the **immunoglobulins**. One of these (IgG or γ-globulin) is present to the extent of 1.2–1.8 g/100 ml. Also present in amounts greater than 200 mg per 100 ml are **α-** and **β-lipoproteins**, the **$\alpha_1$ antitrypsin**, **$\alpha_2$-macroglobulin**, **haptoglobin**, **transferrin**, and fibrinogen.

Plasma proteins have many functions. One of them, fullfilled principally by serum albumin, is to impart enough osmotic pressure to plasma to match that of the cytoplasm of cells. The heart-shaped human serum albumin consists of a single 65 kDa chain of 585 amino acid residues coiled into 28 helices.[g] Three homologous repeat units or domains each contain six disulfide bridges, suggesting that gene duplication occurred twice during the evolution of serum albumins. The relatively low molecular mass and high density of negative charges on the surface make serum albumin well adapted for the role of maintaining osmotic pressure. However, serum albumin is not essential to life.

Human serum albumin. From He and Carter. [g]

Over 50 mutant forms have been found and at least 30 persons have been found with no serum albumin in their blood.[h,i] These analbuminemic individuals are healthy and have increased concentrations of other plasma proteins.

A second major function of plasma proteins is transport. Serum albumin binds to and carries many sparingly soluble metabolic products, including fatty acids, tryptophan, cysteine, steroids, thyroid hormones, $Ca^{2+}$, $Cu^{2+}$, $Zn^{2+}$, other metal ions, bilirubin, and various drugs. There are also many more specialized transporter proteins. **Transferrin** carries iron and **ceruloplasmin** (an $\alpha_2$ globulin) transports copper. **Transcortin** carries corticosteroids and progesterone, while another protein carries sex hormones. **Retinol-binding protein** carries vitamin A and **cobalamin-binding proteins** vitamin $B_{12}$. **Hemopexin** carries heme to the liver, where the iron can be recovered.[j] Haptoglobin binds hemoglobin released from broken red cells and also assists in the recycling of the iron in the heme.[k] **Lipoproteins** (see Table 21-1) carry phospholipids, neutral lipids, and cholesterol esters. Most of the mass of these substances is lipid.

Immunoglobulins, $\alpha_1$-trypsin inhibitor and **$\alpha_2$-macroglobulin**,[k] ten or more blood clotting factors; and proteins of the **complement system** all have protective functions that are discussed elsewhere in this book. Hormones, many of them proteins, are present in the blood as they are carried to their target tissues. Many serum proteins have unknown or poorly understood functions. Among these are the **acute phase proteins**, whose concentrations rise in response to inflammation or other injury.

---


[a] Allison, A. C., ed. (1974) *Structure and Function of Plasma Proteins*, Vol. 1, Plenum, New York

[b] Allison, A. C., ed. (1976) *Structure and Function of Plasma Proteins*, Vol. 2, Plenum, New York

[c] Putnam, F. W., ed. (1975) *The Plasma Proteins*, 2nd ed., Vol. 1 and 2, Academic Press, New York

[d] Blomback, B., and Hanson, L. A., eds. (1979) *Plasma Proteins*, John Wiley, Chichester, Oklahoma

[e] Geisow, M. J., and Gordon, A. H. (1978) *Trends Biochem. Sci.* **3**, 169–171

[f] Smith, E. L., Hill, R. L., Lehman, I. R., Lefkowitz, R. J., Handler, P., and White, A. (1983) in *Principles of Biochemistry, Mammalian Biochemistry*, 7th ed., McGraw-Hill, New York, pp. 3–37

[g] He, X. M., and Carter, D. C. (1992) *Nature* **358**, 209–215

[h] Peters, T., Jr. (1996) *All About Albumin: Biochemistry, Genetics and Medical Applications*, Academic Press, San Diego, California

[i] Madison, J., Galliano, M., Watkins, S., Minchiotti, L., Porta, F., Rossi, A., and Putnam, F. W. (1994) *Proc. Natl. Acad. Sci: USA* **91**, 6476–6480

[j] Satoh, T., Satoh, H., Iwahara, S.-I., Hrkal, Z., Peyton, D. H., and Muller-Eberhard, U. (1994) *Proc. Natl. Acad. Sci: U.S.A.* **91**, 8423–8427

[k] Feldman, S. R., Gorias, S. L., and Pizzo, S. V. (1985) *Proc. Natl. Acad. Sci: U.S.A.* **82**, 5700–5704

have been deduced from the sequences of nucleotides in DNA. Sequences of some small peptide hormones and antibiotics are shown in Fig. 2-4 and that of a 412-residue protein in Fig. 2-6. The molecular mass of a protein can be estimated from the chain length by assuming that each residue adds 100–115 Da.

The amino acid composition varies greatly among proteins. A typical globular protein contains all or most of the 20 amino acids. The majority are often present in roughly similar amounts but His, Cys, Met, Tyr, and Trp tend to be less abundant than the others. Specialized proteins sometimes have unusual amino acid compositions. For example, collagen of connective tissue contains 33 mole% glycine and 21% of proline + hydroxyproline residues; the major proteins of saliva contain 22% of glutamate + glutamine and 20–45% proline.[89] Cell walls of plants contain both high proline and high glycine polypeptides. One from petunias is 67% glycine.[90] Silk fibroin contains 45% glycine and 29% alanine. A DNA repair protein of yeast has 13 consecutive aspartate residues.[91] The tough eggshell (chorion) of the domesticated silkmoth *Bombyx mori* contains proteins with ~30% cysteine.[92] Many proteins consist, in part, of repeated short sequences. For example, the malaria-causing *Plasmodium falciparum* in its sporozoite stage is coated with a protein that contains 37 repeats of the sequence NANP interspersed with 4 repeats of NDVP.[93] These two sequences have been indicated with single-letter abbreviations for the amino acids.

With a large number of protein and DNA sequences available, it has become worthwhile to compare sequences of the same protein in different species or of different proteins within the same or different species. Computer programs make it possible to recognize **similarities** and **homologies** between sequences even when deletions and insertions have occurred.[88,88a,94–97] The term homology has the precise biological definition "*having a common evolutionary origin*," but it is often used to describe any close similarity in sequence.[98] Among a pair of homologous proteins, a change at a given point in a sequence may be either **conservative**, meaning that a residue of similar character (large, small, positively charged, nonpolar, etc.) has been substituted, or it may be **nonconservative**.

## D. The Architecture of Folded Proteins

All proteins are made in the same way but as the growing peptide chains peel off from the ribosome, each of the thousands of different proteins in a living cell folds into its own special **tertiary structure**.[88a] The number of possible conformations of a protein chain is enormous. Consider a 300-residue polypeptide which could stretch in fully extended form for ~100 nm. If the chain were folded back on itself about 13 times it could form a 7-nm square sheet about 0.5 nm thick. The same polypeptide could form a thin helical rod 45 nm long and ~1.1 nm thick. If it had the right amino acid sequence it could be joined by two other similar chains to form a collagen-type triple helix of 87 nm length and about 1.5 nm diameter (Fig. 2-7).

The highly folded **globular proteins** vary considerably in the tightness of packing and the amount of internal water of hydration.[99,100] However, a density of ~1.4 g $cm^{-3}$ is typical. With an average mass per residue of 115 Da our 300-residue polypeptide would have a mass of 34.5 kDa or $5.74 \times 10^{-20}$ g and a volume of 41 $nm^3$. This might be approximated by a cube 3.45 nm in width, a "brick" of dimensions 1.8 x 3.6 x 6.3 nm, or a sphere of diameter 4.3 nm. Although protein molecules are usually very irregular in shape,[101] for purposes of calculation idealized ellipsoid and rod shapes are often assumed (Fig. 2-7).

It is informative to compare these dimensions with those of the smallest structures visible in cells; for example, a bacterial flagellum is ~13 nm in diameter and a cell membrane ~8–10 nm in thickness. Bricks of the size of the 300-residue polypeptide could be used to assemble a bacterial flagellum or a eukaryotic microtubule. Helical polypeptides may extend through cell membranes and project on both sides, while a globular protein of the same chain length may be almost completely embedded in the membrane.

### 1. Conformations of Polypeptide Chains

To understand how a polypeptide chain folds we need to look carefully at the possible conformations of the peptide units. Since each peptide unit is nearly planar, we can think of a polypeptide as a chain of flat units fastened together as in Fig. 2-8. Every peptide unit is connected to the next by the α-carbon of an amino acid. This carbon provides two single bonds to the chain and rotation can occur about both of them (except in the cyclic amino acid proline). To specify the conformation of an amino acid unit in a polypeptide chain, we must describe the torsion angles about both of these single bonds.[11,76,102] These angles are indicated by the symbols **ϕ (phi)** and **ψ (psi)** and are assigned the value 180° for the fully extended chain as shown in Fig. 2-8. Each angle is taken as zero for the impossible conformation in which the two chain ends are in the eclipsed conformation. By the same token, the torsion angle **ω (omega)** around the C–N bond of the amide is 0° for a planar *cis* peptide linkage and 180° for the usual *trans* linkage.

Since both ϕ and ψ can vary for each residue in a protein, there are a large number of possible conformations. However, many are excluded because they bring certain atoms into collision. This fact can be established readily by study of molecular models.

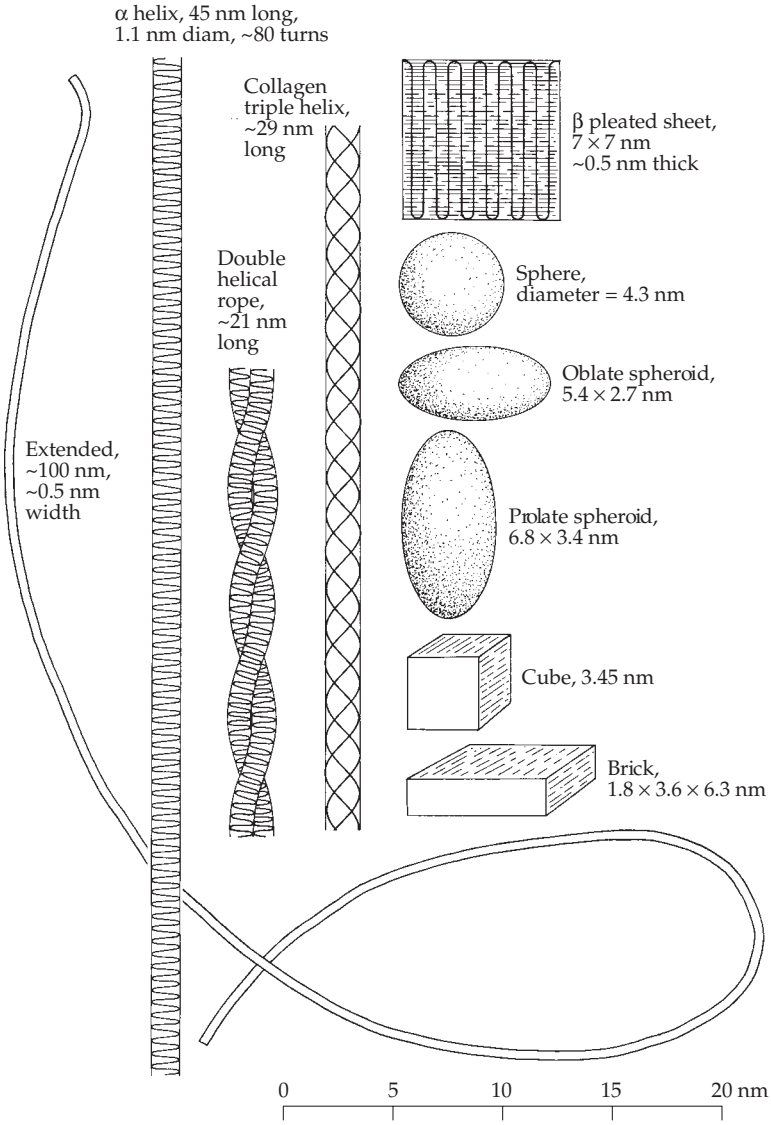



**Figure 2-7** Some idealized shapes that a 34.5 kDa protein molecule of 300 amino acids might assume.

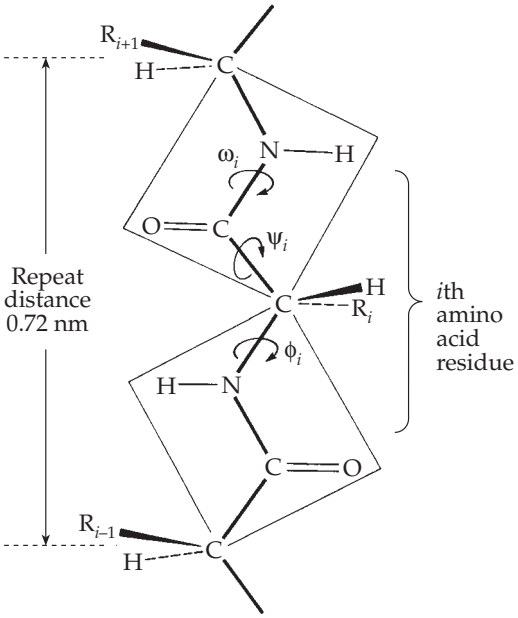



**Figure 2-8** Two peptide units in the completely extended β conformation. The torsion angles $\phi_i$, $\psi_i$, and $\omega_i$ are defined as 0° when the main chain atoms assume the *cis* or eclipsed conformation. The angles in the completely extended chain are all 180°. The distance from one α carbon atom ($C_\alpha$) to the next in a peptide chain is always 0.38 nm, no matter how the chain is folded.

Using a computer, it is possible to study the whole range of possible combinations of $\phi$ and $\psi$. This has been done for the peptide linkage by Ramachandran. The results are often presented as plots of $\phi$ vs $\psi$ (Ramachandran plots or **conformational maps**)[102a,103,103a] in which possible combinations of the two angles are indicated by blocked out areas. The original Ramachandran plots were made by representing the atoms as hard spheres of appropriate van der Waals radii, but the version shown in Fig. 2-9 was calculated using a complex potential energy function to represent the van der Waals attraction and the repulsion from close contact.[103] This map was calculated for poly-L-alanine but it would be very similar for most amino acids.

Notice the two areas connected by a higher energy bridge on the left side of Fig. 2-9. The upper area contains the pairs of torsion angles for the extended **β structures** as well as for **collagen**. The lower area contains allowed conformations for the *right-handed helices.* As can be seen from Fig. 2-10, most of the observed conformations of peptide units in a real protein fall into these regions. Glycyl residues are an exception. Since glycine has no β-carbon atom, the conformations are less restricted. Out of nearly 1900 non-glycine residues in well-determined protein structures, 66 were found in disallowed areas of the Ramachandran diagram.[104a] These were often accommodated by local distortions in bond angles. The positions at which such steric strain occurs are often in regions concerned with function.[104b] One residue, which lies in a disallowed region in Fig. 2-10 is asparagine 297 of aspartate aminotransferase (Fig. 2-6). It is located

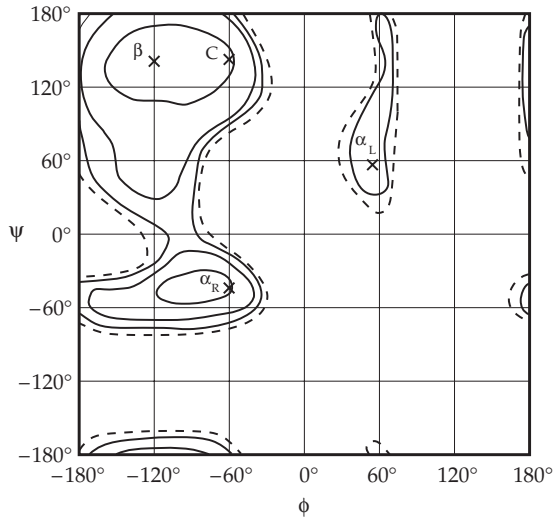



**Figure 2-9** Potential energy distribution in the ϕ–ψ plane for a pair of peptide units with alanyl residues calculated using potential parameters of Scheraga and Flory. Contours are drawn at intervals of 1 kcal (4.184 kJ) per mol going down from 0 kcal per mol. The zero contour is dashed. From Ramachandran *et al*.[104] The points marked x are for the four ideal structures: twisted β structure (β), collagen (C), right-handed α helix ($\alpha_R$), and the less favored left-handed α helix ($\alpha_L$).

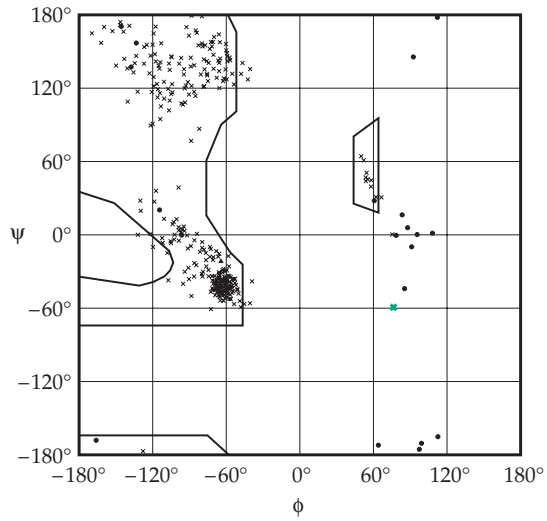



**Figure 2-10** Ramachandran plot for cytosolic aspartate aminotransferase. The angles ϕ and ψ were determined experimentally from X-ray diffraction data at 0.16 nm resolution and model building. The majority of conformations are those of α helices or of β structure. Glycine residues are indicated by filled circles, while all other residues are denoted by an "×". One of these (green) lies quite far from an allowed area and must give rise to localized strain.[104b] Extreme lower limit "allowed" regions by the hard-sphere criteria are shown in outline. From coordinates of Arthur Arnone *et al*. (unpublished).[106]

adjacent to the coenzyme site. The possible conformations of proline residues are limited. The angle ϕ is always −60 ± 20°, while ψ for the residue adjacent to the proline N can be either ~150° or ~ −30°.[105] Typical ϕ, ψ angles for some regular peptide structures are given in Table 2-3.

Conformations of side chain groups are designated by a series of torsion angles designated $\chi_i$.[11] Within proteins there are strong preferences for certain $\chi_i$,$\chi_2$ pairs.[107] Torsion angles within proline rings are considered in Chapter 5,A,6.

## 2. The Extended Chain β Structures

As was first pointed out by Pauling and Corey,[108,109] an important structural principle is that within proteins the maximum possible number of hydrogen bonds involving the C=O and N–H groups of the peptide chain should be formed. One simple way to do this is to line up fully extended chains (ϕ = ψ = 180°) and to form hydrogen bonds between them. Such a structure

**TABLE 2-3**
**Approximate Torsion Angles for Some Regular Peptide Structures[a]**

| Structure | | | ϕ (deg.) | ψ (deg.) |
|---|---|---|---|---|
| Hypothetical fully extended polyglycine chain[b] | | | − 180 | +180 |
| β-Poly(L-alanine) in antiparallel-chain pleated sheet | | | − 139 | +135 |
| Parallel-chain pleated sheet | | | − 119 | +113 |
| Twisted β strand | | | − 120 | +140 |
| Polyglycine II | | | − 80 | +150 |
| Poly(L-proline) II | | | − 78 | +149 |
| Collagen[c] | | | − 60 ± 15 | − 140 ± 15 |
| Right-handed α helix[d] | | | − 57 | − 47 |
| Left-handed α helix | | | + 57 | + 47 |
| β Bends: | Type I, | residue 2 | − 60 | − 30 |
| | | residue 3 | − 90 | 0 |
| | Type II, | residue 2 | − 60 | 120 |
| | | residue 3 | 80 | 0 |
| | Type III, | residue 2 & 3 | − 60 | − 30 |
| β Bulges[e]: | "Classical" β bulge | residue 1 | − 100 | − 45 |
| | G1 bulge | residue 1 | 85 | 0 |

[a] From Liébecq, C., ed. (1992) *Biochemical Nomenclature*, Portland Press, London and Chapel Hill (for the International Union of Biochemistry and Molecular Biology).
[b] Torsion angles for the fully extended chain can be designated either +180° or −180°, the two being equivalent. They are given as ϕ = −180° and ψ = +180° to facilitate comparison with the other structures.
[c] Ramachandran, G. N. (1967) *Treatise on Collagen*, Vol. 1, p. 124, Academic Press, New York
[d] Both ϕ and ψ are quite variable but ϕ + ψ = −104°.
[e] For residues other than these indicated the torsion angles are about those of typical β structures.

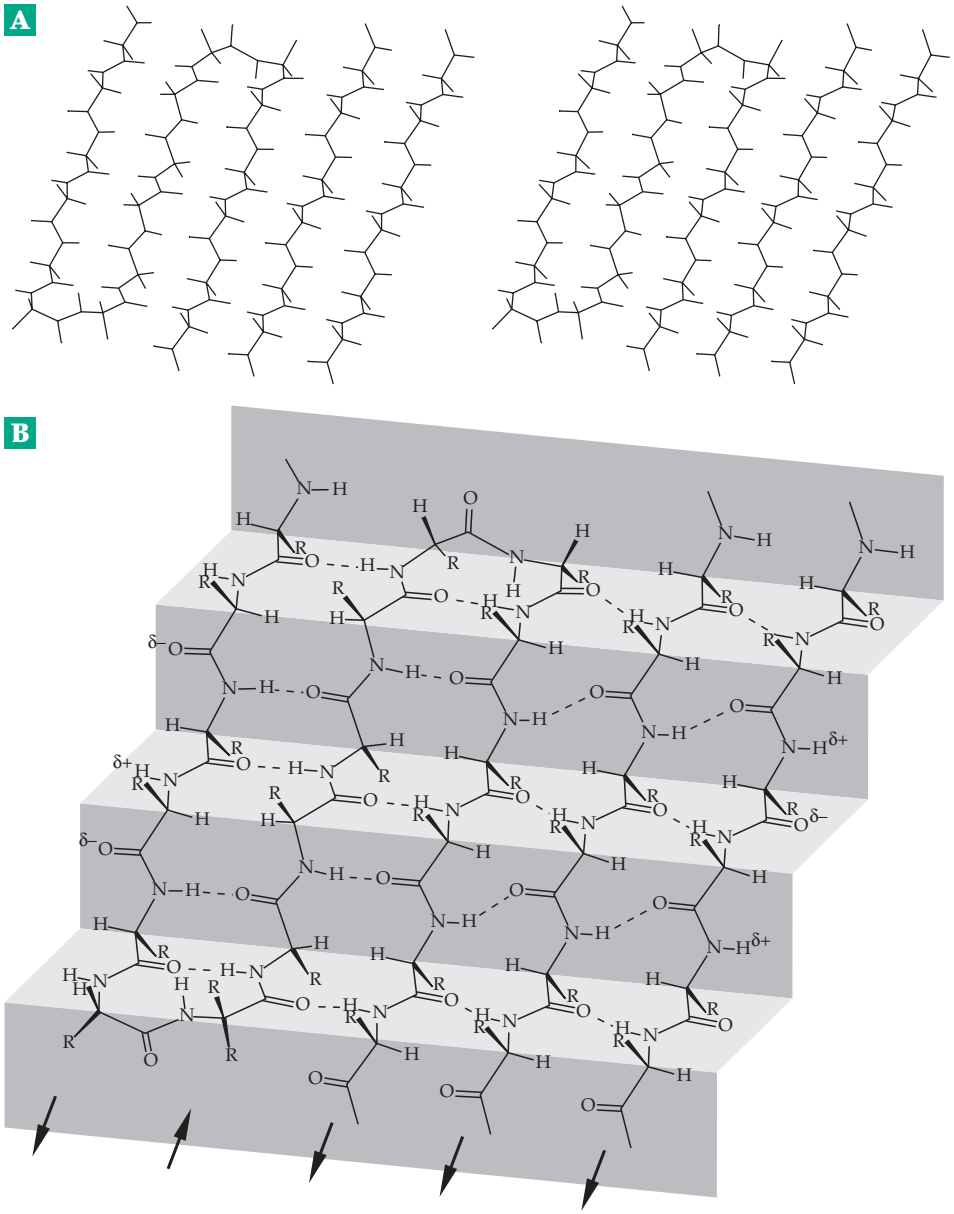



**Figure 2-11** The extended chain β pleated sheet structures. (A) Stereoscopic drawing without atomic symbols. (B) Drawing with atomic symbols. At the left is the *antiparallel* structure. The 0.70 nm spacing is slightly decreased from the fully extended length. The amino acid side chains (R) extend alternately above and below the plane of the "accordion pleated" sheet. The pairs of linear hydrogen bonds between the chains impart great strength to the structure. The chain can fold back on itself using a "β turn" perpendicular to the plane of the pleated sheet. The *parallel* chain structure (right side) is similar but with a less favorable hydrogen bonding arrangement. Arrows indicate chain directions.

exists for polyglycine and resembles that in Fig. 2-11. Notice that on the left side of this figure, the adjacent chains run in opposite directions; hence, the term **antiparallel β structure**. The antiparallel arrangement not only gives the best hydrogen bond formation between chains but also permits a single chain to fold back on itself giving a compact hairpin loop.

***Pleated sheets.*** While a fully extended polyglycine chain is possible, the side chains of other amino acids cannot be accommodated without some distortion of the structure. Thus, the peptide chains in silk fibroin have a repeat distance of 0.70 nm compared with the 0.72 nm for the fully extended chain (Fig. 2-8). Pauling and Corey[108] showed that this shortening of the chain could result from rotation of angle $\phi$ by ~40° (to –140°) and rotation of $\psi$ in the opposite direction by ~45° (to +135°) to give a slightly puckered chain. The resulting multichain structure (shown in Fig. 2-11) is known as a **pleated sheet**. As in this figure, both

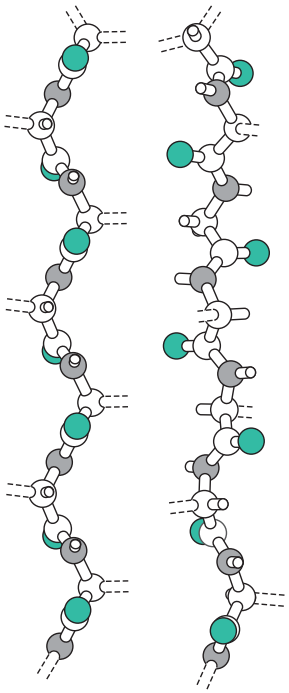


**Figure 2-12** Straight (left) and twisted (right) peptide chains in extended β conformations. From Chothia.[110]

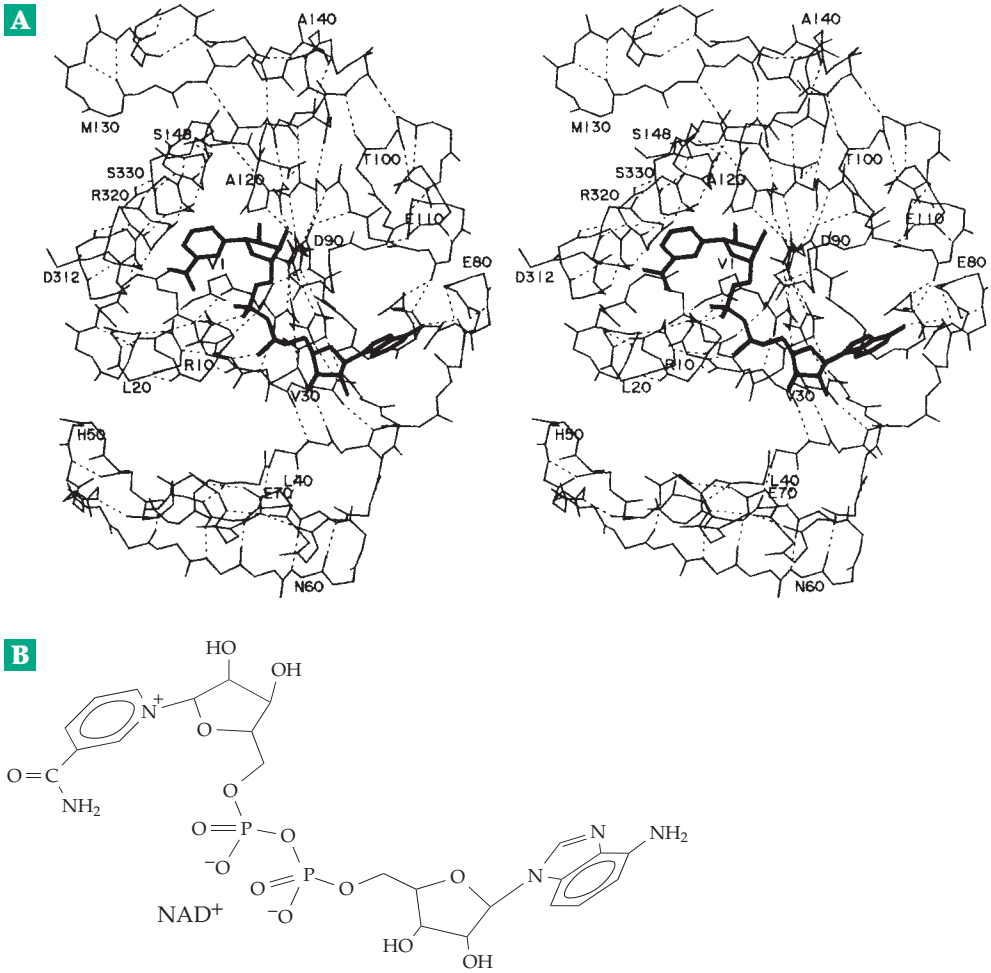



**Figure 2-13** (A) Stereoscopic view of the nucleotide binding domain of glyceraldehyde phosphate dehydrogenase. The enzyme is from *Bacillus stearothermophilus* but is homologous to the enzyme from animal sources. Residues are numbered 0–148. In this wire model all of the main chain C, O, and N atoms are shown but side chains have been omitted. The large central twisted β sheet, with strands roughly perpendicular to the page, is seen clearly; hydrogen bonds are indicated by dashed lines. Helices are visible on both sides of the sheet. The coenzyme $NAD^+$ is bound at the end of the β sheet toward the viewer. Note that the two phosphate groups in the center of the $NAD^+$ are H-bonded to the N terminus of the helix beginning with R10. From Skarzynski *et al.*[111a] (B) Structural formula for $NAD^+$.

parallel and antiparallel strands are often present in a single β sheet within a protein.

The β structure is one of the most important **secondary structures** in proteins. It occurs in about 80% of the soluble globular proteins whose structures have been determined. In many cases almost the entire protein is made up of β structure. Single strands of extended polypeptide chain are sometimes present within globular proteins but more often a chain folds back on itself to form a hairpin loop. A second fold may be added to form an antiparallel "β meander"[102] and additional folds to form β sheets. Beta structures are found in silk fibers (Box 2-B) as well as in soluble proteins.

***Twisted sheets.*** X-ray diffraction studies have shown that β pleated sheets are usually not flat but are twisted. In a twisted sheet the individual polypeptide chains make a shallow left-handed helix. However, when successive carbonyl groups are viewed along the direction of the chain, a **right-handed twist** is seen (Fig. 2-12).[110] Such twisted β sheets are often found in the globular proteins. An example (Fig. 2-13) is the "nucleotide-binding" domain of a dehydrogenase enzyme. The twist of the sheet is seen clearly in this stereoscopic view. When such chains are associated into β sheets, whether parallel or antiparallel, and are viewed in a direction perpendicular to the chains and looking down the edge of the sheet, a left-handed "propeller" is seen. Such a propeller is visible in the

drawing of carboxypeptidase A (Fig. 2-14).

The cause of the twist in β sheets appears to lie in noncovalent interactions between hydrogen atoms on the β-carbon atoms of side chains and the peptide backbone atoms. For side chains of most L- amino acids these interactions provide a small tendency towards the observed right-handed twist.[111] Nonplanarity in the amide groups (Fig. 2-5) may also contribute. Interstrand interactions seem to be important.[111b]

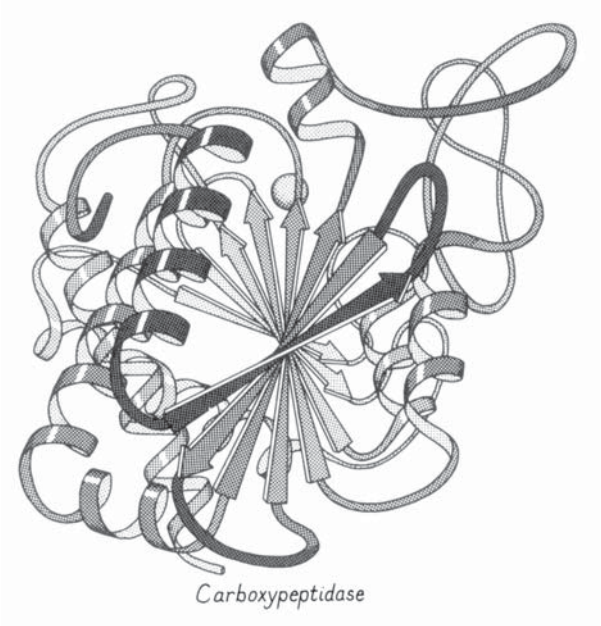



**Figure 2-14** A "ribbon" drawing of the 307- residue protein-hydrolyzing enzyme carboxypeptidase A. In this type of drawing wide ribbons are used to show β strands and helical turns while narrower ribbons are used for bends and loops of the peptide chains. The direction from the N terminus to C terminus is indicated by the arrowheads on the β strands. No individual atoms are shown and side chains are omitted. Courtesy of Jane Richardson.[117]

***Properties of β sheets.*** In antiparallel β sheets, nonpolar residues are often present on one side of the sheet and polar residues on the other. The nonpolar side may be buried in the protein, perhaps backed up against another β sheet of similar structure as in the carbohydrate-binding **lectin** shown in Fig 2-15 to give a **β sandwich**. To accommodate this packing arrangement, nonpolar and polar residues tend to alternate in the amino acid sequence. The facing nonpolar surfaces of the two sheets are essentially smooth. The β strands in one sheet lie at an angle of ~30° to those in the other sheet. The twist of the strands allows pairs of nonpolar side chains from the two sheets to maintain good van der Waals contact with each other for a considerable distance along the strands. Beta sheets of silks (Box 2-B) and of immunoglobulin domains (Fig 2-16B) are also thought to be associated in a back-to-back fashion. *Parallel* β structures have been found only when there are five or more strands, some of which may be antiparallel to the others. The parallel β structure is apparently less stable than the antiparallel structure. Parallel strands are usually buried in the protein, being surrounded by either other extended strands or helices.

Chains of hydrogen bonds and amide linkages pass across β sheets perpendicular to the chain direction. There are partial positive and negative charges along the outside edges of the sheet where these hydrogen bond chains terminate (Fig. 2-11). The polarity of these chains alternates and the positive end of one peptide bond is relatively near to the negative end of the next along the edges of the sheets. These "unsatisfied ends" of hydrogen bond chains are often sites of

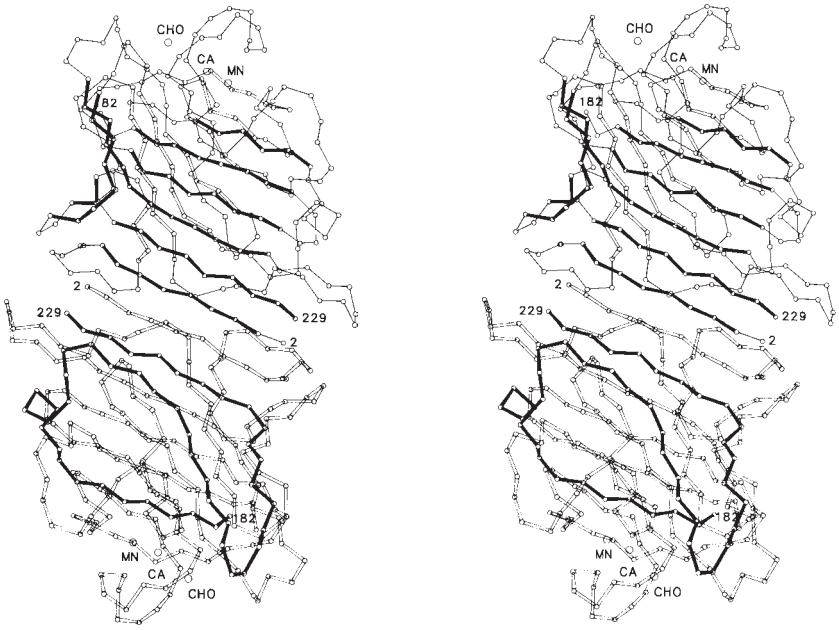
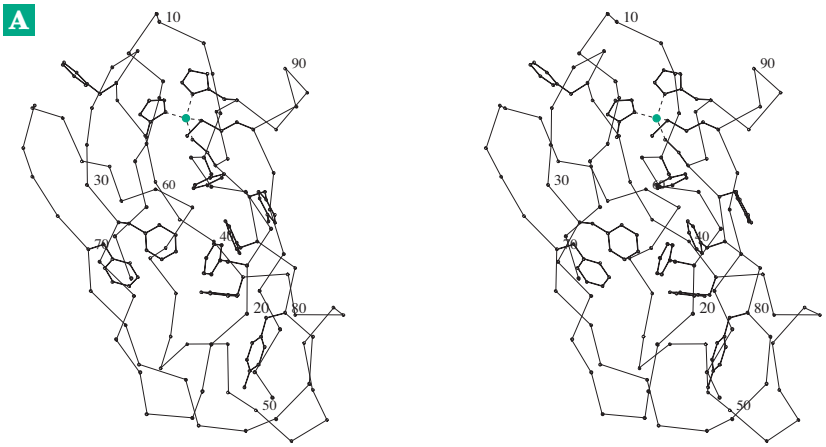
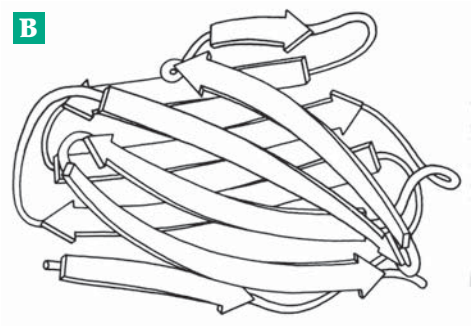



**Figure 2-15** A stereoscopic **alpha-carbon plot** showing the three-dimensional structure of favin, a sugar-binding lectin from the broad bean (*Vicia faba*). In this plot only the α-carbon atoms are shown at the vertices. The planar peptide units are represented as straight line segments. Side chains are not shown. The protein consists of two identical subunits, each composed of a 20-kDa α chain and a 20-kDa β chain. The view is down the twofold rotational axis of the molecule. In the upper subunit the residues involved in the front β sheet are connected by double lines, while those in the back sheet are connected by heavy solid lines. In the lower subunit the α chain is emphasized. Notice how the back β *sheet* (not the chain) is continuous between the two subunits. Sites for bound $Mn^{2+}$ (MN), $Ca^{2+}$ (CA), and sugar (CHO) are marked by larger circles. From Reeke and Becker.[112]



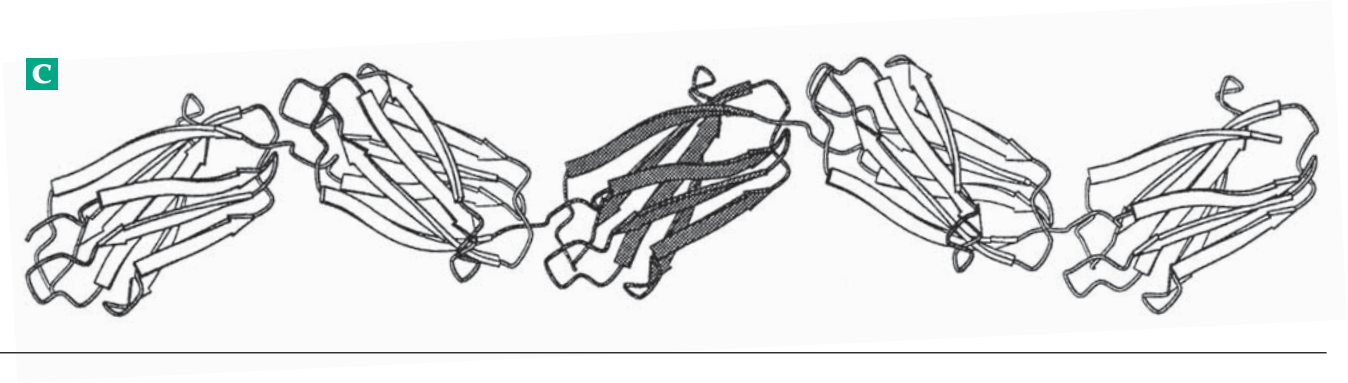




**Figure 2-16** Beta cylinders. (A) Stereoscopic α-carbon plot of plastocyanin, a copper-containing electron-transferring protein of chloroplasts. The copper atom at the top is also shown coordinated by the nitrogen atoms of two histidine side chains. The side chains of the aromatic residues phenylalanines 19, 29, 35, 41, 70, and 82 and tyrosines 80 and 83 are also shown. Most of these form an internal cluster. From Guss and Freeman.[116] (B) Ribbon drawing of immunoglobulin fold. This is a common structure in domains of the immunoglobulins and in many other extracellular proteins. Two layers of antiparallel β sheet are stacked face to face to form a flattened barrel. One disulfide bridge is always present and is represented as a thick rod. From J. Richardson.[117] (C) Five tandem fibronectin type III domains. These domains, which are found in the muscle protein titin as well as in fibronectin, resemble immunoglobulin domains but lack disulfide bridges. From Erickson.[117a] Figure courtesy of Harold P. Erickson.

interaction of polar groups from side chains. Thus, OH groups of serine or threonine residues, amide groups of asparagine and glutamine residues, etc., often fold back and hydrogen-bond to these ends.

Edges of β sheets can also serve as binding sites for other polar molecules. For example, substrates bind to an edge of a β sheet in the active sites of trypsin and other proteases (Chapter 12). Some proteins, e.g. the lectin shown in Fig. 2-15, form dimers by joining identical edges of a β sheet in antiparallel orientation.[112]

***Cylinders and barrels.*** The twisted β sheets of proteins are often curved to form structures known as **β cylinders** or **β barrels** (Fig. 2-16).[113,114] Simple cylinders formed by parallel β strands form the backbones of the electron transport protein plastocyanin, the enzyme superoxide dismutase, the oxygen carrier hemerythin (Fig. 16-20), transporter proteins that carry hydrophobic ligands,[115] and the immunoglobulins in which each domain contains a β barrel.

The eight-stranded β cylinder of plastocyanin (Fig. 2-16A) is somewhat flattened and can also be regarded as a β sandwich.[116,118] However, the β barrel of triose phosphate isomerase (see Fig. 2-28) is surrounded by eight α helices which provide additional stability and a high symmetry. Bacterial outer membranes contain pores created by very large β cylinders within proteins called **porin**s.[119,120] The one shown in Fig. 8-20 has 16 strands.

A single β strand can also be wound into a cylinder with the hydrogen bonds running parallel to the helix axis. A right-handed **parallel β helix** of this type has been found in the bacterial enzyme **pectate lyase**.[121,122] The polypeptide chains of the 353-residue protein contain seven complete turns of about 22

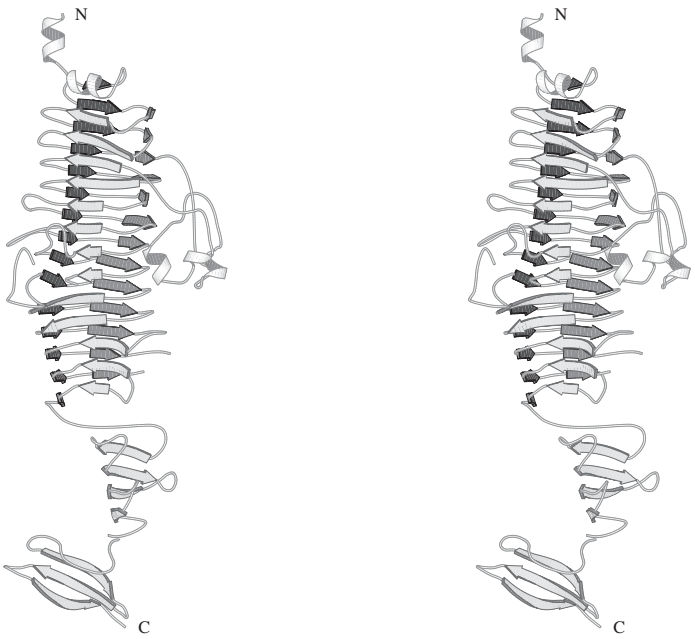



**Figure 2-17** Wire model of the tailspike protein of bacteriophage P22 of *Salmonella*. Three of these fish-shaped molecules associate as a trimer to form the spike. From Steinbacher *et al*.[123]

residues each which form a cylindrical parallel β sheet. The cylinder is folded inward into roughly an L-shape with packed side chains filling the remaining inside space. The tailspikes of bacteriophage P22, a virus of *Salmonella*, are formed from 666-residue protein subunits which contain 13 turns of parallel β helix (Fig. 2-17).[123]

The tendency for a β sheet to fold into a cylinder is encouraged in antiparallel β structures by the existence of a common irregularity called the β **bulge**.[124,125] As illustrated in Fig. 2-18, a β bulge contains an extra residue inserted into one of the chains. In the second type of bulge shown in Fig. 2-18, the extra residue is glycine with torsion angles of about $\phi = 85°$, $\psi = 0°$, which are possible only for glycine. In the two β cylinders of trypsin, chymotrypsin, and elastase (Fig. 12-9), there are seven β bulges.

Beta structures are found in many small peptides. The hormone oxytocin (Fig. 2-4), the antibiotics gramicidin S (Fig. 2-4) and valinomycin (Fig. 8-22), and the mushroom peptide antamanide (Box 28-B) are among these. The cyclic structures of these compounds favor formation of antiparallel β strands with sharp turns at the ends. Polypeptide antibiotics that have alternating

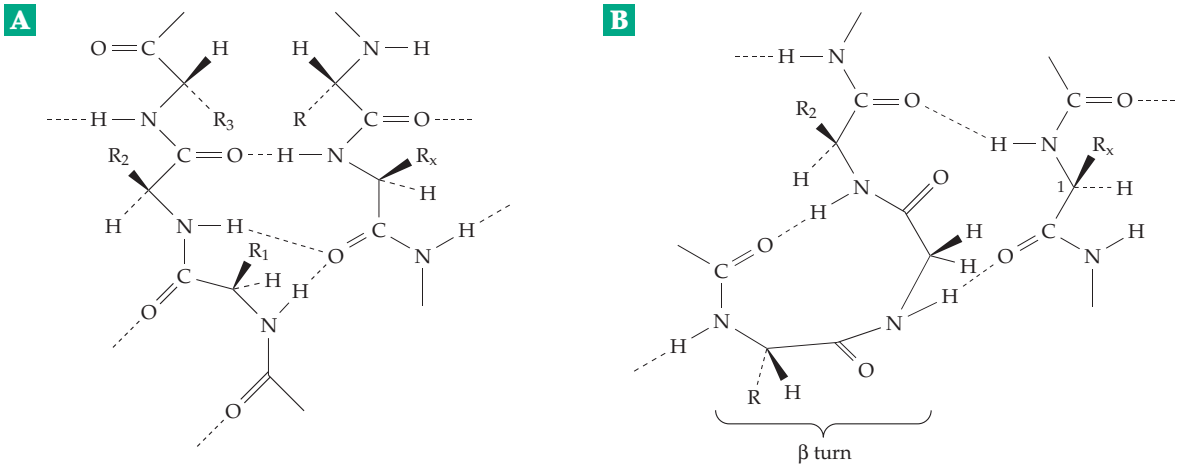



**Figure 2-18** Typical β bulges in antiparallel pleated sheets. The residues $R_1$, $R_2$, and $R_x$ identify the bulges. (A) A "classic" β bulge, in which $\phi_1$ and $\psi_1$ are nearly those of an α helix while other torsion angles are approximately those of regular β structures. (B) The G1 bulge in which the first residue is glycine with $\phi_1 = 85°$, $\psi_1 = 0°$. It is attached to a type II β turn of which the glycine (labeled 1) is the third residue.

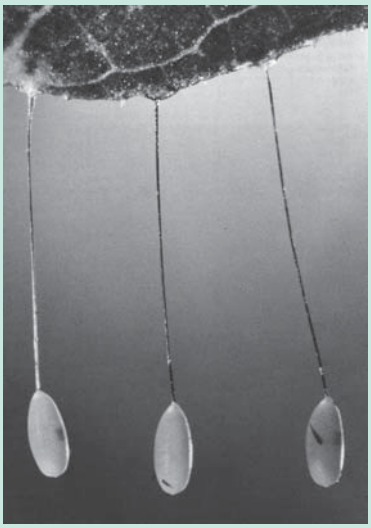
sequences of D and L residues can be coiled into a β helix[126,127] or a pair of polypeptide chains arranged in an antiparallel fashion can form a double stranded β helix with an 0.3 to 0.4-nm hydrophilic pore through the center. These peptide antibiotics appear to exert their antibacterial action by creating pores through cell membranes and allowing ions to pass through without control.

***Beta propellers.*** Another major folding pattern is a circular array of four to eight "blades" that form a propeller-like structure. Each blade is a small, roughly triangular four-stranded antiparallel β sheet (See Figs. 11-7 and 15-23). Sequences that fold into these blades can often, but not always, be recognized as **WD repeats**. These are typically 44- to 60-residue sequences that have the sequence GH (Gly-His) about 11–24 residues from the N terminus and WD (Trp-Asp) at the C terminus.[127a,127b] This repeat sequence encodes the

## BOX 2-B SILKS

The green lacewing fly *Chrysopa flava* lays its eggs on 1 cm silk stalks glued to the undersides of leaves. It has been proposed, as for other silks, that the peptide chains in the stalks are aligned perpendicular to the long dimension of the fiber and are folded back on themselves many times to form a β sheet with only ~8 residues between folds.[a] The chains of **silk fibroin**, the major protein of silkworm silk, contain 50 repeats of the sequence:[b]

GAGAGSGAAG(SGAGAG)$_8$Y



Courtesy of Ralph Buchsbaum

All of the alanine and serine side chains of this sequence presumably protrude on one side of a β sheet while the other side has only the hydrogen atoms of the glycine. This permits an efficient stacking of the sheets with interdigitation of the side chains of the alanine and serine residues of adjacent sheets.

However, DNA sequences of silk genes have revealed a greater complexity. The silk fibroin sequence suggests that at bends between β strands there are often S – S bridges and the crystalline β sheet domains are interspersed with 100–200 residue segments of amorphous protein[b–d] whose coiled chains can be stretched greatly. The protein from *Bombyx mori* consists of 350-kDa heavy chains linked by a disulfide bond to 25-kDa light chains.[e] A silkworm **housing silk** contains polyalanine sequences, e.g. (Ala)$_{10-14}$. These evidently form α-helical regions.[f] Silk molecules are synthesized and stored in glands as a concentrated solution of apparently globular molecules. The silk is extruded through a spinneret, whose diameter is ~10 μm, after which the silkworm stretches the silk, causing it to form a stiff fiber. Apparently the stretching of the folded polypeptide chains permits the β-sheet-forming sequences to find each other and form oriented crystalline regions.[d]

Spiders may produce silk from as many as seven different types of glands. Dragline silk, by which a spider itself may descend, is stronger than steel but because of the coiled amorphous regions may be stretched by 35%.[g–l] The most elastic silks are those of the catching spirals of orb-webs.[i,j,l] They can be stretched 200% and contain a variety of repeated sequences, including GPCC(X)$_n$. The latter is similar to sequences in elastin (Section 4), and is able to form type II β bends, which may have proline in position 2 and must have glycine in position 3 (Fig. 2-24). A chain of repeated β bend motifs can form a flexible spring, a β spiral as proposed for elastin. Genes for spider silk have been cloned and are being used to engineer new proteins with commercial uses, e.g., to help anchor cells in regenerating body tissues.[d,h,j]


[a] Geddes, A. J., Parker, K. D., Atkins, E. D. T., and Beighton, E. (1968) *J. Mol. Biol.* **32**, 343–368
[b] Garel, J.-P. (1982) *Trends Biochem. Sci.* **7**, 105–108
[c] Vollrath, R. (1992) *Sci. Am.* **266**(Mar), 70–76
[d] Calvert, P. (1998) *Nature (London)* **393**, 309–311
[e] Mori, K., Tanaka, K., Kikuchi, Y., Waga, M., Waga, S., and Mizuno, S. (1995) *J. Mol. Biol.* **251**, 217–228
[f] van Beek, J. D., Beaulieu, L., Schäfer, H., Demura, M., Asakura, T., and Meier, B. H. (2000) *Nature (London)* **405**, 1077–1079
[g] Simmons, A. H., Michal, C. A., and Jelinski, L. W. (1996) *Science* **271**, 84–87
[h] Hinman, M. B., and Lewis, R. V. (1992) *J. Biol. Chem.* **267**, 19320–19324
[i] Hayashi, C. Y., and Lewis, R. V. (1998) *J. Mol. Biol.* **275**, 773–784
[j] Guerette, P. A., Ginzinger, D. G., Weber, B. H. F., and Gosline, J. M. (1996) *Science* **272**, 112–115
[k] Spek, E. J., Wu, H.-C., and Kallenbach, N. R. (1997) *J. Am. Chem. Soc.* **119**, 5053–5054
[l] Hayashi, C. Y., and Lewis, R. V. (2000) *Science* **287**, 1477–1479

fourth β strand of one blade followed by the first three strands of the next blade. This overlap snaps the blades together to form the propeller. At least one tight turn is present in each blade. A seven-bladed propeller is present in the β subunits of the regulatory GTP-hydrolyzing **G proteins**, which couple extracellular signals to intracellular enzymes and ion channels (Fig. 11-7). A similar β propeller is present in **clathrin**, which forms cage-like enclosures around endocytic vesicles (Chapter 8).[127c] Six-bladed propellers are predicted to occur in many extracellular proteins.[127d] A β propeller binds the coenzyme PQQ in bacterial dehydrogenases (Fig. 15-23).

## 3. Helices

The **alpha helix** represents the second major structural element of soluble proteins[108,128] and is also found in many fibrous proteins, including those of muscle and hair. In the α helix both ϕ and ψ are about −50 to −60° and except at the N-terminal helix end, *each NH is hydrogen bonded to the fourth C=O further down the chain*. All of the N–H and C=O groups of the peptide linkages lie roughly parallel to the helix axis; the N–H groups point toward the N terminus of the chain and the carbonyl groups toward the C terminus (Fig. 2-19A).

The number of amino acid units per turn of the helix is ~3.6, with five turns of the helix containing 18 residues. The **pitch** (repeat distance) of the helix, which can be determined experimentally from X-ray diffraction data, is 0.54 nm. Polar coordinates for the α helix have been tabulated.[130] With L-amino acids, the right-handed helix, is more stable than the left-handed helix which has so far not been found in proteins. Frequently, however, a few residues have the ϕ, ψ angles of this helix. The ϕ, ψ angles of the α helix are given in Table 2-3 as −57°, −47°, but are much more variable in real helices. In erythrocruorin, for which an accurate structure determination has been made,[131]

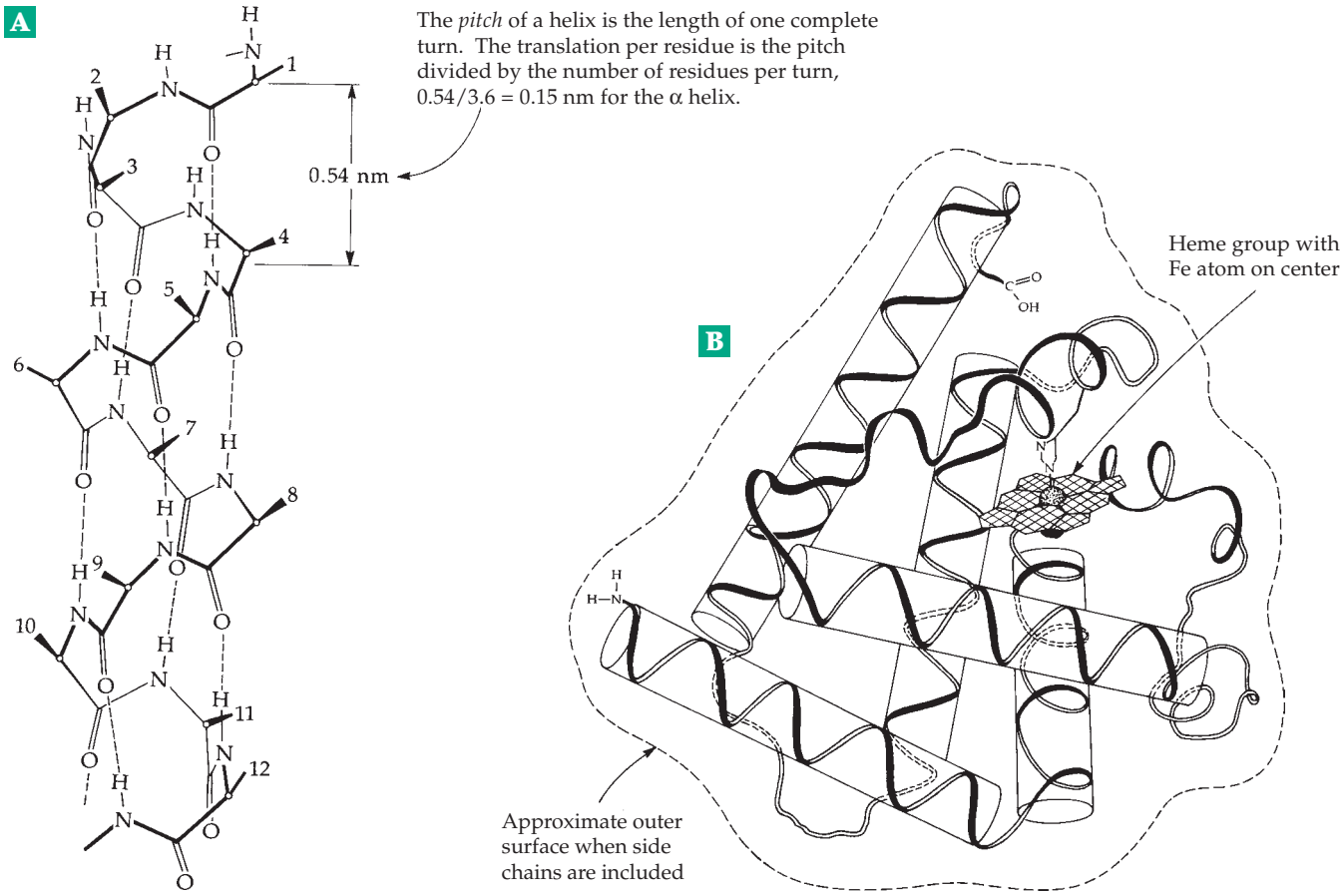



**Figure 2-19** The α helix. (A) The right-handed α helix with vertical hydrogen bonds indicated by dotted lines. The positions of the amino acid side chains are indicated by the numbers. (B) The conformation of the peptide backbone of myoglobin.[129] Five long α helices are indicated as rods. Several shorter helices can also be seen. The overall size of the molecule is approximately 4.4 × 4.4 × 2.5 nm.

the average values are $\phi = -64°$, $\psi = -40°$. More important is the observation that $\phi + \psi = 104°$ within about $\pm 10°$ for most residues in helices of this and other proteins.[132,133] The deviation from the ideal dimensions results in part from the hydrogen bonding of water molecules or polar protein side chains to the carbonyl oxygen atoms in the helix.[131]

Helix formation is spontaneous for peptides as short as 13 residues in water.[134,135] Although the difference in thermodynamic stability between an α helix and an unfolded **"random coil"** conformation is small, poly-L-alanine peptides form helices in water. Glycine destabilizes helices, presumably because of the increased entropy of the unfolded chain which results from the wider range of the conformational angles $\phi$ and $\psi$ for glycyl residues. Proline destabilizes helices even more because its restricted $\phi$ and $\psi$ angles cause the helix to be kinked.[136,137] However, prolyl residues are often present at ends of helices. Other amino acids all fit into helices but may stabilize or destabilize the helix depending upon immediately neighboring groups.[138–140] Bulky side chain groups with a low dielectric constant stabilize helices by strengthening the hydrogen bonds within the helices.[141]

Helices of smaller and larger diameter than that of the α helix are possible. The most important is the **$3_{10}$ helix** (or $3.0_{10}$ helix), which has exactly three residues per turn.[140,142–144] Each NH forms a hydrogen bond to the third C=O on down the chain; thus, the $3_{10}$ helix is tighter than the α helix. Although long $3_{10}$ helices are seldom found in proteins, a single turn of this tighter helix frequently occurs as a "defect" at the end of an α helix. A polymer of α-aminoisobutyric acid forms long $3_{10}$ helices because the α-dimethyl side chains constrain $\phi$ and $\psi$ to appropriate values.[145] Short helical peptide chains in water may exist as a mixture of α and $3_{10}$ forms in equilibrium.[146] The **π helix**, with 4.4 residues per turn is of a larger diameter than the α helix and has only been found in proteins as a single turn, usually at a C terminus.[132,132a]

***Properties of helices.*** The dipoles of the backbone amide linkages of an α helix are all oriented in the same direction. The positive end of each dipole is associated through hydrogen bonding with the negative end of another. This leaves three partial positive charges at the N terminus of the helix and three partial

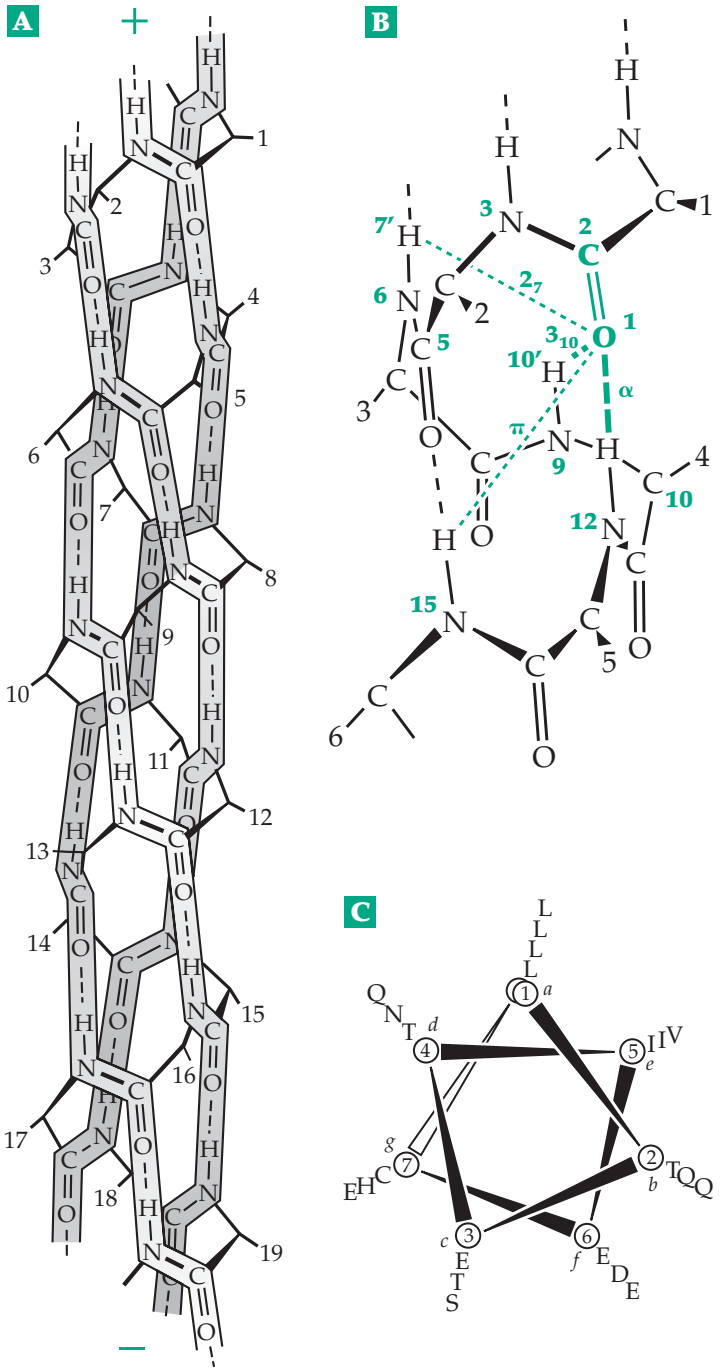



**Figure 2-20** (A) The α helix showing the three chains of interconnected amide groups and hydrogen bonds with partial net positive and negative charges at the ends. These chains run across the turns of the helical polypeptide backbone. (B) Scheme illustrating hydrogen-bonding pattern for $2_7$ ribbon, $3_{10}$ helix, α helix, and π helix. (C) The α helix represented as a helical wheel. Imagine viewing the helix from the N-terminal end of the segment with the lines corresponding to the backbone of the peptide. Residues, which are designated by single letters, are spaced 100° apart since there are 3.6 residues/turn. The peptide shown is a 22-residue sequence from a "leucine-zipper" domain of a protein that participates in gene regulation. From Fathallah-Shaykh *et al.*[145a]

negative charges at the C terminus and creates a "macrodipole." It has been estimated that each end of the helix carries one-half an elementary unit of charge.[147–148b] However, this may be an overestimate. The partial charges are connected by three chains of hydrogen bonds that run across the turns of the helix as is indicated in Fig. 2-20A. These chains are polarized and also possess a large **hyperpolarizability**,[149] i.e., the polarization of the helix increases more rapidly than in direct proportion to an applied electrical field.

In a $3_{10}$ helix, there are just two chains of hydrogen bonds across the polypeptide chain and running the length of the helix. The characteristic hydrogen bonding of the α helix and that of the $3_{10}$ helix are often both present within a helix. Irregularities with $3_{10}$ type hydrogen bonds, arising from interactions with amino acid side chains or with solvent molecules, may cause a helix to be kinked or curved.[132] Side chains of polar residues, including those of Asn, Asp, Thr, and Ser (and less often Glu, Gln, or His), frequently fold back and hydrogen bond to the NH and CO groups that carry the partial charges at the helix ends (Fig. 2-20). The side chain of the third residue in the helix may also hydrogen bond to the NH of the first residue.[133,150–157] The hydrogen bonding of a negatively charged side chain group to the N-terminal end of the helix or of a positively charged group to the C-terminal end provides an **N-cap** or a **C-cap** which helps to stabilize the helix by strengthening its hydrogen bonds.[158] However, the most frequent residue at the C-terminal end is glycine.[151,155,159] Helices often point toward active sites of enzymes and interactions of the helix dipoles with substrates undergoing reaction may be important to the mechanism of action of these catalysts.[148,149,160]

***Stacking of helices in proteins.*** Many proteins are made up almost entirely of α helices. One of these, **myoglobin**, was the first protein for which the complete three-dimensional structure was worked out by X-ray crystallography.[129] Myoglobin is a small oxygen-carrying protein of muscle. Its structure is closely related to that of hemoglobin of blood. Its 153 amino acid residues are arranged in eight different α-helical segments containing from 7 to 26 residues each. These rodlike helices are stacked together in an irregular fashion as shown in Fig. 2-19. Serum albumin (Box 2-A) has 28 helices organized into three homologous domains. In contrast, the filamentous bacterial viruses have protein coats made up of small subunits, each coiled as a single α helix. These are packed in a regular manner to form the rodlike virus coats (Fig. 7-7).

***Coiled coils.*** In a large family of proteins, two right-handed α helices are coiled around each other in a left-handed **superhelix** (Fig. 2-21).[161–167] This **coiled coil** structure was first suggested by Francis Crick[166] to account for the fact that the X-ray diffraction pattern of the **keratins** of skin and hair indicated a pitch of 0.51 nm rather than the 0.54 nm expected for an α helix.

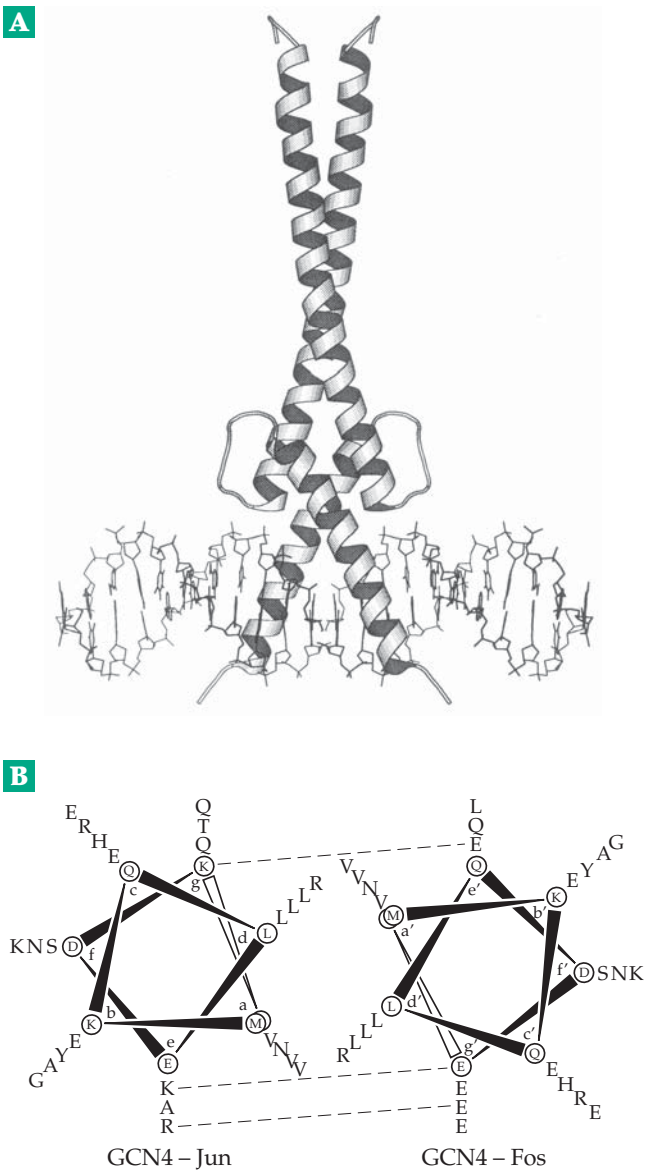



**Figure 2-21** (A) Ribbon drawing of the transcription factor called Max in a complex with DNA. The C termini of the peptide chains are at the top.[169] Courtesy of S. K. Burley. (B) Helical wheel representation of residues 2–31 of the coiled coil portion of the leucine zipper (residues 249–281) of the related transcription factor GCN4 from yeast. The view is from the N terminus and the residues in the first two turns are circled. Heptad positions are labeled *a*–*g*. Leucine side chains at positions *d* interact with residues *d′* and *e′* of the second subunit which is parallel to the first. However, several residues were altered to give a coiled coil that mimics the structure of the well-known heterodimeric oncoproteins Fos and Jun (see Chapter 11). This dimer is stabilized by ion pairs which are connected by dashed lines. See John et al.[172]

Crick suggested that two supercoiled α helices inclined at an angle of 20° to each other would produce the apparent shortening in the helix pitch and would also permit nonpolar side chains from one strand to fit into gaps in the surface of the adjacent strand, a **knobs-into-holes** bonding arrangement. Helical strands tend to coil into ropes because a favorable interstrand contact can be repeated along the length of the strands only if the strands coil about each other. A coiled coil can often be recognized by **heptad repeats**:

$$(\mathbf{a}\text{-b-c-}\mathbf{d}\text{-e-f-}\mathbf{g})_n$$
$$1\ 2\ 3\ 4\ 5\ 6\ 7$$

Here, residues *a*, *d*, and *g* (1, 4, and 7) often carry nonpolar side chains. These come together in the coiled coil as is illustrated in the **helical wheel** representations in Fig. 2-21B and provide a longitudinal hydrophobic strip along the helix.[167] Charged groups are often present in other locations and in such a way as to provide electrostatic stabilization through interactions between residues within a single α helix[170-171a] or between the pair of helices. The latter type of interaction also determines whether the coiled-coil consists of a parallel or antiparallel pair and whether the two helices are of identical or of differing amino acid sequence.[171b]

Recent attention has been focused on a DNA-binding structure called the **leucine zipper**. A pair of parallel α helices are joined as a coiled coil at one end but flare out at the other end to bind to DNA. In the yeast transcription factor GCN4, whose three-dimensional structure has been determined to high resolution (Fig. 2-21B),[172] the *d* position of the coiled coil is occupied by leucine and the *e* and *g* positions are often occupied by charged groups that form stabilizing ion pairs. Residues at positions *b*, *c*, and *f* are generally on the outside and exposed to solvent.[168,171,173] The coiled coil flares out at the C-terminal ends and carries DNA-binding groups. The structure of a related transcription factor is shown in Fig. 2-21A.[169]

The muscle proteins **myosin**[174] and **tropomyosin** also both consist of pairs of identical chains oriented in the same direction. The two 284-residue tropomyosin chains each contain 40 heptads and are linked by a single disulfide bridge. X-ray crystallographic studies[175,176] and electron microscopy[177] show that the molecule is a rod of 2.0 nm diameter and 41 nm length, the dimensions expected for the coiled coil. However, as with other "regular" protein structures, there are some irregularities. Myosin chains (Chapter 19) contain 156 heptads.

Another group of proteins have parallel coiled coils flanked by nonhelical domains in subunits that associate as filaments. These include the **keratins** of skin as well as the intermediate fiber proteins of the cytoskeleton (Chapter 7).[164,178] Natural coiled coils often have a parallel orientation, but synthetic peptides have been designed to form antiparallel coiled coils.[179,179a]

***Helix bundles.*** A third peptide chain can be added to a coiled coil to form a triple-stranded bundle.[180–183] An example is the glycoprotein **laminin** found in basement membranes. It consists of three peptide chains which, for ~600 residues at their C-terminal ends, form a three-stranded coil with heptad repeats.[182,184] Numerous proteins are folded into four helical segments that associate as **four-helix bundles** (Fig. 2-22).[185–188] These include electron carriers, hormones, and structural proteins. The four-helix bundle not only is a simple packing arrangement, but also allows interactions between the + and – ends of the macro-dipoles of the helices.

Membranes contain many largely α-helical proteins. Cell surface receptors often appear to have one, two, or several membrane-spanning helices (see Chapter 8). The single peptide chain of the bacterial light-operated ion pump **bacteriorhodopsin** (Fig. 23-45) folds back upon itself to form seven helical rods just long enough to span the bacterial membrane in which it functions.[189] Photosynthetic reaction centers contain an α helix bundle which is formed from two different protein subunits (Fig. 23-31).[190] A recently discovered **α,α barrel** contains 12 helices. Six parallel helices form an inner barrel and 6 helices antiparallel to the first 6 form an outer layer (see Fig. 2-29).[191–193]

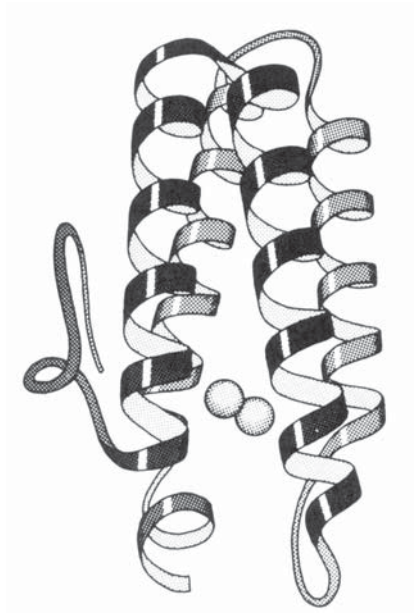


**Figure 2-22** Ribbon drawing of an up-and-down four-helix bundle in **myohemerythrin**. The two spheres represent the two iron atoms which carry an $O_2$ molecule. They are coordinated by histidine and aspartate side chains. Courtesy of J. Richardson.[117]

## 4. Polyglycine II and Collagen

In a second form of polyglycine each amino acid residue is rotated 120° from the preceding one about a 3-fold screw axis as is shown in the end view of Fig. 2-23A. The angle ψ is about 150° while ϕ is about –80° for each residue. The distance along the axis is 0.31 nm/residue and the repeat distance is 0.93 nm. The molecules can coil into either a right-handed or a left-handed helix. In this structure, the N–H and C=O groups protrude perpendicular to the axis of the helix and, as in the β structure, can form H-bonds with adjacent chains.

Poly (L-proline) assumes a similar helical structure. However, the presence of the bulky side chain groups induces a *left-handed* helical twist. **Collagen**, the principal protein of connective tissue, basement membranes, and other structures, is the most abundant protein in the animal body. Its fundamental unit of structure is a triple helix of overall dimensions 1.4 x 300 nm which resembles the structure of polyglycine II but contains only three chains.[195] The left-handed helices of the individual chains are further wound into a right-handed superhelix (Fig. 2-23B,C). Collagen contains 33% glycine and 21% (proline + hydroxyprolines). The reason for the high glycine content is that bulky side chain groups cannot fit inside the triple helix. A hydrogen-bonded cylinder of hydration surrounds each triple helix and is anchored by hydrogen bonds to the –OH groups of the hydroxyproline residues.[196] Sequences of several collagen chains have been established. One of these contains ~1050 residues, ~1020 of which consist of ~340 triplets of sequence GlyXY where Y is usually proline and X is often hydroxyproline (Hyp).[197,198] The commonest triplet is Gly-Hyp-Pro.

There are several types of collagen. In one type, two identical chains of one kind are coiled together with a third dissimilar chain to form the triple helix. Several of these triple helices associate to form 8-nm microfibrils (Fig 2-23D).[199] Once synthesized, collagen is extensively modified and crosslinked. See Chapter 8.

Collagen-like triple helices also occur within other proteins. One of these is protein C1q, a component of the **complement system** of blood (Chapter 31). This protein interacts with antibodies to trigger a major aspect of the immune response. C1q has six subunits, each made up of three different polypeptide chains of about 200 residues apiece. Beginning a few residues from the N termini, there are over 80 residues in each chain with collagen-like sequences. The three chains apparently form a triple helix within each subunit. However, the C-terminal portions are globular in nature.[200] Collagen-like tails also are present on some forms of the enzyme acetylcholinesterase (see Chapter 12C,10). The **extensins** of plant cell walls contain 4-hydroxyproline and evidently have a structure related to that of collagen.[201,202] Shorter 4- to 8-residue segments of left-handed polyproline helix are found in many proteins.[203]

## 5. Turns and Bends

To form a globular protein, a polypeptide chain must repeatedly fold back on itself. The turns or bends by which this is accomplished can be regarded as a third major secondary structural element in proteins. Turns often have precise structures, a few of which are illustrated in Fig. 2-24. As components of the loops of polypeptide chains in active sites, turns have a special importance for the functioning of enzymes and other proteins. In addition, tight turns are often sites for modification of proteins after their initial synthesis (Section F).

The **β turn** (Fig. 2-24) is often found in hairpin or reverse turns at the edges of β sheets (Fig. 2-11) and at other locations.[204–212] If all four residues that contribute to β bends are counted, they constitute about one-third of the amino acid residues in most proteins.[124] In many β turns, the C=O of the first residue hydrogen bonds to the NH of the fourth residue. This hydrogen bond may be part of the hydrogen bond network of a β pleated sheet. The peptide unit between α-carbon atoms 2 and 3 of the turn is perpendicular to the sheet. There are two possibilities for the orientation of this peptide unit. In a *type I turn*, the C=O is down when the turn is viewed as in Fig. 2-24, while the side chains of residues 2 and 3 point upwards or outward on the opposite side of the bend. In a *type II turn*, the C=O is up and the NH down. Residue 3 is always glycine in a type II turn because the side chain would collide with the C=O group if any other amino acid were present. As is seen in Fig. 2-24, a *trans*-proline can fit at position 2 in a type II turn[206] as well as in type I turns. A *cis*-proline residue can fit at position 2 or position 3 in a type I β turn.[137,213]

The *type III β turn* is similar to a type I turn but has the ϕ, ψ angles of a $3_{10}$ helix and the two chains emerging from the turn are not as nearly parallel as they are in type I turns. Beta turns of the less common types I′, II′, and III′ have a left-handed twist. As can be seen in Fig. 2-24, this permits a better match to the twist of strands in a β sheet. Unless glycine is present, these bends are less stable because of steric hindrance.[214,215] Polar side chain groups such as those of aspartate or asparagine often form hydrogen bonds to the central peptide units of β turns.[214]

The tight γ turn[215] and the proline-containing β turn shown in Fig. 2-24 are thought to be major components of the secondary structure of **elastin**.[216–218] This stretchable polymer, which consists largely of nonpolar amino acids, is the most abundant protein of the elastic fibers of skin, lungs, and arteries. The

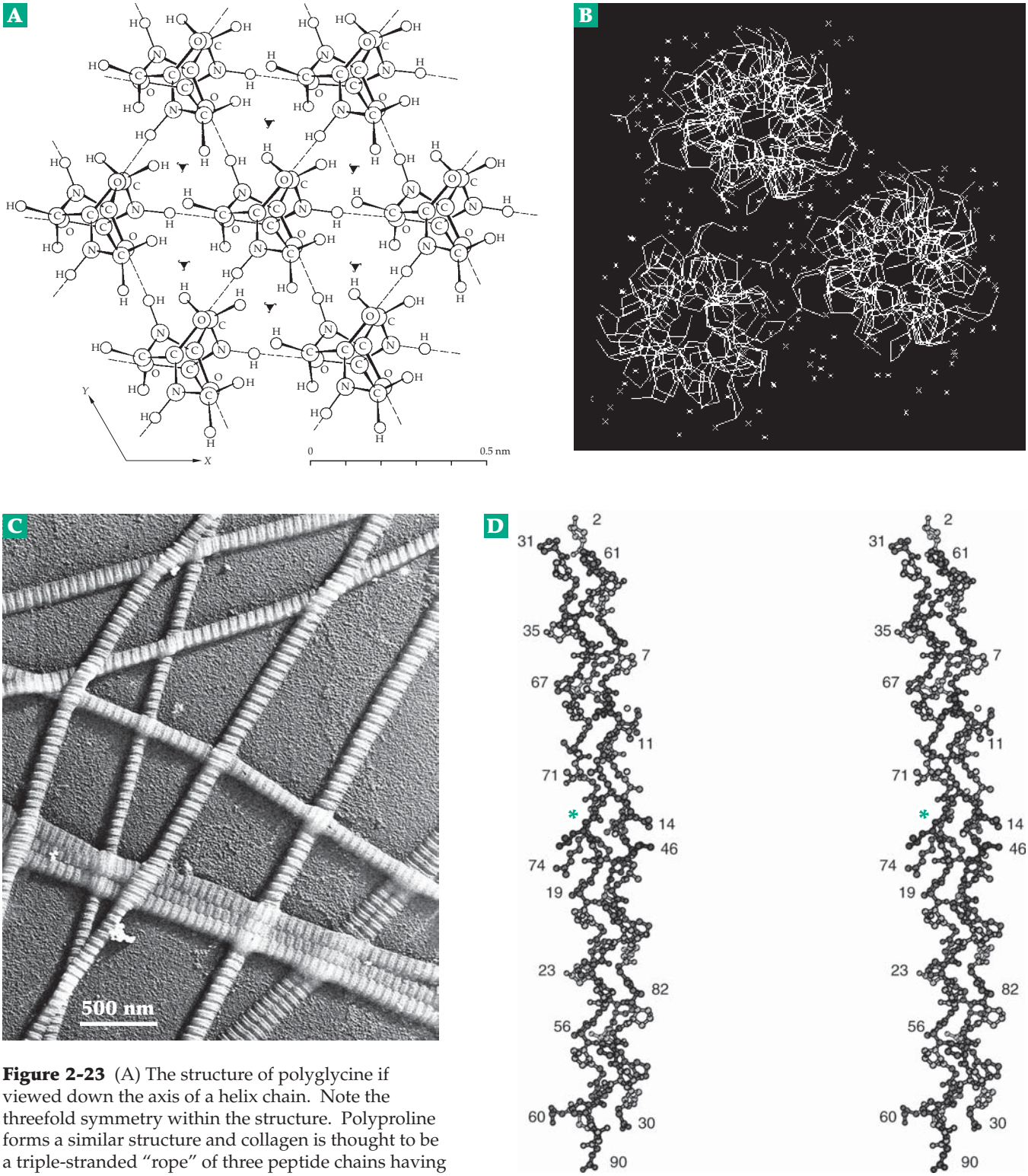



**Figure 2-23** (A) The structure of polyglycine if viewed down the axis of a helix chain. Note the threefold symmetry within the structure. Polyproline forms a similar structure and collagen is thought to be a triple-stranded "rope" of three peptide chains having essentially the same structure but in addition containing a right-handed supercoil. This is illustrated by the crystal structure of a collagen-like peptide shown in B and C. (B) View similar to that in (A), but showing how three chains form the triple-stranded ropes separated by a cylinder of hydration. The structure is shown as a wire model. The x's are water molecules. They form an extensive network of H-bonds to one another and to peptide groups. From Bella *et al.*[194] C) Electron micrograph of collagen fibrils, each of which consists of many triple-helical units. Deposited from suspension and shadow cast with chromium. Courtesy of Jerome Gross, M.D. (D) Stereoscopic view of a collagen-like model peptide. Each of the three parallel 30-residue chains contains a (Pro-Hyp-Gly)$_3$ "cap" at each end and the 12-residue sequence Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala-Gly-Pro-Hyp-Gly in the center. This sequence is found in human type III collagen and includes a site (green asterisk) of known mutations (see Chapter 8). From Kramer *et al.*[194a] Courtesy of Helen Berman.

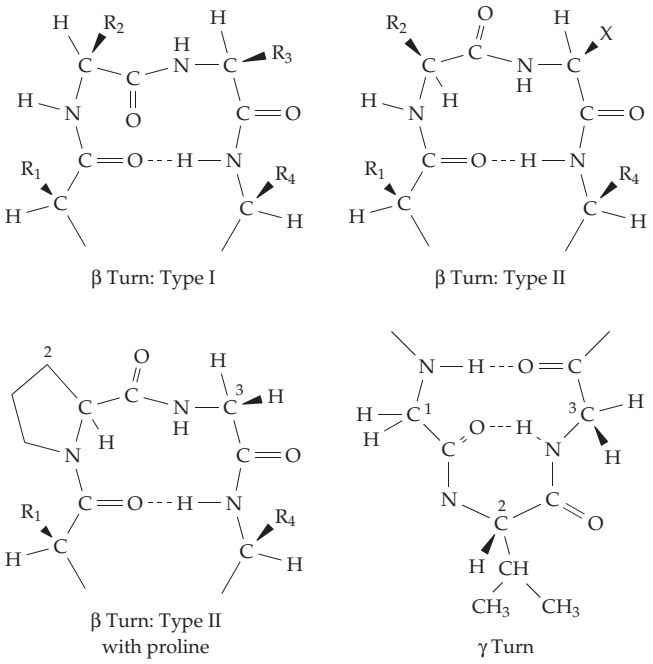



**Figure 2-24** Tight turns found in polypeptide chains. Two types of β turn are shown. A third variant, the type III or $3_{10}$ turn resembles the type I turn but has the ϕ, ψ angles of a $3_{10}$ helix. Type II β turns containing proline and tighter γ turns are thought to be major structural components of elastin. Another β turn, lacking the hydrogen bond has a *cis*-proline residue at position 3.

70-kDa chains are thought to have 70- to 75-residue regions in which the polypeptide folds back on itself repeatedly with a large number of bends in a broad left-handed **β spiral**.[216,218a] The consensus sequence VPGVP, which tends to form a type II β turn with proline in position 2 (Fig. 2-24) is present in long tandem repeats e.g., $(VPGVG)_{11}$. These extensible regions alternate with short α helices which are crosslinked to other chains. Similar structures are present in silks (Box 2-B) and in proteins of wheat gluten.[217a,218b]

Besides hairpin turns and broader U-turns, a protein chain may turn out and fold back to reenter a β sheet from the opposite side. Such **crossover connections**, which are necessarily quite long, often contain helices. Like turns, crossover connections have a handedness and are nearly always right-handed (Fig. 2-25).[117,219] Most proteins also contain poorly organized loops on their surfaces. Despite their random appearance, these loops may be critical for the functioning of a protein.[220] In spite of the complexity of the folding patterns, peptide chains are rarely found to be knotted.[221]

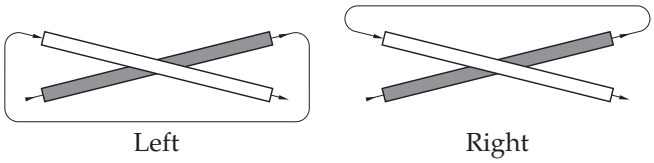



**Figure 2-25** Right- and left-handed crossover connections in proteins. These connections are nearly always right-handed. The broad arrows represent β strands. The crossover often contains a helix. Units of two adjacent β strands (βαβ units) with an α helix between are found frequently in globular proteins.

## 6. Domains, Subunits, and Interfaces

Many proteins are organized into tightly folded globular "domains" consisting of 50–150 amino acid residues.[117,222–227] Smaller proteins may have 2 or 3 domains but large proteins may have more. For example, the immunoglobulin "heavy chains" (Chapter 31) have four or five domains similar to that shown in Fig. 2-16B. The enormous 3000-kDa muscle protein **titin** contains 260 domains, many of which are of the immunoglobulin typeFig. 2-16,C.[228,117a] In most proteins the polypeptide chain folds to form one domain, then passes through a "hinge" to the next. In others, the C terminus or the N terminus of the polypeptide folds back across two or more domains as a kind of "strap" that helps to hold the protein together. Even when a protein contains only one domain, it often consists of two distinct lobes with a cleft between them. Many proteins, e.g. hemoglobin (Fig. 7-25), consist of subunits about the size of the globular domains in larger proteins.

Structural domains of proteins are sometimes encoded by a single coding segment of DNA i.e., by a single exon in a split gene. Domains of this type may have served as **evolutionarily mobile modules** that have spread to new proteins and multiplied during evolution. For example, the immunoglobulin structural domain is found not only in antibodies but also in a variety of cell surface proteins.[229–232]

Whether we deal with domains connected by a flexible hinge or with subunits, there are **interfaces** between the different parts of the protein. These interfaces are often formed largely of nonpolar groups. However, they frequently contain a small number of hydrogen bonds that bridge between one domain and another or between one subunit and another. In the case of hemoglobin, important changes occur in this hydrogen bonding and movement occurs along one of the interfaces between two subunits. Likewise the active sites of enzymes are often located at interfaces between domains. During catalysis, movement and reorganization of the hydrogen bonds and side chain packing in the interfaces may take place.[232a]

## 7. Packing of Side Chains

In Fig. 2-19B myoglobin is pictured as a cluster of α-helical rods surrounding the heme core. This picture is incomplete because the space between the rods and inside the molecule is tightly packed with amino acid side chains almost all of which are hydrophobic. The same is true for the β barrels of Fig. 2-16, which are filled largely with nonpolar side chains. As the structures of more and more proteins have been determined, a consistent pattern has emerged. Within the interior of proteins the side chain groups are packed together remarkably well.[99,100,222,233–236] Although occasional holes are present, they are often filled by water molecules.[99,237a] The **packing density**, the volume enclosed by the van der Waals envelope divided by the total volume, is ~0.75 for the interior of the lysozyme and ribonuclease molecules compared with the theoretical value of 0.74 for close-packed spheres. However, regions with many hydrogen bonds may be less tightly packed.[238]

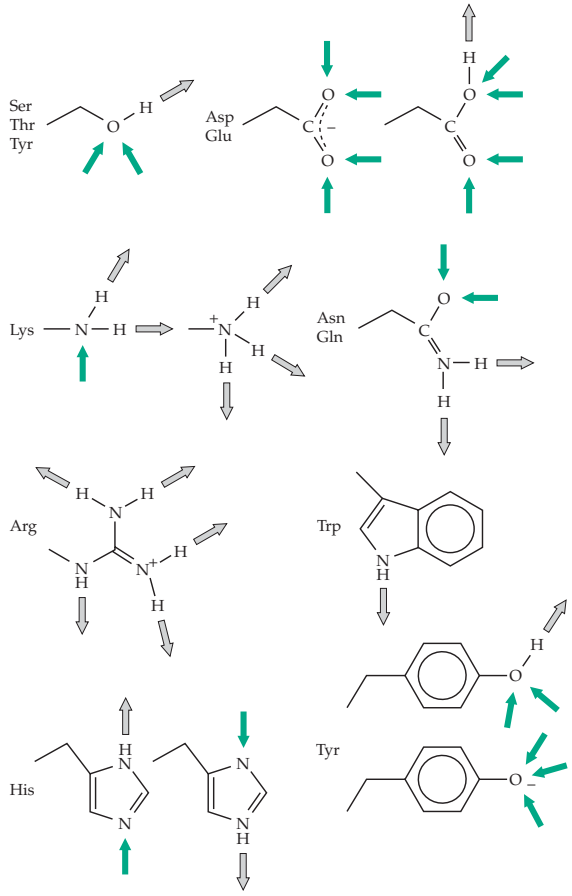



**Figure 2-26** Some of the possibilities for hydrogen bonding of side chain groups in proteins. Oxygen atoms can and frequently do form up to three hydrogen bonds at once. Open arrows point *from* H-atoms and *toward* electron donor pairs.

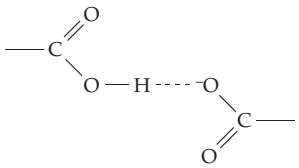
The interiors of proteins often contain large numbers of aromatic side chains which are frequently associated as pairs or as **aromatic clusters**.[239,240] For example, see the structure of plastocyanin in Fig. 2-16A. Rings may lie perpendicular one to another or be "stacked" face to face but offset. Oxygen atoms often lie in contact with the edges of aromatic rings.[241] The planar guanidinium groups of arginine side chains often stack against aromatic rings and amide groups may sometimes do the same.[242] It has been suggested that both aromatic: aromatic and aromatic:oxygen interactions may be associated with additional stabilization of the protein by ~4- to 8- kJ/mol. Tyrosine side chain –OH groups often stabilize ends of β strands by forming H-bonds to backbone atoms.[243]

Most polar groups are on the surfaces of proteins, and those that are not are almost always hydrogen bonded to other groups in the interior.[244] While most nonpolar groups are inside proteins, they are also present in the outer surfaces where they are often clustered into **hydrophobic regions** or "**patches**." The latter may be sites of interaction with other proteins or with lipid portions of membranes.

## 8. The Network of Internal Hydrogen Bonds

The fact that nonpolar groups tend to be buried in the interior of proteins suggests that the inside of a protein might be a flexible blob of oily material. In fact, the nonpolar groups tend to be densely packed and aromatic rings often impart considerable rigidity to the hydrophobic cores. Buried polar groups, which are invariably hydrogen bonded to other side chain groups, to amide groups of the polypeptide backbone,[244–247] or to buried water molecules, form a well-defined internal network. When a series of closely related proteins, for example, those having the same function in several different species, are compared, the hydrogen bonded network is often nearly invariant. This suggests a functional significance. The hydrogen bonding possibilities for some of the side chain groups in proteins are indicated in Fig. 2-26.

Charged side chains are sometimes buried in the interior of proteins, usually together with an ion of the opposite charge to form a **hydrogen-bonded ion pair**.[248,248a,248b] However, there is sometimes a single charge that is "neutralized" by the interaction with dipoles of polar groups.[217] Sometimes an undissociated carboxyl group is hydrogen bonded to a carboxylate ion.[249]



A buried carboxyl group of this type will display a p$K_a$ value far higher than the normal value.

Arginine side chains are large and able to form multiple hydrogen bonds[247,250] as well as salt linkages[251] to different parts of a folded peptide. Cations of arginine and lysine are never buried unless in ion pairs. Protons from strong acids have long been known to bind to centers of benzene rings and water can form weak hydrogen bonds to the centers of aromatic rings.[252,253,253a] Such bonds also occur within proteins and often involve the binding of guanidinium groups or of inorganic cations to indole rings of tryptophans.[253b,253c] Protonated imidazole groups may also bind to aromatic rings.[253d]

Another important aspect of the structure of proteins is the presence of **hydrogen-bonded water molecules** in pockets and cracks. These molecules, as well as a much larger number of water molecules bound at the outer surface, are clearly visible from X-ray studies. They often occur singly, bonded to the ends of amide groups, especially to the carbonyl ends. Internal bends of the peptide chain are almost always hydrated.[254] These water molecules often make two or more hydrogen bonds to different parts of the protein or to other water molecules. Clusters of water molecules,[49,255,255a] sometimes in the form of pentagonal rings,[256] are often present. NMR spectroscopy has shown that water molecules bound to protein *surfaces* exchange rapidly with the bulk water in which the protein is dissolved; the "residence time" on the protein is typically less than a nanosecond. *Interior* water molecules have much longer residence times of $10^{-2}$ to $10^{-8}$ s for a small protein.[257] They may be regarded as part of the protein structure.

## E. Folding Patterns and Protein Families

Proteins are folded in many ways. We have already considered several simple patterns: the antiparallel β cylinder (Fig. 2-16), the 2-helix coiled coil (Fig. 2-21) and the 3- and 4-helix bundles (Fig. 2-22). Another simple motif that has been found repeatedly is the **helix–turn–helix** or **helix–loop–helix** in which two helices at variable angles, one to another and with a turn or short loop between them, form a structural unit. DNA-binding **repressors** and **transcription factors** (see Fig. 2-21 and also Chapter 5) often contain this motif as do many $Ca^{2+}$-binding proteins. Proteins containing 3–6 helical segments, often fold into a roughly polyhedral shape.[258,259] An example is myoglobin (Fig. 2-19B).

### 1. Complex Folding Patterns

Proteins often contain elements of both α and β structure. One of the first of the complex folding motifs to be recognized was a nucleotide-binding domain identified by Rossmann and associates.[260–262] This Rossman fold contains six parallel β strands which alternate with six helices. The result of the parallel β structure is that the helices are also parallel and that their amino-terminal ends, which carry partial positive charges, are aligned in approximately the same directions. The positive end of one of the helices lies behind the negatively charged phosphate groups[148] which characteristically bind at the edge of the sheet containing the C termini of the β strands. This can be seen in Fig. 2-13, which shows binding of the coenzyme NAD to the nucleotide binding domain of **glyceraldehyde phosphate dehydrogenase**. Similar nucleotide-binding domains are found in many other dehydrogenases whose members constitute a **protein family**.

Figure 2-27 depicts **topology diagrams** for the Rossman fold and for two related families of proteins. These families bind the nucleotides called GTP and ATP, respectively. Both are structural relatives of NAD. A major part of the structure of all of the proteins in these families consist of β–α units, each one containing a β strand followed by a helix. They are

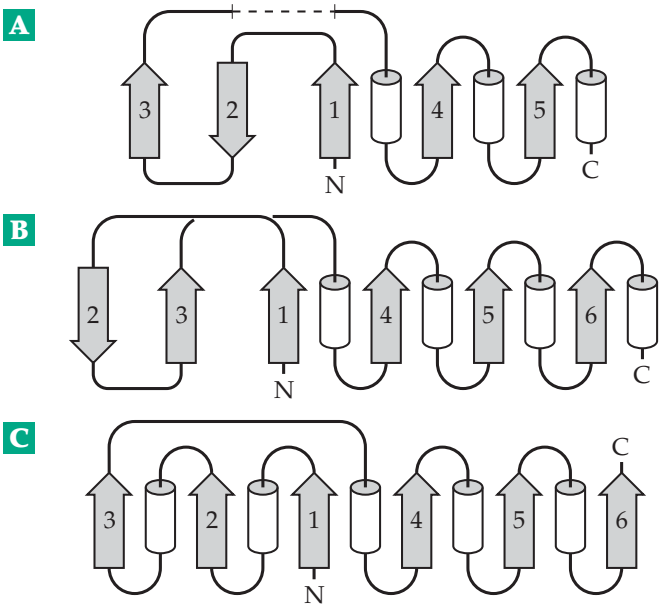



**Figure 2-27** Topologies of the folds of three families of nucleotide binding α/β proteins. Cylinders represent α helices and arrows β strands. (A) The ATPase fold for the clathrin-uncoating ATPase; (B) The G-protein fold that binds GTP and is found in *ras* proteins; (C) The Rossmann fold that binds NAD in several dehydrogenases. From Brändén.[262]

**Figure 2-28** The eight-fold α/β barrel structure of triose phosphate isomerase. From Richardson. (A) Stereoscopic view. (B) Ribbon drawing. Courtesy of Jane Richardson.[117]

classified as α/β proteins. The Rossman fold is composed of six β–α units. Recently a ribonuclease inhibitor protein with 15 consecutive β–α units has been characterized.[263] Each β–α unit contains several residues of leucine. This **leucine-rich repeat** occurs in many other proteins as well.[264,264a,264b]

The **α/β barrel** shown in Fig. 2-28 consists of 8 consecutive β–α units in a symmetric array.[265,266] By 1995 over 40 of these barrels had been identified in a diverse group of enzymes. One bifunctional enzyme contains two α/β barrels. Although the nature of the reaction catalyzed varies, the active site is always found in the center of the barrel at the C-terminal ends of the 8 parallel β strands and therefore between the N termini of the surrounding helices. The enzyme sequences show no homology and frequent occurrence of the 8-stranded barrel may reflect the fact that it is a natural packing arrangement of β–α units. However, a 10-stranded barrel of this type has also been found.[267]

This barrel can be compared with that of the 12-helix α,α barrel of a fungal glucoamylase whose structure is shown in Fig. 2-29. Numerous more complex folding patterns have been discovered. They have been classified by Jane Richardson.[117,122] Many of the proteins described by these folding patterns can be grouped into families and "superfamilies."[227] Chothia suggested that there may be about 1,000 families in nature;[268] over 700, with over 360 distinct folds have been identified.[268a]

## 2. Symmetry

A sometimes puzzling feature of protein structure is the widespread occurrence of an approximate two-fold axis of symmetry. This often arises as a natural result of association of a pair of irregular subunits (Chapter 7). The association is such that rotation

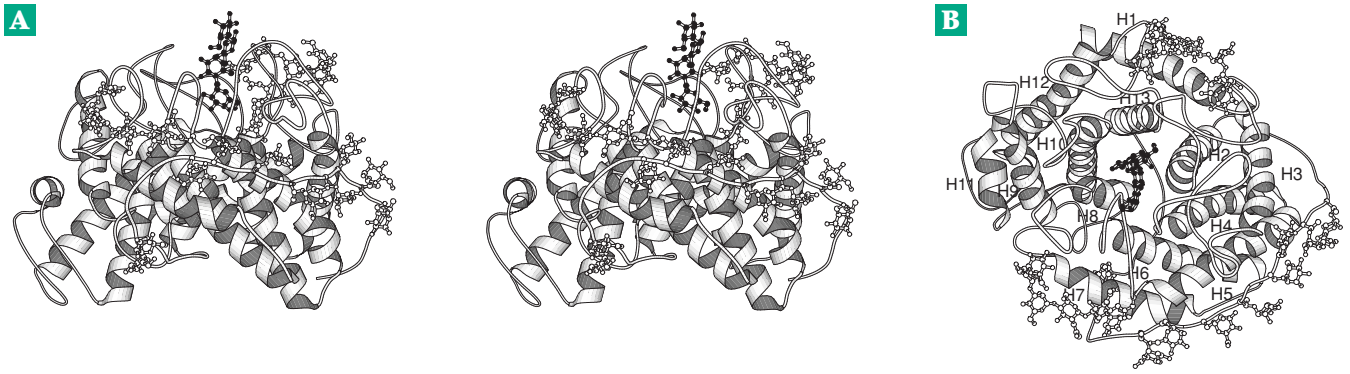



**Figure 2-29** Structure of the α,α barrel of a fungal enzyme glucoamylase. (A) side view (stereoscopic); (B) top view. The active site, which cleaves glucose units from the ends of starch chains, is in the depression in the center of the barrel. Here it is occupied by an inhibitor. See Aleshin *et al*.[192] Courtesy of Alexander Aleshin.

about the twofold axis will cause the two subunits to exchange positions and to remain in an identical chemical environment. Approximate symmetry is often observed also *within* single peptide chains. For example, in the Rossman fold (Figs. 2-13, 2-27), an approximate twofold axis passes between the center strands of the β sheet residues and relates the two flanking helices, which begin with residues R10 and T100, respectively. The bound $NAD^+$ also possesses an approximate twofold axis, but it is not quite symmetrically placed at the end of the β sheet. Both phospho groups are seen to interact with the N terminus of the helix beginning at residue 10. The small bacterial protein ferredoxin (Fig. 16-16B) contains two iron-sulfur clusters related by an approximate 2-fold axis. The two β cylinders of elastase (Fig. 12-9) as well as the two sides of the flattened β barrel of copper–zinc superoxide dismutase[269] are approximately related by twofold axes. The enzyme thiosulfate: cyanide sulfurtransferase (Eq. 24-46) is remarkably symmetric but the active site is located in just one half. The widespread existence of this approximate symmetry suggests a biological significance that remains to be discovered.

## 3. Effects of Sequence on Folding

Studies of synthetic polypeptides as well as examination of known protein structures reveal that some amino acids, e.g., Glu, Ala, Leu, tend to promote α helix formation. Others, such as Tyr, Val, and Ile, are more often present in β structure, while Gly, Pro, and Asn are likely to be found in bends.[270,270a,270b] The frequencies with which particular amino acids appear in helices, β structure, or turns were first compiled by

**TABLE 2-4**
**Classification of Protein Residues According to Their Tendencies to Form α Helix, β Structure, and β Turns[a]**

| Amino acid | $P_\alpha$ | Helix-forming tendency | $P_\beta$ | β structure-forming tendency | $P_t$ |
|---|---|---|---|---|---|
| $Glu^-$ | 1.51 | ++ | 0.37 | br+ | 0.44 |
| Met | 1.45 | ++ | 1.05 | + | 0.67 |
| Ala | 1.42 | ++ | 0.83 | i | 0.57 |
| Leu | 1.21 | ++ | 1.30 | + | 0.53 |
| Lys+ | 1.16 | + | 0.74 | br | 1.01 |
| Phe | 1.13 | + | 1.38 | + | 0.71 |
| Gln | 1.11 | + | 1.10 | + | 0.56 |
| Trp | 1.08 | + | 1.37 | + | 1.11 |
| Ile | 1.08 | + | 1.60 | ++ | 0.58 |
| Val | 1.06 | + | 1.70 | ++ | 0.30 |
| $Asp^-$ | 1.01 | w | 0.54 | br+ | 1.26 |
| $His^+$ | 1.00 | w | 0.87 | i | 0.69 |
| $Arg^+$ | 0.98 | i | 0.93 | i | 1.00 |
| Thr | 0.83 | i | 1.19 | + | 1.00 |
| Ser | 0.77 | i | 0.75 | br | 1.56 |
| Cys | 0.70 | i | 1.19 | + | 1.17 |
| Tyr | 0.69 | br | 1.47 | ++ | 1.25 |
| Asn | 0.67 | br | 0.89 | i | 1.68 |
| Pro | 0.57 | br+ | 0.55 | br+ | 1.54 |
| Gly | 0.57 | br+ | 0.75 | br | 1.68 |

++ = strong former; + = former; w = weak former; i = indifferent; br = breaker; br+ = strong breaker

[a] The conformational parameters $P_\alpha$, $P_\beta$, and $P_t$ (β turn) are the frequencies of finding a particular amino acid in an α helix, β structure, or β turn (in 29 proteins of known structure) divided by the average frequency of residues in those regions. Residues are arranged in order of decreasing tendency toward helix formation. From Chou, P. V. and Fasman, G. D (1974) *Biochemistry* **13**, 222–245.

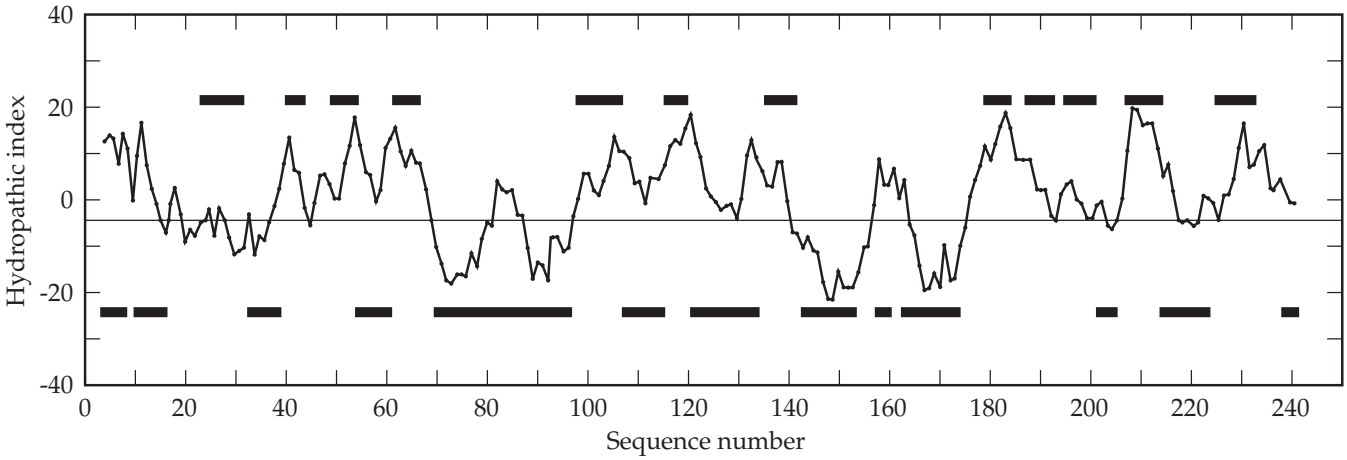



**Figure 2-30** Plot of hydropathy index versus sequence number for bovine chymotrypsinogen. The indices for individual residues have been averaged nine at a time. The solid bars at the top of the plot mark interior regions as determined by crystallography. The solid bars below the plot indicate regions that are on the outside of the molecule. From Kyte and Doolittle.[280]

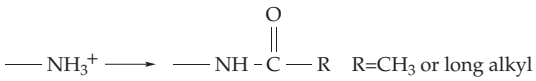
Chou and Fasman[270] (Table 2-4). These frequencies also differ significantly between parallel and antiparallel β sheets. Some of the preferences are readily understandable. For example, glycine is too flexible to stabilize a helix, but it can occur in helices between residues that are better helix formers. Because the peptide linkage to its nitrogen lacks an NH group, proline can fit into only one of three positions at the N terminus of a helix. For the strong helix formers, the β-$CH_2$ groups fit into the helix well and provide stabilization through van der Waals interactions. However, the β-methyl groups of valine and isoleucine cause crowding. For these residues, a β structure is more favorable than a helix.[271] Side chains of Asp, Asn, Ser, Thr, and Cys can hydrogen bond to backbone amide groups and can either stabilize or destabilize a helix or β sheet depending upon their location.

It appears that the folding pattern of a peptide is encoded in the sequence itself.[272–274] Thus, when several residues that favor helix formation are clustered together, a helix may form. Chou and Fasman suggested that when four helix formers out of six residues are clustered, nucleation of a helix takes place. The helix can then be elongated in both directions until terminated by a proline or other "helix breaker"; additional folding can then occur. If three out of five β formers are clustered, a β strand may form. If random folding brings two or more of these strands together they may associate to form the nucleus for a β sheet. Some success has been achieved using this approach in predicting whether a given residue will be found in a helix, a β strand or a loop.[275] However, prediction of complete folding patterns is much more difficult. Many new approaches are being explored.[275–279d] The problem is an important one. Although three-dimensional structures have been determined for thousands of proteins, sequences are known for hundreds of thousands. The number is growing rapidly. Being able to predict correctly a three-dimensional protein structure will be a major scientific accomplishment with many practical consequences.

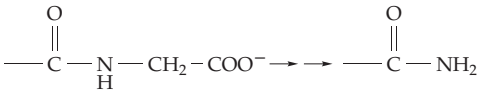
A way of examining the entire sequence for polar or nonpolar character of the side chains was introduced by Kyte and Doolittle.[280,281] A **hydropathy index** based on the polarity of the side chains of a given residue and of its neighbors in the sequence is plotted against residue number (Fig. 2-30). Helices are often found to have an **amphipathic** character, hydrophobic on one side and hydrophilic on the other. Such helices can be characterized by a plot such as that in Fig. 2-30. The hydrophobic side of an amphipathic helix can pack against a hydrophobic core of a protein, can lie against a membrane, or can be aligned with other helices to form coiled coils (Fig. 2-21) or to give a hydrophilic channel with an outer hydrophobic surface that can fit into a cell membrane.

## F. Chemical Modification and Crosslinking

Some newly synthesized proteins, upon folding, are ready to "go to work" immediately but others must be modified. Pieces are frequently cut from the ends or out of the center of a folded peptide chain. Sometimes the amino and carboxyl termini are converted to nonionic groups, e.g.,



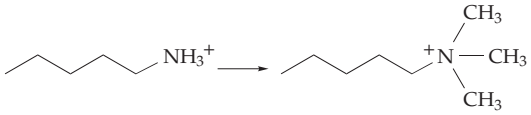

(2-12)



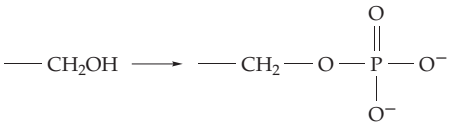

(2-13)

Side chains may be modified in a very large number of different ways.[282–284] These include acetylation and other kinds of acylation (Eq. 2-12),[285–287] **methylation** (Eqs. 2-14, 2-15), **phosphorylation** (Eq. 2-16), **phosphoadenylation**,[288] formation of **sulfate esters** (Eq. 2-17),[289,290] and hydrolysis (Eqs. 2-18, 2-19).[291] In at least a few proteins some L-amino acid residues are converted to D-.[292–294]

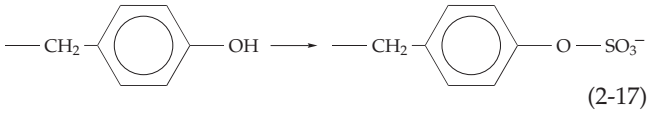
(2-14)



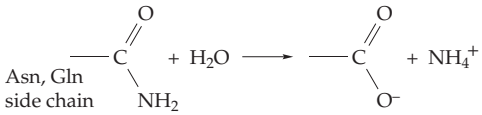

(2-15)



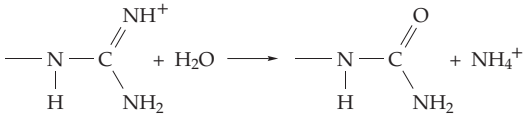

(2-16)




(2-17)

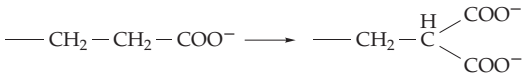



(2-18)




(2-19)

Usually, only one or a small number of side chains of a protein is modified. However, glycoproteins may contain many different attached sugar molecules. An example is the glucoamylase shown in Fig. 2-29. In several proteins involved in blood clotting (Fig. 12-17), as many as 10 glutamic acid side chains are carboxylated (Eq. 2-20).

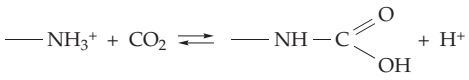


(2-20)

The 216-residue hen egg yolk storage protein **phosvitin** contains 123 serine residues, most of which have been phosphorylated (Eq. 2-16).[295] A basic protein of the myelin sheath of neurons contains as many as 6 specific residues of **citrulline** (Eq. 2-19).[296] An adhesive protein from the foot of a marine mollusk contains ~80 repeated sequences containing hydroxy-proline 2,3-dihydroxyproline and **3,4-dihydroxyphenylalanine** (Dopa).[297,298]

Some modification reactions alter the electrical charge on a side chain and, as a consequence, can affect the ways in which the protein interacts with other molecules. Negative charges added in formation of clusters of γ-carboxyglutamates (Eq. 2-20) create strong **calcium ion-binding centers** in the modified proteins. Acylation of N termini or of Ser, Thr, or Cys side chains by long-chain fatty acids provides hydrophobic "tails" able to anchor proteins to membrane surfaces.[285,286] Addition of polyprenyl groups to cysteine side chains near the C-termini has a similar effect.[299,300] These and more complex membrane anchors are considered in Chapter 8. While modified amino acids can be found at many places in a protein, they are often located at turns. For example, serine or threonine residues in turns are often phosphorylated or glycosylated. Modification of proteins is dealt with further in Chapter 10 and at other points in the book. Glycosylation of proteins is considered in Chapter 20.

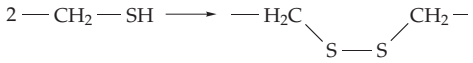
In addition to deliberate enzyme-catalyzed processes, there are nonenzymatic processes that alter proteins. These include the degradative reactions described in Section 5 and also *reversible* reactions that may be physiologically important. For example, the N-terminal amino groups of peptides, and other amino groups of low p$K_a$ can form **carbamates** with bicarbonate (Eq. 2-21).[301–303] This provides an important mechanism of carbon dioxide transport in red blood cells (Chapter 7) and a way by which $CO_2$ pressure can control some metabolic processes.



(2-21)

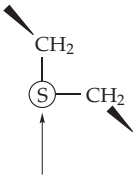
## 1. Disulfide Crosslinkages

The linking together of two different parts of a peptide chain or of two different peptide chains is extremely important to living beings.[304,305] One of the most widespread of crosslinkages is the **disulfide bridge**. It forms spontaneously when two –SH groups of cysteine side chains are close together and are oxidized by $O_2$ or some other reagent (Eq. 2-22).



(2-22)

The disulfide group is inherently chiral, with the preferred torsion angle about the S–S bond being ~90°. Both right- and left-handed disulfide bridges occur in proteins.[305,306]

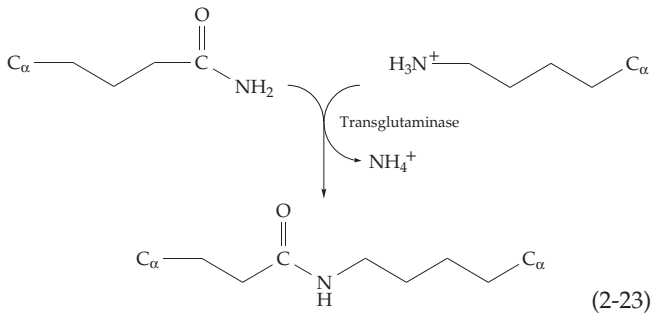


Two S atoms viewed down S—S axis

Disulfide linkages are frequently present in proteins that are secreted from cells but are less common in enzymes that stay within cells. Perhaps because the latter are in a protected environment, the additional stabilization provided by disulfide bridges is not needed. Disulfide bridges are not only found within single-subunit proteins but they also link different peptide chains. For example, the four chains of each immunoglobulin molecule (Chapter 31) are joined by disulfide bridges and each domain within the chains is stabilized by a disulfide bridge. One of the most highly crosslinked proteins known occurs in the keratin matrix of hair. Breakage of the –S–S– linkages of this protein is an essential step in the chemical "permanent wave" process. A thiol compound is used to reductively cleave the crosslinks and after the hair is reset new crosslinks are formed by air oxidation.

## 2. Other Crosslinkages

Another common crosslink is an amide formed between the γ-carboxyl group of a glutamic acid side chain and an amino group from a lysine residue.[307] This **isopeptide** linkage is formed from a residue of gluta-mine through the action of the enzyme **transglutaminase** (Eq. 2-23). Isopeptide crosslinks are found in hair, skin, connective tissue, and blood clots.




(2-23)

Occasionally an isopeptide linkage joins amino groups of lysine side chains to the C-terminal carboxyl groups of other peptide chains to give branched chains, e.g. see **ubiquitin** (Box 10-C). Elaborate crosslinks derived from lysine are found in collagen and elastin (Chapter 8). Dityrosine linkages formed by oxidative joining of the aromatic rings of tyrosine are found in insect cuticle and in the plant cell wall extensins[202] (Chapters 20, 25).

## G. Dynamic Properties of Proteins

Sometimes using energy from the cleavage of chemical bonds, sometimes depending only upon energy provided by the random bombardment by solvent molecules, proteins perform their specific functions with amazing speed. A question that has long intrigued biochemists is "to what extent do proteins stretch or flex, unfold and refold, or undergo other conformational movements during their action?" To answer this and related questions, many techniques are being applied to study the dynamic properties of proteins.[308,309]

### 1. Motion of Backbone and Side Chains

Even in the *crystalline state* there is evidence of movement. In the images constructed from X-ray or neutron diffraction experiments side chains on the surfaces of protein molecules are often not clearly visible because of rapid rotational movement. Some segments of the polypeptide chain may be missing from the image. However, side chain groups within the core of a domain are usually seen clearly. They probably move only in discrete steps. However, they may sometimes shift rapidly between different conformations, all of which maintain a close-packed interior.[310–312]

Studies of nuclear magnetic resonance spectra (Chapter 3) and of polarization of fluorescence (Chapter 23), have shown that there is rapid though restricted rotational movement of side chains of proteins *in solution*. Even buried phenylalanine and tyrosine side chains often rotate rapidly whereas movement of the bulkier tryptophan rings is more limited. Peptide NH protons in unfolded polypeptide chains undergo rapid acid or base catalyzed exchange with $H_2O$.[313] However, in globular proteins, the rate of exchange of buried NH protons is often orders of magnitude slower.[314–316] For most proteins there appears to be little tendency to unfold completely and then refold; the major domains hold together tightly. However, there may be local unfolding, for example, of a helix at the surface of a protein, which will allow more rapid exchange. Using NMR techniques (Chapter 3), rates of exchange of all of the individual peptide NH protons within small proteins can be measured.[315] Cracks may open up in proteins. This is suggested by the fact that $O_2$, $I_2$, and certain other small molecules are apparently able to penetrate the protein freely and to quench the fluorescence of buried tryptophan side chains (Chapter 23).

Since packing density tends to be lower at active sites than in the bulk of the protein, it is probable that more conformational alterations occur near active sites than elsewhere.[317] Lumry and Rosenberg[318] suggested that the "defects" of poor packing and poor hydrogen bonding of some regions of a folded peptide chain provide a store of potential energy that can be important to the functioning of a protein. Even in a very well-packed protein domain there are defects. Some atoms are compressed by the folding of the peptide chain and are closer together by over 0.04 nm than predicted by the van der Waals radii. These packing defects have been estimated to destablize the protein by as much as 250 kJ/mol.[319] Details of structural heterogeneity within several proteins for which very precise structural data are available have been described.[311]

### 2. Conformational Changes

We have seen that some polypeptides assume an extended β conformation while others form helices. In some cases, the same protein can do both. For example, hair can be stretched greatly, the α helices of the keratin molecules uncoiling and assuming a β conformation with hydrogen bonds *between* chains instead of *within* a single chain. Thus, a polymer may have more than one conformational state in which the folding and hydrogen bonding are different.[320,321] With soluble proteins more than one folded conformation is possible with different sets of hydrogen bonds and internal hydrophobic interactions. Some of the conformations of a globular protein are more stable than others, and a protein will ordinarily assume one of the energetically most favorable conformations. However, there may be other conformations of almost equal energy.

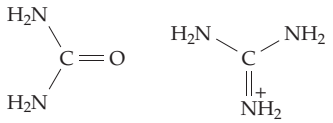
A large body of evidence suggests that many proteins do exist in two or more different but well-defined

conformational states and that the ability of a protein to undergo easy conversion from one to another is of great biological significance. In some cases such as that of hemoglobin (Chapter 7), there are changes in the interactions between subunits. Alterations in the hinge regions between domains of immunoglobulins have been seen. In many enzymes, including the dehydrogenases (Chapter 15), kinases (Chapter 12), and aspartate aminotransferase (Fig. 2-6), a cleft between two domains appears to open and close.[322] In other cases, more subtle alterations in conformational state, involving mainly changes in the internal hydrogen bonded network together with small localized changes in chain folding, have been observed.

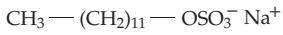
Conformational alterations in proteins are probably facilitated by the fact that some hydrogen-bonded groups are found within the hydrophobic interior. *All of the buried hydrogen atoms suitable for hydrogen bond formation are ordinarily hydrogen bonded to an electron donor group. However, because oxygen atoms in proteins each have two unshared electron pairs, there are, in general, more electron donor groups than there are hydrogen atoms to which they can bind. This sets the stage for a competition between electronegative centers for particular proton suitable for hydrogen bonding and provides a molecular basis for the easy triggering of conformational changes.*[323]

## 3. Denaturation and Refolding

An extreme conformational alteration is the **denaturation** of proteins, which may be caused by heating or by treatment with reagents such as strong acids and bases, **urea**, **guanidinium chloride**, and **sodium dodecyl sulfate** (SDS).




Urea Guanidinium chloride




Sodium dodecyl sulfate (SDS)

Denaturation leads to unfolding of a protein to a more random conformation. In the denatured state the amide groups of the peptide chain form hydrogen bonds with surrounding water molecules and with denaturants such as urea or the guanidinium ion rather than with each other.[324,325] Denaturants also diminish the strength of the hydrophobic interactions that promote folding.[326] Characteristic biochemical activities are lost and physical properties such as sedimentation constant, viscosity, and light absorption are altered. The ease of denaturation of proteins and the fact that denaturation is sometimes reversible show that the energy differences between the folded conformations and the open random coil conformation are usually not great.[274,327,328] However, it has been difficult to establish the amount of stabilization of a folded polypeptide chain provided by buried hydrogen bonds[328a,328b,328c] or the role of cooperatively formed hydrogen-bonded chains[328c,328d] or clusters.[328c,328d,328e]

Complete denaturation of a protein was generally regarded as an irreversible process prior to 1956 when Anfinsen showed that denatured ribonuclease (Chapter 12) could refold spontaneously.[273] This 124-residue protein contains four disulfide (–S–S–) bridges and thus is tied firmly together. When these bridges are broken reductively in the presence of a denaturing agent, the enzyme becomes inactive. Anfinsen found that upon reoxidation under appropriate conditions, complete activity reappeared. The molecules had folded spontaneously into the correct conformation, the one in which the correct *one* of 105 (7 x 5 x 3 x 1) possible pairings of the eight –SH groups present needed to reform the four disulfide bridges had taken place. This observation has had an important influence on thinking about protein synthesis and folding of polypeptide chains into biologically active molecules.

A puzzling problem was posed by Levinthal many years ago.[329] We usually assume that the peptide chain folds into one of the most stable conformations possible. However, proteins fold very rapidly. Even today, no computer would be able, in our lifetime, to find by systematic examination the thermodynamically most stable conformation.[328] It would likewise be impossible for a folding protein to "try out" more than a tiny fraction of all possible conformations. Yet folded and unfolded proteins often appear to be in a thermodynamic equilibrium! Experimental results indicate that denatured proteins are frequently in equilibrium with a **compact denatured state** or "molten globule" in which hydrophobic groups have become clustered and some secondary structures exists.[330–336] From this state the polypeptide may rearrange more slowly through other **folding intermediates** to the final "native conformation."[336a,336b]

It is generally assumed that within cells the folding of the peptide chain commences while the chain is still being synthesized on a ribosome. The growing chain probably folds rapidly in a random way until it finds one or more stable conformations which serve as folding intermediates for slower conversion to the finished protein.[337,338] However, any process within a cell is affected by the complex intracellular environment.[339] Folding can be catalyzed or inhibited by proteins known as **molecular chaperones**. Folding may sometimes require isomerization of one or a few proline residues from *trans* to *cis*.[340,341] For example, during the refolding of ribonuclease the isomerization of Pro 93 appears to be a rate-limiting step.[342,343] Such

isomerizations may be assisted by **peptidyl-prolyl-(*cis–trans*) isomerases** (Box 9-F). Disulfide linkages are sometimes formed incorrectly.[344,345] A **protein-disulfide isomerase** catalyzes cleavage and reformation of these bridges, allowing the protein to find the most stable crosslinking arrangement. The actions of these enzymes and of molecular chaperones are considered further in Chapters 10 and 12.

## 4. Effects of pH and Solvent

Because polypeptide chains contain many acidic and basic amino acid side chains, the properties of proteins are greatly influenced by pH. At low pH carboxylates, $–S^–$, and imidazole groups accept protons causing the overall net charge on the macromolecule to be strongly positive. At high pH protons are lost and the protein becomes negatively charged. Electrostatic repulsion between like charges may cause proteins to denature at low or high pH. More stable proteins may be very soluble at low or high pH. Proteins often have a minimum solubility and a maximum stability near the **isoelectric point,** the pH at which the net charge is zero.[336,346,347] Activities of enzymes, abilities to bind specifically to other proteins, and

## BOX 2-C THE NOBEL PRIZES

Many young scientists dream of one day winning a Nobel Prize. Although denounced by some, the much sought and highly publicized award has, since 1901, been given to an outstanding group of scientists. Many of these have made major contributions to biochemistry or to techniques important to biochemists. Here is a partial list.

| Year | Name | Prize[a] | Discovery or subject of study |
|---|---|---|---|
| 1901 | Wilhelm C. Röntgen | Physics | Discovery of X-rays |
| 1902 | Emil H. Fischer | Chemistry | Synthesis of sugars and purines |
| 1903 | Svante A. Arrhenius | Chemistry | Electrolytic dissociation; a founder of physical chemistry |
| 1903 | Antoine Henri Becquerel; Marie S. Curie, and Pierre Curie | Physics | Discovery and study of radioactivity |
| 1906 | Camillo Golgi and S. Ramon y Cajal | Physiology/Medicine | Discovery of Golgi apparatus |
| 1907 | Edward Buchner | Chemistry | Biochemistry, cell-free fermentation |
| 1910 | Albrecht Kossel | Physiology/Medicine | Isolation of nucleic acid bases |
| 1914 | Max von Laue | Physics | Discovery of X-ray diffraction by crystals |
| 1915 | Richard M. Willstätter | Chemistry | Plant pigments, especially chlorophyll |
| 1915 | William H. and William L. Bragg | Physics | Analysis of crystal structure by X-rays |
| 1919 | Jules Bordet | Physiology/Medicine | Discovery of blood complement; complement fixation test |
| 1920 | Walther H. Nernst | Chemistry | Thermochemistry |
| 1922 | Archibald V. Hill and Otto F. Meyerhof | Physiology/Medicine | Chemistry of muscle contraction |
| 1923 | Frederick G. Banting and John J. R. Macleod | Physiology/Medicine | Discovery of insulin and treatment of diabetes |
| 1926 | Theodor Svedburg | Chemistry | Study of high $M_r$ compounds, development of ultracentrifuge |
| 1927 | Heinrich O. Wieland | Chemistry | Bile acids |
| 1928 | Adolf O. R. Windaus | Chemistry | Sterols and vitamins |
| 1929 | Frederick G. Hopkins and Christiaan Eijkman | Physiology/Medicine | Discovery of vitamins, tryptophan, vitamin $B_1$ |
| 1929 | Arthur Harden and Hans A. S. von Euler-Chelpin | Chemistry | Fermenting enzymes, fermentation of sugars |
| 1930 | Karl Landsteiner | Physiology/Medicine | Blood groups A, B, O, Rh |
| 1930 | Hans Fischer | Chemistry | Structures and chemistry of porphyrins, chlorophyll |
| 1931 | Otto H. Warburg | Physiology/Medicine | Respiratory enzymes |
| 1933 | Thomas H. Morgan | Physiology/Medicine | Chromosome theory and chromosome maps |
| 1934 | George R. Minot, William P. Murphy, and George H. Whipple | Physiology/Medicine | Treatment of pernicious anemia |
| 1936 | Henry H. Dale and Otto Loewi | Physiology/Medicine | Acetylcholine release at nerve endings |
| 1937 | Albert von Szent-Györgyi | Physiology/Medicine | Vitamin C |
| 1937 | Walter N. Haworth and Paul Karrer | Chemistry | Carbohydrate structures, structures of carotenoids, flavins, vitamin A |
| 1938 | Richard Kuhn | Chemistry | Carotenoids and vitamins |
| 1939 | Gerhard Domagk | Physiology/Medicine | Prontosil, first antibacterial sulfa drug |
| 1939 | Adolf F. J. Butenandt and L. Ruzicka | Chemistry | Isolation and study of sex hormones, study of polymethylenes, terpenes |
| 1943 | E. A. Doisy and Carl Henrik Dam | Physiology/Medicine | Isolation and study of vitamin K |
| 1945 | Alexander Fleming, Ernst B. Chain, and Howard W. Florey | Physiology/Medicine | Discovery and structure of penicillin |
| 1945 | A. J. Virtanen | Chemistry | Nutritional chemistry |
| 1946 | James B. Sumner, J. H. Northrop, and W. M. Stanley | Chemistry | Crystallization of enzymes and virus proteins |
| 1947 | Carl F. Cori and Gerty T. Cori | Physiology/Medicine | Glycogen metabolism, the Cori cycle |
| 1947 | B. A. Houssay and Robert Robinson | Chemistry | Investigation of plant alkaloids |
| 1948 | Arne W. K. Tiselius | Chemistry | Electrophoresis, study of serum proteins |
| 1950 | Phillip S. Hench, Edward C. Kendall, and Tadeus Reichstein | Physiology/Medicine | ACTH |
| 1952 | Selman A. Waksman | Physiology/Medicine | Discovery of streptomycin |
| 1952 | A. J. P. Martin and Richard L. M. Synge | Chemistry | Paper chromatography |
| 1953 | Fritz A. Lipmann and Hans A. Krebs | Physiology/Medicine | Coenzyme A, citric acid cycle |
| 1954 | Linus C. Pauling[b] | Chemistry | The nature of the chemical bond |
| 1955 | A. H. T. Theorell | Physiology/Medicine | Oxidative enzymes |
| 1955 | Vincent du Vigneaud | Chemistry | Synthesis of biotin and oxytocin |
| 1957 | Daniel Bovet | Physiology/Medicine | First antihistamine drug |
| 1957 | Alexander R. Todd | Chemistry | Work on nucleotides coenzymes |
| 1958 | Joshua Lederberg, George W. Beadle, and Edward L. Tatum | Physiology/Medicine | One gene-one enzyme hypothesis from genetic studies of neurospora |
| 1958 | Frederick Sanger[b] | Chemistry | Protein sequencing, insulin |

numerous physiological processes are controlled by effects of pH on proteins.

The solvent for most proteins in nature is water. However, many enzymes function well in organic solvents if they retain only a small amount of essential structural and catalytic water.[348]

## 5. Irreversible Damage to Proteins

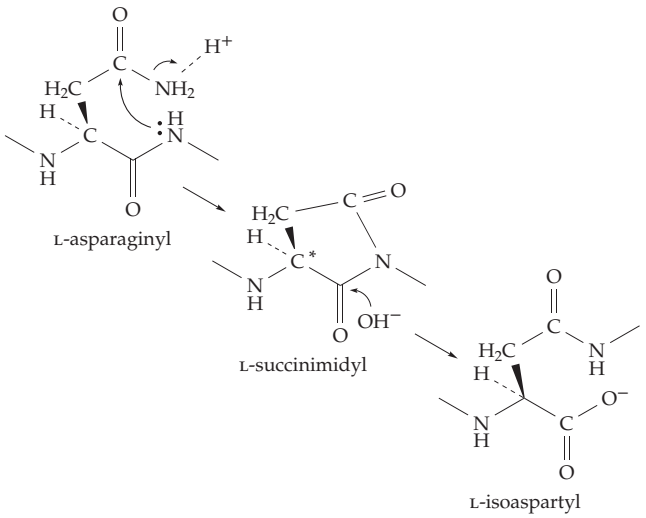
Every protein in an organism has its own characteristic lifetime. No sooner is it synthesized than degradation begins. Enzyme-catalyzed cleavage of the peptide linkages leads to turnover of proteins but before this occurs a number of spontaneous damaging reactions may alter the protein. Prevalent among these is **deamidation** of asparaginyl residues to aspartyl or

### BOX 2-C THE NOBEL PRIZES (continued)

| Year | Name | Prize[a] | Discovery or subject of study |
|---|---|---|---|
| 1959 | Arthur Kornberg and Severo Ochoa | Physiology/Medicine | Enzymatic synthesis of DNA |
| 1960 | F. M. Burnet and P. Medawar | Physiology/Medicine | Immunological tolerance in animals |
| 1961 | Melvin Calvin | Chemistry | Photosynthesis using $^{14}CO_2$ |
| 1962 | Francis H. C. Crick, James D. Watson, and Maurice H. F. Wilkins | Physiology/Medicine | Molecular structure of DNA |
| 1962 | John C. Kendrew and Max F. Perutz | Chemistry | Structures of protein by X-ray diffraction |
| 1962 | Linus C. Pauling[b] | Peace | Ending atmospheric testing of nuclear weapons |
| 1964 | Konrod E. Bloch and Feodor Lynen | Physiology/Medicine | Pathways of cholesterol biosynthesis |
| 1964 | Dorothy M. Crowfoot Hodgkin | Chemistry | X-ray structures, vitamin $B_{12}$ |
| 1965 | Andre Lwolf, Jacques Monod, and Francois Jacob | Physiology/Medicine | Messenger RNA, regulation of transcription |
| 1967 | Ragnar Granit, Haldan Keffer Hartline, and George Wald | Physiology/Medicine | Chemistry of vision |
| 1967 | Manfred Eigen and Norrish Porter | Chemistry | Study of high-speed chemical reactions |
| 1968 | Robert W. Holley | Physiology/Medicine | RNA sequence |
| 1968 | H. Gobind Khorana and Marshall W. Nirenberg | Physiology/Medicine | The genetic code |
| 1969 | Max Delbrück, Alfred D. Hershey, and Salvador E. Luria | Physiology/Medicine | Replication and genetic structures of viruses |
| 1970 | Julius Axelrod, Bernard Katz, and Ulf von Euler | Physiology/Medicine | Transmission of nerve impulses, noradrenaline |
| 1970 | Luis F. Leloir | Chemistry | Role of sugar nucleotides in biosynthesis of carbohydrates |
| 1971 | Earl W. Sutherland, Jr. | Physiology/Medicine | Cyclic AMP |
| 1972 | Gerald M. Edelman and Rodney R. Porter | Physiology/Medicine | Structure of antibodies |
| 1972 | Christian B. Anfinsen, Sanford Moore, and William H. Stein | Chemistry | Ribonuclease, structure and activity |
| 1974 | Albert Claude, Christian R. de Duve, and George E. Palade | Physiology/Medicine | Cell structure |
| 1975 | David Baltimore, Renato Dulbecco, and Howard M. Temin | Physiology/Medicine | Reverse transscriptase |
| 1975 | John W. Cornforth and Vladimir Prelog | Chemistry | Stereochemistry of organic molecules and of enzymatic reactions |
| 1977 | Rosalyn S. Yalow, Roger C. L. Guillemin, and Andrew V. Schally | Physiology/Medicine | Radioimmunoassay thyrotropin-releasing hormone |
| 1978 | Daniel Nathans, Werner Arber, and Hamilton O. Smith | Physiology/Medicine | Restriction enzymes |
| 1978 | Peter Mitchell | Chemistry | Biological energy transfer |
| 1980 | Paul Berg, Walter Gilbert, and Frederick Sanger[b] | Chemistry | Recombinant DNA, methods of sequence determination for DNA |
| 1982 | Sune K. Bergstrom, Bergt I. Samuelsson, and John R. Vane | Physiology/Medicine | Isolation and study of prostaglandins |
| 1982 | Aaron Klug | Chemistry | Development of crystallographic electron microscopy |
| 1983 | Barbara McClintock | Physiology/Medicine | Gene transposition |
| 1984 | Niels K. Jerne, Georges J. F. Köhler, and Cesar Milstein | Physiology/Medicine | Cellular basis of immunology |
| 1984 | R. Bruce Merrifield | Chemistry | Solid-phase synthesis of peptides |
| 1985 | Joseph L. Goldstein and Michael S. Brown | Physiology/Medicine | Control of cholesterol metabolism |
| 1986 | Rita Levi-Montalcini and Stanley Cohen | Physiology/Medicine | Nerve growth factor |
| 1987 | Susumu Tonegawa | Physiology/Medicine | Genetics of antibody formation |
| 1988 | Johan Diesenhofer, Robert Huber, and Hartmut Michel | Chemistry | Three-dimensional structure of a photosynthetic reaction center |
| 1988 | Gertrude Elion, George Hitchings, and James Black | Physiology/Medicine | Principles of drug treatment and design of many important drugs |
| 1989 | J. Michael Bishop, Harold E. Varmus, and Joseph E. Murray | Physiology/Medicine | Origin of retroviral oncogenes |
| 1989 | Sidney Altman and Thomas R. Cech | Chemistry | Catalytic RNA |
| 1991 | Erwin Neher and Bert Sakmann | Physiology/Medicine | Functioning of single ion channels in cells |
| 1991 | Richard R. Ernst | Chemistry | High-resolution NMR |
| 1992 | Edmond H. Fischer and Edwin G. Krebs | Physiology/Medicine | Reversible protein phosphorylation in biological regulation |
| 1992 | Rudolph A. Marcus | Chemistry | Theory of electron transfer reactions |
| 1993 | Richard J. Roberts and Phillip A. Sharp | Physiology/Medicine | Discovery of split genes |
| 1993 | Michael Smith and Kary B. Mullis | Chemistry | Oligonucleotide-directed mutagenesis and polymerase chain reaction |
| 1994 | Alfred Gilman and Martin Rodbell | Physiology/Medicine | "G-proteins" |
| 1995 | Edwin B. Lewis, Christiane Nusslein-Volhard, and Eric Wieschaus | Physiology/Medicine | Homeotic mutations in *Drosophila* |
| 1995 | Paul Crutzen, Sherwood Rowland, and Mario Molina | Chemistry | Damage to the stratospheric ozone layer |
| 1996 | Peter C. Doherty and Rolf M. Zinkernagel | Physiology/Medicine | Specificity of cell-mediated immune response |
| 1997 | Paul D. Boyer and John E. Walker | Chemistry | Mechanism of ATP synthesis |
| 1997 | Jens C. Skou | Chemistry | Discovery of $Na^+$, $K^+$-ATPase |
| 1997 | Stanley B. Prusiner | Physiology/Medicine | Discovery of prions |
| 1998 | Robert F. Furchgott, Louis J. Ignarro, and Ferid Murad | Physiology/Medicine | Nitric oxide as a signaling molecule |
| 1999 | Günter Blobel | Physiology/Medicine | Intrinsic signals that govern transport and localization of proteins |
| 2000 | Arvid Carlsson | Physiology/Medicine | Identification of dopamine as signaling molecule in brain |
| | Paul Greengard | Physiology/Medicine | Discovery of the dopamine signaling cascade |
| | Eric R. Kandel | Physiology/Medicine | Molecular basis of learning |

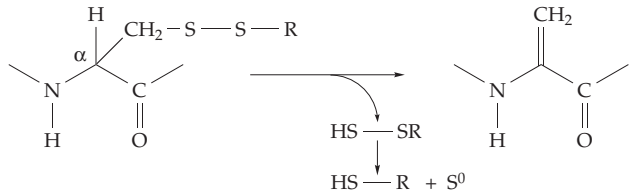
[a] Prizes are given in Physics, Chemistry, Physiology or Medicine, Literature and Peace.
[b] Sanger and Pauling each have been awarded two Nobel Prizes.

**isoaspartyl** groups (Eq. 2-24).[349,350] Aspartyl residues can undergo the same type of cyclization at low pH. The sequence Asn-Gly is especially susceptible to rearrangement according to Eq. 2-24. However, the peptide torsion angles may be more important in determining whether deamidation occurs. The intermediate succinimide may be racemized easily at the chiral center marked by the asterisk in Eq. 2-24. Thus, the Asn-Gly sequence represents a weak linkage, which is nevertheless present in many proteins.




(2-24)

Cystine residues in disulfide bridges are also not completely stable but undergo β elimination[305,351] (Chapter 13) at slightly alkaline pH values according to Eq. 2-25. The free thiol group formed (HSR) is still attached to the protein but may sometimes be in a position to enter into thiol-disulfide exchange reactions with other S–S bridges causing further degradation. Methionine side chains in proteins can be oxidized to sulfoxides: $-CH_2-CH_2-\overset{\overset{\displaystyle O}{\|}}{S}-CH_3$ and hydroxyl groups can be introduced into aromatic rings by oxidation. Hydrogen peroxide and other oxidants may be responsible for such oxidation within cells (see Chapter 18). The very long-lived proteins of the lens of the eye are especially susceptible to deamidation, racemization, oxidation, and accumulation of covalently attached blue fluorescent compounds.[352–354]



(2-25)

## H. Design and Engineering of Proteins

Methods of chemical synthesis of polypeptides and of cloning and mutating genes now allow us to alter peptide sequences at will and to design completely new proteins.[355–356b] The methods are discussed in Chapters 3, 5, and 26. The following are examples.

Peptides that form α helices that associate as coiled coils,[357] or as three- or four-helix tetrameric bundles[179a,358–360] or amphipathic helices that associate with lipid bilayers have been made.[355,361] More difficult has been the design of proteins that form β sheets.[362–364a] Many efforts are being made to understand protein stability[365–367] by systematic substitutions of one residue for another. Addition of new disulfide linkages at positions selected by study of three-dimensional structures sometimes stabilizes enzymes.[368–371] On the other hand elimination of unnecessary cysteine residues can enhance stability by preventing β elimination[351] and replacement of asparagine by threonine can improve the thermostability of enzymes by preventing deamidation.[372,373]

Artificial mutants of subtilisin[373] and of many other enzymes are helping us to understand the mechanisms of catalysis. Artificially prepared temperature-sensitive mutants (Chapter 26) and naturally occurring mutants are providing new insight into pathways of folding of proteins whose three-dimensional structures are known. For these studies it is necessary to follow a rational strategy in deciding which of the astronomical number of possible mutant forms may be of interest. For example, residues within active sites may be changed. Introduction of mutations at random locations can also be useful in finding regions of interest.[374] When mutant proteins can be crystallized, an exact understanding of the effect of the mutation on the structure is possible. Sometimes two or more mutations may have to be introduced to obtain the desired modification.

It is also possible to incorporate a range of "unnatural" amino acids in specific sites in a polypeptide sequence and to observe resulting effects on a protein's properties.[375] A quite different approach is to design polypeptides that mimic a natural peptide but consist of D-amino acids. The peptide chain is *reversed*, i.e., the N terminus becomes the C terminus and every peptide linkage is also reversed. The amino acid side chains preserve their relationships one to another and the backbone atoms tend to preserve their hydrogen bonding pattern. Peptides made in this way tend to be resistant to cleavage by enzymes. Some may be useful as drugs.[376,377]

## References


1. Glendening, E. D., Faust, R., Streitwieser, A., Vollhardt, K. P. C., and Weinhold, F. (1993) *J. Am. Chem. Soc.* **115**, 10952–10957
2. Gao, Q., Weber, H.-P., Craven, B. M., and McMullan, R. K. (1994) *Acta Crystallogr.* **B50**, 695–703
3. Hegstrom, R. A., and Kondepudi, D. K. (1990) *Sci. Am.* **262**(Jan), 108–115
4. Cahn, R. S., Ingold, C., and Prelog, V. (1966) *Angew. Chem. Int. Ed. Engl.* **5**, 385–415
5. Eliel, E. L., Wilen, S. H., and Mander, L. N. (1994) *Stereochemistry of Organic Compounds*, Wiley, New York
6. Bentley, R. (1969) *Molecular Asymmetry in Biology*, Vol. 1, Academic Press, New York (pp. 49–56)
7. Blackwood, J. E., Gladys, C. L., Loening, K. L., Petrarca, A. E., and Rush, J. E. (1968) *J. Am. Chem. Soc.* **90**, 509–510
8. Hanson, K. R. (1966) *J. Am. Chem. Soc.* **88**, 2731–2742
9. Lambert, J. B. (1970) *Sci. Am.* **222**(Jan), 58–70
10. Klyne, W., and Prelog, V. (1960) *Experientia* **16**, 521–523
11. Liébecq, C., ed. (1992) *Biochemical Nomenclature*, Portland Press, London and Chapel Hill, North Carolina (for the International Union of Biochemistry and Molecular Biology)
12. Chiang, Y., Kresge, A. J., and Tang, Y. S. (1984) *J. Am. Chem. Soc.* **106**, 460–462
13. Beak, P., Fry, F. S., Jr., Lee, J., and Steele, F. (1976) *J. Am. Chem. Soc.* **98**, 171–179
14. Beak, P. (1977) *Acc. Chem. Res.* **10**, 186–192
15. Metzler, D. E., Harris, C. M., Johnson, R. J., Siano, D. B., and Thomson, J. A. (1973) *Biochemistry* **12**, 5377–5392
16. Katritzky, A. R., and Karelson, M. (1991) *J. Am. Chem. Soc.* **113**, 1561–1566
17. Blomberg, F., Maurer, W., and Rüterjans, H. (1977) *J. Am. Chem. Soc.* **99**, 8149–8159
18. Walters, D. E., and Allerhand, A. (1980) *J. Biol. Chem.* **255**, 6200–6204
19. Fersht, A. (1977) *Enzyme Structure and Mechanism*, Freeman, San Francisco, California
20. Gabler, R. (1978) *Electrical Interactions in Molecular Biophysics*, Academic Press, New York (p. 188)
21. Israelachvili, J. N., and McGuiggan, P. M. (1988) *Science* **241**, 795–800
22. Salem, L. (1962) *Can. J. Biochem. Physiol.* **40**, 1287–1298
23. Tsao, Y.-H., Fennell Evans, D., and Wennerström, H. (1993) *Science* **262**, 547–550

23a. Kurihara, K., and Kunitake, T. (1992) *J. Am. Chem. Soc.* **114**, 10927–10933

23b. Yang, L., Valdeavella, C. V., Blatt, H. D., and Pettitt, B. M. (1996) *Biophys. J.* **71**, 3022–3029

24. Warshel, A. (1987) *Nature (London)* **330**, 15–16
25. Honig, B., and Nicholls, A. (1995) *Science* **268**, 1144–1149
26. Meot-Ner, M. M. (1988) *J. Am. Chem. Soc.* **110**, 3075–3080
27. Loewenthal, R., Sancho, J., Reinikainen, T., and Fersht, A. R. (1993) *J. Mol. Biol.* **232**, 574–583
28. Gristina, A. G., Oga, M., Webb, L. X., and Hobgood, C. D. (1985) *Science* **228**, 990–993
29. Jeffrey, G. A., and Saenger, W. (1991) *Hydrogen Bonding in Biological Structures*, Springer-Verlag, Berlin
30. Gilli, P., Bertolasi, V., Ferretti, V., and Gastone, G. (1994) *J. Am. Chem. Soc.* **116**, 909–915

30a. Jeffrey, G. A. (1997) *An Introduction to Hydrogen Bonding*, Oxford Univ. Press, New York

31. Derewenda, Z. S., Derewenda, U., and Kobos, P. M. (1994) *J. Mol. Biol.* **241**, 83–93

31a. Wahl, M. C., and Sundaralingam, M. (1997) *Trends Biochem. Sci.* **22**, 97–102

31b. Bella, J., and Berman, H. M. (1996) *J. Mol. Biol.* **264**, 734–742

32. Schulz, G. E., and Schirmer, R. H. (1979) *Principles of Protein Structure*, Springer, New York (p. 36)

32a. Platts, J. A., Howard, S. T., and Bracke, B. R. F. (1996) *J. Am. Chem. Soc.* **118**, 2726–2733

33. Taylor, R., Kennard, O., and Versichel, W. (1983) *J. Am. Chem. Soc.* **105**, 5761–5766
34. Murray-Rust, P., and Glusker, J. P. (1984) *J. Am. Chem. Soc.* **106**, 1018–1025
35. Görbitz, C. H., and Etter, M. C. (1992) *J. Am. Chem. Soc.* **114**, 627–631
36. Nelson, D. D., Jr., Fraser, G. T., and Klemperer, W. (1987) *Science* **238**, 1670–1674

36a. Ghanty, T. K., Staroverov, V. N., Koren, P. R., and Davidson, E. R. (2000) *J. Am. Chem. Soc.* **122**, 1210–1214

36b. Cornilescu, G., Ramirez, B. E., Frank, M. K., Clore, G. M., Gronenborn, A. M., and Bax, A. (1999) *J. Am. Chem. Soc.* **121**, 6275–6279

36c. Benedict, H., Shenderovich, I. G., Malkina, O. L., Malkin, V. G., Denisov, G. S., Golubev, N. S., and Limbach, H.-H. (2000) *J. Am. Chem. Soc.* **122**, 1979–1988

37. Fersht, A. R., Shi, J., Knill-Jones, J., Lowe, D. M., Witkinson, A. J., Blow, D. M., Brick, P., Carter, P., Woye, M. M. Y., and Winter, G. (1985) *Nature (London)* **314**, 235–238
38. Eriksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) *Science* **255**, 178–183

38a. Takano, K., Yamagata, Y., Kubota, M., Funahashi, J., Fujii, S., and Yutani, K. (1999) *Biochemistry* **38**, 6623–6629

39. Lim, W. A., Hodel, A., Sauer, R. T., and Richards, F. M. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 423–427
40. Schuster, P., Zundel, G., and Sandorfy, C. (1976) *The Hydrogen Bond*, North Holland Publ. Co., Amsterdam (3 vols.)
41. Luck, W. A. P., and Kleeberg, H. (1978) in *Photosynthetic Oxygen Evolution* (Metzner, H., ed), pp. 1–29, Academic Press, London
42. Edsall, J. T. (1983) *Trends Biochem. Sci.* **8**, 29–31
43. Grunwald, E. (1986) *J. Am. Chem. Soc.* **108**, 5719–5726
44. Westhof, E. (1993) *Water and Biological Macromolecules*, CRC Press, Boca Raton, Florida
45. Kusalik, P. G., and Svishchev, I. M. (1994) *Science* **265**, 1219–1221
46. Libnau, F. O., Toft, J., Christy, A. A., and Kvalheim, O. M. (1994) *J. Am. Chem. Soc.* **116**, 8311–8316
47. Benson, S. W., and Siebert, E. D. (1992) *J. Am. Chem. Soc.* **114**, 4269–4276
48. Liu, K., Cruzan, J. D., and Saykally, R. J. (1996) *Science* **271**, 929–933
49. Pauling, L. (1935) *J. Am. Chem. Soc.* **57**, 2680
50. Wittebort, R. J., Usha, M. G., Ruben, D. J., Wemmer, D. E., and Pines, A. (1988) *J. Am. Chem. Soc.* **110**, 5668–5671
51. Li, J., and Ross, D. K. (1993) *Nature (London)* **365**, 327–329
52. Dyke, T. R., Mack, K. M., and Muenter, J. S. (1977) *J. Chem. Phys.* **66**, 498–510
53. Saenger, W. (1979) *Nature (London)* **279, 280**, 343–344, 848
54. Saenger, W. (1987) *Ann. Rev. Biophys. Biophys. Chem.* **16**, 93–114
55. Pribble, R. N., and Zwier, T. S. (1994) *Science* **265**, 75–79
56. Gregory, J. K., Clary, D. C., Liu, K., Brown, M. G., and Saykally, R. J. (1997) *Science* **275**, 814–817
57. Quiocho, F. A., Sack, J. S., and Vyas, N. K. (1987) *Nature (London)* **329**, 561–564
58. Pflugrath, J. W., and Quiocho, F. A. (1988) *J. Mol. Biol.* **200**, 163–180
59. Kauzmann, W. (1959) *Adv. Prot. Chem.* **14**, 1–63
60. Jencks, W. P. (1987) *Catalysis in Chemistry and Enzymology*, Dover, Mineola, New York (p. 393)
61. Tanford, C. (1980) *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2nd ed., Wiley, New York
62. Abraham, M. H. (1980) *J. Am. Chem. Soc.* **102**, 5910–5912
63. Ben-Náim, A. (1980) *Hydrophobic Interactions*, Plenum, New York
64. Murphy, K. P., Privalov, P. L., and Gill, S. J. (1990) *Science* **247**, 559–561
65. Muller, N. (1992) *Trends Biochem. Sci.* **17**, 459–463
66. Spolar, R. S., Livingstone, J. R., and Record, J., MT. (1992) *Biochemistry* **31**, 3947–3955
67. Stites, W. E., Meeker, A. K., and Shortle, D. (1994) *J. Mol. Biol.* **235**, 27–32
68. Hecht, D., Tadesse, L., and Walters, L. (1993) *J. Am. Chem. Soc.* **115**, 3336–3337
69. Gill, S. J., and Wadsö, I. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2955–2958
70. Sharp, K. A., Nicholls, A., Fine, R. F., and Honig, B. (1991) *Science* **252**, 106–109
71. Van Oss, C. J. (1994) *Interfacial Forces in Aqueous Media*, Dekker, New York

71a. Hummer, G., Garde, S., García, A. E., Pohorille, A., and Pratt, L. R. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8951–8955

72. Israelachvili, J., and Wennerström, H. (1996) *Nature (London)* **379**, 219–225
73. Radzicka, A., and Wolfenden, R. (1988) *Biochemistry* **27**, 1664–1670
74. Kellis, J. T., Jr., Nyberg, K., Sali, D., and Fersht, A. R. (1988) *Nature (London)* **333**, 784–786
75. Sanger, F. (1988) *Ann. Rev. Biochem.* **57**, 1–28
76. Creighton, T. E. (1993) *Proteins, Structures and Molecular Principles*, 2nd ed., Freeman, New York
77. Kyte, J. (1995) *Structure in Protein Chemistry*, Garland Publ., New York
78. Guillemin, R., and Burgus, R. (1972) *Sci. Am.* **227**(Nov), 24–33
79. Hochster, R. M., Kates, M., and Quastel, J. H., eds. (1973) *Metabolic Inhibitors*, Vol. 3, Academic Press, New York (p. 312)
80. Ramachandran, G. N., and Kolaskar, A. S. (1973) *Biochim. Biophys. Acta.* **303**, 385–388
81. MacArthur, M. W., and Thornton, J. M. (1996) *J. Mol. Biol.* **264**, 1180-1195

81a. Milner-White, E. J. (1997) *Protein Sci.* **6**, 2477–2482

82. Moffat, J. B. (1973) *J. Theor. Biol.* **40**, 247–258
83. Fersht, A. R. (1971) *J. Am. Chem. Soc.* **93**, 3504–3515
84. Karle, I. L. (1974) *Biochemistry* **13**, 2155–2162
85. Ovchinnikov, Y. A. (1973) *FEBS Lett.* **29**, 31–33
86. Doonan, S. (1974) *FEBS Lett.* **38**, 229–233
87. Bhown, A. S. (1990) *Handbook of Proteins*, A & M Publ., Birmingham, Alabama
88. Boguski, M., and McEntyre, J. (1994) *Trends Biochem. Sci.* **19**, 71

88a. Branden, C., and Tooze, J. (1991) *Introduction to Protein Structure*, Garland Publ., New York

89. Clements, S., Mehansho, H., and Carlson, D. M. (1975) *J. Biol. Chem.* **260**, 13471–13477
90. Condit, C. M., and Meagher, R. B. (1986) *Nature (London)* **323**, 178–181
91. Reynolds, P., Weber, S., and Prakash, L. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 168–172
92. Rodakis, G. C., and Kafatos, F. C. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3551–3555
93. Young, J. F., Hockmeyer, W. T., Gross, M., Ballou, W. R., Wirtz, R. A., Trosper, J. H., Beaudoin, R. L., Hollingdale, M. R., Miller, L. H., Diggs, C. L., and Rosenberg, M. (1985) *Science* **228**, 958–962

## References


94. Doolittle, R. F. (1987) *Of URFS and ORFS: A Primer on How to Analyze Derived Amino Acid Sequences*, Univ. Science Books, Mill Valley, California
95. Doolittle, R. F. (1989) *Trends Biochem. Sci.* **14**, 244–245
96. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444–2448
97. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233**, 123–138
98. Lewin, R. (1987) *Science* **237**, 1570
99. Fleming, P. J., and Richards, F. M. (2000) *J. Mol. Biol.* **299**, 487–498
100. Gekko, K., and Hasegawa, Y. (1986) *Biochemistry* **25**, 6563–6571
101. Goodsell, D. S., and Olson, A. J. (1993) *Trends Biochem. Sci.* **18**, 65–68
102. Schultz, G. E., and Schirmer, R. H. (1979) *Principles of Protein Structure*, Springer-Verlag, New York
102a. Ramachandran, G. N., Ramakrishnam, C., and Sasisekharan, V. (1963) *J. Mol. Biol.*, **7**, 95–99
103. Ramachandran, G. N., Venkatachalam, C. M., and Krimm, S. (1966) *Biophys. J.* **6**, 849–872
103a. Karplus, P. A. (1996) *Protein Sci.* **5**, 1406–1420
104. Sasisekharian, V., Lakshminarayanan, A. V., and Ramachandran, G. N. (1967) in *Conformation of Biopolymers*, Vol. 2 (Ramachandran, G. N., ed), p. 641–654, Academic Press, New York
104a. Gunasekaran, K., Ramakrishnan, C., and Balaram, P. (1996) *J. Mol. Biol.* **264**, 191–198
104b. Herzberg, O., and Moult, J. (1991) *Proteins* **11**, 223–229
105. Milner-White, E. J., Bell, L. H., and Maccallum, P. H. (1992) *J. Mol. Biol.* **228**, 725–734
106. Rhee, S. (1990) *High resolution X-ray diffraction studies of cytosolic aspartate aminotransferase* [Ph.D. Dissertation], University of Iowa, Iowa City, Iowa
107. Dunbrack, R. L., Jr., and Karplus, M. (1993) *J. Mol. Biol.* **230**, 543–574
108. Pauling, L., and Corey, R. B. (1951) *Proc. Natl. Acad. Sci. U.S.A.* **37**, 729–740
109. Rich, A. (1994) *Nature (London)* **371**, 285
110. Chothia, C. (1973) *J. Mol. Biol.* **75**, 295–302
111. Chou, K.-C., Némethy, G., and Scheraga, H. A. (1983) *Biochemistry* **22**, 6213–6221
111a. Skarzynski, T., Moody, P. C. E., and Wonacott, A. J. (1987) *J. Mol. Biol.* **193**, 171–187
111b. Wang, L., O'Connell, T., Tropsha, A., and Hermans, J. (1996) *J. Mol. Biol.* **262**, 283–293
112. Reeke, G. N., Jr., and Becker, J. W. (1986) *Science* **234**, 1108–1111
113. Murzin, A. G., Lesk, A. M., and Chothia, C. (1994) *J. Mol. Biol.* **236**, 1369–1381
114. Murzin, A. G., Lesk, A. M., and Chothia, C. (1994) *J. Mol. Biol.* **236**, 1382–1400
115. LaLonde, J. M., Bernlohr, D. A., and Ranaszak, L. J. (1994) *FASEB J.* **8**, 1240–1241
116. Guss, J. M., and Freeman, H. C. (1983) *J. Mol. Biol.* **169**, 521–563
117. Richardson, J. S. (1981) *Adv. Prot. Chem.* **34**, 167–339
117a. Erickson, H. P. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10114–10118
118. Redinbo, M. R., Cascio, D., Choukair, M. K., Rice, D., Merchant, S., and Yeates, T. O. (1993) *Biochemistry* **32**, 10560–10567
119. Weiss, M. S., and Schulz, G. E. (1993) *J. Mol. Biol.* **231**, 817–824
120. Kreusch, A., and Schulz, G. E. (1994) *J. Mol. Biol.* **243**, 891–905
121. Yoder, M. D., and Jurnak, F. (1995) *FASEB J.* **9**, 335–342
122. Richardson, J. S. (1994) *FASEB J.* **8**, 1237–1239
123. Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R., and Reinemer, P. (1994) *Science* **265**, 383–385
124. Richardson, J. S., Getzoff, E. D., and Richardson, D. C. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2574–2578
125. Chan, A. W. E., Hutchinson, E. G., Harris, D., and Thornton, J. M. (1993) *Protein Sci.* **2**, 1574–1590
126. Urry, D. W. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1610–1614
127. Nicholson, L. K., and Cross, T. A. (1989) *Biochemistry* **28**, 9379–9385
127a. Smith, T. F., Gaitatzes, C., Saxena, K., and Neer, E. J. (1999) *Trends Biochem. Sci.* **24**, 181–185
127b. Garcia-Higuera, I., Fenoglio, J., Li, Y., Lewis, C., Panchenko, M. P., Reiner, O., Smith, T. F., and Neer, E. J. (1996) *Biochemistry* **35**, 13985–13994
127c. ter Naar, E., Harrison, S. C., and Kirchhausen, T. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1096–1100
127d. Springer, T. A. (1998) *J. Mol. Biol.* **283**, 837–862
128. Pauling, L. (1993) *Protein Sci.* **2**, 1060–1063
129. Kendrew, J. C. (1969) *Sci. Am.* **205**(Dec), 96–110
130. Arnott, S., and Dover, S. D. (1967) *J. Mol. Biol.* **30**, 209–212
131. Steigemann, W., and Weber, E. (1979) *J. Mol. Biol.* **127**, 309–338
132. Barlow, D. J., and Thornton, J. M. (1988) *J. Mol. Biol.* **201**, 601–619
132a. Rajashankar, K. R., and Ramakumar, S. (1996) *Protein Sci.* **5**, 932–946
133. Harper, E. T., and Rose, G. D. (1993) *Biochemistry* **32**, 7605–7609
134. Chakrabartty, A., Schellman, J. A., and Baldwin, R. L. (1991) *Nature (London)* **351**, 586–588
135. Creamer, T. P., and Rose, G. D. (1995) *Protein Sci. U.S.A.* **4**, 1305–1314
136. Dempsey, C. E. (1992) *Biochemistry* **31**, 4705–4712
137. Ramachandran, G. N., and Mitra, A. K. (1976) *J. Mol. Biol.* **107**, 85–92
138. Blaber, M., Zhang, X.-j, and Matthews, B. W. (1993) *Science* **260**, 1637–1639
139. Chakrabartty, A., Kortemme, T., and Baldwin, R. L. (1994) *Protein Sci.* **3**, 843–852
140. Toniolo, C., and Benedetti, E. (1991) *Trends Biochem. Sci.* **16**, 350–353
141. Avbelj, F., and Moult, J. (1995) *Biochemistry* **34**, 755–764
142. Karpen, M. E., De Haseth, P. L., and Neet, K. E. (1992) *Protein Sci.* **1**, 1333–1342
143. Fiori, W. R., Miick, S. M., and Millhauser, G. L. (1993) *Biochemistry* **32**, 11957–11962
144. Smythe, M. L., Huston, S. E., and Marshall, G. R. (1993) *J. Am. Chem. Soc.* **115**, 11594–11595
145. Bavosa, A., Benedetti, E., Di Blasio, B., Pavone, V., Pedone, C., Toniolo, C., and Bonora, G. M. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1988–1992
145a. Fathallah-Shaykh, H., Wolf, S., Wong, E., Posner, J. B., and Furneaux, H. M. (1991) *Proc. Natl. Acad. Sci. U.S.A* **88**, 3451–3454
146. Miick, S. M., Martinez, G. V., Flori, W. R., Todd, A. P., and Millhauser, G. L. (1992) *Nature (London)* **359**, 653–655
147. Wada, A. (1976) *Adv. Biophys.* **9**, 1–63
148. Hol, W. G. J., van Duijnen, P. T., and Berendsen, H. J. C. (1978) *Nature (London)* **273**, 443–446
148a. Hol, W. G. J., Halie, L. M., and Sander, C. (1981) *Nature (London)* **294**, 532–536
148b. Lockhart, D. J., and Kim, P. S. (1992) *Science* **257**, 947–951
149. Levine, B. F., and Bethea, C. G. (1976) *J. Chem. Phys.* **65**, 1989–1993
150. Gray, T. M., and Matthews, B. W. (1984) *J. Mol. Biol.* **175**, 75–81
151. Richardson, J. S., and Richardson, D. C. (1988) *Science* **240**, 1648–1652
152. Nicholson, H., Anderson, D. E., Dao-pin, S., and Matthews, B. W. (1991) *Biochemistry* **30**, 9816–9828
153. Åqvist, J., Luecke, H., Quiocho, F. A., and Warshel, A. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2026–2030
154. Lyu, P. C., Wemmer, D. E., Zhou, H. X., Pinker, R. J., and Kallenbach, N. R. (1993) *Biochemistry* **32**, 421–425
155. Viguera, A. R., and Serrano, L. (1995) *J. Mol. Biol.* **251**, 150–160
156. Doig, A. J., Chakrabartty, A., Klingler, T. M., and Baldwin, R. L. (1994) *Biochemistry* **33**, 3396–3403
157. Doig, A. J., Macarthur, M. W., Stapley, B. J., and Thornton, J. M. (1997) *Protein Sci.* **6**, 147–155
158. Sali, D., Bycroft, M., and Fersht, A. R. (1988) *Nature (London)* **335**, 740–743
159. Aurora, R., and Rose, G. D. (1998) *Protein Sci.* **7**, 21–38
160. Metzler, D. E. (1979) *Adv. Enzymol.* **50**, 1–40
161. Lupas, A. (1996) *Trends Biochem. Sci.* **21**, 375–382
161. Lupas, A. (1996) *Trends Biochem. Sci.* **21**, 375–382
162. Kohn, W. D., Mant, C. T., and Hodges, R. S. (1997) *J. Biol. Chem.* **272**, 2583–2586
163. Fraser, R. D. B., MacRae, T. P., Parry, D. A. D., and Suzuki, E. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1179–1183
164. Johnson, L. D., Idler, W. W., Zhou, X.-M., Roop, D. R., and Steinert, P. M. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1896–1900
165. Cooper, D., and Sun, T.-T. (1986) *J. Biol. Chem.* **261**, 4646–4654
166. Crick, F. H. C. (1953) *Acta Crystallogr.* **6**, 689–697
167. Vazquez, S. R., Kuo, D. Z., Bositis, C. M., Hardy, L. W., Lew, R. A., and Humphreys, R. E. (1992) *J. Biol. Chem.* **267**, 7406–7410
168. O'Shea, E. K., Llemm, J. D., Kim, P. S., and Alber, T. (1991) *Science* **254**, 539–544
169. Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) *Nature (London)* **363**, 38–44
170. Lyu, P. C., Marky, L. A., and Kallenbach, N. R. (1989) *J. Am. Chem. Soc.* **111**, 2733–2734
171. Graddis, T. J., Myszka, D. G., and Chaiken, I. M. (1993) *Biochemistry* **32**, 12664–12671
171a. Fairman, R., Chao, H.-G., Lavoie, T. B., Villafranca, J. J., Matsueda, G. R., and Novotny, J. (1996) *Biochemistry* **35**, 2824–2829
171b. Kohn, W. D., Kay, C. M., and Hodges, R. S. (1998) *J. Mol. Biol.* **283**, 993–1012
172. John, M., Briand, J.-P., Granger-Schnarr, M., and Schnarr, M. (1994) *J. Biol. Chem.* **269**, 16247–16253
173. Zhou, N. E., Kay, C. M., and Hodges, R. S. (1992) *Biochemistry* **31**, 5739–5746
174. Cohen, C., and Parry, D. A. D. (1986) *Trends Biochem. Sci.* **11**, 245–248
175. Phillips, G. N., Jr., Fillers, J. P., and Cohen, C. (1986) *J. Mol. Biol.* **192**, 111–131
176. Mo, J., Holtzer, M. E., and Holtzer, A. (1993) *Protein Sci.* **2**, 128–130
177. Xie, X., Rao, S., Walian, P., Hatch, V., Phillips, G. N., Jr., and Cohen, C. (1994) *J. Mol. Biol.* **236**, 1212–1226
178. Eckert, R. L. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1114–1118
179. Gernert, K. M., Surles, M. C., Labean, T. H., Richardson, J. S., and Richardson, D. C. (1995) *Protein Sci.* **4**, 2252–2260
179a. Monera, O. D., Zhou, N. E., Lavigne, P., Kay, C. M., and Hodges, R. S. (1996) *J. Biol. Chem.* **271**, 3995–4001
180. Lovejoy, B., Choe, S., Cascio, D., McRorie, D. K., DeGrado, W. F., and Eisenberg, D. (1993) *Science* **259**, 1288–1293

## References


181. Harbury, P. B., Kim, P. S., and Alber, T. (1994) *Nature (London)* **371**, 80–83
182. Efimov, V. P., Nepluev, I. V., Sobolev, B. N., Zurabishvili, T. G., Schulthess, T., Lustig, A., Engel, J., Haener, M., Aebi, U., Venyaminov, S. Y., Potekhin, S. A., and Mesyanzhinov, V. V. (1994) *J. Mol. Biol.* **242**, 470–486
183. Lieberman, M., Tabet, M., and Sasaki, T. (1994) *J. Am. Chem. Soc.* **116**, 5035–5044
184. Utani, A., Nomizu, M., Timpl, R., Roller, P. P., and Yamada, Y. (1994) *J. Biol. Chem.* **269**, 19167–19175
185. Robinson, C. R., and Sligar, S. G. (1993) *Protein Sci.* **2**, 826–837
186. Steif, C., Weber, P., Hinz, H.-J., Flossdorf, J., Cesareni, G., and Kokkinidis, M. (1993) *Biochemistry* **32**, 3867–3876
187. Redfield, C., Boyd, J., Smith, L. J., Smith, R. A. G., and Dobson, C. M. (1992) *Biochemistry* **31**, 10431–10437
188. Harris, N. L., Presnell, S. R., and Cohen, F. E. (1994) *J. Mol. Biol.* **236**, 1356–1368
189. Chou, K.-C., Carlacci, L., Maggiora, G. M., Parodi, L. A., and Schulz, M. W. (1992) *Protein Sci.* **1**, 810–827
190. Rees, D. C., Komiya, H., Yeates, T. O., Allen, J. P., and Feher, G. (1989) *Ann. Rev. Biochem.* **58**, 607–633
191. Aleshin, A. E., Hoffman, C., Firsov, L. M., and Honzatko, R. B. (1994) *J. Mol. Biol.* **238**, 575–591
192. Aleshin, A. E., Firsov, L. M., and Honzatko, R. B. (1994) *J. Biol. Chem.* **269**, 15631–15639
193. Harris, E. M. S., Aleshin, A. E., Firsov, L. M., and Honzatko, R. B. (1993) *Biochemistry* **32**, 1618–1626
194. Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) *Science* **266**, 75–81
194a. Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B., and Berman, H. M. (1999) *Nature Struct. Biol.* **6**, 454–457
195. Brodsky, B., and Shah, N. K. (1995) *FASEB J.* **9**, 1537–1546
196. Redeker, V., Levilliers, N., Schmitter, J.-M., Le Caer, J.-P., Rossier, J., Adoutte, A., and Bré, M.-H. (1994) *Science* **266**, 1688–1691
197. Miller, A. (1982) *Trends Biochem. Sci.* **7**, 13–18
198. Eyre, D. R. (1980) *Science* **207**, 1315–1322
199. Parry, D. A. D., and Craig, A. S. (1979) *Nature (London)* **282**, 213–215
200. Porter, R. R., and Reid, K. B. M. (1979) *Adv. Prot. Chem.* **33**, 1–71
201. Ashford, D., and Neuberger, A. (1980) *Trends Biochem. Sci.* **5**, 245–248
202. Chen, J., and Varner, J. E. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4399–4403
203. Adzhubei, A. A., and Sternberg, M. J. E. (1994) *Protein Sci.* **3**, 2395–2410
204. Richardson, J. S. (1985) *Nature (London)* **316**, 102–103
205. Hutchinson, E. G., and Thornton, J. M. (1994) *Protein Sci.* **3**, 2207–2216
206. Venkatachalam, C. M. (1968) *Biopolymers* **6**, 1425–1436
207. Perczel, A., Foxman, B. M., and Fasman, G. D. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8210–8214
208. Yan, Y., Tropsha, A., Hermans, J., and Erickson, B. W. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7898–7902
209. Perczel, A., McAllister, M. A., Császár, P., and Csizmadia, I. G. (1993) *J. Am. Chem. Soc.* **115**, 4849–4858
210. Mattos, C., Petsko, G. A., and Karplus, M. (1994) *J. Mol. Biol.* **238**, 733–747
211. Hynes, T. R., Hodel, A., and Fox, R. O. (1994) *Biochemistry* **33**, 5021–5030
212. Haque, T. S., Little, J. C., and Gellman, S. H. (1994) *J. Am. Chem. Soc.* **116**, 4105–4106
213. Gierasch, L. M., Deber, C. M., Madison, V., Niu, C.-H., and Blout, E. R. (1981) *Biochemistry* **20**, 4730–4738
214. Rees, D. C., Lewis, M., and Lipscomb, W. N. (1983) *J. Mol. Biol.* **168**, 367–387
215. Sapse, A.-M., Mallah-Levy, L., Daniels, S. B., and Erickson, B. W. (1987) *J. Am. Chem. Soc.* **109**, 3526–3529
216. Urry, D. W. (1993) *Angew. Chem. Int. Ed. Engl.* **32**, 819–841
217. Sandberg, L. B., Soskel, N. T., and Leslie, J. G. (1981) *N. Engl. J. Med.* **304**, 566–579
217a. Tatham, A. S., and Shewry, P. R. (2000) *Trends Biochem. Sci.* **25**, 567–571
218. Gotte, L., Volpin, D., Horne, R. W., and Mammi, M. (1976) *Micron* **7**, 95–102
218a. Urry, D., Luan, C.-H., and Peng, S. (1995) *Ciba. Found. Symp.* **192**, 4–30
218b. van Dijk, A. A., de Boef, E., Bekkers, A., van Wijk, L. L., van Swieten, E., Hamer, R. J., and Robillard, G. T. (1997) *Protein Sci.* **6**, 649–656
219. Sternberg, M. J. E., and Thornton, J. M. (1976) *J. Mol. Biol.* **105**, 367–382
220. Leszczýnski, J. F., and Rose, G. D. (1986) *Science* **234**, 849–855
221. Liang, C., and Mislow, K. (1994) *J. Am. Chem. Soc.* **116**, 11189–11190
222. Chothia, C. (1984) *Ann. Rev. Biochem.* **53**, 537–572
223. Resnick, D., Pearson, A., and Krieger, M. (1994) *Trends Biochem. Sci.* **19**, 5–8
224. Shoyab, M., Plowman, G. D., McDonald, V. L., Bradley, J. G., and Todaro, G. J. (1989) *Science* **243**, 1074–1076
225. Doolittle, R. F. (1995) *Ann. Rev. Biochem.* **64**, 287–314
226. Berman, A. L., Kolker, E., and Trifanov, E. N. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4044–4047
227. Orengo, C. A., Jones, D. T., and Thornton, J. M. (1994) *Nature (London)* **372**, 631–634
228. Pan, K.-M., Damodaran, S., and Greaser, M. L. (1994) *Biochemistry* **33**, 8255–8261
229. Doolittle, R. F., and Bork, P. (1993) *Sci. Am.* **269**(Oct), 50–56
230. Baron, M., Norman, D. G., and Campbell, I. D. (1991) *Trends Biochem. Sci.* **16**, 13–17
231. Mulichak, A. M., Tulinsky, A., and Ravichandran, K. G. (1991) *Biochemistry* **30**, 10576–10588
232. Bork, P., Holm, L., and Sander, C. (1994) *J. Mol. Biol.* **242**, 309–320
232a. Bogan, A. A., and Thorn, K. S. (1998) *J. Mol. Biol.* **280**, 1–9
233. Tsai, J., Taylor, R., Chothia, C., and Gerstein, M. (1999) *J. Mol. Biol.* **290**, 253–266
233a. Gerstein, M., and Chothia, C. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10167–10172
234. Varadarajan, R., and Richards, F. M. (1992) *Biochemistry* **31**, 12315–12327
235. Allewell, N. (1987) *Trends Biochem. Sci.* **12**, 417–418
236. Lim, W. A., Farruggio, D. C., and Sauer, R. T. (1992) *Biochemistry* **31**, 4324–4333
237. Hubbard, S. J., and Argos, P. (1994) *Protein Sci.* **3**, 2194–2206
237a. Hubbard, S. J., and Argos, P. (1996) *J. Mol. Biol.* **261**, 289–300
238. Hunt, N. G., Gregoret, L. M., and Cohen, F. E. (1994) *J. Mol. Biol.* **241**, 214–225
239. Hunter, C. A., and Sanders, K. M. (1990) *J. Am. Chem. Soc.* **112**, 5525–5534
240. Burley, S. K., and Petsko, G. A. (1985) *Science* **229**, 23–28
241. Thomas, K. A., Smith, G. M., Thomas, T. B., and Feldmann, R. J. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4843–4847
242. Flocco, M. M., and Mowbray, S. L. (1994) *J. Mol. Biol.* **235**, 709–717
243. Hemmingsen, J. M., Gernert, K. M., Richardson, J. S., and Richardson, D. C. (1994) *Protein Sci.* **3**, 1927–1937
244. Teeter, M. M. (1991) *Ann. Rev. Biophys. Biophys. Chem.* **20**, 577–600
245. McDonald, I. K., and Thornton, J. M. (1994) *J. Mol. Biol.* **238**, 777–793
246. Bordo, D., and Argos, P. (1994) *J. Mol. Biol.* **243**, 504–519
247. Shimoni, L., and Glusker, J. P. (1995) *Protein Sci.* **4**, 65–74
248. Dao-pin, S., Anderson, D. E., Baase, W. A., Dahlquist, F. W., and Matthews, B. W. (1991) *Biochemistry* **30**, 11521–11529
248a. Kumar, S., and Nussinov, R. (1999) *J. Mol. Biol.* **293**, 1241–1255
248b. Strop, P., and Mayo, S. L. (2000) *Biochemistry* **39**, 1251–1255
249. Sawyer, L., and James, M. N. G. (1982) *Nature (London)* **295**, 79–80
250. Borders, C. L., Jr., Broadwater, J. A., Bekeny, P. A., Salmon, J. E., Lee, A. S., Eldridge, A. M., and Pett, V. B. (1994) *Protein Sci.* **3**, 541–548
251. Mrabet, N. T., Van den Broeck, A., Van den brande, I., Stanssens, P., Laroche, Y., Lambeir, A.-M., Matthiissens, G., Jenkins, J., Chiadmi, M., van Tilbeurgh, H., Rey, F., Janin, J., Quax, W. J., Lasters, I., De Maeyer, M., and Wodak, S. J. (1992) *Biochemistry* **31**, 2239–2253
252. Dougherty, D. A. (1996) *Science* **271**, 163–168
253. Mitchell, J. B. O., Nandi, C. L., McDonald, I. K., Thornton, J. M., and Price, S. (1994) *J. Mol. Biol.* **239**, 315–331
253a. De Wall, S. L., Meadows, E. S., Barbour, L. J., and Gokel, G. W. (1999) *J. Am. Chem. Soc.* **121**, 5613–5614
253b. Gallivan, J. P., and Dougherty, D. A. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9459–9464
253c. Wouters, J. (1998) *Protein Sci.* **7**, 2472–2475
253d. Fernández-Recio, J., Romero, A., and Sancho, J. (1999) *J. Mol. Biol.* **290**, 319–330
254. Rose, G. D., Young, W. B., and Gierasch, L. M. (1983) *Nature (London)* **304**, 654–657
255. Watenpaugh, K. D., Margulis, T. N., Sieker, L. C., and Jensen, L. H. (1978) *J. Mol. Biol.* **122**, 175–190
255a. Nakasako, M. (1999) *J. Mol. Biol.* **289**, 547–564
256. Teeter, M. M. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6014–6018
257. Otting, G., Liepinsh, E., and Wüthrich, K. (1991) *Science* **254**, 974–980
258. Murzin, A. G., and Finkelstein, A. V. (1988) *J. Mol. Biol.* **204**, 749–770
259. Chothia, C. (1989) *Nature (London)* **337**, 204–205
260. Buehner, M., Ford, G. C., Moras, D., Olsen, K. W., and Rossmann, M. G. (1974) *J. Mol. Biol.* **90**, 25–49
261. Rossmann, M. G., Liljas, A., Branden, C.–I., and Banaszak, L. J. (1975) *The Enzymes*, 3rd ed., Vol. 11, Academic Press, New York (p. 61)
262. Brändén, C. (1990) *Nature (London)* **346**, 607–608
263. Kobe, B., and Deisenhofer, J. (1995) *Nature (London)* **374**, 183–186
264. Kobe, B., and Deisenhofer, J. (1994) *Trends Biochem. Sci.* **19**, 415–421
264a. Kajava, A. V. (1998) *J. Mol. Biol.* **277**, 519–527
264b. Kobe, B., and Kajava, A. V. (2000) *Trends Biochem. Sci.* **25**, 509–515
265. Farber, G. K., and Petsko, G. A. (1990) *Trends Biochem. Sci.* **15**, 228–234
266. Reardon, D., and Farber, G. K. (1995) *FASEB J.* **9**, 497–503
267. Uhlin, U., and Eklund, H. (1994) *Nature (London)* **370**, 533–539
268. Chothia, C. (1992) *Nature (London)* **357**, 543–544
268a. Zhang, C., and DeLisi, C. (1998) *J. Mol. Biol.* **284**, 1301–1305
269. McLachlan, A. D. (1980) *Nature (London)* **285**, 267–268

## References


270. Chou, P. Y., and Fasman, G. D. (1974) *Biochemistry* **13**, 222–245
270a. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1997) *Biochemistry* **36**, 10923–10929
270b. Spek, E. J., Olson, C. A., Shi, Z., and Kallenbach, N. R. (1999) *J. Am. Chem. Soc.* **121**, 5571–5572
271. Minor, D. L., Jr., and Kim, P. S. (1994) *Nature (London)* **367**, 660–663
272. Richards, F. M. (1991) *Sci. Am.* **264**(Jan), 54–63
273. Anfinsen, C. B. (1973) *Science* **181**, 223–230
274. Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D., and Chan, H. S. (1995) *Protein Sci.* **4**, 561–602
275. Rost, B., Schneider, R., and Sander, C. (1993) *Trends Biochem. Sci.* **18**, 120–123
276. Wako, H., and Blundell, T. L. (1994) *J. Mol. Biol.* **238**, 693–708
277. Yi, T.-M., and Lander, E. S. (1993) *J. Mol. Biol.* **232**, 1117–1129
278. Monge, A., Friesner, R. A., and Honig, B. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5027–5029
279. Smith, C. K., Withka, J. M., and Regan, L. (1994) *Biochemistry* **33**, 5510–5517
279a. Lacroix, E., Viguera, A. R., and Serrano, L. (1998) *J. Mol. Biol.* **284**, 173–191
279b. Gotoh, O. (1996) *J. Mol. Biol.* **264**, 823–838
279c. Dandekar, T., and Argos, P. (1996) *J. Mol. Biol.* **256**, 645–660
279d. Jones, D. T. (1999) *J. Mol. Biol.* **292**, 195–202
280. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
281. Reyes, V. E., Phillips, L., Humphreys, R. E., and Lew, R. A. (1989) *J. Biol. Chem.* **264**, 12854–12858
282. Wold, F. (1986) *Trends Biochem. Sci.* **11**, 58–59
283. Graves, D. J., Martin, B. L., and Wang, J. H. (1994) *Co-and post-translational modification of proteins*, Oxford Univ. Press, New York
283a. Angeletti, R. H., ed. (1998) *Protein, Analysis and Design*, Academic Press, San Diego, California
284. Yan, S. C. B., Grinnell, B. W., and Wold, F. (1989) *Trends Biochem. Sci.* **14**, 264–268
285. Berthiaume, L., Deichaite, I., Peseckis, S., and Resh, M. D. (1994) *J. Biol. Chem.* **269**, 6498–6505
286. McIlhinney, R. A. J. (1990) *Trends Biochem. Sci.* **15**, 387–391
287. Violand, B. N., Schlittler, M. R., Lawson, C. Q., Kane, J. F., Siegel, N. R., Smith, C. E., Kolodziej, E. W., and Duffin, K. L. (1994) *Protein Sci.* **3**, 1089–1097
288. Hilz, H., Fanick, N., and Klapproth, K. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6267–6271
289. Hortin, G., Fok, K. F., Toren, P. C., and Strauss, A. W. (1987) *J. Biol. Chem.* **262**, 3082–3085
290. Bateman, A., Solomon, S., and Bennett, H. P. J. (1990) *J. Biol. Chem.* **265**, 22130–22136
291. Takahara, H., Tsuchida, M., Kusubata, M., Akutsu, K., Tagami, S., and Sugawara, K. (1989) *J. Biol. Chem.* **264**, 13361–13368
292. Kreil, G. (1994) *J. Biol. Chem.* **269**, 10967–10970
293. Mor, A., Amiche, M., and Nicolas, P. (1992) *Trends Biochem. Sci.* **17**, 481–485
294. Kreil, G. (1994) *Science* **266**, 996–997
295. Prescott, B., Renugopalakrishnan, V., Glimcher, M. J., Bhushan, A., and Thomas, G. J., Jr. (1986) *Biochemistry* **25**, 2792–2798
296. Wood, D. D., and Moscarello, M. A. (1989) *J. Biol. Chem.* **264**, 5121–5127
297. Waite, J. H., Housley, T. J., and Tanzer, M. L. (1985) *Biochemistry* **24**, 5010–5014
298. Taylor, S. W., Waite, J. H., Ross, M. M., Shabanowitz, J., and Hunt, D. F. (1994) *J. Am. Chem. Soc.* **116**, 10803–10804
299. Clarke, S. (1992) *Ann. Rev. Biochem.* **61**, 355–386
300. Marshall, C. J. (1993) *Science* **259**, 1865–1866
301. Mroz, E. A. (1989) *Science* **243**, 1615
302. Rothgeb, T. M., England, R. D., Jones, B. N., and Gurd, R. S. (1978) *Biochemistry* **17**, 4564–4571
303. Lorimer, G. H. (1983) *Trends Biochem. Sci.* **8**, 65–68
304. Friedman, M. (1977) *Protein Crosslinking*, Vol. 2, Plenum, New York
305. Torchinsky, Y. M. (1981) *Sulfur in Proteins*, Pergamon, Oxford
306. Harrison, P. M., and Sternberg, M. J. E. (1994) *J. Mol. Biol.* **244**, 448–463
307. Folk, J. E., and Finlayson, J. S. (1977) *Adv. Prot. Chem.* **31**, 1–133
308. Gurd, F. R. N., and Rothgeb, T. M. (1979) *Adv. Prot. Chem.* **33**, 73–165
309. Karplus, M., and McCammon, J. A. (1983) *Ann. Rev. Biochem.* **53**, 263–300
310. Caspar, D. L. D., Clarage, J., Salunke, D. M., and Clarage, M. (1988) *Nature (London)* **332**, 659–662
311. Smith, J. L., Hendrickson, W. A., Honzatko, R. B., and Sheriff, S. (1986) *Biochemistry* **25**, 5018–5027
312. Frauenfelder, H., Sligar, S. G., and Wolynes, P. G. (1991) *Science* **254**, 1598–1603
313. Buck, M., Radford, S. E., and Dobson, C. M. (1994) *J. Mol. Biol.* **237**, 247–254
314. Kim, K.-S., Fuchs, J. A., and Woodward, C. K. (1993) *Biochemistry* **32**, 9600–9608
315. Rohl, C. A., and Baldwin, R. L. (1994) *Biochemistry* **33**, 7760–7767
316. Spyracopoulos, L., and O'Neil, J. D. J. (1994) *J. Am. Chem. Soc.* **116**, 1395–1402
317. Huber, R. (1979) *Trends Biochem. Sci.* **4**, 271–276
318. Lumry, R., and Rosenberg, A. (1974) *Colloques Internationaux du C.N.R.S.* **246**, 53
319. Mao, B., Pear, M. R., McCammon, J. A., and Northrup, S. H. (1981) *J. Mol. Biol.* **151**, 199–202
320. Brack, A., and Spach, G. (1981) *J. Am. Chem. Soc.* **103**, 6319–6323
321. Frauenfelder, H., Parak, F., and Young, R. D. (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 451–479
322. Gerstein, M., Lesk, A. M., and Chothia, C. (1994) *Biochemistry* **33**, 6739–6749
323. Kretsinger, R. H., and Nockolds, C. E. (1973) *J. Biol. Chem.* **248**, 3313–3326
324. Liepinsh, E., and Otting, G. (1994) *J. Am. Chem. Soc.* **116**, 9670–9674
325. Scholtz, J. M., Barrick, D., York, E. J., Stewart, J. M., and Baldwin, R. L. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 185–189
326. Kamoun, P. P. (1988) *Trends Biochem. Sci.* **13**, 424–425
327. Dill, K. A., and Shortle, D. (1991) *Ann. Rev. Biochem.* **60**, 795–825
328. Baker, D., and Agard, D. A. (1994) *Biochemistry* **33**, 7505–7509
328a. Sharp, K. A., and Englander, S. W. (1994) *Trends Biochem. Sci.* **19**, 526–529
328b. Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) *FASEB J.* **10**, 75–83
328c. Sippl, M. J. (1996) *J. Mol. Biol.* **260**, 644–648
328d. Meyer, E. (1992) *Protein Sci.* **1**, 1543–1562
328e. Hill, R. B., Hong, J.-K., and DeGrado, W. F. (2000) *J. Am. Chem. Soc.* **122**, 746–747
329. Levinthal, C. (1968) *J. Chim. Phys.-Chim. Biol.* **65**, 44–45
330. Lattman, E. E., Fiebig, K. M., and Dill, K. A. (1994) *Biochemistry* **33**, 6158–6166
331. Mark, A. E., and van Gunsteren, W. F. (1992) *Biochemistry* **31**, 7745–7748
332. Peng, Z., and Kim, P. S. (1994) *Biochemistry* **33**, 2136–2141
333. Hagihara, Y., Tan, Y., and Goto, Y. (1994) *J. Mol. Biol.* **237**, 336–348
334. Yu, Y., Makhatadze, G. I., Pace, C. N., and Privalov, P. L. (1994) *Biochemistry* **33**, 3312–3319
335. Dobson, C. M., Evans, P. A., and Radford, S. E. (1994) *Trends Biochem. Sci.* **19**, 31–37
336. Yang, A.-S., and Honig, B. (1994) *J. Mol. Biol.* **237**, 602–614
336a. Baker, D. (2000) *Nature (London)* **405**, 39–42
336b. Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M., and Karplus, M. (2000) *Trends Biochem. Sci.* **25**, 331–339
337. Udgaonkar, J. B., and Baldwin, R. L. (1988) *Nature (London)* **335**, 694–699
338. Roder, H., Elöve, G. A., and Englander, S. W. (1988) *Nature (London)* **335**, 700–704
339. Gething, M.-J., and Sambrook, J. (1992) *Science* **355**, 33–45
340. Odefey, C., Mayr, L. M., and Schmid, F. X. (1995) *J. Mol. Biol.* **245**, 69–78
341. Chazin, W. J., Kördel, J., Drakenberg, T., Thulin, E., Brodin, P., Grundström, T., and Forsén, S. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2195–2198
342. Lin, L.-N., and Brandts, J. F. (1987) *Biochemistry* **26**, 3537–3543
343. Herning, T., Yutani, K., Taniyama, Y., and Kikuchi, M. (1991) *Biochemistry* **30**, 9882–9891
344. Chatrenet, B., and Chang, J.-Y. (1993) *J. Biol. Chem.* **268**, 20988–20996
345. Talluri, S., Rothwarf, D. M., and Scheraga, H. A. (1994) *Biochemistry* **33**, 10437–10449
346. Pace, C. N. (1990) *Trends Biochem. Sci.* **15**, 14–17
347. Yang, A.-S., and Honig, B. (1993) *J. Mol. Biol.* **231**, 459–474
348. Hartsough, D. S., and Merz, K. M., Jr. (1993) *J. Am. Chem. Soc.* **115**, 6529–6537
349. Aswad, D. W., and Johnson, B. A. (1987) *Trends Biochem. Sci.* **12**, 155–158
350. Brennan, T. V., and Clarke, S. (1993) *Protein Sci.* **2**, 331–338
351. Volkin, D. B., and Klibanov, A. M. (1987) *J. Biol. Chem.* **262**, 2945–2950
352. Wells-Knecht, M. C., Huggins, T. G., Dyer, D. G., Thorpe, S. R., and Baynes, J. W. (1993) *J. Biol. Chem.* **268**, 12348–12352
353. Luthra, M., Ranganathan, D., Ranganathan, S., and Balasubramanian, D. (1994) *J. Biol. Chem.* **269**, 22678–22682
354. Lubec, G., Weninger, M., and Anderson, S. R. (1994) *FASEB J.* **8**, 1166–1169
355. Kaiser, E. T. (1987) *Trends Biochem. Sci.* **12**, 305–309
355a. Beasley, J. R., and Hecht, M. H. (1997) *J. Biol. Chem.* **272**, 2031–2034
356. Oxender, D. E., and Fox, C. F., eds. (1987) *Protein Engineering*, Liss, New York
356a. Dahiyat, B. I., and Mayo, S. L. (1997) *Science* **278**, 82–87
356b. DeGrado, W. F., Summa, C. M., Pavone, V., Nastri, F., and Lombardi, A. (1999) *Ann. Rev. Biochem.* **68**, 779–819
357. Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O'Neil, K. T., and DeGrado, W. F. (1995) *Science* **270**, 935–941
358. Nautiyal, S., Woolfson, D. N., King, D. S., and Alber, T. (1995) *Biochemistry* **34**, 11645–11651
359. Betz, S. F., Liebman, P. A., and DeGrado, W. F. (1997) *Biochemistry* **36**, 2450–2458
360. Handel, T. M., Williams, S. A., and DeGrado, W. F. (1993) *Science* **261**, 879–885
361. Struthers, M. D., Cheng, R. P., and Imperiali, B. (1996) *Science* **271**, 342–345
362. Yan, Y., and Erickson, B. W. (1994) *Protein Sci.* **3**, 1069–1073
363. Quinn, T. P., Tweedy, N. B., Williams, R. W., Richardson, J. S., and Richardson, D. C. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8747–8751
364. Hecht, M. H. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8729–8730
364a. Kortemme, T., Ramírez-Alvarado, M., and Serrano, L. (1998) *Science* **281**, 253–256

## References

365. Kellis, J. T., Jr., Nyberg, K., and Fersht, A. R. (1989) *Biochemistry* **28**, 4914–4922
366. Goldenberg, D. P. (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 481–507
367. Alber, T. (1989) *Ann. Rev. Biochem.* **58**, 765–798
368. Katz, B. A., and Kossiakoff, A. (1986) *J. Biol. Chem.* **261**, 15480–15485
369. Pantoliano, M. W., Ladner, R. C., Bryan, P. N., Rollence, M. L., Wood, J. F., and Poulos, T. L. (1987) *Biochemistry* **26**, 2077–2082
370. Wetzel, R. (1987) *Trends Biochem. Sci.* **12**, 478–482
371. Mitchinson, C., and Wells, J. A. (1989) *Biochemistry* **28**, 4807–4815
372. Ahern, T. J., Casal, J. I., Petsko, G. A., and Klibanov, A. M. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 675–679
373. Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G., and Wells, J. A. (1986) *Science* **233**, 659–663
374. Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E., and Hutchison, C. A., III. (1989) *Nature (London)* **340**, 397–440
375. Mendel, D., Ellman, J. A., Chang, Z., Veenstra, D. L., Kollman, P. A., and Schultz, P. G. (1992) *Science* **266**, 1798–1802
376. Brady, L., and Dodson, G. (1994) *Nature (London)* **368**, 692–694
377. Liu, N., Deillon, C., Klauser, S., Gutte, B., and Thomas, R. M. (1998) *Protein Sci.* **7**, 1214–1220

## Study Questions

1. Name all of the isometric tripeptides which could be formed from one molecule each of tyrosine, alanine, and valine.

2. What functional groups are found in protein side chains? Of what importance to protein structure and function are (a) hydrophobic groups, (b) acidic and basic groups, (c) sulfyhydryl groups?

3. If placed in water and adjusted to a pH of 7, will the following migrate toward the anode or the cathode if placed in an electrical field?
(a) Aspartic acid, (b) alanine, (c) tyrosine, (d) lysine, (e) arginine, (f) glutamine.

4. Draw the following hydrogen-bonded structures:
   (a) A dimer of acetic acid.
   (b) A tyrosine–carboxylate bond in the interior of a protein.
   (c) A phosphate–guanidinium ion pair in an enzyme–substrate complex.

5. Contrast the properties of the amino acids with those of the saturated fatty acids with respect to solubility in water and in ether and to physical state.

6. Describe in as much detail as you can the characteristic properties of (a) β sheets, (b) α helices, (c) turns in peptide chains, and (d) collagen.

7. Predict whether the following peptide segments will be likely to exist as an α helix or as part of a β structure within a protein:
   (a) Poly-L-leucine
   (b) Poly-L-valine
   (c) Pro-Glu-Met-Val-Phe-Asp-Ile
   (d) Pro-Glu-Ala-Leu-Phe-Ala-Ala

8. Describe three ways in which a side chain of a serine residue can fold back and hydrogen bond to a C=O or N–H group of the backbone and two ways by which an asparagine side chain can do the same. There are yet other possibilities.

9. Compare structural features and properties of the following proteins: silk fibroin, α-keratin, collagen, and bovine serum albumin.

10. In what way do the solubilities of proteins usually vary with pH? Why?

11. Compare the following: the diameters of (a) a carbon atom in an organic molecule (b) a bacterial cell, e.g. of *E. coli* (c) a human red blood cell (d) a ribosome (e) the length of a peptide unit in an extended polypeptide chain (f) the length of the carbon atom chain in an 18-carbon fatty acid.

12. Compare: (a) the length of a peptide unit (residue) in a polypeptide in an extented (β) conformation. (b) the length by which an α helix is extended by the addition of one amino acid unit (c) the length of one turn of an α helix. (d) the diameter of an α helix (both using atom centers in the backbones and using van der Waals radii) for a poly-L-alanine helix.

13. What are: albumins, globulins, protamines, scleroproteins, glycoproteins, lipoproteins?

14. Where are the following found and what are their functions? Gamma globulin, hemocyanin, pepsin, glucagon, ferritin, phosphorylase.

15. List the nutritionally essential amino acids for human beings. Compare these needs with those of other species, including lactic acid bacteria, malaria parasites, green plants, etc.

16. Define: chiral, enantomer, diastereomer, epimer, anomer (see Chapter 4), prochiral (see Chapter 9). What is meant by the statement that biochemical reactions are stereochemically specific? Why is such stereospecificity to be expected in organisms (which are constructed from asymmetric units)? See Chapter 9 for further discussion.

17. What are disulfide bridges and of what significance are they in protein structure?

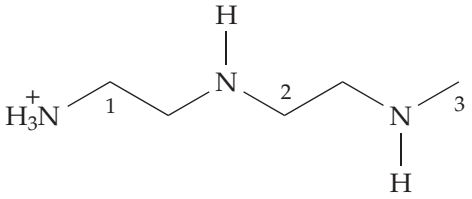
18. What is meant by "denaturation" of a protein? Mention several ways in which denaturation can be brought about. How is denaturation explained in terms of structure?

19. A chain of L-amino acids can form either a right-handed or a left-handed helix. From the Ramachandran diagram in Fig. 2-9, can you say anything about the relative stabilities of right and left-handed helices? What do you predict for polyglycine?

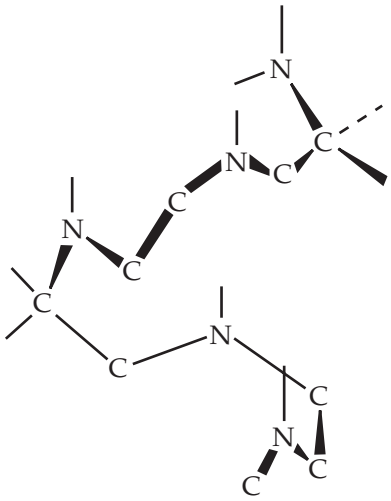
20. What similarities and differences would you predict for two proteins of identical amino acid sequence but one made from all L-amino acids and the other from all D-amino acids?

## Study Questions

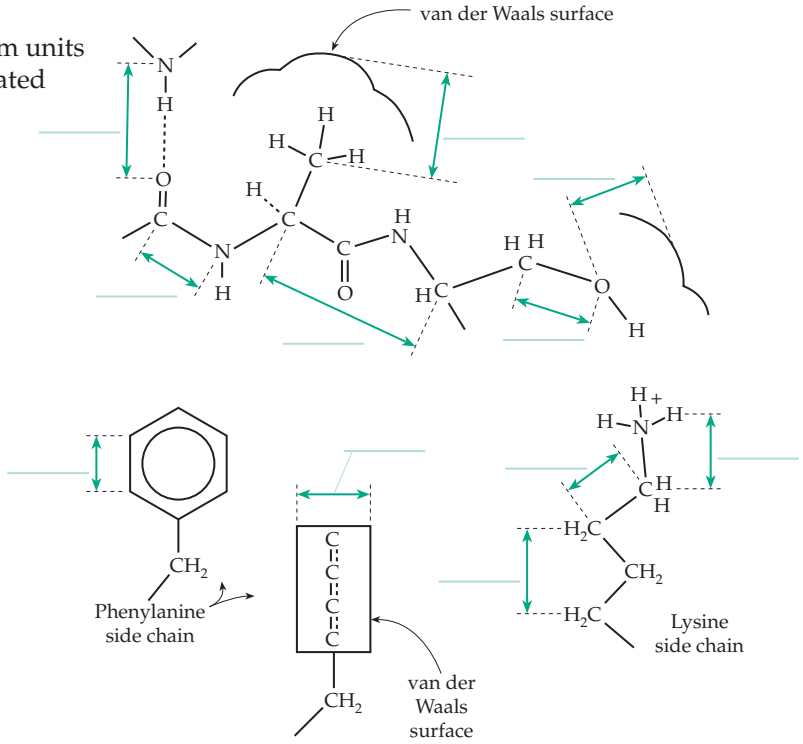
21. Complete the following peptide structure for L-seryl-L-valyl-L-asparaginyl—etc. Extend the chain in the C-terminal direction to form a **beta turn** with the chain coming back to form a **beta sheet**. Add a third segment of peptide *parallel* to the folded back chain to form a 3-stranded beta sheet. Indicate all hydrogen bonds correctly. Draw the side chains of the seryl and asparaginyl residues in positions 1 and 3 so that they form proper hydrogen bonds to groups in the peptide backbone.




22. Complete the following structure to form a short segment of alpha helix. Extend the chain by at least three residues. Form all hydrogen bonds correctly. Add at least three side chains with the correct chirality at the alpha carbon positions. Add one electrically charged side chain and show how it interacts with the peptide backbone at either the C- or N-terminus to help *stabilize* the helix.

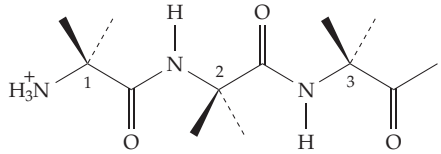


23. What is the relationship between Ångstrom units (Å) and nanometers (nm)? Give the indicated distance in Å or nm.

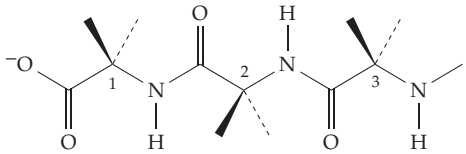


## Study Questions

24. Draw three residues of a polypeptide chain constructed of L-amino acids using the top template of the two below. Now, using the second template, whose polypeptide chain begins at the right, use the same three residues, *numbering from the C-terminus* using D-amino acids. This is know as a *retro*-inverso polypeptide.




Polypeptide

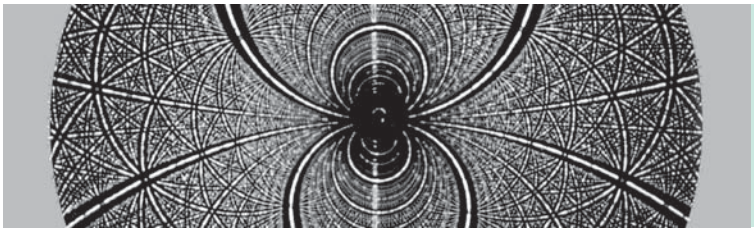



*Retro*-inverso polypeptide

Compare the outer surface of of the standard polypeptide and its *retro*-inverso analog. How would your answer be affected by the presence of threonine or isoleucine in the peptides?

If the two polypeptides were folded into a hairpin loop of β structure how would the exterior surfaces compare? How would the hydrogen bonding compare?

What possible value can you imagine for *retro*-inverso polypeptides in design of drugs? See Brady and Dodson, *Nature* **368**, 692–694 (1994) or Guichard *et al., Proc. Natl. Acad. Sci. USA* **91**, 9765–9769 (1994)



A Laue X-ray diffraction pattern from a protein crystal. A stationary crystal is irradiated with very intense white, multiwavelength X rays from a synchrotron source. The diffraction pattern is rich in information. A single 0.1 ms X-ray pulse may provide a pattern with enough information to determine a three-dimensional structure. The pattern consists of thousands of diffraction spots arranged on intersecting rings. The coordinates of the diffraction spots together with their measured intensities provide the necessary information for structure determination. Courtesy of Louise Johnson.

# Contents


95 **A. Understanding pH and Electrical Charges on Macromolecules**
95 1. Strengths of Acids and Bases: the pKa's
96 2. Titration Curves
97 3. Buffers
98 **B. Isolating Compounds**
98 1. Fractionation of Cells and Tissues
100 2. Separations Based on Molecular Size, Shape, and Density
100 *Dialysis, ultrafiltration, and perfusion chromatography*
100 *Centrifugation*
101 3. Separations Based on Solubility
102 4. Separation by Partition
103 5. Ion Exchange Chromatography
104 6. Affinity Chromatography
106 7. Electrophoresis and Isoelectric Focusing
108 **C. Determining the Relative Molecular Mass, $M_r$**
108 1. Ultracentrifugation
108 *Sedimentation velocity*
112 2. Gel Filtration and Gel Electrophoresis
112 3. Mass Spectrometry
115 **D. Determining Amino Acid Composition and Sequence**
115 1. Cleavage of Disulfide Bridges
116 2. Hydrolysis and Other Chain Cleavage Reactions
116 *Selective enzymatic hydrolysis*
117 *Nonenzymatic cleavages*
118 *Separating the peptides*
118 3. Determining Amino Acid Sequence
118 *The Edman degradation*
119 *Protein sequences from the genes*
119 *Mass spectrometry in sequencing*
119 4. Locating Disulfide Bridges
120 5. Detecting Products
122 6. Absorption of Light
123 **E. Quantitative Determinations and Modification Reactions of Side Chain Groups**
123 1. Reactions of Amino Groups
125 2. Reactions of SH Groups
126 3. Reactions of Other Side Chains
127 4. Affinity Labeling
127 **F. Synthesis of Peptides**
127 1. Solid-Phase Peptide Synthesis
128 2. Combinatorial Libraries
129 **G. Microscopy**
132 **H. X-ray and Neutron Diffraction**
137 **I. Nuclear Magnetic Resonance (NMR)**
137 1. Basic Principles of NMR Spectroscopy
138 *Band widths*
138 *The chemical shift*
139 *Scalar coupling (J coupling)*
140 2. Nuclei Other Than Hydrogen
140 *Deuterium (2H)*
140 *Carbon - 13*
140 *Nitrogen - 15*
140 *Phosphorus - 31*
140 *Fluorine - 19*
141 *Some Other Nuclei*
141 3. Fourier Transform Spectrometers and Two-Dimensional NMR
141 *Free induction decay*
141 *Relaxation times T1 and T2*
142 *Two-dimensional and multidimensional NMR spectra*
145 4. Three-Dimensional Structures and Dynamics of Proteins
145 *Assignment of resonances*
145 *NOEs and distance constraints*
147 5. Other Information from NMR Spectra
147 *NMR titrations*
148 *Observing exchangable protons*
148 *Exchange rates of amide protons*
149 *Solid-state NMR and other topics*
149 **J. The Protein Data Bank, Three-Dimensional Structures, and Computation**

150 **References**
157 **Study Questions**


**Boxes**

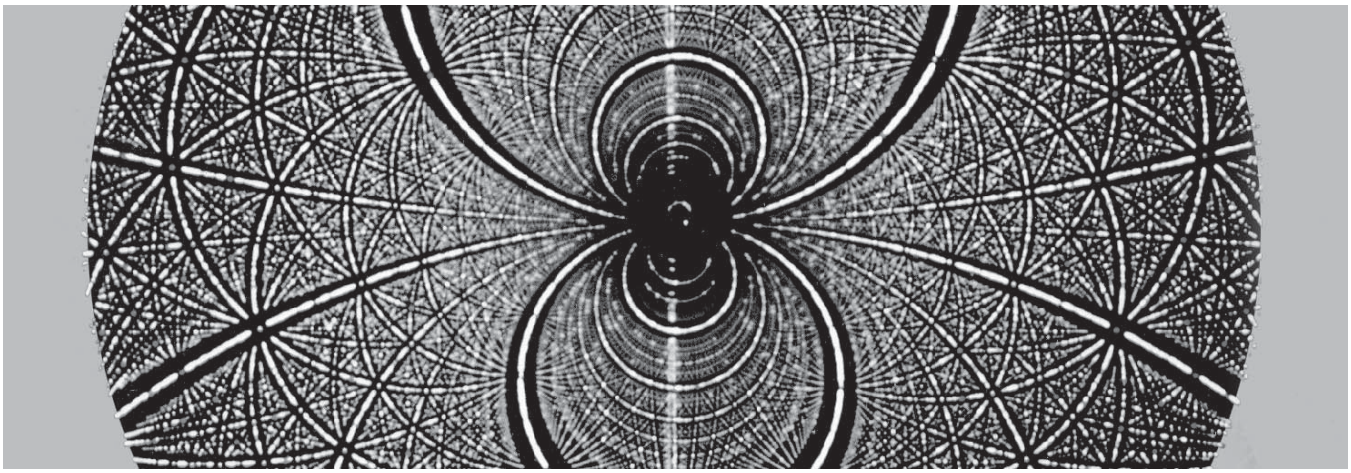

102 Box 3-A Quantitative Estimation of Protein Concentrations
107 Box 3-B Sorting and Analyzing Single Cells
110 Box 3-C Isotopes in Biochemical Investigations
116 Box 3-D The Proteome
121 Box 3-E Ninhydrin
122 Box 3-F Biosensors and Electronic Noses


**Tables**


99 Table 3-1 Practical $pK_a$ Values for Some Useful Buffer Compounds at 25°C and Ionic Strength 0.1
117 Table 3-2 Specifities of Commonly Used Protein-Hydrolyzing Enzymes
141 Table 3-3 Approximate Chemical Shift Ranges in $^1$H- and in $^{15}$N- NMR Spectra
149 Table 3-4 Selected World Wide Web Servers Related to Protein Structures and Sequences

# Determining Structures and Analyzing Cells

# 3



How have chemists deduced the thousands of structural formulas that we write for the substances found in nature? The answer is far too complex to give here in detail. However, the separation of compounds, the analysis of mixtures, and the unraveling of structures remain essential parts of biochemistry. A "minireview" of methods, with emphasis on proteins, follows. Additional procedures having to do primarily with carbohydrates, nucleic acids, or lipids are given in Chapters 4, 5, and 8, respectively.

## A. Understanding pH and Electrical Charges on Macromolecules

Proteins, nucleic acids, and carbohydrates all contain acidic or basic functional groups. The strengths of the acidic groups vary over a broad range from that of the strongly acidic phosphate and sulfate esters to that of the very weakly acidic alcoholic – OH group. The net electrical charge, as well as the spatial distribution of the charged groups, affects the properties of a macromolecule greatly. Therefore, it will be worthwhile for us to consider some aspects of acid–base chemistry before discussing other topics.

### 1. Strengths of Acids and Bases: the p$K_a$'s

The strength of an acid is usually described by the acid dissociation constant $K_a$

$$HA = A^- + H^+ \tag{3-1}$$

$$K_a = [A^-][H^+] / [HA] \tag{3-2}$$

For strong acids $K_a$ is high and for weak acids it is low. Since the values of $K_a$ vary by many orders of magnitude it is customary to use as a measure of the acid strength p$K_a$. This is the negative logarithm of $K_a$ (p$K_a$ = –log $K_a$) . For very strong acids p$K_a$ is less than zero, while very weak acids have p$K_a$ values as high as 15 or more.

In the biochemical literature *the strength of a base is nearly always given by the p$K_a$ of the conjugate acid.* Thus, $A^-$ in Eq. 3-1 is a base and HA its conjugate acid. The base could equally well be uncharged A and its conjugate acid $HA^+$. In both cases Eq. 3-2 would hold. This defines the strength of both the acid HA and the base $A^-$. It follows that strong bases have weak conjugate acids with high p$K_a$ values, while weak bases have strong conjugate acids with low p$K_a$ values.

For a compound containing several acidic groups we define a series of consecutive dissociation constants $K_{1a}$, $K_{2a}$, $K_{3a}$, etc. For the sake of simplicity we will omit the a's and call these $K_1$, $K_2$, $K_3$, ⋯.

$$H_3A \xrightarrow{K_1} H_2A \xrightarrow{K_2} HA \xrightarrow{K_3} A \tag{3-3}$$

If there are $n$ consecutive dissociation constants there will be $n + 1$ ionic species $H_3A$, $H_2A$, etc. Notice that $H_3A$ could be a neutral molecule with $H_2A$, HA, and A bearing changes of –1, –2, and –3, respectively. Alternatively, $H_3A$ might carry 1, 2, or 3 positive changes. In every case the mathematical expressions will be the same. For this reason the charges have been deliberately omitted from Eq. 3-3 and others that follow. Each constant in Eq. 3-3 is defined as in Eq. 3-2, i.e., $K_2 = [HA][H^+] / [H_2A]$, etc. Keep in mind that there

are significant differences between the *apparent equilibrium constants* (concentration equilibrium constants) that we ordinarily use and thermodynamic equilibrium constants that are obtained by extrapolation to zero ionic strength.[1] A related complication is the uncertainity associated with the measurement of pH. This is often considered a measurement of hydrogen ion activity but this is not a correct statement. (The matter is considered briefly in Chapter 6). However, for all practical purposes, in the range of about pH 4–10 the pH can be equated with $-\log [H^+]$.

Often only one of the ionic forms of Eq. 3-3 will be important in a biochemical reaction. A particular ionic species may be the substrate for an enzyme. Likewise, an enzyme–substrate complex in only a certain state of protonation may react to given products. In these cases, and whenever pH affects an equilibrium, it is useful to relate the concentration $[A_i]$ of a given ionic form of a compound to the total of all ionic forms $[A]_t$ using Eqs. 3-4 and 3-5.

$$[A_i] = [A]_t / F_i \tag{3-4}$$

$$[A]_t = [A] + [HA] + [H_2A] + \cdots [H_nA] \tag{3-5}$$

For Eq. 3-4, $A_1 = H_nA$, $A_2 = H_{n-1}A$, etc. and $F_1$, $F_2$, etc. are the *Michaelis pH functions*,[2,3] which were proposed by L. Michaelis in 1914. For the case represented by Eq. 3-3 there are four ionic species and therefore four Michaelis pH functions which have the following form (Eq. 3-6). Here, $K_1$, $K_2$, etc. are the usual consecutive acid dissociation constants.

$$F_1 = 1 + K_1/[H^+] + K_1K_2/[H^+]^2 + K_1K_2K_3/[H^+]^3$$

$$F_2 = [H^+]/K_1 + 1 + K_2/[H^+] + K_2K_3/[H^+]^2$$

$$F_3 = [H^+]^2/K_1K_2 + [H^+]/K_2 + 1 + K_3/[H^+]$$

$$F_4 = [H^+]^3/K_1K_2K_3 + [H^+]^2/K_2K_3 + [H^+]/K_3 + 1 \tag{3-6}$$

If there are only three ionic forms the first three of these equations will apply if the final term is dropped from each. The student should be able to verify these equations and to write the appropriate pH functions for other cases. Since these relationships are met so often in biochemistry it is worthwhile to program a computer to evaluate the Michaelis pH functions and to apply them as needed. From Eq. 3-4 it can be seen that *the reciprocal of the Michaelis pH function for a given ionic form represents the fraction of the total compound in that form* and that the sum of these reciprocals for all the ionic forms is equal to one. Examples of the use of the Michaelis pH functions in this book are given in Eq. 6-50, which relates the Gibbs energy of hydrolysis of ATP to the pH, and in Eqs. 9-55 to 9-57, which describe the pH dependence of enzymatic action.

Notice that in Eq. 3-3 single arrows have been used rather than the pairs ($\rightleftarrows$) that are often employed to indicate reversible equilibria. This is done so that *the direction of the arrow indicates whether we are using a dissociation constant or an association constant*. The use of single arrows in this manner to indicate how the equilibrium constants are to be written is a good practice when dealing with complex equilibria.

## 2. Titration Curves

When a neutral amino acid is titrated with acid, the carboxylate groups become protonated and acid is taken up. Likewise titration with base removes protons from the protonated amino groups and base is taken up. If we plot the number of equivalents of acid or base that have reacted with the neutral amino acid versus pH, a titration curve such as that shown in Fig. 3-1 is generated. The curve for histidine contains three steps; the first corresponds to the titration of the carboxylate group with acid, the second to the titration of the protonated imidazole of the side chain, and the third to the titration of the protonated amino group with base. Each step is characterized by a midpoint that is equal to the $pK_a$ value for the group being titrated. The ends of the curve at low and high pH, which are drawn with a dashed line, are obtained only after corrections have been applied to the data. If only the equivalents of acid or base *added* rather than the number *reacted* are plotted, we obtain the solid line shown in Fig. 3-1. We see that at the low pH end there is no distinct end point. As we add more acid to try to complete the titration, the correction that must be applied to the data becomes increasingly greater. The difference between the dashed and solid lines reflects the fact that at the low pH end much of the acid added is used to simply lower the pH. Therefore, we have a large free $[H^+]$. Similarly, at the high pH end we have a high free $[OH^-]$.

The exact shapes of the ends of the titration curve depend heavily on the total concentration. Likewise, the magnitude of the correction required to obtain a plot of equivalents of acid or base reacted varies with the concentration and is smaller the higher the concentration of the substance being titrated (see problems 2 and 3 at the end of this chapter). An important rule in doing acid-base titrations, especially when very small samples are available, is to use the highest possible concentration of sample in the smallest possible volume and to titrate with relatively concentrated acid or base. Because of the difficulty of adequately correcting titration curves at low pH it is hard to estimate the $pK_a$ values of the carboxyl groups of amino acids accurately from titration. An additional experimental difficulty exists at the high pH end because of the tendency for basic

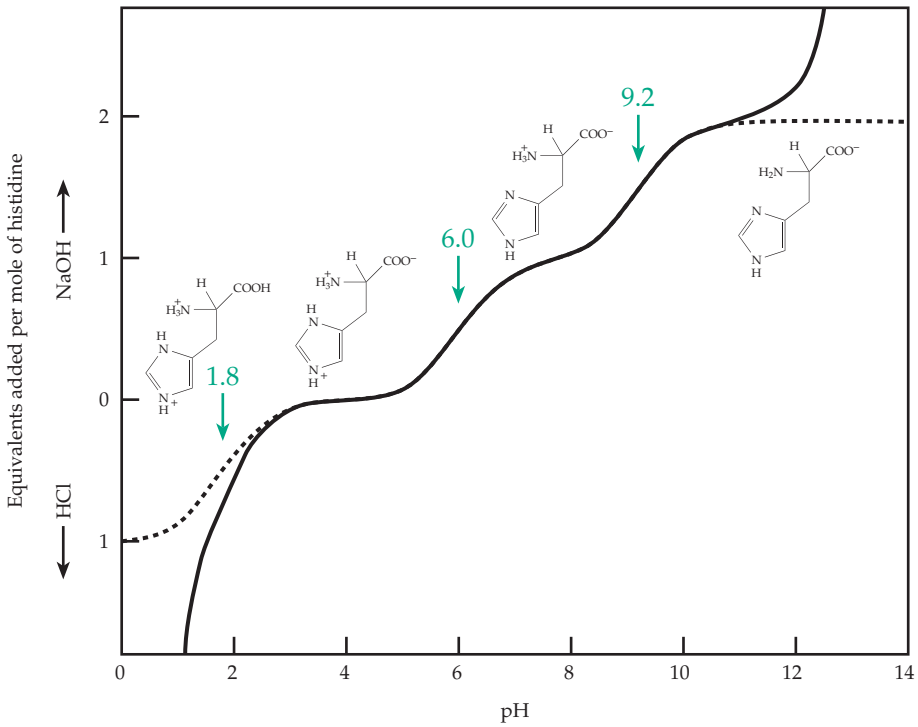



**Figure 3-1** Titration curve for histidine. The solid line represents the uncorrected titration curve for 3 mM histidine monohydrochloride titrated with 0.2 M HCL to lower pH or with 0.2 M NaOH to higher pH assuming p$K_a$ values of 1.82, 6.00, and 9.17. The dashed line represents the corrected curve showing the number of protons bound or lost per mole of histidine monohydrochloride.

solutions to absorb $CO_2$ from the air. Sometimes formaldehyde is added to shift the apparent p$K_a$ of the amino groups to lower values and make the titration more satisfactory.

Despite the difficulties, real proteins can be titrated successfully to estimate the numbers of carboxyl, histidine, tyrosine, and other groups.[4] An example is shown in Fig. 3-2. The experimental data have been fitted with a theoretical curve based on the p$K_a$'s of the carboxyl, histidine, tyrosine, and amino groups as determined by NMR measurements.[5] Account has been taken of the effect of the electrical field created by the many charged groups in distorting the curve from that obtained by summing the theoretical titration curves of the component groups. However, when there are multiple acid–base groups that are close together in a protein, a more complex situation involving tautomerism arises. This is discussed in Chapter 6.

Titration curves based on plots of light absorption versus pH or of NMR chemical shifts versus pH (see Fig. 3-29) are often useful. They have the important advantage that no special correction for free acid or base is needed at low or high pH.

## 3. Buffers

A mixture of a weak acid HA and of its **conjugate base** A constitutes a buffer which resists changes in pH. This can be seen most readily by taking logarithms of both sides of Eq. 3-2. By replacing log $K$ with $-$p$K_a$ and log [$H^+$] with $-$pH and rearranging we obtain Eq. 3-7 (the Henderson–Hasselbalch equation). It is sometimes useful to rewrite this as Eq. 3-8, where $\alpha$ is the fraction of the acid that is dissociated at a given pH.

$$\text{pH} - \text{p}K_a = \log([A]/[HA]) \tag{3-7}$$

$$\text{pH} - \text{p}K_a = \log[\alpha/(1-\alpha)];$$

$$\alpha = [A] / [HA] + [A] = \frac{10^{(\text{pH}-\text{p}K_a)}}{1 + 10^{(\text{pH}-\text{p}K_a)}} \tag{3-8}$$

Logarithms to the base 10 are used in both equations. These equations are useful in preparing buffers and in thinking about what fraction of a substance exists in a given ionic form at a particular value of pH. From Eq. 3-7 it is easy to show that when the pH is near the p$K_a$ relatively large amounts of acid or base must be added to change the pH if the concentrations of the buffer pair A and HA are high.

Buffers are often added to maintain a constant pH in biochemical research[6] and naturally occurring buffer systems within body fluids and cells are very important (Box 6-A). Among the most important natural buffers are the proteins themselves, with the imidazole groups of histidine side chains providing much of the buffering capacity of cells around pH 7 (Figs. 3-1 and 3-2). Table 3-1 lists some useful biochemical buffers and their p$K_a$ values. Here are a few

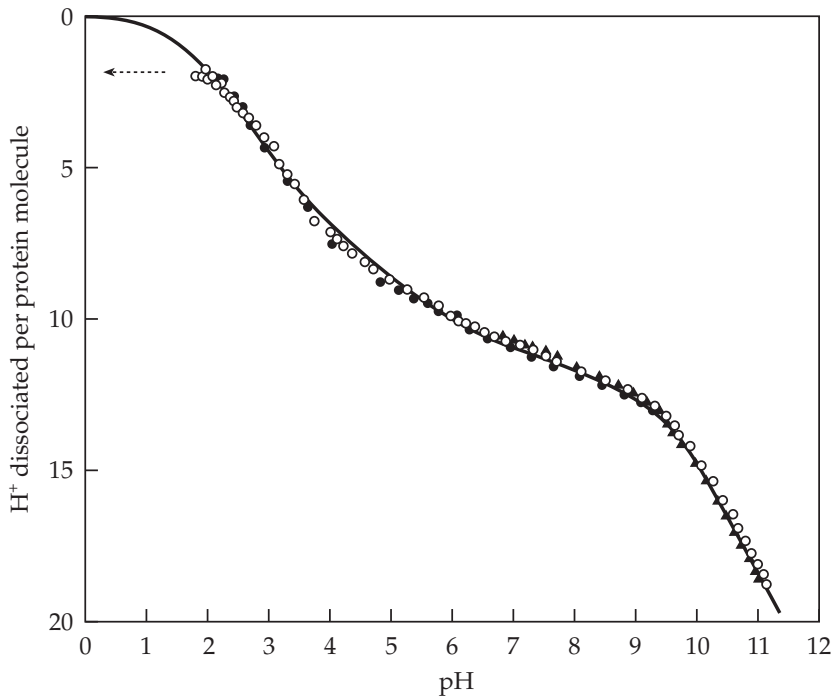



**Figure 3-2** Acid–base titration curve for hen lysozyme at 0.1 ionic strength and 25°C. ○, initial titration from the pH attained after dialysis; ●, back titration after exposure to pH 1.8; ▲, back titration after exposure to pH 11.1. The solid curve was constructed on the basis of "intrinsic" p$K_a$ values based on NMR data. From Kuramitsu and Hamaguchi[5]

practical hints about buffer preparation. Buffers containing monovalent ions tend to change pH with dilution less than do those with multivalent ions such as $HPO_4^{2-}$ or $HP_2O_7^{3-}$. The p$K_a$ values of carboxylic acids and of phosphoric acid, or of its organic derivatives, change very little with change in temperature. The pH of a buffer prepared with such components is nearly independent of temperature (Table 3-1). However, the p$K_a$ of the $-NH_3^+$ group changes greatly with temperature. Buffer composition can be calculated readily from Eq. 3-8 and p$K_a$ values from Table 3-1. It is convenient to keep in the laboratory standardized (to ~1% error) 1 M HCl and 1 M NaOH for use in buffer preparation. Compositions calculated from Eq. 3-8 will usually yield buffers of pH very close to those expected. Final adjustment with HCl may be needed if the pH is more than one unit away from a p$K_a$ value. When two buffering materials are present, the composition should be calculated independently for each. The measurement of pH should always be done with great care because it is easy to make errors. Everything depends upon the reliability of the standard buffers used to calibrate the pH meter.[7] Often, especially during isolation of small compounds, it is desirable to work in the neutral pH region with volatile buffers, e.g., trimethylamine and $CO_2$ or ammonium bicarbonate, which can be removed by vacuum evaporation or lyophilization.[8]

## B. Isolating Compounds

Before structural work can begin, pure substances must be separated from the complex mixtures in which they occur in cells and tissues.[4,13–24] Often, a substance must be isolated from a tissue in which it is present in a very low concentration. After it is isolated in pure form, if it is a large molecule, it must often be cut up into smaller pieces which are separated, purified, and identified. Accurate quantitative analysis is required to determine the ratios of these fragments. Considerable ingenuity may then have to be exercised in putting the pieces of the jigsaw puzzle back together to determine the structure of the "native" molecule. Many books, a few of which are cited here,[4,13–23,25–45] provide instructions. There are also journals and other periodicals dedicated to biochemical methods.[46–53]

### 1. Fractionation of Cells and Tissues

A fresh tissue or a paste of packed cells of a microorganism, usually collected by centrifugation, may be the starting material.[23,54–57] Tissue is often ground in a kitchen-type blender or, for gentler treatment, in a special **homogenizer**. The popular Potter–Elvehjem homogenizer is a small apparatus in which a glass or plastic pestle rotates inside a tight-fitting mortar tube (see standard laboratory equipment catalogs for pictures). Microbial cells are frequently broken with supersonic oscillation (**sonication**) or in special pressure cells. It is important to pay attention to the pH, buffer composition, and, if subcellular organelles are to be separated, the osmotic pressure. To preserve the integrity of organelles, 0.25 M sucrose is frequently used as the suspending medium, and $MgCl_2$ as well as a metal complexing agent such as ethylenediaminetetraacetate (EDTA) (Table 6-10) may be added. Soluble enzymes are often extracted without addition of sucrose, but reducing compounds such as glutathione (Box 11-B), mercaptoethanol, or dithiothreitol (Eq. 3-23) may be added. The crude **homogenate** may be strained and is usually centrifuged briefly to remove cell fragments and other "debris." Large-scale purification of

**TABLE 3-1**
**Practical p$K_a$ Values for Some Useful Buffer Compounds at 25°C and Ionic Strength 0.1[a]**

| Compound | p$K_a$ | Grams per mole | d(p$K_a$) / dT | Charge on conjugate base |
|---|---|---|---|---|
| Citric acid (p$K_1$) | | 192 | | −1 |
| Formic acid | 3.7 | 0 | −1 | |
| Citric acid (p$K_2$) | 4.45 | 192 | −0.0016 | −2 |
| Acetic acid | 4.64 | 60 | 0.0002 | −1 |
| Succinic acid (p$K_2$) | 5.28 | 118 | 0 | −2 |
| Citric acid (p$K_3$) | 5.80 | 192 | 0 | −3 |
| 3,3-Dimethylglutaric acid | 5.98 | 160 | 0.006 | −2 |
| Piperazine (p$K_1$) (5.68) | 6.02 | 86 | | 0 |
| Cacodylic acid (dimethylarsinic acid) | 6.1 | 138 | | −1 |
| MES[b] | 6.1 | 195 | −0.011 | |
| BIS-TRIS[b] | 6.41 | 209 | −.017 | |
| Carbonic acid (p$K_1$) | 6.4[c] | | | |
| Pyrophosphoric acid (p$K_3$) | 6.76 | 178 | | -2 |
| Phosphoric acid (p$K_2$) | 6.84 | 98 | −.0028 | −2 |
| PIPES[b] | 6.90 | 353 | −.0085 | −2 |
| Imidazole | 7.07 | 68 | −.020 | 0 |
| BES[b] | 7.06 | 213 | −.016 | −1 |
| Diethylmalonic acid | 7.2 | 136 | | |
| TES[b] | 7.37 | 279 | −.020 | −1 |
| HEPES[b] | 7.46 | 238 | −.014 | −1 |
| N-Ethylmorpholine | 7.79 | 115 | −.022 | 0 |
| Triethanolamine | 7.88 | 149 | −.020 | 0 |
| TRICINE[b] | 8.02 | 178 | −.021 | −1 |
| TRIS[b] | 8.16 | 121 | −.031 | 0 |
| Glycylglycine | 8.23 | 132 | −.028 | −1 |
| BICINE[b] | 8.26 | 163 | −.018 | −1 |
| 4-Phenolsulfonic acid | 8.70 | 174 | −.013 | −2 |
| Diethanolamine | 9.00 | 105 | −.024 | 0 |
| Ammonia | 9.2 | | −0.031 | 0 |
| Boric acid | 9.2 | | −0.008 | mixed |
| Pyrophosphoric acid (p$K_4$) | 9.41 | 178 | | -3 |
| Ethanolamine | 9.62 | 61 | −.029 | 0 |
| Glycine (p$K_2$) | 9.8 | 75 | −0.025 | −1 |
| Piperazine (p$K_2$) | 9.82 | 86 | | 0 |
| Carbonic acid (p$K_2$) | 10.0 | | −0.009 | −2 |
| Piperidine | 11.1 | 85 | | 0 |

[a] Based on compilation by Ellis and Morrison[9] with additional data from Good *et al.*[10,11] and Dawson *et al.* [12] The Good buffers have dipolar ionic constituents. Since no form is without electrically charged groups they are unlikely to enter and disrupt cells.

[b] Abbreviations used:

| | |
|---|---|
| BES | *N,N*-Bis(2-hydroxyethyl)-2-aminoethanesulfonic acid |
| BICINE | *N,N*-Bis(2-hydroxyethyl)glycine |
| BIS-TRIS | Bis(2-hydroxyethyl)iminotris(hydroxymethyl)methane |
| HEPES | *N*-2-Hydroxyethylpiperazine-*N′*-2-ethanesulfonic acid |
| MES | 2-(*N*-Morpholino)ethanesulfonic acid |
| PIPES | Piperazine-*N,N′*-bis(2-ethanesulfonic acid) |
| TES | *N*-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid |
| TRICINE | *N*-Tris(hydroxymethyl)methylglycine |
| TRIS | Tris (hydroxymethyl) aminomethane |

[c] For $CO_2$ (solid) + $H_2O \rightarrow H^+ + HCO_3^-$, apparent p$K_a$.

proteins is often initiated with such a crude homogenate.

Cell organelles are also often separated by centrifugation. In one procedure a homogenate in 0.25 M sucrose (**isotonic** with most cells) is centrifuged for 10 min at a field of 600–1000 times the force of gravity (600–1000 *g*) to sediment nuclei and whole cells. The supernatant fluid is then centrifuged another 10 min at ~10,000 *g* to sediment mitochondria and lysosomes. Finally, centrifugation at ~100,000 *g* for about an hour yields a pellet of microsomes (p. 14),[58] which contains both membrane fragments and ribosomes. Each of the separated components can be resuspended and recentrifuged to obtain cleaner preparations of the organelles. The sedimented particles can often be solubilized by chemical treatment, for example, by the addition of either ionic or nonionic detergents. Membrane proteins can be isolated following solubilization in this way (Chapter 8). The soluble supernatant fluid remaining after the highest speed centrifugation provides the starting material for isolation of soluble enzymes and many small molecules.

## 2. Separations Based on Molecular Size, Shape, and Density

The simplest way to separate very large dissolved molecules is to let the small ones pass through a suitable sieve which may be a membrane with holes or a bed of gel particles. If the size of the particles approaches that of the holes in the sieve, the rate of passage will depend upon shape as well as size.

***Dialysis, ultrafiltration, and perfusion chromatography.*** In dialysis[59] and ultrafiltration,[60] a thin membrane, e.g., made of cellulose acetate (cellophane) and containing holes 1–10 nm in diameter (typically 5 nm), is used as a semipermeable barrier. Small molecules pass through but large ones are retained. Dialysis depends upon diffusion and can be hastened by adequate stirring. Ultrafiltration requires a pressure difference across the membrane. The more sophisticated procedures of gel filtration and perfusion chromatography were introduced in 1959.[61–65] A column is packed with material such as the crosslinked dextran Sephadex, polyacrylamide gels (such as the Bio-Gel P Series), or agarose gels (e.g., Bio-Gels A and Sepharoses). These come in the form of soft beads, the interior network of which is a three-dimensional network of polymer strands (Fig. 4-10). Recently porous beads of hard crosslinked polystyrene, glass, or various other silicate materials have been employed.[66–68] The interstices between strands, whose size depends upon the degree of crosslinking introduced chemically into the gel, are small enough to exclude large molecules but to admit smaller ones. If a mixture of materials of different molecular size is passed through such a column the smaller molecules are retarded because of diffusion into the gel, while the larger molecules pass through unretarded (Fig. 3-3). Sephadex G-25 excludes all but salts and compounds no larger than a simple sugar ring. Sephadex G-200, which is much less crosslinked, permits separation of macromolecules in the range of 5–200 kDa. As is explained in Section B, gel filtration also provides an important way of estimating $M_r$ for proteins and other macromolecules.

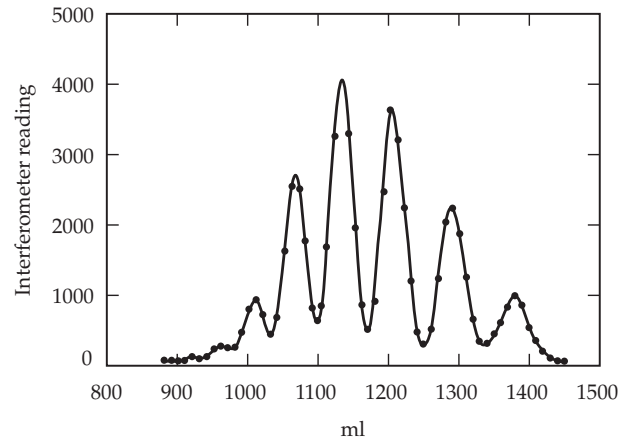



**Figure 3-3** Separation of oligosaccharides by gel filtration. The sugars dissolved in distilled water were passed through a column of Sephadex G-25. The peaks contain (right to left) glucose, cellobiose, cellotriose, etc. From Flodin and Aspberg.[64]

***Centrifugation.*** Centrifuges of many sizes and speeds are used in the laboratory to remove debris as well as to collect precipitated proteins and other materials at various steps in a purification scheme. The most remarkable are the **ultracentrifuges** which produce forces greater than $4 \times 10^5$ times that of gravity. They can be used both for separation of molecules and for determination of $M_r$ (see Section C).

When macromolecules in a solution are subjected to an ultracentrifugal field they are accelerated rapidly to a constant velocity of sedimentation. This is expressed as a **sedimentation constant *s***, which is the rate (cm/s) per unit of centrifugal force. The unit of *s* is the second but it is customary to give it in Svedberg units, S (1S= $10^{-13}s$). Sizes of particles are often cited by their S values. The sedimentation constant is affected by the sizes, shapes, and densities of the particles as is discussed further in Section C. If carried out at constant velocity an equilibrium will eventually be attained in which sedimentation is just balanced by diffusion and a smooth concentration gradient forms from the top to the bottom of the centrifuge cell or tube. Concentration gradients can also be formed by centrifuging a

concentrated solution of small molecules.[13,69,70] Such a concentration gradient is also a **density gradient** which can be made very steep. For example, a gradient with over a 10% increase in density from top to bottom can be created using 6 M **cesium chloride** (CsCl) and is widely used in DNA separations. If DNA is added prior to centrifugation it will come to rest in a narrow band or bands determined by the **buoyant densities** of the species of DNA present (see Chapter 5). After centrifugation, which is usually done in a plastic tube, a hypodermic needle is inserted through the bottom of the tube and the contents are pumped or allowed to flow by gravity into a fraction collector.

Another type of gradient centrifugation (**zone centrifugation**) utilizes a preformed gradient to stabilize bands of cell fragments, organelles, or macromolecules as they sediment.[71–74] For example, RNA can be separated into several fractions of differing sedimentation constants in a centrifuge tube that contains **sucrose** ranging in concentration from 25% at the bottom to 5% at the top. This is prepared by a special mixing device or "gradient maker" prior to centrifugation. The solution of RNA is carefully layered on the top, the tube is centrifuged at a high speed for several hours, and the different RNA fractions separate into slowly sedimenting sharp bands. A 20–60% gradient of sucrose or glycerol may be used in a similar way to separate organelles.[58]

## 3. Separations Based on Solubility

Some fibrous proteins are almost insoluble in water and everything else can be dissolved away. More often soluble proteins are precipitated from aqueous solutions by adjustment of the pH or by addition of large amounts of salts or of organic solvents. The solubility of any molecule is determined both by the forces that hold the molecules together in the solid state and by interactions with solvent molecules and with salts or other solutes that may be present.[58a]

Proteins usually have many positively and negatively charged groups on their surfaces. If either a positive or a negative charge predominates at a given pH the protein particles will tend to repell each other and to remain in solution. However, near the isoelectric point (pI), the pH at which the net charge is zero (see Section 7), the solubility will usually be at a minimum. The pH of a tissue extract may be adjusted carefully to the pI of a desired protein. Any protein that precipitates can be collected by centrifugation and redissolved to give a solution enriched in the protein sought. Some proteins, such as those classified (by an old system) as **globulins**, are insoluble in water but are readily **"salted in"** by addition of low concentrations (e.g., up to 0.1 M) of salts. Low concentrations of salts increase the solubility of most proteins because the salt ions interact with the charged groups on the protein surfaces and interfere with strong electrostatic forces that are often involved in binding protein molecules together in the solid state. Some salts, including $CaCl_2$ and NaSCN, which bind to proteins, are especially effective in salting in.[75,76] Addition of *high* concentrations of salt causes precipitation of most proteins from aqueous solutions. The most effective and most widely used materials for this **"salting out"** of proteins are $(NH_4)_2SO_4$ and $Na_2SO_4$. Because the salt ions interact so strongly with water, the protein molecules interact less with water and more with each other. A similar intramolecular effect may cause the stabilization of proteins of halophilic bacteria by 1–4 M KCl.[77]

Different proteins precipitate at different concentrations of an added salt. Hence, a fraction of proteins precipitating between two different concentrations of salt can be selected for further purification (Fig. 3-4). Protein concentrations can be estimated as described in Box 3-A.

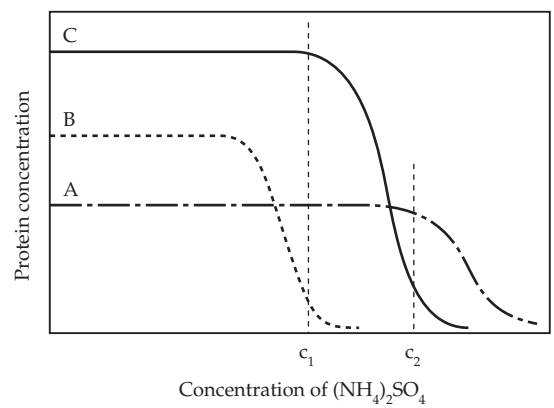



**Figure 3-4** Hypothetical behavior of a solution containing three proteins, A, B, and C, upon ammonium sulfate fractionation. The concentration of protein remaining in the solution is plotted against ammonium sulfate concentration (usually expressed as % saturation). Addition of ammonium sulfate to concentration $c_1$ will precipitate largely protein B, which can be removed by centrifugation. Addition of additional salt to $c_2$ will precipitate largely protein C, while A remains in solution.

Precipitation methods are popular first steps in purification of proteins because they can be carried out on a large "batch" scale. The amounts of ammonium sulfate used are often expressed as percentage saturation, i.e., as a percentage of the amount required to saturate the solution (4.1 M at 25°). Convenient tables are available[12,78] that permit one to weigh out the correct amount of solid ammonium sulfate to give a desired percentage saturation or to go from one

percentage saturation to a higher one.

Proteins are often stabilized by low concentrations of simple alcohols or ketones[76] and by higher concentrations of polyhydroxy alcohols, such as glycerol[77] and sucrose,[78] and also by certain inert, synthetic polymers such as **polyethyleneglycol** (PEG).[79] The latter is a widely used precipitant. The polyhydroxyalcohols and PEG are all hydrated but tend not to interact strongly with the protein molecules. On the other hand, simple alcohols may denature proteins by their interaction with nonpolar regions.[77]

## 4. Separation by Partition

Many of the most important separation methods are based on repeated equilibration of a material between two separate phases, at least one of which is usually liquid. Small molecules may be separated by **countercurrent distribution** in which a material is repeatedly partitioned between two immiscible liquid phases, one more polar than the other. New portions of both liquids are moved by a machine in a "countercurrent manner" between the equilibration steps or are moved continuously through coiled tubes.[80–82,82a]

A similar result is accomplished by using as one phase a solid powder or fine "beads" packed in a vertical column or spread in a thin layer on a plate of glass. The methods are usually referred to as **chromatography**, a term proposed by Tswett to describe separation of materials by color. In 1903 Tswett passed solutions of plant leaf pigments (chlorophylls and carotenes) in nonpolar solvents such as hexane through columns of alumina and of various other adsorbents and observed separation of colored bands which moved down the column as more solvent was passed through. Individual

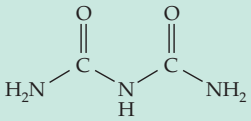
### BOX 3-A QUANTITATIVE ESTIMATION OF PROTEIN CONCENTRATIONS

Biochemists often need to estimate the content of protein in a sample. For example, in devising a purification procedure for an enzyme it is customary to estimate the number of units of enzyme activity (as defined in Chapter 9) per milligram of protein (U/ mg). As progress is made in the purification this ratio increases. It becomes constant with respect to additional purification attempts, when a homogeneous enzyme is obtained.

One of the most widely used and most sensitive protein assays (for 0.1–1 mg/ml of protein) is the colorimetric procedure of **Lowry**.[a–c] It makes use of a phosphomolybdic–phosphotungstic acid reagent (the Folin–Ciocalteu reagent) which is reduced by proteins in the presence of alkaline $Cu^{2+}$ to characteristic "blue oxides" whose color can be monitored at 750 nm. Much of the color comes from the reducing action of tyrosine and tryptophan. The color yield varies greatly from protein to protein and users may have trouble with reproducibility. A related method utilizes **bicinchoninic acid** which forms a purple color (measured at 362 nm) with the $Cu^{+1}$ that is formed by reduction of alkaline $Cu^{2+}$ by the protein.[d–f] This reagent is easier to use than that of the Lowry procedure and gives stable and reproducible readings.

A third widely used procedure, introduced by **Bradford**[g] and modified by others, measures the binding of the dye Coomassie brilliant blue whose peak absorption shifts from 465 nm to 595 nm upon binding. The change occurs within two minutes and is stable. However, the color yield varies from one protein to another.

Less sensitive but very simple and precise is measurement of the light absorption around 280 nm. This is discussed in the main text in Section D,6. For a typical protein an absorbance of 1.0 at 280 nm corresponds to a protein concentration of 1 mg / ml.[h] The very old **biuret method** is also useful for samples containing 1–10 mg / ml of protein. The violet color that arises upon addition of copper sulfate to an alkaline solution of a peptide or protein is recorded at 540–560 nm.[h] The color is especially intense for longer polypeptides. The name of the method arises from the fact that biuret gives a similar color[i] (see also Eq. 6-85).



Biuret

[a] Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275
[b] Peterson, G. L. (1979) *Anal. Biochem.* **100**, 201–220
[c] Larson, E., Howlett, B., and Jagendorf, A. (1986) *Anal. Biochem.* **155**, 243–248
[d] Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Kenk, D. C. (1985) *Anal. Biochem.* **150**, 76–85
[e] Davis, L. C., and Rodke, G. A. (1987) *Anal. Biochem.* **161**, 152–156
[f] Hill, H. D., and Straka, J. G. (1988) *Anal. Biochem.* **170**, 203–208
[g] Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
[h] Fruton, J. S., and Simmonds, S. (1958) *General Biochemistry*, 2nd ed., Wiley, New York (p. 130)
[i] Layne, E. (1957) *Methods Enzymol.* **III**, 450–451

pure pigments could be **eluted** from the column by continued passage of solvent. This important method is called **adsorption chromatography**. It is assumed that the pigments are absorbed on the surface rather than being disssolved in the solid material. A related method is **hydrophobic interaction chromatography** of proteins.[83–86] The packing material is similar to that described in Section 6 (Affinity Chromatography) but bears long-chain alkyl groups that can interact with the hydrophobic patches on surfaces of proteins. A very different adsorbent that is very useful in separation of proteins is carefully prepared microcrystalline **hydroxylapatite.**[87,88] It presumably functions in part by ion exchange.

Column packing materials such as **silica gel** contain a large amount of water, and separation involves partition between an immobilized aqueous phase in the gel and a mobile, often organic, solvent flowing through the column. Usually materials elute sooner when they are more soluble in the mobile phase than in the aqueous phase. These methods are closely related to perfusion chromatography, which is described in Section 2.

Aromatic amino acids, lipids, and many other materials can be separated on **reversed-phase** columns in which nonpolar groups, usually long-chain alkyl groups, are covalently attached to silica gel, alumina, or other inert materials.[66,80] The mobile phase is a more polar solvent, often aqueous, and gradually made less polar by addition of an organic solvent. In reversed-phase chromatography more polar compounds migrate faster through the system than do nonpolar materials, which experience hydrophobic interaction with the solid matrix.

Many traditional chromatographic methods including reversed-phase chromatography have been adapted for use in automatic systems which employ columns of very finely divided solid materials such as silica, alumina, or ion exchange materials coated onto fine glass beads.[89–91] These **high-performance liquid chromatographic** (**HPLC**) systems often utilize pressures as high as 300 atmospheres. Separations are often sharper and faster than with other chromatographic methods.[92–97] Reversed-phase columns in which the solid matrix may carry long (e.g. $C_{18}$) hydrocarbon chains have been especially popular for separation both of peptides and of proteins. Proteins may also be separated by gel filtration, ion exchange, or other procedures with HPLC equipment.

A sheet of high-quality filter paper containing adsorbed water serves as the stationary phase in **paper chromatography**. However, **thin-layer chromatography**, which employs a layer of silica gel or other material spread on a glass or plastic plate, has often supplanted paper chromatography because of its rapidity and sharp separations (Fig. 3-5).[16,96a,98–100] An approach that requires no stationary phase at all is **field flow fractionation**.[101] Here a suitable external field (e.g., electrical, magnetic, and centrifugal) or a thermal gradient is imposed on the particles flowing through a narrow channel.

For volatile materials **vapor phase chromatography** (gas chromatography) permits equilibration between the gas phase and immobilized liquids at relatively high temperatures. The formation of volatile derivatives, e.g., methyl esters or trimethylsilyl derivatives of sugars, extends the usefulness of the method.[103,104] A method which makes use of neither a gas nor a liquid as the mobile phase is **supercritical fluid chromatography**.[105] A gas above but close to its critical pressure and temperature serves as the solvent. The technique has advantages of high resolution, low temperatures, and ease of recovery of products. Carbon dioxide, $N_2O$, and xenon are suitable solvents.

## 5. Ion Exchange Chromatography

Separation of molecules that contain electrically charged groups is often accomplished best by ion exchange chromatography.[105a] The technique depends upon interactions between the charged groups of the molecules being separated and fixed ionic groups on

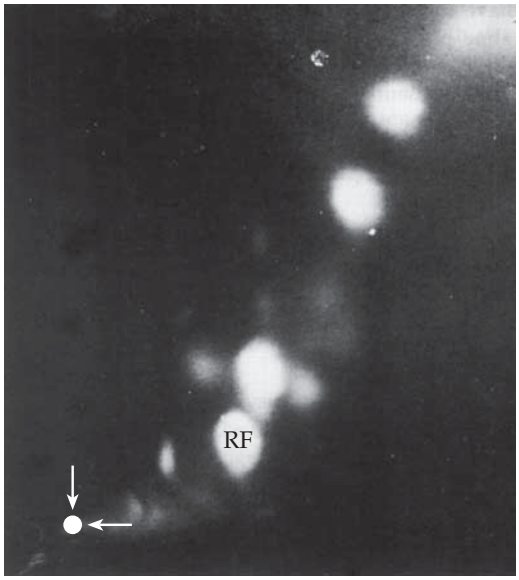



**Figure 3-5** Photograph of a two-dimensional thin layer (silica gel) chromatogram of a mixture of flavins formed by irradiation of ~10 μg of the vitamin riboflavin. The photograph was made by the fluorescence of the compounds under ultraviolet light. Some riboflavin (RF) remains. The arrows indicate the location of the sample spot before chromatography. Chromatography solvents: a mixture of acetic acid, 2-butanone, methanol, and benzene in one direction and *n*-butanol, acetic acid, and water in the other. See Treadwell *et al*.[102]

an immobile matrix. Separation depends upon small differences in p$K_a$ values and net charges and upon varying interactions of nonpolar parts of the molecules being separated with the matrix. Since changes in pH can affect both the charges on the molecules being separated and those of the ion exchange material, the affinities of the molecules being separated are strongly dependent on pH. For example, proteins and most amino acids are held tightly by cation exchangers at low pH but not at all at high pH.

Aqueous solutions are usually employed and the columns are packed with beads of **ion exchange resins**, porous materials containing bound ionic groups such as $-SO_3^-$, $-COO^-$, $-NH_3^+$, or quaternary nitrogen atoms. Synthetic resins based on a cross-linked polystyrene are usually employed for separation of small molecules. For larger molecules chemical derivatives of cellulose or of crosslinked dextrans (Sephadex), agarose, or polyacrylamide are more appropriate. Positively charged ions, such as amino acids in a low pH solution, are placed on a cation exchange resin such as Dowex 50, which contains dissociated sulfonic acid groups as well as counter ions such as $Na^+$, $K^+$, or $H^+$. The adsorbed amino acids are usually eluted with buffers of increasing pH containing sodium or lithium ions. The procedure, which was developed by Moore and Stein,[106–109] is widely used for automatic quantitative analysis of amino acid mixtures obtained by hydrolysis of a protein or peptide (Fig. 3-6).[110–113]

Ion exchange chromatography of proteins and peptides is often done with such ion exchange materials as carboxymethyl-Sephadex and phosphocellulose, which carry negatively charged side chains or diethylaminoethylcellulose (DEAE-cellulose), which carries positively charged amino groups.[114] These materials do not denature proteins or entrap them and have a large enough surface area to provide a reasonable absorptive capacity. The mobile phase is usually buffered. For anion exchangers the pH should be above ~4.4 to keep most carboxylate side chains on the proteins ionized. The pH may be increased to pH 7–10, where most histidine imidazolium ions have dissociated, increasing the mobility of many proteins. For cation exchange the pH is usually buffered below pH 6 or 7 (Fig. 3-6).[115,116]

## 6. Affinity Chromatography

In this technique the chromatographic absorbent is designed to make use of specific biochemical inter-

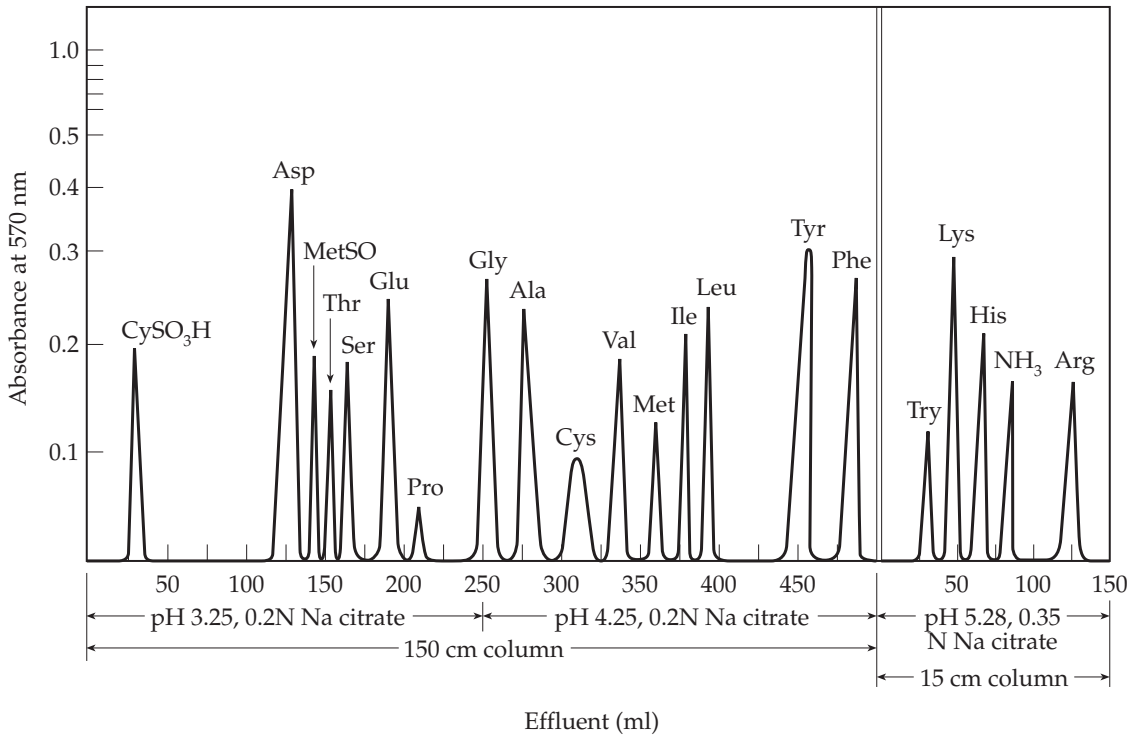
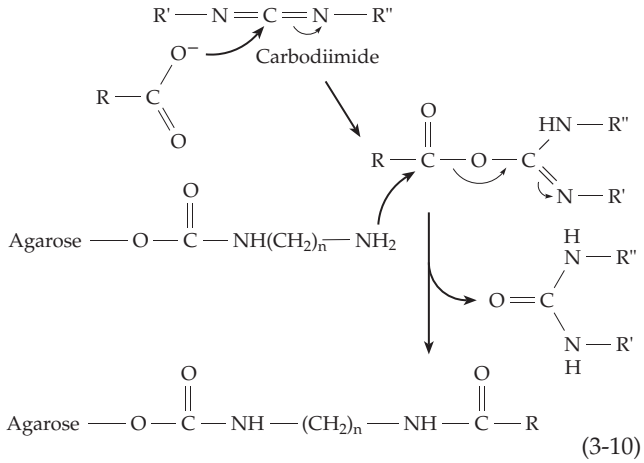



**Figure 3-6** Separation of amino acids by cation-exchange chromatography on a sulfonated polystyrene resin in the $Na^+$ form by the method of Moore and Stein.[110] The amino acids were detected by reaction with ninhydrin (Box 3-C); areas under the peaks are proportional to the amounts. Two buffers of successively higher pH are used to elute the amino acids from one column, while a still higher pH buffer is used to separate basic amino acids on a shorter column. From Robyt and White.[13]

actions to "hook" selectively a particular macromolecule or group of macromolecules.[117–122] Affinity chromatography is used in many ways, including the purification of enzymes, antibodies, and other proteins that bind tightly to specific small molecules.

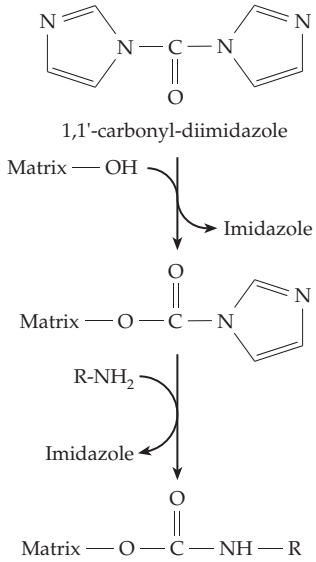
Because of their open gel structure (Fig. 4-10) agarose derivatives in bead form provide a good solid support matrix. The hydroxyl groups of the agarose are often linked to amino compounds. In one widely used procedure[119] the agarose is treated with cyanogen bromide (Br – C ≡N) in base to "activate" the carbohydrate (Eq. 3-9, step *a*). Then the amino compound is added (step *b*). The isourea product shown is the major one with agarose gels but dextran-based matrices tend to react by steps *c*–*e*.

Absorbents containing a large variety of R groups of shapes specifically designed to bind to the desired proteins can be made in this way. If the coupling is done with a diamine [R = $(CH_2)_n$—$NH_2$], the resulting ω-aminoalkyl agarose can be coupled with other compounds by reaction with **carbodiimide** (Eq. 3-10). For reaction in nonaqueous medium dicyclohexylcarbodiimide (R″ = cyclohexyl) is often used, but for linking groups to agarose a water-soluble reagent such as 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide is recommended.[120] Carbodiimides are widely used for forming amide or phosphodiester linkages in the laboratory. The formation of an amide with a side chain of ω-aminoalkyl agarose can be pictured as in Eq. 3-10.

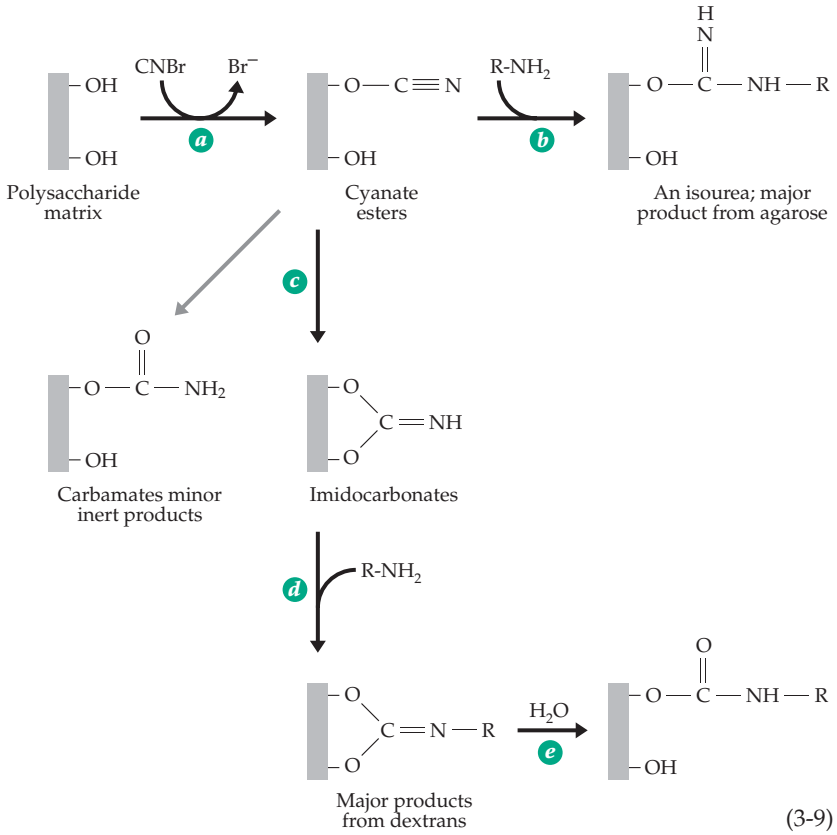



(3-10)

Many other means of preparation of adsorbents for affinity chromatography are also available.[121,122] For example, 1,1'-carbonyl-diimidazole can be used to couple a diamine to the matrix (Fig. 3-11). This reagent has the advantage that it does not depend upon the relatively unstable isourea linkages formed by Eq. 3-9 to hold the specific affinity ligands.[121]

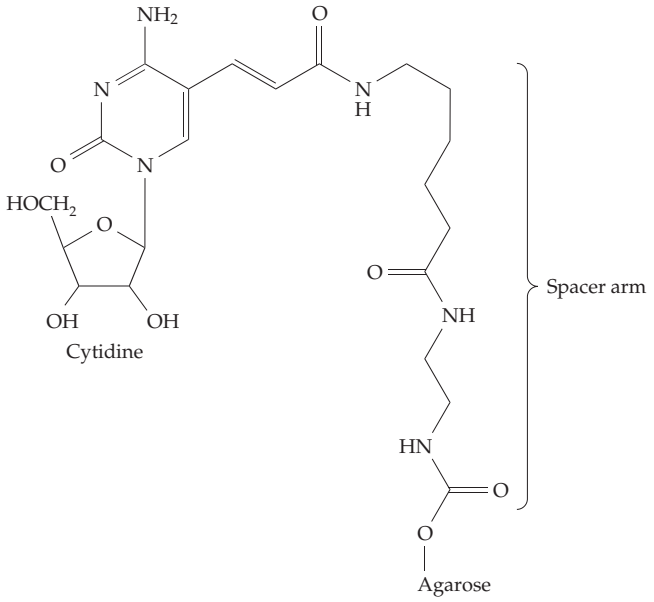



(3-11)




(3-9)

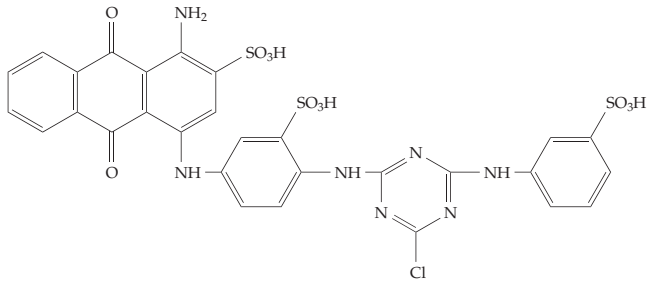
An example of a successful application of affinity chromatography is the isolation of the enzyme cytidine deaminase from cells of *E. coli*. Cytidine was linked covalently via long **spacer** arms to the agarose beads as in the following diagram:




A cell extract was subjected to ammonium sulfate fractionation and the dialyzed protein was then poured through the affinity column which held the cytidine deaminase molecules because of their affinity for the cytidine structures that were bound to the agarose. The protein was eluted with a borate buffer; the borate formed complexes with the adjacent hydroxyl groups of the cytidine and thereby released the protein. After passage of the protein through an additional column of DEAE-Sephadex the deaminase had been purified 1700-fold compared to the crude extract.[123]

Another technique is to engineer genes to place a polyhistidine "tag" at the C terminus of a protein chain. Commercial cloning vehicles and kits are available for this purpose.[124] The protein produced when the engineered gene is expressed can be captured by the affinity of the polyhistidine tag for $Cu^{2+}$, $Ni^{2+}$, $Co^{2+}$, or $Zn^{2+}$ held in chelated form on an affinity column.[124–127]

A surprising discovery was that certain dyes, for example Cibachron blue, when covalently coupled to a suitable matrix, often bind quite specifically to proteins that have a nonpolar binding pocket near a positive charge.[128] This includes many enzymes that act on nucleotides.




Cibacron blue F3GA

## 7. Electrophoresis and Isoelectric Focusing

The methods considered in this section make use of movement of molecules in an electrical field. Separation depends directly upon differences in the net charge carried by molecules at a fixed pH. The net charge for compounds containing various combinations of acidic and basic groups can be estimated by considering the $pK_a$ of each group and the extent to which that group is dissociated at the selected pH using Eqs. 3-3 to 3-5. At some pH, the **isoelectric point** (pI), a molecule will carry no net charge and will be immobile in an electric field. At any other pH it will move toward the anode (+) or cathode (–). The pH at which the protein carries no net charge in the complete absence of added electrolytes is called the **isoionic point**.[129]

**Electrophoresis**, the process of separating molecules, and even intact cells[130] (Box 3-B), by migration in an electrical field, is conducted in many ways.[28,131–140] In **zone electrophoresis**, a tiny sample of protein solution, e.g., of blood serum, is placed in a thin line on a piece of paper or cellulose acetate. The sheet is moistened with a buffer and electrical current is passed through it. An applied voltage of a few hundred volts across a 20-cm strip suffices to separate serum proteins in an hour. To hasten the process and to prevent diffusion of low-molecular-weight materials, a higher voltage may be used. Two to three thousand volts may be applied to a sample cooled by water-chilled plates. Large-scale electrophoretic separations may be conducted in beds of starch or of other gels.

One of the most popular and sensitive methods for separation of proteins is electrophoresis in a column filled with **polyacrylamide** or **agarose gel** or on a thin layer of gel on a plate. The method depends upon both electrical charge and molecular size and has been referred to as **electrophoretic molecular sieving**.[28,133,135–137,141–143] Polyacrylamide gel electrophoresis is often carried out in the presence of ~1% of the denaturing detergent sodium docecyl sulfate which coats the polypeptide chain rather evenly. This method, which is often referred to as **SDS–PAGE,** has the advantage of breaking up complex proteins composed of more than one subunit and sorting the resultant monomeric polypeptide chains according to molecular mass (see Box 3-C). A disulfide-reducing reagent (see Eq. 3-23) such as ~1% 2-mercaptoethanol is usually present but may be omitted to permit detection of crosslinked peptides.

**Capillary electrophoresis** is increasingly popular[144–149a] and can be used to separate attomole amounts.[150] It can be used not only for separation of proteins but also for rapid estimation of the net charge on a protein.[151] The separation is conducted in tubes with internal diameters as small as 10–15 μm and as short as 1 cm. Multiple channels cut into a glass chip

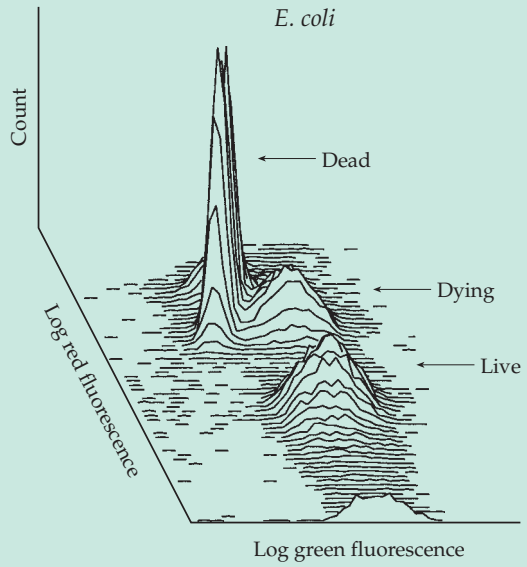
## BOX 3-B SORTING AND ANALYZING SINGLE CELLS

It is often important to examine and analyze individual cells.[a] For example, large numbers of single blood cells can be tested for the presence of specific antigenic determinants that arise by mutation. This permits assessment of the frequency of these mutations.[b] The complex chemical processing of neuropeptides can be studied on the contents of a single neuron (Chapter 30) using mass spectrometry.[c]

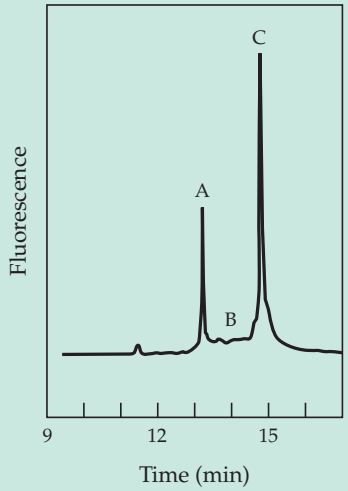
Several methods for separating cells have been devised. These include electrophoresis[d] or use of magnetic microspheres.[b] Micromanipulation can sometimes be used to select single cells for analysis. The most impressive technique is **flow cytometry**,[e,f] which is used daily on human blood samples in clinical laboratories. A suspension of cells is passed at a high rate of flow through a narrow capillary of ~0.2 mM diameter. The sample stream, which is surrounded by a larger "sheath" stream, has a smaller diameter of ~20 µm. One or more laser beams are used to record information about each cell over a period of a few microseconds as the cells pass by at a rate of as much as $10^5$ cells / s.

Flow cytometers developed from simpler cell counters, but now they are used to record cell size (from light scattering), optical absorbance, fluorescence, and phosphorescence. The optical properties are often enhanced by staining. The use of two dyes that fluoresce at different wavelengths permits the construction of two-dimensional plots as in the accompanying figure.

Capillary electrophoresis is one of the techniques able to separate constituents of single cells and is illustrated in the second figure.




Flow cytometric histogram of fluorescently labeled live and dead *E. coli* bacteria. The dye kit (BacLightTM) that was used stains membrane-compromised dead bacteria with a red fluorescing dye and live bacteria with a green fluorescing dye. Cells were analyzed on an EPICS XL cytometer (Coulter Corporation). By integrating the area under each population it was possible to discern the percentage of dead (54), dying (16), and live (30) bacteria within a mixed population. The use of fluorescent dyes such as these has proven useful for studying various mechanisms employed by the food industry for killing microorganisms in food products and for studying a variety of bacteria derived from seawater, soil, plant materials, and laboratory-grown cultures. Courtesy of Kristi Harkins.




Electropherogram of major proteins from a single erythrocyte. Peaks A, B, and C are carbonic anhydrase (~7 amol), methemoglobin (~ 5 amol), and hemoglobin $A_0$ (~ 450 amol), as identified from migration times relative to standards.[a] Courtesy of Edward S. Yeung.


[a] Yeung, E. S. (1994) *Acc. Chem. Res.* **27**, 409–414

[b] Jovin, T. M., and Arndt-Jovin, D. J. (1980) *Trends Biochem. Sci.* **5**, 214–219

[c] Li, K. W., Hoek, R. M., Smith, F., Jiménez, C. R., van der Schors, R. C., van Veelen, P. A., Chen, S., van der Greef, J., Parish, D. C., Benjamin, P. R., and Geraerts, W. P. M. (1994) *J. Biol. Chem.* **269**, 30288–30292

[d] Bauer, J. (1994) *Cell Electrophoresis*, CRC Press, Boca Raton, Florida

[e] Shapiro, H. M. (1995) *Practical Flow Cytometry*, 3rd ed., Wiley-Liss, New York

[f] Darzynkiewicz, Z., Robinson, J. P., and Crissman, H. A., eds. (1994) *Flow Cytometry*, 2nd ed., Academic Press, San Diego

can be used.[152]

Whereas in conventional zone electrophoresis most of the electrical current is carried by the buffer, in **isotachophoresis**[153,154] the ions being separated carry most of the current. In **isoelectric focusing**,[28,155–157] a pH gradient is developed electrochemically in a vertical column or on a thin horizontal plate between an anode and a cathode. The pH gradient in a column is stabilized by the presence of a density gradient, often formed with sucrose, and the apparatus is maintained at a very constant temperature. Proteins within the column migrate in one direction or the other until they reach the pH of the isoelectric point where they carry no net charge and are "focused" into a narrow band. As little as 0.01 pH unit may separate two adjacent protein bands which are located at positions corresponding to their isoelectric points. A newer development is the use of very narrow pH gradients that are immobilized on a polyacrylamide matrix.[158–160] With this technique some hemoglobin mutants differing only in substitution of one neutral amino acid for another have been separated.[161] Special techniques are needed for highly basic proteins.[162]

A two-dimensional method in which proteins are separated by isoelectric focusing (preferably with an immobilized pH gradient) in the first dimension and by SDS-gel electrophoresis in the second has become a popular and spectacularly successful method for studying complex mixtures of proteins (Box 3-C).[163–166] Over 2000 proteins can be separated on a single plate. A similar procedure but without SDS can be used to examine undenatured proteins.[167–169] Computer-assisted methods are being developed to catalog the thousands of proteins being identified in this way[170–172] and also to allow rapid identification of spots by mass spectrometry. The technique can be applied to intact proteins in subpicomole quantities, even in whole cell lysates,[150,173,173a] or an enzyme such as trypsin can be used to cut the proteins into pieces on the gel plate and the mixtures of peptides can be analyzed by mass spectrometry.[174–176] Capillary electrophoresis or capillary isoelectric focusing can be applied before samples are sent to the mass spectrometer.

## C. Determining the Relative Molecular Mass, $M_r$

The evaluation of $M_r$ is often of critical importance. Minimum values of $M_r$ can often be computed from the content of a minor constituent, e.g., the tryptophan of a protein or the iron of hemoglobin. However, physicochemical techniques provide the basis for most measurements.[177] Observations of osmotic pressure or light scattering can also be used and provide determinations of $M_r$ that are simple in principle, but which have pitfalls.[178]

### 1. Ultracentrifugation

Some of the most reliable methods for determining $M_r$ depend upon **analytical ultracentrifuges**. These instruments, capable of generating a centrifugal field as much as 4 x $10^5$ times that of gravity, were developed in the 1920s and 1930s by T. Svedburg and associates in Uppsala, Sweden.[179–181] Driven by oil turbines, the instruments were expensive and difficult to use, but by 1948 a reliable electrically driven machine, the Beckman Model E ultracentrifuge, came into widespread use. It has had a major impact on our understanding of proteins, on methods of purification of proteins, and on our understanding of interactions of protein molecules with each other and with small molecules.[182,183] Nevertheless, it was not until 1990 that a truly "user-friendly" analytical untracentrifuge became available.[180,184–186] The Beckman Model XL-A centrifuge has a very small rotor driven by an air-cooled induction motor and is computer controlled. Data are recorded automatically in digital form and computer programs are available to carry out the necessary computations. The instrument can record ultraviolet-visible spectra at multiple radial positions in the sample cell (Fig. 3-7).

A straightforward determination of $M_r$ is obtained by centrifuging until an equilibrium distribution of the molecules of a protein or other macromolecular material is obtained and by recording the variation in concentration from the center to the periphery of the centrifuge cell[177,183,185,187–190] (see also Section A,2). Using short cells, this **sedimentation equilibrium** can be attained in 1–5 hours instead of the 1–2 days needed with older instruments. For a single component system the concentration distribution at equilibrium is given by Eq. 3-12.

$$c(r) = c(a) \exp\left[M\,(1-\bar{v}\rho)\,\omega^2\,(r^2-a^2)\,/\,2RT\right] \qquad (3\text{-}12)$$

Here $c(r)$ is the concentration c at the radial position $r$ (measured from the centrifuge axis), $a$ is the radial distance of the meniscus, $M$ is the molecular mass in daltons, and $\bar{v}$ is the partial specific volume in ml/gram. For most proteins $\bar{v}$ varies from 0.69–0.75. It is the reciprocal of the density of the particle. Rho (ρ) is the density (g/ml) of the solvent. A plot of log $c(r)$ against $r^2$ is a straight line of slope $M\,(1-\bar{v}\rho)\,/\,2RT$. The computer can also accommodate mixtures of proteins of differing molecular masses, interacting mixtures, etc.[185,191]

***Sedimentation velocity.*** The relative molecular mass $M_r$ can also be measured from observation of the velocity of movement of the boundary (or boundaries for multicomponent systems) between solution and solvent from which the macromolecules have sedi-

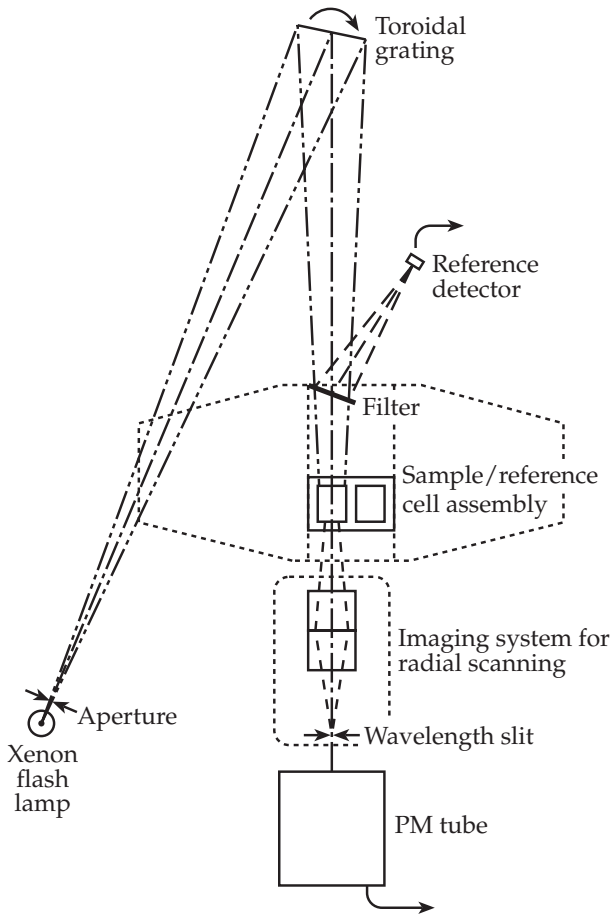



**Figure 3-7** The scanning absorption optical system of the Beckman Optima™ XL-A ultracentrifuge. Courtesy of Beckman Coulter.

mented. This boundary, which can be visualized by optical methods, is quite sharp initially, but it broadens with time, because of diffusion, as the macromolecules sediment.

A molecule in a centrifuge is acted upon not only by the applied centrifugal force but also by an opposing **buoyant force** that depends upon the difference in density of the sedimenting particles and the solvent and by a frictional drag, which is proportional to a **frictional coefficient** $f$. Setting the sum of these forces to zero for the hydrodynamic steady state yields Eq. 3-13, which defines the **sedimentation constant** $s$.

$$s = \frac{v}{\omega^2 r} = \frac{M(1-\bar{v}\rho)}{N f} \qquad (3\text{-}13)$$

Here $f$ is the frictional coefficient which is difficult to predict or to measure but is often assumed to be the same as the frictional coefficient that affects **diffusion**. It can be obtained from the diffusion coefficient $D$ using Eq. 3-14.

$$D(\text{cm}^2\ \text{s}^{-1}) = k_B T/f = RT/N f \qquad (3\text{-}14)$$

By combining Eqs. 3-13 and 3-14 we obtain the Svedberg equation:

$$M = \frac{RTs}{D(1-\bar{v}\rho)} \qquad (3\text{-}15)$$

Here $R$ is in the cgs (cm-gram-second) unit of 8.31 x $10^7$ erg $\text{mol}^{-1}\ \text{deg}^{-1}$. Using this equation the relative molecular mass $M_r$, which is numerically the same as $M$, can be evaluated from the sedimentation constant $s$. Since $s$, $D$, and $\bar{v}$ must all be measured with care, the method is demanding. It is often necessary to measure $s$ and $D$ at several concentrations and to extrapolate to infinite dilution. It is also customary to correct the data to give the values $s^\circ_{20,w}$ and $D^\circ_{20,w}$ expected at 20°C in pure water at infinite dilution.

To the extent that we can regard protein molecules as spherical we can substitute for $f$ in Eq. 3-13 the frictional coefficient of a sphere:

$$f_{\text{sphere}} = 6\pi\eta r_h \qquad (3\text{-}16)$$

Here $r_h$ is the hydrated radius or **Stokes radius** of the protein. On this assumption $s$ will be expected to increase with the relative molecular mass approximately as $M_r^{2/3}$. A plot of log $s$ against log $M_r$ should be a straight line. Figure 3-8 shows such a plot for a number of proteins. The plots for nucleic acids, which can often be approximated as rods rather than spheres, fall on a different line from those of proteins. Furthermore, the sedimentation constant falls off more rapidly with increasing molecular mass than it should for spheres.

From analysis of a variety of well-characterized proteins, Squire and Himmel[192] observed that if proteins are assumed to contain 0.53 g $H_2O$ per gram of protein and to have a mean value for $\bar{v}$ of 0.730 g / $\text{cm}^3$ the value of $M_r$ can be predicted by Eq. 3-17 with the standard deviation indicated. Here, S is the sedimentation constant in Svedberg units ($10^{-13}$s).

$$M_r = 6850\ S^{3/2}\ )\ 0.090\ M_r \qquad (3\text{-}17)$$

For proteins with various values of $\bar{v}$, Eq. 3-18 applies.

$$M_r = 922\ [S/(1-\bar{v}\rho)]^{3/2}\ ) \pm 0.066\ M_r \qquad (3\text{-}18)$$

## BOX 3-C ISOTOPES IN BIOCHEMICAL INVESTIGATIONS

Both stable[a] and radioactive[b–e] isotopes are widely used in chemical and biological investigations. The study of metabolism was revolutionized by the introduction of isotopic tracers. In one of the first biological experiments with the stable isotope $^{15}N$ (detected by mass spectrometry), Schoenheimer and associates in 1937 established the previously unsuspected turnover of protein in living tissues (Chapter 24, Section B). In 1937 Ruben *et al.* reported the uptake of radioactive $^{11}CO_2$ by plants.[f] A few years later Calvin and associates first traced the pathway of carbon in photosynthesis using the much longer lived $^{14}CO_2$ (Box 17-F). Wood and Werkman, in 1941, employed the stable isotope $^{13}C$ in studies of bacterial and mammalian metabolism (Box 17-C). The radioactive $^{32}P$ and $^{35}S$ have served to elucidate the metabolism of phosphorus and sulfur. Titrium ($^{3}H$) has been used to label many organic substances including thymine, which has been used extensively in the study of nucleic acid metabolism. Radioactive isotopes provide the basis for sensitive analytical procedures such as **radioimmunoassays** of minute quantities of hormones (Box 31-D). Through **radioautography** these isotopes facilitate numerous analytical procedures (see accompanying photo) and have provided the basis for important end-group methods used in sequence determination of polynucleotides (Eq. 5-24).

Several isotopes used in biochemistry are listed in the following table. For each radioactive isotope, the half-life is given, as is the type of particle emitted, and the energy of the particle. Gamma rays, such as those given off in decay of $^{125}I$ or $^{131}I$, are very penetrating and easy to count precisely, as is the energetic β radiation from $^{32}P$. On the other hand, $^{3}H$ (tritium) is relatively difficult to detect[g] but its weak β particle, which can travel only a short distance through a sample, makes it uniquely suitable for radioautography on a microscopic scale. Positrons ($\beta^+$) travel some distance, e.g., up to a few millimeters in the case of $^{13}N$. They are then destroyed by reacting with an electron to produce a pair of γ rays of energy 0.511 MeV, equal to the sum of the rest masses of an electron plus a positron. The half-life (Eq. 9-4) determines the isotopic abundance needed to achieve a given radiation rate, a practical matter in providing a sufficient rate of decay to permit counting with an acceptably low statistical error. Even very short-lived isotopes such as $^{13}N$, have proved useful as tracers.[h] The amount of an isotope giving $3.7 \times 10^{10}$ disintegrations per second (this is 1 g of pure radium, 0.3 mg of $^{3}H$, or 0.22 g of $^{14}C$) is known as the **curie** (Ci). One millicurie (mCi) provides $2.22 \times 10^{9}$ disintegrations / min (dpm). Radiolabeled substances ordinarily contain only a small fraction of the unstable isotope together with a larger number of unlabeled molecules. Compounds are usually sold in millicurie or microcurie quantities and with a stated specific activity as mCi $mmol^{-1}$. For example, a compound labeled at a single position with $^{3}H$ and having a specific activity of 50 mCi $mmol^{-1}$ would contain about 0.17% $^{3}H$ at that position.

| Isotope | Half-life | Maximum energy of radiation (MeV) β | γ |
|---|---|---|---|
| $^{2}H$ (deuterium) | Stable | | |
| $^{3}H$ (tritium) | 12.26 years | 0.018 | |
| $^{11}C$ | 20.4 min | | ($\beta^+$) |
| | 0.511 | | |
| $^{13}C$ | Stable | | |
| $^{14}C$ | 5730 years | 0.156 | |
| $^{13}N$ | 9.96 min | 1.2 | ($\beta^+$) |
| | 0.511 | | |
| $^{15}N$ | Stable | | |
| $^{15}O$ | 20.4 min | | ($\beta^+$) |
| | 0.511 | | |
| $^{18}O$ | Stable | | |
| $^{18}F$ | 110 min | | ($\beta^+$) |
| | 0.511 | | |
| $^{22}Na$ | 2.6 years | 0.55 | |
| | 1.28 | | |
| $^{32}P$ | 14.3 days | 1.71 | |
| $^{35}S$ | 87.2 days | 0.167 | |
| $^{36}Cl$ | $3 \times 10^{5}$ years | 0.716 | |
| $^{40}K$ | $1.3 \times 10^{9}$ years | 1.4 | |
| | 1.5 | | |
| $^{45}Ca$ | 165 days | 0.26 | |
| $^{59}Fe$ | 45 days | 0.46 | |
| | 1.1 | | |
| $^{65}Zn$ | 250 days | 0.32 | |
| | 1.14 | | |
| $^{90}Sr$ | 29 years | 0.54 | |
| $^{125}I$ | 60 days | 0.036 | |
| $^{131}I$ | 8.06 days | 0.61 | |
| | 0.36 | | |

Because of the development of new NMR techniques and improvements in mass spectrometry stable isotopes, such as $^{2}H$, $^{13}C$, $^{19}F$, and $^{31}P$, are being used more frequently to study metabolism.[j] Carbon-13 containing compounds can fulfill many

## BOX 3-C ISOTOPES IN BIOCHEMICAL INVESTIGATIONS (continued)

Radioautogram showing the separation of proteins of *E. coli* labeled with $^{14}C$ amino acids. From O'Farrell.[i] Twenty-five µl of sample containing 180,000 cpm and ~10 µg of protein were subjected to isoelectric focusing in a 2.5 x 130-mm tube containing polyacrylamide gel to separate proteins according to isoelectric point. The gel was then extruded from the column and was placed on one edge of a slab of polyacrylamide gel. Then SDS electrophoresis in the second dimension separated the proteins according to size. Over 1000 spots could be seen in the original radioautogram, which was obtained by placing a piece of photographic film over the gel slab and exposing it to the radiation for 875 hours. For details see O'Farrell.[i]

of the same tracer functions as $^{14}C$. Even the radioactive $^{3}H$ nucleus can be utilized for *in vivo* NMR.[k,l] Although radioisotope labeling is very sensitive, it gives little information unless compounds are isolated and laboriously degraded to determine the positions of the labels. NMR spectroscopy is less sensitive but can give direct chemical information about the positions of $^{13}C$ in compounds within living cells. A compound containing only $^{13}C$ in one or in many positions can be safely administered to human individuals as well as to other organisms and spectra of products that arise can be observed. High-resolution deuterium NMR spectroscopy has been used to follow $^{2}H$ incorporated at C-1, C-2, or C-6 positions in glucose.[m]

As a result of metabolic reactions an isotope may appear at more than one position in a product, yielding two or more isotope isomers or **isotopomers**. These are seen individually by NMR spectroscopy and the concentration and isotope labeling patterns of the labeled compounds can be followed over a period of time. The use of this **isotopomer analysis** in studies of the citric acid cycle is illustrated in Box 17-C and its use in studies of glucose metabolism is considered in Chapter 17, Section L.

A change in isotopic mass, especially from $^{1}H$ to $^{2}H$ or $^{3}H$, often produces a strong effect on reaction rates and the study of **kinetic isotope effects** has provided many insights into the mechanisms of enzymatically catalyzed reactions. Isotopes have permitted a detailed understanding of the stereochemistry of enzymatic reactions, an impressive example being the synthesis and use of chiral acetate (Chapter 13)[n] and chiral phosphate groups (Chapter 12). Specific isotopic properties provide the basis for NMR (Section G).


[a] Matwiyoff, N. A., and Ott, D. G. (1973) *Science* **181**, 1125–1132
[b] Wang, C. H., Willis, D. L., and Loveland, W. D. (1975) *Radiotracer Methodology in the Biological, Environmental, and Physical Sciences*, Prentice-Hall, Englewood Cliffs, New Jersey
[c] Wang, Y., ed. (1969) *Handbook of Radioactive Nuclides*, CRC Press, Cleveland, Ohio
[d] Thornburn, C. C. (1972) *Isotopes and Radiation in Biology*, Butterworth, London
[e] Slater, R. J., ed. (1990) *Radioisotopes in Biology: A Practical Approach*, IRL Press, Oxford
[f] Ruben, S., Hassid, W. Z., and Kamen, M. D. (1939) *J. Am. Chem. Soc.* **61**, 661–663
[g] Bransome, J., ed. (1970) *Liquid Scintillation Counting*, Grune & Stratton, New York
[h] Cooper, A. J. L. (1985) *Adv. Enzymol.* **57**, 251–356
[i] O'Farrell, P. H. (1975) *J. Biol. Chem.* **250**, 4007–4021
[j] Wolfe, R. R. (1992) *Radioactive and Stable Isotope Tracers in Biomedicine*, Wiley, New York
[k] Newmark, R. D., Un, S., Williams, P. G., Carson, P. J., Morimoto, H., and Klein, M. P. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 583–587
[l] Bergerat, A., Guschlbauer, W., and Fazakerley, G. V. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6394–6397
[m] Aguayo, J. B., Gamcsik, M. P., and Dick, J. D. (1988) *J. Biol. Chem.* **263**, 19552–19557
[n] Cornforth, J. W. (1976) *Science* **193**, 121–125

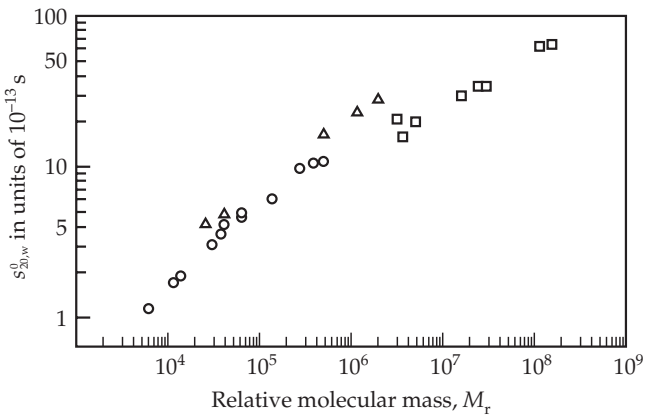



**Figure 3-8** Plots of the logarithm of the sedimentation constant *s* against the logarithm of the molecular weight for a series of proteins and nucleic acids: (○) globular proteins, (△) RNA, and (□) DNA. Proteins include a lipase (milk), cytochrome *c*, ribonuclease (pancreatic), lysozyme (egg white), follicle-stimulating hormone, bacterial proteases, human hemoglobin, prothrombin (bovine), malate dehydrogenase, γ-globulin (horse), tryptophanase (*E. coli*), glutamate dehydrogenase (chicken), and cytochrome *a*. Double-stranded DNA molecules are those of bacteriophage ϕX174 (replicative form), $T_7$, $\lambda_{b2}$, $T_2$, and $T_4$, and that of a papilloma virus. The RNA molecules are tRNA, rRNA, and mRNA of *E. coli*, and that of turnip yellow mosaic virus.[190a,191a]

## 2. Gel Filtration and Gel Electrophoresis

Several newer methods of molecular mass determination were developed in the 1960s–1980s. One is gel filtration. A column of gel beads such as Sephadex is prepared carefully and is calibrated by passing a series of protein solutions through it. The volume $V_e$ at which a protein peak emerges from the column can be expressed as the sum of two terms (Eq. 3-19) in

$$V_e = V_o + \sigma V_i \tag{3-19}$$

which $V_o$ is the **void volume**, i.e., the elution volume that is observed for very large particles that are completely excluded from the gel, and $V_i$ is the internal volume within the beads of gel. The value of σ is inversely related to the diffusion constant *D*, which for a spherical particle *D* is related by Eq. 3-20 to the Stokes radius $r_h$. This equation comes directly from Eqs. 3-16 and 3-14.

$$r_h = \frac{k_B T}{6\pi\eta D} \tag{3-20}$$

This suggests a proportionality between σ and the molecular radius. In fact, Eq. 3-21, in which a and b are constants provides a fairly good approximation for σ and $V_e$ is correlated approximately with log $M_r$ as shown in Fig. 3-9.[193,194]

$$\sigma = a \log r_h + b \tag{3-21}$$

A series of reference proteins of known molecular masses are used to calibrate the column and $M_r$ for an unknown protein is estimated from its position on the graph.[195,196] Another modification of the method depends upon chromatography in a high concentration of the denaturing salt guanidinium chloride. The assumption is made that proteins are denatured into random coil conformations in this solvent.[196]

Probably the most widely used method for determining the molecular mass of protein subunits is gel electrophoresis in the presence of the denaturing detergent sodium dodecyl sulfate (SDS). The protein molecules are not only denatured but also all appear to become more or less evenly coated with detergent.[197] The resulting rodlike molecules usually show a uniform dependence of electrophoretic mobility on molecular mass (plotted as log $M_r$). An example is shown in Fig. 3-10. Again, the molecular mass of the protein under investigation is estimated by comparison of its rate of migration with that of a series of marker proteins.[195,198]

## 3. Mass Spectrometry

Mass spectrometry has played a role in biochemistry since the early 1940s when it was introduced for use in following isotopic labels during metabolism.[199–200c] However, it was not until the 1990s that suitable commercial instruments were developed to permit mass spectrometry using two new methods of ionization. The techniques are called **matrix-assisted laser desorption / ionization time-of-flight (MALDI-TOF)** and **electrospray ionization (ESI)** mass spectrometry.

In the MALDI technique a pulsed laser beam strikes a solid sample and heats, vaporizes, and ionizes compounds with little decomposition.[201–209] Proteins or other biopolymers are mixed with a "matrix" that absorbs the heat of the laser beam. The protein sample together with the matrix is dried. Most proteins form crystals and the laser beam is directed toward individual protein crystals or aggregates. Various materials are used for the matrix. Compounds as simple as glycerol, succinic acid, or urea can be used with an infrared laser. For proteins an ultraviolet nitrogen laser tuned to 337 nm is usually employed with an ultraviolet light-absorbing matrix such as hydroxybenzoic acid, 2,5-dihydroxybenzoic acid, α-hydroxy-

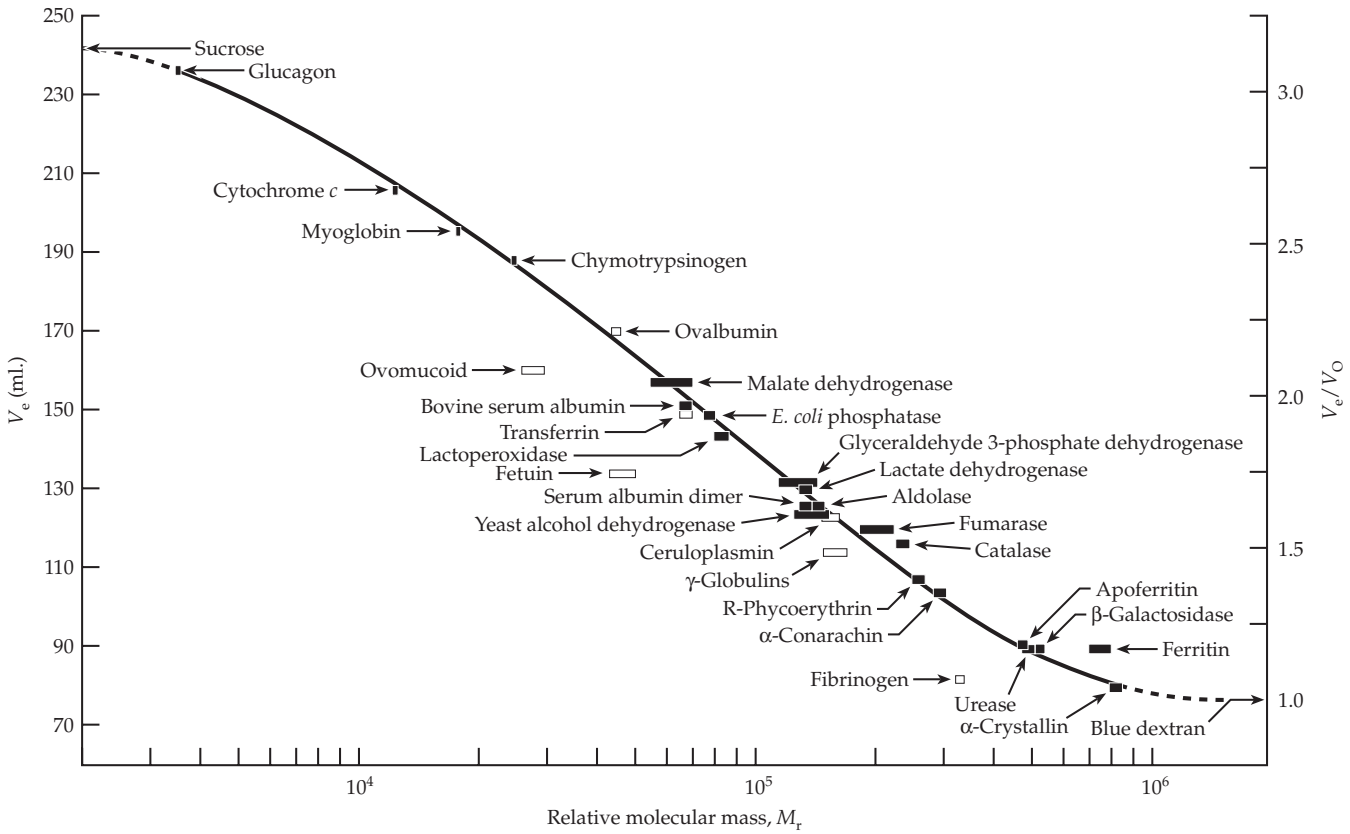



**Figure 3-9** Elution volume of various proteins on a column of Sephadex G-200 as a function of molecular mass. The right-hand vertical axis shows the ratio of the elution volumes to that of blue dextran, a high-molecular-mass polysaccharide that is excluded from the internal volume. After Andrews.[193]

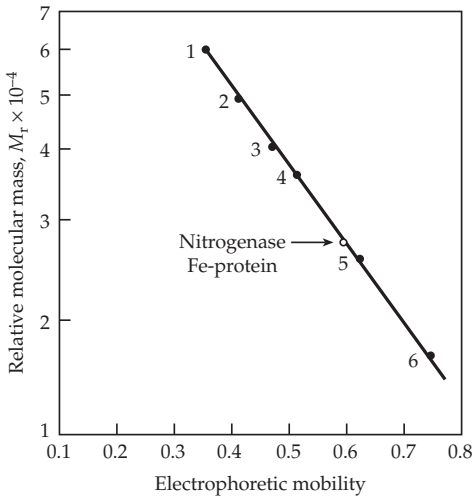



**Figure 3-10** Estimation of the molecular mass of the polypeptide chain of the nitrogenase Fe-protein using SDS-polyacrylamide electrophoresis; from a set of four standard curves. The marker proteins are (1) catalase, (2) fumarase, (3) aldolase, (4) glyceraldehyde-phosphate dehydrogenase, (5) α-chymotrypsinogen A, and (6) myoglobin. (o) indicates position of azoferredoxin. From Nakos and Mortenson.[195]

cinnamic acid, or sinapinic acid (Chapter 25). The matrix ionizes, desorbs from the surface, and transfers energy to the crystalline protein, causing it to ionize and desorb from the surface. Oligosaccharides and oligonucleotides can be ionized in a similar way.

MALDI spectra are relatively simple (Fig. 3-11), often containing a single major peak corresponding to the singly charged molecular ion $[M + H]^+$ of mass $m + 1$ and perhaps a doubly charged molecular ion $[M + 2H]^{2+}$ of mass $(m + 2)/2$. For oligomeric proteins the major peak is often that of the monomer with weaker peaks for oligomers. The instrument can also be adjusted to generate negative ions whose detection is useful for study of phosphorylated peptides, many oligosaccharides, and oligonucleotides. With a TOF spectrometer there is no upper limit to the mass range and masses of over 100 kDa can be measured to about ±0.1%. Femtomole quantities can be detected.

The MALDI method is especially useful for complex mixtures of peptides and can be utilized in peptide sequencing. The technique is also appropriate for studying mixtures of glycoproteins. Negative-ion MALDI can be applied to oligonucleotide mixtures. Further improvements in resolution in both MALDI

and ESI methods are anticipated as a result of developement of Fourier transform mass spectrometers.[202]

In ESI mass spectrometry[201,203–205,210–213] the sample, dissolved in an appropriate solvent (usually a 50:50 mixture of methanol and water for proteins), is infused directly into the ionization chamber of the spectrometer through a fused silica capillary. At the end of the capillary the solution is subjected to electrical stress created by a voltage difference of about 5 kV between the electrospray needle and the sampling orifice (the counter-electrode). The process results in the formation of singly and / or multiply charged molecular ions which are guided into the analyzer for mass analysis. For proteins every arginine, lysine, and histidine may bind a hydrogen ion to form a variety of positive ions. A 100-kDa protein may easily bind 100 protons bringing

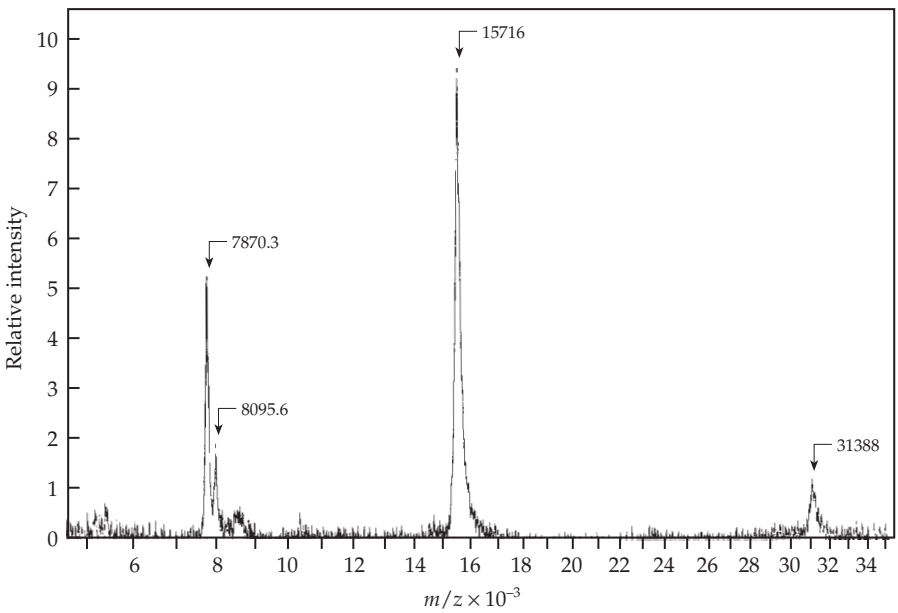



**Figure 3-11** Matrix-assisted laser desorption / ionization time-of-flight (MALDI-TOF) mass spectrum of bovine erythrocyte Cu-Zn superoxide dismutase averaged over ten shots with background smoothing. One-half µl of solution containing 10 pmol of the enzyme in 5 mM ammonium bicarbonate was mixed with 0.5 µl of 50 mM α-cyanohydroxycinnamic acid dissolved in 30% (v / v) of acetonitrile-0.1% (v / v) of trifluoroacetic acid. The mixture was dried at 37° C before analysis. The spectrum shows a dimer of molecular mass of 31,388 Da, singly charged and doubly charged molecular ions at 15,716, and 7870 Da, respectively. The unidentified ion at mass 8095.6 may represent an adduct of the matrix with the doubly charged molecular ion. Courtesy of Louisa Tabatabai.

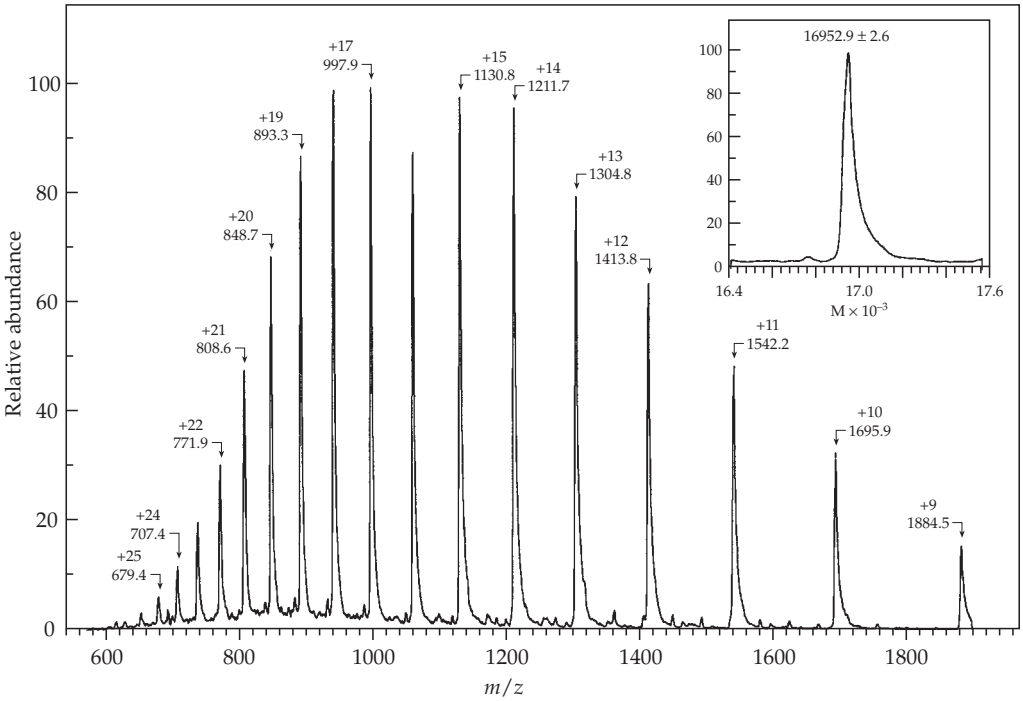



**Figure 3-12** Positive ion electrospray mass spectrum of horse apomyoglobin ($M_r$ 16,950.4). The net charge on each ion as well as the mass to charge ratio $m/z$ is indicated at the top of each peak. The inset shows a computer "deconvolution" of the spectrum with the calculated value of molecular mass. Courtesy of Kamel Harrata.

the $m/z$ ratio for the fully charged protein to 1000, well below the maximum $m/z$ ratio of ~2400 for a typical quadrupole mass spectrometer. Since not all basic groups are protonated the spectra consist of families of peaks of differing $m/z$ (Fig. 3-12). For a single pure protein the molecular mass can be calculated from the ratio of $m/z$ values from any pair of adjacent peaks. It is better, especially if there is a mixture of proteins, to use a computer to extract $M_r$.[214] The accuracy can be quite high, typically ±0.01%: one mass unit in 10,000.

Some complexity arises from the fact that each carbon atom in a protein contains about 1% of $^{13}C$. This means that for a protein of mass >10 kDa there will be a confusing array of peaks in the mass spectrum and it may be difficult to pick out the relatively minor "monoisotopic" peak that arises from molecules containing only $^{12}C$, $^{1}H$, $^{14}N$, $^{16}O$, and $^{32}S$. In fact, the peak representing the most abundant mass will be a few mass units higher than the monoisotopic peak.[215,216] (see Study Question 15). New computer programs have been devised to assist in the analysis. Use of **$^{13}C$ and $^{15}N$-depleted nutrients** also extends the applicability of mass spectrometry.[217]

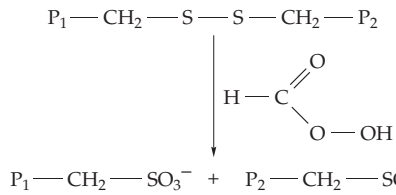
Electrospray mass spectrometry utilizes a soft ionization technique at nearly atmospheric pressure. As a result, intact molecular ions are formed in high yield. The instrument can be interfaced readily to HPLC or capillary electrophoresis columns and sub-femtomole amounts of proteins can be detected. A disadvantage is that salt concentrations must be kept low (< mM) and that the protein tends to bind $Na^+$, $K^+$, and anions that may confuse interpretation of spectra.

## D. Determining Amino Acid Composition and Sequence

When they are first isolated, proteins are usually characterized by $M_r$, isoelectric point, and other easily measured properties. Among these is the amino acid composition[112] which can be determined by completely hydrolyzing the protein to the free amino acids. Later, it is important to establish the primary structure or amino acid sequence.[51] This has been accomplished traditionally by cutting the peptide chain into smaller pieces that can be characterized easily. However, most protein sequences are now deduced initially from the corresponding DNA sequences, but further chemical characterization is often needed.

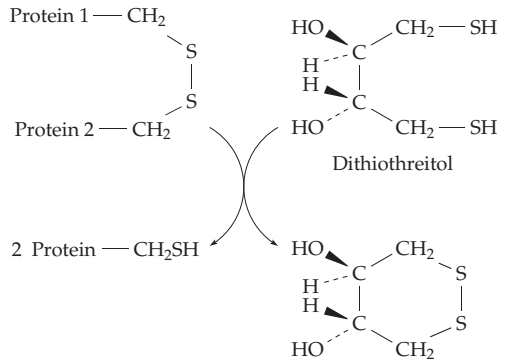
### 1. Cleavage of Disulfide Bridges

Before a polypeptide chain can be degraded it is usually necessary to break any disulfide bridges.[218–220] For some proteins such as the keratins of hair, these linkages must be broken even to get the protein into solution. Oxidation with performic acid (Eq. 3-22) has been used on ribonuclease but is not often employed because the performic acid also oxidizes tryptophan residues.




(3-22)

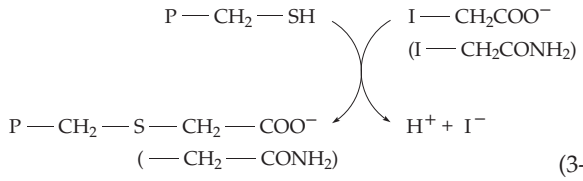
Reduction by **dithiothreitol** or **dithioerythritol** (Eq. 3-23) is usually successful.




(3-23)

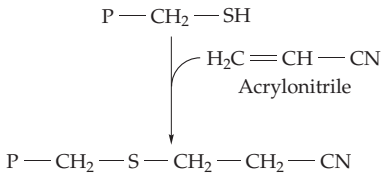
Upon oxidation these dithiols cyclize to form stable disulfides, driving the reaction to completion. These same compounds are also widely used to protect SH groups in enzymes against accidental oxidation by oxygen and to dissolve highly crosslinked insoluble proteins. Mercaptoethanol, $HS–CH_2–CH_2–OH$, may be used for the same purposes but requires higher concentrations and has a disagreeable odor.

Once cleaved, disulfide bridges may be prevented from reforming by conversion of the resulting thiol groups to stable derivatives, e.g., with iodoacetate and iodacetamide (Eq. 3-24).



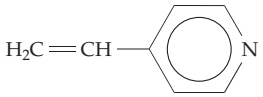

(3-24)

A better reagent is acrylonitrile (Eq. 3-25).




(3-25)

In modern sequencing methods vinylpyridine, which reacts in a similar way, is often used. It can be detected during amino acid analysis or sequencing after derivatization with phenylisothiocyanate (Eq. 3-30).



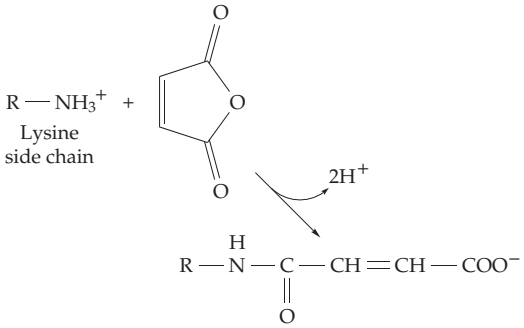

Vinylpyridine

## 2. Hydrolysis and Other Chain Cleavage Reactions

Most biopolymers are inherently unstable with respect to cleavage to monomer units by reaction with water. Hydrolysis can be catalyzed by protons, by hydroxyl ions, or by the protein-hydrolyzing enzymes that are discussed in Chapter 12. Complete hydrolysis of proteins is usually accomplished by heating under nitrogen with 6 M HCl at 150°C for 65 min. Some amino acids, especially tryptophan, are destroyed and the amides in the side chains of asparagine and glutamine are converted to the free acids. Some peptide linkages such as Val-Val are very resistant and tend to be incompletely hydrolyzed. No procedure has been found which gives the ideal complete hydrolysis. Use of 4 M methanesulfonic acid containing 3-(2-aminoethyl)indole instead of 6 M HCl gives less decomposition of tryptophan.[221] Base-catalyzed hydrolysis of proteins also gives good yields of tryptophan but causes extensive racemization of amino acids.[222]

Complete enzymatic digestion of proteins can be accomplished with a mixture of enzymes including proteases produced by fungi (Pronase). However, the enzymes attack each other, making quantitative analysis difficult. The problem can be circumvented by immobilizing the hydrolytic enzymes in a column of agarose gel. The protein to be hydrolyzed is passed through the gel and the constituent amino acids emerge from the bottom of the column.[223,224]

***Selective enzymatic hydrolysis.*** The traditional strategy in sequence determination is to cut protein chains into smaller pieces which can be separated by chromatography or electrophoresis and sequenced individually. Enzymatic cleavage is especially useful because of its specificity. **Trypsin**, a so-called **endopeptidase**, cleaves peptide chains at a rapid rate only if the carbonyl group of the amide linkage cleaved is contributed by one of the basic amino acids lysine, arginine, or aminoethylcysteine (see Fig. 12-10). If the protein is treated with maleic anhydride (Eq. 3-26),




(3-26)

### BOX 3-D THE PROTEOME

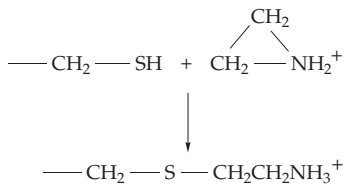
The ability to separate rapidly and detect minute amounts of proteins has spawned a new concept: The **proteome** is envisioned as a record of all proteins being actively synthesized by a cell – or of all genes being actively "expressed."[a–c] The concept evolved from efforts to automate the cataloging of spots on two-dimensional gels such as that shown in Box 3-B where each spot represents a single protein. The ability to unambiguously identify the spots by mass spectroscopy has brought new optimism to the attempt to use gels automatically to analyze all of the proteins formed by a cell.[a,c] The methods are potentially very important to developing new diagnostic procedures for human medicine. Watching changes in the proteome, including posttranslational modifications in proteins, as cells develop and grow will provide new insights into biochemical regulation.

A similar concept is that of a **complete transcriptional map**, a record of all of the different RNA molecules being synthesized by a cell.[d,e] These include many different mRNAs, each of which may give rise to more than one protein, as well as many RNAs with other functions.

---

[a] Kahn, P. (1995) *Science* **270**, 369–370
[b] Swinbanks, D. (1995) *Nature (London)* **378**, 653
[c] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko.A, Boucherie, H., and Mann, M. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14440–14445
[d] Richard, G.-F., Fairhead, C., and Dujon, B. (1997) *J. Mol. Biol.* **268**, 303–321
[e] Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., and Davis, R. W. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10614–10619

the lysine residues are protected and trypsin will cleave only at the Arg-X positions. If the resultant peptides are separated and held at pH 3.5 overnight, the blocking groups are hydrolyzed off and a second trypsin treatment can be used to cleave at the Lys-X positions.[225] The number of cleavage sites for trypsin can be increased by converting an –SH group to a positively charged one by aminoethylation (Eq. 3-27).[226] The reaction can be accomplished either with ethyleneimine (caution: carcinogen) as shown in this equation or by bromoethylamine, which eliminates $Br^-$ to form ethyleneimine.




(3-27)

Because of its specificity for basic residues, trypsin converts a protein into a relatively small number of **tryptic peptides** which may be separated and characterized. Trypsin acts primarily on denaturated proteins, and to obtain good results the disulfide bridges must be broken first. **Chymotrypsin** is less specific than trypsin and **pepsin** is even less specific (Table 3-2). Nevertheless, they can be used to cut a peptide chain into smaller fragments whose sequences can be determined. To establish the complete amino acid sequence for a protein, "overlapping" peptide fragments must be found that contain sequences from ends of two different tryptic fragments. In this way the tryptic peptides can be placed in the order in which they occurred in the native protein. This tedious procedure is rarely used today. Peptide sequencing is still important but is usually coordinated with gene sequencing, X-ray structure determination, or mass spectroscopy which minimize the need for overlapping fragments.

While trypsin cuts the peptide linkages Lys-X and Arg-X, a fungal protease cleaves only X-Lys.[227] A protease from the submaxillary glands of mice cleaves only Arg-X,[228] one from *Staphylococcus* specifically at Glu-X,[229,230] and one from kidneys at Pro-X.[231]

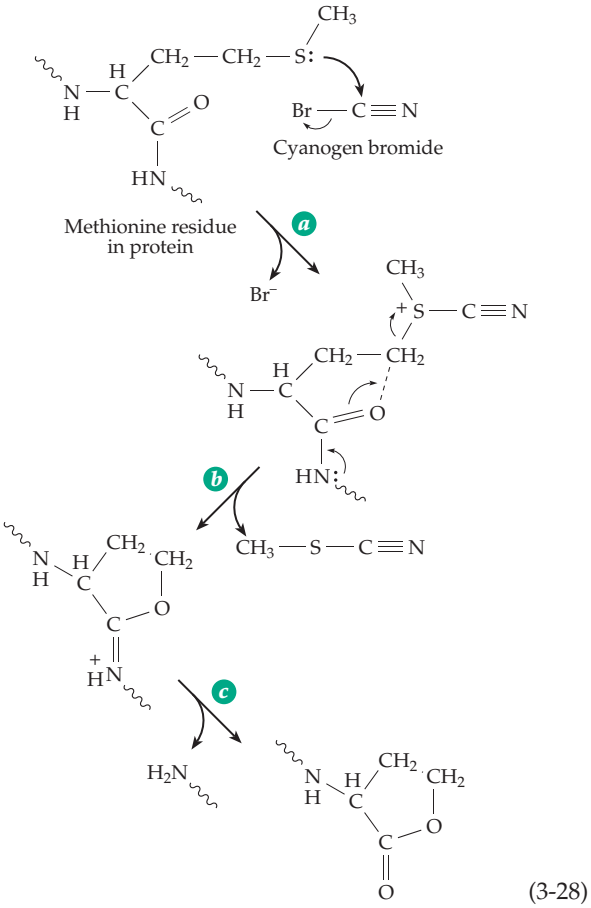
Several enzymes catalyze stepwise removal of amino acids from one or the other end of a peptide chain. **Carboxypeptidases**[232] remove amino acids from the carboxyl-terminal end, while **aminopeptidases** attack the opposite end. Using chromatographic methods, the amino acids released by these enzymes may be examined at various times and some idea of the sequence of amino acids at the chain ends may be obtained. A **dipeptidyl aminopeptidase** from bovine spleen cuts dipeptides one at a time from the amino terminus of a chain. These can be converted to volatile trimethylsilyl derivatives and identified by mass spectrometry.[233] If the chain is shortened by one residue using the Edman degradation (Section 3) and the dipeptidyl aminopeptidase is again used, a different set of dipeptides that overlaps the first will be obtained and a sequence can be deduced. Carboxypeptidase Y can be used with MALDI mass spectrometry to deduce the C-terminal amino acid sequence for a peptide. However, Ile and Leu cannot be distinquished.

**TABLE 3-2**
**Specificities of Commonly Used Protein-Hydrolyzing Enzymes**

| | | |
|---|---|---|
| Trypsin | Lys-X, Arg-X | X not Pro |
| Chymotrypsin | | |
| rapidly: | Phe-X, Tyr-X, Trp-X | X not Pro |
| slowly: | Y-X<br>Y=Leu, Asn, Gln, His, Met, Ser, Thr | X not Pro |
| *Staphylococcus aureus* protease V-8 | Glu-X | X not Pro |
| Clostripain | Arg-X | |
| Pepsin | | |
| preferentially: | X-Phe-X, X-Tyr-X, X-Leu-X | |
| less so: | X-Ala-X | |
| Thermolysin | | |
| rapidly: | X-Y<br>Y=Ile, Leu, Val, Ala, Phe, Met | |
| slowly: | X-Y<br>Y=Tyr, Gly, Thr, Ser | |

***Nonenzymatic cleavages.*** Of the various nonenzymatic methods that have been proposed, one has been outstandingly useful. Cyanogen bromide, N ≡ C – Br, cleaves peptide chains adjacent to methionine residues. The sulfur of methionine displaces the bromide ion (Eq. 3-28) and because of a favorable spatial relationship, the resulting sulfonium compound undergoes C–S bond cleavage through participation of the adjacent peptide group (Eq. 3-28, step *b*). The C=N of the product is then hydrolyzed with cleavage of the peptide chain in step *c*.

The linkage Asp-Gly can often be cleaved specifically by treatment with hydroxylamine at high pH.[195] Procedures for specific cleavage of tryptophanyl bonds have been de-




(3-28)

vised.[234] The Asp-Pro linkage is susceptible to cleavage by trifluoroacetic acid, which is used in the automated Edman degradation employed in peptide sequencing (Eq. 3-30). Cleavage by trifluoroacetic acid can be used to generate peptides for subsequent sequence determination.

***Separating the peptides.*** A procedure that has been very important in the development of protein chemistry is **peptide mapping** or "fingerprinting." The procedure begins with cleavage of the disulfide linkages, denaturation, and digestion with trypsin or some other protease. The sizes and amino acid compositions of the resulting series of peptides are characteristic of the protein under study. The mixture of peptides is placed on a thin layer plate and subjected to chromatography in one direction, then to electrophoresis in the other direction, with the peptides separating into a characteristic pattern or fingerprint. Fingerprinting has been especially useful in searching for small differences in protein structure, for example, between genetic variants of the same protein (Fig. 7-27). Currently, the peptides are usually separated on ion exchange or gel filtration columns, by reversed-phase HPLC, or by capillary electrophoresis and are then often passed, in subpicomole amounts, into a mass spectrometer.[176,235,236]

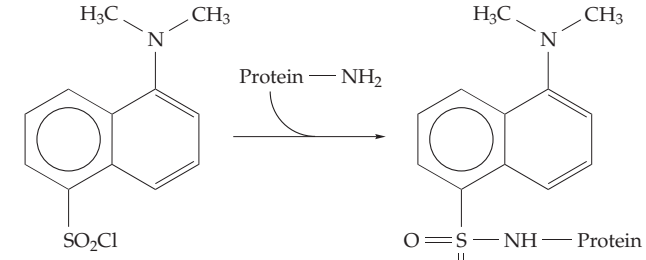
## 3. Determining Amino Acid Sequence

The covalent structure of insulin was established by Frederick Sanger in 1953 after a 10-year effort. This was the first protein sequence determination.[237,238] Sanger used partial hydrolysis of peptide chains whose amino groups had been labeled by reaction with 2,4-dinitrofluorobenzene[239] to form shorter end-labeled fragments. These were analyzed for their amino acid composition and labeled and hydrolyzed again as necessary. Many peptides had to be analyzed to deduce the sequence of the 21-residue and 30-residue chains that are joined by disulfide linkages in insulin.[237,238]

The Sanger method is mainly of historic interest, although end-labeling may still be used for various purposes. A more sensitive labeling reagent than was used by Sanger is **dansyl chloride**. It reacts to form a sulfonamide linkage that is stable to acid hydrolysis and is brilliantly fluorescent (Eq. 3-29). The related reagent dimethylaminoazobenzene-4'-sulfonyl chloride gives highly colored derivatives easily seen on thin-layer chromatography plates.[240]

***The Edman degradation.*** One of the most important reagents for sequence analysis is **phenylisothiocyanate,** whose use was developed by P. Edman.[241–243] This reagent also reacts with the N-terminal amino group of peptides (Eq. 3-30, step *a*). The resulting adduct undergoes cyclization with cleavage of the peptide linkage (Eq. 3-30, step *b*) under acidic conditions. After rearrangement (step *c*) the resulting **phenylthiohydantoin** of the N-terminal amino acid can be identified. The procedure can then be repeated on the shortened peptide chain to identify the amino acid residue in the second position. With careful work the Edman degradation can be carried down the chain for several tens of residues.

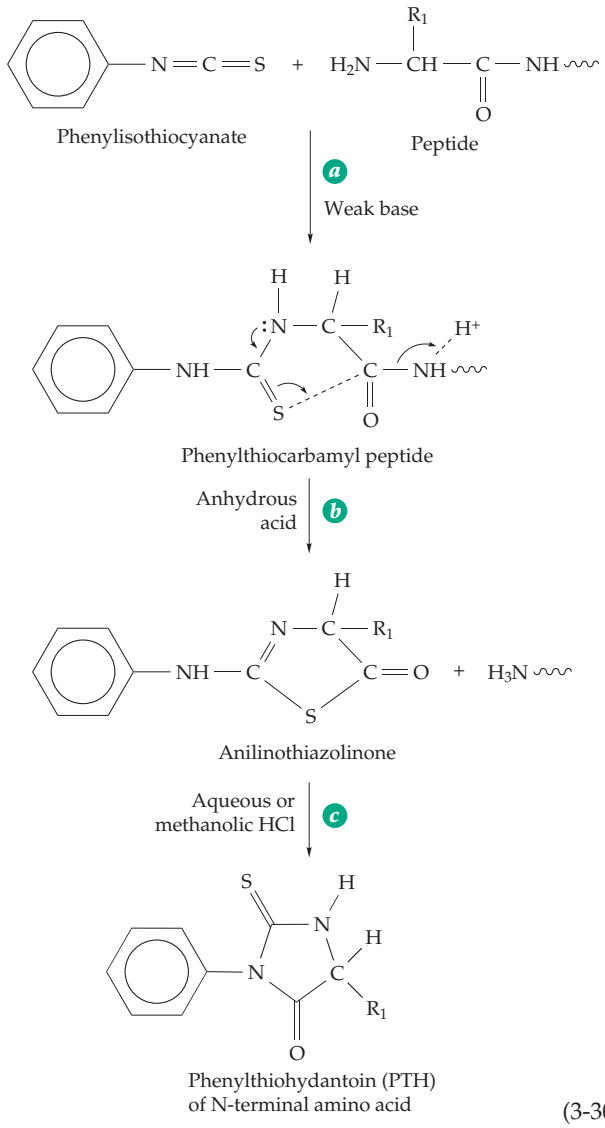
Ingenious **protein sequenators** have been devised to carry out the Edman degradation automatically.[242,244–246] Each released phenylthiohydantoin is then identified by HPLC or other techniques. Commercial sequenators have often required 5–20 nmol of peptide but new microsequenators can be used with amounts as low as 5–10 picomoles or less.[247,248]




(3-29)

Microsequencers permit sequence analysis on minute amounts of protein. Microsequencing can be used in conjunction with two-dimensional electrophoretic separations of proteins such as that shown in Box 3-C. The proteins in the polyacrylamide gel are electrophoretically transferred onto a porous sheet (membrane) of an inert material such as polyvinyl difluoride.[249–251] After staining, a selected spot is cut out and placed into the sequencer. To avoid the problems associated with blocked N termini, the protein may be treated with proteases on the membrane and the resulting peptide fragments may then be separated on a narrow-bore HPLC column and sequenced.[240]

Because many proteins are modified at the N terminus, blocking application of the Edman degradation, it would be useful to have a similar method for sequencing from the C terminus. It has been difficult to devise a suitable strategy, but there has been some success.[252–254]




(3-30)

***Protein sequences from the genes.*** Complete sequences of large numbers of genes have been determined and the corresponding sequences of proteins can be read directly from those of the corresponding genes. One method for sequencing a gene is to isolate a specific messenger RNA, which does not contain intervening sequences. A DNA copy (cDNA) is made from the mRNA and is used to ascertain the sequence of the encoded protein. The genomic DNA is also often sequenced. Introns are recognized by the nucleotide sequences at their ends and the correct amino acid sequence for the encoded protein is deduced.

In many instances, however, a gene can be identified only after part of the protein, often an N-terminal portion has been sequenced. This knowledge permits synthesis of an **oligonucleotide probe** that can be used to locate the gene (Chapter 5). Nucleotide sequences can be verified by comparison with sequences of tryptic or other fragments of a protein. Similarly, protein sequences are often checked by sequencing the corresponding genes as well as by study of X-ray structures. Substantial numbers of errors are made in sequencing of both DNA and protein so that checking is important.

***Mass spectrometry in sequencing.*** Proteins can also be sequenced by mass spectrometry or by a combination of Edman degradation and mass spectrometry.[213,255,255a] Until recently the peptides had to be converted to volatile derivatives by extensive methylation and acetylation or by other procedures. However, newer ionization methods including MALDI (Fig. 3-11) and ESI (Fig. 3-12) have made it possible to obtain mass spectra on unmodified peptides. In one procedure a nonspecific protease cleaves a peptide chain into a mixture of small oligopeptides which are separated by HPLC into 20–40 fractions, each of which may contain 10–15 peptides but which can be sent directly into the ionization chamber of the mass spectrometer.[256] Peptides can be generated from a protein using immobilized enzymes, separated on a chromatographic column, and introduced sequentially into the mass spectrometer. Examination of peptide mixtures by mass spectrometry provides a way of verifying sequences deduced from DNA sequencing.[257] Mass spectrometry is also used widely to study covalently modified proteins.[173,216,257a,257b] As a rule, these cannot be recognized from gene sequences.

## 4. Locating Disulfide Bridges

A final step in sequencing is often the location of S–S bridges. The reduced and alkylated protein can be cleaved enzymatically (e.g., with elastase, pepsin, or thermolysin) to relatively small fragments, each of which contains no more than one modified cysteine.

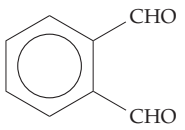
The same enzymatic cleavage can then be applied to the unreduced enzyme. Pairs of peptide fragments remain linked by the S–S bridges. These crosslinked pairs can be separated, the disulfide bridges cleaved, and the resulting peptides identified, each as one of the already sequenced fragments. Mass spectrometry provides a rapid method for their identification.[258]

Another elegant way of locating S–S bridges employs **diagonal electrophoresis**. Electrophoresis of the digest containing the crosslinked pairs is conducted in one direction on a sheet of filter paper. Then the paper is exposed to performic acid vapor to cleave the bridges according to Eq. 3-22 and electrophoresis is conducted in the second direction and the paper is sprayed with ninhydrin. The spots falling off the diagonal are those that participated in S–S bridge formation. They can be associated in pairs from their positions on the paper and can be identified with peptides characterized during standard sequencing procedures.[259]

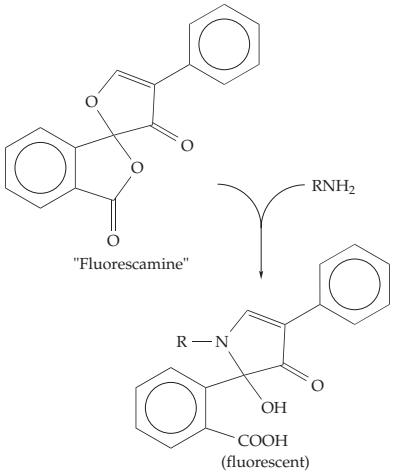
Diagonal electrophoresis and its relative diagonal chromatography are useful for other purposes as well. After electrophoresis or chromatography is conducted in one direction, the paper or thin-layer plate may be sprayed with a reagent that will react with some components or may be irradiated with light before the separation is repeated in the second direction (Fig. 3-5).[102,260]

## 5. Detecting Products

Important to almost all biochemical activity is the ability to detect, and to measure quantitatively, tiny amounts of specific compounds. "Color reagents," which develop characteristic colors with specific compounds, are especially popular. For example, **ninhydrin** (Box 3-E) can be used as a "spray reagent" to detect a small fraction of a micromole of an amino acid or peptide in a spot on a chromatogram. It can also be used for a quantitative determination, the color being developed in a solution. More sensitive than absorption of light (color) is fluorescence. **Fluorescamine** (Eq. 3-31) reacts with any primary amine to form a highly fluorescent product. As little as 50 pmol of amino acid can be determined quantitatively.[261] A yet more sensitive fluorogenic reagent for detection of amino acids, peptides, and amines of all types is *o*-phthaldialdehyde.[262,263]

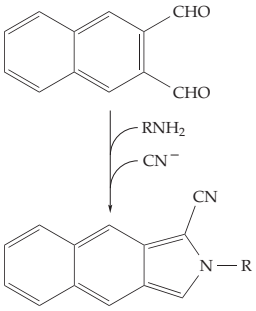



*o*-Phthaldialdehyde



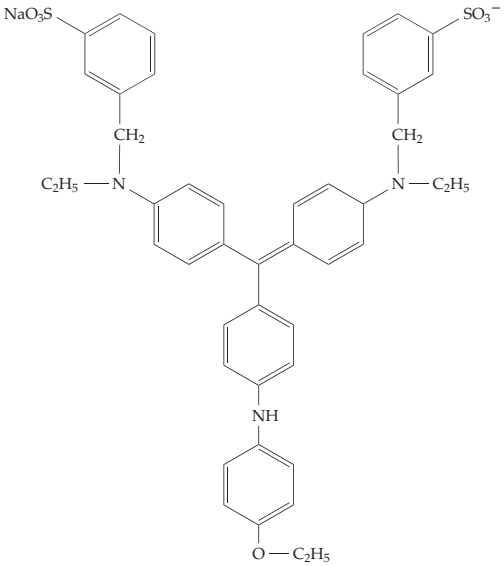

(3-31)

Reaction with naphthalene 2,3-dicarboxaldehyde (Eq. 3-32) increases the limit of detection 100-fold or more.[264]




(3-32)

Detection of proteins on thin-layer plates, gel slabs, or membranes is often accomplished by staining with a dye,[265–267] the most widely used being Coomassie brilliant blue.[268] Various silver-containing stains may also be used. After separation of a protein mixture by electrophoresis and transfer to an inert membrane,

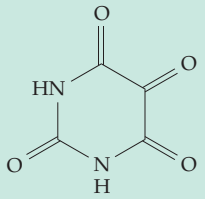



Coomassie brilliant blue

## BOX 3-E NINHYDRIN

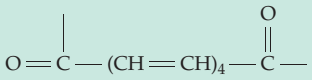
Ninhydrin (1,2,3-indantrione monohydrate) forms Schiff bases (ketimines) with amino acids. These react in ways similar to those of Schiff bases of pyridoxal phosphate (Chapter 14). Decarboxylation of the ketimines followed by hydrolysis of the resulting aldimines yields an intermediate amine that can couple with a second molecule of ninhydrin to form a characteristic purple color.[a,b] First reported by Ruhemann in 1910, the intermediate amine can also be hydrolyzed to free ammonia. Therefore, to ensure maximum color yield ninhydrin solutions for quantitative analysis usually contain reduced ninhydrin, which can react with free $NH_3$ and ninhydrin to form Ruhemann's purple (see scheme). The reaction has been widely used in chromatography and in quantitative amino acid analysis and also as a convenient spray reagent for paper and thin-layer chromatography. While α-amino acids react most readily, primary amines and peptides also form Ruhemann's purple. In these cases a proton rather than $CO_2$ is lost from the ketimine. When pyridoxamine (Chapter 14) on chromatograms reacts, a bright orange product, presumably the aldimine, appears. Secondary amines, such as proline, give a yellow color.

Both the ninhydrin reaction and pyridoxal phosphate-catalyzed decarboxylation of amino acids (Chapter 14) are examples of the **Strecker degradation**. Strecker reported in 1862 that alloxan causes the decarboxylation of alanine to acetaldehyde, $CO_2$, and ammonia.[c]

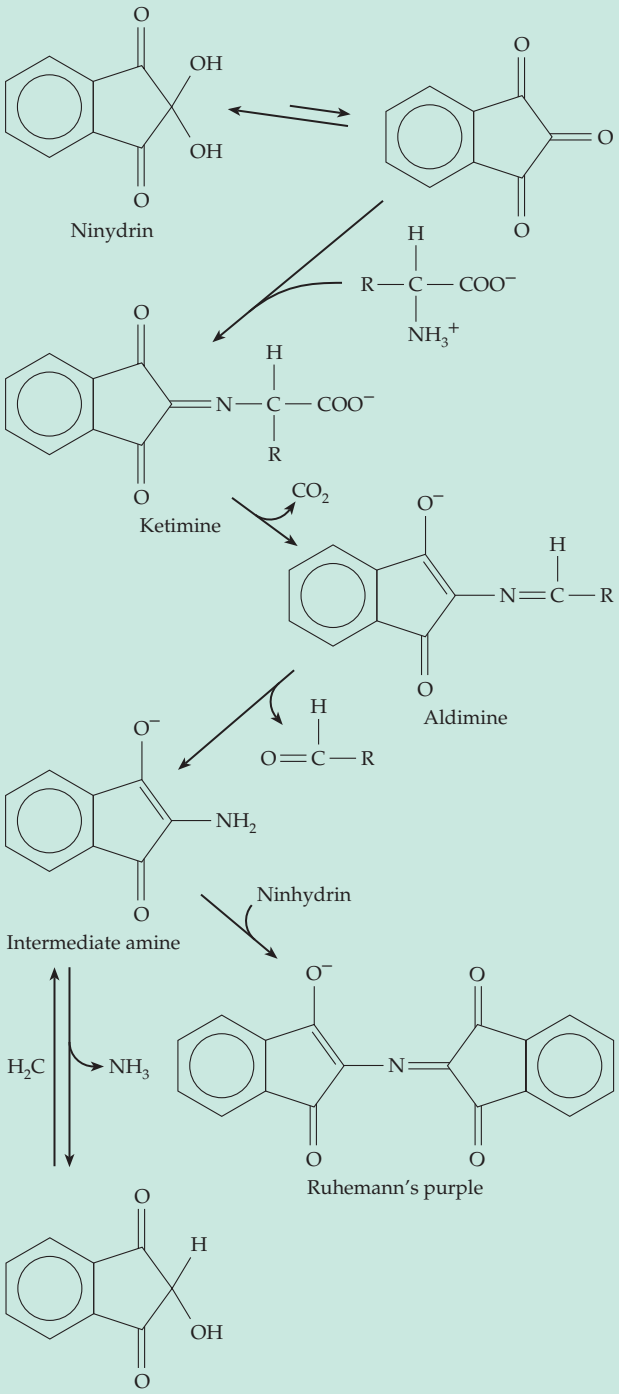


Alloxan

Many other carboxyl compounds, e.g., those of the general structure




and *p*-nitrosalicylaldehyde also cause the Strecker degradation.[d]




[a] Wigfield, D. C., and Croteau, S. M. (1980) *Biochem. Edu.* **8**, 26–27
[b] Friedman, M., and Williams, L. D. (1974) *Bioorg. Chem.* **3**, 267–280
[c] Strecker, A. (1862) *Annalen* **123**, 363–365
[d] Schonberg, A., and Moubacher, R. (1952) *Chem. Rev.* **50**, 261–277

the resulting protein "blots" can be stained with specific antibodies.[269,270] Flame ionization detectors can measure as little as a few picomoles of almost any substance leaving a vapor-phase chromatographic column. The importance of developing new, more sensitive analytical methods by which the quantity of material investigated can be scaled down can hardly be overemphasized. Increasingly sensitive methods of detection, including mass spectrometry, now permit measurement of fmol ($10^{-15}$ mol) quantities in some cases. With this ability the output of neurotransmitters from a single neuron in the brain can be measured and the contents of single cells can be analyzed.

## 6. Absorption of Light

Side chains of the three aromatic amino acids phenylalanine, tyrosine, and tryptophan absorb ultraviolet light in the 240- to 300-nm region, while histidine and cystine absorb to a lesser extent. Figure 3-13 shows the absorption spectrum of a "reference compound" for tyrosine. There are three major absorption bands, the first one at 275 nm being a contributor to the well-known 280-nm absorption band of proteins. There is a much stronger absorption band at about 240 nm. Sensitive methods for estimating protein concentration depend upon the measurement of this absorption together with that from other side chains at around 280 or 230 nm.[13,271–274] There is an even stronger absorption band at 192 nm. However, at these wavelengths even air absorbs light and experimental difficulties are extreme. At 280 nm, and even more at 230 nm, it is easy to contaminate samples with traces of light-absorbing material invisible to the eye. Therefore, most estimations of protein concentration from light absorption depend upon the 280-nm band.

Figure 3-14 shows the spectra of *N*-acetyl ethyl esters of all three of the aromatic amino acids and of cystine. To a first approximation, the absorption spectra of proteins can be regarded as a summation of the spectra of the component amino acids. However, the absorption bands of some residues, particularly of tyrosine and tryptophan, are shifted to longer wavelengths than those of the reference compounds in water. This is presumably a result of being located within nonpolar regions of the protein. Notice that the spectra for tyrosine, phenylalanine, and cystine in Fig.

### BOX 3-F BIOSENSORS AND ELECTRONIC NOSES

A new approach to detection of molecules of biological interest is the development of biosensors. These are small devices that detect the binding of specific molecules to a **receptor** which is in intimate contact with a specially prepared surface that serves as a **transducer**. The receptor might be a layer of enzyme, antibody, hormone receptor, lectin, or oligonucleotide. Binding of substrate, antigen, hormone, sugar, or complementary polynucleotide strand, respectively, induces a response consisting of some kind of electrical or optical signal.[a–d] If the sensor is constructed on a semiconductor chip changes in an imposed potential difference may be detected.[a,e] However, changes in optical properties are more often observed. Fluorescence of dyes incorporated into the tranducing layer may be induced by binding of a molecule to a protein that undergoes an allosteric modification (see Chapter 9).[f] Many biosensors measure **surface plasmon resonance**, a change in the evanescent wave that develops in a surface when a light beam at the angle of total reflectance strikes the surface. This induces a change in the dielectric constant which can be measured.[f–m] Biosensors are used to estimate binding constants and also rate constants. However, read the article by Schuck and Milton[m] for tests of the validity of kinetic data. Biosensors can serve as "electronic noses." One possible application is in the analysis of compounds in human breath as an aid to medical diagnosis. Over 400 volatile organic compounds have been identified in breath using gas chromatography and mass spectrometry.[n]


[a] Briggs, J. (1987) *Nature (London)* **329**, 565–566
[b] Zurer, P. (1997) *Chem. Eng. News* **September 15**, 7
[c] Kress-Rogers, E., ed. (1997) *Handbook of Biosensors and Electronic Noses: Medicine, Food, and the Environment*, CRC Press, Boca Raton, Florida
[d] Cunningham, A. J. (1998) *Introduction to Bioanalytical Sensors*, Wiley, New York
[e] McConnell, H. M., Owicki, J. C., Parce, J. W., Miller, D. L., Baxter, G. T., Wada, H. G., and Pitchford, S. (1992) *Science* **257**, 1906–1912
[f] Marvin, J. S., Corcoran, E. E., Hattangadi, N. A., Zhang, J. V., Gere, S. A., and Hellinga, H. W. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4366–4371
[g] Raether, H. (1988) *Surface Plasmons, Springer Tracts in Modern Physics*, Vol. 111, Springer-Verlag, Berlin
[h] Peterlinz, K. A., Georgiadis, R. M., Herne, T. M., and Tarlov, M. J. (1997) *J. Am. Chem. Soc.* **119**, 3401–3402
[i] Hendrix, M., Priestley, E. S., Joyce, G. F., and Wong, C.-H. (1997) *J. Am. Chem. Soc.* **119**, 3641–3648
[j] Salamon, Z., Brown, M. F., and Tollin, G. (1999) *Trends Biochem. Sci.* **24**, 213–219
[k] Chao, H., Houston, M. E., Jr., Grothe, S., Kay, C. M., O'Connor-McCourt, M., Irvin, R. T., and Hodges, R. S. (1996) *Biochemistry* **35**, 12175–12185
[l] McNally, A. J., Mattsson, L., and Jordan, F. (1995) *J. Biol. Chem.* **270**, 19744–19751
[m] Schuck, P., and Minton, A. P. (1996) *Trends Biochem. Sci.* **21**, 458–460
[o] Phillips, M. (1992) *Sci. Am.* **267**(July), 74–79

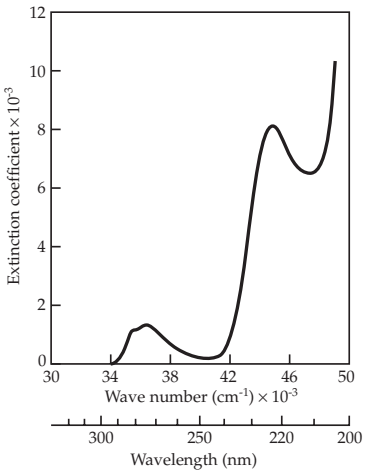



**Figure 3-13** The absorption spectrum of *N*-acetyltyrosine ethyl ester in an aqueous phosphate buffer of pH 6.8. Absorbance (as molar extinction coefficient, Eq. 23-5) is plotted against increasing energy of light quanta in units of wave number. The more commonly used wavelength scale is also given. Spectra are most often presented with the low wavelength side to the left. In the convention adopted here the energy of a quantum increases to the right. There are three $\pi-\pi^*$ electronic transitions that give rise to absorption bands of increasing intensity. The third $\pi-\pi^*$ transition of the aromatic ring is at ~ 52,000 $cm^{-1}$ (192 nm) and reaches a molar extinction coefficient of ~ 40,000. The $n-\pi^*$ and $\pi-\pi^*$ transitions of the amide group in this compound also contribute to the high energy end of the spectrum (see Chapter 23 for further discussion).

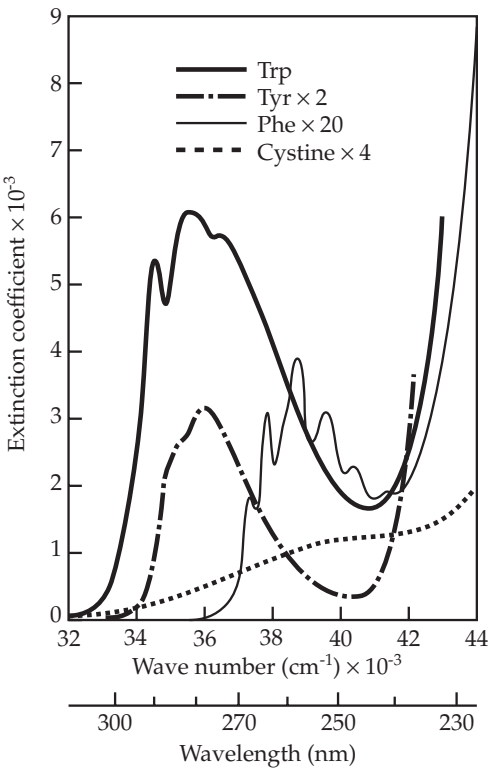



**Figure 3-14** The spectra of the first electronic transitions of the *N*-acetyl derivatives of the ethyl esters of phenylalanine, tyrosine, and tryptophan together with that of the dimethyl ester of cystine in methanol at 25°C. The spectra for the Tyr, Phe, and cystine derivatives have been multiplied by the factors given on the graph.[272]

3-14 have been multiplied by factors of 2 to 20. It is evident that if all of the light-absorbing side chains were present in equal numbers tryptophan would dominate the absorption band and that phenylalanine would contribute little except some small wiggles. The molar extinction coefficient ε can be estimated from the numbers of residues of each type per molecule as follows:[274]

$$\varepsilon_{280}\ (M^{-1}cm^{-1}) = 5500\ (\text{no. Trp}) + 1490\ (\text{no. Tyr}) + 125\ (\text{no. cystine}) \quad (3\text{-}33)$$

For proteins of unknown composition, a useful approximation is that a solution containing 1 mg / ml of protein has an absorbance at 280 nm of about 1.0.

## E. Quantitative Determinations and Modification Reactions of Side Chain Groups

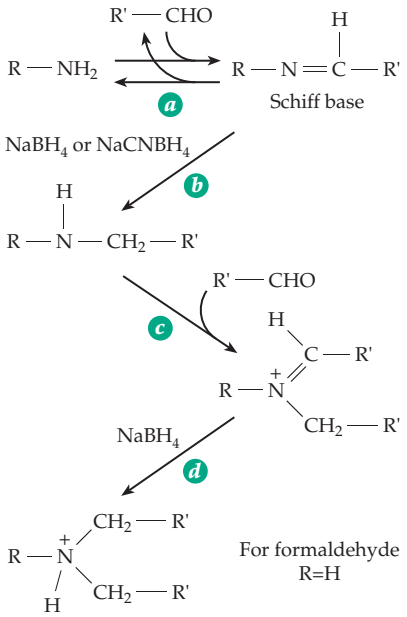
The functional groups present in the side chains of proteins include $-NH_2$, –SH, S—S, –OH, $-COO^-$, the imidazole group of histidine, the guanidine group of arginine, the phenolic group of tyrosine, the indole ring of tryptophan, and the $-S-CH_3$ group of methionine. These are able to enter into a great variety of chemical reactions, most of which make use of the nucleophilic properties of these groups. The reactions are most often those of nucleophilic **addition** or nucleophilic **displacement**. The basic chemistry of these reactions often parallels biochemical reactions that are discussed in Chapters 12 and 13. In many instances, the reactions are nonspecific; amino, thiol, and hydroxyl groups may all react with the same reagent. The usefulness of the reactions depends to a large extent on the discovery of conditions under which there is some selectivity. It is also important that the reactions be complete. Only a few reactions will be considered here; these and others have been reviewed by Glazer *et al.*[225,275]

### 1. Reactions of Amino Groups

The numerous amino groups of lysine residues and of the N termini of peptide chains usually pro-

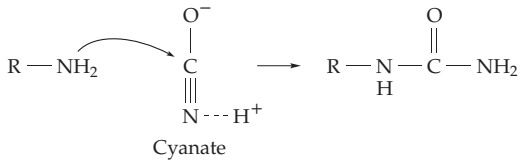
trude into the aqueous surroundings of a protein. Chemical modification can be done in such a way as to preserve the net positive charge which amino groups carry at most pH values, to eliminate the positive charge leaving a neutral side chain, or to alter the charge to a negative value. Alterations of these charges can greatly affect interactions of the protein molecules with each other and with other substances.

Amino groups react reversibly with carbonyl compounds to form Schiff bases.[275–277] Reduction of the latter by sodium borohydride or sodium cyanoborohydride causes an irreversible change (Eq. 3-34, steps *a* and *b*). Cyanoborohydride is specific for Schiff bases and does not reduce the carbonyl compound. However, side products may cause problems.[277,278] Depending upon which carbonyl compound is used, the net positive charge on the amino group may be retained or it may be replaced with a different charge by this "reductive alkylation" sequence. Formaldehyde will react according to Eq. 3-34 in two steps to give a dimethyl amino group with no change of net charge.[279] Pyridoxal phosphate (Chapter 14) is converted by Eq. 3-34 into a fluorescent label. With a limited amount of pyridoxal phosphate only one or a few lysine residues may be labeled, often at active centers of enzymes. Schiff bases formed from glyceraldehyde in Eq. 3-34 can undergo the Amadori rearrangement (Eq. 4-8) to form stable products which, however, can be reconverted to the original amino groups upon acid hydrolysis. The borohydride reduction product of Eq. 3-34 with glyceraldehyde can be reconverted to the original amine by periodate oxidation (Eq. 4-11).[277]



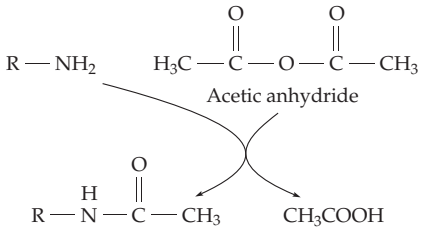

(3-34)

Another addition reaction of amino groups is **carbamoylation** with sodium cyanate (Eq. 3-35). A displacement reaction by an amino group on an acid




(3-35)

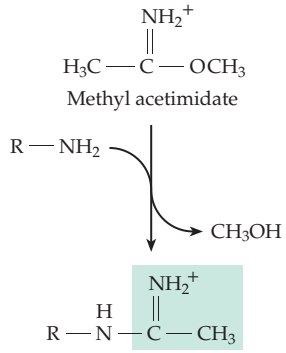
anhydride such as acetic anhydride (Eq. 3-36) leads to **acylation**, a nonspecific reaction which is also undergone by thiol, hydroxyl, and other groups. When acetic anhydride is used, the net positive charge of an



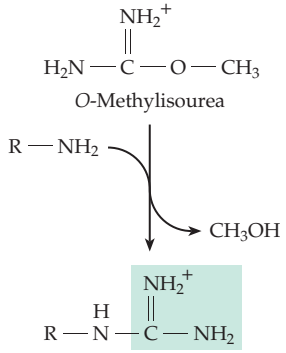

(3-36)

amino group is lost. However, the product obtained with succinic anhydride or maleic anhydride (Eq. 3-26) carries a negative charge. In the latter case, the modification can readily be reversed by altering the pH.

Both **amidination** (Eq. 3-37) and **guanidination** (Eq. 3-38) lead to retention of the positive charge.

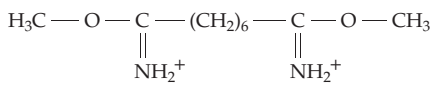



(3-37)




(3-38)

Bifunctional imidoesters such as **dimethylsuberimidate** may be used to establish whether or not two different proteins or subunits are close together in a complex or in a supramolecular structure such as a membrane or ribosome.

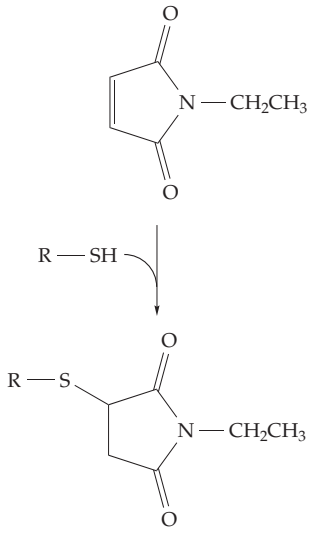



Dimethylsuberimidate

Another useful reaction of amino side chains is that with dansyl chloride (Eq. 3-29). Many lysine derivatives can be determined quantitatively by amino acid analysis.[280]

## 2. Reactions of SH Groups

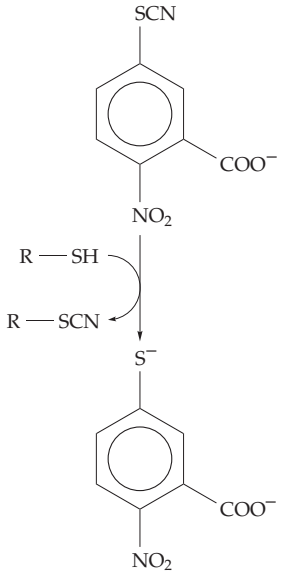
In addition to the alkylation with iodoacetate (Eq. 3-24), sulfhydryl groups can react with *N*-ethylmaleimide (Eq. 3-39).[281] This reaction blocks the SH groups irreversibly and has often been used in attempts to establish whether or not a thiol group plays a role in the functioning of a protein. Loss of function in the




(3-39)

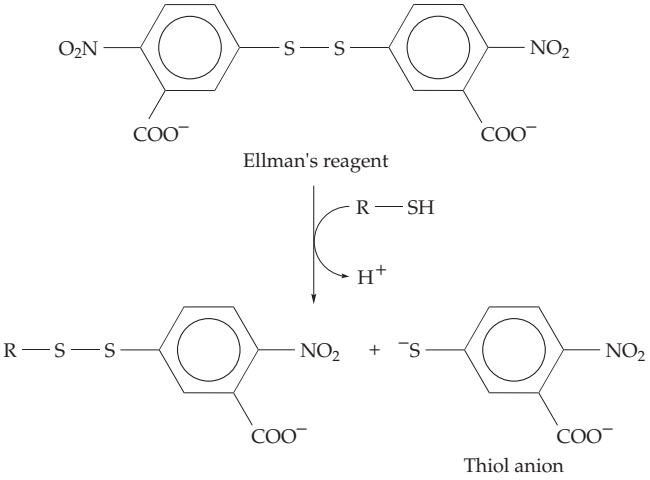
presence of this "sulfhydryl reagent" may mean that an SH group has an essential role or it could be a result of the bulk of the group added. The *N*-ethylmaleimide group is large and could prevent proper contact between an enzyme and substrate or between two proteins. To avoid the possible effect of excessive bulk, it is useful to convert the SH to the small thiocyanate group – SCN (Eq. 3-40).[282]

Ellman's reagent 5,5'-dithiobis(2-nitrobenzoic acid; DTNB) reacts quantitatively with – SH groups (Eq. 3-41) to form mixed disulfides with release of a thiolate anion that absorbs light at 412 nm with a molar extinction



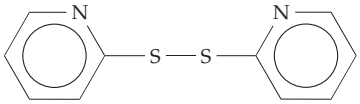

(3-40)

coefficient of 14,150 $M^{-1}$ $cm^{-1}$.[283] While DTNB has been widely used to determine the content of – SH groups in proteins, there are some disadvantages.



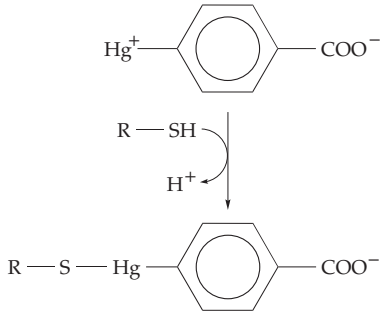

(3-41)

Pyridyldisulfides such as 2-pyridyldisulfide or the isomeric 4-pyridyldisulfide react more completely and with greater selectivity.[281]




Thiol groups have a high affinity for mercury ions including organic mercury derivatives, which are widely used in the determination of protein structures by X-ray crystallography (Section F). Titration of SH groups in proteins is often accomplished with

*p*-mercuribenzoate (Eq. 3-42). The reaction may be followed spectrophotometrically at 250 – 255 nm, a region in which the mercaptide product absorbs strongly.

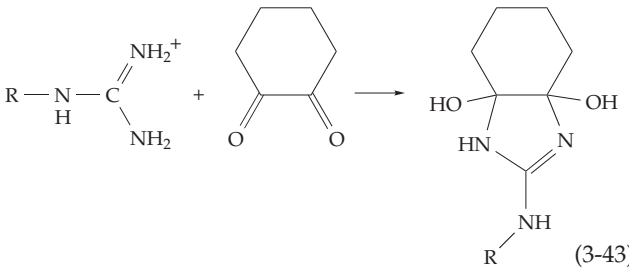



(3-42)

## 3. Reactions of Other Side Chains

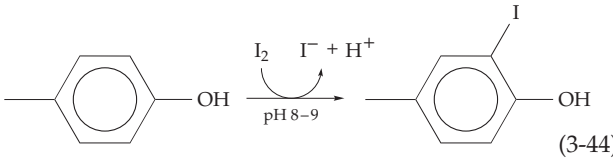
There are no highly selective reactions for –OH, –COO⁻, or imidazole groups. However, some hydroxyl groups in active sites of enzymes are unusually reactive in nucleophilic addition or displacement and can be modified by acylation, phosphorylation, or in other ways. Carboxyl groups, which are exceedingly numerous on protein surfaces, can be modified by treating with a water-soluble carbodiimide (Eq. 3-10) in the presence of a high concentration of an amine such as the ethyl ester of glycine. The imidazole groups of residues of histidine can often be selectively destroyed by dye-sensitized photooxidation (Ch. 12, Section D,5) or can be acylated with ethoxyformic anhydride.

Compounds with two adjacent carbonyl groups such as 1,2-cyclohexanedione (Eq. 3-43) react selectively with guanidinium groups from arginine residues in proteins. Under certain conditions the product indicated in Eq. 3-43 predominates. Related reagents are derived from camphorquinone.[284]



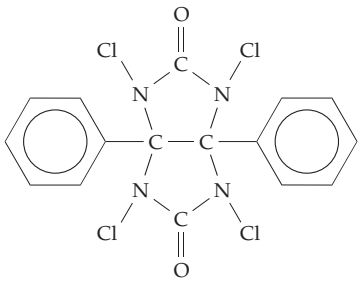

(3-43)

The phenolic group of tyrosine undergoes iodination (Eq. 3-44), acylation, coupling with diazonium compounds, and other reactions.



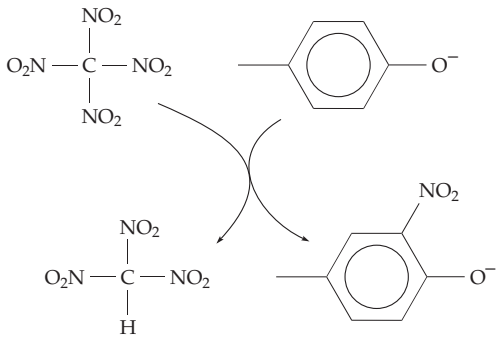

(3-44)

The following sparingly soluble chloroamide together with I⁻ will also iodinate tyrosine and can be used to incorporate radiolabeled iodine into proteins.[285,286]




1,3,4,6-Tetrachloro-3α,6α-diphenylglycouril

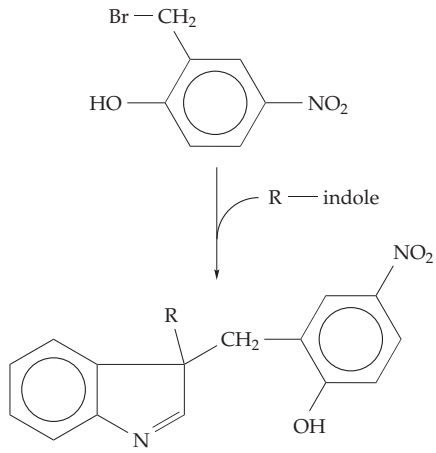
Tetranitromethane reacts slowly with tyrosyl groups to form 3-nitrotyrosyl groups (Eq. 3-45). The by-product **nitroform** is intensely yellow with $\varepsilon_{350}$ = 14,400. The reagent also oxidizes SH groups and reacts with other anionic groups.




Nitroform

(3-45)

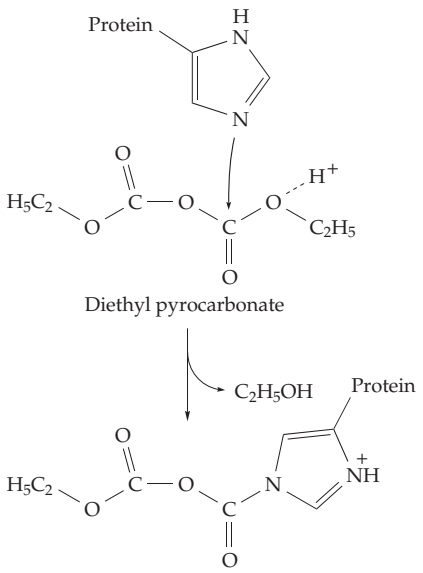
Koshland devised the following reagent for the indole rings of tryptophan residues (Eq. 3-46).




(3-46)

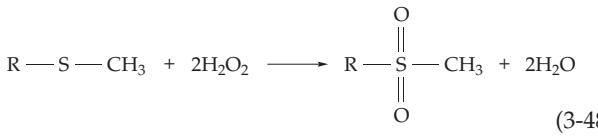
Imidazole, lysine amino groups, and tyrosine hydroxyl groups react with **diethylpyrocarbonate**

(Eq. 3-47) at low enough pH (below 6) that the reaction becomes quite selective for histidine.[70,287] Reactivity with this reagent is often used as an indication of histidine in a protein.[288–290]. The reaction may be monitored by observation of NMR resonances of imidazole rings.[288,290]




(3-47)

The thioether side chains of methionine units in proteins can be oxidized with hydrogen peroxide to the corresponding sulfones (Eq. 3-48). They can also be alkylated, e.g., by $CH_3I$ to form $R—S^+(CH_3)_2$.




(3-48)

## 4. Affinity Labeling

To identify groups that are part of or very near to the active site of a protein, reagents can be designed that carry a reactive chemical group into the active site.[291] The related **photoaffinity labeling**[292,293] is also widely used (see also Chapter 23).

# F. Synthesis of Peptides

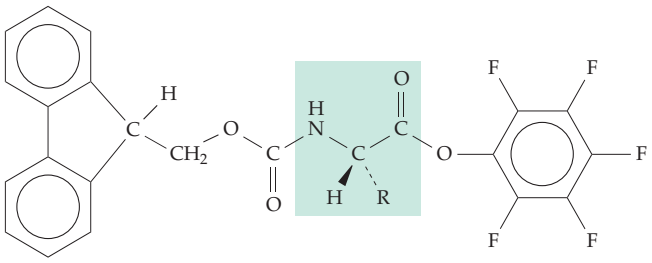
The synthesis of peptides of known sequence in the laboratory is extremely important to biochemical research. For example, we might want to know how the effects of a peptide hormone are altered by replacement of one amino acid in a particular position by another. The synthetic methods must be precise[294–298] and because there are so many steps the yield should be 98% or better for every step. Even so, it is still impractical to synthesize very large peptides. Those that have been made, such as the hormone insulin and the enzyme ribonuclease, have been obtained in low yields and have been difficult to purify.[298] It is usually more practical to obtain large peptides from natural sources. It is often practical to clone a suitable piece of DNA in a bacterial plasmid and to set up biological production of the desired peptide (Chapter 26). On the other hand, for smaller peptides, laboratory synthesis is feasible. Even for large peptides it is useful because it permits incorporation of unnatural amino acids as well as isotopic labels.

The general procedure for making a peptide in the laboratory is to "block" the amino group of what will become the N-terminal amino acid with a group that can be removed later. The subsequent amino acid units "activated" at their carboxyl end are then attached one by one. The chemical activation is often accomplished by conversion of the carboxyl group of the amino acid to an anhydride. At the end of the synthesis, the blocking group must be removed from the N terminus and also from various side chain groups such as those of cysteine and lysine residues. In many respects, this procedure is analogous to the biological synthesis of proteins whose basic chemistry is discussed in Chapter 29.

## 1. Solid-Phase Peptide Synthesis

Modern methods of peptide synthesis began with the solid-phase method introduced by Merrifield[299] in 1962 (Fig. 3-15). To begin the synthesis a suitably protected amino acid is covalently linked to a polystyrene bead. The blocking *t*-butoxycarbonyl (Boc) group is removed as isobutene by an elimination reaction to give a bound amino acid with a free amino group. This can then be coupled to a second amino acid with a blocked amino group using dicyclohexylcarbodiimide (Eq. 3-10). The removal of the blocking group and addition of a new amino acid residue can then be repeated as often as desired. The completed peptide is removed from the polystyrene by action of a strong acid such as HF.

Advantages of the Merrifield procedure are that the peptide is held tightly and can be washed thoroughly at each step. Problems arise from repeated use of trifluoroacetic acid and the need to use HF or other strong acid to cleave the peptide from the matrix and also to remove blocking benzyl groups that must be present on many side chain groups. Newer variations of the procedure include a more labile linkage to a polyamide type of polymer and use of blocked amino acids.[297,300–301a] These "active esters" will spontaneously condense with the free amino group of the growing peptide and with suitable catalysis will eliminate




Blocked amino acid

pentafluorophenol. The fluorenylmethoxycarbonyl (Fmoc) blocking group is removed under mildly basic conditions. The whole procedure has been automated in commercially available equipment.

Smaller peptides may be joined to form longer ones.[298] Also useful is enzymatic synthesis. Protein-hydrolyzing enzymes under appropriate conditions will form peptide linkages, for example, joining together oligopeptides.[302,303] Other new methods have been devised to join unprotected peptides.[304,305]

Semisynthetic approaches can also be used to place unnatural amino acids into biologically synthesized proteins through the use of suppressor transfer RNAs (Chapter 29).[306]

## 2. Combinatorial Libraries

Many chemists devote all of their efforts to the synthesis of new compounds, including polypeptides, that might be useful as drugs. Traditionally, this has involved the tedious preparation of a large number of compounds of related structure which can be checked individually using various biochemical or biological tests. In recent years a new approach using "combinatorial chemistry" has become very popular and is continually being adapted for new purposes.[307–310] There are several approaches to creating a combinatorial library.

In "split synthesis" procedures a solid-phase

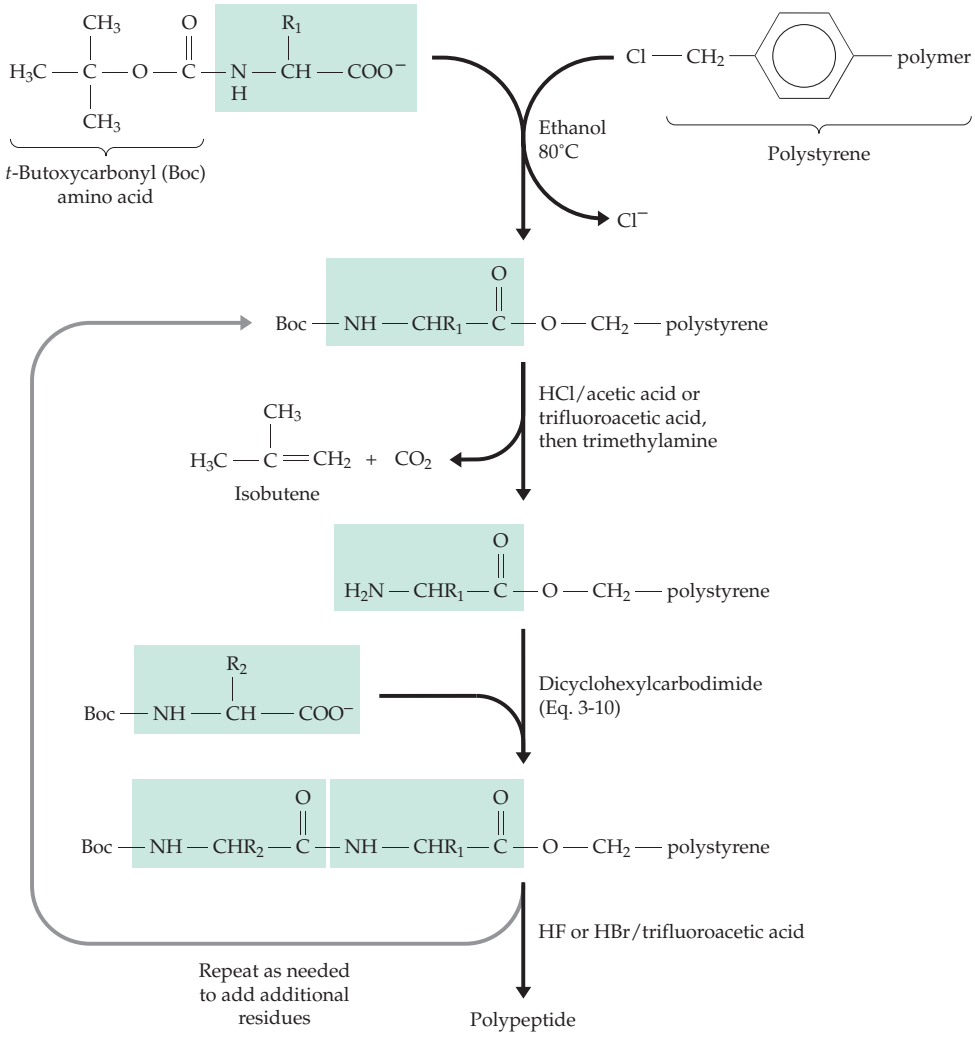



**Figure 3-15** Procedure for solid-phase peptide synthesis devised by Merrifield.[299]

synthesis is conducted on beads. For example, a family of peptides, each with the same C terminus, can be started on a large number of beads. After the first amino acid residue is attached the beads are divided into up to 20 equal portions and different amino acids are added to each portion. The beads can then be mixed and again subdivided. The third residue will again contain many different amino acids attached to each of the different amino acids in the second position. By repeating the procedure again, perhaps for many steps, a "library" of random peptides with each bead carrying a single compound will be formed.

To test whether a polypeptide or other compound carried on a given bead has a derived biological activity, such as the ability to inhibit a certain enzyme, various assays that require only one bead can be devised. However, if a particular bead carries a compound of interest, how can it be identified? The bead carries only a small amount of compound but it may be possible using microsequencing procedures to identify it. An alternative procedure is to use an encoding method to identify the beads.

An alternative to the "one bead–one peptide" approach is to incorporate random sequences of a DNA segment into a gene that can be used to "display" the corresponding peptide sequence. This is illustrated in Fig. 3-16. A protein segment (which may be a random sequence) can be displayed either on the major coat proteins along the shaft or on the minor coat proteins at the end of the bacterial virus fd (see Fig. 7-7). In the case of random insertions, each virus particle may display a different sequence (as many as $10^8$). Peptides may be selected by binding to a desired receptor or monoclonal antibody and the DNA encapsulated in the virus particle can be used to produce more peptides for identification purposes.[311–316] Many other ingenious systems for constructing and testing libraries of peptides[317–320] and other molecules[319,321,322] are being devised. One of these involves a photolithography procedure for immobilizing macromolecules in a regular addressable array, e.g., in a 0.5-mm checkerboard pattern, on a flat surface.[323,324]

## G. Microscopy

The light microscope[325] was developed around 1600 but serious studies of cell structure (histology) did not begin until the 1820s. By 1890, microscope lenses had reached a high state of perfection[326] but the attainable resolution was limited by the wavelength of light. For 450 nm blue light the limit is about 300 nm and for ultraviolet light, viewed indirectly, about 200 nm.[325,327] By the 1940s the electron microscope with its far superior resolving power had overshadowed the light microscope.

For both light and electron microscopy, the preparation of thin sections of cells is a very important technique. Only with very thin sections is the image sufficiently focused. However, **confocal scanning optical microscopy**, invented in the 1950s but not used commercially until much later,[328,329] provided an alternative solution to the focusing problem. A conical beam of light focused to a point is scanned across the sample and the transmitted light (or light emitted by fluorescence) passes through a small "pinhole" aperture located in the primary image plane to a photomultiplier tube where its intensity is recorded. The illuminating beam is moved to scan the entire field sequentially. A series of pinholes in a spinning disk may accomplish the same result. The focal plane can

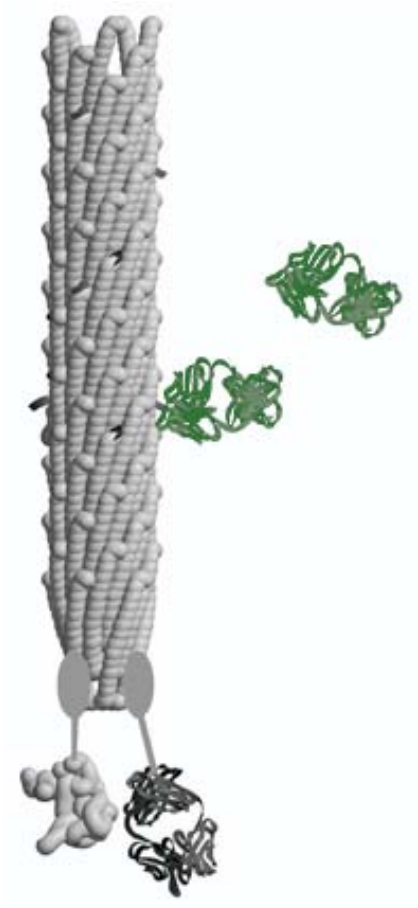


**Figure 3-16** Model of bacteriophage fd engineered to display peptides as inserts in the coat proteins of the virus. The native virus structure is shown in gray; proteins not present in the native virus are shown black or green. Inserted near the N-termini of some major coat proteins is a 6-residue peptide. To one of these peptides a specific Fab antibody fragment (green) has bound from solution, and a second Fab is shown nearby. The N-terminal region of a minor coat protein at the end of the virion has been engineered to display a (different) Fab fragment. Steric constraints are less stringent for inserts in the minor proteins, but fewer copies per virion are possible. Reprinted with permission from Barbas, *et al.*[313a]

be varied so that an image of a thick object such as a cell can be optically sectioned into layers of less than 1 µm thickness. Stereoscopic pairs can also be generated (Fig. 3-17).[330]

A newer development in confocal microscopy is the use of two-photon and three-photon excitation of the fluorescent molecules that occur naturally within cells using short pulses of short-wavelength high-energy laser light.[331] Distribution of such compounds as NADH,[332] DNA, and the neurotransmitter serotonin[333] can be observed without damaging cells. Individual storage granules, each containing ~5 x $10^8$ molecules of serotonin in a concentration of ~50 mM, can be seen.[334] Another new instrument, the **near-field scanning optical microscope** (NSOM),[335,336] is a lensless instrument in which the illuminating beam passes through a very small (e.g., 100 nm diameter) hole in a probe that is scanned in front of the sample. It may extend the limit of optical microscopy to ~1/50 the wavelength of the light.

Since it first became commercially available in 1939, the electron microscope has become one of the most important tools of cell biology.[337,338] The practical resolution is about 0.4 nm, but recent developments in scanning electron microscopy have resulted in resolution of 0.14 nm.[339] Of major importance was the development around 1950 of microtomes and knives capable of cutting thin (20–200 nm) sections of tissues embedded in plastic.[340] A bacterium such as *E. coli* can be sliced into as many as 10 thin longitudinal slices (see Fig. 1-4) and a eukaryotic cell of 10 µm diameter into 100 slices. Serial sections can be examined to determine three-dimensional structures. For some results see Bubel.[341]

If a slice of fresh (frozen) tissue is examined directly, little is seen because most of the atoms found in cells are of low atomic mass and scatter electrons weakly and uniformly. Therefore, thin sections must be "stained' with atoms of high atomic mass, e.g., by treatment with potassium permanganate or osmium tetroxide. Tissues must also be "fixed" to prevent disruption of cell structures during the process of removal of water and embedding in plastic. Fixatives such as formaldehyde react with amino groups and other groups of proteins and nucleic acids. Some proteins are precipitated in place and digestive enzymes that otherwise would destroy much of the fine structure of the cell are inactivated. Glutaraldehyde (a five-carbon dialdehyde) is widely used to fix and crosslink protein molecules in the tissue. The methods continue to be improved.[342]

Small particles, including macromolecules, may be "shadowed." Chromium or platinum can be evaporated in a vacuum from an angle onto the surface of the specimen, Individual DNA molecules can be "seen" in this way.[343] In fact, only the "shadows" are seen and they are 2–3 times wider than the DNA molecules. In the **negative contrast** method a thin layer of a solution containing the molecules to be examined, together with an electron-dense material such as 1% sodium phosphotungstate, is spread on a thin carbon support film. Upon drying, a uniform electron-dense layer is formed. Where the protein molecules lie, the phosphotungstate is excluded, giving an image of the protein molecule.

Surfaces of cells, slices, or intact bacteria can be coated with a deposit of platinum or carbon. The coating, when removed, provides a "negative" **replica** which can be examined in the microscope. Alternatively, a thin plastic replica can be made and can be shadowed to reveal topography. In "freeze fracturing" and "freeze etching," fresh tissue, which may contain glycerol to prevent formation of large ice crystals, is frozen rapidly. Such frozen cells can often be revived; hence, they may be regarded as still alive until the moment that they are sliced! The frozen tissue is placed in a vacuum chamber within which it is sliced or fractured with a cold knife. If desired, the sample can be kept in the vacuum chamber at about –100°C for a short time, during which some water molecules evaporate from the surface. The resultant etching reveals a fine structure of cell organelles and membranes in sharp relief. After etching, a suitable replica is made and examined (Fig. 1-15A and E). Fracturing tends to take place through lipid portions of cell membranes.

Small viruses, bacterial flagella, ribosomes, and even molecules can be seen by electron microscopy. However, to obtain a clear image in three dimensions requires a computer-based technique of **image reconstruction** or **electron microscope tomography**, which was developed initially by Aaron Klug and associates.[344–349] A sample is mounted on a goniometer, a device that allows an object to be tilted at exact angles. Electron

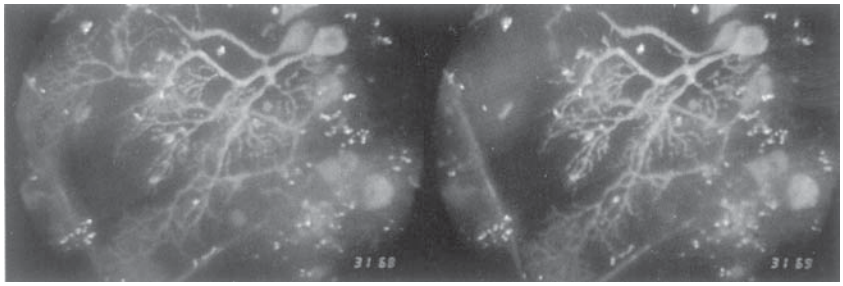


**Figure 3-17** Confocal micrograph showing a forty-micrometer stereo slice in a 90-µm thick section of mouse cerebellum.[330] Courtesy of A. Boyde.

micrographs are prepared with the sample untilted and tilted in several directions at various angles, e.g., up to 90° in 10° increments. The micrographs are digitized and a computer is used to reconstruct a three-dimensional image.

In **electron crystallography**[350] micrographs of two-dimensional crystalline arrays of molecules or larger particles are prepared. A Fourier transform of the micrograph gives a diffraction pattern which can be treated in a manner similar to that usual for X-ray diffraction to give a three-dimensional image. An important milestone in use of this technique was the determination of the structure of bacteriorhodopsin (Fig. 23-45) at 0.3-nm resolution. Bear in mind that X-ray crystallography can also be viewed as a form of microscopy.

Invention of the **scanning tunneling microscope** (STM) by Binnig and Roher[351–353] initiated a new revolution in microscopy. The STM and similar scanned probe microscopes examine surfaces by moving a fine probe mounted in a piezoelectric x,y,z-scanner[354] across the surface to be examined. The tiny tungsten probe of the STM is so fine that its tip may consist of a single atom. When a small voltage is applied a minuscule quantum mechanical tunnelling current flows across the small gap between the probe and the surface and a high-resolution image, sometimes at atomic resolution, is created from the recorded variation in current.[354-359] The STM theoretically responds only to surfaces that conduct electrons, but nonconducting samples have been imaged at high humidity; presumably by conductance of electrons or ions through the surface water layer. The success of the STM spurred the development of many other types of scanned probe microscopes. Among these, the **atomic force microscope** (AFM; Fig. 3-18) has been especially useful for biological materials, including proteins and nucleic acids. The AFM moves a fine-tipped stylus directly across the sample surface or, alternatively, vibrates the probe above the surface. The small up-and-down movements of the stylus are recorded[359–366] and thereby create a topographic or force-field map of the sample. AFM images contain three-dimensional information and can be used to view individual molecules (Fig. 3-19).[367,367a] **Chemical force microscopy** is sensitive to adhesion and friction as a function of the interaction between defined chemical groups on the tip and sample.[368]

An emerging field is force spectroscopy, in which the AFM measures interaction forces between and within individual molecules.[369–371a] Under development are NMR microscopes (Section I). There is continual effort to see small objects more directly and more clearly!

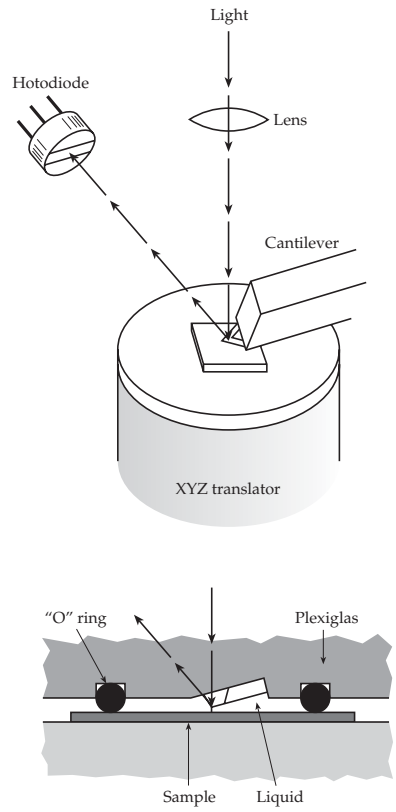



**Figure 3-18** Schematic diagram of the atomic force microscope.[360] Courtesy of Paul Hansma.

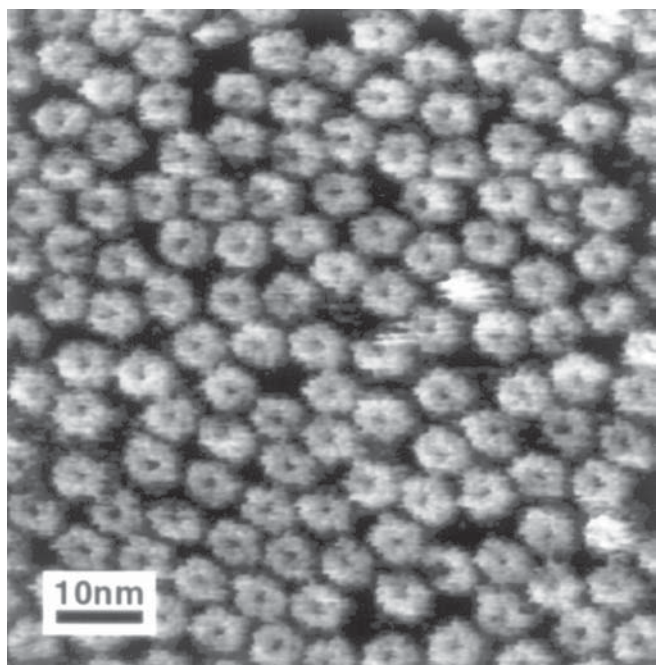



**Figure 3-19** AMF images of cholera toxin[364] (see also Box 11-A). Courtesy of Z. Shao.

## H. X-ray and Neutron Diffraction

One of the most important techniques by which we have learned bond lengths and angles and precise structures of small molecules is X-ray diffraction. Today, this technique, which involves measurements of the scattering of X-rays by crystalline arrays of molecules, is being used with spectacular success to study macromolecules of biochemical and medical importance.[372–378]

X-rays were described by Röntgen in 1896 but there was uncertainty as to their wave nature.[274] It was not until 1912 that the wavelengths of X-rays had been measured and it was recognized that they were appropriate for the use of X-rays in structure determination. Consider the fact that with a conventional light microscope we cannot distinguish two small objects that are much closer together than the distance represented by the wavelength of the light with which we observe them. This is about 460 nm for blue light. By comparison the wavelength of the copper $K_\alpha$ radiation, which is used in protein crystallography, is 0.1542 nm entirely appropriate for seeing the individual atoms of which matter is composed. Recognizing this, W. L. Bragg in Cambridge, England, in 1913 used X-ray diffraction to establish the structures of NaCl, KCl, and KBr in the crystalline state. The science of X-ray crystallography had been founded.[379]

In 1926, James Summer crystallized the enzyme urease (Chapter 16) and crystallization of other enzymes soon followed.[380] In 1934, J. B. Bernal brought back to Cambridge from Uppsala, Sweden, some crystals of pepsin almost 2 mm long that had been grown in T. Svedberg's laboratory. Bernal and Crowfoot showed that these delicate crystals, which contained almost 50% water, gave a sharp diffraction pattern when they were protected by enclosure in a narrow capillary tube containing some of the mother liquor from which the crystals had been grown.[381,382] After this, diffraction patterns were obtained for many protein crystals, and in 1937 Max Perutz chose for his thesis work at Cambridge the X-ray crystallography of hemoglobin. The project seemed hopeless at times, but in 1968, 31 years later, Perutz had determined the structure of hemoglobin.[383–386]

If X-rays could be focused easily, could one build an X-ray microscope that would permit the immediate viewing of molecular structures? X-ray holograms at the molecular level have been obtained.[387,388] However, currently the only practical X-ray microscope for protein structures involves the measurement of diffraction patterns created by the scattering of X-rays (or of neutrons) from the crystalline lattice. The details of this procedure can be found in other sources.[372–375,378,389–390] It is sufficient to point out here that a pattern of many

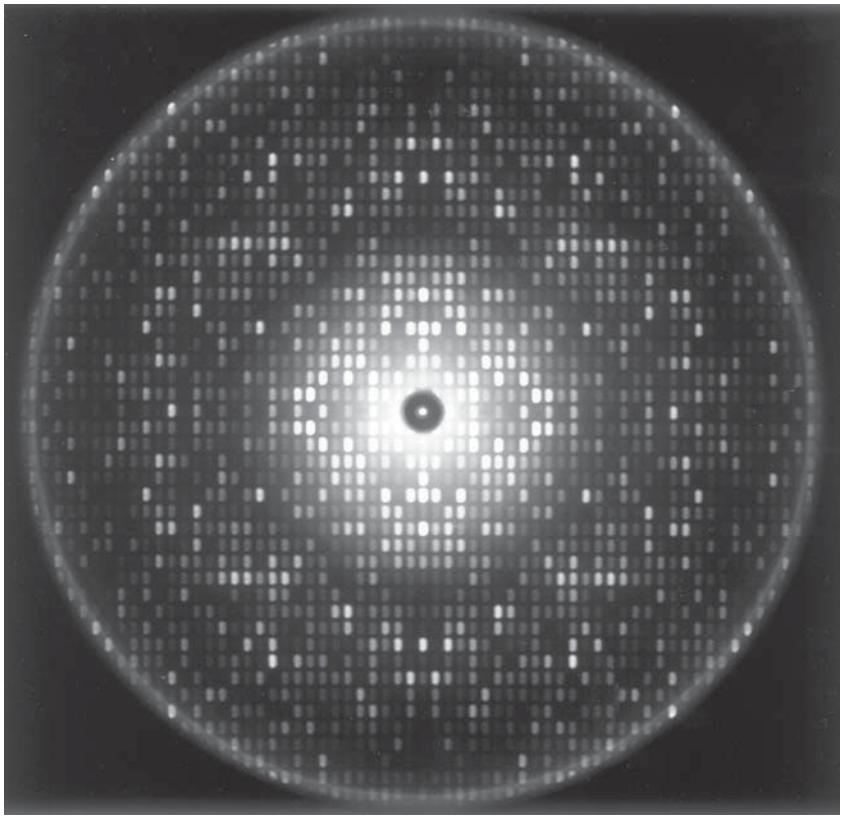


**Figure 3-20** An X-ray diffraction photograph such as was used to determine the structure of hemoglobin. This precession photograph was obtained from a crystal of human deoxyhemoglobin by rotating the crystal along two different axes in a defined manner before a narrow X-ray beam. The film also was moved synchronously. The periodicity observed in the photograph is a result of a diffraction phenomenon arising from the periodic arrangement of atoms in the crystal. The distances of the spots from the origin (center) are inversely related to the distances between planes of atoms in the crystal. In this photograph (which shows only two dimensions of the three-dimensional diffraction pattern) the spots at the periphery represent a spacing of 0.28 nm. If the intensities of the spots are measured, and if the phases of the harmonic functions required for the Fourier synthesis can be assigned correctly, the structure can be deduced to a resolution of 0.28 nm from a set of patterns of this type. In the case of human deoxyhemoglobin, a complete set of data would consist of about 16,000 spots. With modern equipment the resolution limit can be extended to better than 0.15 nm (or more than 100,000 unique spots). Photograph courtesy of Arthur Arnone.

spots, such as that in Fig. 3-20, is obtained. In a typical determination of a protein structure from 10,000 to several times that number of spots must be measured. The needed information is contained in the coordinates of the spots (i.e., in the angles through which the scattering occurs) and in the intensities of the spots. The structure is obtained by a Fourier synthesis in which a large number of mathematical functions in the form of three-dimensional sinusoidal waves are summed. The wavelengths of these functions are determined by the positions of the spots in the diffraction pattern. The amplitudes of the waves are related to the intensities of the spots.[374] However, the waves are not all in phase and the phases must be learned in some other way. This presents a difficult problem. Mathematical methods have been devised that automatically determine the phases for small molecules and usually allow the structures to be established quickly. For proteins it is more difficult.

In 1954, Perutz introduced the **isomorphous replacement method** for determining phases. In this procedure a heavy metal, such as mercury or platinum, is introduced at one or more locations in the protein molecule. A favorite procedure is to use mercury derivatives that combine with SH groups. The resulting heavy metal-containing crystals must be isomorphous with the native, i.e., the molecules must be packed the same and the dimensions of the crystal lattice must be the same. However, the presence of the heavy metal alters the intensities of the spots in the diffraction pattern and from these changes in intensity the phases can be determined. Besides the solution to the phase problem, another development that was absolutely essential was the construction of large and fast computers. It would have been impossible for Perutz to determine the structure of hemoglobin in 1937, even if he had already known how to use heavy metals to determine phases.

The first protein structure to be learned was that of myoglobin, which was established by Kendrew *et al*. in 1960.[391–393] That of the enzyme lysozyme was deduced by Blake *et al*. in 1965.[394] Since then, new structures have appeared at an accelerating rate so that today we know the detailed architecture of over 6000 different proteins[395] with about 300 distinctly different folding patterns.[396] New structures are being determined at the rate of about one per day. X-ray diffraction has also been very important to the study of naturally or artifically oriented fibrous proteins[397] and provided the first experimental indications of the β structure of proteins.

Suppose that you have isolated a new protein. How can you learn its three-dimensional structure? The first step is crystallization of the protein, something that a biochemist may be able to do. Crystallization is done in many ways, often by the slow diffusion of one solution into another or by the slow removal of solvent through controlled evaporation.[374,398] Ammonium sulfate and polyethylene glycol are two commonly used precipitants. The presence of the neutral detergent β-octyl glucoside improves some crystals.[399] Droplets of protein solution mixed with the precipitant are often suspended on microscope cover glasses in small transparent wells or are placed in depression plates within closed plastic boxes.[374,400,401] In either case, a reservoir of a solution with a higher concentration of precipitant is present in the same compartment. Water evaporates from the samples into the larger reservoir, concentrating the protein and causing its crystallization. Crystals grow slightly better in a spacecraft than on Earth.[402,403] Some proteins, notably myosin from muscle, crystallize well only after reductive methylation of all lysine side chains to dimethyllysine with formaldehyde and sodium borohydride (Eq. 3-34).[278]

The next step is for a protein crystallographer to mount a small perfect crystal in a closed silica capillary tube and to use an X-ray camera to record diffraction patterns such as that in Fig. 3-20. These patterns indicate how perfectly the crystal is formed and how well it diffracts X-rays. The patterns are also used to calculate the dimensions of the unit cell and to assign the crystal to one of the seven **crystal systems** and one of the 65 enantiomorphic **space groups**. This provides important information about the relationship of one molecule to another within the unit cell of the crystal. The unit cell (Fig. 3-21) is a parallelopiped

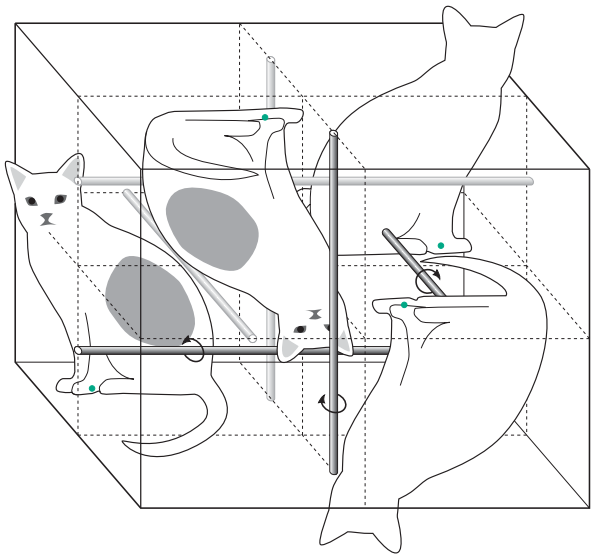


**Figure 3-21** Diagram showing an asymmetric unit as it might appear in a unit cell of space group $P2_12_12_1$. This unit cell has three pairs of nonintersecting twofold **screw axes** which are marked by the shaded rods. These are designated by arrows and the symbol. Two asymmetric units are related one to another by rotation around a twofold axis together with translation by one-half the dimension of the unit cell.

whose sides are parallel to crystallographic axes and which, by translation in three dimensions, gives rise to the entire crystal. The unit cell must contain at least one **asymmetric unit**, the smallest unit of structure that lacks any element of crystal symmetry. The asymmetric unit may consist of one or a small number of protein subunits.

In crystals of the triclinic system all of the asymmetric units are aligned in the same manner and there are no axes of symmetry. Monoclinic crystals have a single **twofold** or **dyad axis** (see Fig. 7-11). The unit cell might contain two molecules related one to another by the dyad or one dimeric molecule with the dyad located as in Fig. 7-11C. Orthorhombic crystals have three mutually perpendicular twofold axes and the unit cell is a rectangular solid. Trigonal, tetragonal, and hexagonal crystal systems have three, four and sixfold axes of symmetry, respectively, while the cubic crystal contains four threefold axes along with diagonals of the cube as well as two-fold axes passing through the faces (see Fig. 15-14). Within each crystal system there are several space groups. An example is the orthorhombic space group $P2_12_12_1$, which is often met with small organic molecules and proteins. In this space group the unit cell contains three mutually perpendicular but nonintersecting twofold screw axes (Fig. 3-21). The position of one molecule in the unit cell is related to the next by both a 180° degree rotation about the twofold axes and a translation of one-half the length of the unit cell. There is such a screw axis for each of the three directions.[372]

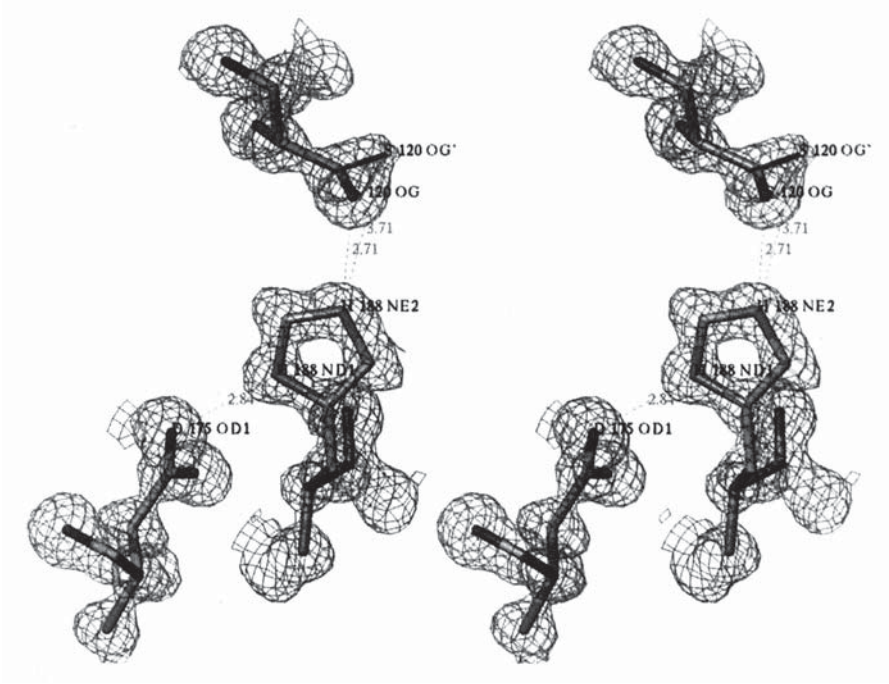



**Figure 3-22** Stereoscopic view of a section of the structure of cutinase from the fungus *Fusarium solani* determined to a resolution of 0.10 nm. The three amino acid residues shown are serine 120 (top), histidine 188, and aspartate 175 (lower left). The structure is presented as a contour map with a "wire mesh" drawn at a "cutoff" level of density equal to 1 σ above the average, where σ is the root mean square density of the entire map. The side chains of these three residues constitute the "catalytic triad" in the active site of this enzyme (see Chapter 12). At this resolution more than one conformation of a group may often be seen. For example, the gamma oxygen (OG) of S120 is seen in two positions, the major one being toward His 188. When the map is drawn with a lower contour level the N-H proton on His 188 that is hydrogen bonded to Asp 175 can also be seen.[410] Courtesy of Christian Cambillau.

The third step in the structure determination is collection of the X-ray diffraction data. This may be done with a **diffractometer** in which a narrow collimated pencil source of X-rays is aimed at the crystal and the intensities and positions of the diffracted beams are measured automatically. The computer-controlled diffractometer is able to measure the angles to within less than one-hundredth of a degree. If sufficient time is allowed, very weak spots can be counted. Today, diffractometers are more likely to be used for preliminary measurements, while the major data collection is done with an **area detector**, an instrument that collects many reflections simultaneously. There are some difficulties. For example, during the long periods of irradiation needed to measure the weak spots with a diffractometer the protein crystals decompose and must be replaced frequently. Data collection may last for months. The newest methods utilize more powerful X-ray sources and often **synchrotron radiation**, which delivers very short and extremely intense pulses of X-rays and allows data to be collected on very small well-formed crystals.[404]

The fourth step is the preparation of isomorphous crystals of heavy metal-containing derivatives. The heavy metal may be allowed to react with the protein before crystallization or may be diffused into preformed crystals. A variety of both cationic and anionic metal complexes, even large $Ta_6Br_{12}^{2+}$ tantalum clusters, have been used.[405] Two or more different heavy metal derivatives are often required for calculation of the phases. The heavy metal atoms must be present at only a very small number of locations in the unit cell.

An entire data set must be collected for each of these derivatives. The evaluation of the phases from these data is a complex mathematical process which usually involves the calculation first of a "difference Patterson projection."[406] This is derived by Fourier transformation of the differences between the scattering intensities from the native and heavy atom-containing crystals. The Patterson map is used to locate the coordinates of the heavy metal atoms which are then refined and used to compute the phases for the native protein.

Alternative methods of solving the phase problem are also used now. When a transition metal such as Fe, Co, or Ni is present in the protein, anomolous scattering of X-rays at several wavelengths (from synchrotron radiation) can be used to obtain phases. Many protein structures have been obtained using this multiple wavelength anomalous diffraction (MAD phasing) method.[404,407,408] Selenocysteine is often incorporated into a protein that may be produced in bacteria using recombinant DNA procedures. Crystals are prepared both with protein enriched in Se and without enrichment. In some ways it is better to incorporate tellurium ($^{127}$Te) in telluromethionine.[409]

In the fifth step of an X-ray structure determination the **electron density map** is calculated using the intensities and phase information. This map can be thought of as a true three-dimensional image of the molecule revealed by the X-ray microscope. It is usually displayed as a stereoscopic view on a computer graphics system (Fig. 3-22). It is also often prepared in the form of a series of transparencies mounted on plastic sheets. Each sheet represents a layer, perhaps 0.1 nm thick, with contour lines representing different levels of electron density.

Using the electron density map a three-dimensional model of the protein can be built. For years, the customary procedure was to construct a model at a scale of 2 cm = 0.1 nm using an **optical comparator**, but

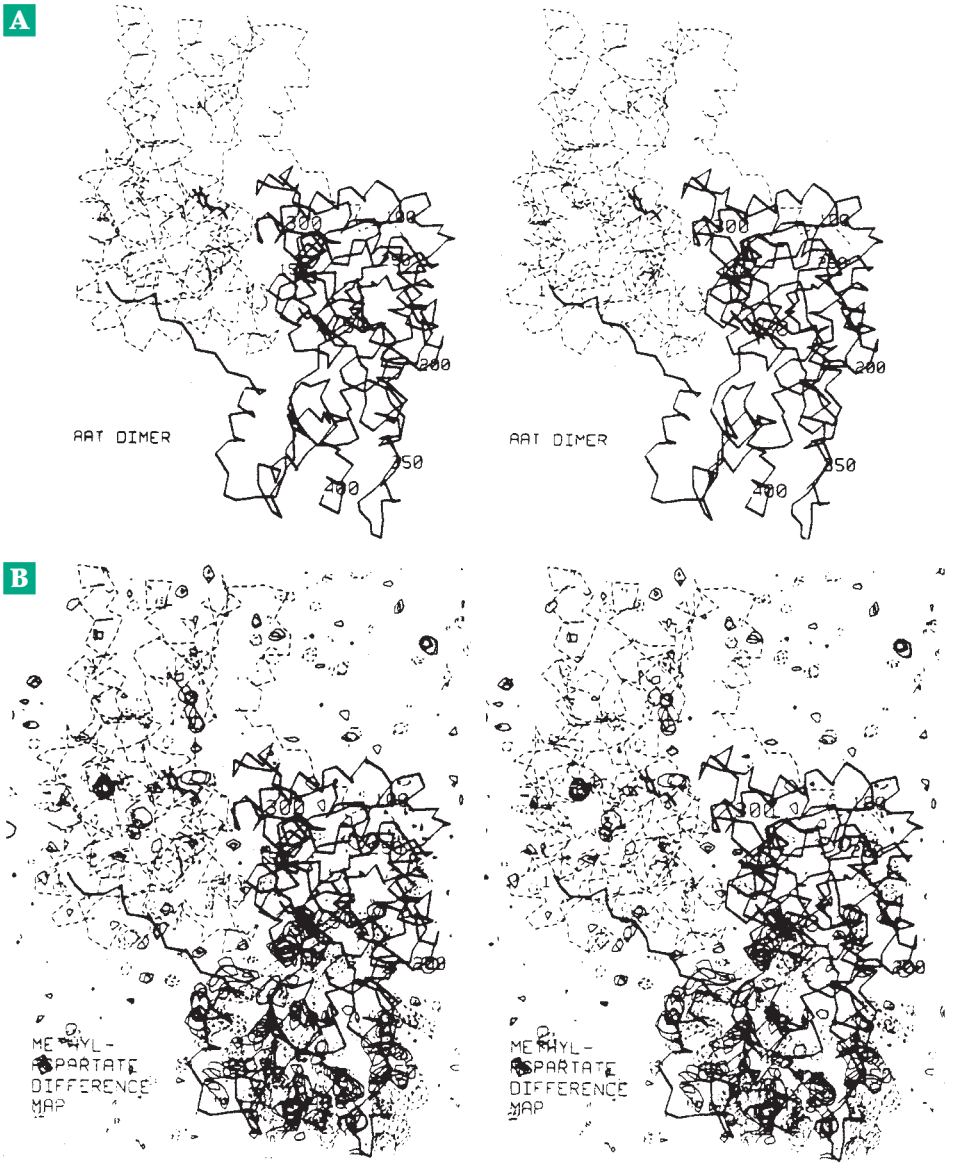



**Figure 3-23** (A) Stereoscopic α-carbon plot of the cystolic aspartate aminotransferase dimer viewed down its dyad symmetry axis. Bold lines are used for one subunit (subunit 1) and dashed lines for subunit 2. The coenzyme pyridoxal 5'-phosphate (Fig. 3-24) is seen most clearly in subunit 2 (center left). (B) Thirteen sections, spaced 0.1 nm apart, of the 2-methylaspartate difference electron density map superimposed on the α-carbon plot shown in (A). The map is contoured in increments of ± 2σ (the zero level omitted), where σ = root mean square density of the entire difference map. Positive difference density is shown as solid contours and negative difference density as dashed contours. The alternating series of negative and positive difference density features in the small domain of subunit 1 (lower right) show that the binding of L-2-methylaspartate between the two domains of this subunit induces a right-to-left movement of the small domain. (Continues)

crystallographers now use computer graphics systems. The three-dimensional image of the electron density map and a computer-generated atomic model are superimposed on the computer screen. An example is shown in Fig. 3-22. When the superposition has been completed, the coordinates of all atoms are present in the memory of the computer.

The final step in the structure determination is **refinement** using various mathematical methods. From the coordinates of the atoms in the model the expected diffraction pattern is computed and is compared with that actually observed. The differences between predicted and observed density are squared and summed. The sum of the squares constitutes an error function which is then minimized by moving the various atoms in the model short distances while keeping bond lengths, angles, and van der Waals distances within acceptable limits and recalculating the error function. This complex "refinement" procedure must be repeated literally hundreds of thousands of times with every part of the structure being varied. Structures at very high resolution (0.07–0.1 nm) may reveal multiple positions for hydrogen-bonded side chains[410a,b] as well as hydrogen atoms (Fig. 3-22) and even bonding electrons.[410b]

Once the three-dimensional structure of the protein is known, further experiments are usually done using the X-ray diffraction technique. Since protein crystals contain channels of solvent between the packed molecules, it is usually possible to diffuse small molecules into the crystal. These may be substrates, inhibitors, or allosteric effectors. Diffraction data are collected after diffusion of the small molecules into the crystal and a **difference electron density map** may be calculated. This may show exactly where those molecules were bound. An example is shown in Fig. 3-23. Sometimes difference maps not only show the binding to an enzyme of a substrate or other small molecule but also reveal conformational changes in proteins. Such is the case for Fig. 3-23, which shows the binding of α-methylaspartate, an inhibitor that behaves initially like a substrate and goes part way through the reaction sequence for aspartate aminotransferase until further reaction is blocked by the methyl group. This difference map also shows that the part of the protein to the left of the binding site in the figure has moved.[411] Using the X-ray data it was possible subsequently to deduce the nature of the conformational change.

Examination of the effect of temperature on the diffraction pattern of a protein can give direct information about the mobility of different parts of the mole-

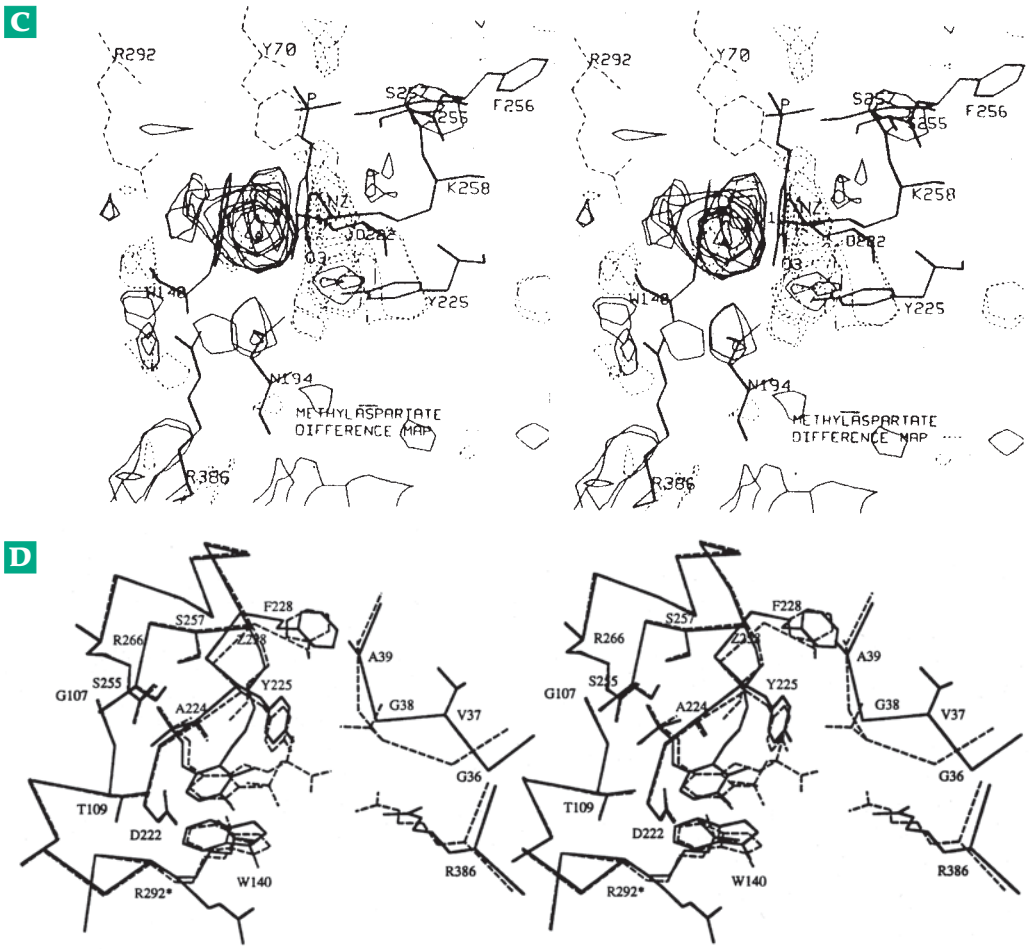
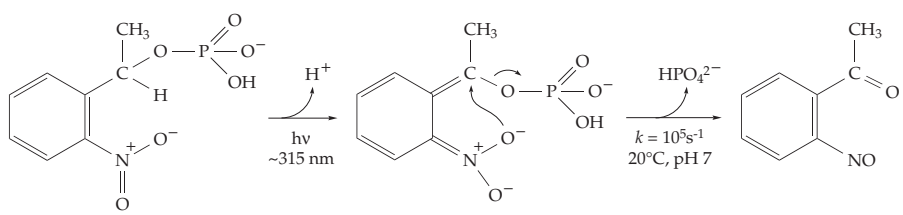



**Figure 3-23** (Cont.) (C) Nine sections, spaced 0.1 nm apart, of a part of the 2-methylaspartate difference map superimposed on the atomic model shown in (A) and (D). The coenzyme is shown as the internal aldimine with Lys 258 (see Fig. 14-6, 14-10). The positive and negative contours on the two sides of the coenzyme ring indicate that the coenzyme tilts over to form the external aldimine when substrates react.[413] (D) Superimposed structure of the active site of the enzyme in its free form as in (A) (bold lines) and the refined structure of the α-methylaspartate complex, (dashed lines).[411] This illustrates the tilting of the coenzyme ring, which is also shown in Eq. 14-39 and Fig. 14-10. Courtesy of Arthur Arnone and Sangkee Rhee.

cules.[412] This complements information obtainable in solution from NMR spectra.

Because hydrogen atoms contain only one electron, and therefore scatter X-rays very weakly, they are usually not seen at all in X-ray structures of proteins. However, neutrons are scattered strongly by hydrogen atoms and **neutron diffraction** is a useful tool in protein structure determination.[414,415] It has been used to locate tightly bonded protons that do not exchange with $^2H_2O$ as well as bound water ($^2H_2O$).

The development of synchrotron radiation as an X-ray source[404,416–418] has permitted accumulation of data for electron density difference maps in less than 1 s and it is expected that such data can eventually be acquired in ~1 ps.[419–421] If a suitable photochemical reaction can be initiated by a picosecond laser flash, a substrate within a crystalline enzyme can be watched as it goes through its catalytic cycle. An example is the release of inorganic phosphate ions from a "caged phosphate" (Eq. 3-49) and study of the reaction of the released phosphate with glycogen phosphorylase (Chapter 12).[422,423]




"Caged phosphate" (3-49)

However, caged substrates usually must diffuse some distance before reacting, so very rapid events cannot be studied. An alternative approach is to diffuse substrates into crystals at a low temperature at which reaction is extremely slow but a substrate may become seated in an active site ready to react. In favorable cases such "frozen" Michaelis complexes may be heated by a short laser pulse to a temperature at which the reaction is faster and the steps in the reaction may be observed by X-ray diffraction.[424,425]

## I. Nuclear Magnetic Resonance (NMR)

Organic chemists and biochemists alike have long relied on NMR spectroscopy to assist in identification and determination of structures of small compounds. Most students have some familiarity with this technique and practical information is available in many places.[426–428] Measurements can be done on solids, liquids, or gases but are most often done on solutions held in special narrow NMR tubes. Volumes of samples are typically 0.5 ml (for a 1–5-mM solution containing 1–5 μmol of protein) but less for small molecules and higher concentrations. Newer techniques allow use of a volume as small as 5 nl and containing <0.1 nmol of sample.[429] For many years progress in biochemical application of NMR was slow, but a dramatic increase in the power of the spectrometers, driven in great measure by the revolution in computer technology, has made NMR spectroscopy a major force in the determination of structures and functions of proteins and nucleic acids.[430,431]

### 1. Basic Principles of NMR Spectroscopy

The basis of NMR spectroscopy lies in the absorption of electromagnetic radiation at radiofrequencies by atomic nuclei.[426,427,432–437] All nuclei with odd mass numbers (e.g., $^1H$, $^{13}C$, $^{15}N$, $^{17}O$, $^{19}F$, and $^{31}P$), as well as those with an even mass number but an odd atomic number, have magnetic properties. Absorption of a quantum of energy $E = h\nu$ occurs only when the nuclei are in the strong magnetic field of the NMR spectrometer and when the frequency ν of the applied electromagnetic radiation is appropriate for "resonance" with the nucleus being observed. In the widely used "500-megahertz" NMR spectrometers the liquid helium-cooled superconducting electromagnet has a field strength of 11.75 tesla (T). In this field a proton resonates at ~500 megahertz (MHz) and nuclei of $^{31}P$, $^{13}C$, and $^{15}N$ at ~202, 125, and 50 MHz, respectively. At 500 MHz the energy of a quantum is only $E = 3.3 \times 10^{-33} \times 10^8$ J = 0.2 J mol$^{-1}$, more than four orders of magnitude less than the average energy of thermal motion of molecules (3.7 kJ mol$^{-1}$). Thus, the spin transitions induced in the NMR spectrometer have no significant effect on the chemical properties of molecules.

The resonance frequency ν at which absorption occurs in the spectrometer is given by Eq. 3-50, where $H_o$ is the strength of the external magnetic field, $\mu$ is the magnetic moment of the nucleus being investigated, and $h$ is Planck's constant. The basis for NMR spectroscopy lies in the fact that nuclei in different positions in a molecule resonates at slightly different frequencies. In a protein each one of the hundreds or thousands of protons resonate at its own frequency. With older NMR instruments a spectrum at the constant magnetic field $H_o$ can be obtained by varying the frequency and observing the values at which absorption occurs, much as is done for ultraviolet, visible, and infrared spectra (Chapter 23). With newer pulsed NMR spectrometers the measurement is done differently but the spectra

$$\nu = \mu H_o / h \qquad (3\text{-}50)$$

look the same. The higher the magnetic field, the greater the variation in resonance frequency and the higher the sensitivity. The most powerful commercial NMR spectrometers currently available operate at about 750 MHz for $^1H$ and a few higher frequency instruments have been built.

The **proton NMR** spectrum of the coenzyme pyridoxal phosphate in $^2H_2O$ is shown in Fig. 3-24 as obtained with a 60-MHz spectrometer. Four things can be measured from such a spectrum: (1) the **intensity** (area under the band). In a proton NMR spectrum, areas are usually proportional to the numbers of equivalent protons giving rise to absorption bands; (2) the **chemical shift**, the difference in frequency between the peak observed for a given proton and a peak of some standard reference compound. In Fig. 3-24 the reference peak is at the right edge; (3) the **width** at half-height (in hertz), a quantity that can provide information about molecular motion and about chemical exchange; and (4) **coupling constants** which measure interactions between nearby magnetic nuclei. These are extremely important to the determination of structures of both small and large molecules.

With a magnetic field of $H_o$ = 11.75 T and a 500-MHz oscillator, the positions of proton resonances in organic compounds are spread over a range of ~10,000 Hz. This is 20 parts per million (ppm) relative to 500 MHz. Positions of individual resonances are usually given in ppm and are always measured in terms of a shift from the resonance position of some standard substance. For protons this is most often **tetramethylsilane (TMS)**, an inert substance that can be added directly to the sample in its glass tube. Biochemists often use $^2H_2O$ as solvent and the water-soluble sodium 3-trimethylsilyl 1-propane sulfonate (**DSS** or Tier's salt) as a standard. Its position is insignificantly different from that of TMS. For NMR spectra measured in $^2H_2O$, the "pD" of the medium is sometimes indicated. It has often been taken as the pH meter reading plus 0.4. However, because of uncertainty about the meaning of pD, most workers cite the apparent pH measured with a glass electrode and standardized against aqueous buffers.[438] It is important to describe how the measurement was made when publishing results.

***Band widths.*** The narrowness of a band in an NMR spectrum is limited by the **Heisenberg uncertainty principle**, which states that $\Delta E \times \Delta t = h/2\pi$, where $h$ is Planck's constant, $\Delta E$ is the uncertainty in the energy, and $\Delta t$ is the lifetime of the magnetically excited state. Since $E = h\nu$ for electromagnetic radiation, $E$ is directly proportional to the width of the absorption band (customarily measured at one-half its full height). The magnetic nucleus is well shielded from external influences and the lifetime of its excited state tends to be long. Hence, $\Delta\nu$ is small, often amounting to less than 0.2 Hz. This fact is very favorable for the success of high-resolution proton magnetic resonance. However, bands are often much broader for large macromolecules.

***The chemical shift.*** In a molecule such as TMS, the electrons surrounding the nuclei "shield" the nucleus so that it does not experience the full external magnetic field. For this reason, absorption occurs at a high frequency (high energy). Protons that are bound

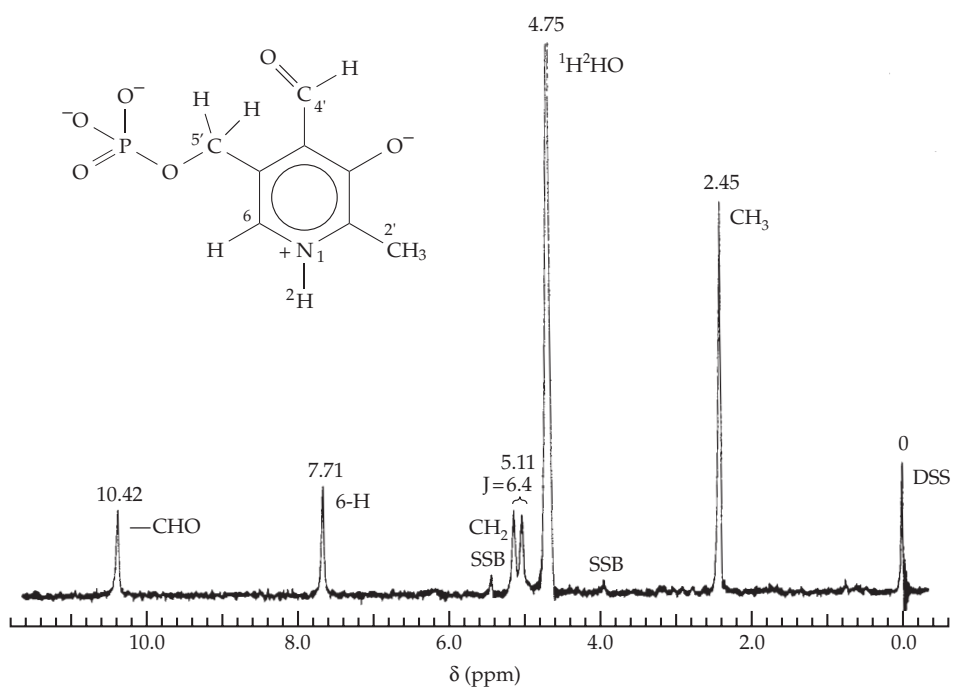



**Figure 3-24** The 60-MHz proton magnetic resonance spectrum of pyridoxal 5'-phosphate at neutral pH (apparent pH = 6.65). The internal standard is DSS. Chemical shifts in parts per million are indicated beside the peaks. Spectrum courtesy of John Likos.

to an atom deficient in electrons (because of attachment to electron withdrawing atoms or groups) are **deshielded**. The greater the deshielding, the further **downfield** from the TMS position is the NMR peak.

The magnitude of this chemical shift may be stated in hertz, but it is most often expressed in ppm as δ (Eq. 3-51). The value of δ is the shift in frequency relative to frequency of the oscillator in parts per million and is independent of the field strength. It still depends upon use of a particular reference standard which must be stated when a δ value is given. In the spectrum shown in Fig. 3-24, the methyl protons appear 2.45 ppm below the DSS peak but still at a relatively high field. Characteristic chemical shift ranges for other protons (Table 3-3) extend to ~20 ppm.

$$\delta\,(\text{ppm}) = \frac{\Delta\nu\,(\text{Hz}) \times 10^6}{\nu\,(\text{Hz})\text{ of oscillator}} \qquad (3\text{-}51)$$

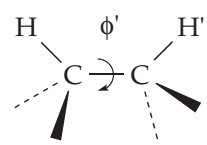
Aromatic rings lead to strong deshielding of attached protons because of a **ring current** induced in the circulating π electrons. Thus, in Fig. 3-24 the peaks of the methylene protons which are adjacent to the aromatic ring occur at 5.10 and 5.12 ppm. The 6-H, which is bound directly to the ring, is more strongly deshielded and appears at 7.71 ppm. The hydrogen of the aldehyde groups is deshielded as a result of a similar "diamagnetic electronic circulation" in the carbonyl group. Its peak is even further downfield at 10.4 ppm. Ring current and other effects on chemical shifts are important in NMR spectroscopy of proteins. Aromatic proton resonances sometimes stand out because they have been shifted far downfield. Ring current effects on chemical shifts can be predicted quite accurately if three-dimensional structures are known.[439] Additionally, computer programs are available for predicting them.[440–443]

Hydrogen bonding has a very large effect on the chemical shift of protons. The resonance of a strongly hydrogen-bonded proton is usually shifted downfield from its position in non-hydrogen-bonding media. This is especially true for hydrogen bonds to charged groups, e.g., the NH of a histidine or tryptophan side chain hydrogen bonded to a carboxylate group, a situation often met in the active sites of proteins. The chemical shift of $^{13}C$ in the carboxyl group is also affected.[444,445] While **ring current shifts** can be predicted quite well, it is much more difficult to predict the total chemical shift.[446,447]

***Scalar coupling (J coupling).*** The energy of the spin transition of a hydrogen nucleus is strongly influenced by the local presence of other magnetic nuclei, e.g., other protons that are covalently attached to the same or an adjacent atom. These neighboring protons can be in either of the two spin states, a fact that results in easily measured differences in the energy of the NMR transition under consideration. This spin–spin interaction (coupling) leads to a splitting of NMR bands of protons into two or more closely spaced bands. The ethyl group often appears in NMR spectra as a "quartet" of four evenly spaced peaks that arise from the $CH_2$ group and a triplet of peaks arising from the $CH_3$ protons. Protons attached to the same carbon (**geminal** protons), and in similar environments, do not ordinarily split each other's peaks, while the protons on the neighboring carbon do.

The **coupling constant** *J* is the difference in hertz between the successive peaks in a multiplet. It is a field-independent quantity and the same no matter what the frequency of the spectrometer. In Fig. 3-24 the peak of the methylene protons is split by $^1H–^{31}P$ coupling, with a value of $J \sim 6.4$ Hz. While spin–spin coupling is most pronounced when magnetic nuclei are close together in a structure, the effect can sometimes be transmitted through up to five covalent bonds. The technique of **double irradiation** or **spin decoupling** can be used to detect spin coupling. The sample is irradiated at the resonance frequency of one of the nuclei involved in the coupling, while the spectrum is observed in the frequency region of the other nucleus of the coupled pair. Under these conditions the multiplet collapses into a singlet and the mutual coupling of the two nuclei is established. The coupling can be seen directly in appropriate two-dimensional NMR spectra.

The coupling constant between two **vicinal protons** which are attached to adjacent carbon atoms (or other atoms) depends upon the torsion angle.



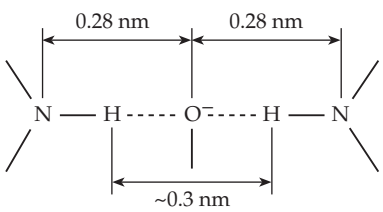

The **Karplus equation** (Eq. 3-52) relates *J* to the torsion angle ϕ′ (so labeled to distinguish it from peptide torsion angle ϕ; Fig. 2-8).

$$J_{H,H'} \quad A\ \cos^2\phi' + B\cos\phi' + C \qquad (3\text{-}52)$$

This equation was predicted on theoretical grounds, but the constants, *A*, *B*, and *C* are empirical.[450] Other forms of the equation, some of them simplified, have also been proposed.[451] The Karplus relationship is often used to estimate time-averaged torsion angles in peptides. For a $C_\alpha H–NH$ torsion angle the parameters *A*, *B* and *C* of Eq. 3-52 are ~ 6.4, 1.4, and 1.9, respectively. For a $C_\alpha H - C_\beta H$ they are ~ 9.5, 1.6, and

1.8, respectively.[452] Coupling constants between $^{1}H$ and $^{13}C$ or $^{15}N$ are of importance in determination of three-dimensional structure of proteins.[453]

**The nuclear Overhauser effect (NOE)** is the result of transfer of magnetization from one nucleus to a nearby nucleus directly through space rather than via *J*-coupling.[427,454] This was observed first as a result of irradiation of a resonance in a one-dimensional spectrum resulting in an increased intensity of the resonance of the nearby nucleus. Magnetization transfer can occur from a given nucleus to one or more nearby nuclei. Each such transfer that is detected is usually referred to simply as an NOE. For an NOE to be observable the two nuclei must be very close together, <0.5 nm (5 Å). The strength of the magnetization transfer falls off approximately as the sixth power of the interatomic distance. Consider two hydrogen atoms, both tightly hydrogen-bonded to an intervening oxygen atom, e.g., of a carboxylate or phosphate group. The expected H–H distance would be ~0.3 nm and a strong NOE between them would be anticipated. Two nonbonded hydrogen atoms (e.g., on methyl groups of amino acid side chains) can be as close together as 0.24 nm at van der Waals contact and could show a very strong NOE. However, contact is rarely this close in proteins unless in a hydrogen bond.




## 2. Nuclei Other than Hydrogen

***Deuterium ($^{2}H$).*** The natural abundance is very low so that use of $^{2}H$-labeled compounds is practical for study of metabolism, e.g., for following an $^{2}H$ label in glucose into products of fermentation[455] or in mammalian blood flow.[456] Deuterium NMR has been used extensively to study lipid bilayers (Chapter 8).

***Carbon - 13.*** Use of $^{13}C$ in NMR developed slowly because of the low natural abundance of this isotope. Another complication was the occurrence of $^{13}C$–$^{1}H$ coupling involving the many protons normally present in organic compounds. The latter problem was solved by the development of **wide-band proton decoupling** (noise decoupling). With a natural abundance of only 1.1%, $^{13}C$ is rarely present in a molecule at adjacent positions. Thus, $^{13}C$–$^{13}C$ coupling does not introduce complexities and in a noise-decoupled natural abundance spectrum each carbon atom gives rise to a single peak. Even so, $^{13}C$ NMR spectroscopy was not practical until pulsed Fourier transform (FT) spectrometers were developed (Section 2).

Chemical shifts in $^{13}C$ spectra are often 200 ppm or more downfield relative to TMS. The effects of substituents attached to a carbon atom are often additive when two or more substituents are attached to the same atom.[457,458] It is often necessary (but costly) to prepare compounds enriched in $^{13}C$ beyond the natural abundance. For proteins this may be done by growing an organism on a medium containing [$^{13}C$]glucose or a single amino acid enriched in $^{13}C$.[459,460] Using metabolites enriched in $^{13}C$, it is also possible to observe metabolism of living tissues directly. For example, glycogen synthesis from $^{13}C$-containing glucose has been observed in a human leg muscle using a wide-bore magnet and surface coils for transmitting and receiving.[461] This topic is discussed further in Box 17-C.

***Nitrogen - 15.*** Despite difficulties associated with low natural abundance (0.37%) and low sensitivity, $^{15}N$ NMR is practical and with isotopically enriched samples has become very important. Proteins with a high content of $^{15}N$ can be produced easily and inexpensively from cloned genes in bacterial plasmids. For example, cells of *E. coli* can be grown on a minimal medium containing [$^{15}N$] $NH_4Cl$. Since $^{13}C$ can also be introduced in a similar way it is possible to incorporate both isomers simultaneously. Production of uniformly labeled protein containing $^{15}N$ and / or $^{13}C$ provides the basis for multidimensional isotope-edited spectra necessary for protein structure determination (next section) and for study of tautomerization of histidine rings (Eq. 2-6).[460,462–464] $^{15}N$ chemical shifts of groups in proteins are spread over a broad range (Table 3-3).[465]

***Phosphorus - 31.*** NMR spectroscopy using $^{31}P$, the ordinary isotope of phosphorus, also has many uses.[466] Application of $^{31}P$ NMR to living tissues has been extraordinarily informative[467] and is dealt with in Chapter 6. The many phosphorus nuclei in nucleotides, coenzymes, and phosphorylated metabolites and proteins are all suitable objects of investigation by NMR techniques.

***Fluorine - 19.*** Although not abundant in nature, $^{19}F$ gives an easily detected NMR signal and can be incorporated in place of hydrogen atoms into many biochemical compounds including proteins.[468,469] In one study genetic methods were used to place 3-fluorotyrosine separately into eight positions in the lac repressor protein (Chapter 29).[470] Measurements of $^{19}F$ NMR spectra were used to study domain movement. Active site groups of enzymes can be modified to incorporate $^{19}F$.[471] Binding of fluorinated substrates can be studied.[472] Nontoxic $^{19}F$-containing compounds are useful as intracellular pH indicators, the NMR spectrometer

serving as the pH meter (Box 6-A).[473,474] An atom of fluorine attached to an aromatic ring is highly and predictibly sensitive to inductive effects of substituents in the para position to the fluorine.[475] For example, fluorine at the 6-position in pyridoxal phosphate (Fig. 14-4) can be observed in enzymes and reports changes in coenzyme structure.[476,477]

***Some other nuclei.*** Here are a few reported uses of NMR on other nuclei. $^{3}$He, binding into little cavities in fullerenes;[478] $^{11}$B, binding of boronic acids to active sites;[479] $^{23}$Na, measurement of intracellular [$Na^+$];[480–482] $^{35}$Cl and $^{37}$Cl, binding to serum albumin;[483] $^{113}$Cd, reporter that can replace $Zn^{2+}$ (no magnetic moment) in active sites of many enzymes and in nonenzymatic systems as well;[484] Tl, replacing $K^+$ in enzyme binding sites;[485] $^{17}$O, study of dynamics of protein hydration;[486] and $^{77}$Se, observation of acetylchymotrypsin intermediate.[487]

**TABLE 3-3**
**Approximate Chemical Shift Ranges in $^{1}$H- and in $^{15}$N- NMR Spectra**

| Group | $^{1}$H chemical shift (ppm from TMS) | $^{15}$N chemical shift (ppm from liquid $NH_3$) |
|---|---|---|
| $-CH_3$ | 0–4.0 | |
| $-CH_2-$ | 1.1–4.4 | |
| –CH | 2.4–5 | |
| Peptide αH random coil[a] | 3.9–5.0 | |
| –OH[b] | ~ 5–6 | |
| $>C=C<^{H}$ | 5–8 | |
| $-NH_2$ | | 31– 37 |
| NH | | |
| Peptide | 7–12 | 103 – 142 |
| Aromatic H | 7–9 | |
| Imidazole | | |
| $C^{\varepsilon 1}-H$ | ~7.7 | |
| $C^{\delta 2}-H$ | ~7.0 | |
| N–H | ~10 | 165 – 180 |
| Imidazolium | | |
| $C^{\varepsilon 1}-H$ | ~8.7 | |
| $C^{\delta 2}-H$ | ~7.4 | |
| N–H | 10–18 | |
| –CHO Aldehyde | 9.4–10.4 | |
| –COOH | 11.3–12.2 | |
| Indole NH | ~10 | 130 – 145 |
| Guanidinium | | |
| $-N^{\eta}H_2$ | | 69 –77 |
| $-N^{\varepsilon}H$ | | 31–37 |

[a] See Wishart *et al.*[448]
[b] See Linderstrøm-Lang[449]

## 3. Fourier Transform Spectrometers and Two-Dimensional NMR

Although NMR spectroscopy was widely used by the 1950s it was revolutionized by two developments, pioneered by Richard Ernst in the mid 1960s and 1970s.[488] The first of these was pulsed Fourier transform (FT) spectroscopy, which permits rapid accumulation of high-resolution spectra. In an FT NMR spectrometer a strong pulse of radiofrequency (RF) radiation is delivered to the sample over a period of a few microseconds and its effects are observed at all frequencies simultaneously. Although 1–2 or more seconds must be allowed before the next pulse is delivered, one complete NMR spectrum is obtained with each pulse. Often the results of hundreds, thousands, or even hundreds of thousands of pulses are added to provide greater sensitivity. With good temperature control this may be accomplished over periods of minutes to days.[489,490]

***Free induction decay.*** The strong exciting RF pulse is delivered with an orientation at right angles to that of the static field $H_o$ of the magnet and whose direction defines the *z* axis. As a result of this pulse, the magnetization of a nucleus is tilted away from the *z* axis and *precesses around the z axis* at its resonance (Larmor) frequency, which is ~500 MHz for $^{1}$H in a 500-MHz spectrometer. The frequency that is measured is actually a difference from the "carrier frequency" of the RF pulse. The precessing magnetization of the nuclei has a component in the *xy* plane which induces an electrical signal in the coil of the NMR probe which defines the *y* axis. This signal, which contains the Larmor frequencies of all of the nuclei of a given element, is recorded as a function of time over a period of a few seconds. The signal decays away exponentially. However, this curve of **free induction decay** (FID) is not smooth but contains within it all of the Larmor frequencies. If enough points (perhaps 500 in a 2-s acquisition) are recorded and stored in the computer's memory, Fourier transformation of the data will produce the frequency-dependent NMR spectrum.[489,490]

***Relaxation times $T_1$ and $T_2$.*** When a very strong pulse of electromagnetic radiation is applied in the NMR spectrometer, virtually all of the nuclei are placed in the magnetically excited state. If another pulse were applied immediately, little energy would be absorbed because the system is **saturated**. In the

older "continuous-wave" NMR spectrometers, the energy is always kept small so that little saturation occurs. However, in FT NMR instruments, the strong pulses lead to a high degree of saturation. Application of repeated pulses would produce no useful information were it not for the fact that the excited nuclei soon relax back to their equilibrium energy distribution. Relaxation occurs through interactions of the nuclei with fluctuating magnetic fields in the environment. For organic molecules in solution the fluctuations that are most often effective in bringing about relaxation are the result of moving electrical dipoles in the immediate vicinity. Even so, relaxation of protons in water requires seconds.

Relaxation of nuclear magnetic states is characterized by two relaxation times. The longitudinal or **spin-lattice relaxation time** $T_1$ measures the rate of relaxation of the net magnetic vector of the nuclei in the direction of $H_o$. The transverse or **spin–spin relaxation time** $T_2$ measures the relaxation in the *xy* plane perpendicular to the direction of $H_o$. The two relaxation times can be measured independently. In general, $T_2 < T_1$. For solids, $T_2$ is quite short ($\sim 10^{-5}$ s) whereas relaxation times of seconds are observed in solutions. This lengthening of the lifetime of the excited state in going from solid to liquid leads to a narrowing of absorption lines and explains why NMR bands in liquids are often narrow. However, an increase in viscosity or a loss of fluidity in a membrane leads to broadening.

How can $T_2$ and $T_1$ be measured? $T_2$ for fluids can often be estimated from the width of the band $\Delta\nu$ at half-height (Eq. 3-53). However, pulsed NMR methods are usually employed, the measurement of $T_1$ being especially easy.

$$T_2 \approx 1/\pi\,\Delta\nu \tag{3-53}$$

An attempt is often made to relate $T_1$ and $T_2$ to the molecular dynamics of a system. For this purpose a relationship is sought between $T_1$ or $T_2$ and the **correlation time $\tau_c$** of the nuclei under investigation. The correlation time is the time constant for exponential decay of the fluctuations in the medium that are responsible for relaxation of the magnetism of the nuclei. In general, $1/\tau_c$ can be thought of as a rate constant made up of the sum of all the rate constants for various independent processes that lead to relaxation. One of the most important of these ($1/\tau_1$) is for molecular tumbling.

$$1/\tau_1 \approx (3k_BT)/4\pi\eta r^3 \tag{3-54}$$

This equation is closely related to that of rotational diffusion (Eq. 9-35). Another term is the reciprocal of the **residence time** $\tau_m$, the mean time that a pair of dipoles are close enough together to lead to relaxation.

In the usual solvents at room temperature, $\tau_c$ is of the order of $10^{-12}$ to $10^{-10}$ s. Thus, relaxation rates in solutions are considerably faster than the frequencies of radiation absorbed in the NMR spectrometer ($\sim 10^8$ $s^{-1}$). Relaxation is relatively ineffective and $T_1$ and $T_2$ are usually large and equal. Bands remain sharp. As the correlation time increases (as happens, for example, if the viscosity is increased), $T_1$ and $T_2$ decrease with $T_1$ reaching a minimum when $\tau_c \sim \nu$, the frequency of the absorbed radiation. Lines are broadened and hyperfine lines (from coupling between nuclei) cannot be resolved. As $\tau_c$ is increased further, $T_2$ reaches a constant low value, while $T_1$ rises again. NMR measurements can be made in the region where $\tau_c$ exceeds $\nu$, a circumstance that is favored by the use of high-frequency spectrometers. On the other hand, in fluids it is more customary to work in the range of "extreme motional narrowing" at low values of $\tau_c$. Both $T_1$ and $T_2$ rise as the mobility of the molecules increases.

A limitation of use of NMR measurements of proteins comes from the increase in tumbling time with increasing size of the molecules. Since $1/\tau_r$ is often the most important term in the relaxation rate constant, only small proteins of mass <20 kDa give very sharp bands. Nevertheless, usable spectra are often obtainable on proteins ten times this size.

A practical problem in $^{13}C$ NMR arises from slow relaxation (long $T_1$). Partial saturation is attained and signal intensities are reduced for those carbon atoms for which relaxation is especially ineffective. Relaxation times can be measured separately for each carbon atom in a molecule and can yield a wealth of information about the **segmental motion** of groups within a molecule. Although the relationships between relaxation times and molecular motion are complex, they are often relatively simple for $^{13}C$. Carbon atoms are usually surrounded by attached hydrogen atoms, and dipole–dipole interactions with these hydrogen atoms cause most of the nuclear relaxation. For a carbon atom attached to N equivalent protons in a molecule undergoing rapid tumbling, Eq. 3-55 holds. Where $h = h/2\pi$ and $\gamma_C$ and $\gamma_H$ are the magnetogyric ratios of carbon and hydrogen nuclei.

$$1/T_1 \approx \frac{Nh^2\gamma_C{}^2\gamma_H{}^2\tau_{eff}}{r^6} \tag{3-55}$$

This equation permits a calculation of an effective correlation time $\tau_{eff}$ for each carbon atom.[491] $T_1$ and $T_2$ can also be evaluated for individual $^{15}N$ or $^{13}C$ nuclei in labeled proteins.[491]

***Two-dimensional and multidimensional NMR spectra.*** Proteins have such complex NMR spectra that, except for small regions at the upfield and downfield

ends (see Fig. 3-26A), it is impossible to interpret one-dimensional spectra. A solution to this problem came from the development by Jeener, Ernst, and Freeman of methods of displaying NMR spectra in two dimensions.[428,490,492,493] All two-dimensional and multidimensional NMR methods make use of one basic procedure: After the initial RF pulse, a second or a series of subsequent RF pulses are introduced. This is done before the nuclei have had time to relax completely. The time $t_1$ from the initial pulse to the second pulse is called the **evolution period** and allows accumulation of information about NOEs or *J*- coupling, whether homonuclear (e.g. $^1H–^1H$ or $^{13}C–^{13}C$) or heteronuclear (e.g., $^1H–^{13}C$ or $^1H–^{15}N$).

Two-dimensional spectra usually require hours or days of acquisition because separate FIDs are collected for a series of many different values of $t_1$. This provides a second timescale for the experiment. The second pulse is often used to rotate the directions of magnetization of the nuclei that are being observed from the *xy* plane into the *yz* plane but with the *z* component at 180° to the H° vector. One very important type of two-dimensional plot is the **NOESY** (NOE spectroscopy) spectrum, which detects NOEs between all excited nuclei that are close enough together (Fig. 3-25B). The transfer of magnetization occurs during a **mixing time** Δ, which follows the second pulse. For a protein Δ may be 25–300 ms. A third 90° pulse returns the *z* components of the magnetization to be parallel with the *y* axis and the FID is collected over the period $t_2$. The amplitude of the resonances detected is modulated by the frequencies that existed during the evolution

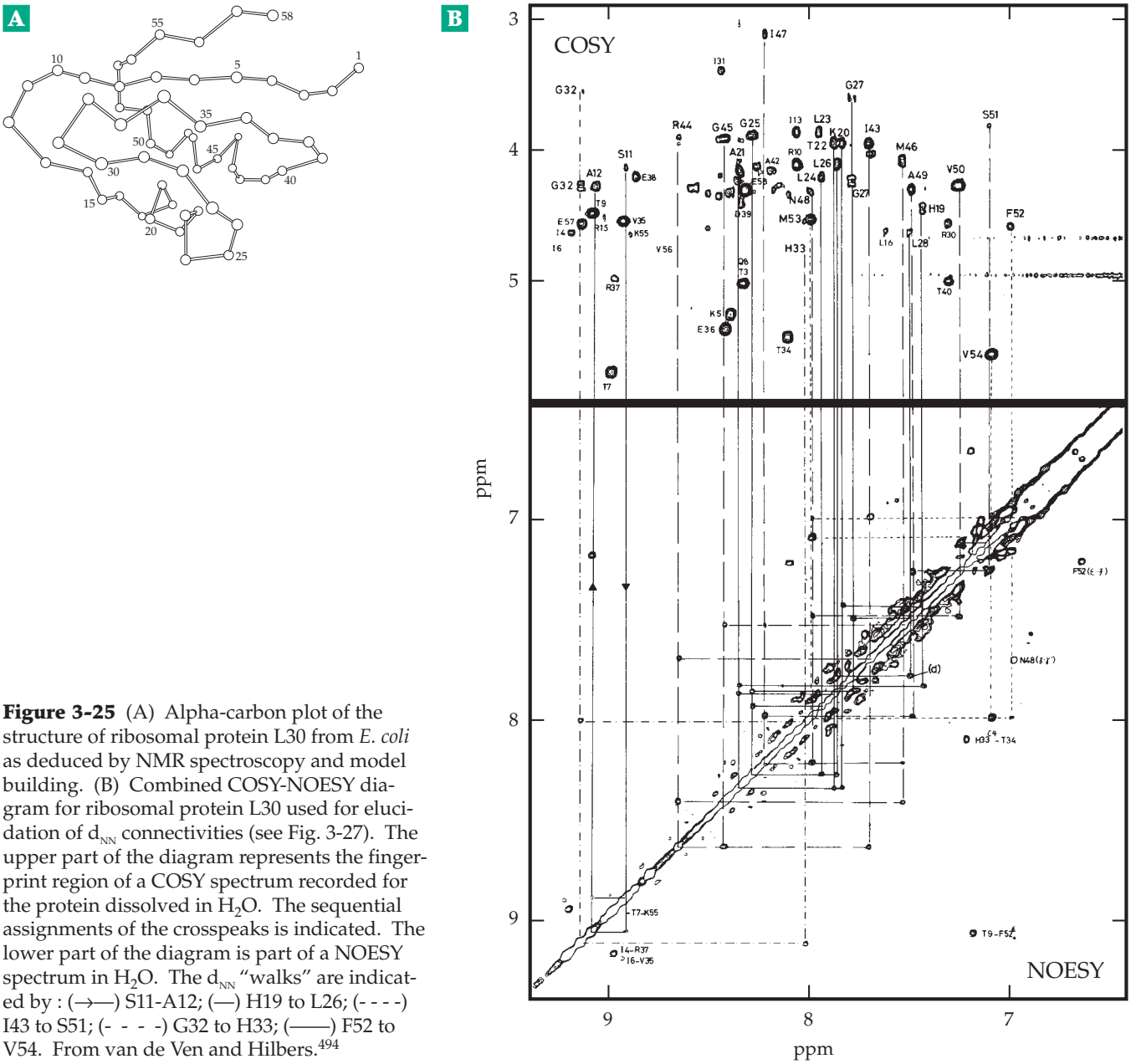



**Figure 3-25** (A) Alpha-carbon plot of the structure of ribosomal protein L30 from *E. coli* as deduced by NMR spectroscopy and model building. (B) Combined COSY-NOESY diagram for ribosomal protein L30 used for elucidation of $d_{NN}$ connectivities (see Fig. 3-27). The upper part of the diagram represents the fingerprint region of a COSY spectrum recorded for the protein dissolved in $H_2O$. The sequential assignments of the crosspeaks is indicated. The lower part of the diagram is part of a NOESY spectrum in $H_2O$. The $d_{NN}$ "walks" are indicated by : (→—) S11-A12; (—) H19 to L26; (- - - -) I43 to S51; (- - - -) G32 to H33; (——) F52 to V54. From van de Ven and Hilbers.[494]

period. After Fourier analysis[436,490] a two-dimensional plot with two frequency axes is generated and is usually displayed as a contour plot. Along the diagonal of Fig. 3-25B are peaks representing the one-dimensional spectrum. All of the peaks off of the diagonal are NOEs which can be related back to the peaks on the diagonal as shown by the horizontal and vertical lines.

A second important two-dimensional method is **correlation spectroscopy (COSY)**, in which the pairs of off-diagonal peaks result from spin–spin coupling. A related method called **TOCSY** provides correlations that extend through more than three bonds. The COSY plot in Fig. 3-26B is for the synthetic cyclic decapeptide *cyclo*-($\Delta^3$-Pro–D-*p*-Cl-Phe–D-Trp–Ser–Tyr–D-Trp–N-Me-Leu–Arg–Pro–β-Ala). It was obtained in six hours on a 500-MHZ instrument. The region marked I reveals couplings of protons on α-carbons to those on adjacent β carbons within the same residue ($J_{\alpha\beta}$; Fig. 3-27) and other couplings within the side chain. Each amino acid has a characteristic pattern. From careful study of this region it is possible to correlate each α-H resonance with a particular amino acid side chain. However, some residues are difficult to distinguish, e.g., His, Trp, Phe, and Tyr have similar $J_{\alpha\beta}$ values. Region II of Fig. 3-26B reveals connectivities of α and β hydrogens to N-H protons in the 7–9 ppm region. Each α-H is coupled to the N-H of the same residue ($J_{\alpha H}$, Fig. 3-27). A section of a COSY plot is also shown in Fig. 3-25B and indicates how resonances can be related to those in the NOESY plot made on the same sample.

Of great importance in the determination of protein structures is the use of $^{15}$N- or $^{13}$C-enriched samples to obtain **isotope-edited** spectra. For example in **HSQC** or in **$^{15}$N-multiple quantum cohenence (HMQC)** spectra we see only NH protons in a plot of $^1$H chemical shift in one dimension versus the $^{15}$N chemical shift of the attached

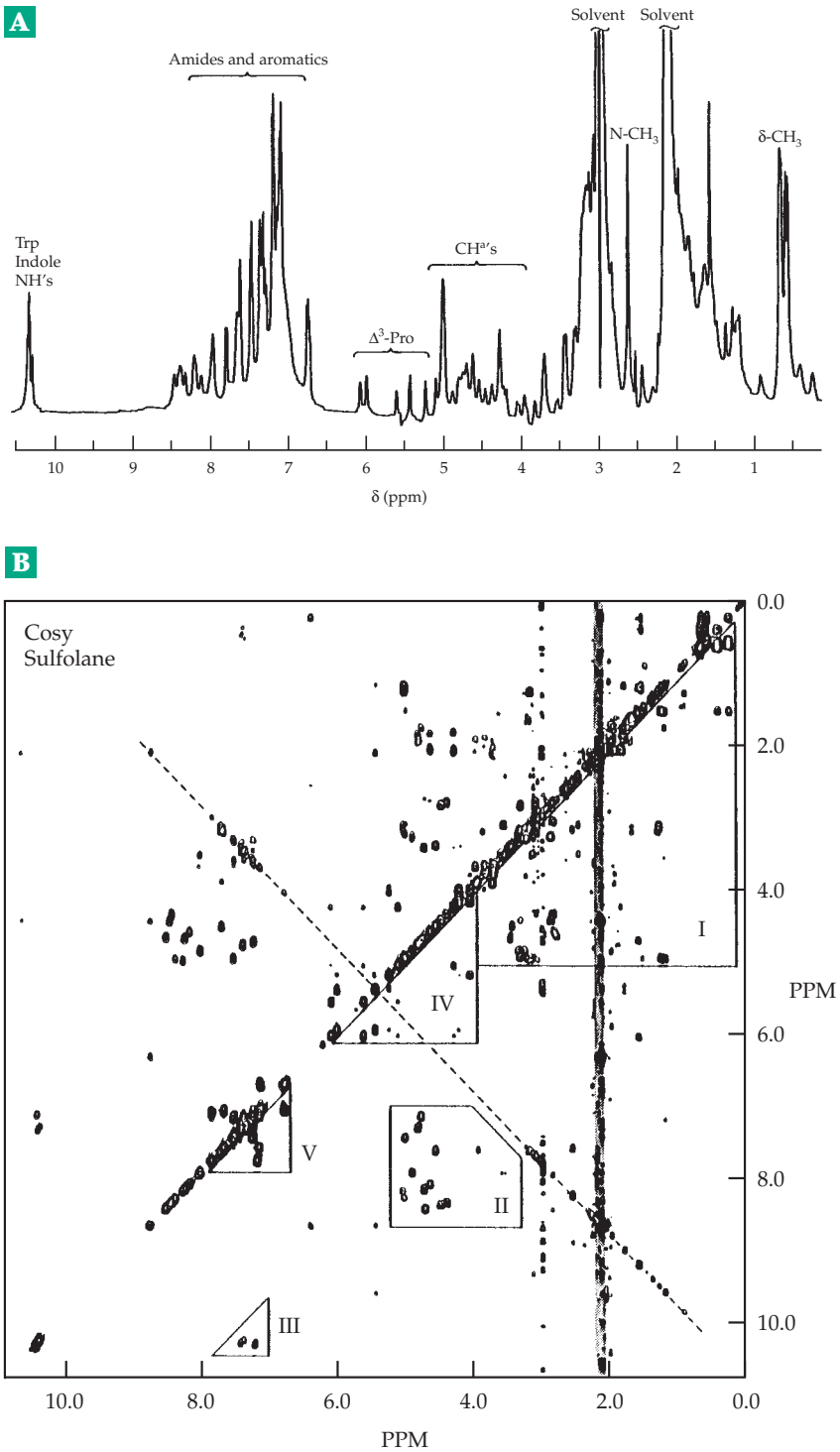



**Figure 3-26** Proton NMR spectra of a cyclic decapeptide analog of the gonadotropin-releasing hormone in the solvent sulfolane at 500 MHz. (A) One-dimensional spectrum. This figure also illustrates upfield methyl group regions, α-hydrogen, amide, aromatic, and downfield (10–20 ppm) regions. The indole NH resonances are shifted downfield by the ring current of the indole. (B) COSY spectrum plotted as a contour map. The outlined areas represent five unique *J*-coupled regions; area I, the $C^\alpha$ to $C^\beta$ to $C^\gamma$ etc., side chain connectivities; area II, the NH to $C^\alpha$ connectivities; area III, the indole NH of tryptophan; area IV, the connectivities of the $\Delta^3$-Prol residue; and area V, the aromatic resonances. Peaks that appear on the solid vertical line and the dashed diagonal lines are artifacts. From Baniak *et al.*[495]

nitrogen atom in the other (Fig. 3-28). Furthermore, as shown in this figure, particular types of NH bonds (peptide, $-NH_2$, imidazole, indole, amide, and guanidinium) appear in different regions. There is only one peptide NH per residue and, for a small protein, each may be separately visible.

## 4. Three-Dimensional Structures and Dynamics of Proteins

The first three-dimensional structure of a small protein was determined solely from NMR measurements in 1984. To date hundreds of "NMR structures" have been deduced. For small proteins of $M_r < 10{,}000$ two-dimensional COSY and NOESY spectra can suffice. For larger proteins use of $^{15}$N- and/or $^{13}$C-enriched proteins is essential.[428,430,496] This permits generation of a third and even a fourth frequency axis and three- and four dimensional NMR. For example, using various complex pulse sequences an $^{15}$N-correlated $^1$H – $^2$H NOESY spectrum can be generated. This shows directly which NOEs arise from NH protons with resonances in the amide region of an HSQC or HMQC spectrum (Fig. 3-28). With both $^{15}$N and $^{13}$C present, many additional coupling patterns and *J* values can be observed.[497–502] It is also possible to measure NOEs from atoms in a protein to those of a relatively weakly bound ligand such as a coenzyme or substrate analog and to determine the conformation of the bound ligand from this **transferred NOE**.[503–505]

***Assignment of resonances.*** After acquisition of the necessary data, which may require 10–15 mg of protein, the observed resonances must be assigned to specific amino acid residues in the peptide chain. The connectivities of the individual CH and NH groups that have been identified in the COSY spectrum and information about the relationship of one atom to another atom nearby in space are required. The closest neighbors to either αH or peptide NH protons are often protons in a neighboring residue ($d_{\alpha N}$, $d_{NN}$, Fig. 3-27). As was pointed out in the preceding section, NOE correlations obtained from plots such as that in Fig. 3-25B provide much of the information needed to establish which resonances belong to each residue in a known sequence.[494] The three-dimensional structure of the ribosomal protein L30 of *E. coli*. (Fig. 3-25A) was deduced entirely by NMR spectroscopy. A downfield part of the NOESY and COSY plots used is shown in Fig. 3-25B. This figure also shows how cross-peaks in the NOESY spectrum were correlated with identified COSY peaks. It is helpful initially to hunt for unique dipeptides that can be identified in the NOESY spectrum. Ambiguities that arise can be resolved by use of various additional techniques.

***NOEs and distance constraints.*** NOESY plots also contain the essential information needed to determine which side chains *distant* in the sequence are close together in space. A NOE observed for a pair of nuclei falls off as the inverse sixth power of the distance between them. For this reason, NOEs are observed only for pairs of atoms closer than about 0.4 nm. It is possible, in principle, to calculate distances between nuclei from the NOE intensity, but this is not accurate. Often, the NOE cross-peaks are grouped into three categories that correspond to maximum possible distances of 0.25, 0.30, and 0.40 nm. These can be related for the most part to intraresidue (e.g., $d_{N\alpha'}$ of Fig. 3-27) sequential (e.g., $d_{\alpha N'}$ and $d_{NN}$, Fig. 3-27) and long range backbone–backbone distances. These values constitute a series of **distance constraints** which are applied while making an automated computer search for a folding pattern that will meet these constraints and at the same time have acceptable torsion angles and good side chain packing throughout. This process makes use of distance geometry algorithms and other methods.[430,506,507] An early success was the solution of a 75-residue amylase inhibitor independently by crystallographers[508] and NMR spectroscopists.[509] The NMR structure was based on 401 NOE distance constraints, 168 distance constraints imposed by hydrogen bonds and 50 torsion angles deduced from *J* values. Recently, refinement of NMR structures has been done as in X-ray crystallography.[507] Some structures have been refined using both NMR and X-ray data.[442]

The spectra in Figs. 3-25, 3-26 and 3-28 are for relatively small proteins. Spectra of larger proteins are more complex and lines are broader. Many techniques are used to simplify spectra. The NMR spectrum of a protein is simplified considerably if the protein is denatured by heating, and $^1$H NMR spectra of "random coil" proteins can be predicted well from tables of standard chemical shifts for the individual amino acids.[510] Many amide NH protons exchange with solvent rapidly, making it easier to assign the remaining peaks. However, nearly all NH peaks will be seen in an HMQC or HSQC (Fig. 3-28) spectrum. Partial, or even complete, substitution of deuterium for hydrogen will also simplify spectra.[511–513] Microorganisms that will grow in a medium rich in $D_2O$ can be used as sources of partially deuterated proteins. Because the remaining protons usually have $^2$H rather than $^1$H as a neighbor, dipolar line broadening is reduced and sharper resonances are observed.[514] Substitution of $^{15}$N for $^{14}$N in the backbone amide groups can also yield spectra with narrower lines.[515] **Isotope-edited** NMR spectra allow simplification of complex two-dimensional spectra by observation of only those protons attached to an isotopically labeled nucleus, e.g., $^{13}$C or $^{15}$N.[516–518] Measurement of $^{15}$NH-$C_{\alpha H}$ *J* couplings facilitates structure determinations.[519]

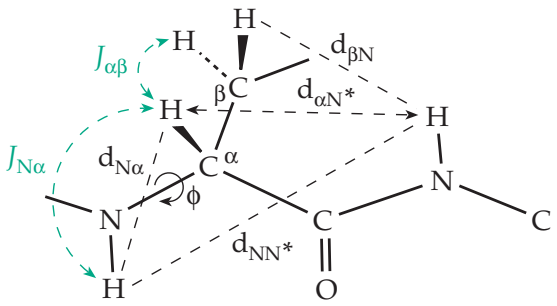



**Figure 3-27** Illustration of some distances ($d_{NN}$, $d_{\alpha N}$ and $d_{\beta N}$ obtained from NOESY spectra and some coupling constants ($J_{\alpha\beta}$ and $J_{N\alpha}$ obtained from $^1$H COSY spectra or *J*-resolved spectra. The coupling constants and proton chemical shifts provide a "fingerprint" for each residue and $J_{N\alpha}$ may also provide an estimated value for torsion angle f. The distances establish residue-to-residue connectivities as well as distance constraints that may permit a calculation of three-dimensional structure. Additional coupling constants can be measured for $^{15}$N- or $^{13}$C- enriched proteins. Coupling from β-hydrogens to other side chain hydrogens provides "fingerprint" information about individual residues.

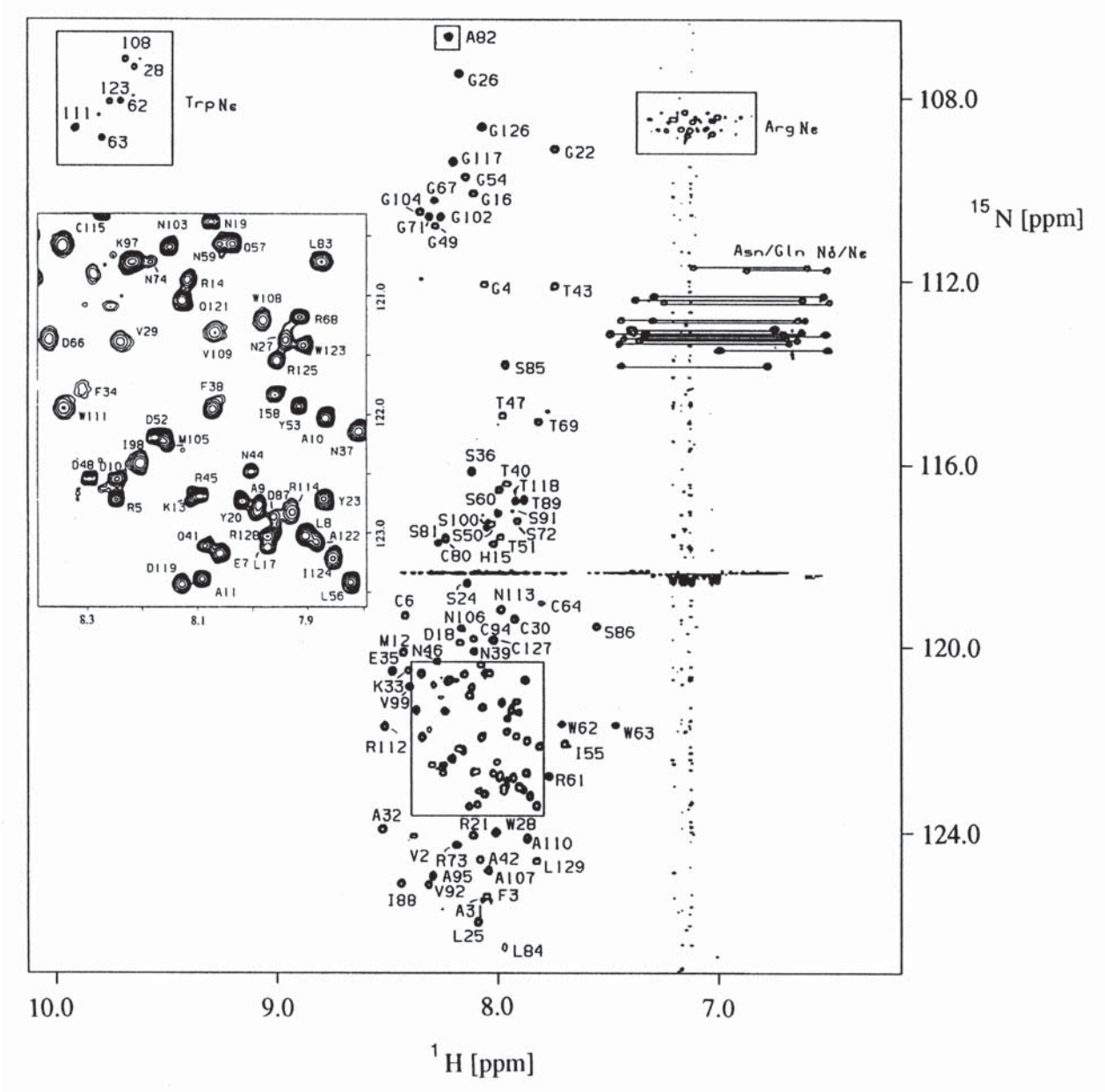



**Figure 3-28** A $^{15}$N – $^1$H HSQC spectrum of partially denatured 129-residue hen lysozyme. Boxes enclose the tryptophan indole region (upper left), the arginine side chain N$^{\varepsilon}$ region (upper left), and a portion of the amide NH region (lower center and enlarged in the insert). Resonances of pairs of hydrogen atoms in side chain (Asn and Gln) amide groups are indicated by horizontal lines. From Buck *et al.*[524]

Many aspects of the dynamics of the action of enzymes of this size can be studied by NMR spectroscopy. New approaches allow study of proteins up to ~30 kDa in size.[519a] Computer programs can analyze data and make automatic assignments of resonances.[520,521]

Also useful are techniques for measuring NMR spectra on solids, including microcrystalline proteins[522] (see also Box 9-C). In some cases, e.g., for heme proteins,[522,523] the crystals can be oriented in a magnetic field permitting measurements of NMR spectra with more than one orientation of the crystals. This potentially affords more information than the usual techniques. For example, growth of bacteria in $^2H_2O$ media containing $^1$H-containing pyruvate yields proteins with almost complete deuteration in the C-α an C-β positions but with highly protonated methyl groups. This gives rise to good – $CH_3$ to – $CH_3$ NOEs and provides other advantages in both NMR spectroscopy and mass spectrometry.[512] Isotopically enriched amino acids, e.g., $^{13}$C-enriched leucine,[525] or isoleucine containing $^{15}$N, $^{13}$C, and $^2$H as well[526] can be fed to growing bacteria. The use of paramagnetic shifts by ions such as $Gd^{3+}$ may be helpful.[527] Very large shifts are sometimes induced in heme proteins by the $Fe^{3+}$ of the heme that is embedded within the protein.[528]

## 5. Other Information from NMR Spectra

***NMR titrations.*** How do pH changes affect NMR resonances? Resonances of $^1$H nuclei close enough to a proton with a p$K_a$ in the pH region under study will experience a shift, which may be either upfield or downfield when the proton dissociates. The resonance of $^{13}$C in a carboxyl group will shift downfield when the proton on the carboxyl group dissociates. If the proton that dissociates is tightly hydrogen bonded in a protein, its rate of dissociation may be *slow* compared to the NMR frequency used. If so, the original peak will decrease in value as the pH is raised and a new peak will appear at a position characteristic of the dissociated form. However, protons attached to N or O are usually in *rapid* exchange with the solvent. In this case, the NMR resonance of the nucleus being observed will move continuously from one chemical shift value at low pH to a different one at high pH. In an intermediate case the resonance will shift and broaden.

Both the $C^{\delta 2}$-H and $C^{\varepsilon 1}$-H protons of histidine can often be seen in proteins (Fig. 3-29A).[529–531] As is shown in Fig. 3-29 A and B, their chemical shifts are strongly dependent upon the state of protonation of the ring nitrogen atoms. At low pH, the positive charge that is shared by the two NH groups attracts electrons away from both CH positions, causing deshielding of the CH protons. The effect is greater for the $C^{\varepsilon 1}$-H than for $C^{\delta 2}$-H protons (Table 3-3). If the groups being titrated do not interact strongly with other nearby basic or

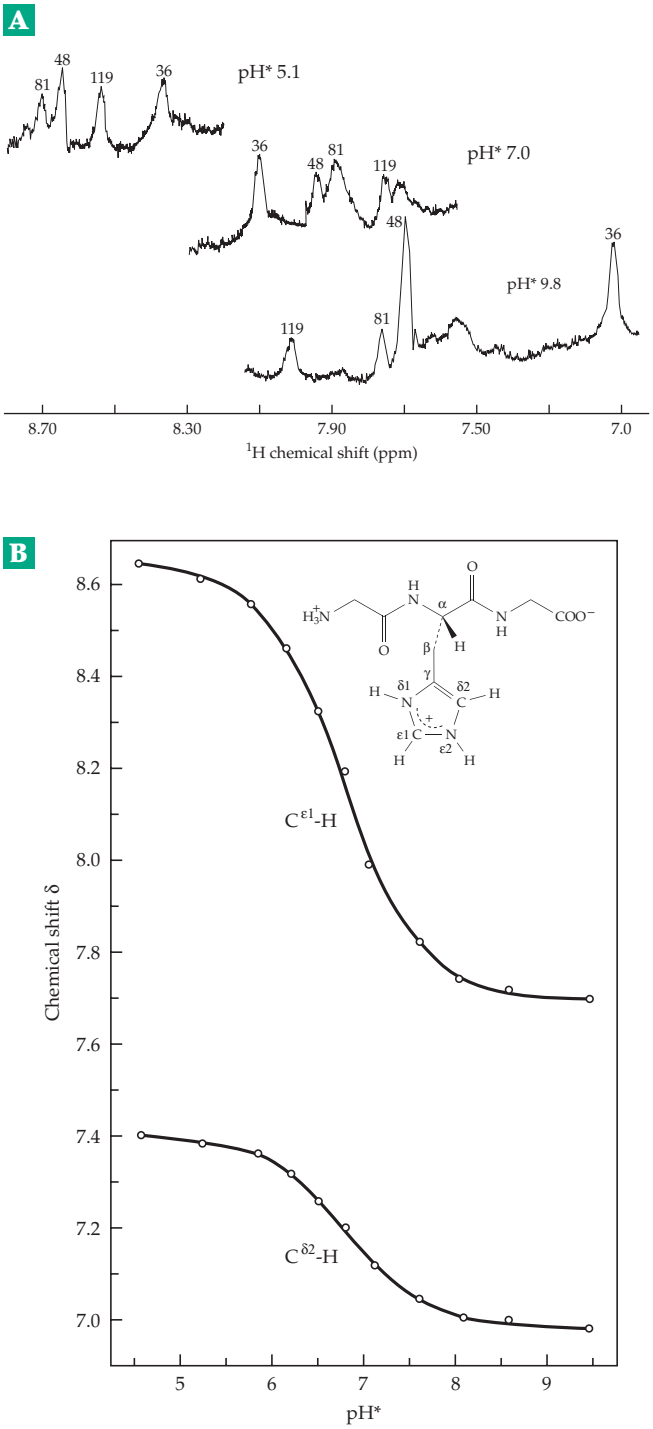



**Figure 3-29** (A) $^1$H NMR spectra of human myoglobin in $D_2O$ showing the C-H$^{\varepsilon 1}$ resonances of the imidazole rings of histidines 36, 48, 81, and 119 at three values of the apparent pH (pH*). From Bothelho and Gurd.[529] (B) $^1$H NMR titration curves for the histidine $C^{\varepsilon 1}$ and C-$^{\delta 2}$ ring protons of glycyl-L-histidylglycine. Data obtained at 100 MHz using 0.1 M tripeptide in $D_2O$ containing 0.3 M NaCl. The pH values (labeled pH*) are uncorrected glass electrode pH meter readings of $D_2O$ solutions using an electrode standardized with normal $H_2O$ buffers. Chemical shifts are downfield from TMS. Courtesy of J. L. Markley.[438]

acidic groups, p$K_a$ values can be estimated for individual histidines.[530,532] With $^{15}$N-containing proteins the tautomeric states ($NH^{\varepsilon 2}$ vs $NH^{\delta 1}$) of the imidazole rings can also be deduced.[533] Observation of imidazole rings shows that in proteins p$K_a$ values of imidazolium groups are sometimes less than five and sometimes greater than ten.[463,534]

***Observing exchangable protons.*** $^1$H spectra of proteins are often recorded in $D_2O$ because of interference from the very strong absorption of $H_2O$ at approximately 4.8 ppm. However, resonances of some rapidly exchanging protons are lost. Special pulse sequences, as well as improvements in spectrometer design, allow many of these resonances to be seen in $H_2O$.[535–537] At the far downfield end (δ>10) of $^1$H NMR spectra there are often weak peaks arising from NH protons of imidazole or indole side chains that can be observed in $H_2O$.[463,538–540] These resonances are often shifted 2–5 ppm downfield from the positions given in Table 3-3. The $^1$H resonance for a carboxyl (– COOH) proton is shifted to 20 ppm or more in spectra of very strongly hydrogen-bonded anionic complexes such as the malonate dianion.[541–544] (see Chapter 9, Section D,4). The strongest hydrogen bonds cause the greatest downfield shifts because the negative charge pulls the $^1$H proton away from the electrons of the atom to which it is attached, deshielding the proton.[545]

A good example is provided by the imidazole NH proton of the active site of trypsin and related serine proteases (Chapter 12), which is seen at ~16 ppm. Another example is provided by aspartate aminotransferase, whose $^1$H NMR spectrum in a dilute aqueous phosphate buffer is shown in Fig. 3-30. The peak labeled A is the resonance of the NH proton on the ring of the pyridoxal phosphate coenzyme (marked in Fig. 14-6) and peak B belongs to an adjacent imidazole group of histidine 143. Both of these protons move with pH changes around a p$K_a$ of ~ 6.2 which is associated with the Schiff base proton 6–8 nm away.[534,546–551] Peak A moves upfield 2.0 ppm and peak B downfield 1.0 ppm when the pH is raised around this p$K_a$. These hydrogen-bonded protons act as sensitive "reporters" of the electronic environment of the active site. Many proteins contain carboxylate or phosphate groups in their active sites and observation of NMR resonances of protons hydrogen bonded to them or to groups in substrates or inhibitors may be a useful technique for study of many enzymes and other proteins.

***Exchange rates of amide protons.*** The NH protons of the peptide backbone can be observed in $H_2O$ in the 6 –11 ppm region. If the spectrum is recorded for a sample in $D_2O$, many of these resonances disappear gradually as the protons of the peptide units exchange with the deuterium ions of the medium. A study of the observed exchange rates can shed light on the dynamics of protein molecules in solution. Exchange of amide protons in proteins with $^2$H or $^3$H has been studied on a relatively slow timescale by such techniques as observation of infrared vibrations of the amide group (see Chapter 23).[552,553] However, the development of two-dimensional NMR dramatically improved the ability to study proton exchange.[552,554–556] The measurements may involve use of quenching by rapid solvent change,[556] special pulse sequences,[555] and study of $T_1$ relaxation rates (seconds timescale). Recently, electrospray mass spectrometry has also been exploited.[557,558] Exchange patterns for large proteins may be followed,[559] and proteolytic fragmentation into short peptides may be used after various lengths of exchange time to investigate the exchange in specific regions of a protein.[216,560] Unfolding of a protein under denaturing conditions can be studied, as can refolding of a completely denatured protein.[561]

Amide hydrogen exchange is usually discussed in terms of a model proposed by Linderstrøm-Lang in 1955.[554,556] He suggested that portions of protein molecules unfold sporadically to allow rapid exchange which can be catalyzed by $H^+$, $HO^-$, or other acids or

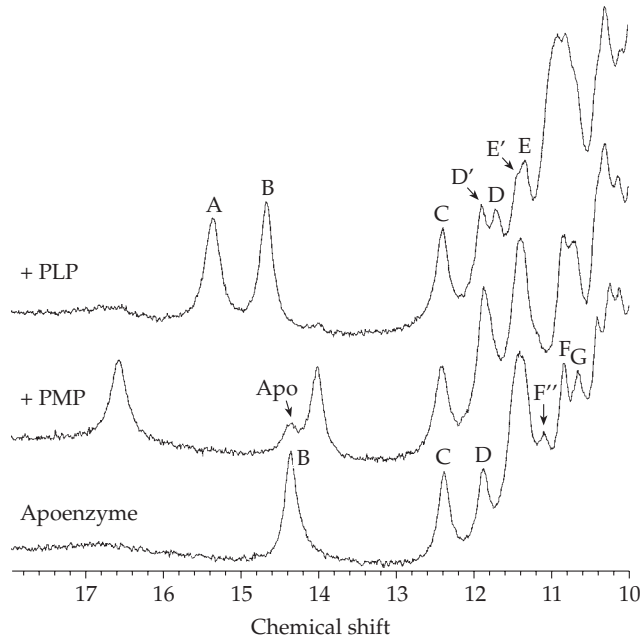



**Figure 3-30** Spectra of the pyridoxal phosphate (PLP), pyridoxamine phosphate (PMP) and apoenzyme forms of pig cytosolic aspartate aminotransferase at pH 8.3, 21°C. Some excess apoenzyme is present in the sample of the PMP form. Spectra were recorded at 500 $MH_2$. Chemical shift values are in parts per million relative to that of $H_2O$ taken as 4.80 ppm at 22°C. Peak A is from a proton on the ring nitrogen of PLP or PMP, peaks B and D are from imidazole NH groups of histidines 143 and 189 (see Fig. 14-6), and peaks C and D′ are from amide NH groups hydrogen bonded to carboxyl groups.

bases. The pH dependence of exchange for a given amide NH can be described as follows:

$$k_{ex} = k_H [H^+] + k_{OH} [OH^-] + k_W \quad (3\text{-}56)$$

where $k_H$, $k_{OH}$, and $k_W$ are rate constants for acid-catalyzed, base-catalyzed, and the very slow water-catalyzed exchange.[554,556,562] While many amide protons exchange rapidly, hydrogen-bonded NH protons in well-packed hydrophobic core regions exchange slowly,[554,563] sometimes remaining in the protein for years in a $D_2O$ solution. Some unusually stable small proteins such as the seed protein **crambin** show little exchange. The $C^{\varepsilon 1}$ protons of histidine imidazoles also exchange slowly with $D_2O$ from the medium. The exchange rates[529,564] are rapid at higher temperatures, with an average half-life of about 11 min at 65°C. Different residues may exchange at different rates. Binding of substrates or inhibitors can stabilize the protein, slowing all exchange rates of both amide and imidazole groups.

***Solid-state NMR and other topics.*** Little can be said about these topics, but NMR measurement on solid crystalline materials is now practiced and is providing a wealth of information. It is being applied more often to biochemically related problems.[465,542,565 – 567]

NMR spectroscopy of nucleic acids is discussed briefly in Chapter 5. An important medical application of NMR is in **imaging**, a topic dealt with in Box 30-A.

## J. The Protein Data Bank, Three-Dimensional Structures, and Computation

The x,y,z coordinates of all atoms in published, refined three-dimensional structures have been deposited in the Protein Data Bank (Table 3-4).[568–571] Many other related databases are available,[572] e.g., covering molecular modeling,[573] gene sequences, proteome data,[574] and much, much more. A good way to keep up to date is to read the "computer corner" in *Trends in Biochemical Sciences (TIBS).* Most databases can be reached on the World Wide Web.[572] A selected list is

**TABLE 3-4**
**Selected World Wide Web Servers Related to Protein Structures and Sequences**[a]

| Question / area | Tool | Access / URL |
|---|---|---|
| Database search by comparison of 3D structures | Dali server | *http://www.embl-heidelberg.de/dali/dali.html* |
| Structural classification of proteins | SCOP | *http://www.bio.cam.ac.uk/scop/* |
| Summary and analysis of PDP structures | PDBSum | *http://www.biochem.ucl.ac.uk/bsm/pdbsum* |
| Retrieve 3D coordinates | Protein Data Bank | *http://www.rcsb.org/pdb/* |
| SWISS-PROT sequence database, Swissmodel homology modeling, etc. | ExPASy | *http://expasy.hcuge,ch/* |
| Molecular graphics viewer for PCs and workstations | RasMol | *http://www.bernstein-plus-sons.com/software/rasmol* |
| World Wide Web -Entrez and Molecular Modeling Database access | | *http://www.ncbi.nlm.nih.gov* |
| *Protein Science* Kinemages | MAGE and PREKIN | *http://www.prosci.uci.edu/kinemages/Kinemagelndex.html* |
| Pedro's Biomolecular Research Tools | | *http://www.fmi.ch/biology/research_tools.html* |
| Predict secondary structure from sequence | Predict Protein server | *http://www.sander.embl-heidelberg.de* |
| Browse databanks in molecular biology | SRS server | *http://www.embl-heidelberg/srs/srsc* |
| The Human Genome Database | | *http://gdbwww.gdb.org/* |
| Image Library of Biological Macromolecules | | *http://www.imb-jena.de/IMAGE.html* |
| Bacterial Nomenclature | | *http://www.gbf-braunschweig.de/DSMZ/bactnom/bactnam.htm* |

[a] From Holm and Sander,[570] Hogue *et al.*,[573] Laskowski *et al.*,[580] and Walsh *et al.*[572]

given in Table 3-4. The widely used viewer called RasMol can be used with your PC or MacIntosh computer[575–576a] or UNIX workstation[569,575]. Another way to view macromolecules is to use the *Protein Science* Kinemages ("kinetic images") using the program MAGE.[577,578]

Although it is mentioned in a few places, this book does not begin to describe the rapid growth of computation in biochemistry, biophysics, and biology in general. Very fast methods of protein structure determination are being developed.[579,582] One of the major goals in current computation is to predict folding patterns of proteins from their sequences.[582,583] By comparing sequences we can often guess an approximate structure but accurate predictions are still not possible. Having a structure, we would like to predict properties and reactivities and to be able to guess how two more macromolecules interact to form macromolecular complexes. We would like to understand the complex chain of nonpolar and electrostatic interactions that underlie the fundamental properties of catalysis, movement, and responsiveness of organisms. Many computers and ingeneous minds are working to help us match theory with reality in these areas. Read the current journals!

## References


1. Strang, R. (1981) *Trends Biochem. Sci.* **6**, VII–VIII
2. Michaelis, L. (1922) *Die Wasserstoffionenkonzentration*, p. 48, Springer-Verlag, Berlin
3. Dixon, M., and Webb, E. C., eds. (1979) *Enzymes*, 3rd ed., pp.138–163, Academic Press, New York
4. Dryer, R. L., and Lata, G. F. (1989) *Experimental Biochemistry*, Oxford Univ. Press, New York
5. Kuramitsu, S., and Hamaguchi, K. (1980) *J. Biochem.* **87**, 1215 – 1219
6. Beynon, R. J., and Easterby, J. S. (1996) *Buffer Solutions: The Basics*, IRL Press, Oxford
7. Bates, R. G. (1964) *Determination of pH*, Wiley, New York
8. Stoll, V. S., and Blanchard, J. S. (1990) in *Guide to Protein Purification* (Deutscher, M. P., ed), pp. 24–38, Academic Press, San Diego, California
9. Ellis, K. J., and Morrison, J. F. (1982) *Methods Enzymol.* **87**, 405–426
10. Good, N. E., Winget, G. D., Winter, W., Connolly, T. N., Izawa, S., and Singh, R. M. M. (1966) *Biochemistry* **5**, 467–477
11. Ferguson, W. J., Braunschweiger, K. I., Braunschweiger, W. R., Smith, J. R., McCormick, J. J., Wasmann, C. C., Jarvis, N. P., Bell, D. H., and Good, N. E. (1980) *Anal. Biochem.* **104**, 300–310
12. Dawson, R. M. C., Elliot, D. C., Elliot, W. H., and Jones, K. M. (1986) *Data for Biochemical Research*, 3rd ed., Oxford Science Pub., London
13. Robyt, J. F., and White, B. J. (1990) *Biochemical Techniques, Theory and Practice*, 2nd ed., Waveland Press, Inc., Prospect Height, Ill
14. Cooper, T. G. (1977) *The Tools of Biochemistry*, Wiley, New York
15. Alexander, R. R., and Griffiths, J. M. (1992) *Basic Biochemical Methods*, 2nd ed., Wiley, New York
16. Wilson, K., and Walker, J., eds. (1994) *Principles and Techniques of Practical Biochemistry*, 4th ed., Cambridge Univ. Press, New York
17. Boyer, R. F. (1992) *Modern Experimental Biochemistry*, 2nd ed., Benjamin-Cummings Publ., Redwood City, California
18. le Maire, M., Chabaud, R., and Herve, G. (1991) *Laboratory Guide to Biochemistry, Enzymology, and Protein Physical Chemistry*, Plenum, New York
19. Fini, C., Floridi, A., and Finelli, V. N., eds. (1989) *Laboratory Methodology in Biochemistry*, CRC Press, Boca Raton, Florida
20. Scopes, R. K. (1993) *Protein Purification: Principles and Practice*, 3rd ed., Springer, New York
21. Harris, E. L. V., and Angal, S. (1990) *Protein Purification: A Practical Approach*, IRL Press, Oxford, 2 volumes
22. Deutscher, M. P., ed. (1990) *Guide to Protein Purification: Methods in Enzymology*, Vol. 182, Academic Press, San Diego, California
23. Celis, J. E., ed. (1994) *Cell Biology: A Laboratory Handbook*, Academic Press, New York (3 volumes)
24. Fasman, G. D., ed. (1989) *CRC Practical Handbook of Biochemistry and Molecular Biology*, CRC Press, Boca Raton, Florida
25. Rosenberg, I. M. (1996) *Protein Analysis and Purification*, Birkhäuser, Cambridge, Massachusetts
26. Walker, J. M., ed. (1996) *The Protein Protocols Handbook*, Humana Press, Totowa, New Jersey
27. Marshak, D. R., Kadonaga, J. T., Burgess, R. R., Knuth, M. W., Brennan, W. A., Jr., and Lin, S.-H. (1996) *Strategies for Protein Purification and Characterization*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
28. Creighton, T. E., ed. (1989) *Protein Structure: A Practical Approach*, IRL Press, Oxford
29. Marshak, D. R., Kadonaga, J. T., Burgess, R. R., Knuth, M. W., Brennan, W. A., and Lin, S.-H. (1995) *Strategies for Protein Purification and Characterization*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
30. Crabb, J. W., ed. (1995) *Techniques in Protein Chemistry*, Vol. VI, Academic Press, San Diego, California
31. Walker, J. M., ed. (1994) *Basic Protein and Peptide Protocols*, Humana Press, Totowa, New Jersey
32. Bollag, D. M., and Edelstein, S. J. (1990) *Protein Methods*, Wiley-Liss, New York
33. Creighton, T. E., ed. (1989) *Protein Structure and Protein Function: A Practical Approach*, IRL Press, Oxford
34. Copeland, R. A. (1994) *Methods for Protein Analysis. A Practical Guide to Laboratory Protocols*, Chapman and Hall, New York
35. Matsuidaira, P. T., ed. (1995) *A Practical Guide to Protein and Peptide Purification for Microsequencing*, Academic Press, San Diego, CA
36. Hugli, T. E., ed. (1989) *Techniques in Protein Chemistry*, Academic Press, New York
37. Means, G. E., and Feeney, R. E. (1973) *Chemical Modification of Proteins*, Holden-Day, San Francisco
38. Keleti, G., and Lederer, W. H. (1974) *Handbook of Micromethods for the Biological Sciences*, Van Nostrand-Reinhold, New York
39. Switzer, R. L., and Garrity, L. F. (1999) *Experimental Biochemistry:Theory and Exercises in Fundamental Methods*, 3rd ed., Freeman, New York
40. Schulz, G. E., and Schirmer, R. H. (1979) *Principles of Protein Structure*, Springer-Verlag, New York
41. Scopes, R. K. (1994) *Protein Purification. Principles of Protein Structure*, Springer-Verlag, New York
42. Villafranca, J., ed. (1990) *Current Research in Protein Chemistry: Techniques, Structure, and Function*, Academic Press, New York
43. Coligan, J., Dunn, B., Ploegh, H., Speicher, D., and Wingfield, P., eds. (1995) *Current Protocols in Protein Science*, Wiley, New York
44. Matsudaira, P. T., ed. (1989) *A Practical Guide to Protein and Peptide Purification for Microsequencing*, Academic Press, San Diego, California
45. Doonan, S., ed. (1996) *Protein Purification Protocols*, Humana Press, Totowa, New Jersey
46. Little, P., and Hughes, S., eds. (1992) *Technique*, Vol. 4, Academic Press, London
47. *Methods: A Companion to Methods in Enzymology*, 1990–present, Academic Press, San Diego, California
48. Wankat, P. C., Van Oss, C. J., and Henry, J. D., Jr., eds. (1992) *Separation and Purification Methods*, Dekker, New York
49. *protein expression and purification*, (1990 – present), Academic Press, San Diego, California
50. Coligan, J., Dunn, B., Ploegh, H., Speicher, D., and Wingfield, P., eds. (1996–present) *Current Protocols in Protein Science*, Wiley, New York
51. Hirs, C. H. W., and Timasheff, S. N., eds. (1983) *Methods in Enzymology, Enzyme Structure*, Vol. 91, Academic Press, New York
52. Doolittle, R. F., ed. (1990) *Methods in Enzymology, Molecular Evolution: Computer Analysis of Protein and Nucleic Acid*, Vol. 183, Academic Press, New York
53. Bollag, D. E., Rozycki, M. D., and Edelstein, S. J., eds. (1996) *Protein Methods*, 2nd ed., Wiley, New York
54. Birnie, G. D., and Fox, S. M., eds. (1969) *Subcellular Components*, Butterworth, London
55. Estabrook, R. W., and Pullman, M. E., eds. (1967) *Methods in Enzymology*, Vol. 10, Academic Press, New York
56. Kaback, H. R. (1971) in *Methods in Enzymology*, Vol. 22 (Jakoby, W. B., ed), pp. 99–120, Academic Press, New York
57. Dingle, J. T. (1972) *Lysosomes. A Laboratory Handbook*, North-Holland Publ., Amsterdam
58. Rasmussen, N. (1996) *Trends Biochem. Sci.* **21**, 319–321

58a. Jenkins, W. T. (1998) *Protein Sci.* **7**, 376–382

59. McPhie, P. (1971) in *Methods in Enzymology*, Vol. 22 (Jakoby, W. B., ed), pp. 23–32, Academic Press, New York
60. Blatt, W. F. (1971) in *Methods in Enzymology*, Vol. 22 (Jakoby, W. B., ed), pp. 39–49, Academic Press, New York
61. Porath, J., and Flodin, P. (1959) *Nature (London)* **183**, 1657–1659

# References


62. Porath, J. (1978) *Trends Biochem. Sci.* **3**, N100
63. Lathe, G. H. (1978) *Trends Biochem. Sci.* **3**, N99–N100
64. Flodin, P., and Aspberg, K. (1961) *Biol. Struct. Funct. Proc. IUB/IUBS Int. Sym*, 1st ed., Vol. 1, pp. 345-349 (p. 346)
65. Weir, M., and Sparks, J. (1987) *Biochem. J.* **245**, 85
66. Regnier, F. E. (1991) *Nature (London)* **350**, 634–635
67. Afeyan, N. B., Gordon, N. F., and Mazsaroff, I. (1990) *J. Chromatogr.* **519**, 1–29

67a. Wu, C.-s. (1999) *Column Handbook for Size Exclusion Chromatography*, Academic Press, San Diego, California

68. Liapis, A. I., and McCoy, M. A. (1992) *J. Chromatogr.* **599**, 87–104
69. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry, Part II*, Freeman, San Francisco, California (pp. 549–685)
70. Kyte, J. (1995) *Structure in Protein Chemistry*, Garland Publ., New York
71. Brakke, M. K. (1953) *Arch. Biochem. Biophys.* **45**, 275–290
72. Britten, R. J., and Roberts, R. B. (1960) *Science* **131**, 32–33
73. Martin, R. G., and Ames, B. N. (1961) *J. Biol. Chem.* **236**, 1372–1379
74. Meselson, M., Stahl, F. W., and Vinograd, J. (1957) *Proc. Natl. Acad. Sci. U.S.A.* **43**, 581–588
75. da Silva, J. R. R. F. (1978) in *New Trends in Bio-organic Chemistry* (Williams, R. J. P., and da Silva, J. R. R. F., eds), Academic Press, London
76. Arakawa, T., and Timasheff, S. N. (1982) *Biochemistry* **21**, 6545–6552
77. Rao, J. K. M., and Argos, P. (1981) *Biochemistry* **20**, 6536–6543
78. Green, A. A., and Hughes, W. L. (1955) in *Methods in Enzymology*, Vol. 1 (Colowick, S. P., and Kaplan, N. O., eds), pp. 67–90, Academic Press, New York
79. Dawson, R. M. C., Elliott, D. C., Elliott, W. H., and Jones, K. M. (1969) *Data for Biochemical Research*, 2nd ed., Clarendon Press, Oxford
80. Ito, Y. (1982) *Trends Biochem. Sci.* **7**, 47–50
81. Ito, Y. (1987) *Nature (London)* **326**, 419–420
82. Mandava, N. B., and Ito, Y., eds. (1988) *Countercurrent Chromatography*, Dekker, New York

82a. Menet, J.-M., and Thiébaut, D., eds. (1999) *Countercurrent Chromatography*, Dekker, New York

83. Er-el, Z., Zaidenzaig, Y., and Shaltiel, S. (1972) *Biochem. Biophys. Res. Commun.* **49**, 383–390
84. Arakawa, T., and Timasheff, S. N. (1982) *Biochemistry* **21**, 6536–6544
85. Atha, D. H., and Ingham, K. C. (1981) *J. Biol. Chem.* **256**, 12108–12117
86. Kenney, A., and Fowell, S., eds. (1992) *Practical Protein Chromatography*, Humana Press, Totowa, New Jersey
87. Tiselius, A., Hjerten, S., and Levin, O. (1956) *Arch. Biochem. Biophys.* **65**, 132
88. Wolf, W. J., and Sly, D. A. (1964) *J. Chromatogr.* **15**, 247–250
89. Regnier, F. E. (1987) *Science* **238**, 319–323
90. Hutchens, T. W., and Porath, J. (1987) *Biochemistry* **26**, 7199–7204
91. Molnár, I., Horváth, C., and Jatlow, P. (1978) *Chromatographia* **11**, 260–265
92. Hamilton, R. J., and Sewell, P. A. (1982) *Introduction to High Performance Liquid Chromatography*, 2nd ed., Chapman and Hall, London
93. Berridge, J. C. (1985) *Techniques for the Automated Optimization of HPLC Separations*, Wiley, New York
94. Simpson, C. F. (1982) *Techniques in Liquid Chromatography*, Wiley, New York
95. Mant, C. T., and Hodges, R. S. (1991) *High-Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis, and Conformation*, CRC Press, Boca Raton, Florida
96. Gooding, K. M., and Regnier, F. E., eds. (1990) *HPLC of Biological Macromolecules*, Dekker, New York

96a. Beesley, T. E., Buglio, B., and Scott, R. P. W. (2000) *Quantitative Chromatographic Analysis*, Dekker, New York

97. Oliver, R. W. A., ed. (1989) *HPLC of Macromolecules: A Practical Approach*, IRL Press, Oxford
98. Fried, B., and Sherma, J., (1999) *Thin-Layer Chromatography*, 4th ed., Dekker, New York
99. Zlatkis, A., and Kaiser, R. E. (1977) *High Performance Thin-Layer Chromatography*, Elsevier Scientific-North Holland, Amsterdam
100. Sherma, J., and Fried, B., eds. (1996) *Practical Thin-Layer Chromatography: A Multidisciplinary Approach*, CRC Press, Boca Raton, Florida
101. Liu, M.-K., Li, P., and Giddings, J. C. (1993) *Protein Sci.* **2**, 1520–1531
102. Treadwell, G. E., Cairns, W. L., and Metzler, D. E. (1968) *J. Chromatogr.* **35**, 376–388
103. Novotny, M. V. (1989) *Science* **264**, 51–57
104. Schomvurg, G. (1990) *Gas Chromatography: A Practical Course*, VCH Publ., Cambridge, UK
105. Lee, M. L., and Markides, K. E. (1987) *Science* **235**, 1342–1347

105a. Muraviev, D., Gorshkov, V., and Warshawsky, A. (1999) *Ion Exchange*, Dekker, New York

106. Moore, S., and Stein, W. H. (1949) *Cold Spring Harbor Symp. Quant. Biol.* **14**, 179–190
107. Moore, S., and Stein, W. H. (1949) *J. Biol. Chem.* **178**, 53–77
108. Manning, J. M. (1993) *Protein Sci.* **2**, 1188–1191
109. Svasti, J. (1980) *Trends Biochem. Sci.* **5**, 8–9
110. Moore, S., and Stein, W. H. (1951) *J. Biol. Chem.* **192**, 663–681
111. Moore, S., and Stein, W. H. (1954) *J. Biol. Chem.* **211**, 893–906
112. Moore, S., and Stein, W. H. (1963) in *Methods in Enzymology*, Vol. 6 (Colowick, S. P., and Kaplan, N. O., eds), pp. 819–831, Academic Press, New York
113. Yamamoto, S., Nakanishi, K., and Matsuno, R. (1988) *Ion Exchange Chromatography of Proteins*, Dekker, New York
114. Peterson, E. A. (1970) *Cellulosic Ion Exchangers*, Elsevier-North Holland, Amsterdam
115. Nölting, B., Golbik, R., Neira, J. L., Soler-Gonzalez, A. S., Schreiber, G., and Fersht, A. R. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 826–830
116. Lundell, N., and Markides, K. (1992) *Chromatographia* **34**, 369–375
117. Dean, P. D. G., Johnson, W. S., and Middle, F. A., eds. (1985) *Affinity Chromatography, A Practical Approach*, IRL Press, Oxford
118. Parikh, I., and Cuatrecasas, P. (1985) *Chem. Eng. News* **63** (Aug 26), 17–32
119. Porath, J. (1968) *Nature (London)* **218**, 834–838
120. Inman, J. K. (1974) in *Methods in Enzymology*, Vol. 34 (Jakoby, W. B., and Wilchek, M., eds), pp. 30–58, Academic Press, New York
121. Bethell, G. S., Ayers, J. S., Hearn, M. T. W., and Hancock, W. S. (1981) *J. Chromatogr.* **219**, 361–372
122. Wilchek, M., and Miron, T. (1987) *Biochemistry* **26**, 2155–2161
123. Ashley, G. W., and Bartlett, P. A. (1984) *J. Biol. Chem.* **259**, 13615–13620
124. Hengen, P. N. (1995) *Trends Biochem. Sci.* **20**, 285–286
125. Loo, T. W., and Clarke, D. M. (1995) *J. Biol. Chem.* **270**, 21449–21452
126. Janknecht, R., de Martynoff, G., Lou, J., Hipskind, R. A., Nordheim, A., and Stunnenberg, H. G. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8972–8976
127. Hochuli, E., Dobeli, H., and Schacher, A. (1987) *J. Chromatogr.* **411**, 177–184
128. Turner, A. J. (1981) *Trends Biochem. Sci.* **6**, 171–173
129. Vesterberg, O. (1971) in *Methods in Enzymology*, Vol. 22 (Jakoby, W. B., ed), pp. 389–412, Academic Press, New York
130. Bauer, J. (1994) *Cell Electrophoresis*, CRC Press, Boca Raton, Florida
131. Allison, S. A., and Tran, V. T. (1995) *Biophys. J.* **68**, 2261–2270
132. Shuster, L. (1971) in *Methods in Enzymology*, Vol. 22 (Jakoby, W. B., ed), pp. 412–433, Academic Press, New York
133. Chrambach, A., and Rodbard, D. (1971) *Science* **172**, 440–451
134. Andrews, A. T. (1986) *Electrophoresis: Theory, Techniques and Biochemical and Clinical Applications*, 2nd ed., Clarendon-Oxford Univ. Press, Oxford
135. Gordon, A. H. (1980) *Electrophoresis of Proteins in Polyacrylamide and Starch Gels*, Revised ed., Elsevier-North Holland, Amsterdam
136. Chrambach, A., Dunn, M. J., and Radola, B. J., eds. (1987) *Advances in Electrophoresis*, Vol. I, VCH Publ., Cambridge, UK
137. Serwer, P., Easom, R. A., Hayes, S. J., and Olson, M. S. (1989) *Trends Biochem. Sci.* **14**, 4–7
138. *electrophoresis*, (1995–present), VCH Publ., New York
139. *advances in electrophoresis*, (1987–present), VCH Publ., Cambridge
140. Hawcroft, D. M. (1997) *Electrophoresis: The Basics*, IRL Press, Oxford
141. Hengen, P. N. (1995) *Trends Biochem. Sci.* **20**, 202–203
142. Garfin, D. E. (1990) *Methods Enzymol.* **182**, 425–441
143. Hames, B. D., and Rickwood, D., eds. (1990) *Gel Electrophoresis of Proteins: A Practical Approach*, 2nd ed., IRL Press, Oxford
144. Righetti, P. G., ed. (1996) *Capillary Electrophoresis in Analytical Biotechnology*, CRC Press, Boca Raton, Florida
145. Camilleri, P., ed. (1993) *Capillary Electrophoresis: Theory and Practice*, CRC Press, Inc., Boca Raton
146. Altria, K. D., ed. (1996) *Capillary Electrophoresis Guidebook*, Humana Press, Totowa, New Jersey
147. Landers, J. P., ed. (1994) *Handbook of Capillary Electrophoresis*, CRC Press, Boca Raton, Florida
148. Landers, J. P. (1993) *Trends Biochem. Sci.* **18**, 409–414
149. Wehr, T., Rodríguez-Díaz, R., and Zhu, M. (1998) *Capillary Electrophoresis of Proteins*, Dekker, New York

149a. Yao, S., Anex, D. S., Caldwell, W. B., Arnold, D. W., Smith, K. B., and Schultz, P. G. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5372–5377

150. Valaskovic, G. A., Kelleher, N. L., and McLafferty, F. W. (1996) *Science* **273**, 1199–1202
151. Gao, J., Gomez, F. A., Härter, R., and Whitesides, G. M. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12027–12030
152. Harrison, D. J., Fluri, K., Seiler, K., Fan, Z., Effenhauser, C. S., and Manz, A. (1993) *Science* **261**, 895–897
153. Radola, B. J., and Graesslin, D., eds. (1977) *Electrofocusing and Isotachophoresis*, de Gruyter, Berlin
154. Bocek, P., Dem, M., Gebauer, P., and Dolnik, V. (1987) *Analytical Isotachophoresis*, VCH Publ., Cambridge, UK

## References


155. Righetti, P. G. (1983) *Isoelectric Focusing: Theory, Methodology and Applications*, Elsevier - North Holland Biomedical Press, Amsterdam
156. Catsimpoolas, N., and Drysdale, J., eds. (1977) *Biological and Biomedical Applications of Isoelectric Focusing*, Plenum, New York
157. Egen, N. B., Bliss, M., Mayersohn, M., Owens, S. M., Arnold, L., and Bier, M. (1988) *Anal. Biochem.* **172**, 488–494
158. Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Görg, A., Westermeier, R., and Postel, W. (1982) *J. Biochem. Biophys. Methods* **6**, 317–339
159. Righetti, P. G., Gianazza, E., and Gelfi, C. (1988) *Trends Biochem. Sci.* **13**, 335–338
160. Righetti, P. G. (1990) *Immobilized pH Gradients: Theory and Methodology*, Elsevier, Amsterdam
161. Whitney, J. B., III, Cobb, R. R., Popp, R. A., and O'Rourke, T. W. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7646–7650
162. Madden, M. S. (1995) *Anal. Biochem.* **229**, 203–206
163. Anderson, N. (1979) *Nature (London)* **278**, 122–123
164. Wade, N. (1981) *Science* **211**, 33–35
165. Henslee, J. G., and Srere, P. A. (1979) *J. Biol. Chem.* **254**, 5488–5491
166. Hanash, S. M., and Strahler, J. R. (1989) *Nature (London)* **337**, 485–486
167. Neukirchen, R. O., Schlosshauer, B., Baars, S., Jackle, H., and Schwarz, U. (1982) *J. Biol. Chem.* **257**, 15229–15234
168. Racine, R. R., and Langley, C. H. (1980) *Nature (London)* **283**, 855–857
169. Celis, J. E., and Bravo, R., eds. (1984) *Two-Dimensional Gel Electrophoresis of Proteins: Methods and Applications*, Academic Press, New York
170. Celis, J. E., and Bravo, R. (1981) *Trends Biochem. Sci.* **6**, 197–201
171. Garrels, J. I. (1989) *J. Biol. Chem.* **264**, 5269–5282
172. Krauss, M. R., Collins, P. J., and Blose, S. H. (1989) *Nature (London)* **337**, 669–670
173. Wood, T. D., Chen, L. H., Kelleher, N. L., Little, D. P., Kenyon, G. L., and McLafferty, F. W. (1995) *Biochemistry* **34**, 16251–16254

173a. McLafferty, F. W., Fridriksson, E. K., Horn, D. M., Lewis, M. A., and Zubarev, R. A. (1999) *Science* **284**, 1289 – 1290

174. Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko.A, Boucherie, H., and Mann, M. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14440–14445
175. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) *Nature (London)* **379**, 466–469
176. Hess, D., Covey, T. C., Winz, R., Brownsey, R. W., and Aebersold, R. (1993) *Protein Sci.* **2**, 1342–1351
177. Van Holde, K. E. (1985) *Physical Biochemistry*, 2nd ed., Prentice-Hall, Engelwood Cliffs, New Jersey
178. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry*, Freeman, San Francisco, California
179. Svedburg, T., and Pederson, K. O. (1940) *The Ultracentrifuge*, Oxford Univ. Press, London
180. Schachman, H. K. (1989) *Nature (London)* **341**, 259–260
181. Schachman, H. K. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., and Horton, J. C., eds), pp. 3–15, Royal Society for Chemistry, Cambridge, UK
182. Schachman, H. K. (1959) *Ultracentrifugation in Biochemistry*, Academic Press, New York
183. Rivas, G., and Minton, A. P. (1993) *Trends Biochem. Sci.* **18**, 284–287
184. Harding, S. E., Rowe, A. J., and Horton, J. C., eds. (1992) *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, CRC Press, Boca Raton, Florida
185. Hansen, J. C., Lebowitz, J., and Demeler, B. (1994) *Biochemistry* **33**, 13155–13163
186. Ralston, G. (1993) *Introduction to Analytical Ultracentrifugation*, Beckman Instruments, Fullerton, California
187. Stafford, W. F., and Schuster, T. M. (1995) in *Introduction to Biophysical Methods for Protein and Nucleic Acid Research* (Glasel, J. A., and Deutscher, M. P., eds), pp. 111–145, Academic Press, San Diego
188. Bowen, T. J. (1970) *An Introduction to Ultracentrifugation*, Wiley (Interscience), New York
189. Bothwell, M. A., Howlett, G. J., and Schachman, H. K. (1978) *J. Biol. Chem.* **253**, 2073–2076
190. Pollet, R. J., Haase, B. A., and Standaert, M. L. (1979) *J. Biol. Chem.* **254**, 30–33

190a. *CRC Handbook of Biochemistry*, (1968) Chem. Rubber Publ. Co., Cleveland, Ohio

191. Jacobsen, M. P., Wills, P. R., and Winzor, D. J. (1996) *Biochemistry* **35**, 13173–13179

191a. Colowick, S. P., and Kaplan, N. O., eds., (1968) *Methods Enzymol.* **12B**, 388–389

192. Squire, P. G., and Himmel, M. E. (1979) *Arch. Biochem. Biophys.* **196**, 165–177
193. Andrews, P. (1965) *Biochem. J.* **96**, 595–606
194. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry*, Freeman, San Fransico, California (pp. 670–684)
195. Nakos, G., and Mortenson, L. (1971) *Biochemistry* **10**, 455–458
196. Mann, K. G., and Fish, W. W. (1972) *Methods Enzymol.* **26C**, 28–42
197. Igou, D. K., Lo, J. T., Clark, D. S., Mattice, W. L., and Younathan, E. S. (1974) *Biochem. Biophys. Res. Commun.* **60**, 140–145
198. Weber, K., Pringle, J. R., and Osborn, M. (1972) *Methods Enzymol.* **26C**, 3–29
199. Wood, H. G., Werkman, C. H., Hemingway, A., and Nier, A. O. (1941) *J. Biol. Chem.* **139**, 377–381
200. Biemann, K. (1995) *Protein Sci.* **4**, 1920–1927

200a. Niessen, W. M. A. (1998) *Liquid Chromatography - Mass Spectrometry*, 2nd ed., Dekker, New York

200b. Chapman, J. R., ed. (2000) *Mass Spectrometry of Proteins and Peptides*, Hamana Press, Totowa, New Jersey

200c. Burlingame, A. L., Carr, S. A., and Baldwin, M. A., eds. (2000) *Mass Spectrometry in Biology and Medicine*, Hamana Press, Totowa, New Jersey

201. Senko, M. W., and McLafferty, F. W. (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23**, 763–785
202. Li, Y., Hunter, R. L., and McIver, R. T., Jr. (1994) *Nature (London)* **370**, 393–395
203. Siuzdak, G. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11290–11297
204. Caprioli, R. M., and Suter, M. J.-F. (1995) in *Introduction to Biophysical Methods for Protein and Nucleic Acid Research* (Glasel, J. A., and Deutscher, M. P., eds), pp. 147-204, Academic Press, San Diego
205. Burlingame, A. L., and Carr, S. A., eds. (1996) *Mass Spectrometry in the Biological Sciences*, Humana Press, Totowa, New Jersey
206. Karas, M., and Hillenkamp, F. (1988) *Anal. Chem.* **60**, 2299–2303
207. Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., and Yoshida, T. (1988) *Rapid Commun. Mass Spectrom.* **2**, 151–156
208. Beavis, R. C., and Chait, B. T. (1989) *Rapid Commun. Mass Spectrom.* **3**, 233–237
209. Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991) *Anal. Chem.* **63**, 1193A–1198A
210. Smith, R. D., Loo, J. A., Edmonds, C. G., Barinaga, C. J., and Udseth, H. R. (1990) *Anal. Chem.* **62**, 882–899
211. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1990) *Mass Spectrom. Rev.* **9**, 37
212. Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D., and Alice, M. B. (1968) *J. Chem. Phys.* **49**, 2240
213. Mann, M., and Wilm, M. (1995) *Trends Biochem. Sci.* **20**, 219–224
214. Mann, M., Meng, C. K., and Fenn, J. B. (1989) *Anal. Chem.* **61**, 1702–1708
215. Covey, T. R., Bonner, R. F., Shushan, B. I., and Henion, J. (1988) *Rapid Commun. Mass Spectrom.* **2**, 249–256
216. Miranker, A., Kruppa, G. H., Robinson, C. V., Aplin, R. T., and Dobson, C. M. (1996) *J. Am. Chem. Soc.* **118**, 7402–7403
217. Marshall, A. G., Senko, M. W., Li, W., Li, M., Dillon, S., Guan, S., and Logan, T. M. (1997) *J. Am. Chem. Soc.* **119**, 433–434
218. Torchinsky, Y. M. (1981) *Sulfur in Proteins*, Pergamon, Oxford
219. Friedman, M. (1973) *The Chemistry and Biochemistry of the Sulfhydryl Group in Amino Acids, Peptides and Proteins*, Pergamon, Oxford
220. Jocelyn, P. C. (1972) *Biochemistry of the SH Group*, Academic Press, New York
221. Simpson, R. J., Neuberger, M. R., and Liu, T.-Y. (1976) *J. Biol. Chem.* **251**, 1936–1940
222. Hugli, T. E., and Moore, S. (1972) *J. Biol. Chem.* **247**, 2828–2834
223. Chin, C. C. Q., and Wold, F. (1974) *Anal. Biochem.* **61**, 379–391
224. Royer, G. P., Schwartz, W. E., and Liberatore, F. A. (1977) *Methods Enzymol.* **47E**, 40–45
225. Glazer, A. N. (1976) in *The Proteins*, 3rd ed., Vol. II (Neurath, H., and Hill, R. L., eds), pp. 1–103, Academic Press, New York
226. Plapp, B. V., Raftery, M. A., and Cole, R. D. (1967) *J. Biol. Chem.* **242**, 265–270
227. Doonan, S., and Fahmy, H. M. A. (1975) *Eur. J. Biochem.* **56**, 421–426
228. Mitchell, W. M. (1977) *Methods Enzymol.* **47E**, 165–169
229. Gadasi, H., Maruta, H., Collins, J. H., and Korn, E. D. (1979) *J. Biol. Chem.* **254**, 3631–3636
230. Drapeau, G. R. (1977) *Methods Enzymol.* **47E**, 189–191
231. Koida, M., and Walter, R. (1976) *J. Biol. Chem.* **251**, 7593–7599
232. Hayashi, R. (1977) *Methods Enzymol.* **47E**, 84–93
233. Krutzsch, H. C., and Pisano, J. J. (1977) *Methods Enzymol.* **47E**, 391–404
234. Mahoney, W. C., and Hermodson, M. A. (1979) *Biochemistry* **18**, 3810–3814
235. Hancock, W. S., ed. (1996) *New Methods in Peptide Mapping for the Characterization of Proteins*, CRC Press, Boca Raton, Florida
236. Lundblad, R. L. (1995) *Techniques in Protein Modification*, CRC Press, Boca Raton, Florida
237. Sanger, F. (1949) *Biochem. J.* **45**, 563–574
238. Sanger, F. (1988) *Ann. Rev. Biochem.* **57**, 1–28
239. Ruthen, R. (1993) *Sci. Am.* **269**(Oct), 30–31
240. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E., and Kent, S. B. H. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6970–6974
241. Edman, P. (1956) *Acta Chem. Scand.* **10**, 761–768
242. Edman, P., and Begg, C. (1967) *Eur. J. Biochem.* **1**, 80–91
243. Niall, H. D. (1977) *Nature (London)* **268**, 279–280
244. Schlesinger, D. H., ed. (1988) *Macromolecular Sequencing and Synthesis*, Liss, New York
245. Shively, J. E., ed. (1986) *Methods of Protein Microcharacterization*, Humana Press, Clifton, New Jersey

# References


246. Bhown, A. S. (1988) *Protein/Peptide Sequence Analysis: Current Methodoligies*, CRC Press, Boca Raton, Florida
247. Hunkapiller, M. W., and Hood, L. E. (1983) *Science* **219**, 650–659
248. Totty, N. F., Waterfield, M. D., and Hsuan, J. J. (1992) *Protein Sci.* **1**, 1215–1224
249. Gershoni, J. M. (1985) *Trends Biochem. Sci.* **10**, 103–106
250. Moos, M., Jr., Nguyen, N. Y., and Liu, T.-Y. (1988) *J. Biol. Chem.* **263**, 6005–6008
251. Baldo, B. A., and Tovey, E. R., eds. (1989) *Protein Blotting*, S. Karger, Farmington, CT
252. Wellner, D., Panneerselvam, C., and Horecker, B. L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1947–1949
253. Bailey, J. M., Nikfarjam, F., Shenoy, N. R., and Shively, J. E. (1992) *Protein Sci.* **1**, 1622–1633
254. Bailey, J. M., Shenoy, N. R., Ronk, M., and Shively, J. E. (1992) *Protein Sci.* **1**, 68–80
255. Johnson, R. S., and Walsh, K. A. (1992) *Protein Sci.* **1**, 1083–1091
255a. Keough, T., Youngquist, R. S., and Lacey, M. P. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7131–7136
256. Hunt, D. F., Shabanowitz, J., Yates, J. R., III, Zhu, N.-Z., Russell, D. H., and Castro, M. E. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 620–623
257. Takao, T., Hitouji, T., Shimonishi, Y., Tanabe, T., Inouye, S., and Inouye, M. (1984) *J. Biol. Chem.* **259**, 6105–6109
257a. Jaffe, H., Veeranna, and Pant, H. C. (1998) *Biochemistry* **37**, 16211–16224
257b. Whitelegge, J. P., Gundersen, C. B., and Faull, K. F. (1998) *Protein Sci.* **7**, 1423–1430
258. Yazdanparast, R., Andrews, P. C., Smith, D. L., and Dixon, J. E. (1987) *J. Biol. Chem.* **262**, 2507–2513
259. Brown, J. R., and Hartley, B. S. (1966) *Biochem. J.* **101**, 214–228
260. Svobodova, S., Hais, I. M., and Kostir, J. V. (1953) *Chem. Listy* **47**, 205
261. Chen, R. F., Smith, P. D., and Maly, M. (1978) *Arch. Biochem. Biophys.* **189**, 241–250
262. Lee, K. S., and Drescher, D. G. (1979) *J. Biol. Chem.* **254**, 6248–6251
263. Chen, R. F., Scott, C., and Trepman, E. (1979) *Biochim. Biophys. Acta.* **576**, 440–455
264. Lunte, S. M. *et al.* (1989) *Anal. Biochem.* **178**, 202–207
265. Sharma, Y., Rao, C. M., Rao, S. C., Krishna, A. G., Somasundaram, T., and Balasubramanian, D. (1989) *J. Biol. Chem.* **264**, 20923–20927
266. Wilson, C. M. (1983) *Methods Enzymol.* **91**, 236–247
267. Merril, C. R. (1990) *Methods Enzymol.* **182**, 477–488
268. Finn, F. M., and Hofmann, K. (1976) in *The Proteins*, 3rd ed., Vol. II (Neurath, H., and Hill, R. L., eds), pp. 105–253, Academic Press, New York
269. Burnette, W. N. (1981) *Anal. Biochem.* **112**, 195–203
270. Karey, K. P., and Sirbasku, D. A. (1989) *Anal. Biochem.* **178**, 255–259
271. Strickland, E. H., Billups, C., and Kay, E. (1972) *Biochemistry* **11**, 3657–3662
272. Metzler, D. E., Harris, C., Yang, I.-Y., Siano, D., and Thomson, J. A. (1972) *Biochem. Biophys. Res. Commun.* **46**, 1588–1597
273. Morton, R. A. (1975) *Biochemical Spectroscopy*, Wiley, New York
274. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) *Protein Sci.* **4**, 2411–2423
275. Glazer, A. N., Delange, R. J., and Sigman, D. S. (1976) *Chemical Modification of Proteins: Selected Methods and Analytical Procedures*, Elsevier, New York
276. Feeney, R. E., Blankenhorn, G., and Dixon, H. B. F. (1975) *Advances in Protein Chemistry*, Vol. 29, Academic Press, New York (pp. 135–203)
277. Acharya, A. S., and Manjula, B. N. (1987) *Biochemistry* **26**, 3524–3530
278. Rypniewski, W. R., Holden, H. M., and Rayment, I. (1993) *Biochemistry* **32**, 9851–9858
279. Jentoft, N., and Dearborn, D. G. (1979) *J. Biol. Chem.* **254**, 4359–4365
280. Hennecke, M., and Plapp, B. N. (1984) *Anal. Biochem.* **136**, 110–118
281. Brocklehurst, K. (1979) *Int. J. Biochem* **10**, 259–274
282. Degani, Y., and Patchornik, A. (1974) *Biochemistry* **13**, 1–11
283. Riddles, P. W., Blakeley, R. L., and Zerner, B. (1983) *Methods Enzymol.* **91**, 49–60
284. Pande, C. S., Pelzig, M., and Glass, J. D. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 895–899
285. Fraker, P. J., and Speck, J. C., Jr. (1978) *Biochem. Biophys. Res. Commun.* **80**, 849–857
286. Tolan, D. R., Lambert, J. M., Boileau, G., Fanning, T. G., Kenny, J. W., Vassos, A., and Traut, R. R. (1980) *Anal. Biochem.* **103**, 101–109
287. Miles, E. W. (1977) *Methods Enzymol.* **47**, 431–442
288. Kawata, Y., Sakiyama, F., Hayashi, F., and Kyogoku, Y. (1990) *Eur. J. Biochem.* **187**, 255–262
289. Secundo, F., Carrea, G., D'Arrigo, P., and Servi, S. (1996) *Biochemistry* **35**, 9631–9636
290. Miura, S., Tomita, S., and Ichikawa, Y. (1991) *J. Biol. Chem.* **266**, 19212–19216
291. Colman, R. F. (1983) *Ann. Rev. Biochem.* **52**, 67–91
292. Chowdhry, V., and Westheimer, F. H. (1979) *Ann. Rev. Biochem.* **48**, 293 – 325
293. Jelenc, P. C., Cantor, C. R., and Simon, S. R. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3564–3568
294. Simmonds, R. J. (1992) *Chemistry of Biomolecules*, CRC Press, Boca Raton, Florida
295. Bodanszky, M., Bodanszky, A., and Trost, B. M., eds. (1994) *The Practice of Peptide Synthesis*, 2nd ed., Springer-Verlag, New York
296. Lloyd-Williams, P., Albericio, F., and Giralt, E. (1997) *Chemical Approaches to the Synthesis of Peptides and Proteins*, CRC Press, Boca Raton, Florida
297. Kent, S. B. H. (1988) *Ann. Rev. Biochem.* **57**, 957–989
298. Kaiser, E. T., Mihara, H., Laforet, G. A., Kelly, J. W., Walters, L., Findeis, M. A., and Sasaki, T. (1989) *Science* **243**, 187–192
299. Merrifield, B. (1986) *Science* **232**, 341–347
300. Atherton, E., Logan, C. J., and Sheppard, R. C. (1981) *J. Chem. Soc., Perkin, I.*, 538–546
301. Atherton, E., and Sheppard, R. C. (1985) *J. Chem. Soc. Chem. Commun.*, 165–166
301a. Miller, S. C., and Scanlan, T. S. (1998) *J. Am. Chem. Soc.* **120**, 2690–2691
302. You-shang, Z. (1983) *Trends Biochem. Sci.* **8**, 16–17
303. Kullman, W. (1987) *Enzymatic Peptide Synthesis*, CRC Press, Boca Raon, Florida
304. Liu, C.-F., and Tam, J. P. (1994) *J. Am. Chem. Soc.* **116**, 4149–4153
305. Tam, J. P., Lu, Y.-A., Liu, C.-F., and Shao, J. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 12485–12489
306. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989) *Science* **244**, 182–188
307. Borman, S. (1996) *Chem. Eng. News* **74** (Feb 12), 29–35
308. Youngquist, R. S., Fuentes, G. R., Lacey, M. P., and Keough, T. (1995) *J. Am. Chem. Soc.* **117**, 3900–3906
309. Cortese, R., ed. (1996) *Combinatorial Libraries: Synthesis, Screening and Application Potential*, Walter de Gruyter & Co., New York
310. Ellman, J., Stoddard, B., and Wells, J. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2779–2782
311. Kishchenko, G., Batliwala, H., and Makowski, L. (1994) *J. Mol. Biol.* **241**, 208–213
312. Scott, J. K. (1992) *Trends Biochem. Sci.* **17**, 241–245
312a. Barbas, C. F., Burton, D. B., Scott, J. K., and Silverman, G. J. (2000) *Phage Display: A Laboratory Manual* Cold Spring Harbor Press, Cold Spring Harbor, New York
313. Marvin, D. A., (1998) *Current Opinions in Structural Biology* **8**, 150–158
313a. Barbas, C. F., Burton, D. B., Scott, J. K., and Silverman, G. J. (2000) *Phage Display: A Laboratory Manual* Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
314. McConnell, S. J., and Hoess, R. H. (1995) *J. Mol. Biol.* **250**, 460–470
315. Katz, B. A. (1995) *Biochemistry* **34**, 15421–15429
316. Allen, J. B., Walberg, M. W., Edwards, M. C., and Elledge, S. J. (1995) *Trends Biochem. Sci.* **20**, 511–516
317. Gates, C. M., Stemmer, W. P. C., Kaptein, R., and Schatz, P. J. (1996) *J. Mol. Biol.* **255**, 373–386
318. Bastos, M., Maeji, N. J., and Abeles, R. H. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6738–6742
319. Han, H., Wolfe, M. M., Brenner, S., and Janda, K. D. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6419–6423
320. Chu, Y.-H., Dunayevskiy, Y. M., Kirby, D. P., Vouros, P., and Karger, B. L. (1996) *J. Am. Chem. Soc.* **118**, 7827–7835
321. Déprez, B., Williard, X., Bourel, L., Coste, H., Hyafil, F., and Tartar, A. (1995) *J. Am. Chem. Soc.* **117**, 5405–5406
322. Kim, R. M., Manna, M., Hutchins, S. M., Griffin, P. R., Yates, N. A., Bernick, A. M., and Chapman, K. T. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10012–10017
323. Sundberg, S. A., Barrett, R. W., Pirrung, M., Lu, A. L., Kiangsoontra, B., and Holmes, C. P. (1995) *J. Am. Chem. Soc.* **117**, 12050–12057
324. Baum, R. M. (1991) *Chem. Eng. News* **69**, 21–22
325. Slayter, E. M., and Slayter, H. S. (1992) *Light and Electron Microscopy*, Cambridge Univ. Press, New York
326. Bracegirdle, B. (1989) *Trends Biochem. Sci.* **14**, 464–468
327. Matthews, C. K., and van Holde, K. E. (1996) *Biochemistry*, 2nd ed., Benjamin, Menlo Park, California
328. Lichtman, J. W. (1994) *Sci. Am.* **271**(Aug), 40–45
329. Shotton, D., and White, N. (1989) *Trends Biochem. Sci.* **14**, 435–439
330. Boyde, A. (1985) *Science* **230**, 1270–1276
331. Gura, T. (1997) *Science* **276**, 1988–1990
332. Bennett, B. D., Jetton, T. L., Ying, G., Magnuson, M. A., and Piston, D. W. (1996) *J. Biol. Chem.* **271**, 3647–3651
333. Pennisi, E. (1997) *Science* **275**, 480–481
334. Maiti, S., Shear, J. B., Williams, R. M., Zipfel, W. R., and Webb, W. W. (1997) *Science* **275**, 530–532
335. Pool, R. (1988) *Science* **241**, 25–26
336. Betzig, E., and Chichester, R. J. (1993) *Science* **262**, 1422–1425
337. Dykstra, M. J. (1992) *Biological Electron Microscopy*, Plenum, New York
338. Hand, A. R. (1995) in *Introduction to Biophysical Methods for Protein and Nucleic Acid Research* (Glasel, J. A., and Deutscher, M. P., eds), pp. 205-260, Academic Press, San Diego
339. Nellist, P. D., McCallum, B. C., and Rodenburg, J. M. (1995) *Nature (London)* **374**, 630–632

## References


340. Schliwa, M. (1997) *Nature (London)* **387**, 764
341. Bubel, A., ed. (1989) *Microstructure and Function of Cells*, 1st ed., Ellis Harwood, New York
342. Penman, S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5251–5257
343. Morel, G., ed. (1995) *Visualization of Nucleic Acids*, CRC Press, Boca Raton, Florida
344. Caspar, D. L. D., and DeRosier, D. J. (1982) *Science* **218**, 653–655
345. Frank, J., ed. (1992) *Electron Tomography: Three-Dimensional Imaging with the Transmission Electron Microscope*, Plenum, New York
346. Skoglund, U., and Daneholt, B. (1986) *Trends Biochem. Sci.* **11**, 499–503
347. Lambert, O., Boisset, N., Taveau, J.-C., and Lamy, J. N. (1994) *J. Mol. Biol.* **244**, 640–647
348. Taveau, J.-C., Boisset, N., Lamy, J., Lambert, O., and Lamy, J. N. (1997) *J. Mol. Biol.* **266**, 1002–1015
349. De Rosier, D. J. (1997) *Nature (London)* **386**, 26–27
350. Dorset, D. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1791–1794
351. Binnig, G., and Rohrer, H. (1987) *Rev. Mod. Phys.* **59**, 615–625
352. Binnig, G., and Rohrer, H. (1985) *Sci. Am.* **253**, 50–56
353. Wiesendanger, R. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12749–12750
354. Wickramasinghe, H. K. (1989) *Sci. Am.* **261**(Oct), 98–105
355. Hansma, P. K., Elings, V. B., Marti, O., and Bracker, C. E. (1988) *Science* **242**, 209–216
356. Pool, R. (1990) *Science* **247**, 634–636
357. Edstrom, R. D., Yang, X., Lee, G., and Evans, D. F. (1990) *FASEB J.* **4**, 3144–3151
358. Engel, A. (1991) *Ann. Rev. Biophys. Biophys. Chem.* **20**, 79–108
359. Amato, I. (1997) *Science* **276**, 1982–1985
360. Prater, C. B., Butt, H. J., and Hansma, P. K. (1990) *Nature (London)* **345**, 839–840
361. Hansma, H. G., and Hoh, J. H. (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23**, 115–139
362. Radmacher, M., Tillmann, R. W., Fritz, M., and Gaub, H. E. (1992) *Science* **257**, 1900–1905
363. Karrasch, S., Hegerl, R., Hoh, J. H., Baumeister, W., and Engel, A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 836–838
364. Mou, J., Yang, J., and Shao, Z. (1995) *J. Mol. Biol.* **248**, 507–512
365. Yip, C. M., and Ward, M. D. (1996) *Biophys. J.* **71**, 1071–1078
366. Mou, J., Sheng, S., Ho, R., and Shao, Z. (1996) *Biophys. J.* **71**, 2213–2221
367. Kuznetsov, YuG., Malkin, A. J., Land, T. A., DeYoreo, J. J., Barba, A. P., Konnert, J., and McPherson, A. (1997) *Biophys. J.* **72**, 2357–2364
367a. Czajkowsky, D. M., and Shao, Z. (1998) *FEBS Letters* **430**, 51–54
368. Noy, A., Frisbie, C. D., Rozsnyai, L. F., Wrighton, M. S., and Lieber, C. M. (1995) *J. Am. Chem. Soc.* **117**, 7943–7951
369. Boland, T., and Ratner, B. D. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5297–5301
370. Miles, M. (1997) *Science* **277**, 1845–1847
371. Hansma, H. G., Laney, D. E., Bezanilla, M., Sinsheimer, R. L., and Hansma, P. K. (1995) *Biophys. J.* **68**, 1672–1677
371a. Fisher, T. E., Oberhauser, A. F., Carrion-Vazquez, M., Marszalek, P. E., and Fernandez, J. M. (1999) *Trends Biochem. Sci.* **24**, 379–384
372. Blundell, T. L., and Johnson, L. N. (1976) *Protein Crystallography*, Academic Press, New York
373. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry*, Freeman, San Francisco, California (pp. 687–811)
374. McPherson, A. (1982) *Preparation and Analysis of Protein Crystals*, Wiley, New York
375. Wyckoff, H. W., Hirs, C. H. W., and Timasheff, S. N., eds. (1985) *Methods in Enzymology*, Vol. 114,115, Academic Press, Orlando, Florida
376. Allewell, N. M., and Trikha, J. (1995) in *Introduction to Biophysical Methods for Protein and Nucleic Acid Research* (Glasel, J. A., and Deutscher, M. P., eds), pp. 381–431, Academic Press, San Diego
377. Verlinde, C. L. M. J., Merritt, E. A., Van Den Akker, F., Kim, H., Feil, I., Delboni, L. F., Mande, S. C., Sarfaty, S., Petra, P. H., and Hol, W. G. J. (1994) *Protein Sci.* **3**, 1670–1686
378. Woolfson, N. M. (1997) *An Introduction to X-ray Crystallography*, 2nd ed., Cambridge Univ. Press, New York
379. Thomas, J. M. (1993) *Nature (London)* **364**, 478–482
380. Dounce, A. L., and Allen, P. Z. (1988) *Trends Biochem. Sci.* **13**, 317–320
381. Bernal, J. D., and Crowfoot, D. (1934) *Nature (London)* **133**, 794–795
382. Glusker, J. P. (1994) *Protein Sci.* **3**, 2465-2469
383. Perutz, M. F., Muirhead, H., Cox, J. M., Goaman, L. G. G., Mathews, F. S., McGandy, E. L., and Webb, L. E. (1968) *Nature (London)* **219**, 29–32
384. Perutz, M. (1992) *Protein Structure: New Approaches to Disease and Therapy*, W.H. Freeman, San Fransico and New York
385. Eisenberg, D. (1994) *Protein Sci.* **3**, 1625–1628
386. Perutz, M. F. (1978) *Sci. Am.* **239**(Dec), 92–125
387. Tegze, M., and Faigel, G. (1996) *Nature (London)* **380**, 49–51
388. Fadley, C. S., and Len, P. M. (1996) *Nature (London)* **380**, 27–28
389. Rhodes, G. (1993) *Crystallography Made Crystal Clear*, Academic Press, New York
389a. Glusker, J. P., and Trueblood, K. N. (1985) *Crystal Structure Analysis A Primer*, 2nd ed., Oxford Univ. Press, Oxford
389b. Drenth, J. (1994) *Principles of Protein X-ray Crystallography*, Springler-Verlag, New York
389c. McRee, D. E. (1999) *Practical Protein Crystallography*, Academic Press, San Diego, California
390. Sweet, R. M. (1985) *Methods Enzymol.* **114**, 19–46
391. Kendrew, J. C. (1963) *Science* **139**, 1259–1266
392. Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C. (1960) *Nature (London)* **185**, 422–427
393. Rossman, M. G. (1994) *Protein Sci.* **3**, 1731–1733
394. Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., and Sarma, V. R. (1965) *Nature (London)* **206**, 757–761
395. Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L., and Thornton, J. M. (1997) *Trends Biochem. Sci.* **22**, 488–490
396. Gerstein, M., and Levitt, M. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11911–11916
397. Marvin, D. A., and Nave, C. (1982) in *Structural Molecular Biology* (Davies, D. B., Saenger, W., and Danyluk, S. S., eds), pp. 3–44, Plenum, New York
398. Ducruix, A., and Giegé, R., eds. (1999) *Crystallization of Nucleic Acids and Proteins*, 2nd ed., IRL Press, Oxford
399. McPherson, A., Koszelak, S., Axelrod, H., Day, J., Williams, R., Robinson, L., McGrath, M., and Cascio, D. (1986) *J. Biol. Chem.* **261**, 1969–1975
400. McPherson, A., Jr. (1976) *J. Biol. Chem.* **251**, 6300–6303
401. Ray, W. J., Jr., and Puvathingal, J. M. (1986) *J. Biol. Chem.* **261**, 11544–11549
402. Day, J., and McPherson, A. (1992) *Protein Sci.* **1**, 1254–1268
403. Koszelak, S., Day, J., Leja, C., Cudney, R., and McPherson, A. (1995) *Biophys. J.* **69**, 13–19
404. Service, R. F. (1999) *Science* **285**, 1342–1346
405. Knäblein, J., Neuefeind, T., Schneider, F., Bergner, A., Messerschmidt, A., Löwe, J., Steipe, B., and Huber, R. (1997) *J. Mol. Biol.* **270**, 1–7
406. Glusker, J. P. (1984) *Trends Biochem. Sci.* **9**, 328–330
407. Guss, J. M., Merritt, E. A., Phizackerley, R. P., Hedman, B., Murata, M., Hodgson, K. O., and Freeman, H. C. (1988) *Science* **241**, 806–811
408. Walter, R. L., Ealick, S. E., Friedman, A. M., Blake, R. C., II, Proctor, P., and Shoham, M. (1996) *J. Mol. Biol.* **263**, 730–751
409. Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A., Prade, L., Neuefeind, T., Moroder, L., and Huber, R. (1997) *J. Mol. Biol.* **270**, 616–623
410. Longhi, S., Czjzek, M., Lamzin, V., Nicolas, A., and Cambillau, C. (1997) *J. Mol. Biol.* **268**, 779–799
410a. Yamano, A., Heo, N.-H., and Teeter, M. M. (1997) *J. Biol. Chem.* **272**, 9597–9600
410b. Lamzin, V. S., Morris, R. J., Dauter, Z., Wilson, K. S., and Teeter, M. M. (1999) *J. Biol. Chem.* **274**, 20753–20755
411. Rhee, S., Silva, M. M., Hyde, C. C., Rogers, P. H., Metzler, C. M., Metzler, D. E., and Arnone, A. (1997) *J. Biol. Chem.* **272**, 17293–17302
412. Frauenfelder, H., Petsko, G. A., and Tsernoglou, D. (1979) *Nature (London)* **280**, 558–563
413. Arnone, A., Rogers, P. H., Hyde, C. C., Briley, P. D., Metzler, C. M., and Metzler, D. E. (1985) in *Transaminases* (Christen, P., and Metzler, D. E., eds), pp. 138–155, Wiley, New York
414. Wlodawer, A. (1982) *Prog. Biophys. and Mol. Biol.* **40**, 115–159
415. Kossiakoff, A. A. (1985) *Ann. Rev. Biochem.* **54**, 1195–1227
416. Butler, D. (1994) *Nature (London)* **371**, 469
417. Hellemans, A. (1997) *Science* **277**, 1214–1215
418. Glanz, J. (1995) *Science* **267**, 1904–1906
419. Pool, R. (1988) *Science* **241**, 295
420. Eisenberger, P., and Suckewer, S. (1996) *Science* **274**, 201–202
421. Neutze, R., and Hajdu, J. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5651–5655
422. Hajdu, J., and Johnson, L. N. (1990) *Biochemistry* **29**, 1669–1678
423. Duke, E. M. H., Wakatsuki, S., Hadfield, A., and Johnson, L. N. (1994) *Protein Sci.* **3**, 1178–1196
424. Jacoby, M. (1997) *Chem. Eng. News* **Dec 8**, 5
425. Rischel, C., Rousse, A., Uschmann, I., Albouy, P.-A., Geindre, J.-P., Audebert, P., Gauthier, J.-C., Förster, E., Martin, J.-L., and Antonetti, A. (1997) *Nature (London)* **390**, 490–492
426. Rattle, H. (1995) *An NMR Primer for Life Scientists*, Partnership Press, Tuckerton, NJ
427. Roberts, G. C. K., ed. (1993) *NMR of Macromolecules. A Practical Approach*, IRL Press, Oxford
428. Evans, J. N. S. (1995) *Biomolecular NMR Spectroscopy*, Oxford University Press, New York
429. Olson, D. L., Peck, T. L., Webb, A. G., Magin, R. L., and Sweedler, J. V. (1995) *Science* **270**, 1967–1969
430. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York
431. Nageswara Rao, B. D., and Kemple, M. D., eds. (1996) *NMR as a structural tool for macromolecules: Current status and future directions*, Plenum, New York
432. Bovey, F. A. (1969) *Nuclear Magnetic Resonance Spectroscopy*, Academic Press, New York

# References


433. Sanders, J. K. M., and Hunter, B. K. (1993) *Modern NMR Spectroscopy*, 2nd ed., Oxford Univ. Press, London
434. Dybowski, C., and Lichter, R. L., eds. (1987) *NMR Spectroscopy Techniques*, Vol. 5, Dekker, New York
435. Bangerter, B. W. (1995) in *Introduction to Biophysical Methods for Protein and Nucleic Acid Research* (Glasel, J. A., and Deutscher, M. P., eds), pp. 317-379, Academic Press, San Diego
436. Freeman, R. (1997) *A Handbook of Nuclear Magnetic Resonance*, 2nd ed., Addison Wesley Longman Limited, Essex, England
437. Emsley, J. W., ed. (1996) *Encyclopedia of NMR*, John Wiley & Sons, Ltd., Sussex, England
438. Markley, J. L. (1975) *Acc. Chem. Res.* **8**, 70–80
439. Perkins, S. J. (1982) in *Biological Magnetic Resonance*, Vol. 4 (Berliner, L. J., and Reuben, J., eds), pp. 193-334, Plenum Press, New York and London
440. Osapay, K., and Case, D. A. (1991) *J. Am. Chem. Soc.* **113**, 9436–9444
441. Szilágyi, L. (1995) *Prog. Nucl. Magn. Reson. Spectrosc.* **27**, Part 4, 326–443
442. Williamson, M. P., Kikuchi, J., and Asakura, T. (1995) *J. Mol. Biol.* **247**, 541–546
443. Sitkoff, D., and Case, D. A. (1997) *J. Am. Chem. Soc.* **119**, 12262–12273
444. Gu, Z., Zambrano, R., and McDermott, A. (1994) *J. Am. Chem. Soc.* **116**, 6368–6372
445. Gu, Z., Ridenour, C. F., Bronnimann, C. E., Iwashita, T., and McDermott, A. (1996) *J. Am. Chem. Soc.* **118**, 822–829
446. de Dios, A. C., Pearson, J. G., and Oldfield, E. (1993) *Science* **260**, 1491–1496
447. Sulzbach, H. M., Schleyer, P. V. R., and Schaefer, H. F., III (1994) *J. Am. Chem. Soc.* **116**, 3967–3972
448. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) *Biochemistry* **31**, 1647–1651
449. Linderstrøm-Lang, K. (1955) *Chemical Society Special Publication*, No. 2, 1–20
450. Barfield, M., and Karplus, M. (1969) *J. Am. Chem. Soc.* **91**, 1–16
451. Karplus, M. (1963) *J. Am. Chem. Soc.* **85**, 2870–2871
452. Bax, A. (1989) *Ann. Rev. Biochem.* **58**, 223–256
453. Wang, A. C., and Bax, A. (1996) *J. Am. Chem. Soc.* **118**, 2483–2494
454. Stonehouse, J., Adell, P., Keeler, J., and Shaka, A. J. (1994) *J. Am. Chem. Soc.* **116**, 6037–6038
455. Aguayo, J. B., Gamcsik, M. P., and Dick, J. D. (1988) *J. Biol. Chem.* **263**, 19552–19557
456. Ackerman, J. J. H., Ewy, C. S., Kim, S.-G., and Shalwitz, R. A. (1987) *Ann. N.Y. Acad. Sci.* **508**, 89–98
457. Breitmaier, E., and Voelter, W. (1987) *Carbon-13 NMR Spectroscopy*, 3rd ed., VCH Publishers, New York
458. Beckmann, N. (1995) *Carbon-13 NMR Spectroscopy of Biological Systems*, Academic Press, New York
459. Gronenborn, A. M., Clore, G. M., Schmeissner, U., and Wingfield, P. T. (1986) *Eur. J. Biochem.* **161**, 37–43
460. Clore, G. M., Bax, A., Driscoll, P. C., Wingfield, P. T., and Gronenborn, A. M. (1990) *Biochemistry* **29**, 8172–8184
461. Jue, T., Rothman, D. L., Shulman, G. I., Tavitian, B. A., DeFronzo, R. A., and Shulman, R. G. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4489–4491
462. Markley, J. L. (1989) *Methods Enzymol.* **176**, 12–64
463. Pelton, J. G., Torchia, D. A., Meadow, N. D., and Roseman, S. (1993) *Protein Sci.* **2**, 543–558
464. Rajagopal, P., Waygood, E. B., and Klevit, R. E. (1994) *Biochemistry* **33**, 15271–15282
465. Solum, M. S., Altmann, K. L., Strohmeier, M., Berges, D. A., Zhang, Y., Facelli, J. C., Pugmire, R. J., and Grant, D. M. (1997) *J. Am. Chem. Soc.* **119**, 9804–9809
466. Verkade, J. G., and Quin, L. D., eds. (1987) *Phosphorus-31 NMR Spectroscopy in Stereochemical Analysis*, VCH Publishers, Deerfield Beach, Florida
467. Ugurbil, K., Kingsley-Hickman, P. B., Sako, E. Y., Zimmer, S., Mohanakrishnan, P., Robitaille, P. M. L., Thoma, W. J., Johnson, A., Foker, J. E., From, A. H. L. (1987) *Ann. N.Y. Acad. Sci.* **508**, 265–286
468. Feeney, J., McCormick, J. E., Bauer, C. J., Birdsall, B., Moody, C. M., Starkmann, B. A., Young, D. W., Francis, P., Havlin, R. H., Arnold, W. D., and Oldfield, E. (1996) *J. Am. Chem. Soc.* **118**, 8700–8706
469. Lau, E. Y., and Gerig, J. T. (1997) *Biophys. J.* **73**, 1579–1592
470. Jarema, M. A. C., Lu, P., and Miller, J. H. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2707–2711
471. Slebe, J. C., and Martinez-Carrion, M. (1990) *J. Biol. Chem.* **253**, 2093–2097
472. Briley, P. A., Eisenthal, R., Harrison, R., and Smith, G. D. (1990) *Biochem. J.* **163**, 325–331
473. Roberts, J. K. M., Wade-Jardetzky, N., and Jardetzky, O. (1981) *Biochemistry* **20**, 5389–5394
474. Deutsch, C. J., and Taylor, J. S. (1987) *Ann. N.Y. Acad. Sci.* **508**, 33
475. Gutowsky, H. S., McCall, D. W., McGarvey, B. R., and Meyer, L. H. (1952) *J. Am. Chem. Soc.* **74**, 4809–4817
476. Chang, Y. C., and Graves, D. J. (1985) *J. Biol. Chem.* **260**, 2709–2714
477. Scott, R. D., Chang, Y.-C., Graves, D. J., and Metzler, D. E. (1985) *Biochemistry* **24**, 7668–7681
478. Saunders, M., Jiménez-Vázquez, H. A., Cross, R. J., Mroczkowski, S., Freedberg, D. I., and Anet, F. A. L. (1994) *Nature (London)* **367**, 256–257
479. Adebodun, F., and Jordan, F. (1988) *J. Am. Chem. Soc.* **110**, 309–310
480. Springer, C. S., Jr. (1987) *Ann. Rev. Biophys. Biophys. Chem.* **16**, 375–399
481. Wittenberg, B. A., and Gupta, R. K. (1985) *J. Biol. Chem.* **260**, 2031–2034
482. Jelicks, L. A., and Gupta, R. K. (1989) *J. Biol. Chem.* **264**, 15230–15235
483. Price, W. S., Ge, N.-H., Hong, L.-Z., and Hwang, L.-P. (1993) *J. Am. Chem. Soc.* **115**, 1095–1105
484. Wang, S. M., and Gilpin, R. K. (1983) *Anal. Chem.* **55**, 493–497
485. Markham, G. D. (1986) *J. Biol. Chem.* **261**, 1507–1509
486. Denisov, V. P., and Halle, B. (1995) *J. Mol. Biol.* **245**, 682–687
487. Mullen, G. P., Dunlap, R. B., and Odom, J. D. (1986) *Biochemistry* **25**, 5625–5632
488. Aldhous, P. (1991) *Nature (London)* **353**, 689
489. Becker, E. D., and Farrar, T. C. (1972) *Science* **178**, 361–368
490. Bax, A., and Lerner, L. (1986) *Science* **232**, 960–967
491. Yamazaki, T., Muhandiram, R., and Kay, L. E. (1994) *J. Am. Chem. Soc.* **116**, 8266–8278
492. Farrar, T. C., and Becker, E. D. (1971) *Pulse and Fourier Transform NMR*, Academic Press, New York
493. Wüthrich, K. (1989) *Science* **243**, 45–50
494. van de Ven, F. J. M., and Hilbers, C. W. (1986) *J. Mol. Biol.* **192**, p. 419-441
495. Baniak, E. L., Rivier, J. E., Struthers, R. S., Hagler, A. T., and Gierasch, L. M. (1987) *Biochemistry* **26**, 2642–2656
496. Wüthrich, K. (1990) *J. Biol. Chem.* **265**, 22059–22062
497. Bax, A., and Grzesiek, S. (1993) in *NMR of Proteins* (Clore, G. M., and Gronenborn, A. M., eds), pp. 33–52, CRC Press, Boca Raton, Florida
498. Clore, G. M., and Gronenborn, A. M. (1994) *Protein Sci.* **3**, 372–390
499. Hoitnik, C. W. G., Driscoll, P. C., Hill, H. A. O., and Canters, G. W. (1994) *Biochemistry* **33**, 3560–3571
500. Yamazaki, T., Tochio, H., Furui, J., Aimoto, S., and Kyogoku, Y. (1997) *J. Am. Chem. Soc.* **119**, 872–880
501. Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., and Kay, L. E. (1994) *J. Am. Chem. Soc.* **116**, 11655–11666
502. Hu, J.-S., and Bax, A. (1997) *J. Am. Chem. Soc.* **119**, 6360–6368
503. Song, S., Velde, D. V., Gunn, C. W., and Himes, R. H. (1994) *Biochemistry* **33**, 693–698
504. Barsukov, I. L., Lian, L.-Y., Ellis, J., Sze, K.-H., Shaw, W. V., and Roberts, G. C. K. (1996) *J. Mol. Biol.* **262**, 543–558
505. Wu, W.-J., Anderson, V. E., Raleigh, D. P., and Tonge, P. J. (1997) *Biochemistry* **36**, 2211–2220
506. Metzler, W. J., Hare, D. R., and Pardi, A. (1989) *Biochemistry* **28**, 7045–7052
507. McDowell, L. M., Lee, M., McKay, R. A., Anderson, K. S., and Schaefer, J. (1996) *Biochemistry* **35**, 3328–3334
508. Pflugrath, J. W., Wiegand, G., Huber, R., and Vértesy, L. (1986) *J. Mol. Biol.* **189**, 383–386
509. Kline, A. D., Braun, W., and Wüthrich, K. (1986) *J. Mol. Biol.* **189**, 377–382
510. McDonald, C. C., and Phillips, W. D. (1969) *J. Am. Chem. Soc.* **91**, 1513–1521
511. Nietlispach, D., Clowes, R. T., Broadhurst, R. W., Ito, Y., Keeler, J., Kelly, M., Ashurst, J., Oschkinat, H., Domaille, P. J., and Laue, E. D. (1996) *J. Am. Chem. Soc.* **118**, 407–415
512. Rosen, M. K., Gardner, K. H., Willis, R. C., Parris, W. E., Pawson, T., and Kay, L. E. (1996) *J. Mol. Biol.* **263**, 627–636
513. Venters, R. A., Farmer, B. T., II, Fierke, C. A., and Spicer, L. D. (1996) *J. Mol. Biol.* **264**, 1101–1116
514. LeMaster, D. M., and Richards, F. M. (1988) *Biochemistry* **27**, 142–150
515. Bax, A., Kay, L. E., Sparks, S. W., and Torchia, D. A. (1989) *J. Am. Chem. Soc.* **111**, 408–409
516. Torchia, D. A., Sparks, S. W., and Bax, A. (1989) *Biochemistry* **28**, 5509–5524
517. Lowry, D. F., Redfield, A. G., McIntosh, L. P., and Dahlquist, F. W. (1988) *J. Am. Chem. Soc.* **110**, 6885–6886
518. Stockman, B. J., Reily, M. D., Westler, W. M., Ulrich, E. L., and Markley, J. L. (1989) *Biochemistry* **28**, 230–236
519. Qian, H., Mayo, S. L., and Morton, A. (1994) *Biochemistry* **33**, 8167–8171
519a. Riek, R., Pervushin, K., and Wüthrich, K. (2000) *Trends Biochem. Sci.* **25**, 462–468
520. Günter, P., Braun, W., Billeter, M., and Wüthrich, K. (1989) *J. Am. Chem. Soc.* **111**, 3997–4004
521. Oh, B.-H., Westler, W. M., and Markley, J. L. (1989) *J. Am. Chem. Soc.* **111**, 3083–3085
522. Cross, T. A., and Opella, S. J. (1983) *J. Am. Chem. Soc.* **105**, 306–308
523. Rothgeb, T. M., and Oldfield, E. (1981) *J. Biol. Chem.* **256**, 1432–1446
524. Buck, M., Schwalbe, H., and Dobson, C. M. (1995) *Biochemistry* **34**, 13219–13232
525. Nicholson, L. K., Kay, L. E., Baldisseri, D. M., Arango, J., Young, P. E., Bax, A., and Torchia, D. A. (1992) *Biochemistry* **31**, 5253–5263
526. Gardner, K. H., and Kay, L. E. (1997) *J. Am. Chem. Soc.* **119**, 7599–7600
527. Lenkinski, R. E., Dallas, J. L., and Glickson, J. D. (1979) *J. Am. Chem. Soc.* **101**, 3071–3077

## References


528. Carver, J. A., and Bradbury, J. H. (1984) *Biochemistry* **23**, 4890–4905
529. Bothelho, L. H., and Gurd, F. R. N. (1978) *Biochemistry* **17**, 5188–5196
530. Bothelho, L. H., Friend, S. H., Matthew, J. B., Lehman, L. D., Hanania, G. I. H., and Gurd, F. R. N. (1978) *Biochemistry* **17**, 5197–5205
531. Morino, Y., Yamasaki, M., Tanase, S., Nagashima, F., Akasaka, K., Imoto, T., and Miyazawa, T. (1984) *J. Biol. Chem.* **259**, 3877–3882
532. Cocco, M. J., Kao, Y.-H., Phillips, A. T., and Lecomte, J. T. J. (1992) *Biochemistry* **31**, 6481–6491
533. Bachovin, W. W., Wong, W. Y. L., Farr-Jones, S., Shenvi, A. B., and Kettner, C. A. (1988) *Biochemistry* **27**, 7689–7697
534. Mollova, E. T., Metzler, D. E., Kintanar, A., Kagamiyama, H., Hayashi, H., Hirotsu, K., and Miyahara, I. (1997) *Biochemistry* **36**, 615–625
535. Guéron, M., Plateau, P., and Decorps, M. (1991) *Progr. in NMR Spectr.* **23**, 135–209
536. Plateau, P., and Guéron, M. (1982) *J. Am. Chem. Soc.* **104**, 7310–7311
537. Sklenar, V., and Bax, A. (1987) *J. of Magnetic Resonance* **74**, 469–479
538. Bullitt, E., and Makowski, L. (1995) *Nature (London)* **373**, 164–167
539. Griffin, J. H., Cohen, J. S., and Schechter, A. N. (1973) *Biochemistry* **12**, 2096–2099
540. Tong, H., and Davis, L. (1995) *Biochemistry* **34**, 3362–3367
541. Gunnarsson, G., Wennerström, H., Egan, W., and Forsén, S. (1976) *Chem. Phys. Lett.* **38**, 96–99
542. McDermott, A., and Ridenour, C. F. (1996) in *Encyclopedia of NMR* (Emsley, J. W., ed), pp. 3820–3824, J. Wiley & Sons, Ltd., Sussex, UK
543. Zhao, Q., Abeygunawardana, C., Gittis, A. G., and Mildvan, A. S. (1997) *Biochemistry* **36**, 14616–14626
544. Perrin, C. L., and Nielson, J. B. (1997) *J. Am. Chem. Soc.* **119**, 12734–12741
545. Tjandra, N., and Bax, A. (1997) *J. Am. Chem. Soc.* **119**, 8076–8082
546. Kintanar, A., Metzler, C. M., Metzler, D. E., and Scott, R. D. (1991) *J. Biol. Chem.* **266**, 17222–17229
547. Metzler, C. M., Metzler, D. E., Kintanar, A., Scott, R. D., and Marceau, M. (1991) *Biochem. Biophys. Res. Commun.* **178**, 385–392
548. Firsov, L. M., Neustroev, K. N., Aleshin, A. E., Metzler, C. M., Metzler, D. E., Scott, R. D., Stoffer, B., Christensen, T., and Svensson, B. (1994) *Eur. J. Biochem.* **223**, 293–302
549. Metzler, D. E., Metzler, C. M., Mollova, E. T., Scott, R. D., Tanase, S., Kogo, K., Higaki, T., and Morino, Y. (1994) *J. Biol. Chem.* **269**, 28017–28026
550. Metzler, D. E., Metzler, C. M., Scott, R. D., Mollova, E. T., Kagamiyama, H., Yano, T., Kuramitsu, S., Hayashi, H., Hirotsu, K., and Miyahara, I. (1994) *J. Biol. Chem.* **269**, 28027–28033
551. Metzler, D. E. (1997) *Methods Enzymol.* **280**, 30–40
552. Englander, S. W., Mayne, L., Bai, Y., and Sosnick, T. R. (1997) *Protein Sci.* **6**, 1101–1109
553. Kossiakoff, A. A. (1982) *Nature (London)* **296**, 713–721
554. Rohl, C. A., and Baldwin, R. L. (1994) *Biochemistry* **33**, 7760–7767
555. Gemmecker, G., Jahnke, W., and Kessler, H. (1993) *J. Am. Chem. Soc.* **115**, 11620–11621
556. Zhang, Y.-Z., Paterson, Y., and Roder, H. (1995) *Protein Sci.* **4**, 804–814
557. Zhang, Z., and Smith, D. L. (1993) *Protein Sci.* **2**, 522–531
558. Wagner, D. S., Melton, L. G., Yan, Y., Erickson, B. W., and Anderegg, R. J. (1994) *Protein Sci.* **3**, 1305–1314
559. Zhang, Z., Post, C. B., and Smith, D. L. (1996) *Biochemistry* **35**, 779–791
560. Zappacosta, F., Pessi, A., Bianchi, E., Venturini, S., Sollazzo, M., Tramontano, A., Marino, G., and Pucci, P. (1996) *Protein Sci.* **5**, 802–813
561. Kim, K.-S., Fuchs, J. A., and Woodward, C. K. (1993) *Biochemistry* **32**, 9600–9608
562. Eriksson, M. A. L., Härd, T., and Nilsson, L. (1995) *Biophys. J.* **69**, 329–339
563. Spyracopoulos, L., and O'Neil, J. D. J. (1994) *J. Am. Chem. Soc.* **116**, 1395–1402
564. Markley, J. L. (1975) *Biochemistry* **14**, 3546–3554
565. Schmidt-Rohr, K., and Spiess, H. W. (1994) *Multidimensional Solid-State NMR and Polymers*, Academic Press, San Diego, California
566. Schmidt-Rohr, K. (1996) *J. Am. Chem. Soc.* **118**, 7601–7603
567. Wu, C. H., Ramamoorthy, A., Gierasch, L. M., and Opella, S. J. (1995) *J. Am. Chem. Soc.* **117**, 6148–6149
568. Meyer, E. F. (1997) *Protein Sci.* **6**, 1591–1597
569. Stampf, D. R., Felder, C. E., and Sussman, J. L. (1995) *Nature (London)* **374**, 572–573
570. Holm, L., and Sander, C. (1995) *Trends Biochem. Sci.* **20**, 478–480
571. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol.* **247**, 536–540
572. Walsh, L., Newell, M., Ruczaj, H., Reid, T., and Desmond, P., eds. (1997) *Trends Guide to the Internet*, Elsevier, Cambridge, UK
573. Hogue, C. W. V., Ohkawa, H., and Bryant, S. H. (1996) *Trends Biochem. Sci.* **21**, 226–229
574. Wilkins, M. R., Hochstrasser, D. F., Sanchez, J.-C., Bairoch, A., and Appel, R. D. (1996) *Trends Biochem. Sci.* **21**, 496–497
575. Peitsch, M. C., Wells, T. N. C., Stampf, D. R., and Sussman, J. L. (1995) *Trends Biochem. Sci.* **20**, 82–84
576. Sayle, R. A., and Milner-White, E. J. (1995) *Trends Biochem. Sci.* **20**, 374–376
576a. Bernstein, H. J. (2000) *Trends Biochem. Sci.* **25**, 453–454
577. Sokolik, C. W. (1995) *Trends Biochem. Sci.* **20**, 122–124
578. Hunt, T. (1996) *Trends Biochem. Sci.* **21**, 74
579. Service, R. F. (2000) *Science* **287**, 1954–1956
580. Laskowski, R. A., Hutchinson E. G., Michie, A. D., Wallace, A. C., Jones, M. L., and Thornton, J. M. (1997) *Trends Biochem. Sci* **22**, 488–490
581. Cohn, E. J., and Edsall, J. T. (1943) *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*, Van Nostrand-Reinhold, Princeton, New Jersey (pp. 90–93)
582. Abbott, A. (2000) *Nature (London)* **408**, 130–132
583. Simmerling, C., Lee, M. R., Ortiz, A. R., Kolinski, A., Skolnick, J., and Kollman, P. A. (2000) *J. Am. Chem. Soc.* **122**, 8392–8402

## Study Questions

1. a. From the expression for the dissociation constant of an acid, HA (Eq. 3-1), derive the logarithmic form $pH - pK_a = \log[A^-] / [HA] = \log_{10} \alpha/(1-\alpha)$ (Eq. 3-11), where $\alpha$ is the fraction of the acid in the ionized form.

   b. The apparent dissociation constant $K_a$ for the $H_2PO_4^-$ ion at 25°C and 0.5 M total phosphate concentration is $1.380 \times 10^{-7}$ M. What will be the pH of a solution 0.025 M in $KH_2PO_4$ and 0.25 M in $Na_2HPO_4$? This is a National Bureau of Standards buffer (see Bates[7]).

   c. Suppose that you wanted to prepare a buffer of pH = 7.00 at 25°C from anhydrous $KH_2PO_4$ ($M_r$ = 136.09) and $Na_2HPO_4$ ($M_r$ = 141.98). If you placed 3.40 g of $KH_2PO_4$ in a 1 liter volumetric flask, how much anhydrous $Na_2HPO_4$ would you have to weigh and add before making to volume to obtain the desired pH? If you wanted to have the correct pH to ± 0.01 unit, how accurately would you have to weigh your salts? NOTE: It is quicker to prepare a buffer of precise pH this way than it is to titrate a portion of buffer acid to the desired pH with sodium hydroxide.

2. The apparent $pK_a$ for 0.1 M formic acid is 3.70 at 25°C.

   a. Concentrated HCl was added to a liter of 0.1 M sodium formate until a pH of 1.9 was attained. Calculate the concentration of formate ion and that of unionized formic acid.

   b. Calculate the hydrogen ion concentration.

   c. How many equivalents of HCl had to be added in part a to bring the pH to 1.9?

3. Exactly 0.01-mol portions of glycine were placed in several 100-ml volumetric flasks. The following exact amounts of HCl or NaOH were added to the flasks, the solutions were made to volume with water, mixed, and the pH measured. Calculate the $pK_a$ values for the carboxyl and amino groups from the following, making as many independent calculations of each p*K*a as the data permit. At low pH values *you must correct for the free hydrogen ion concentration* (see question 2c).[581]

| Flask No. | Mol HCl | Mol NaOH | pH |
|---|---|---|---|
| 1 | 0.010 | | 1.71 |
| 2 | 0.009 | | 1.85 |
| 3 | 0.006 | | 2.25 |
| 4 | 0.002 | | 2.94 |
| 5 | | 0.002 | 9.00 |
| 6 | | 0.004 | 9.37 |
| 7 | | 0.005 | 9.60 |

4. Using the $pK_a$ values from problem 3, construct the theoretical titration curve showing the equivalents of $H^+$ or $OH^-$ reacting with 1 mol of glycine as a function of pH. Note that the shape of this curve is independent of the $pK_a$. Sketch similar curves for glutamic acid ($pK_a$'s equal 2.19, 4.25, and 9.67), histidine ($pK_a$'s equal 1.82, 6.00, and 9.17) and lysine ($pK_a$'s equal 2.18, 8.95, and 10.53).

   Compare your plot for glycine with a plot of 1 M acid or base added to 0.01 mol of glycine in 100 ml of water. You may also compare your curves with those for glycine published in other textbooks.

5. Make a table of characteristic $pK_a$ values for acidic and basic groups in proteins. Which of these groups contribute most significantly to the titration curves of proteins?

6. If placed in water and adjusted to a pH of 7, will the following migrate toward the anode or the cathode if placed in an electrical field? (a) Aspartic acid, (b) alanine, (c) tyrosine, (d) lysine, (e) arginine, and (f) glutamine

7. The tripeptide L-Ala – L-His – L-Gln had the following $pK_a$ values: 3.0 ($\alpha$-COOH), 9.1 ($\alpha$-$NH_3^-$), and 6.7 (imidazolium).

   a. What is the isoelectric pH (pI) of the peptide, i.e., the pH at which it will carry no net charge? Hint, the pI for amino acids is usually given approximately as the arithmetic mean of two $pK_a$ values.[581]

   b. Draw the structures of the ionic forms of the peptide that occur at pH 5 and at pH 9. At each pH compute the fraction of the peptide in each ionic form.

## Study Questions

8. a. Write the structure for glycyl-L-tryptophanyl-L-prolyl-L-seryl-L-lysine.

   b. What amino acids could be isolated from it following acid hydrolysis?

   c. Following alkaline hydrolysis?

   d. After nitrous acid treatment followed by acid hydrolysis

   e. In an electrolytic cell at pH 7.0 would the peptide migrate toward the cathode or toward the anode? What is the approximate isoelectric point of the peptide?

   f. If a solution of this peptide were adjusted to pH 7, and then titrated with sodium hydroxide in the presence of 10% formaldehyde, how many equivalents of base would be required per mole of peptide to raise the pH to 10?

9. A peptide was shown to contain only L-lysine and L-methionine. Titration of the peptide showed 3 free amino groups for each free carboxyl group present, and each amino group liberated 1 mole of $N_2$ when the peptide was treated with $HNO_2$ in the Van Slyke apparatus. When the deaminated peptide was hydrolyzed completely in acid and the hydrolyzate again treated with $HNO_2$, the same amount of $N_2$ is liberated as that derived from the intact peptide. A sample of the original peptide was treated with excess dinitrofluorobenzene to give a dinitrophenyl (DNP) peptide, which was shown spectrophotometrically to contain three DNP groups per free carboxyl group. When this DNP-peptide was completely hydrolyzed, the following products were found: a colorless compound containing S ($A_1$) and a yellow compound containing S ($A_3$). Partial hydrolysis of the DNP-peptide yields $A_1$, $A_2$, $A_3$, plus 4 additional yellow compounds, $B_1$, $B_2$, $B_3$, and $B_4$. On complete hydrolysis, $B_1$ yields $A_1$, $A_2$, and $A_3$; $B_2$ yields $A_1$ and $A_2$; $B_3$ yields $A_1$ and $A_3$; and $B_4$ yields $A_3$ only.

   What is the most probable structure of the original peptide?

10. (a) Explain two advantages of the isotope dilution method of analysis. (b) From the following data, calculate the amount of cyclic AMP (cAMP) present per ml of human gluteus maximus muscle cells. Cells were treated with $^{32}P$-enriched cAMP, S.A. = 50 μCi / μmol for 0.2 h (all cAMP was taken up by cells), cells were homogenized, and the soluble cAMP was isolated and purified. The specific activity of the isolated cAMP = 10 μCi / μmol. The total amount of cAMP added was $1.0 \times 10^{-7}$ mol per ml of cells.

11. The figure in Box 3-C shows a high-resolution separation of the soluble proteins of *E. coli*. The investigator labeled the proteins with $^{14}C$-containing amino acids.

    a. How would you carry out the labeling experiment?

    b. What other isotope(s) could be used to label proteins? What chemical form(s) would you use? What limitations might there be?

    c. What soluble components of an *E. coli* cell sonicate might interfere with the two-dimensional separation, and how could they be removed?

    d. What technique(s) other than radioactive labeling could be used for locating proteins?

    e. If *all* the soluble proteins of *E. coli* were detected, about how many separate proteins would you expect to see?

    f. Indicate two or more properties of the resolution technique which are most significant in making it applicable to a system containing a very large number of proteins.

12. $^{35}S$ is a beta emitter, with no gamma or other type of radiation. It has the following properties: $t_{1/2}$ = 86.7 days, $\varepsilon_{max}$ = 0.168 MeV.

    a. Write the equation for the radiochemical decomposition of $^{35}S$.

    b. Discuss the advantages and limitation of the use of $^{35}S$ as an isotopic tracer.

Each cotton fiber is a single cell seed hair, ~30 mm in length. The dry fiber is ~95% cellulose, which constitutes the secondary cell wall (See Fig. 20-4,D) and is also present in the primary cell wall. The fibers, which are ~30 nm in length, consist of many parallel chains (Fig. 4-5) ~5 µm in length, each containing ~10,000 glucose residues. Van der Waals forces, together with one hydrogen bond per glucose, contribute to the stability of the tightly packed fiber. Cotton is one of the major agricultural crops, ~87 million bales, each ~220 kg, being produced annually world-wide. Photo and information from A. D. French and M. A. Godshall, Southern Regional Research Center, USDA, New Orleans, LA.

# Contents


161 ..... **A. Structures and Properties of Simple Sugars**
162 ..... 1. The Variety of Monosaccharides
163 ..... *Ways of indicating configuration*
164 ..... *Natural derivatives of sugars*
166 ..... 2. Conformations of Five- and Six-Membered Ring Forms
167 ..... 3. Characteristic Reactions of Monosaccharides
167 ..... **B. Glycosides, Oligosaccharides, Glycosylamines, and Glycation**
169 ..... **C. Polysaccharides (Glycans)**
170 ..... 1. Conformations of Polysaccharide Chains
172 ..... 2. The Glucans
175 ..... 3. Other Homopolysaccharides
175 ..... 4. Heteropolysaccharides of the Animal Body
177 ..... 5. Plant Heteropolysaccharides
179 ..... 6. Polysaccharides of Bacterial Surfaces
180 ..... **D. Glycoproteins and Proteoglycans**
181 ..... 1. *O*-Linked Glycosyl Groups
182 ..... 2. Asparagine-Linked Oligosaccharides
186 ..... 3. Glycoproteins in Biological Recognition
186 ..... *Lectins and other carbohydrate-binding proteins*
186 ..... *Carbohydrate-binding sites*
186 ..... *Binding of viruses and bacteria to cells*
187 ..... *Aggregation and adherence of cells*
187 ..... *Growth and differentiation*
188 ..... *Recognition and adhesion by leukocytes*
188 ..... **E. Some Special Methods**
188 ..... 1. Release of Oligosaccharides from Glycoproteins
189 ..... 2. Hydrolysis
189 ..... 3. Methylation
190 ..... 4. Periodate Oxidation (Smith Degradation)
190 ..... 5. Nuclear Magnetic Resonance

192 ..... **References**
196 ..... **Study Questions**


**Boxes**


171 ..... Box 4-A Cyclodextrins
178 ..... Box 4-B Silicon: An Essential Trace Element
184 ..... Box 4-C The Blood Group Determinants
191 ..... Box 4-D Antifreeze and Ice-Nucleation Proteins


**Tables**


170 ..... Table 4-1 Some of the Many Polysaccharides Found in Nature

# Sugars, Polysaccharides, and Glycoproteins

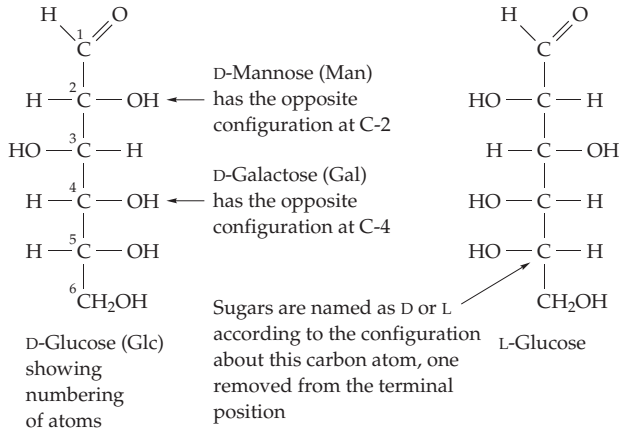
# 4

We are all familiar with sugars, important components of our diet which are present in fruits, honey, table sugar, and syrups. Within our bodies the simple sugar **D-glucose** is an essential source of energy. It is present in blood at a relatively constant concentration of 5.5 mM and is carried to all tissues. A **polysaccharide** called **glycogen**, a polymeric form of glucose, provides a reserve of readily available energy within our cells. **Starches** and other polysaccharides store energy within plants. Polysaccharides also have major *structural* functions in nature. **Cellulose**, another polymer of glucose, forms the fibers of cotton, plant cell walls, and wood. Both the tough exoskeletons of anthropods and the cell walls of fungi depend for their strength on the nitrogen-containing polysaccharide **chitin**. Polysaccharides that carry many negative charges, such as **hyaluronan,** form a protective layer between animal cells, while **pectins** play a similar role in plants.

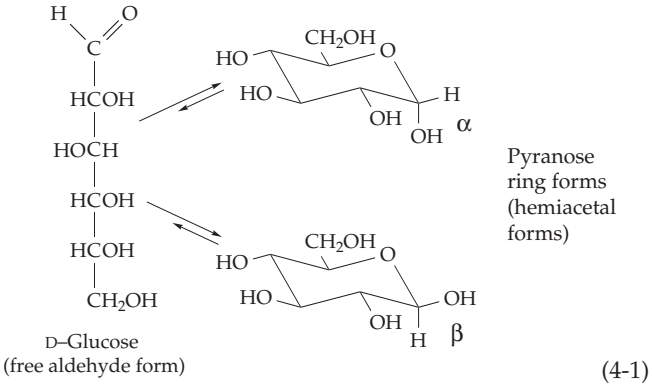
A third function of sugar residues (**glycosyl groups**) is in biological recognition and communication. The outer surfaces of cells are nearly covered by covalently attached **oligosaccharides**, small polymeric arrays of sugar rings. Some are attached to side chain groups of proteins and others to lipids to form **glycoproteins** and **glycolipids,** respectively. Because of the variety of different sugars, the various ways in which they can be linked, and their ability to form oriented hydrogen bonds, these oligosaccharides provide much of the chemical code for identifying cells. This coding enables cells to attach to each other in correct ways during development of a multicelled organism. It helps to activate our immune system to attack parasites and also helps bacteria to attack us! We are only beginning to understand the numerous critical functions of the glycosyl groups of cell surfaces.

## A. Structures and Properties of Simple Sugars

The simple sugars, or **monosaccharides**, are polyhydroxyaldehydes (**aldoses**) or polyhydroxyketones (**ketoses**).[1–5] All have the composition $(CH_2O)_n$, hence the family name **carbohydrate**. A typical sugar, and the one with the widest distribution in nature, is glucose.

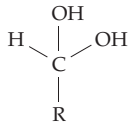



The carbonyl group of this and other sugars is highly reactive and a characteristic reaction is **addition** of electron-rich groups such as –OH. If a sugar chain is long enough (4–6 carbon atoms) one of the hydroxyl groups of the same molecule can add to the carbonyl group to form a cyclic **hemiacetal** or ring form, which reaches an equilibrium with the free aldehyde or ketone form (Eq. 4-1). The six-membered rings formed in this way (**pyranose** rings) are especially stable, but five-membered **furanose** rings also exist in many carbohydrates.



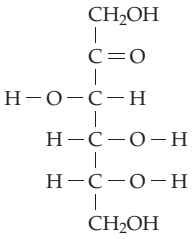

(4-1)

The ring forms of the sugars are the monomers used by cells to form polysaccharides. Indeed, it is the natural tendency of 5- and 6-carbon sugars (**pentoses** and **hexoses**) to cyclize that permits formations of stable sugar polymers from the reactive and unstable monomers. When a sugar cyclizes a new chiral center is formed at the **anomeric carbon atom**, the atom that was present in the original carbonyl group. The two configurations about this carbon atom are designated α and β as indicated in Eq. 4-1. In an equilibrium mixture, *ring forms of most sugars predominate over open chains.*[6–13] Thus, at 25°C in water glucose reaches an



Covalent hydrate
of R—CHO

equilibrium with ~0.001% free aldehyde, 0.004% covalent hydrate of the aldehyde (see also Eq. 13-1), 39% α-pyranose form, 61% β-pyranose form, and 0.15% each of the much less stable α and β furanose forms.[9,12] The ketose sugar **fructose** in solution exists as ~73% β-pyranose, 22% β-furanose, 5% α-furanose, and 0.5% open chain form (Fig. 4-1).[6,10,13] Although polysaccharides are composed almost exclusively of sugar residues in ring forms, the open chain forms are sometimes metabolic intermediates.



D-Fructose (Fru), a ketose that is a close structural and metabolic relative of D-glucose. It occurs in honey and fruit juices in free form, in the disaccharide sucrose (table sugar) as a 5-membered furanose ring, and in other oligosaccharides and polysaccharides.

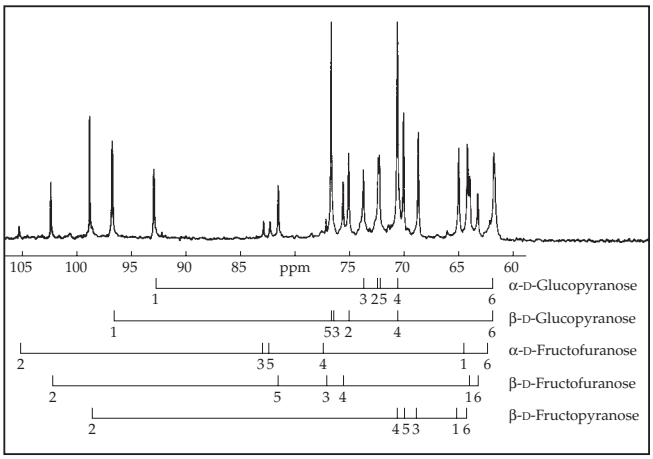



**Figure 4-1** Natural abundance $^{13}$C-NMR spectrum of honey showing content of α and β pyranose ring forms of glucose and both pyranose and furanose ring forms of fructose. Open chain fructose, with a peak at 214 ppm, was present in a trace amount. From Prince *et al.*[6]

Because of their many polar hydroxyl groups, most sugars are very soluble in water. However, hydrogen bonds between molecules stabilize sugar crystals making them insoluble in nonpolar solvents. Intermolecular hydrogen bonds between chains of sugar rings in cellulose account for much of the strength and insolubility of these polysaccharides.

## 1. The Variety of Monosaccharides

Sugars contain several chiral centers and the various diastereomers are given different names. The commonly occurring sugars **D-glucose**, **D-mannose**, and **D-galactose** are just three of the 16 diastereomeric aldohexoses. The Fischer projection formulas for the entire family of eight D-aldoses with 3–6 carbon atoms are given in Fig. 4-2. Several of these occur only rarely in nature.

Monosaccharides are classified as D or L according to the configuration at the chiral center farthest from the carbonyl group. If the –OH group attached to this carbon atom lies to the right when the sugar is oriented according to the Fischer convention, the sugar belongs to the D family. The simplest of all the chiral sugars is glyceraldehyde. The family of aldoses in Fig. 4-2 can be thought of as derived from D-glyceraldehyde by upward extension of the chain. Besides D-glucose, D-mannose, and D-galactose, the most abundant naturally occurring sugar shown in Fig. 4-2 is **D-ribose**, a major component of RNA. Another abundant aldose is **D-xylose**, a constituent of the polysaccharides of wood.

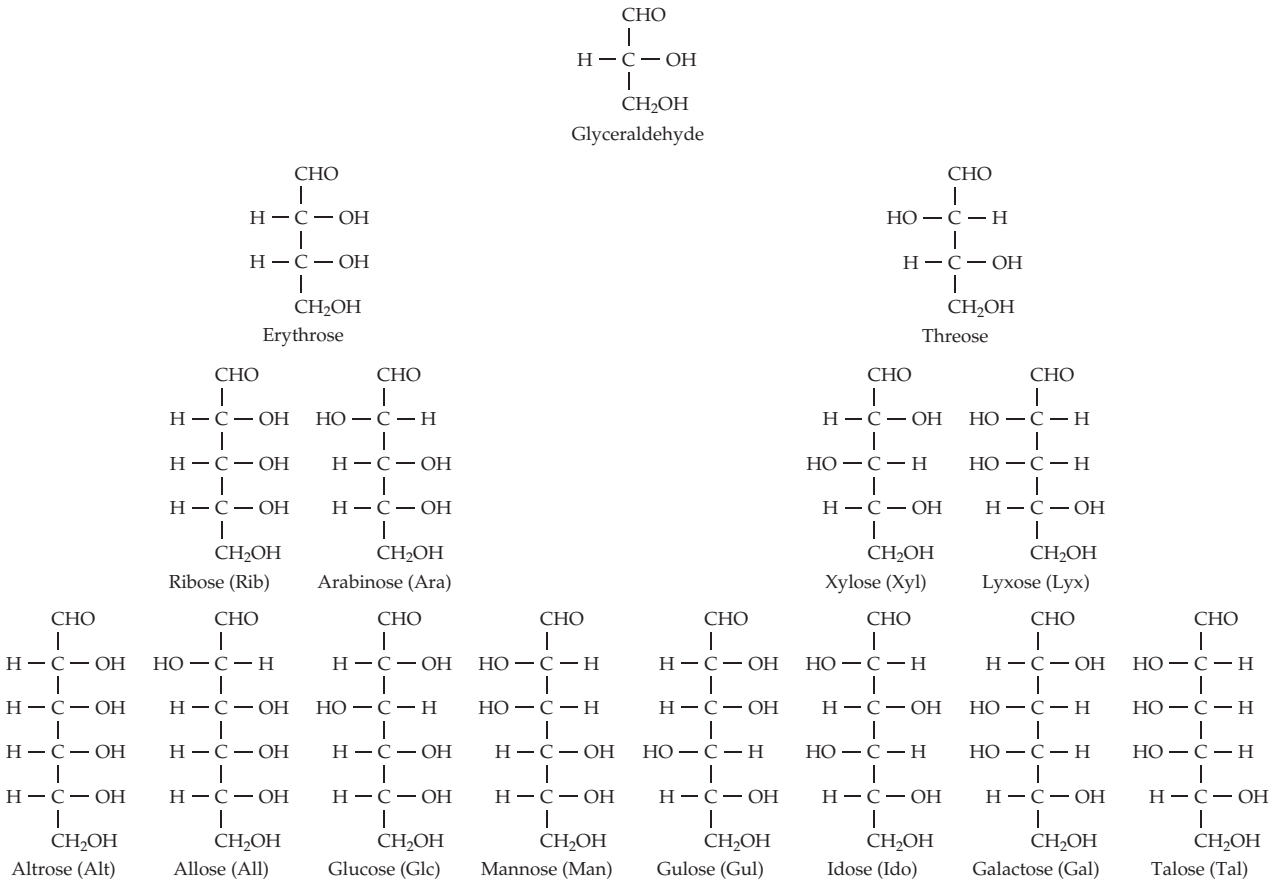
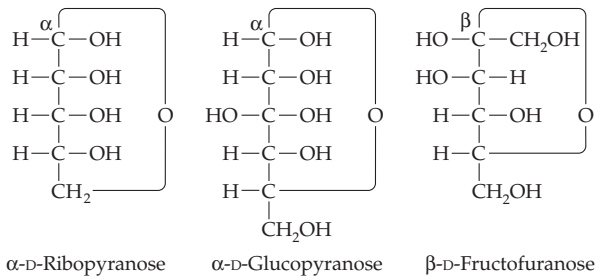



**Figure 4-2** Formulas for the D-aldoses. Prefixes derived from the names of these aldoses are used in describing various other sugars including ketoses. The prefixes include *erythro*, *threo*, *arabino*, *ribo*, *galacto*, *manno*, etc., D-Fructose can be described in this manner as D-*arabino*-hexulose. The prefixes refer to the configurations of a series of consecutive but not necessarily contiguous chiral centers.[2] Thus, 3-*arabino*-hexulose is an isomer of fructose with the carbonyl group at the 3 position. The prefix deoxy, which means "lacking oxygen," is often used to designate a modified sugar in which an –OH has been replaced by –H, e.g., 2-deoxyribose.

For each D sugar there is an L sugar which is the complete mirror image or enantiomer of the D form. Although **L-arabinose** occurs widely in plants, and some derivatives of L sugars are present in glycoproteins, most naturally occurring sugars have the D configuration. A pair of sugars, such as glucose and galactose, differing in configuration at only one of the chiral centers are known as **epimers**. The D-ketose sugars with the carbonyl group in the 2 position (Fig. 4-3) are also abundant in nature, often occurring in the form of phosphate esters as intermediates in metabolism.

***Ways of indicating configuration.*** The Fischer projection formulas used in Figs. 4-2 and 4-3 are convenient in relating the sugar structures by their individual carbon configurations to each other, but they give an unrealistic three-dimensional picture. According to the Fischer convention each carbon atom must be viewed with both vertical bonds projecting behind the atom viewed. In fact, the molecule cannot assume such a conformation because the chain folds back on itself, bringing many atoms into collision.

Ring forms of sugars are also often drawn according to the Fischer convention; making use of elongated bent lines to represent ordinary simple bonds:




In this representation the hydroxyl at the anomeric

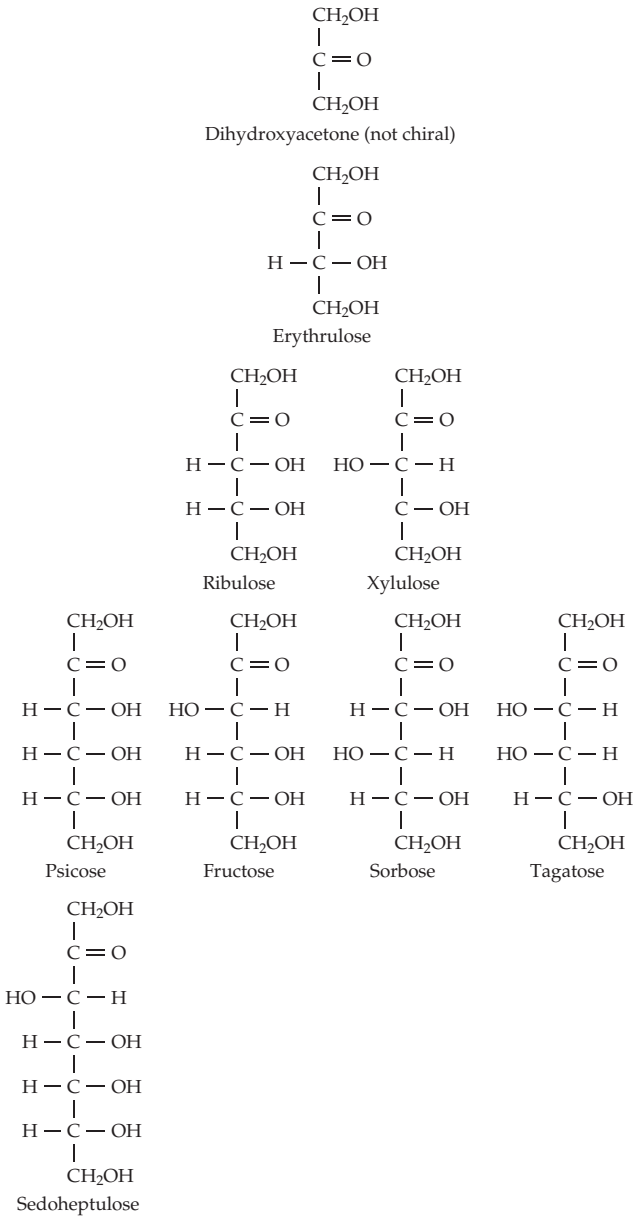
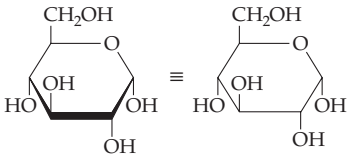



**Figure 4-3** Formulas for the open forms of the D-ketoses.

carbon atom is always on the right side of the molecules for α forms in the D series of sugars and at the left side for the β forms. For the L sugars the opposite is true. For example, since α-D-glucopyranose and α-L-glucopyranose are enantiomers, their Fischer formulas must be mirror images.

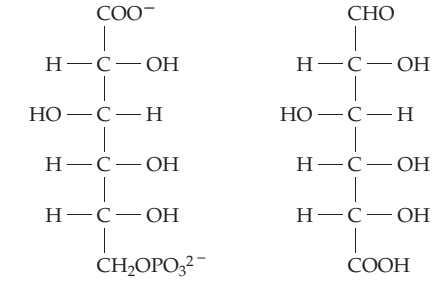
Simplified sugar rings are often drawn with **Haworth structural formulas**. The lower edge of the ring, which may be shown as a heavy line, is thought of as projecting out toward the reader and the other edge as projecting behind the plane of the paper.



α-D-Glucopyranose (Haworth structural formula)

Haworth structures are easy to draw and unambiguous in depicting configurations,[14] but they also do not show the spatial relationships of groups attached to other rings correctly. For this reason conformational formulas of the type described in Section 2 and shown in Fig. 4-4 are used most often in this book.

***Natural derivatives of sugars.*** The aldehyde group of an aldose can be oxidized readily to a carboxyl group to form an **aldonic acid**. Among the several aldonic acids that occur naturally is 6-phosphogluconic acid, which is pictured here as the 6-phosphogluconate ion:



6-Phosphogluconate ion  Glucuronic acid (GlcA)

The enzymatically formed **uronic acids** have –COOH in the terminal position. This is position 6 in **glucuronic acid**, whose structure is given here and as a ring in Fig. 4-4. Sugar chains with –COOH at both ends are called **aldaric acids**, e.g., glucaric acid. The –OH group in the 2 position of glucose may be replaced by $-NH_2$ to form 2-amino-2- deoxyglucose, commonly called **glucosamine** (GlcN), or by $-NH-CO-CH_3$ to form ***N*-acetylglucosamine** (GlcNAc). Similar derivatives of other sugars exist in nature. In many poly-saccharides, sulfate groups are attached in ester linkage to the sugar units. The sulfo ($-SO_3^-$) sugar **6-sulfo-α-D-quinovose** (Fig. 4-4) is found in lipids of photosynthetic membranes.[15,16]

The sugar alcohols, in which the carbonyl group has been reduced to –OH, also occur in nature. For example, **D-glucitol** (D-sorbitol), the sugar alcohol obtained by reducing either D-glucose or L-sorbose (Eq. 4-2), is a major product of photosynthesis and widely distributed in bacteria and throughout the eukaryotic kingdom. It is present in large amounts in berries of the mountain ash and in many other fruits. It exists in a high concentration in human semen and

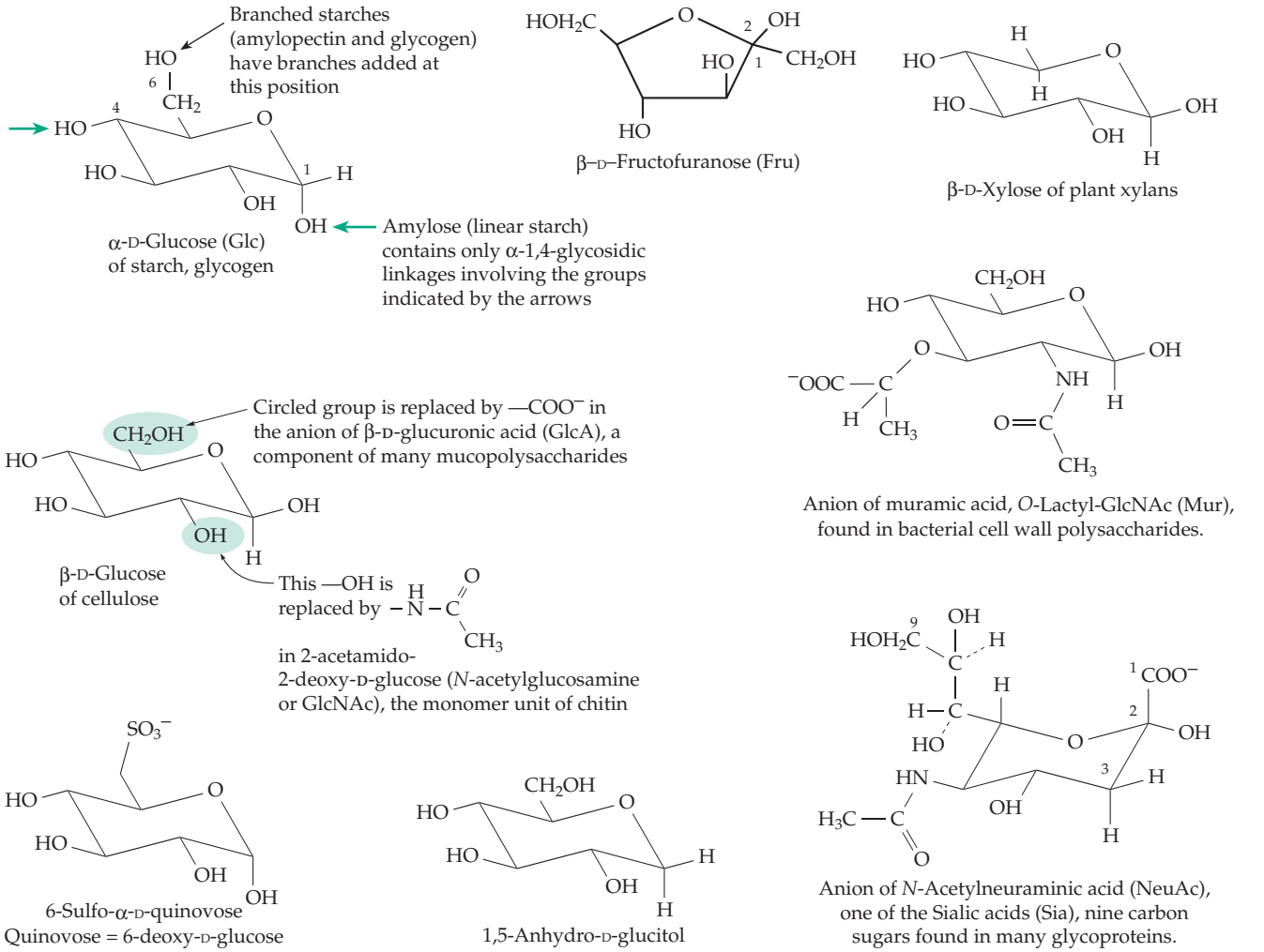
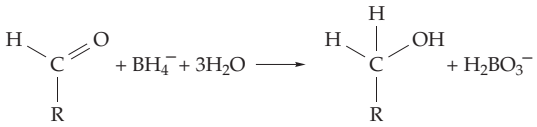
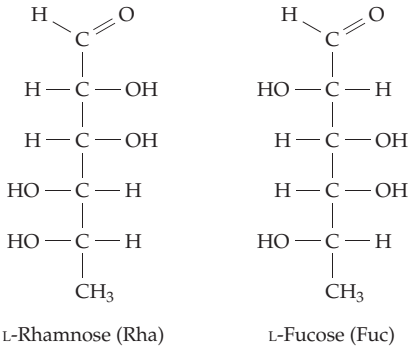



**Figure 4-4** Some simple sugars and sugar derivatives in ring forms. Most of these are present in polysaccharides.



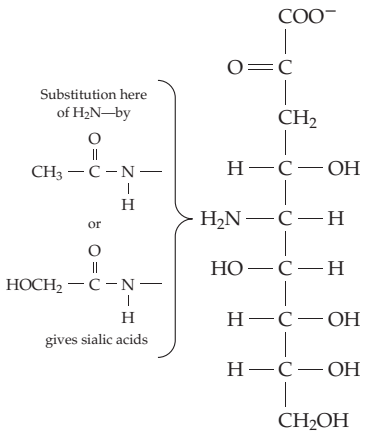

(4-2)




accumulates in lenses of diabetics. D-Glucitol and other sugar alcohols arise in some fungi during metabolism of the corresponding sugars.[17] Mannitol, another product of photosynthesis, is also present in many organisms.[18] Another polyol that is found in human blood in significant concentrations (0.06–0.25 mM) is 1,5-anhydro-D-glucitol (Fig. 4-4). It is largely of dietary origin.[19]

Two common 6-deoxy sugars which lack the hydroxyl group at C-6 are **rhamnose** and **fucose**. Both are of the "unnatural" L configuration but are derived metabolically from D-glucose and D-mannose, respectively.

Vitamin C (ascorbic acid, Box 18-D) is another important sugar derivative. **Neuraminic acid** is a 9-carbon sugar made by transferring a 3-carbon piece onto a hexosamine. Its *N*-acetyl and *N*-glycolyl derivatives are called **sialic acids**. Their names may be abbreviated NeuAc and NeuGl, respectively, or simply as Sia (see also Fig. 4-4).

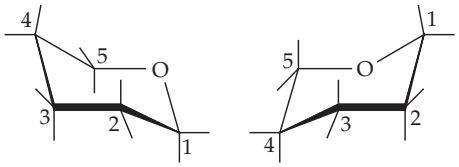



Neuraminic acid anion (Neu)

The sialic acids are prominent constituents of the glycoproteins of cell surfaces. More than 30 modified forms, for example, with added methyl or acetyl groups, are known.[20–24]

## 2. Conformations of Five- and Six-Membered Ring Forms

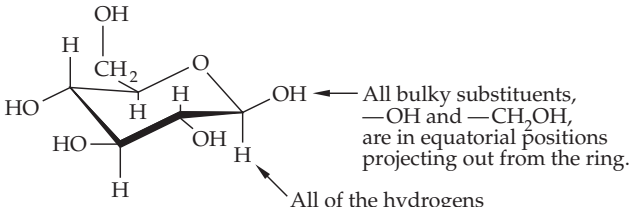
Single-bonded six-membered molecular rings, such as those in cyclohexane and in sugars, most often assume a **chair (C)** conformation. For sugars in pyranose ring forms, there are two possible chair conformations,[4,25,26] which are designated $^4C_1$(or C1) and $^1C_4$ (or 1C). The superscript and subscript numbers on the designations indicate which atoms are above and




$^4C_1$ (or C1) $^1C_4$ (or 1C)

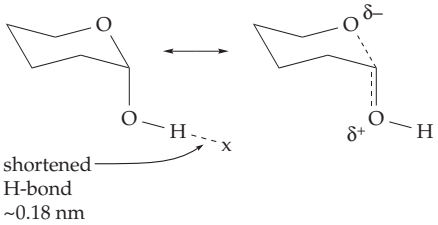
below the plane of the other four ring atoms. *These conformers are not easily interconvertible unless the ring is opened by a chemical reaction,* e.g., that of Eq. 4-1. However, manipulation with an atomic force microscope has shown that stretching of single polysaccharide chains can cause interconversion of the two chair conformations of pyranose rings, which are separated by an energy barrier of ~46 kJ/mol.[26a]

Most sugars occur in the chair conformation that places the largest number of substituents in equatorial positions and is therefore most stable thermodynamically. For D-aldoses this is usually the $^4C_1$ conformation:




β-D-Glucose, a pyranose ring form in a *chair* conformation.

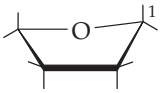
For L-aldoses it is the $^1C_4$ form. There are exceptions to this rule. For example, α-D-idopyranose as well as iduronate rings assume the $^1C_4$ conformation because this conformation places the maximum number of bulky groups in equatorial positions.[27,28] It is noteworthy that electronegative substituents on the anomeric carbon atom of a sugar ring often prefer an axial orientation. This **anomeric effect** is also reflected in a shortening by about 0.01 nm of hydrogen bonds involving the hydrogen atom of the anomeric axial OH group. These effects can be explained partially as a result of coulombic repulsion of the two C–O dipoles and of resonance of the following type:[29–31]




In addition to the chair conformations of six-membered rings the less stable **boat (B)** conformations are also possible. The six boat forms are smoothly interconvertible through intermediate **twist** (**T**) forms, which are also called **skew** (**S**) forms.[3,5] Since the

Boat Skew or twist

internal angle in a pentagon is 108° (close to the tetrahedral angle) we might anticipate a nearly planar five-membered ring. However, eclipsing of hydrogen atoms on adjacent carbons prevents formation of such a flat structure. One of the atoms may be buckled out of the plane of the other four about 0.05 nm, into an **envelope conformation** (e.g., $^1E$), or only three atoms may be in a plane, as in a twist conformation.

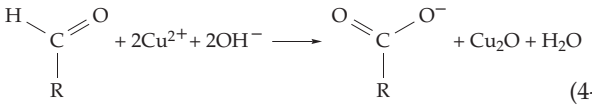



The $^1E$ envelope conformation

Any one of the five atoms of the ring can be either above or below the plane defined by the other four in the envelope conformation. The energy barriers separating them are very low, and in cyclopentane or in proline all of the envelope conformations are freely interconvertible through intermediate skew forms.[32] Furanose sugar rings are very flexible but the presence of the bulky substituents reduces the number of possible conformations.[33–36a] See Chapter 5 for further discussion.

### 3. Characteristic Reactions of Monosaccharides

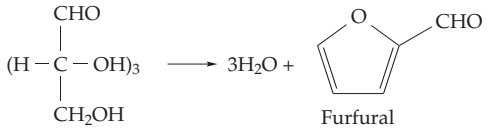
The aldehyde group of aldoses can either be oxidized or reduced and ketoses can be reduced. The best laboratory reagent for reduction is **sodium borohydride** which acts rapidly in neutral aqueous solutions (see Eq. 4-2). Since both $NaB^3H_4$ and $NaB^2H_4$ are available, radioactive or heavy isotope labels can be introduced in this way. The aldehyde groups can be oxidized by a variety of agents to the corresponding aldonic acids, a fact that accounts for the reducing properties of these sugars. In alkaline solution aldoses reduce $Cu^{2+}$ ions to cuprous oxide (Eq. 4-3), silver ions



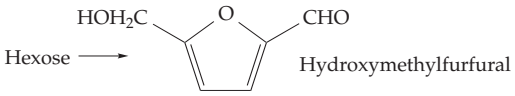

(4-3)

to the free metal, or hexacyanoferrate (III) to hexacyanoferrate (II). These reactions provide the basis for sensitive analytical procedures. Even though the aldoses tend to exist largely as hemiacetals (Eq. 4-1) the reducing property is strongly evident. Oxidation by metal-containing reagents is usually via the free aldehyde, but oxidation by hypobromite $BrO^-$ ($Br_2$ in alkaline solution) yields the lactone, as does enzymatic oxidation (Eq. 15-10).

Sugars are unstable in acid. Boiling with concentrated HCl or $H_2SO_4$ converts pentoses to furfural (Eq. 4-4a) and hexoses to hydroxymethylfurfural (Eq. 4-4b).




(4-4a)

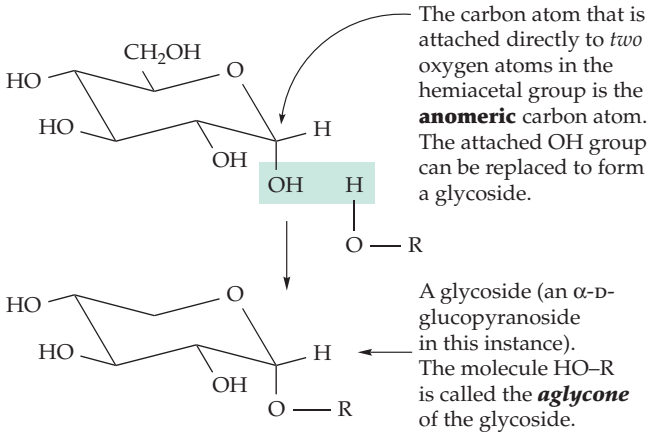



(4-4b)

The aldehydes produced in these reactions can be condensed with various phenols or quinones to give colored products useful in quantitative estimation of sugar content or in visualizing sugars on thin-layer chromatographic plates. Phenol and sulfuric acid yield a product whose absorbance at 470 nm can be used as a measure of the total carbohydrate content of most samples. Resorcinol (1,3-dihydroxybenzene) in 3 M HCl (Seliwanoff's reagent) gives a red precipitate with ketoses. Orcinol (5-methylresorcinol) reacts rapidly with pentoses and with ribonucleosides and ribonucleotides. Since 2-deoxy sugars react slowly this can be used as a test for RNA. Diphenylamine with $H_2SO_4$ gives a blue-green color specifically with 2-deoxy sugars and can be used to test for DNA if the sample is first hydrolyzed.[37]

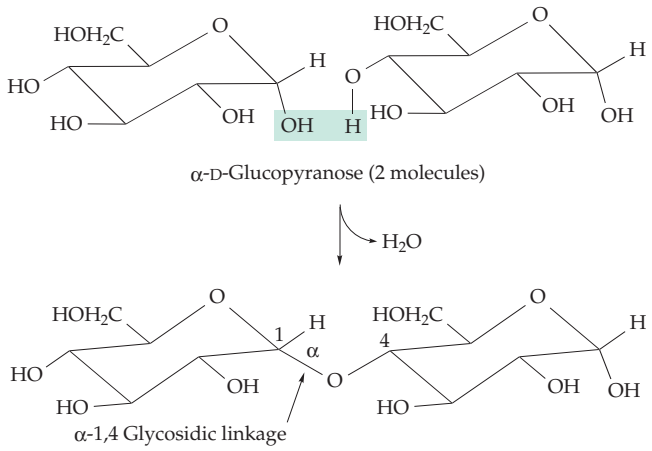
## B. Glycosides, Oligosaccharides, Glycosylamines, and Glycation

The hydroxyl group on the anomeric carbon atom of the ring forms of sugars is reactive and can be replaced by another nucleophilic group such as –O–R from an alcohol. The product is a glycoside (Eq. 4-5). The reaction with methanol occurs readily with acid catalysis under dehydrating conditions, e.g., in 100% methanol.



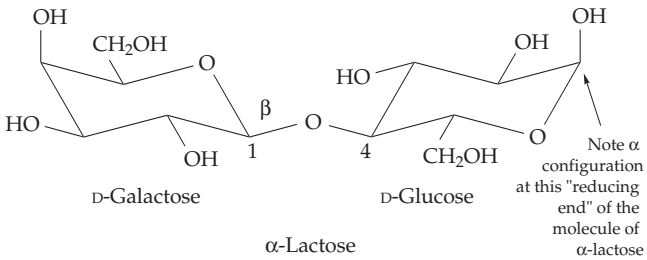

(4-5)

The alcohol in this equation can be a simple one such as methanol or it can be any of the –OH groups of another sugar molecule. For example, two molecules of α-D-glucopyranose can be joined, in an indirect synthesis, to form **maltose** (Eq. 4-6). Maltose is formed by the hydrolysis of starch and is otherwise not found in nature. There are only three abundant naturally occurring **disaccharides** important to the metabolism of plants and animals.[38] They are **lactose** (milk), **sucrose** (green plants), and **trehalose** (fungi and insects).




(4-6)

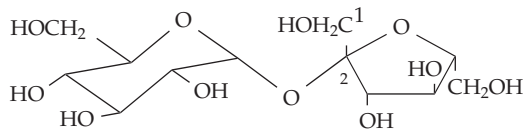
Disaccharides are linked by **glycosidic** (acetal) linkages. The symbol α-1,4, used in Eq. 4-6, refers to the fact that in maltose the glycosidic linkage connects carbon atom 1 (the anomeric carbon atom) of one ring with C-4 of the other and that the configuration about the anomeric carbon atom is α. While the α and β ring forms of free sugars can usually undergo ready interconversion, the configuration at the anomeric carbon atom is "frozen" when a glycosidic linkage is formed. To describe such a linkage, we must state this configuration together with the positions joined in the two rings (see Eq. 4-6). Lactose, whose structure follows, can be described as a disaccharide containing one galactose unit in a β-pyranose ring form and whose anomeric carbon atom (C-1) is joined to the 4 position of glucose, giving a β-1,4 linkage:

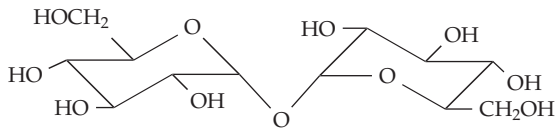



The systematic name for α-lactose, *O*-β-D-galactopyranosyl-(1→4)-α-D-glucopyranose, provides a complete description of the stereochemistry, ring sizes, and mode of linkages. This name may be abbreviated β-D-Gal*p*-(1→4)-α-D-Glc*p*. Since pyranose rings are so common and since most natural sugars belong to the D family, the designations D and *p* are often omitted. It may be assumed that in this book sugars are always D unless they are specifically designated as L. When linkages remain uncertain abbreviated formulas are given. Because the glucose ring in lactose is free to open to an aldehyde and to equilibrate (in solution) with other ring forms, the name lactose does not imply a fixed ring structure for the glucose half. Thus, the name lactose can be abbreviated as βGal-(1→4)-Glc or, more succinctly, as Galβ1→4Glc. However, in crystalline form the sugar exists either as α-lactose or β-lactose. The latter is more soluble and sweeter than α-lactose, which sometimes crystallizes in ice cream upon prolonged storage and produces a "sandy" texture. An isomer of lactose, Galβ1→6Glc or **allolactose**, is an important inducer of transcription in cells of *E. coli* (Chapter 28).

Notice that in the drawing of the lactose structure the glucose ring has been "flipped over" with respect to the orientation of the galactose ring, a consequence of the presence of the β-1,4 linkage. For maltose, where the linkage is α-1,4, the two rings are usually drawn with the same orientation (Eq. 4-6). Maltose can be described as α-D-Glc*p*-(1→4)-D-Glc*p* or more simply as Glcα1→4Glc.

In sucrose and in α,α-trehalose the reducing groups of two rings are joined. Each of these sugars exists in a single form. Sucrose serves as the major transport sugar in green plants, while trehalose plays a similar role in insects, as does D-glucose in our blood.




Trehalose, or "mushroom sugar," is found not only in fungi but also in many other organisms.[39–41]. It serves as the primary transport sugar in the hemolymph of insects and also acts as an "antifreeze" in many species. It accounts for up to 20% of the dry weight of **anhydrobiotic organisms**, which can survive complete dehydration. These include spores of some fungi, yeast cells, macrocysts of *Dictyostelium*, brine shrimp cysts (dried gastrulas of *Artemia salina*), some nematodes, and the resurrection plant. These organisms can remain for years in a dehydrated state. Hydrogen bonding between the trehalose and phosphatidylcholine may stabilize the dry cell membranes.[18,40,41] Although they can be desiccated, fungal spores remain dormant even when considerable water is present. One of the first detectable changes when the spores germinate is a rapid increase in the activity of the enzyme **trehalase**

which hydrolyzes trehalose to glucose.[42,43]

Disaccharides, as well as higher oligosaccharides and polysaccharides, are thermodynamically unstable with respect to hydrolysis, for example, for lactose in aqueous solution:

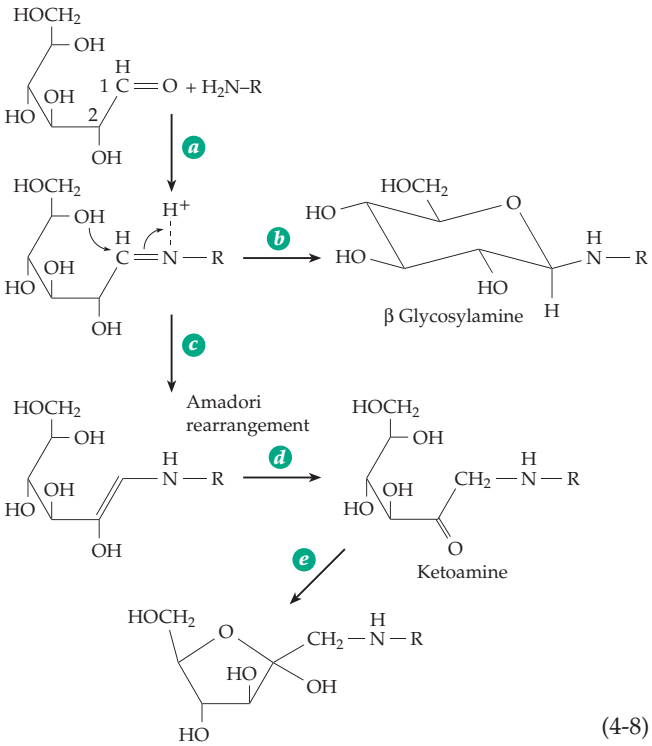
$$\text{Lactose} + H_2O \rightarrow \text{D-glucose} + \text{D-galactose}$$
$$\Delta G^\circ \approx -8.7 \pm 0.2 \text{ kJ mol}^{-1} \text{ at } 25^\circ\text{C} \qquad (4\text{-}7)$$

The corresponding equilibrium constant $K$ = [D-glucose][D-galactose]/[lactose] ≈ 34 M. For other oligosaccharides $K$ varies from 17 to 500 M.[44,45] However, sucrose is far less stable with $K = 4.4 \times 10^4$ M and $\Delta G^\circ = -26.5 \pm 0.3$ kJ mol$^{-1}$ for its hydrolysis.[46] Sucrose is also less stable kinetically and undergoes rapid acid-catalyzed hydrolysis. This fact is exploited when vinegar is used to convert sucrose to the less crystallizable mixture of glucose and fructose during candy-making. In marked contrast, trehalose is extremely resistant to acid-catalyzed hydrolysis.

The joining of additional sugar rings through glycosidic linkages to a disaccharide leads to the formation of **oligosaccharides**, which contain a few residues, and to **polysaccharides**, which contain many residues. Among the well-known oligosaccharides are the substituted sucroses **raffinose**, Gal*p*(1→6) Glc*p*(1→2) Fru*f*, and **stachyose**, Gal*p*(1→6) Gal*p*(1→6) Glc*p*(1→2) Fru*f*. Both sugars are found in many legumes and other green plants in which they are formed by attachment of the galactose rings to sucrose. Oligosaccharides have many functions. For example, gram-negative bacteria often synthesize oligosaccharides of 6–12 glucose units in β-1,2 linkage joined to *sn*-1-phosphoglyceryl groups. They are found in the periplasmic space between the inner and outer cell membranes and may serve to control osmotic pressure.[47] Oligosaccharides of 10–14 α-1,4-linked D-galacturonic acid residues serve as signals of cell wall damage to plants and trigger defensive reactions against bacteria in plants.[48]

Just as alcohols can be linked to sugars by glycoside formation (Eq. 4-5), amines can react similarly to give **glycosylamines** (*N*-glycosides). In this instance it is usually the free aldehyde that reacts via formation of a Schiff base (Eq. 4-8*a*). The latter can cyclize (Eq. 4-8*b*) to the glycosylamine with either the α or β configuration. Another important reaction of Schiff bases of sugars is the **Amadori rearrangement** (Eq. 4-8*c*,*d*), which produces a secondary ketoamine with a 1-oxo-2-deoxy structure.

This can cyclize as in step *e* of Eq. 4-8. In addition, epimerization at position 2 occurs through the reversal of step *d*. The overall reaction of Eq. 4-8 is often called **glycation** to distinguish it from **glycosylation,** the transfer of a glycosyl group. The Amadori rearrangement is important in nitrogen metabolism and in nonenzymatic reactions of sugars. For example, small




(4-8)

amounts of glycosylated hemoglobin and other proteins modified by glycation of the protein amino groups are normally present in the blood.[49–51] People with diabetes, who have a high concentration of blood glucose, have increased amounts of glycated protein. High concentration of either glucose or fructose[41,52] may cause serious problems. For example, modification of the protein **crystallins** of the lens of the eye may lead to cataracts. Similar problems with galactose may accompany galactosemia (Chapter 20).[53] Ketoamines formed by glycation of proteins may undergo crosslinking reactions with side chains of other protein molecules and this may be one cause of aging.[42,43] Other reactions of the ketoamines lead to formation of fluorescent and colored products.[41,43,50,54–56] Oxidation products of sugars also participate in these reactions,[57] and nucleic acid bases also react.[56]

## C. Polysaccharides (Glycans)

Polymers of sugars are present in all cells and serve a variety of functions.[58–60] The simple sugars commonly used in the assembly of polysaccharides include D-glucose, D-mannose, D-galactose, D-fructose, D-xylose, L-arabinose, related uronic acids, and amino sugars (Fig. 4-4). These monomer units can be put together in many ways, either as **homopolysaccharides** containing a single kind of monomer or as **heteropolysaccharides** containing two or more different monomers. Because there are many sugars and many ways in which they can be linked, there is a bewildering

**TABLE 4-1**
**Some of the Many Polysaccharides Found in Nature**

| Name | Source | Monomer | Main linkage | Branch linkages |
|---|---|---|---|---|
| Starch | Green plants | | | |
| Amylose | | D-Glucose | α1,4 | |
| Amylopectin | | D-Glucose | α1,4 | α1,6 |
| Glycogen | Animals, bacteria | D-Glucose | α1,4 | α1,6 |
| Cellulose | Green plants, some bacteria | D-Glucose | β1,4 | |
| Dextrans | Some bacteria | D-Glucose | α1,6 | α1,3 |
| Pullulan | Yeast | D-Glucose | α1,6 + α1,4 | |
| Callose | Green plants | D-Glucose | β1,3 | |
| Yeast glucan | Yeast | D-Glucose | β1,3 | |
| Schizophyllan, curdlan, paramylon | | D-Glucose | β1,3 | β1,6 on every third residue |
| Mannans | Algae | D-Mannose | 1,4 | |
| | Yeast | D-Mannose | α1,6 | |
| Xylans | Green plants | D-Xylose | β1,3 | |
| | Brown seaweed | | | |
| Inulin | Some plant tubers | D-Fructose | β2,6 | |
| Chitin | Fungi, anthropods | *N*-acetyl-D-Glucosamine | β1,4 | |
| **Alternating polysaccharides** | | | | |
| Hyaluronan | Animal connective tissue | Glucuronic acid + *N*-acetylglucosamine | β1,4 | |
| Chondroitin sulfate | | D-Glucosamine *N*-acetyl-D-Galactosamine | β1,3 + β1,4 | |
| Dermatan sulfate | | α-L-Iduronate + *N*-acetyl-D-Galactosamine | β1,3 + β1,4 | |
| Pectin | Higher plants | D-Galactunonate + others | β1,4 + others | |
| Alginate | Seaweed | D-Mannuronate + L-Guluronate | β1,4 + α1,4 | |
| Agar-agar | Red seaweed | Galactose | β1,4 and α1,3 | |
| Carageenan | Red seaweed | Galactose-4-sulfate + 3,6-anhydro-D-Galactose-2-sulfate | β1,4 + α1,3 | |
| Murein | Bacterial cell wall | *N*-acetyl-D-Glucosamine + *N*-acetyl-D-Muramic acid | β1,4 | |

variety of different polysaccharides. Their chains can be linear or helical. The most numerous functional groups present are the free hydroxyl groups, some of which may form additional glycosidic linkages to produce **branched chains**. Polysaccharides may also contain –COOH, $-NH_2$, $-NHCOCH_3$, and other groups. After polymerization, hydroxyl groups are sometimes methylated or converted to sulfate esters or to ketals formed with pyruvic acid. Structural characteristics of some of the major polysaccharides are listed in Table 4-1.

## 1. Conformations of Polysaccharide Chains

Despite the variety of different monomer units and kinds of linkage present, the conformational possibilities for carbohydrate chains are limited. The sugar ring is a rigid unit and the connection of one unit to the next can be specified by means of two torsion angles $\phi$ and $\psi$ just as with peptides.[61–63] To specify a torsion angle four atoms must be selected–the two at the ends of the bond about which rotation is being considered and two others. There is more than one way to define the zero angle for $\phi$ and $\psi$. As illustrated in the drawing on p. 172, $\phi$ may be taken as the H1–C1–O–C4' dihedral angle and $\psi$ as the H4'–C4'–O–C1 angle. The zero angle is when H1 and H4' are eclipsed. A related alternative is to take $\phi$ and $\psi$ as 0° when the two midplanes of the sugar rings are coplanar.

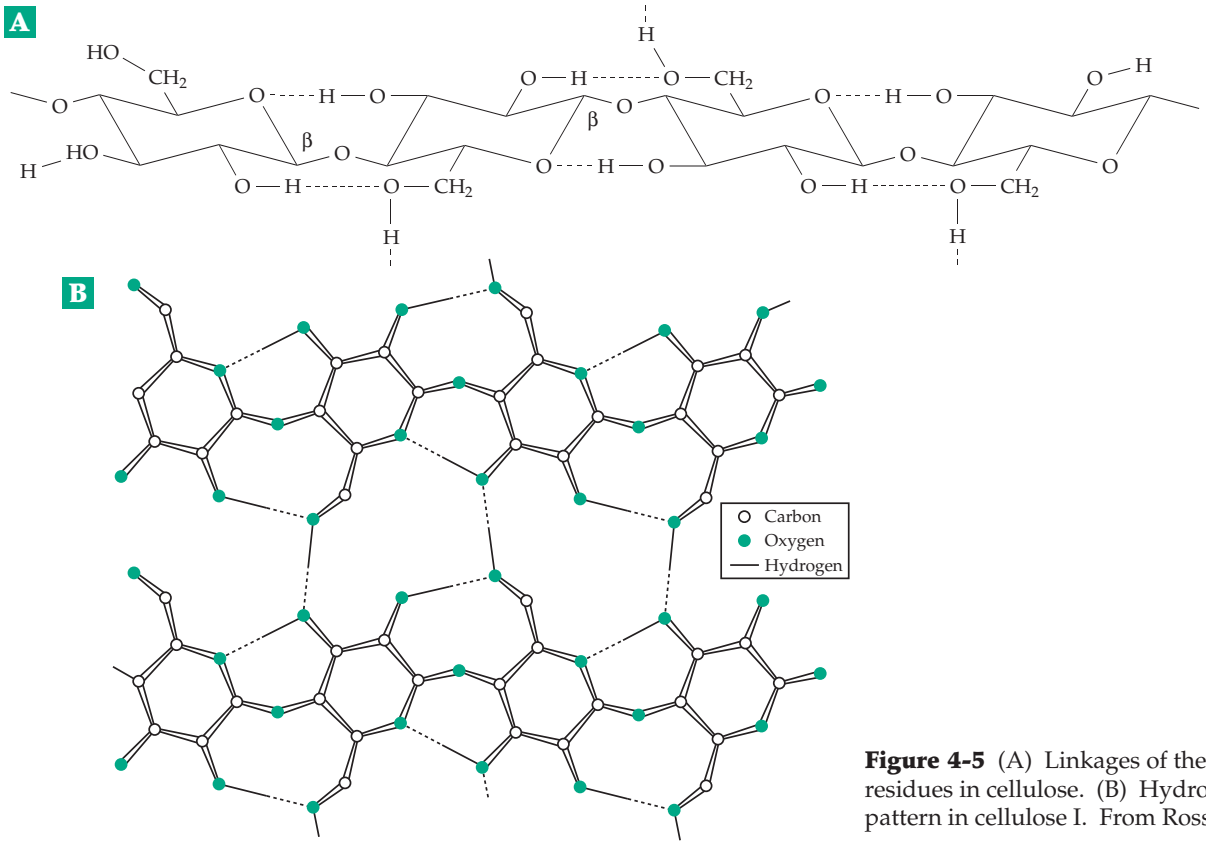
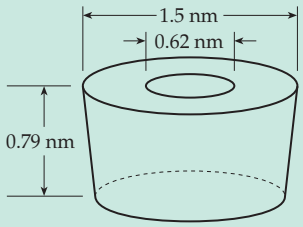



**Figure 4-5** (A) Linkages of the D-glucose residues in cellulose. (B) Hydrogen-bonding pattern in cellulose I. From Ross *et al.*[66]

## BOX 4-A CYCLODEXTRINS

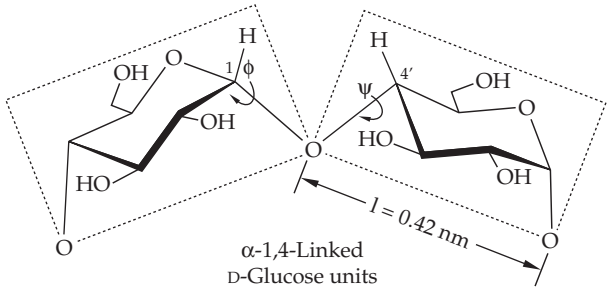
An enzyme produced by some bacteria of the genus *Bacillus* cuts chains of amylose and converts them into tiny rings consisting of six (α-cyclodextrin), seven (β-cyclodextrin), eight (γ-cyclodextrin), or more glucose units. The cyclodextrins, which were first isolated by F. Schardinger in 1903, have intrigued carbohydrate chemists for many years by their unusual properties.[a] They are surprisingly resistant to acid hydrolysis and to attack by amylases and cannot be fermented by yeast.

The torus-like cyclodextrin molecules have an outer polar surface and an inner nonpolar surface.




The small hydrophobic cavities within the cyclodextrins have diameters of 0.50, 0.62, and 0.79 nm, respectively, for the α, β, and γ dextrins.[b] Cavities are potential binding sites for a great variety of both inorganic and organic molecules.[a–g] Complexes of simple alcohols, polyiodides,[c] ferrocene,[d] and many other compounds have been observed. Cyclodextrins can be used to "encapsulate" food additives[b] and their complexes may be useful in separation of enantiomers of drugs.[f,g] They can be chemically modified by adding catalytic groups to serve as enzyme models[h,i] or as color-change indicators sensitive to binding of organic molecules.[j] They can be linked together to form molecular **nanotubes**. Polymer chains can even be threaded through the tubes.[k] What practical applications may yet come from this?


[a] French, D. (1957) *Adv. Carbohydr. Chem. Biochem.* **12**, 189–260
[b] Korpela, T., Mattsson, P., Hellman, J., Paavilainen, S., and Mäkelä, M. (1988–89) *Food Biotechnology* **2**, 199–210
[c] Noltemeyer, M., and Saenger, W. (1980) *J. Am. Chem. Soc.* **102**, 2710–2722
[d] Menger, F. M., and Sherrod, M. J. (1988) *J. Am. Chem. Soc.* **110**, 8606–8611
[e] Hamilton, J. A., and Chen, L. (1988) *J. Am. Chem. Soc.* **110**, 4379–4391
[f] Armstrong, D. W., Ward, T. J., Armstrong, R. D., and Beesley, T. E. (1986) *Science* **232**, 1132–1135
[g] Lipkowitz, K. B., Raghothama, S., and Yang, J. (1992) *J. Am. Chem. Soc.* **114**, 1154–1162
[h] Anslyn, E., and Breslow, R. (1989) *J. Am. Chem. Soc.* **111**, 8931–8932
[i] Granados, A., and de Rossi, R. H. (1995) *J. Am. Chem. Soc.* **117**, 3690–3696
[j] Ueno, A., Kuwabara, T., Nakamura, A., and Toda, F. (1992) *Nature* **356**, 136–137
[k] Harada, A., Kamachi, J. L., and Kamachi, M. (1993) *Nature* **364**, 516–518




The IUB has proposed another convention.[64]

## 2. The Glucans

From glucose alone various organisms synthesize a whole series of polymeric glucans with quite different properties. Of these, **cellulose,** an unbranched β-1,4-linked polyglucose (Fig. 4-5A), is probably the most abundant. It is the primary structural polysaccharide of the cell walls of most green plants.[65] For the whole Earth, plants produce ~$10^{14}$ kg of cellulose per year.

A systematic examination of the possible values for ϕ and ψ shows that for cellulose these angles are constrained to an extremely narrow range which places the monomer units in an almost completely extended conformation.[62] Each glucose unit is flipped over 180° from the previous one. The polymer has a twofold screw axis and there is a slight zigzag in the plane of the rings.[66–68] Remember that in the chair form of glucose all of the –OH groups lie in equatorial positions and are able to form hydrogen bonds with neighboring chains. This feature, together with the rigidity of conformation imposed by the β configuration of the monomer units, doubtless accounts for the ability of cellulose to form strong fibers.

It has been impossible to obtain large single crystals of cellulose. However, from 60 to 90% of native cellulose is thought to be clustered to form the needle-like crystalline microfibrils. These microcrystals of **cellulose I** can be separated from other plant materials by prolonged boiling with dilute NaOH and HCl. Their structure has been established by electron diffraction[67] and by comparison with high-resolution X-ray diffraction structure of the tetrasaccharide β-D-cellotetraose.[66,69] Two closely similar parallel-chained, hydrogen-bonded structures appear to be present. One is shown in Fig. 4-5B. The hydrogen bonds and van der Waals forces bind the chains into sheets which are stacked to form fibers. A typical fiber of plant cellulose has a diameter of 3.5–4 nm and contains 30–40 parallel chains, each made up of 2000–10,000 glucose units. The chain ends probably overlap to form essentially endless fibers that can extend for great distances through the cell wall. They interact with other polysaccharides as is illustrated in Fig. 4-14. A single cotton fiber may be 2–3 cm in length.

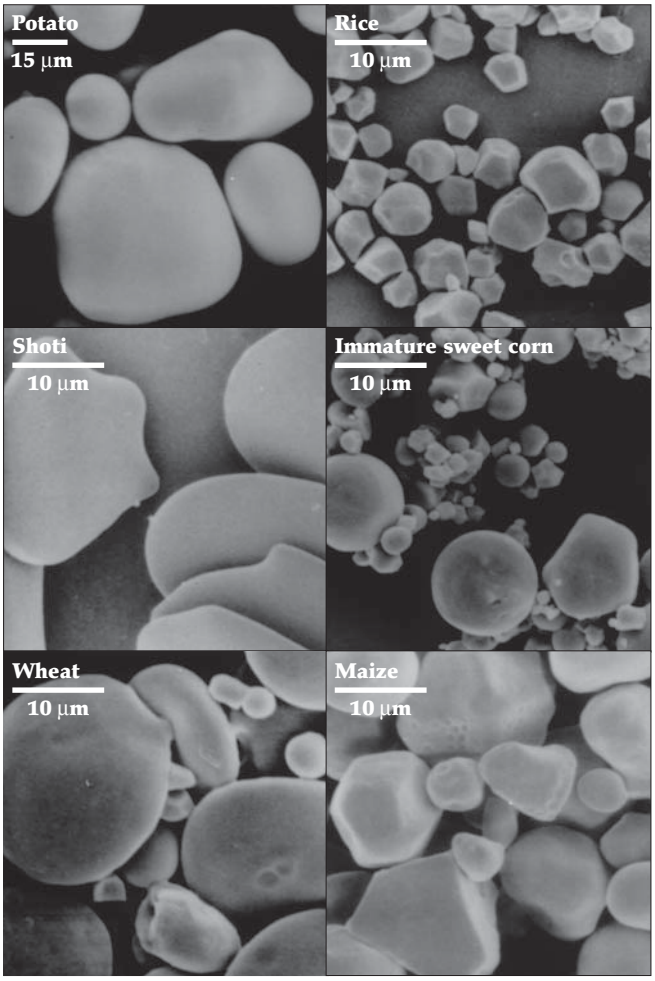



**Figure 4-6** Scanning electron micrographs of starch granules. Magnification 2500× and 3000×. From Jane *et al.*[73]

Cotton thread treated with concentrated NaOH shrinks and has an increased luster. The resulting "Mercerized" cellulose has changed into other crystalline forms. The major one is **cellulose II**, in which the chains in the sheets are antiparallel.[69a] Cellulose II may also occur to some extent in nature. Many other modified celluloses, e.g., **methylcellulose**, in which some –OH groups have been converted to methyl ethers[70] are important commercial products.

**Starch**, another of the most abundant polymers of glucose, is stored by most green plants in a semicrystalline form in numerous small granules. These granules, which are usually formed within colorless membrane-bounded plastids, have characteristic shapes and appearances (Fig. 4-6) that vary from plant to plant. One component of starch, **amylose**, is a linear polymer of many α-D-glucopyranose units in 1,4 linkage (Fig. 4-7) as in maltose. Starch granules always contain a second kind of molecule known as **amylopectin**.[58]

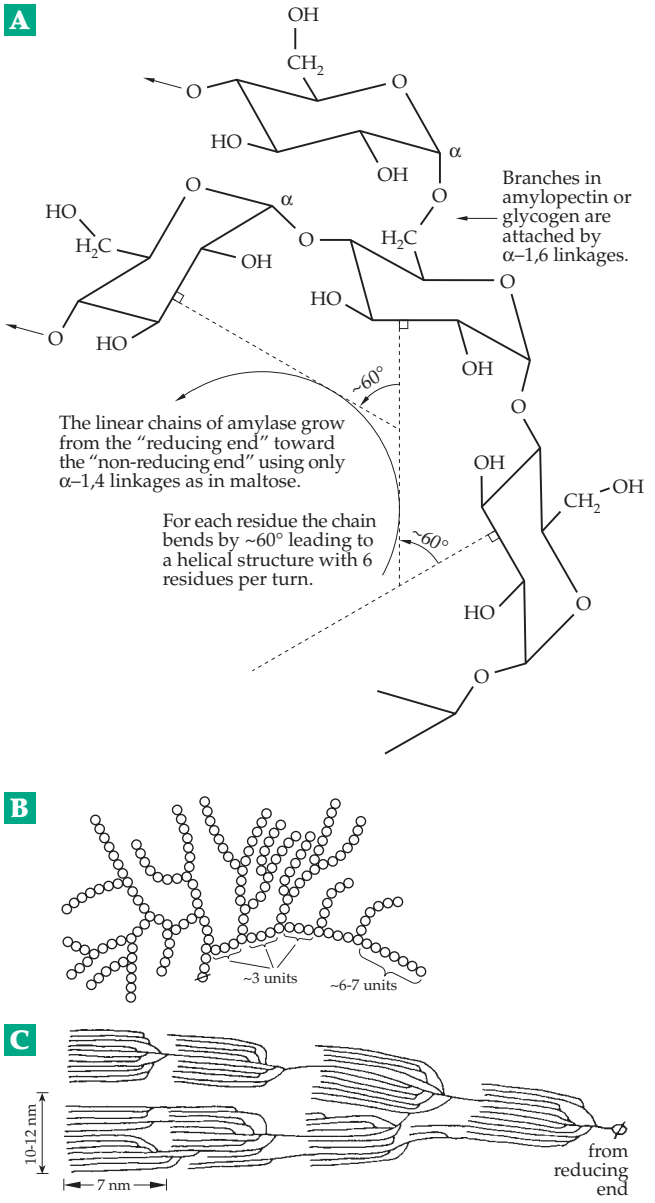


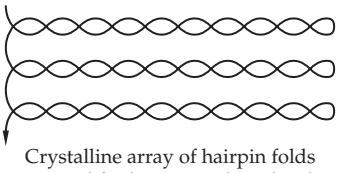

**Figure 4-7** (A) Linkages of the glucose residues in starches and in glycogen. (B) Schematic diagram of the glycogen molecule as proposed originally by K. H. Meyer.[74] The circles represent glucose residues which are connected by α-1,4 linkages and, at the branch points, by α-1,6 linkages. The symbol ϕ designates the reducing group. From D. French.[75] (C) Proposed broomlike clusters in amylopectin. After D. French.[71]

Both amylopectin and **glycogen** (animal starch) consist of highly branched bushlike molecules. Branches are attached to α-1,4-linked chains through α-1,6 linkages (Fig. 4-7). When glucose is being stored as amylopectin or glycogen the many "nonreducing" ends appear to grow like branches on a bush, but when energy is needed the tips are eaten back by enzymatic action (Fig. 11-2). There is only one **reducing end** of a starch molecule where one might expect to find a hemiacetal ring in equilibrium with the free aldehyde. However, this end may be attached to a protein.

Starch granules show a characteristic pattern of growth rings of 0.3- to 0.4-μm thickness with thinner dense layers about 7 nm apart. Study of this layered structure suggests that the individual amylopectin molecules have branches close together in broomlike clusters (Fig. 4-7C).[61,71,72] The amylopectin chains are 120–400 nm long and their relative molecular masses may reach 15–30 million. In addition, starch granules usually contain molecules of the straight-chain amylose, each containing several hundred glucose units and having molecular masses of ~100 kDa. Most starches contain 20–21% amylose but there are special varieties of plants that produce starch with 50–70% amylose. On the other hand, the "waxy" varieties of maize form only amylopectin and lack amylose.

In both starch and glycogen the glucose units of the main chains are linked with α-1,4 linkages. An extended conformation is not possible and the chains tend to undergo helical coiling. One of the first helical structures of a biopolymer to be discovered (in 1943)[76,77] was the left-handed helix of amylose wound around molecules of pentaiodide ($I_5^-$) in the well-known blue starch–iodine complex[78] (Fig. 4-8). The helix contains six residues per turn, with a pitch of 0.8 nm and a diameter of nearly 14 nm. Amylose forms complexes of similar structure with many other small molecules.[79]

Another more tightly coiled double-helical form of amylose has been proposed.[80] Each chain would contain six glucose units per turn and the two chains could be arranged in either parallel or antiparallel directions. The average amylose molecule contains 2000 glucose units and could be stretched to a slender chain over 1 μm long, longer than the crystalline regions observed in starch granules. Thus, the chains within the granules would have to fold back on themselves, possibly in hairpin fashion:



Crystalline array of hairpin folds proposed for linear starch molecules

X-ray diffraction studies support the double-helical structure but suggest a *parallel* orientation of the amylose chains.[81] Since amylose has not been obtained as single crystals the diffraction data do not give a definitive answer. However, if double helices are formed by adjacent branches in amylopectin and glycogen the two strands would be parallel. Starch granules also contain amorphous starch which appears to contain single helices, possibly wrapped around lipid materials.[82]

Glycogen is stored in the cytoplasm of animal cells and to some extent in the lysosomes as enormous 100- to 200-MDa particles. These appear in the electron microscope as aggregates of smaller particles of molecular masses up to 20–40 MDa. A laminated internal structure with surface bumps is suggested by STM microscopy (Fig. 4-9). Biosynthesis of glycogen may be initiated by a 37-kDa protein called **glycogenin**, which remains covalently attached to the reducing end of the glycogen (Chapter 20).[83] Despite the huge molecular masses of glycogen particles, both $^{1}H$ and $^{13}C$ NMR resonances are sharp, indicating a high degree of mobility of the glycosyl units.[84]

**Beta-1,3-linked glucans** occur widely in nature. When a new green plant cell is formed the first polysaccharide to be synthesized is not cellulose but the β-1,3-linked glucose polymer **callose**. Cellulose appears later. Callose is also produced in some specialized plant tissues, such as pollen tubes,[85] and is formed in massive amounts at the site of wounds or of attack by pathogens. The major structural component of the yeast cell wall is a β-1,3-linked glucan with some β-1,6 branches.[86] **Schizophyllan** is a β-1,3-linked glucan with a β-1,6-linked glucosyl group attached to every third residue. A glucan from the coleoptiles of oats contains 30% β-1,3 linkages in a linear chain that otherwise has the structure of cellulose.[87] Other β-1,3-linked glucans serve as energy storage molecules in lower plants and in fungi. Among these are β-1,3-linked glucans such as **paramylon**,[88,89] which is stored by the euglena. A similar polysaccharide, **curdlan**, is formed by certain bacteria.[90]

Some other bacteria, e.g., *Leuconostoc mesenteroides*, make 1,6-linked poly-D-glucose or **dextrans**.[91,92] These always contain some α-1,3-linked branches and may also have α-1,4 and α-1,2 linkages, the structures varying from species to species. Dextrans formed by bacteria such as *Streptococcus mutans* growing on the surfaces of teeth are an important component of dental plaque. Bacterial

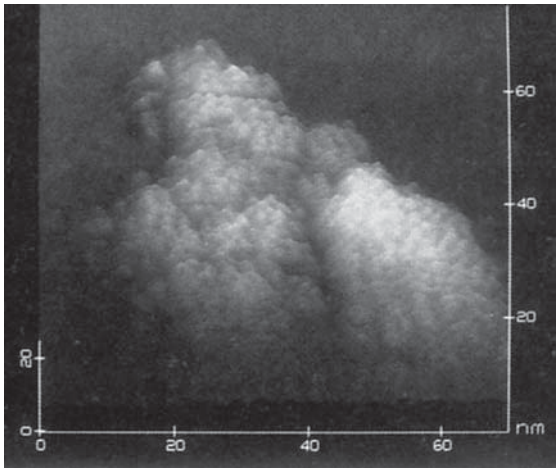



**Figure 4-9** A scanning tunneling microscopic (STM) image of three glycogen molecules. The corrugated surface suggests a laminated structure. These molecules have been purified from tissues by treatment with strong alkali, which breaks the larger aggregates into particles of $M_r = 1-10 \times 10^6$ and diameter 25–30 nm. Courtesy of Fennell Evans.[93]

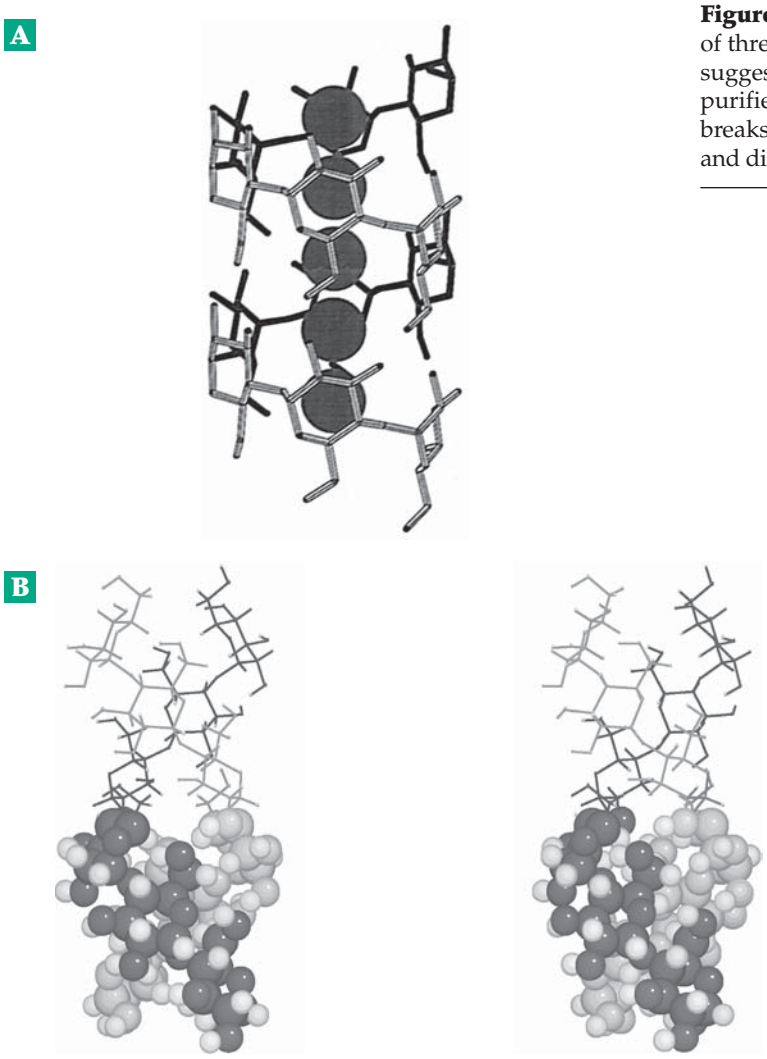



**Figure 4-8** (A) Structure of the helical complex of amylose with $I_3^-$ or $I_5^-$. The iodide complex is located in the interior of the helix having six glucose residues per turn. (B) Model of a parallel-stranded double helix. There are six glucose units per turn of each strand. The repeat period measured from the model is 0.35 nm per glucose unit. Courtesy of Alfred French.

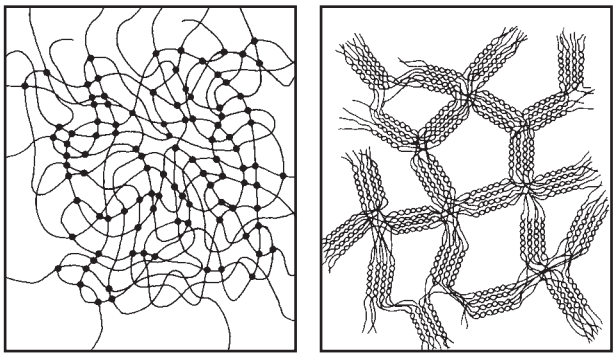


**Figure 4-10** A schematic representation of the gel networks of Sephadex (left) and agarose (right). Note that the aggregates in agarose gels may actually contain 10–$10^4$ helices rather than the smaller numbers shown here. From Arnott *et al*.[95]

dextrans are also produced commercially and are chemically crosslinked to form gels (Sephadex) which are widely used in biochemical separation procedures (Figs. 3-1 and 4-10). A yeast polysaccharide **pullulan** is a regular polymer of maltotriose units with α-1,4 linkages joined in a single chain by α-1,6 linkages.[94] One glucan, called **alternan** is a linear polymer with alternate α-1,6 and 1,3 linkages.

## 3. Other Homopolysaccharides

Cell walls of yeasts contain **mannans** in which the main α-1,6-linked chain carries short branches of one to three mannose units joined in α-1,2, α-1,3, and sometimes α-1,6 linkages.[96] These are covalently linked to proteins.[97] A β-1,4-linked mannan forms microfibrils in the cell wall of some algae such as *Acetabularia* (Fig. 1-11) which do not make cellulose.[98,99] The cell walls of some seaweeds contain a β-1,3-linked **xylan** instead of cellulose. This polysaccharide forms a three-stranded right-handed triple helix.[99,100] Even though xylose is a five-carbon sugar, the polymer contains the thermodynamically more stable six-membered pyranose rings. On the other hand, fructose, a 6-carbon sugar, is present as five-membered furanose rings in **inulin**, the storage polysaccharide of the Jerusalem artichoke and other Compositae, and also in sweet potatoes. The difference has to do with biosynthetic pathways. Furanose rings arise both in inulin and in sucrose because the biosynthesis occurs via the 6-phosphate ester of fructose, making it impossible for the phosphate derivative to form a 6-membered ring.

The major structural polysaccharide in the exoskeletons of arthropods and of other lower animal forms is **chitin**, a linear β-1,4-linked polymer of *N*-acetylglucosamine whose structure resembles that of cellulose. In β chitin the individual parallel chains are linked by hydrogen bonds to form sheets in which parallel chains are held together by NH - - - O = C hydrogen bonds between the carboxamide groups. In the more abundant α chitin the chains in alternate sheets have opposite orientations,[101,102] possibly a result of hairpin folds in the strands. Native chitin exists as microfibrils of 7.25 nm diameter. These contain a 2.8-nm core consisting of 15–30 chitin chains surrounded by a sheath of 27-kDa protein subunits. The microfibrils pack in a hexagonal array, but the structure is not completely regular. Several proteins are present; some of the glucosamine units of the polysaccharide are not acetylated and the chitin core is often calcified.[103] The commercial product **chitosan** is a product of alkaline deacetylation of chitin but it also occurs naturally in some fungi.[102] Chitin is also present in cell walls of yeasts and other fungi. It is covalently bonded to a β-1,3-linked glycan which may, in turn, be linked to a mannoprotein (see Section D,2).[97]

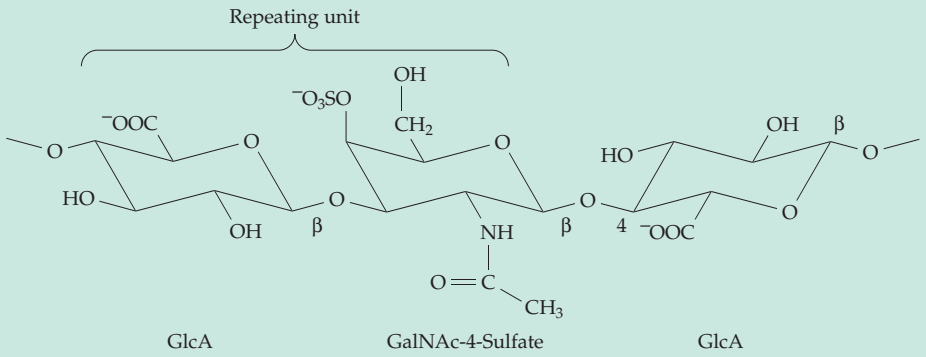
## 4. Heteropolysaccharides of the Animal Body

Many polysaccharides contain repeating units consisting of more than one different kind of monomer.[104,105] Some of these are composed of two sugars in a simple alternating sequence. Examples are **hyaluronan** (hyaluronic acid) and the **chondroitin**, **dermatan**, **keratan**, and **heparan sulfates**. They are important components of the "ground substance" or intracellular cement of connective tissue in animals. Hyaluronan,[106,107] which is abundant in synovial fluid and the vitreous humor of the eye, is a repeating polymer of glucuronic acid and *N*-acetylglucosamine with the structure shown in Fig. 4-11 and $M_r$ of several million. The chondroitin sulfates and dermatan sulfate are similar polymers but with substitution by *N*-acetylgalactosamine and α-L-iduronic acid, respectively, and with sulfate ester groups in the positions indicated in Fig. 4-11.

Hyaluronan solutions are remarkably viscous and at a concentration of only 0.1% can have over 80% of the typical viscosity of biological fluids. This property may result from the presence of hydrogen bonds between the carboxylate, carboxamide, and hydroxyl groups of adjacent sugar residues as in Fig. 4-12. The hydrogen bonds stiffen the chain to give a slender rod. The tetrasaccharide shown in the figure is the repeating unit in a threefold helix.[109] However, the charged molecules do not associate to form strong fibrils like those of cellulose. The chain can be bent easily with breakage of the H-bonds at various positions to give a random coil structure.

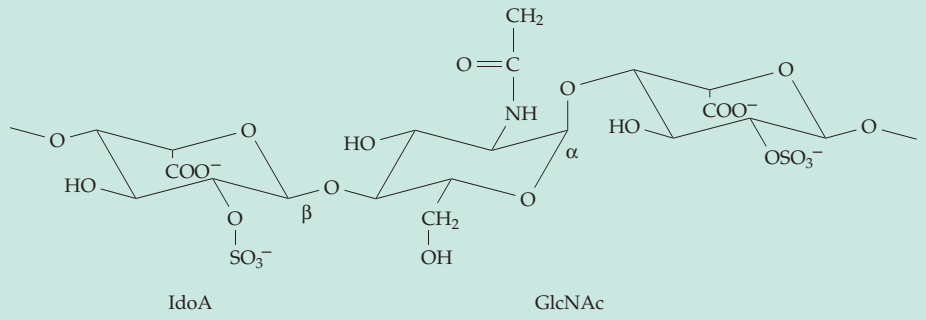
While hyaluronan is not covalently attached to proteins, it is usually anchored to cell surfaces and to

**Type I Chain. Alternating 1→3 and 1→4 linkages; all β, (equatorial).**




| Name | Repeating unit | Variations |
|---|---|---|
| Hyaluronan | $[-4GlcA\beta1\rightarrow3GlcNAc\beta1\rightarrow]_n$ | None. Homogeneous. |
| Chondroitin sulfate | $[-4GlcA\beta1\rightarrow3GalNAc(SO_3^-]\beta1\rightarrow)_n$ | Some GlcA-2-$SO_3^-$, GalNAc-4- or 6-$SO_3^-$or both. |
| Dermatan sulfate | $[-4IdoA\alpha1\rightarrow3GalNAc(4\text{-}SO_3^-]\beta1\rightarrow)_n$ | Some L-IdoA-2-$SO_3^-$. |
| Keratan sulfate | $[-3Gal\beta1\rightarrow4GlcNAc(6\text{-}SO_3^-]\beta1\rightarrow)_n$ | Some Gal-6-$SO_3$, some Sia, Man, Fuc. |

**Type II Chain. All 1→4 linkages, alternating β and α (axial).**




| Name | Repeating unit | Variations |
|---|---|---|
| Heparan sulfate | $[-4\text{-}GlcA\beta1\rightarrow4GlcNAc]_n$ | Some IdoA, sulfation. |
| Heparin | $[-4\text{-L-}IdoA(SO_3^-)\alpha1\rightarrow4GlcN(SO_3^-)\alpha1\rightarrow]_n$ | Some IdoA-2-$SO_3^-$, GlcNAc. 24 different disaccharides possible. |

**Figure 4-11** The repeating disaccharide units of hyaluronan and other glycosaminoglycans. See Fransson[108] and Hardingham and Fosang.[107]

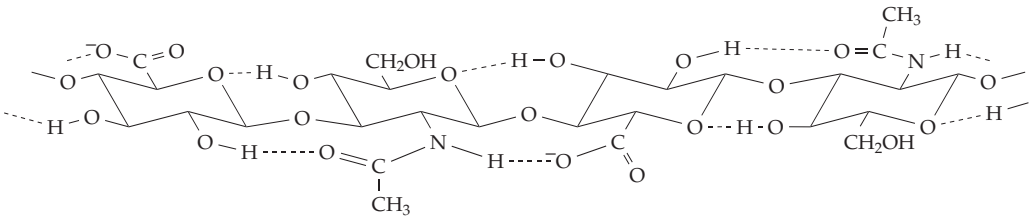


**Figure 4-12** Proposed hydrogen-bonding scheme for the "native" conformation of hyaluronan. See Morris *et al.*[109]

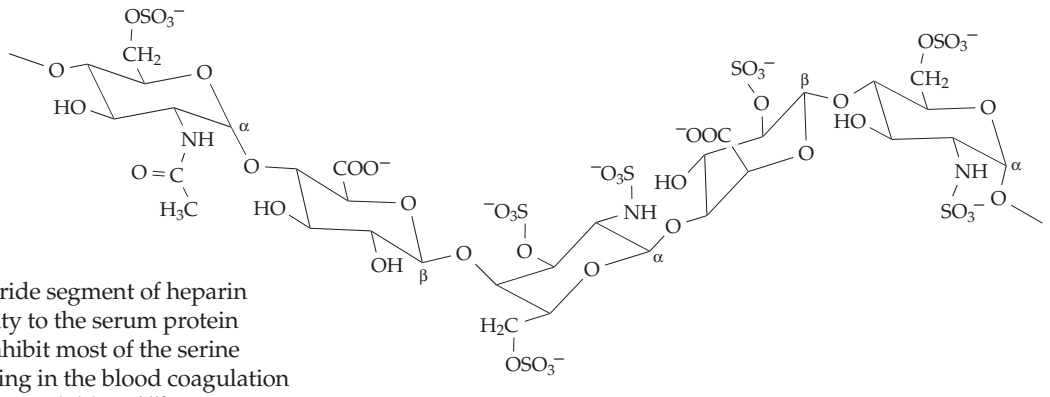


**Figure 4-13** A pentasaccharide segment of heparin which binds with high affinity to the serum protein antithrombin causing it to inhibit most of the serine protease enzymes participating in the blood coagulation process (see Chapter 12). See Lindahl *et al*.[118]

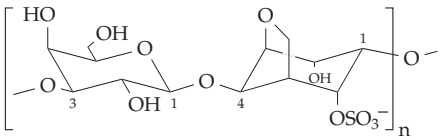
protein "receptors" within the extracellular matrix. In contrast, chondroitin, dermatan, keratan, and heparan sulfates are attached covalently to the proteins at the reducing ends of the polymer chains (see Section D). The attached polymers undergo enzyme-catalyzed chemical alteration. In dermatan most of the glucuronate residues found in chondroitin have been epimerized to iduronate and sulfate groups in ester linkages have been added. Chondroitin sulfate is especially abundant in cartilage; dermatan sulfate is concentrated in skin. Heparan sulfates are more heterogeneous than the other polymers of this group and have been described as "the most complex polysaccharides on the surface of mammalian cells."[110]

In the mast cells of lungs, liver, and other cells heparan sulfate chains are modified further and released into the bloodstream as **heparin**, a material with important anticoagulant properties. Both the amino groups and the 6-hydroxyls of the glucosamine residues of heparin carry sulfate groups. In some units D-glucuronic acid or glucuronate-2-sulfate[111] is present in α-1,4 linkage, but more often L-iduronic acid-2-sulfate is the first unit in the disaccharide.[112–114] The iduronate ring appears to have unusually high conformational flexibility which may influence the biological activity of heparin and related polysaccharides.[115] Because of its anticoagulant property, heparin is an important drug for prevention of blood clot formation.[116] In the United States, in 1976 six metric tons of heparin were used to treat 10 million patients.[117] The anticoagulant activity resides in large part in a nonrepeating pentasaccharide (Fig. 4-13).[112,118] This portion of the heparin, especially if part of a larger octasaccharide,[113] binds to several proteins including the enzyme inhibitors **heparin cofactor II** and **plasma antithrombin III** (Chapter 12). Heparin greatly accelerates the rate at which these proteins bind and inactivate blood clotting factors. See also Section D,1. Lower invertebrates, as well as marine brown algae, contain heavily sulfated fucans which are largely 1,3-linked.[119]

## 5. Plant Heteropolysaccharides

Fibers of cellulose, which run like rods through the amorphous matrix of plant cell walls, appear to be coated with a monolayer of **hemicelluloses**. Predominant among the latter is a **xyloglucan**, which has the basic cellulose structure but with α-1,6-linked xylose units attached to three-fourths of the glucose residues.[104,120,121] L-Fucose may also be present in trisaccharide side chains: L-Fucα1→2Galβ1→ 2Xylα1→.[122] **Pectins** of higher plants contain β-1,4-linked polygalacturonates interrupted by occasional 1,2-linked L-rhamnose residues. Some of the carboxyl groups of these **rhamnogalacturonan** chains are methylated.[123] **Arabinans** and **galactans** are also present in pectin. A possible arrangement of cellulose fibers, hemicelluloses, and pectic materials in a cell wall has been proposed (Fig. 4-14).

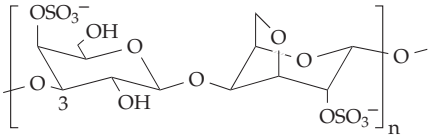
Agarose, an alternating carbohydrate polymer consisting of ~120-kDa chains, is the principal component of agar and the compound that accounts for most of the gelling properties of that remarkable substance. A solid agar gel containing 99.5% water can be formed. Agarose molecules form left-handed double helices with a threefold screw axis, a pitch of 1.90 nm, and a central cavity containing water molecules.[124,125]



Agarose:
[-3-D-Galβ1→4-(3,6-anhydro)-L-Galα1→]$_n$

A similar structure has been established for the gel-forming **carrageenans** from red seaweed. The X-ray data suggest that three of the disaccharide units form one turn of a right-handed helix with a pitch of 2.6 nm. A second chain with a parallel orientation, but displaced by half a turn, wraps around the first helix.[124] Such

double-helical regions provide "tie points" for the formation of gels (Fig. 4-10).[104,127,128]

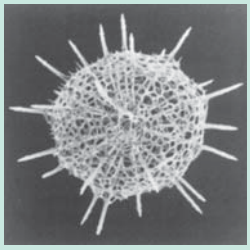



ι-Carrageenan:
[3Gal (4 sulfate) β1→4 (3,6-anhydro) Gal (2-sulfate) α1→]$_n$

Sulfate groups protrude from the structure in pairs and provide binding sites for calcium ions, which stabilize the gel. The presence of occasional extra sulfate groups in these polymers causes kinks in the chains because the derivatized pyranose rings reverse their conformation to the other chair form. This prevents the entire polysaccharide chain from assuming a regular helical structure.[100,104]

**Alginates**, found in cell walls of some marine algae and also formed by certain bacteria, consist in part of a linear β-1,4-linked polymer of D-mannuronate with a cellulose-like structure. Alginates also contain

## BOX 4-B SILICON: AN ESSENTIAL TRACE ELEMENT

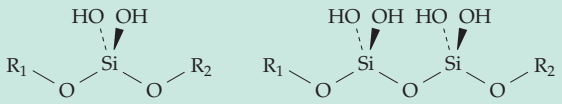


Silica skeleton of a radiolarian[e]

No one can doubt that diatoms, which make their skeletons from $SiO_2$, have an active metabolism of silicon. They can accumulate as much as 0.7 μM soluble silicon, possibly attached to proteins.[a-d]

The radiolaria and sponges often accumulate silicon and limpets make opal base plates for their teeth. Silicon may account for as much as 4% of the solids of certain grasses. Although silicon is usually not considered an essential nutrient for all plants, there is much evidence that it is essential to some and that it is often beneficial.[f] Silicon is found in soil primarily as silicic acid, $H_4SiO_4$, whose concentration ranges from 0.1–0.6 mM. Most of the silicon taken up by plants is deposited within cells, in cell walls, between cells, or in external layers as hydrated $SiO_2$. Presumably the organic components of the plant control the deposition. The $SiO_2$ in some plants takes the form of sharp particles which may have a defensive function. They abrade the enamel surfaces of the teeth of herbivores and can cause other illnesses.[g]

Silicon is essential for growth and development of higher animals,[h-l] and it has been suggested that humans may require 5–20 mg per day.[m] In the chick, silicon is found in active calcification sites of young bone.[i] Silicon-deficient animals have poorly calcified bones and also an elevated aluminum content in their brains.[m] Silicon is present in low amounts in the internal organs of mammals but makes up ~0.01% of the skin, cartilage, and ligaments, in which it is apparently bound to proteoglycans such as chondroitin-4-sulfate, dermatan sulfate, and heparan sulfate (Fig. 4-11).[m,n] These polymers contain ~0.04% silicon or one atom of silicon per 130–280 repeating units of the polysaccharides. Plant pectins contain about five times this amount. The silicon is apparently bound tightly in *ether* linkage. Perhaps orthosilicic acid, $Si(OH)_4$, reacts with hydroxyl groups of the carbohydrates to form bridges between two chains as follows:

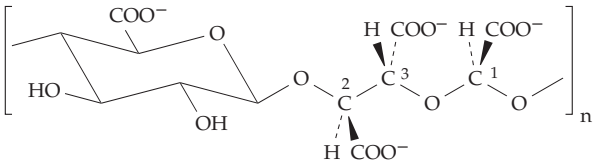



In each of these formulas additional free OH groups are available on the silicon so that it is possible to crosslink more than two polysaccharide chains. Silicon may function as a biological crosslinking agent in connective tissue. **Silaffins**, small polypeptides containing polyamine side chains of modified lysine residues, apparently initiate silica formation from silicic acid in diatoms.[o]


[a] Robinson, D. H., and Sullivan, C. W. (1987) *Trends Biochem. Sci.* **12**, 151–154
[b] Round, F. E. (1981) in *Silicon and Siliceous Structures in Biological Systems* (Simpson, T. L., and Volcani, B. E., eds), pp. 97–128, Springer, New York
[c] Evered, D., and O'Connor, M. (1986) *Silicon Biochemistry*, Wiley, New York
[d] Round, F. E., Crawford, R. M., and Mann, D. G. (1990) *The Diatoms*, Cambridge Univ. Press, Cambridge UK
[e] Buchsbaum, R., Buchsbaum, M., Pearse, J., and Pearse, V. (1987) *Animals Without Backbones*, 3rd ed., Univ. Chicago Press, Chicago
[f] Epstein, E. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11–17
[g] McNaughton, S. J., and Tarrants, J. L. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 790–791
[h] Schwarz, K. (1970) in *Trace Element Metabolism in Animals* (Mills, F., ed), pp. 25–38, Livingstone, Edinburgh, UK
[i] Schwarz, K., and Milne, D. B. (1972) *Nature* **239**, 333–334
[j] Carlisle, E. M. (1972) *Science* **278**, 619–621
[k] Hoekstra, W. H., Suttie, J. W., Ganther, H. E., and Mertz, W., eds. (1974) *Trace Element Metabolism in Animals-2*, University Park Press, Baltimore, Maryland
[l] Carlisle, E. M. (1988) *Science Total Environment* **73**, 95–106
[m] Nielsen, F. H. (1991) *FASEB J.* **5**, 2661–2667
[n] Schwarz, K. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1608–1612
[o] Kröger, N., Deutzmann, R., and Sumper, M. (1999) *Science* **286**, 1129–1132

α-L-guluronate, sometimes in homopolymeric "blocks" and sometimes alternating with the mannuronate residues. Groups of adjacent guluronate units are thought to impart calcium-binding properties to alginates.[129]

Polysaccharides with calcium-binding sites may also serve to initiate deposition of calcium carbonate. For example, the unicellular alga *Pleurochrysis carterae* contains an unusual polysaccharide with the following highly negatively charged repeating unit:




[-4GlcAβ1→2-*meso*-tartrate-3→1-glyoxylate-1→]$_n$

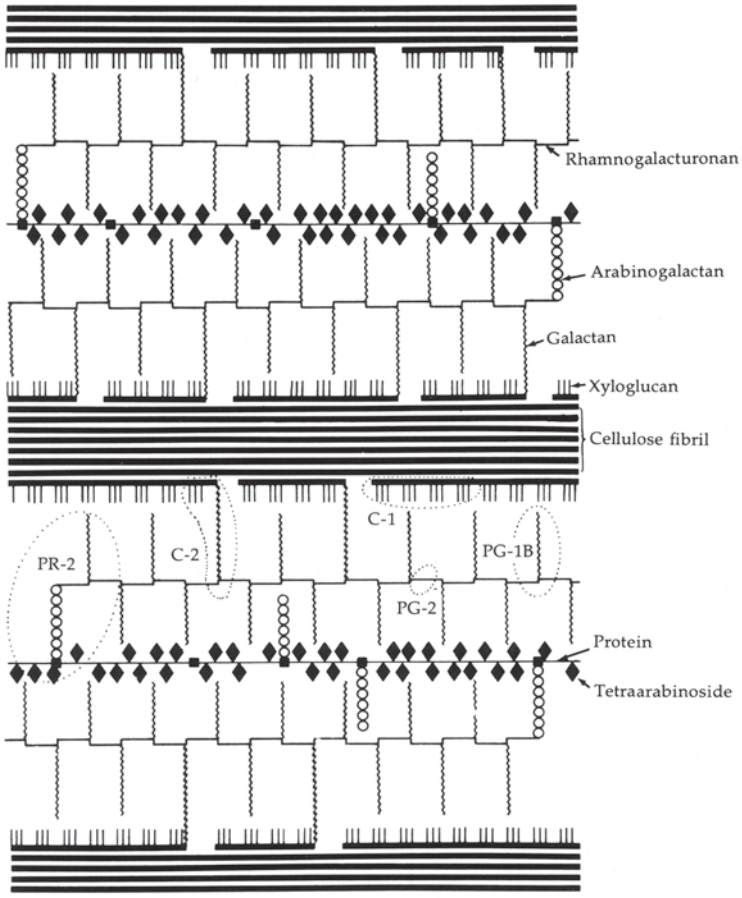


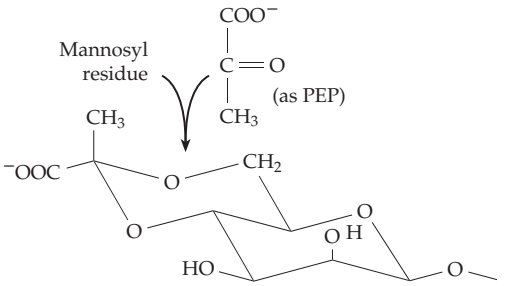

**Figure 4-14** Tentative structure of the walls of suspension-cultured sycamore cells. The wall components are in approximately proper proportions but the distance between cellulose elementary fibrils is expanded to allow room to present the interconnecting structure. There are probably between 10 and 100 cellulose elementary fibrils across a single primary cell wall. From Albersheim *et al.*[126]

*Meso*-tartrate is joined to glucuronic acid in glycosidic linkage and by acetal formation to the aldehyde **glyoxylic acid**, HOOC—CHO, which is also joined in an ether linkage to the next repeating unit.[130] Similar **open acetal linkages** join monosaccharide units in some bacterial lipopolysaccharides and may occur more widely.[130a]

Bacteria form and secrete a variety of heteropolysaccharides, several of which are of commercial value because of their useful gelling properties. **Xanthan gum** (formed by *Xanthomonas campestris*) has the basic cellulose structure but every second glucose residue carries an α-1,3-linked trisaccharide consisting of 6-*O*-acetylmannose, glucuronic acid, and mannose in the following repeating unit:[131,132]

(-4Glcβ1-4Glcβ1-)$_n$
3
↑
6-*O*-Acetyl-Manα1
2
↑
Manβ1-4GlcAβ1

The polymer is further modified by reaction of about half of the mannosyl residues with pyruvate to form ketals (Eq. 4-9).




(4-9)

**Acetan** of *Acetobacter xylinum* has pentasaccharide side chains that contain L-rhamnose.[122] A helical structure for the strands has been observed by atomic force microscopy.

## 6. Polysaccharides of Bacterial Surfaces

The complex structure of bacterial cell walls is discussed in Chapter 8. However, it is appropriate to mention a few bacterial polysaccharides here. The innermost layer of bacterial cell walls is a porous network of a highly crosslinked material known as **peptidoglycan** or **murein** (see Fig. 8-29). The backbone of the peptidoglycan is a β-1,4-linked

alternating polymer of *N*-acetyl-D-glucosamine and *N*-acetyl-D-muramic acid. Alternate units of the resulting chitin-like molecule carry unusual peptides that are attached to the lactyl groups of the *N*-acetyl-muramic acid units and crosslink the polysaccharide chains. In *E. coli* and other gram-negative bacteria the peptidoglycan forms a thin (2 nm) continuous network around the cell. This "baglike molecule" protects the organism from osmotic stress. In addition, gram-negative bacteria have an outer membrane and on its outer surface a complex lipopolysaccharide. The projecting ends of the lipopolysaccharide molecules consist of long carbohydrate chains with repeating units that have antigenic properties and are called **O antigens**. Specific antibodies can be prepared against these polysaccharides, and so varied are the structures that 1000 different "serotypes" of *Salmonella* are known. These are classified into 17 principal groups. For example, group E3 contains the following repeating unit, where *n* may be ~50 on the average. Rha = L-rhamnose.

$$\begin{array}{l} \text{Glc}\alpha 1 \\ \quad\downarrow \\ \quad\ 4 \\ \quad\ (\text{Gal}\beta 1\rightarrow 6\text{Man}1\rightarrow 4\text{Rha}\rightarrow)_n \end{array}$$

Polysaccharides of groups A, B, and D contain the repeating unit

$$\begin{array}{l} \quad\ \text{X} \\ \quad\downarrow \\ (\text{Man}\rightarrow\text{Rha}\rightarrow\text{Gal}\rightarrow)_n \end{array}$$

where X is a 3,6-dideoxyhexose: **paratose** in type A, **abequose** in type B, and **tyvelose** in type D (Fig. 4-15). The existence of the many serotypes depends on the variety of components, on the many types of linkage (α and β, 1→2, 1→3, 1→4, and 1→6) in the repeating units, and on further structural variations at the chain ends.

At the inner end of the O antigen is a shorter polysaccharide "core" whose structure is less varied than that of the outer ends but which is remarkable in containing two sugars found only in bacterial cell walls: a seven-carbon heptose and an eight carbon α-oxo sugar acid, **ketodeoxyoctonate** (KDO). The structures are given in Fig. 4-15 and the arrangement of these sugars in the *Salmonella* lipopolysaccharide is shown in Fig. 8-30. That figure also shows the manner in which the oligosaccharide that bears the O antigen is attached to a lipid anchoring group that is embedded in the outer membrane of the bacteria.

A great variety of polysaccharides are present in the outer layers of other types of bacteria. For example, the mycobacteria have an alternating 5- and 6-linked β-D-Gal*f* polymer attached to their peptidoglycan.

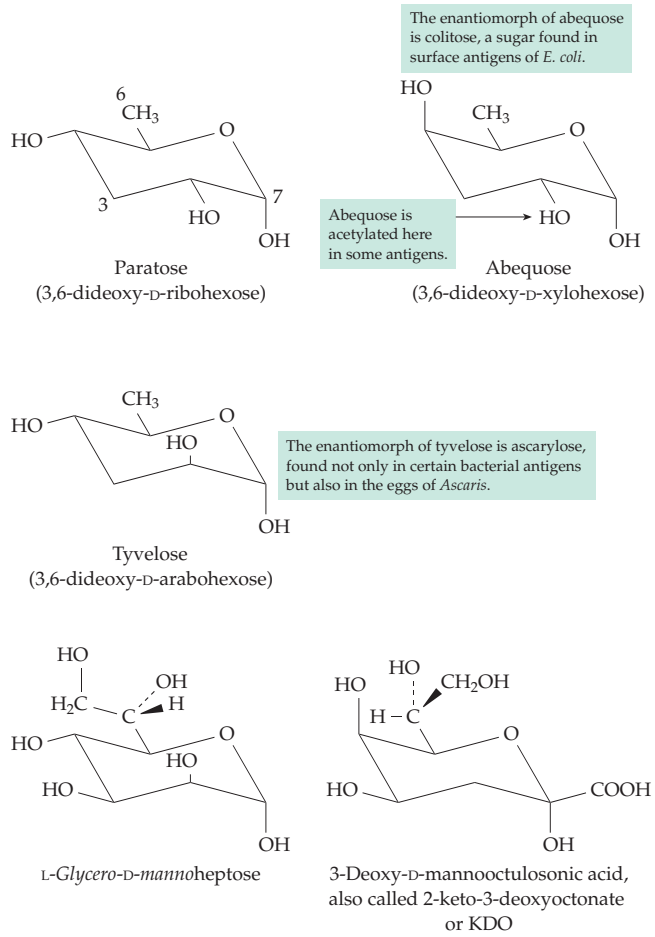



**Figure 4-15** Structures of special sugars found in the "antigens" of the outer cell walls of gram-negative bacteria.

Attached to this galactan are branched penta-D-arabinose units:

$$(\text{Ara}f\beta 1\rightarrow 2\text{Ara}f\alpha 1)_2\rightarrow 3,5\text{Ara}f\alpha\text{—}$$

These are further modified by esterification with mycolic acids.[133] Information on some other cell wall components is given in Chapter 20.

## D. Glycoproteins and Proteoglycans

Many proteins, including almost all of those that are secreted from cells and many that are components of cell surfaces, carry covalently attached oligosaccharides.[134–139] These **glycoproteins** may carry just one or a few, often highly branched, oligosaccharide chains. For example, ribonuclease B has a structure identical to that of ribonuclease A (Fig. 12-25) except for the presence of an oligosaccharide on asparagine 34.[140] In other instances proteins carry a large number of

such chains and the carbohydrate may account for over half of the mass of a glycoprotein. Most carbohydrate chains are attached either as *O*-glycosides with the hydroxyl groups of the side chains of serine, threonine, or other hydroxyamino acid residues or as *N*-glycosyl groups through linkage to the amide groups of asparagine side chains. Both types of linkage may be present in a single protein. Here are some examples.

| | |
|---|---|
| Xylβ1→*O*-Ser(Thr) | Proteoglycans of connective tissue; thyroglobulins |
| →N(ε)—Lys (H) | Some dermatan sulfates |
| Galβ1→*O*-Hydroxylysine (Hydroxyproline) | Collagen, extension |
| L-Araα1→*O*-4-Hydroxy-proline | Plants |
| GalNAcα1→*O*-Ser(Thr) | Many glycoproteins |
| GlcNAcα1→*O*-Ser(Thr) | Glycoproteins of cytoplasmic surfaces |
| GlcNAcβ1→NH-β$CH_2$-Asn | Many glycoproteins<br>Some dermatan and heparan sulfates |

Linkage of a glycosyl group to a carbon atom of an indole ring of tryptophan has also been demonstrated.[141]

## 1. *O*-Linked Glycosyl Groups

In the *O*-linked glycoproteins the sugar that is attached directly to the protein is usually either xylose, galactose, or *N*-acetylgalactosamine, all in the pyranose ring form. Xylose is found only in the intercellular **proteoglycans** which carry the chondroitin, dermatan, and related sulfated polysaccharide chains of connective tissues.[141a,b] Since amino sugars are a major constituent, proteoglycans are often called **glycosaminoglycans**. Chondroitin, dermatan, and heparan sulfates are all attached to "core" proteins by the same linkage, which is illustrated here for chondroitin

Repeating unit of chondroitin as in Fig. 4-11 — Terminal unit

$(\rightarrow 4GlcA\beta1\rightarrow 3GalNA\beta1)_n\rightarrow 4GlcA\beta1$
↓
3Galβ1
↓
3Galβ1
↓
4Xylβ1→*O*-Ser

Some IdoA may be present in the terminal unit of dermatan sulfate.[142] Keratan sulfate has its own core proteins[143] and has different terminal units including the following:[107]

Keratan→Gal→GlcNAc ↘
GalNAc→*O*–Ser
Sia→Gal ↗

The large proteoglycan of human cartilage is built upon the 246 kDa protein **aggrecan**. In the central half of the peptide chain are many Ser-Gly sequences to which about one hundred 10- to 25-kDa chondroitin sulfate chains are attached. About 30 keratan sulfate chains as well as other oligosaccharide groups are also present. These proteoglycan subunits are joined with the aid of 44- to 49-kDa **link protein** to molecules of hyaluronan[144–146] (Fig. 4-16). Several types and sizes of proteoglycan are known.[143,145–147] Dermatan sulfates may be linked to these through either serine or asparagine, depending upon the tissue. The polysaccharide chains of the proteoglycans also bind to collagen fibrils to form a "fiber-reinforced composite material" between cells. Chondroitin and heparan sulfates may be attached at different Ser-Gly sites in a single peptide chain.[148] Degradation of heparan proteoglycans may lead to the shorter free carbohydrate chains found in the circulating heparin. Commercial heparin preparations used as anticoagulants are produced by oxidative destruction of the attached proteins.[113]

A quite different situation holds for **collagen** in which β-galactosyl units and glucosyl-β-galactosyl disaccharide units are attached to side chains of hydroxylysine formed by postsynthetic modification of the original procollagen chain.

Glcα1→2Galβ1→O—CH(–$(CH_2)_2$–$C_\alpha$–$CH_2$–$NH_3^+$) } Side chain of *erythro*-δ-hydroxylysine

A great deal of variation in the amount of glycosylation is observed from one species to another. The human α-1(II) chains of collagen usually carry four disaccharides and four monosaccharide units. In the related collagen-like **extensins**, which are found in plant cell walls, the hydroxyproline (Hyp) side chains are *O*-glycosylated, largely by short oligosaccharides of arabinose in furanose ring form,[150–152] e.g., Ara*f* β1→2Ara*f* β1→2Ara*f* β1→4Hyp. There are as many as 25 repeats of Ser-Pro-Pro-Pro-Pro encoded in an extensin gene. Most of the prolines are hydroxylated and glycosylated. The presence of two or more contiguous proline residues seems to be the signal for the hydroxylation reaction to take place.[152]

Another distinct family of *O*-linked glycoproteins are the **mucins**, which are present in saliva and other

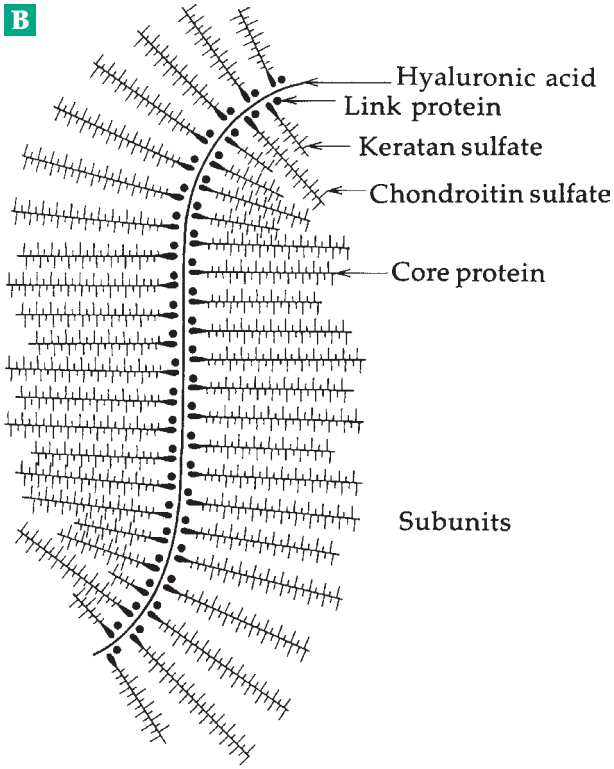



**Figure 4-16** (A) Dark field electron micrograph of a proteoglycan aggregate from bovine articular cartilage (from bearing surfaces of joints). Courtesy of Joseph A. Buckwalter. The filamentous backbone consists of hyaluronic acid, as in (B). The proteoglycan subunits extend from the backbone. From Rosenberg.[149]

mucous secretions. The polypeptides consist largely of serine, threonine, proline, and glycine, with up to one-third of the residues being Ser + Thr. Mucins may contain 70 to 85% carbohydrate linked through *N*-acetylgalactosamine residues to the serine and threonine hydroxyls.[153–160a] Salivary mucins of cows, sheep, and dogs contain largely the following disaccharide, which may contain either *N*-acetyl- or *N*-glycolyl-neuraminic acid:

NeuNAcα2→6GlcNAcβ→Ser(Thr)

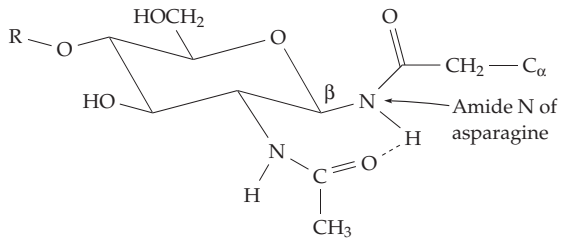
Up to 800 disaccharides are present on a single large protein which is composed of disulfide-linked subunits. The many negatively charged sialic acid (*N*-glycolyl-neuraminic acid) groups are thought to cause expansion and rigidity which increase the viscosity of the protein. Some other mucins contain predominantly large oligosaccharides, which often carry **blood group determinants**.[161,162]

The ABO(H) family of blood group determinants (Box 4-C) are oligosaccharide groups assembled from D-galactose, *N*-acetyl-D-galactosamine, and L-fucose. They are carried on the nonreducing ends of *O*-linked oligosaccharides which may be attached to cell surface proteins, to mucins, or to the sphingolipid known as ceramide (Chapter 8). Attachment to proteins is usually via GalNAc. An example of a blood type A determinant on an *O*-linked oligosaccharide is the following, where the linkage between the galactose and *N*-acetylglucosamine may be either β-1,3 or β-1,4:

GalNAcα1→3Galβ1→3(4)GlcNAcβ1→3GalNAc→*O*-Ser(Thr)
2
↑
Fucα1

## 2. Asparagine-Linked Oligosaccharides

In many glycoproteins oligosaccharides are linked through *N*-acetylglucosamine to the side chain of asparagine.[134,163,164]




This structure also illustrates one of the hydrogen-bonding possibilities through which the sugar can

interact with the protein. Since the site of glycosylation is often at β bends in the surface of the protein, the amide groups of the *N*-acetylglucosamine may alternatively hydrogen bond to amide groups of the peptide.[165] This asparagine linkage is very common. For example, it is present in 97% of glycoproteins of blood plasma[166] and it is also predominent in the glycoproteins of tissue surfaces.[167] There are numerous structures for asparagine-linked oligosaccharides but many contain the following core to which additional glycosyl groups may be attached:

```
        Y3α1
          ↓
          6                                  ±Fucα1
Y2α1→3Manα1                                    ↓
          ↘6                                   6
           Manβ1→4GlcNAβ1→4GlcNAβ1→Asn
          ↗3
Y1α1→2Manα1
```

Here the ±Fuc indicates that this residue is present only on some of the chains.

Notice the three mannose residues on the left side. In the **high mannose type** oligosaccharide, $Y_1$, $Y_2$, and $Y_3$ are additional mannose units. In many instances $Y_1 = Y_2 = Y_3$ = Manα→Man. These "extra" mannose units are put onto the oligosaccharide during the original biosynthesis and before it is attached to the protein (Chapter 20). Some of the mannose units may then be removed during the "processing" of the oligosaccharide in the endoplasmic reticulum and residues of glucosamine, galactose, and sialic acid (Sia) may be added. Thus, $Y_1$ and $Y_2$ in the foregoing structure often become

Siaα2→3(6)Galβ1→4GlcNAcβ1→

Here, the sialic acid may be linked either 2,3 or 2,6 and the GlcNAc either 1,2 or 1,3. Both Y groups may consist of trisaccharides of this type in "biantennary" oligosaccharides and a third trisaccharide ($Y_3$) may be added to form a "triantennary" molecule. An additional *N*-acetylglucosamine is often linked by β-1,4 linkage to the central mannose of the core and fucosyl residues in α-1,6 linkage are often linked to the *N*-acetylglucosamine next to the asparagine.[168] The Galβ1→4GlcNAc disaccharide unit in the above Y group is also called ***N*-acetyllactosamine** because of its relationship to lactose. It is often repeated in long Y groups, e.g., as $(\rightarrow 3Gal\beta1\rightarrow 4Glc\text{-}NAc\beta1\text{—})_n$, the oligosaccharides being called poly-*N*-acetyllactosamino-glycans. These structures are also principal carriers of the ABO blood type determinants on erythrocyte surfaces.[169]

Many *N*-linked oligosaccharides are highly branched. For example, in ovomucoid, a protease inhibitor of hen eggs, "pentaantennary" oligosaccharides have two and three *N*-acetylglucosamine rings, respectively, attached to the terminal mannose units of the oligosaccharide core in 1,2, 1,4 and 1,6 linkages. Another large

```
                    ±NeuAcα2→3Galβ1      ±Fucα1
                                ↘4         ↘3
±NeuAcα2→3Galβ1→4Galβ1 (→4GlcNAcβ1→3Galβ1)p →4GlcNAcβ1
                  ↗3          ↗3                        \
        ±NeuAcα2      ±Fucα1                             \
                                                          \
                    ±NeuAcα2→3Galβ1      ±Fucα1            6
                                ↘4         ↘3
±NeuAcα2→3Galβ1→4Galβ1 (→4GlcNAcβ1→3Galβ1)q →4GlcNAcβ1→4 Manα1                    Asp
                  ↗3          ↗3                        2                         |
        ±NeuAcα2      ±Fucα1                                                      Ala
                                                                                  |
                                                                                  Ala
                    ±NeuAcα2→3Galβ1      ±Fucα1                                   |
                                ↘4         ↘3                                     Ser
±NeuAcα2→3Galβ1→4Galβ1 (→4GlcNAcβ1→3Galβ1)r →4GlcNAcβ1      6                      |
                  ↗3          ↗3                        3 Manβ1→4GlcNAβ1→4GlcNAβ1→Asn
        ±NeuAcα2      ±Fucα1                                                      |
                                                                                  Gln
                    ±NeuAcα2→3Galβ1      ±Fucα1                                   |
                                ↘4         ↘3                                     Thr
±NeuAcα2→3Galβ1→4Galβ1 (→4GlcNAcβ1→3Galβ1)s →4GlcNAcβ1                            |
                  ↗3          ↗3                                                 Val
        ±NeuAcα2      ±Fucα1                                                      |
                                                                                  Ser
                                                  4
                                                  2 Manα1
                    ±NeuAcα2→3Galβ1      ±Fucα1
                                ↘4         ↘3
±NeuAcα2→3Galβ1→4Galβ1 (→4GlcNAcβ1→3Galβ1)t →4GlcNAcβ1
                  ↗3          ↗3
        ±NeuAcα2      ±Fucα1
```

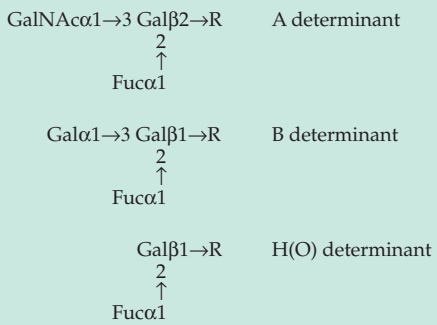
**Figure 4-17** A large *N*-linked glycan from the fertilized egg of the teleost fish *oryzias*.[171] On the average, only two of the "antenna" contain the repeated unit in parenthesis, i.e., p + q + r + s + t = 1.9; an average of about 4.6 residues of sialic acid are present.

## BOX 4-C THE BLOOD GROUP DETERMINANTS

The role of carbohydrates in biological communication is well illustrated by the human blood types.[a,b] According to the ABO system first described by Landsteiner in 1900, individuals are classified into types A, B, AB, and O. Blood of individuals of the same type can be mixed without clumping of cells, but serum from a type O individual contains antibodies that agglutinate erythrocytes of persons of types A and B. Serum of persons of type B causes type A cells to clump and vice versa. Individuals of none of the four types have antibodies against type O erythrocytes. For this reason, persons with type O blood are sometimes inaccurately described as "universal donors."

The ABO blood types are determined by specific **blood group determinants** which are attached to the nonreducing ends of *O*-linked oligosaccharides of surface glycoproteins, mucins, glycolipids, and, to a lesser extent, *N*-linked oligosaccharides. The blood group determinants are found on erythrocytes and all endothelial cells of the body. In about 80% of the population they are also present on glycoproteins of the saliva and other secretions.

The minimal determinant structures, attached to "carrier" R, are as follows:

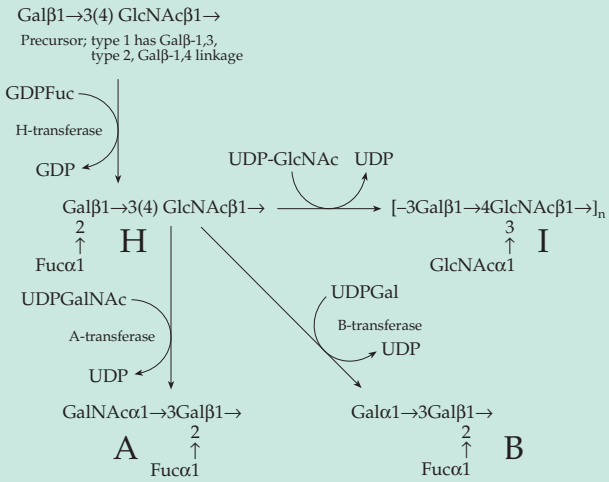



Here R refers to a "carrier oligosaccharide" which can be as simple as

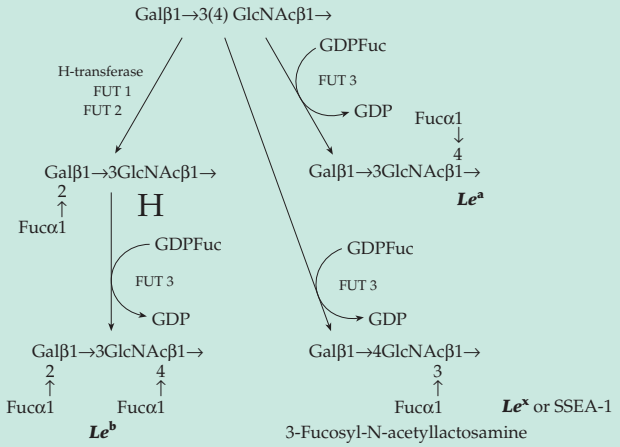
→3(4)GlcNAcβ1→3Galβ1→4Glcβ1—*O*—Ser (Thr, ceramide)

or may be a complex lactosaminoglycan or an oligosaccharide such as that in Fig. 4-17. The minimal determinants can be linked to the carrier oligosaccharide by either a β-1,3 (type I chain) or β-1,4 (type II chain) glycosidic bond.

The genetic basis for the ABO blood groups is well understood. There are three **alleles**, variants of a gene, that encodes a **glycosyltransferase**. In A type individuals, this enzyme transfers *N*-acetylgalactosamine from a carrier molecule, called UDP (Chapter 17), onto the terminal positions of the H(O) determinant. The enzyme specified by the *B* allele transfers galactose. The two enzymes differ in only four amino acid residues but the result is an altered substrate specificity.[a,c,d] The *O* allele produces inactive enzyme as a result of a single base deletion in the gene.[a] The H gene has been identified as that of an α-1,2 fucosyltransferase that transfers α-L-fucose from the carrier GDP to the galactose unit in the foregoing structure.[e] Persons with an inactive *H* transferase gene may have the rare type I, which results from addition of glucosamine branches to the repetitive H antigen structures of polylactosaminoglycans.[f] More often though, the H antigen




is acted upon by a different fucosyltransferase encoded by the *Le* gene, which determines the **Lewis blood group**. This enzyme places α-1,4-linked L-fucose onto the H antigen to give the *Le*[b] antigen. The same fucosyltransferase (which is



**BOX 4-C Continued**

designated FUT 3) acting on type 2 precursor chains forms the *Le*[a] antigen. The same transferase adds a β-1,3-linked fucose to type 2 chains to give the *Le*[x] antigen, which is also called SSEA-1.[g,h] Lactosamine-type precursors carrying α-2,3-linked sialic acid at their nonreducing ends on surfaces of granulocytes, monocytes, and natural killer cells can be converted to sialylated *Le*[x] (s-*Le*[x]).[i,j]

```
NeuAcα2→3Galβ1→4GlcNAc→
          2
          ↑              s-Lex
       Fucα1
```

Individuals with active gene *Se* (for secretion) secrete glycoproteins bearing the blood group substance into saliva and other body fluids. The *Se* gene is another H-transferase present in epithelial cells and salivary glands. In individuals with active *Se* genes (~80% of most populations) most soluble H antigens are converted to *Le*[b] oligosaccharides.[k]

There are at least 12 other well established human blood groups, several of which involve oligosaccharides attached to glycoproteins or glycolipids. The MN antigens consist of a sequence of amino acids near the N terminus of the protein glycophorin (Chapter 8) with attached sialic acid-containing *O*-linked oligosaccharides,[l] e.g., as in the following structure:

```
                NeuAcα2
                      ↓
                      6
NeuAcα2→3Galβ1→3GalNAc→O—Ser (Thr)
          4
          ↑
GalNAcβ1  }in Cad antigen
```

The Cad antigen, also found on glycophorin as well as on gangliosides, has an additional β-1,4-linked GalNAc at the indicated position.[m] The Kell blood group antigens are carried on a 93 kDa glycoprotein of erythrocyte surfaces.[n] The P blood group depends upon surface carbohydrates such as the tetrasaccharide part of the ganglioside called globoside; see Fig. 20-9. Its characteristic antigenic activities are destroyed by treatment with dilute periodate.

Other antigens, including those of the Rh and LW groups, are represented by exposed parts of proteins on the erythrocyte surfaces.[o]


[a] Yamamoto, F., Clausen, H., White, T., Marken, J., and Hakomori, S. (1990) *Nature* **345**, 229–233
[b] Frevert, J., and Ballou, C. E. (1985) *Biochemistry* **24**, 753–759
[c] Feizi, T. (1990) *Trends Biochem. Sci.* **15**, 330–331
[d] Yamamoto, F., and Hakomori, S. (1990) *J. Biol. Chem.* **265**, 19257–19262
[e] Larsen, R. D., Ernst, L. K., Nair, R. P., and Lowe, J. B. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6674–6678
[f] van den Eijnden, D., Koenderman, A., and Schiphorst, W. (1988) *J. Biol. Chem.* **263**, 12461–12471
[g] Mollicone, R., Reguigne, I., Kelly, R. J., Fletcher, A., Watt, J., Chatfield, S., Aziz, A., Cameron, H. S., Weston, B. W., Lowe, J. B., and Oriol, R. (1994) *J. Biol. Chem.* **269**, 20987–20994
[h] Nishihara, S., Narimatsu, H., Iwasaki, H., Yazawa, S., Akamatsu, S., Ando, T., Seno, T., and Ikuyo, N. (1994) *J. Biol. Chem.* **269**, 29271–29278
[i] Natsuka, S., Gersten, K. M., Zenita, K., Kannagi, R., and Lowe, J. B. (1994) *J. Biol. Chem.* **269**, 16789–16794
[j] Murray, B. W., Wittmann, V., Burkart, M. D., Hung, S.-C., and Wong, C.-H. (1997) *Biochemistry* **36**, 823–831
[k] Kelly, R. J., Rouquier, S., Giorgi, D., Lennon, G. G., and Lowe, J. B. (1995) *J. Biol. Chem.* **270**, 4640–4649
[l] Adamany, A., Blumenfeld, O., Sabo, B., and McCreary, J. (1983) *J. Biol. Chem.* **258**, 11537–11545
[m] Gillard, B., Blanchard, D., Bouhours, J.-F., Cartron, J.-P., van Kuik, J. A., Kamerling, J., Vliegenthart, J., and Marcus, D. (1988) *Biochemistry* **27**, 4601–4606
[n] Redman, C., Avellino, G., Pfeffer, S., Mukherjee, T., Nichols, M., Rubinstein, P., and Marsh, W. L. (1986) *J. Biol. Chem.* **261**, 9521–9525
[o] Bloy, C., Hermand, P., Blanchard, D., Cherif-Zahar, B., Goossens, D., and Cartron, J.-P. (1990) *J. Biol. Chem.* **265**, 21482–21487


*N*-linked glycan is pictured in Fig. 4-17. Like many others, it has a number of sialic acid residues at the nonreducing ends and also contains *N*-acetyllactosamine units. The major component of the cell walls of yeast (*S. cerevisiae*) is a mannoprotein that carries long *N*-linked oligosaccharides with highly branched outer chains of over 100 mannose residues[170] (see also Section C,3).

The special importance of sialic acids in glycoproteins may lie in the negative electrical charges and the resultant $Ca^{2+}$-binding properties that they impart to cell surfaces. Some glycoproteins carry sulfate or phosphate groups that have similar effects. The glycoproteins of the slime mold *Dictyostelium* contain both mannose-6-*P* and mannose-6-sulfate residues.[163] Oligosaccharides on the protein subunits of flagella of halobacteria contain sulfate esters of glucuronic acid.[172] Some *N*-linked oligosaccharides carry chains of keratan sulfate in place of the sialic acid in the $Y_1$ and $Y_2$ groups of the oligosaccharide.[173]

## 3. Glycoproteins in Biological Recognition

The clusters of sugar rings that form the oligosaccharides on glycoproteins play a vital role in many aspects of biological recognition.[139,174–176] A good example is provided by the human blood groups whose characteristics are determined largely by oligosaccharides (Box 4-C). The adhesion of viruses, bacteria, and eukaryotic parasites to cell surfaces and of one cell to another in multicellular organisms is also dependent on carbohydrates. Recently, it has become clear that the oligosaccharides of cell surfaces change during growth and development and provide an important mechanism by which cells can recognize each other and respond. Why are carbohydrates used for this purpose? It has been pointed out by Sharon and Lis[175] that four different nucleotides can make only 24 distinct tetranucleotides but that four different monosaccharides can make 35,560 unique tetrasaccharides.

***Lectins and other carbohydrate-binding proteins.*** Much of biological carbohydrate-dependent recognition is a result of interaction of individual glycosyl groups or of oligosaccharides on a glycoprotein with a second protein. In some cases that protein is referred to as a **receptor**; in other cases the glycosyl groups may be called the receptor. Carbohydrate-binding proteins include antibodies, enzymes, and carriers that help sugars to cross cell membranes. In addition, there is a large group of carbohydrate binding proteins called **lectins** (from the latin *lectus*; to select).[24,174,177,178]

The first lectins discovered were proteins of plant seeds with specific sugar-binding properties and the ability to agglutinate erythrocytes. **Ricin**, a very toxic protein from castor beans, was isolated in 1888.[178,179] Perhaps the best known lectin is **concanavalin A**, a protein crystallized by Sumner in 1919.[180] Concanavalin A makes up 2–3% of the protein of the jack bean. It is one of a family of legume lectins that resemble favin, whose structure is shown in Fig. 2-15.[177,181,182] Many lectins, including ricin, have quite different three-dimensional structures but share the common characteristic of having a selective binding site for one or more glycosyl rings. Concanavalin A binds to α-D-mannopyranose or α-D-glucopyranose with unmodified hydroxyl groups at C-3, C-4, and C-6.[183,184] Tighter binding is observed if additional mannose residues are present in an oligosaccharide.[178,183] The protein also has specific binding sites for $Ca^{2+}$ and for a transition metal ion such as $Mn^{2+}$. Soybean lectin binds D-*N*-acetylgalactosamine and D-galactose units, while wheat lectin is specific for D-*N*-acetylglucosamine.

Animal cells also produce lectins.[24,105,185–193] The amebas of the cellular slime mold *Dictyostelium* synthesize a classical lectin called **discoidin I** that binds GalNAc or Gal. It is absent from cells until they are ready to differentiate into an aggregating form (Chapter 1). Then it is produced in abundance.[178] Discoidin I has a second binding site specific for the peptide sequence Arg-Gly-Asp (RGD) which is known to be involved in cell adhesion and which binds to such surface proteins as fibronectin and laminin (Chapter 8). Many animal tissues contain soluble lactose-binding lectins known as **galectins** or S-Lac lectins. The best known member is a dimer of 14-kDa subunits;[187,188] many other related lectins have been found.[189,192] Another family are $Ca^{2+}$-dependent or C-type lectins which are specific for mannose, L-fucose, or other sugars.[105,193] **Lectin domains** are being discovered in many proteins.

***Carbohydrate-binding sites.*** The structures of the sites that recognize and bind sugar rings in lectins, enzymes, transporter proteins, and other carbohydrate-binding proteins vary greatly, as does the tightness of binding. However, there are certain common features: Sugar rings are bound by hydrogen bonds, which are often numerous. An example is the galactose chemoreceptor protein from *E. coli*. It binds both α and β anomers of either D-glucose or D-galactose and is utilized by the bacteria in searching for food (see Chapter 19). The structure of D-glucose bound to this protein is shown in Fig. 4-18. Notice the many hydrogen bonds. Two of the –OH groups have the maximum of three hydrogen bonds apiece. There are three negatively charged aspartate side chains and one positively charged guanidinium group. These provide strong ion–dipole interactions which add strength to the bonds. The presence of ionized groups in varying numbers and constellations is another common feature of protein–carbohydrate interactions. A third common feature is the presence of aromatic rings, which often lie against one face of the sugar. The stereoscopic drawing of Fig. 4-18A shows an indole ring of a tryptophan residue in front and a phenylalanine side chain behind the sugar.[194] Sugars bind to lectins,[181,184,195] to enzymes (Chapter 12), and to antibodies[196] through similar interactions.

***Binding of viruses and bacteria to cells.*** The cholera toxin and a related toxic protein from *E. coli* bind to Sia→Gal groups attached to glycolipids (gangliosides, Chapter 8) of erythrocytes and other cells.[197] The influenza virus gains access to our body cells by first binding through a viral surface **hemagglutinin**. This is a protein that binds specifically to NeuAcα2→6Gal or NeuAcα2→3Gal of cell surface oligosaccharides.[174,175,198] Removal of sialic acid from erythrocyte surfaces abolishes the ability of the influenza and some other viruses to bind. It seems somewhat surprising that a second surface protein on the influenza virus is a **neuraminidase** (or **sialidase**) which catalyzes the removal of sialic acid from cell surface proteins,

destroying unoccupied virus receptors.[199,200] This may facilitate movement of a virus particle through the mucin layer surrounding a cell. Bacteria and other invading parasites also produce neuraminidases.[201] *Trypanosamas cruzi,* the causative agent of chagas disease, employs a **transsialidase** to transfer sialic acid from a Galβ on the host cell onto a protein on the parasite surface. This is essential for successful invasion of the host.[202]

The adherence of cells of *E. coli* to mannosyl units of cell surface proteins may initiate the infections that sometimes occur with this bacterium.[174,203] However, cells of *E. coli* from strains that cause urinary infections bind to Galα1→4Gal on glycolipids that carry the blood group P antigens (Box 4C).[174,204] Neuropathogenic strains of *E. coli* or of *Neisseria meningitidis*, which may cause neonatal meningitis, bind to α-2,8-polysialic acid chains on nerve cells.[135,205,206] *Helicobacter pylori*, the stomach ulcer bacterium, binds to the human Lewis$^b$ blood group antigen (Box 4-C).[207] *Entamoeba histolytica*, which causes amebic dystentery, binds to Gal and GalNAc-containing oligosaccharides such as GalNAcα1→3Galα.[208]

***Aggregation and adherence of cells.*** Differences among cell surfaces are fundamental to the formation of multicellular organisms and to many physiological processes. Proteins, carbohydrates, and lipids all contribute material to exposed cell surfaces. The adhesion of one cell to another is mediated by a group of **adhesion proteins** together with oligosaccharide groups and sometimes polysaccharides. Families of adhesion proteins include **integrins** and **cadhedrins** (discussed in Chapter 8), various members of the **immunoglobulin superfamily**, the **cell differentiation antigens**, (often designated CD44, etc.), the C-type lectins known as **selectins**, and proteoglycans.

Here are two of many known examples of specific cell–cell adhesion. The species-specific reaggregation of dissociated cells of marine sponges (Chapter 1) depends upon a 20-kDa proteoglycan of unique structure[209–211] together with a cell surface receptor protein and calcium ions. The recognition of egg cell surfaces by sperm[212–214] is species specific and depends upon interaction of sperm receptors with *O*-linked oligosaccharides of the extracellular coat of the ovum.

***Growth and differentiation.*** The exact structures of the oligosaccharides and polysaccharides of cell surfaces vary not only with cell and tissue type but also with the position of a cell and with time. Actions of numerous glycosyl transferases alter these saccharide groups as an organism grows and develops. Other enzymes alter them by hydrolytic removal of sugars, by isomerizaion, oxidation, and addition of other components such as phospho, sulfo, and acetal groups. For example, the presence of the H-antigen determinant, whose structure is shown in Box 4-C, is strictly regulated, both temporally and spatially, during vertebrate development.[215,216] The relative amounts of the H determinant vs Le$^x$, sLe$^x$ (Box 4-C)

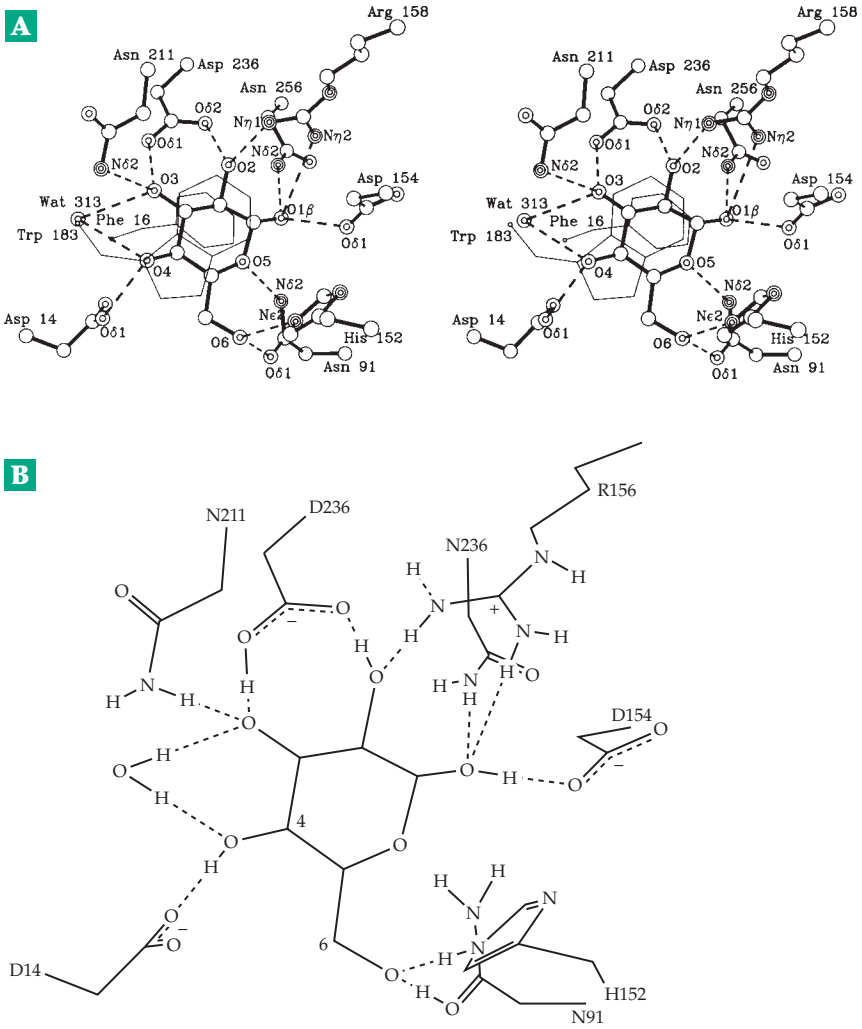
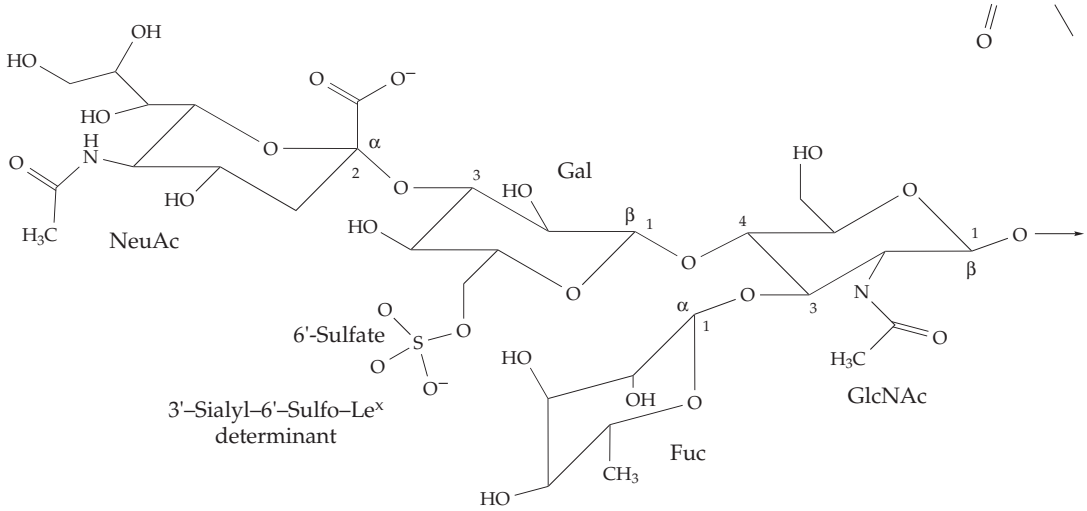



**Figure 4-18** (A) Stereoscopic view of the interactions between the *E. coli* galactose chemoreceptor protein and a bound molecule of D-glucose. (B) Schematic drawing showing many of these interactions and the state of ionization deduced for the aspartate and arginine side chains. From Vyas *et al.*[194]

and other surface groupings is controlled by fucosyltransferases, sialyltransferases, etc. Human cancers often accumulate large amounts of fucosylated glycoproteins and glycolipids carrying $Le^a$, $Le^x$, and sialyl-$Le^x$ antigens[175,216,217] and sialomucins.[218] Glycoproteins help to control many metabolic processes. For example, a protein **calnexin** (a chaperonin, Chapter 10) helps glycoproteins to fold correctly.[219] On the other hand, removal of terminal sialic acid residues from blood plasma proteins leads to rapid removal of the proteins from circulation and to catabolism by liver cells.[175] This process depends upon a receptor protein specific for oligosaccharides with terminal galactosyl residues (Chapter 20).

***Recognition and adhesion by leukocytes.*** A group of three calcium-dependent lectins known as **selectins** bind the sialyl Lewis x ($sLe^x$) antigen and play important roles in adhesion to cells of the vascular endothelium and leukocytes[175,220–224] and also to platelets. Although all of the selectins bind the $sLe^x$ antigen, the binding is weak and these multidomain proteins may simultaneously bind to other ligands such as heparan sulfate.[225] The leukocyte L-selectin (CD62L) binds very tightly to the 3'-sialyl-6'-sulfo-$Le^x$ determinant[220,226] which occurs on mucin-like glycoproteins. The interaction with P selectin helps leukocytes to bind to surfaces on endothelial cells in lymph nodes and sites of chronic or acute inflammation. In a similar way E selectin is synthesized in vascular endothelial cells that have been transiently "activated" by cytokines in response to injury and other inflammatory stimuli. The E selectin binds the sulfated $sLe^x$ antigen on surfaces of neutrophils, monocytes, eosinophils, and basophils.[220,222,227–230] P selectin is stored in secretory granules of platelets and endothelial cells and is released to the cell membrane upon activation by thrombin.[220]

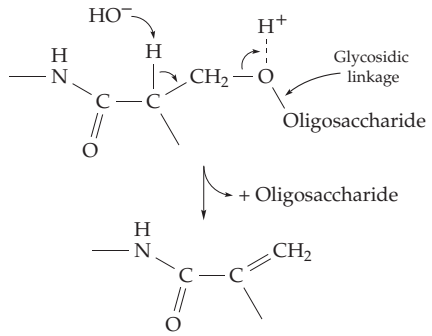



## E. Some Special Methods

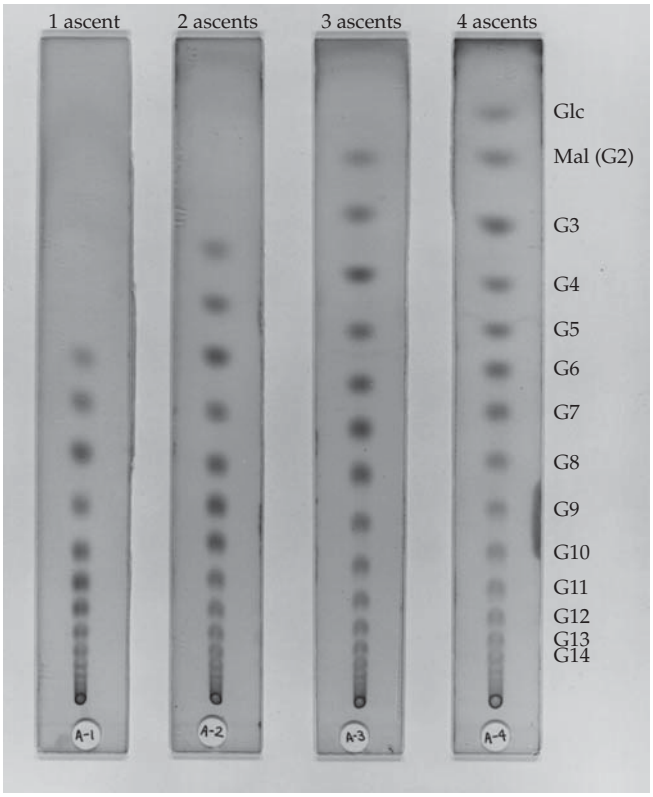
Small monosaccharides and oligosaccharides can be separated readily from polymeric constituents and can be purified further by chromatographic procedures including gel filtration as is illustrated in Fig. 3-3. However, polysaccharides and complex oligosaccharides are harder to purify. A few of these, such as cellulose and glycogen, are sufficiently stable that other materials can be dissolved away from them by prolonged boiling in strongly basic solutions. Complex carbohydrates are usually cut into smaller oligosaccharides or glycopeptides. These may be separated by HPLC, capillary electrophoresis, or thin-layer chromatography[134,231–235] (Fig. 4-19) or by chromatography on immobilized lectins.[234] Quantities of less than 25 picomoles can be separated by use of mass spectrometry with liquid chromatography.[236,237] High resolution Fourier transform mass spectroscopy is very useful in the study of posttranslational glycosylation of proteins.[237a,b]

### 1. Release of Oligosaccharides from Glycoproteins

The *O*-linked oligosaccharides of glycoproteins or glycolipids can be split off from the proteins by β elimination (see also Chapter 13):




(4-10)

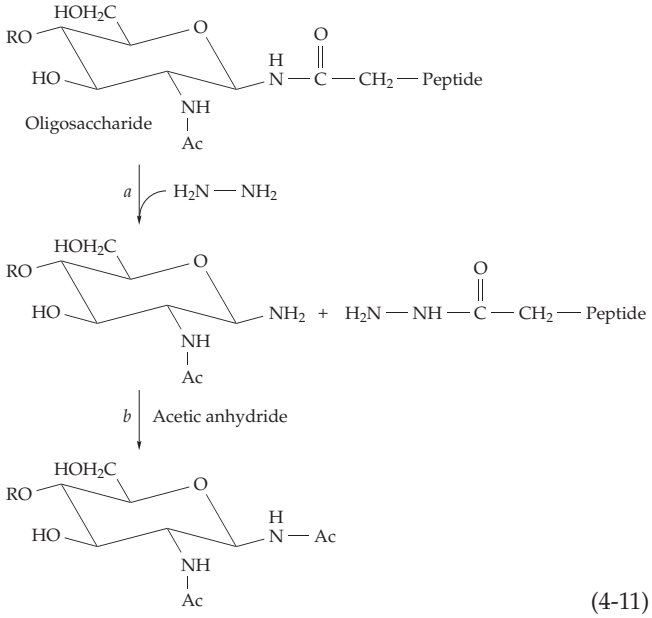



Solvent: MeCN/EtOAc/PrOH-1/$H_2O$
85 20 50 50

**Figure 4-19** Ascending thin-layer chromatography of a mixture of maltooligosaccharides obtained by the hydrolysis of linear starch. G2 (maltose), G3 (maltotriose), G4 (maltotetraose), G5 (maltopentaose) – G14 represent oligosaccharides with the indicated number of glucose residues, all in α-1,4 linkage. In the multiple ascent technique the chromatographic solvent, whose composition (by volume) is indicated, is allowed to ascend the thin-layer plate repeatedly with the plate allowed to dry between ascents. The separation of the higher oligosaccharides is distinctly improved by a larger number of ascents. Photograph courtesy of John Robyt.

Treatment with 0.1–0.5 M NaOH for several hours will completely liberate the oligosaccharides, whose released carbonyl groups may then be reduced with $NaBH_4$, $NaB^2H_4$, or $^3H$-labeled borohydride to form stable sugar alcohols (Eq. 4-2).[238] Asparagine-linked oligosaccharides are often recovered as glycopeptides prepared by complete proteolytic digestion of the denatured glycoprotein. These can be separated by high voltage electrophoresis in borate buffers.[239] The oligosaccharides can be released from the glycopeptides by enzymes such as endo-*N*-acetylglucosaminidase[240] or glycopeptidyl amylase of almonds.[241,242] These release the oligosaccharide as a 1-amino derivative.

When a dry glycoprotein is heated at 105° with anhydrous hydrazine for 8–12 h all of the asparagine-




(4-11)

linked oligosaccharide chains are released (Eq. 4-11).[239,243–245] In addition to the glycosylamine product of step *a*, the corresponding 1–OH and –NH–$NH_2$ compounds are also formed. These are all converted to the stable acetyl derivative (step *b*), after which the oligosaccharides may be characterized by mass spectrometry and high-field NMR spectoscopy.[234,246–248] Crystallization of oligosaccharides in complex mixtures is difficult to impossible.

## 2. Hydrolysis

Most glycosidic linkages are hydrolyzed readily by heating with 1 N mineral acids. The mechanism of the hydrolytic reaction is similar to that employed by the enzyme lysozyme (Chapter 12). Some linkages are unusually sensitive to acid and a few are very resistant. Thus, a variety of conditions may be applied for partial acid hydrolysis as an aid to characterization.[249] Acetolysis, cleavage by acetic anhydride, is also of value.[250] A battery of hydrolytic enzymes specific for sugars that are joined in a given type of glycosidic linkage are available.[251] These are useful in determining sequences of oligosaccharides released from glycoproteins. Radioactive tracer techniques can also be applied.[252]

## 3. Methylation

An important general method in characterization of carbohydrates is the classical **exhaustive methylation** (permethylation). Repeated treatment with a methylating agent such as methyl iodide converts all free OH groups to $OCH_3$ groups. Then, complete acid

hydrolysis, followed by separation of the methylated sugars and their quantitative determination, reveals the relative amounts of **end units** (containing four methoxyl groups), straight **chain units** (containing three methoxyl groups), and **branch points** (containing two methoxyl groups). In addition, the structure of the methylated derivatives provides information on the positions of the linkages in the sugar rings.

After a methylated polysaccharide is subjected to partial hydrolysis, the newly exposed hydroxyl groups (or those created by borohydride reduction of carbonyl groups) can be labeled by ethylation or propylation. One procedure for sequencing complex carbohydrates makes use of high-resolution reversed-phase liquid chromatography to separate the many alkylated oligosaccharides produced by methylation followed by partial acid hydrolysis, reduction, and ethylation.[233] Reductive cleavage of the glycosidic linkages in methylated polysaccharides allows uniquivocal determination of ring size.[253] Branch points may be located by methanolysis of the permethylated polysaccharide followed by conversion of free –OH groups to *p*-bromobenzoate esters. The latter are separated and the circular dichroism (Chapter 23) is measured. Mass spectrometry has also been applied successfully.[246] A simple procedure that can be conducted in any laboratory using thin-layer chromatography is illustrated in Fig. 4-20.[254]

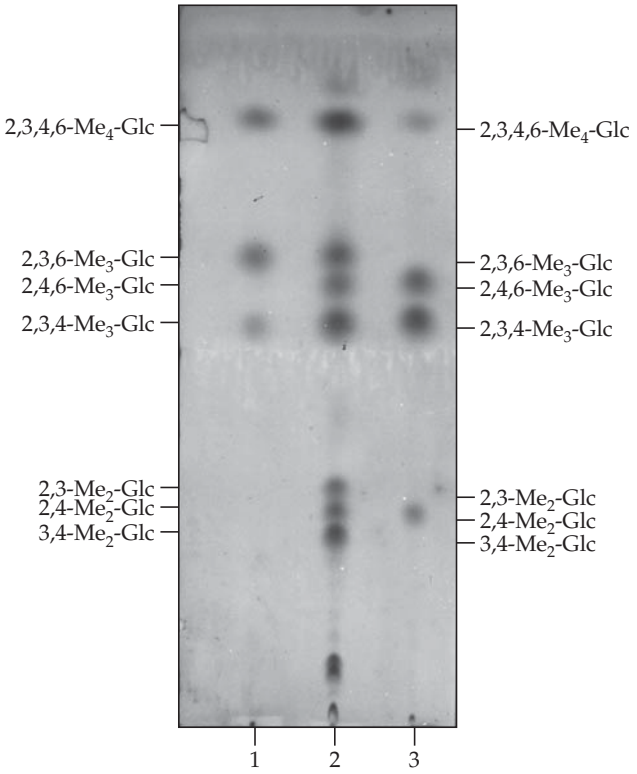
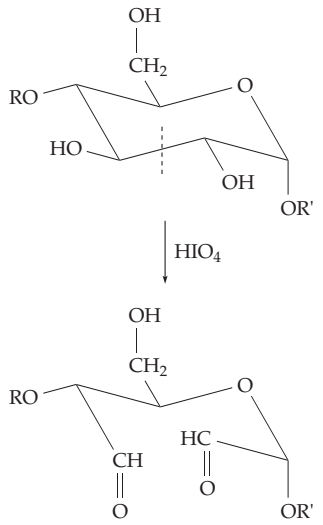



**Figure 4-20** Separation of a mixture of *O*-methylated glucoses by ascending thin-layer chromatography. Whatman K6 TLC plates were used with two ascents of the solvent acetonitrile/chloroform/methanol in the ratio 3/9/2, V/V/V. Courtesy of John Robyt.

## 4. Periodate Oxidation (Smith Degradation)

One of the most important reagents in investigations of carbohydrate structure is periodic acid (or sodium periodate).[255] This reagent oxidatively cleaves C–C bonds bearing adjacent OH or $NH_2$ groups to form dialdehydes (Eq. 4-12). The method is quantitative. After some hours of reaction, excess periodate not consumed in the oxidation can be determined. If three consecutive carbon atoms bear hydroxyl or amino groups, formic acid is liberated from the central atom and can also be measured quantitatively. After destruction of excess periodate the dialdehyde can be reduced by addition of solid sodium borohydride to form stable $CH_2OH$ groups. Following mild acid hydrolysis to split the acyclic acetal linkages, the fragments can be separated and identified. The sequence of reactions is known as the Smith degradation. If sodium borotritide ($NaB^3H_4$) is used for the reduction the fragments will be radioactive and can be located on chromatograms by fluorography. Periodate oxidation can also be used to alter all surface oligosaccharides.[99] Removal of *O*-linked oligosaccharides from glycoproteins can be accomplished by periodate oxidation and alkaline β elimination from the dialdehyde product of Eq. 4-12 if OR′ is part of a serine or other hydroxyamino acid side chain in a protein.[256]



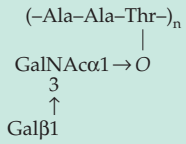

(4-12)

## 5. Nuclear Magnetic Resonance

As in other areas of biochemistry, NMR has become an extremely important tool in carbohydrate research. The mixtures of anomers of various ring

## BOX 4-D ANTIFREEZE AND ICE-NUCLEATION PROTEINS

Fish living in Arctic and Antarctic waters may encounter temperatures as low as –1.9°C. The freezing point depression provided by dissolved salts and proteins in the blood is insufficient to protect the fish from freezing. As winter approaches, they synthesize and accumulate in their blood serum a series of eight or more special antifreeze proteins.[a–d] One type of antifreeze glycoprotein from winter flounder contains the following unit repeated 17–50 times.




Destruction of the galactosyl residues by oxidation with periodate, acetylation of the free hydroxyl groups of the oligosaccharides, or their removal by β elimination all lead to loss of antifreeze activity.

The same fish contain a second series of alanine-rich antifreeze polypeptides that are *not* glycosylated but which exist as amphipathic helices. One of these (Type I) contains ~40 residues in a single helix.[e–i] A third family of antifreeze proteins (Type II), found in the sea raven are globular proteins, rich in cysteine and β structures. They are members of the lectin family.[d,j] A fourth type (Type III) found in the sea pout and some other fishes are 62- to 66-residue globular proteins containing an orthogonal β sandwich structure.[k,l] Messenger RNA molecules coding for the antifreeze proteins are found in the livers of flounder in the winter but are absent in the summer.[m]

Antifreeze proteins, that are 3 – 4 times as effective as those in fish, have been isolated from some insects and other arthropods.[m,n,o] They help beetle larvae to overwinter.[m] The insect proteins have a parallel β helix structure resembling that in Fig. 2-17 and stabilized by S — S bridges.[o,p] Some plants also synthesize antifreeze proteins.[n,q,r] One of these, isolated from carrots, is a member of the leucine-rich-repeat family.[q]

How do antifreeze proteins work? The major effect is to greatly slow the freezing rather than to decrease the freezing point. The proteins apparently accomplish this by binding to the surfaces of small ice crystals and preventing their growth.[d,f,h,i,k,l,s,t] This provides the fish with enough time for the blood to pass back into the liver, in which a high enough temperature is maintained to melt any microcrystals before the blood again circulates through the colder tissues. Some of the proteins have clusters of polar side chains that bind to specific faces of the ice crystals and inhibit growth.[s]

A few fishes tolerate a high internal osmotic pressure and accumulate glycerol in their blood up to a concentration of 0.4 M.[t] Insects may accumulate up to 3 M glycerol and some species utilize various other cryoprotectants, such as mannitol, sorbitol, erythritol, threitol, trehalose, glucose, fructose, proline, and alanine.[u] Some amphibians and reptiles can survive freezing and recover fully. For the most studied wood frog, rapid freezing is fatal, but slow freezing leaves the frog, whose heart ceases to function, with a 200-fold increased glucose concentration and a decreased water content in its organs. It resumes normal activities within 14–24 h of thawing.[t]

Having an effect opposite to that of the antifreeze proteins are surface proteins of some bacteria of the genera *Pseudomonas*, *Erwinia*, and *Xanthomonas*. These proteins provide nuclei for growth of ice crystals from supercooled water.[v,w]

---


[a] Feeney, R. E., Burcham, T. S., and Yeh, Y. (1986) *Ann. Rev. Biophys. Biophys. Chem.* **15**, 59–78
[b] Eastman, J. T., and DeVries, A. L. (1986) *Sci. Am.* **255** (Nov), 106–114
[c] Davies, P. L., and Hew, C. L. (1990) *FASEB J.* **4**, 2460–2468
[d] Gronwald, W., Loewen, M. C., Lix, B., Daugulis, A. J., Sönnichsen, F. D., Davies, P. L., and Sykes, B. D. (1998) *Biochemistry* **37**, 4712–4721
[e] Chakrabartty, A., Ananthanarayanan, V. S., and Hew, C. L. (1989) *J. Biol. Chem.* **264**, 11307–11312
[f] Yang, D. S. C., Sax, M., Chakrabartty, A., and Hew, C. L. (1988) *Nature (London)* **333**, 232–237
[g] Madura, J. D., Wierzbicki, A., Harrington, J. P., Maughon, R. H., Raymond, J. A., and Sikes, C. S. (1994) *J. Am. Chem. Soc.* **116**, 417–418
[h] Sicheri, F., and Yang, D. S. C. (1995) *Nature (London)* **375**, 427–431
[i] Wierzbicki, A., Taylor, M. S., Knight, C. A., Madura, J. D., Harrington, J. P., and Sikes, C. S. (1996) *Biophys. J.* **71**, 8–18
[j] Ng, N. F. L., and Hew, C. L. (1992) *J. Biol. Chem.* **267**, 16069–16075
[k] Chao, H., Sönnichsen, F. D., DeLuca, C. I., Sykes, B. D., and Davies, P. L. (1994) *Protein Sci.* **3**, 1760–1769
[l] Jia, Z., DeLuca, C. I., Chao, H., and Davies, P. L. (1996) *Nature (London)* **384**, 285–288
[m] Gourlie, B., Lin, Y., Price, J., DeVries, A. L., Powers, D., and Huang, R. C. C. (1984) *J. Biol. Chem.* **259**, 14960–14965
[n] Li, N., Chibber, B. A. K., Castellino, F. J., and Duman, J. G. (1998) *Biochemistry* **37**, 6343–6350
[o] Graether, S. P., Kuiper, M. J., Gagné, S. M., Walker, V. K., Jia, Z., Sykes, B. D., and Davies, P. L. (2000) *Nature (London)* **406**, 325 – 328
[p] Liou, Y.-C., Tocilj, A., Davies, P. L., and Jia, Z. (2000) *Nature (London)* **406**, 322–324
[q] Worrall, D., Elias, L., Ashford, D., Smallwood, M., Sidebottom, C., Lillford, P., Telford, J., Holt, C., and Bowles, D. (1998) *Science* **282**, 115–117
[r] Sidebottom, C., Buckley, S., Pudney, P., Twigg, S., Jarman, C., Holt, C., Telford, J., McArthur, A., Worrall, D., Hubbard, R., and Lillford, P. (2000) *Nature (London)* **406**, 256
[s] Knight, C. A., (2000) *Nature (London)* **406**, 249–251
[t] Costanzo, J. P., Lee, R. E., Jr., DeVries, A. L., Wang, T., and Layne, J. R., Jr. (1995) *FASEB J.* **9**, 351–352
[u] Storey, K. B., and Storey, J. M. (1990) *Sci. Am.* **263** (Dec), 92–97
[v] Wolber, P., and Warren, G. (1989) *Trends Biochem. Sci.* **14**, 179–182
[w] Gurian-Sherman, D., and Lindow, S. E. (1993) *FASEB J.* **7**, 1338–1343

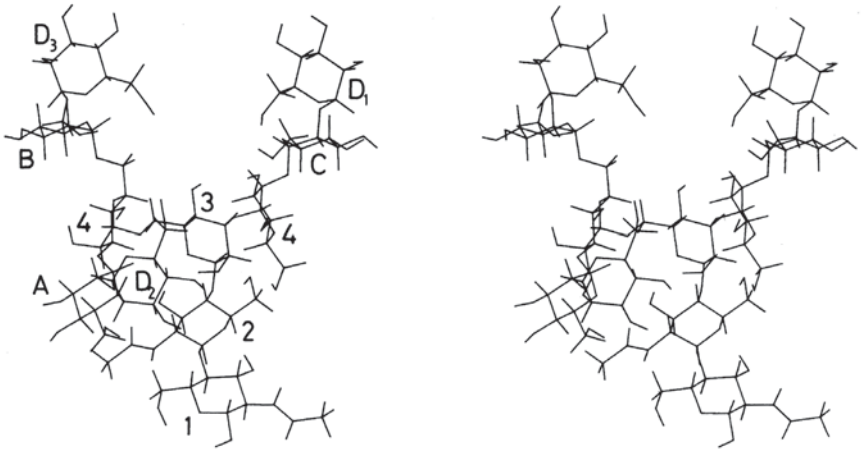



**Figure 4-21** Stereoscopic view of an N-linked oligosaccharide whose structure has been deduced by two-dimensional NMR spectroscopy and energy calculations. This is one of a range of allowed conformations. From Homans *et al.*[257]

sizes (Eq. 4-1) can be analyzed with proton or $^{13}C$ NMR.[10,257] A variety of newer NMR techniques, some of which have been described in Chapter 3, have been applied.[28,258–267] A problem with NMR spectroscopy of these compounds has been the lack of the large number of nuclear Overhauser enhancements (NOEs) that can be observed.[268,269] Nevertheless, when combined with energy calculations it is possible to use NMR measurements to deduce conformations of sugar rings, three-dimensional structures, and the degree of conformational flexibility in various parts of N-linked oligosaccharides (Fig. 4-21). Measurement of C–O–C–C spin-coulping constants is also of value.[270] Use of multidimensional NMR has permitted analysis of mixtures of cellulose oligosaccharides.[269]

## References


1. Pigman, W. W., and Horton, D., eds. (1972) *The Chemistry of Carbohydrates*, 2nd ed., Academic Press, New York
2. Guthrie, R. D. (1974) *Guthrie and Honeyman's Introduction to Carbohydrate Chemistry*, 4th ed., Oxford Univ. Press (Clarendon), London and New York (p. 17)
3. Binkley, R. W. (1988) *Modern Carbohydrate Chemistry*, Dekker, New York
4. Shalenberger, R. S. (1982) *Advanced Sugar Chemistry: Principles of Sugar Stereochemistry*, AVI Publ. Co., Westport, Connecticut
5. El-Khadem, H. S. (1988) *Carbohydrate Chemistry*, Academic Press, San Diego, California
6. Prince, R. C., Gunson, D. E., Leigh, J. S., and McDonald, G. G. (1982) *Trends Biochem. Sci.* **7**, 239–240
7. Wertz, P. W., Garver, J. C., and Anderson, L. (1981) *J. Am. Chem. Soc.* **103**, 3916–3922
8. Ha, S., Gao, J., Tidor, B., Brady, J. W., and Karplus, M. (1991) *J. Am. Chem. Soc.* **113**, 1553–1557
9. Angyal, S. J. (1979) in *Asymmetry of Carbohydrates* (Harmon, R. E., ed), p. 15, Decker, New York
10. Gray, G. R. (1976) *Acc. Chem. Res.* **9**, 418
11. Pierce, J., Serianni, A. S., and Barker, R. (1985) *J. Am. Chem. Soc.* **107**, 2448–2456
12. Maple, S. R., and Allerhand, A. (1987) *J. Am. Chem. Soc.* **109**, 3168–3169
13. Goux, W. J. (1985) *J. Am. Chem. Soc.* **107**, 4320–4327
14. Barker, S. A., and Baggett, N. (1978) *Trends Biochem. Sci.* **3**, 140–141
15. Benson, A. A. (1963) *Adv. Lipid Res.* **1**, 387–394
16. Benning, C., Beatty, J. T., Prince, R. C., and Somerville, C. R. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1561–1565
17. Dijkema, C., Kester, H. C. M., and Visser, J. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 14–18
18. Tarczynski, M. C., Jensen, R. G., and Bohnert, H. J. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2600–2604
19. Yamanouchi, T., Tachibana, Y., Sekino, N., Akanuma, H., Akaoka, I., and Miyashita, H. (1994) *J. Biol. Chem.* **269**, 9664–9668
20. Nadano, D., Iwasaki, M., Endo, S., Kitajima, K., Inoue, S., and Inoue, Y. (1986) *J. Biol. Chem.* **261**, 11550–11557
21. Schauer, R. (1985) *Trends Biochem. Sci.* **10**, 357–360
22. Faillard, H. (1989) *Trends Biochem. Sci.* **14**, 237–241
23. Roth, J., Kempf, A., Reuter, G., Schauer, R., and Gehring, W. J. (1992) *Science* **256**, 673–675
24. Powell, L. D., and Varki, A. (1995) *J. Biol. Chem.* **270**, 14243–14246
25. Stoddart, J. F. (1971) *Stereochemistry of Carbohydrates*, Wiley-Interscience, New York
26. Brady, J. W. (1986) *J. Am. Chem. Soc.* **108**, 8152–8160

26a. Marszalek, P. E., Pang, Y.-P., Li, H., Yazal, J. E., Oberhauser, A. F., and Fernandez, J. M. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7894–7898

27. Venkataraman, G., Sasisekharan, V., Cooney, C. L., Langer, R., and Sasisekharan, R. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6171–6175
28. Hajduk, P. J., Horita, D. A., and Lerner, L. E. (1993) *J. Am. Chem. Soc.* **115**, 9196–9201
29. Wolfe, S., Whangbo, M.-H., and Mitchell, D. J. (1979) *Carbohydrate Res.* **69**, 1–26
30. Kirby, A. J. (1983) *The Anomeric Effect and Related Stereoelectronic Effects at Oxygen*, Springer, Berlin
31. Perrin, C. L., Armstrong, K. B., and Fabian, M. A. (1994) *J. Am. Chem. Soc.* **116**, 715–722
32. Cui, W., Li, F., and Allinger, N. L. (1993) *J. Am. Chem. Soc.* **115**, 2943–2951
33. Westhof, E., and Sundaralingam, M. (1980) *J. Am. Chem. Soc.* **102**, 1493–1500

33a. French, A. D., Dowd, M. K., and Reilly, P. J. (1997) *J. Mol. Struct. (Theochem)*, 395–396

34. Gabb, H. A., and Harvey, S. C. (1993) *J. Am. Chem. Soc.* **115**, 4218–4227
35. Plavec, J., Tong, W., and Chattopadhyaya, J. (1993) *J. Am. Chem. Soc.* **115**, 9734–9746
36. Ellervik, U., and Magnusson, G. (1994) *J. Am. Chem. Soc.* **116**, 2340–2347

36a. French, A. D., and Murphy, V. G. (1973) *Carbohydr. Res.* **27**, 391–406

37. Robyt, J. F., and White, B. J. (1996) *Biochemical Techniques, Theory and Practice*, 2nd ed., Waveland Press, Prospect Heights, Illinois
38. Lipták, A., Fügedi, P., Szurmai, Z., and Harangi, J. (1990) *CRC Handbook of Oligosaccharides. Vol. I: Disaccharides*, CRC Press, Boca Raton, Florida
39. Cerami, A. (1986) *Trends Biochem. Sci.* **11**, 311–314
40. Watkins, N. G., Neglia-Fisher, C. I., Dyer, D. G., Thorpe, S. R., and Baynes, J. W. (1987) *J. Biol. Chem.* **262**, 7207–7212
41. Súarez, G., Rajaram, R., Oronsky, A. L., and Gawinowicz, M. A. (1989) *J. Biol. Chem.* **264**, 3674–3679

## References


42. Takata, K., Horiuchi, S., Araki, N., Shiga, M., Saitoh, M., and Morino, Y. (1988) *J. Biol. Chem.* **263**, 14819–14825
43. Hayase, F., Nagaraj, R. H., Miyata, S., Njoroge, F. G., and Monnier, V. M. (1989) *J. Biol. Chem.* **264**, 3758–3764
44. Goldberg, R. N., and Tewari, Y. B. (1989) *J. Biol. Chem.* **264**, 9897–9900
45. Tewari, Y. B., and Goldberg, R. N. (1989) *J. Biol. Chem.* **264**, 3966–3971
46. Goldberg, R. N., Tewari, Y. B., and Ahluwalia, J. C. (1989) *J. Biol. Chem.* **264**, 9901–9904
47. Weissborn, A. C., Rumley, M. K., and Kennedy, E. P. (1991) *J. Biol. Chem.* **266**, 8062–8067
48. Reymond, P., Grünberger, S., Paul, K., Müller, M., and Farmer, E. E. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4145–4149
49. Nacharaju, P., and Acharya, A. S. (1992) *Biochemistry* **31**, 12673–12679
50. Baynes, J. W., and Monnier, V. M., eds. (1989) *the Maillard Reaction in Aging, Diabetes and Nutrition*, Liss, New York
51. Larsen, M. L., Hørder, M., and Mogensen, E. F. (1990) *N. Engl. J. Med.* **323**, 1021–1025
52. Luthra, M., and Balasubramanian, D. (1993) *J. Biol. Chem.* **268**, 18119–18127
53. Urbanowski, J. C., Cohenford, M. A., and Dain, J. A. (1982) *J. Biol. Chem.* **257**, 111–115
54. Grandhee, S. K., and Monnier, V. M. (1991) *J. Biol. Chem.* **266**, 11649–11653
55. Brownlee, M., Cerami, A., and Vlassara, H. (1988) *N. Engl. J. Med.* **318**, 1315–1320
56. Papoulis, A., Al-Abed, Y., and Bucala, R. (1995) *Biochemistry* **34**, 648–655
57. Wells-Knecht, K. J., Zyzak, D. V., Litchfield, J. E., Thorpe, S. R., and Baynes, J. W. (1995) *Biochemistry* **34**, 3702–3709
58. French, D. (1975) in *International Review of Biochemistry*, Vol. 5 (Whelan, W. J., ed), p. 269, Univ. Park Press, Baltimore, Maryland (Series One)
59. Atkins, E. D. T., ed. (1986) *Polysaccharides, Topics in Structure and Morphology*, VCH Publisher, New York
60. Whistler, R. L., and BeMiller, J. N., eds. (1993) *Industrial Gums*, 3rd ed., Academic Press, San Diego, California
61. Rao, V. S. R., and Sathyanarayana, B. K. (1972) *Biopolymers* **11**, 1379–1394
62. French, A. D., and Dowd, M. K. (1993) in *Cellulostics: Chemical,Biochemical and Materials Aspects* (Kennedy, J. F., Phillips, G. O., and Williams, P. A., eds), pp. 51–56, Ellis Harwood, New York
63. Duda, C. A., and Stevens, E. S. (1993) *J. Am. Chem. Soc.* **115**, 8487–8488
64. Liébecq, C., ed. (1992) *Biochemical Nomenclature*, Portland Press, London and Chapel Hill, North Carolina (for the International Union of Biochemistry and Molecular Biology)
65. Preston, R. D. (1986) *Trends Biochem. Sci.* **11**, 377–380
66. Ross, P., Mayer, R., and Benziman, M. (1991) *Microbiol. Rev.* **55**, 35–58
67. Sugiyama, J., Vuong, R., and Chanzy, H. (1991) *Macromolecules* **24**, 4168–4175
68. Dudley, R. L., Fyfe, C. A., Stephenson, P. J., Deslandes, Y., Hamer, G. K., and Marchessault, R. H. (1983) *J. Am. Chem. Soc.* **105**, 2469–2472
69. Gebler, K., Kraub, N., Steiner, T., Betzel, C., Sandmann, C., and Saenger, W. (1994) *Science* **266**, 1027–1029

69a. Langan, P., Nishiyama, Y., and Chanzy, H. (1999) *J. Am. Chem. Soc.* **121**, 9940–9946

70. Grover, J. A. (1993) in *Industrial Gums*, 3rd ed. (Whistler, R. L., and BeMiller, J. N., eds), pp. 475–504, Academic Press, San Diego, California
71. French, D. (1973) *J. Animal Sci.* **37**, 1048–1061
72. Maddelein, M. L., Libessart, N., Bellanger, F., Delrue, B., D'Hulst, C., Van den Koornhuyse, N., Fontaine, T., Wieruszeski, J. M., Decq, A., and Ball, S. (1994) *J. Biol. Chem.* **269**, 25150–25157
73. Jane, J., Kasemsuwan, T., Leas, S., Zobel, H., and Robyt, J. F. (1994) *Starch/Stärke* **46**, 121–129
74. Meyer, K. H. (1943) *Adv. Enzymol.* **3**, 109–136
75. French, D. (1969) in *Symposium on Foods: Carbohydrates and their Roles* (Schultz, H. W., ed), pp. 26–54, AVI Publ. Co., Westport, Connecticut
76. Rundle, R. E., and French, D. (1943) *J. Am. Chem. Soc.* **65**, 1707–1710
77. Hybl, A., Rundle, R. E., and Williams, D. E. (1965) *J. Am. Chem. Soc.* **87**, 2779–2788
78. Teitelbaum, R. C., Ruby, S. L., and Tobin, J. M. (1980) *J. Am. Chem. Soc.* **102**, 3322–3328
79. Winter, W. T., and Sarko, A. (1974) *Biopolymers* **13**, 1461–1482
80. Kainuma, K., and French, D. (1972) *Biopolymers* **11**, 2241–2250
81. Wu, H.-C., H, and Sarko, A. (1978) *Carbohydr. Res.* **61**, 7–25
82. Gidley, M. J., and Bociek, S. M. (1988) *J. Am. Chem. Soc.* **110**, 3820–3829
83. Lomako, J., Lomako, W. M., Whelan, W. J., Dombro, R. S., Neary, J. T., and Norenberg, M. D. (1993) *FASEB J.* **7**, 1386–1393
84. Chen, W., Avison, M. J., Zhu, X. H., and Shulman, R. G. (1993) *Biochemistry* **32**, 11483–11487
85. Ohana, P., Delmer, D. P., Steffens, J. C., Matthews, D. E., Mayer, R., and Benziman, M. (1991) *J. Biol. Chem.* **266**, 13742–13745
86. Mol, P. C., Park, H., Mullins, J. T., and Cabib, E. (1994) *J. Biol. Chem.* **269**, 31267–31274
87. Yamamoto, R., and Nevins, D. J. (1978) *Carbohydrate Res.* **67**, 275–280
88. Marchessault, R. H., and Deslandes, Y. (1979) *Carbohydrate Res.* **75**, 231–242
89. Kang, M. S., and Cabib, E. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5808–5812
90. Harada, T., Terasaki, M., and Harada, A. (1993) in *Industrial Gums*, 3rd ed. (Whistler, R. L., and BeMiller, J. N., eds), pp. 427–445, Academic Press, San Diego, California
91. Walker, G. J. (1978) in *International Review of Biochemistry*, Vol. 16 (Manners, D. J., ed), p. 75, Univ. Park Press, Baltimore, Maryland
92. de Belder, A. N. (1993) in *Industrial Gums*, 3rd ed. (Whistler, R. L., and BeMiller, J. N., eds), pp. 399–425, Academic Press, San Diego, California
93. Yang, X., Miller, M. A., Yang, R., Evans, D. F., and Edstrom, R. D. (1990) *FASEB J.* **4**, 3140–3143
94. Tsujisaka, Y., and Mitsuhashi, M. (1993) in *Industrial Gums*, 3rd ed. (Whistler, R. L., and BeMiller, J. N., eds), pp. 447–460, Academic Press, San Diego, California
95. Arnott, S., Fulmer, A., Scott, W. E., Dea, I. C. M., Moorehouse, R., and Rees, D. A. (1974) *J. Mol. Biol.* **90**, 269–284
96. Shibata, N., Ikuta, K., Imai, T., Satoh, Y., Satoh, R., Suzuki, A., Kojima, C., Kobayashi, H., Hisamichi, K., and Suzuki, S. (1995) *J. Biol. Chem.* **270**, 1113–1122
97. Kollár, R., Petráková, E., Ashwell, G., Robbins, P. W., and Cabib, E. (1995) *J. Biol. Chem.* **270**, 1170–1178
98. Preston, R. D. (1968) *Sci. Am.* **218**(Jun), 102–108
99. Turvey, J. R. (1978) in *International Review of Biochemistry*, Vol. 16 (Manners, D. J., ed), Univ. Park Press, Baltimore, Maryland
100. Kirkwood, S. (1974) *Ann. Rev. Biochem.* **43**, 401–417
101. Minke, R., and Blackwell, J. (1978) *J. Mol. Biol.* **120**, 167–181
102. Roberts, G. A. F. (1992) *Chitin Chemistry*, MacMillan, London
103. Blackwell, J., and Weih, M. A. (1980) *J. Mol. Biol.* **137**, 49–60
104. Rees, D. A., and Welsh, E. J. (1977) *Angew. Chem. Int. Ed. Engl.* **16**, 214–224
105. Lee, Y. C. (1992) *FASEB J.* **6**, 3193–3200
106. Laurent, T. C., and Fraser, J. R. E. (1992) *FASEB J.* **6**, 2397–2404
107. Hardingham, T. E., and Fosang, A. J. (1992) *FASEB J.* **6**, 861–870
108. Fransson, L. (1987) *Trends Biochem. Sci.* **12**, 406–411
109. Morris, E. R., Rees, D. A., and Welsh, E. J. (1980) *J. Mol. Biol.* **138**, 383–400
110. David, G. (1993) *FASEB J.* **7**, 1023–1030
111. Yamada, S., Murakami, T., Tsuda, H., Yoshida, K., and Sugahara, K. (1995) *J. Biol. Chem.* **270**, 8696–8705
112. Lane, D. A., and Björk, I., eds. (1992) *Heparin And Related Polysaccharides*, Plenum, New York
113. Edens, R. E., Fromm, J. R., Linhardt, R. J., and Weiler, J. M. (1995) *Biochemistry* **34**, 2400–2407
114. Linhardt, R. J., Ampofo, S. A., Fareed, J., Hoppensteadt, D., Mulliken, J. B., and Folkman, J. (1992) *Biochemistry* **31**, 12441–12445
115. Casu, B., Petitou, M., Provasoli, M., and Sinaÿ, P. (1988) *Trends Biochem. Sci.* **13**, 221–225
116. Oates, J. A., and Wood, A. J. J. (1991) *N. Engl. J. Med.* **324**, 1565–1574
117. Rosenberg, R. D., and Lam, L. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1218–1222
118. Lindahl, U., Kusche-Gullberg, M., and Kjelln, L. (1998) *J. Biol. Chem.* **273**, 24979–24982
119. Mulloy, B. A., Ribeiro, A.-C., Alves, A.-P., Vieira, R. P., and Mourao, P. A. S. (1994) *J. Biol. Chem.* **269**, 22113–22123
120. Gibeaut, D., and Carpita, N. C. (1994) *FASEB J.* **8**, 904–915
121. Hayashi, T., and Matsuda, K. (1981) *J. Biol. Chem.* **256**, 11117–11122
122. Millane, R. P. (1992) in *Frontiers in Carbohydrate Research-2* (Chandrasekaran, R., ed), pp. 168–190, Elsevier, London
123. Rolin, C. (1993) in *Industrial Gums*, 3rd ed. (Whistler, R. L., and BeMiller, J. N., eds), pp. 257–293, Academic Press, San Diego, California
124. Arnott, S., Scott, W. E., Rees, D. A., and McNab, C. G. A. (1974) *J. Mol. Biol.* **90**, 253–267
125. Selby, H. H., and Whistler, R. L. (1993) in *Industrial Gums*, 3rd ed. (Whistler, R. L., and BeMiller, J. N., eds), pp. 87–103, Academic Press, San Diego, California
126. Albersheim, P., Bauer, W. D., Keestra, K., Talmadge, K. W. (1973) *Biogenesis of Plant Cell Wall Polysaccharides*, (Loewus, F., ed.), Academic Press, New York
127. Rees, D. A. (1972) *Biochem. J.* **126**, 257–273
128. Therkelsen, G. H. (1993) in *Industrial Gums*, 3rd ed. (Whistler, R. L., and BeMiller, J. N., eds), pp. 145–180, Academic Press, San Diego, California
129. Clare, K. (1993) in *Industrial Gums*, 3rd ed. (Whistler, R. L., and BeMiller, J. N., eds), pp. 105–143, Academic Press, San Diego, California
130. Marsh, M. E., Chang, D. C., and King, G. C. (1992) *J. Biol. Chem.* **267**, 20507–20512

130a. Vinogradov, E., and Bock, K. (1999) *Angew. Chem. Ed. Engl.* **38**, 671–674

131. Sutherland, I. W. (1979) *Trends Biochem. Sci.* **4**, 55–59

## References


132. Kang, K. S., and Pettitt, D. J. (1993) in *Industrial Gums*, 3rd ed. (Whistler, R. L., and BeMiller, J. N., eds), pp. 341–397, Academic Press, San Diego, California
133. Wolucka, B. A., McNeil, M. R., de Hoffmann, E., Chonjnacki, T., and Brennan, P. J. (1994) *J. Biol. Chem.* **269**, 23328–23335
134. Kobata, A. (1984) in *Biology of Carbohydrates*, Vol. 2 (Ginsburg, V., and Robbins, P. W., eds), pp. 87–161, Wiley, New York
135. Lennarz, W. J., ed. (1980) *The Biochemistry of Glycoproteins and Proteoglycans*, Plenum, New York
136. Montreul, J. (1982) in *Comprehensive Biochemistry*, 1st ed., Vol. 19B, Part II (Florkin, M., and Stotz, E. H., eds), pp. 1–188, Elsevier, Amsterdam
137. Rademacher, T. W., Parekh, R. B., and Dwek, R. A. (1988) *Ann. Rev. Biochem.* **57**, 785–838
138. Roberts, D. D., and Mecham, R. P., eds. (1993) *Cell Surface and Extracellular Glycoconjugates: Structure and Function*, Academic Press, San Diego, California
138a. Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J., eds. (1999) *Essentials of Glycobiology*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
139. Fukuda, M., and Hindsgaul, O., eds. (1994) *Molecular Glycobiology*, Oxford Publishing, Oxford
140. Williams, R. L., Greene, S. M., and McPherson, A. (1987) *J. Biol. Chem.* **262**, 16020–16031
141. Hofsteenge, J., Müller, D. R., de Beer, T., Löffler, A., Richter, W. J., and Vliegenthart, J. F. G. (1994) *Biochemistry* **33**, 13524–13530
141a. Iozzo, R. V. (1998) *Ann. Rev. Biochem.* **67**, 609–652
141b. Bernfield, M., Götte, M., Park, P. W., Reizes, O., Fitzgerald, M. L., Lincecum, J., and Zako, M. (1999) *Ann. Rev. Biochem.* **68**, 729–777
142. Sugahara, K., Ohkita, Y., Shibata, Y., Yoshida, K., and Ikegami, A. (1995) *J. Biol. Chem.* **270**, 7204–7212
143. Funderburgh, J. L., Funderburgh, M. L., Brown, S., Vergnes, J.-P., Hassell, J. R., Mann, M. M., and Conrad, G. W. (1993) *J. Biol. Chem.* **268**, 11874–11880
144. Neame, P. J., Christner, J. E., and Baker, J. R. (1987) *J. Biol. Chem.* **262**, 17768–17778
145. Scott, J. E. (1987) *Trends Biochem. Sci.* **12**, 318–321
146. Rosenberg, L., Tang, L.-H., Pal, S., Johnson, T. L., and Choi, H. U. (1988) *J. Biol. Chem.* **263**, 18071–18077
147. Choi, H. U., Johnson, T. L., Pal, S., Tang, L.-H., Rosenberg, L., and Neame, P. J. (1989) *J. Biol. Chem.* **264**, 2876–2884
148. Kokenyesi, R., and Bernfield, M. (1994) *J. Biol. Chem.* **269**, 12304–12309
149. Rosenberg, L. (1975) in *Dynamics of Connective Tissue Macromolecules* (Burleigh, P. M. C., and Poole, A. R., eds), p. 107, North-Holland Publ., Amsterdam
150. Chen, J., and Varner, J. E. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4399–4403
151. Kieliszewski, M., deZacks, R., Leykam, J. F., and Lamport, D. T. A. (1992) *Plant Physiol.* **98**, 919–926
152. Kieliszewski, M. J., O'Neill, M., Leykam, J., and Orlando, R. (1995) *J. Biol. Chem.* **270**, 2541–2549
153. Eckhardt, A. E., Timpte, C. S., Abernethy, J. L., Toumadje, A., Johnson, W. C., and Hill, R. L. (1987) *J. Biol. Chem.* **262**, 11339–11344
154. Timpte, C. S., Eckhardt, A. E., Abernethy, J. L., and Hill, R. L. (1988) *J. Biol. Chem.* **263**, 1081–1088
155. Gupta, R., and Jentoft, N. (1989) *Biochemistry* **28**, 6114–6121
156. Carter, S. R., Slomiany, A., Gwozdzinski, K., Liau, Y. H., and Slomiany, B. L. (1988) *J. Biol. Chem.* **263**, 11977–11984
157. Jentoft, N. (1990) *Trends Biochem. Sci.* **15**, 291–294
158. Butenhof, K. J., and Gerken, T. A. (1993) *Biochemistry* **32**, 2650–2663
159. Toribara, N. W., Roberton, A. M., Ho, S. B., Kuo, W.-L., Gum, E., Hicks, J. W., Gum, J. R., Byrd, J. C., Siddiki, B., and Kim, Y. S. (1993) *J. Biol. Chem.* **268**, 5879–5885
160. Sadler, J. E. (1984) in *"Biology of Carbohydrates"*, Vol. 2 (Ginsburg, V., and Robbins, P. W., eds), pp. 199–288, Wiley, New York
160a. Perez-Vilar, J., and Hill, R. L. (1999) *J. Biol. Chem.* **274**, 31751–31754
161. Fukuda, M., Carlsson, S. R., Klock, J. C., and Dell, A. (1986) *J. Biol. Chem.* **261**, 12796–12806
162. Gerken, T. A., and Jentoft, N. (1987) *Biochemistry* **26**, 4689–4699
163. Freeze, H. H., and Wolgast, D. (1986) *J. Biol. Chem.* **261**, 127–134
164. Carver, J. P., and Brisson, J.-R. (1984) in *Biology of Carbohydrates*, Vol. 2 (Ginsburg, V., and Robbins, P. W., eds), pp. 289–329, Wiley, New York
165. Bush, C. A., Duben, A., and Ralapati, S. (1980) *Biochemistry* **19**, 501–504
166. Finne, J., and Krusius, T. (1979) *Eur. J. Biochem.* **102**, 583–588
167. Järnefelt, J., Finne, J., Krusius, T., and Rauvala, H. (1978) *Trends Biochem. Sci.* **3**, 110–114
168. Kornfeld, R., and Kornfeld, S. (1980) in *The Biochemistry of Glycoproteins and Proteoglycans* (Lennarz, W. J., ed), Plenum, New York
169. Fukuda, M., Dell, A., and Fukuda, M. N. (1984) *J. Biol. Chem.* **259**, 4782–4791
170. Hernández, L. M., Olivero, I., Alvarado, E., and Larriba, G. (1992) *Biochemistry* **31**, 9823–9831
171. Taguchi, T., Seko, A., Kitajima, K., Muto, Y., Inoue, S., Khoo, K.-H., Morris, H. R., Dell, A., and Inoue, Y. (1994) *J. Biol. Chem.* **269**, 8762–8771
172. Wieland, F., Paul, G., and Sumper, M. (1985) *J. Biol. Chem.* **260**, 15180–15185
173. Funderburgh, J. L., Funderburgh, M. L., Mann, M. M., and Conrad, G. W. (1991) *J. Biol. Chem.* **266**, 14226–14231
174. Sharon, N., and Lis, H. (1989) *Science* **246**, 227–229
175. Sharon, N., and Lis, H. (1993) *Sci. Am.* **268**(Jan), 82–89
176. Opdenakker, G., Rudd, P. M., Ponting, C. P., and Dwek, R. A. (1993) *FASEB J.* **7**, 1330–1337
177. Sharon, N. (1993) *Trends Biochem. Sci.* **18**, 221–226
178. Barondes, S. H. (1988) *Trends Biochem. Sci.* **13**, 480–482
179. Sphyris, N., Lord, J. M., Wales, R., and Roberts, L. M. (1995) *J. Biol. Chem.* **270**, 20292–20297
180. Sumner, J. B., and Howell, S. F. (1936) *J. Bacteriol.* **32**, 227–237
181. Bourne, Y., Rougés, P., and Cambillau, C. (1990) *J. Biol. Chem.* **265**, 18161–18165
182. Sharon, N., and Lis, H. (1990) *FASEB J.* **4**, 3198–3208
183. Mandal, D. K., and Brewer, C. F. (1993) *Biochemistry* **32**, 5116–5120
184. Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb, A. J., Dodson, E. J., Papiz, M. Z., Wan, T., and Campbell, J. (1989) *EMBO J.* **8**, 2189–2193
185. Drickamer, K. (1988) *J. Biol. Chem.* **263**, 9557–9560
186. Krusius, T., Gehlsen, K. R., and Ruoslahti, E. (1987) *J. Biol. Chem.* **262**, 13120–13125
187. Gitt, M. A., and Barondes, S. H. (1991) *Biochemistry* **30**, 82–89
188. Barondes, S. H., Cooper, D. N. W., Gitt, M. A., and Leffler, H. (1994) *J. Biol. Chem.* **269**, 20807–20810
188a. Varela, P. F., Solís, D., Díaz-Maurino, T., Kaltner, H., Gabius, H.-J., and Romero, A. (1999) *J. Mol. Biol.* **294**, 537–549
189. Chiu, M. L., Parry, D. A. D., Feldman, S. R., Klapper, D. G., and O'Keefe, E. J. (1994) *J. Biol. Chem.* **269**, 31770–31776
190. Madsen, P., Rasmussen, H. H., Flint, T., Gromov, P., Kruse, T. A., Honoré, B., Vorum, H., and Celis, J. E. (1995) *J. Biol. Chem.* **270**, 5823–5829
191. Cho, M., and Cummings, R. D. (1995) *J. Biol. Chem.* **270**, 5207–5212
192. Gitt, M. A., Wiser, M. F., Leffler, H., Herrmann, J., Xia, Y.-R., Massa, S. M., Cooper, D. N. W., Lusis, A. J., and Barondes, S. H. (1995) *J. Biol. Chem.* **270**, 5032–5038
193. Weis, W. I., Kahn, R., Fourme, R., Drickamer, K., and Hendrickson, W. A. (1991) *Science* **254**, 1608–1615
194. Vyas, N. K., Vyas, M. N., and Quiocho, F. A. (1988) *Science* **242**, 1290–1295
195. Shaanan, B., Lis, H., and Sharon, N. (1991) *Science* **254**, 862–865
196. Cygler, M., Rose, D. R., and Bundle, D. R. (1991) *Science* **253**, 442–445
197. Ångström, J., Teneberg, S., and Karlsson, K.-A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11859–11863
198. Sauter, N. K., Hanson, J. E., Glick, G. D., Brown, J. H., Crowther, R. L., Park, S.-J., Skehel, J. J., and Wiley, D. C. (1992) *Biochemistry* **31**, 9609–9621
199. Janakiraman, M. N., White, C. L., Laver, W. G., Air, G. M., and Luo, M. (1994) *Biochemistry* **33**, 8172–8179
200. Bossart-Whitaker, P., Carson, M., Babu, Y. S., Smith, C. D., Laver, W. G., and Air, G. M. (1993) *J. Mol. Biol.* **232**, 1069–1083
201. Crennell, S. J., Garman, E. F., Laver, E. G., Vimr, E. R., and Taylor, G. L. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9852–9856
202. Colli, W. (1993) *FASEB J.* **7**, 1257–1264
203. Ofek, I., Beachey, E. H., and Sharon, N. (1978) *Trends Biochem. Sci.* **3**, 159–160
204. Lindberg, F., Lund, B., and Normark, S. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1891–1895
205. Cho, J.-W., and Troy, F. A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11427–11431
206. Baumann, H., Brisson, J., Michon, F., Pon, R., and Jennings, H. J. (1993) *Biochemistry* **32**, 4007–4013
207. Falk, P. G., Bry, L., Holgersson, J., and Gordon, J. I. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1515–1519
208. Adler, P., Wood, S. J., Lee, Y. C., Lee, R. T., Petri, W. A., and Schnaar, R. L. (1995) *J. Biol. Chem.* **270**, 5164–5171
209. Misevic, G. N., and Burger, M. M. (1990) *J. Biol. Chem.* **265**, 20577–20584
210. Spillmann, D., Hård, K., Thomas-Oates, J., Vliegenthart, J. F. G., Misevic, G., Burger, M. M., and Finne, J. (1993) *J. Biol. Chem.* **268**, 13378–13387
211. Spillmann, D., Thomas-Oates, J. E., van Kuik, J. A., Vliegenthart, J. F. G., Misevic, G., Burger, M. M., and Finne, J. (1995) *J. Biol. Chem.* **270**, 5089–5097
212. Rosati, F., and De Santis, R. (1980) *Nature (London)* **283**, 762–764
213. Seppo, A., Penttilä, L., Niemelä, R., Maaheimo, H., and Renkonen, O. (1995) *Biochemistry* **34**, 4655–4661
214. Litscher, E. S., Juntunen, K., Seppo, A., Penttilä, L., Niemelä, R., Renkonen, O., and Wassarman, P. M. (1995) *Biochemistry* **34**, 4662–4669

## References


215. Hitoshi, S., Kusunoki, S., Kanazawa, I., and Tsuji, S. (1995) *J. Biol. Chem.* **270**, 8844–8850
216. de Vries, T., Srnka, C. A., Palcic, M. M., Swiedler, S. J., van der Eijnden, D. H., and Macher, B. A. (1995) *J. Biol. Chem.* **270**, 8712–8722
217. Stroud, M. R., Levery, S. B., Mårtensson, S., Salyan, M. E. K., Clausen, H., and Hakomori, S. (1994) *Biochemistry* **33**, 10672–10680
218. Hull, S. R., Sugarman, E. D., Spielman, J., and Carraway, K. L. (1991) *J. Biol. Chem.* **266**, 13580 –13586
219. Ware, F. E., Vassilakos, A., Peterson, P. A., Jackson, M. R., Lehrman, M. A., and Williams, D. B. (1995) *J. Biol. Chem.* **270**, 4697–4704
220. Leppänen, A., Mehta, P., Ouyang, Y.-B., Ju, T., Helin, J., Moore, K. L., van Die, I., Canfield, W. M., McEver, R. P., and Cummings, R. D. (1999) *J. Biol. Chem.* **274**, 24838 – 24848
221. Chandrasekaran, E. V., Jain, R. K., Larsen, R. D., Wlasichuk, K., and Matta, K. L. (1995) *Biochemistry* **34**, 2925–2936
222. Kogan, T. P., Revelle, B. M., Tapp, S., Scott, D., and Beck, P. J. (1995) *J. Biol. Chem.* **270**, 14047–14055
223. Waddell, T. K., Fialkow, L., Chan, C. K., Kishimoto, T. K., and Downey, G. P. (1995) *J. Biol. Chem.* **270**, 15403–15411
224. Bertozzi, C. R., Fukuda, S., and Rosen, S. D. (1995) *Biochemistry* **34**, 14271–14277
225. Norgard-Sumnicht, K., and Varki, A. (1995) *J. Biol. Chem.* **270**, 12012–12024
226. Hemmerich, S., Leffler, H., and Rosen, S. D. (1995) *J. Biol. Chem.* **270**, 12035–12047
227. Yuen, C.-T., Bezouska, K., O'Brien, J., Stoll, M., Lemoine, R., Lubineau, A., Kiso, M., Hasegawa, A., Bockovich, N. J., Nicolaou, K. C., and Feiz, T. (1994) *J. Biol. Chem.* **269**, 1595–1598
228. Patel, T. P., Goelz, S. E., Lobb, R. R., and Parekh, R. B. (1994) *Biochemistry* **33**, 14815–14824
229. Graves, B. J., Crowther, R. L., Chandran, C., Rumberger, J. M., Li, S., Huang, K.-S., Presky, D. H., Familletti, P. C., Wolitzky, B. A., and Burns, D. K. (1994) *Nature (London)* **367**, 532–538
230. Lasky, L. A. (1995) *Ann. Rev. Biochem.* **64**, 113–139
231. Ginsburg, V. E. (1982) *Meth. Enzymol.* **83**, all
232. Churms, S. C. (1990) *CRC Handbook of Chromatography Carbohydrates*, Vol. 2, CRC Press, Boca Raton, Florida
233. Valent, B. S., Darvill, A. G., McNeil, M., Robertsen, B. K., and Albersheim, P. (1980) *Carbohydrate Res.* **79**, 165–192
234. Chaplin, M. F., and Kennedy, J. F., eds. (1994) *Carbohydrate Analysis A Practical Approach*, Oxford University, Oxford
235. Robyt, J. F., and Mukerjea, R. (1994) *Carbohydr. Res.* **251**, 187–202
236. Green, E. D., and Baenziger, J. U. (1989) *Trends Biochem. Sci.* **14**, 168–172
237. Carr, S. A., Huddleston, M. J., and Bean, M. F. (1993) *Protein Sci.* **2**, 183–196
237a. Fridriksson, E. K., Beavil, A., Holowka, D., Gould, H. J., Baird, B., and McLafferty, F. W. (2000) *Biochemistry* **39**, 3369–3376
237b. Burlingame, A. L., Carr, S. A., and Baldwin, M. A., eds. (2000) *Mass Spectrometry in Biology and Medicine*, Hamana Press, Totowa, New Jersey
238. Downs, F., and Pigman, W. (1969) *Biochemistry* **8**, 1760–1766
239. Narasimhan, S., Harpaz, N., Longmore, G., Carver, J. P., Grey, A. A., and Schacter, H. (1980) *J. Biol. Chem.* **255**, 4876–4884
240. Cohen, R. E., and Ballou, C. E. (1980) *Biochemistry* **19**, 4345–4358
241. Plummer, T. H., Jr., and Tarentino, A. L. (1981) *J. Biol. Chem.* **256**, 10243–10246
242. Risley, J. M., and Van Etten, R. L. (1985) *J. Biol. Chem.* **260**, 15488–15494
243. Takasaki, S., Mizuochi, T., and Kobata, A. (1982) *Meth. Enzymol.* **83**, 263–268
244. Parente, J. P., Wieruszeski, J.-M., Strecker, G., Montreuil, J., Fournet, B., van Halbeek, H., Dorland, L., and Vliegenthart, J. F. G. (1982) *J. Biol. Chem.* **257**, 13173–13176
245. Patel, T., Bruce, J., Merry, A., Bigge, C., Wormald, M., Jaques, A., and Parekh, R. (1993) *Biochemistry* **32**, 679–693
246. Sweeley, C. C., and Nunez, H. A. (1985) *Ann. Rev. Biochem.* **54**, 765–801
247. Dua, V. K., Roa, B. N. N., Wu, S.-S., Dube, V. E., and Bush, C. A. (1986) *J. Biol. Chem.* **261**, 1599–1608
248. Tsai, P.-K., Dell, A., and Ballou, C. E. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4119–4123
249. Lindberg, B., Lönngren, L., and Svensson, S. (1975) *Adv. Carbohydr. Chem. Biochem.* **31**, 185–240
250. Guthrie, R. D., and McCarthy, J. F. (1967) *Adv. Carbohydr. Chem. Biochem.* **22**, 11–23
251. Edge, C., Parekh, R., Rademacher, T., Wormald, M., and Dwek, R. (1992) *Nature (London)* **385**, 693–694
252. Varki, A. (1991) *FASEB J.* **5**, 226–235
253. Rolf, D., and Gray, G. R. (1982) *J. Am. Chem. Soc.* **104**, 3539–3541
254. Mukerjea, R., Kim, D., and Robyt, J. F. (1996) *Carbohydrate Res.* **292**, 11–20
255. Bobbitt, J. M. (1956) *Adv. Carbohydr. Chem. Biochem.* **11**, 1–41
256. Rich, A., Nordheim, A., and Wang, A. H.-J. (1984) *Ann. Rev. Biochem.* **53**, 791–846
257. Homans, S. W., Dwek, R. A., and Rademacher, T. W. (1987) *Biochemistry* **26**, 6571–6578
258. Homans, S. W., Pastore, A., Dwek, R. A., and Rademacher, T. W. (1987) *Biochemistry* **26**, 6649 –6655
259. Cumming, D. A., Shah, R. N., Krepinsky, J. J., Grey, A. A., and Carver, J. P. (1987) *Biochemistry* **26**, 6655–6663
260. Reuben, J. (1985) *J. Am. Chem. Soc.* **107**, 1747–1755
261. Cumming, D. A., Dime, D. S., Grey, A. A., Krepinsky, J. J., and Carver, J. P. (1986) *J. Biol. Chem.* **261**, 3208–3213
262. Vuister, G. W., de Waard, P., Boelens, R., Vliegenthart, J. F. G., and Kaptein, R. (1989) *J. Am. Chem. Soc.* **111**, 772–774
263. Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry* **28**, 5978–5984
264. Abeygunawardana, C., and Bush, C. A. (1991) *Biochemistry* **30**, 8568–8577
265. Rivière, M., and Puzo, G. (1992) *Biochemistry* **31**, 3575–3580
266. van Duynhoven, J. P. M., Goudriaan, J., Hilbers, C. W., and Wijmenga, S. S. (1992) *J. Am. Chem. Soc.* **114**, 10055–10056
267. Wu, J., Bondo, P. B., Vuorinen, T., and Serianni, A. S. (1992) *J. Am. Chem. Soc.* **114**, 3499–3505
268. Homans, S. W. (1990) *Biochemistry* **29**, 9110–9118
269. Bose, B., Zhao, S., Stenutz, R., Cloran, F., Bondo, P. B., Bondo, G., Hertz, B., Carmichael, I., and Serianni, A. S. (1998) *J. Am. Chem. Soc.* **120**, 11158–11173
270. Flugge, L. A., Blank, J. T., and Petillo, P. A. (1999) *J. Am. Chem. Soc.* **121**, 7228 – 7238

## Study Questions

1. A nonreducing disaccharide gives an octamethyl derivative with dimethyl sulfate and alkali. On acid hydrolysis, this derivative yields 1 mol of 2,3,4,6-tetramethyl-D-glucose and 1 mol of 2,3,4,6-tetramethyl-D-galactose. The disaccharide is hydrolyzed rapidly by either maltase or lactase (a β-galactosidase).

   Give an adequately descriptive name of the disaccharide, and draw its Haworth projection formula.

2. An aldopentose (A) of the D-configuration on oxidation with concentrated nitric acid gives a 2,3,4-trihydroxypentanedioic acid (a trihydroxyglutaric acid) (B) which is optically inactive. (A) on addition of HCN, hydrolysis, lactonization, and reduction gives two stereoisomeric aldohexoses (C) and (D). (D) on oxidation affords a 2,3,4,5-tetrahydroxy-hexanedioic acid (a saccharic acid) (E) which is optically inactive. Give structures of compounds (A)-(E).

3. What products are formed when periodic acid reacts with sorbitol?

4. A 10.0 g sample of glycogen gave 6.0 millimol of 2,3-di-*O*-methylglucose on methylation and acid hydrolysis.

   a. What percent of the glucose residues in glycogen have chains substituted at the α–1→6 position?

   b. What is the average number of glucose residues per chain?

   c. How many millimols of 2,3,6-tri-*O*-methylglucose were formed?

   d. If the molecular weight of the polysaccharide is $2 \times 10^6$, how many glucose residues does it contain?

   e. How many nonreducing ends are there per molecule or equivalently how many chains are there per molecule?

5. D-Mannitol is a symmetric molecule, yet it is optically active. Explain.

6. When D-glucose is treated with acidic methanol, the first products which can be isolated are mainly methyl furanosides, but after extensive reaction the furanosides disappear and methyl glucopyranosides accumulate. Why?

7. Write the structural formulas for (a) 1,6 anhydro β-D-glucopyranose; (b) 1,6 anhydro β-D-altrose. Compound (b) is many times more stable than compound (a). Explain this on stereochemical grounds.

8. The disaccharide *nigerose* is α-D-Glu*p*-(1→3)-D-glu. Write out its structure. How would you prove this structure using methylation, periodate oxidation, and other methods.

9. Inositol is 1,2,3,4,5,6-hexahydroxycyclohexane. Draw configurational formulas for all possible stereoisomers and indicate which would be expected to be optically active.

10. Why do you suppose that the major form of D-fructose in solution is the pyranose form but D-fructose in sucrose is in the furanose form?

11. D-Xylose is an easily prepared sugar, potentially available in enormous quantity. What is a common source? How can it be obtained from this source?

12. How can xylitol be obtained from xylose? Discuss the stereochemical properties of xylitol.

13. The enzyme xylose isomerase is important industrially. Why?

14. Glucose reacts non-enzymatically with amino acids and proteins, including hemoglobin, egg-white proteins and serum albumin. For example, if glucose is not removed prior to drying, dried egg whites slowly turn brown and develop off-flavors and odors. What do you propose as the most likely first step in the non-enzymatic glucose-protein chemical reaction? How can the first product transform spontaneously into a ketose derivative?

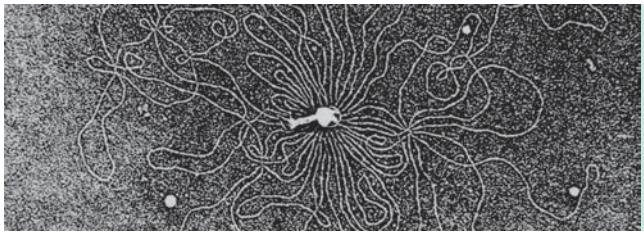
15. What characteristics would you expect in a binding site for a sugar ring in an enzyme, lectin, or other proteins?

16. Using structural formulas, descirbe the two major types of linkage of carbohydrate chains or clusters to proteins to form glycoproteins or proteoglycans.

17. What products would you expect from cellulose as a result of methylation analysis? Periodate oxidation? The Smith degradation? The action of an alpha amylase?

## Study Questions

18. When glycoproteins are treated with alkaline brohydride, amino acid analysis often indicates a decrease in the amount of serine and a corresponding increase in the amount of alanine, or a decrease in threonine with the appearance of α-aminobutyric acid. Explain.



DNA spreading from the broken head of a bacteriophage T2 phage. This classic electron micrograph, published by A. K. Kleinschmidt and coworkers in 1962 (*Biochem. Biophys. Acta.* **61**, 857–864, 1962) was prepared by spreading the phage particles suspended in a protein–salt solution as a mixed monolayer on a water-air interface. The resultant osmotic shock burst the head and confined the DNA as a single thread near the phage ghost. After transfer to a suitable carbon surface, removal of water, and shadowing with platinum, the micrograph was obtained. Courtesy of Albrecht K. Kleinschmidt

# Contents


203 **A. Structure and Chemistry of Nucleotides**
203 1. Names and Abbreviations
204 2. Acid–Base Chemistry and Tautomerism
205 3. Absorption of Ultraviolet Light
207 4. Chemical Reactions of the Bases
207 5. Base Pairs, Triplets, and Quartets
209 *Strengths of base pairs*
209 *Stacking of bases*
211 *Tautomerism and base pairing*
211 6. Conformations of Nucleotides
213 **B. Double Helices**
213 1. The B Form of DNA
213 2. Other Double-Helical Forms of DNA
216 3. The Conformational Flexibility of Double Helices
217 *Rotational and translational movements of bases*
217 *Bends and bulges*
218 *Interactions with ions*
218 **C. The Topology and Dynamics of Nucleic Acids**
218 1. Rings, Catenanes, and Knots
219 2. Supercoiled DNA
222 3. Intercalation
226 4. Polynucleotides with Three or Four Strands
226 *H-DNA*
226 *R-DNA*
227 *Designed third strands*
227 *Tetraplex (quadruplex) structures*
228 5. Junctions
230 **D. Ribonucleic Acids (RNA)**
230 1. RNA Loops and Turns
231 2. Transfer RNAs
232 3 . Properties and Structures of Some Other RNA Molecules
234 4. Modified Nucleotides
234 5. RNA Aptamers
235 **E. The Genetic Code**
236 1. The "Reading Frames"
237 2. Variations
238 3. Palindromes and Other Hidden Messages
239 4. The Base Composition of DNA and RNA
239 **F. Interaction of Nucleic Acids with Proteins**
239 1. The Helix–Turn–Helix Motif
241 2. Other DNA-binding Motifs
241 *Leucine zipper proteins*
241 *Zinc fingers*
241 *Beta ribbons*
243 *The winged helix family*
243 3. RNA-Binding Proteins
244 **G. Viruses and Plasmids**
244 1. Viruses with Single-Stranded DNA
244 2. Viruses with Double-Stranded DNA
247 3. Viruses Containing RNA
248 4. Viruses without Nucleic Acid?
248 5. Life Cycles
248 6. Plasmids and Transposable Genetic Elements
249 **H. Methods of Study**
249 1. Isolation and Separation of Nucleic Acids
249 2. Hydrolysis and Analysis
251 3. Characteristic Reactions of the Bases and Backbone
251 *Reactions of nucleophiles*
253 *Reactions with electrophilic reagents*
254 *Reactions causing cleavage of the sugar-phosphate backbone*
255 4. Melting, Hybridization, and Polynucleotide Probes
258 5. Synthesis of Oligonucleotides and Polynucleotides
260 6. The Polymerase Chain Reaction (PCR)
260 7. Sequence Determination
260 *Preparing the DNA*
260 *Restriction maps and Southern blots*
262 *The Sanger dideoxy method*
264 *The method of Maxam and Gilbert*
264 *Sequencing RNA*
265 *Nearest neighbor analysis*
265 *Understanding sequences*
266 8. Protein–DNA Interactions
266 9. Nuclear Magnetic Resonance

270 **References**
278 **Study Questions**


**Boxes**

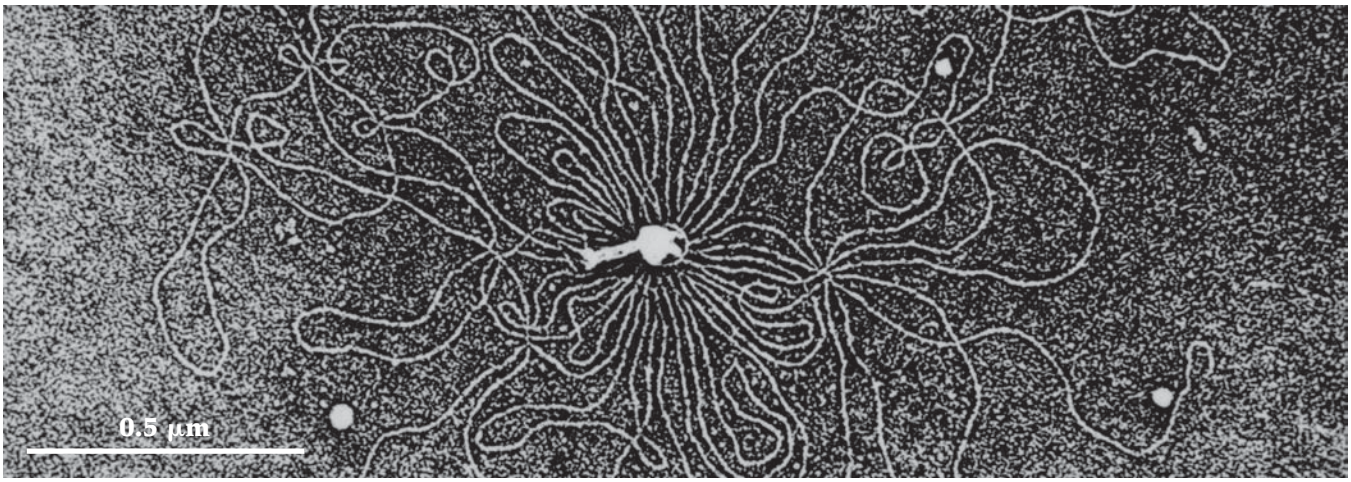

206 Box 5-A The Alkali Metal Ions
224 Box 5-B Antitumor DNA Drugs
235 Box 5-C The RNA World
259 Box 5-D DNA Fingerprinting


**Tables**


203 Table 5-1 Names of Pyrimidine and Purine Bases, Nucleosides, and 5′-Nucleotides
210 Table 5-2 Gibbs Energies of Formation $\Delta G_f°$ at 25°C for Addition of One Base Pair to an Existing RNA Helix
216 Table 5-3 Helix Parameters for Three Types of DNA
232 Table 5-4 Some Specialized RNA Molecules
236 Table 5-5 The Genetic Code
237 Table 5-6 The Sixty-Four Codons of the Genetic Code
245 Table 5-7 Characteristics of Some Individual Viruses and Groups of Viruses

# The Nucleic Acids

# 5




The phosphorus- and nitrogen-containing materials that came to be known as nucleic acids were first isolated from cells around 1870 by Friedrich Miescher but were long regarded as something of a curiosity.[1] Nevertheless, the structures of the monomer units, the **nucleotides**, were established by 1909 and the correct **polynucleotide** structure of the chains of DNA and RNA was proposed by Levene and Tipson in 1935.[2,3]

The nucleotides are made up of three parts:

1. One of the **pyrimidine** or **purine "bases"** uracil, cytosine, adenine, or guanine (Fig. 5-1). All four of these bases are present in RNA, while DNA contains thymine instead of uracil. Atoms in the bases are numbered 1–6 or 1–9.
2. A **sugar**, either D-ribose or D-2-deoxyribose. Carbon atoms in sugars are numbered 1′–5.′
3. **Phosphoric acid**

Although the biological synthesis is indirect, we can imagine that nucleotides are formed from these parts by elimination of two molecules of water as indicated in Eq. 5-1. In nucleic acids the nucleotides are combined through phosphodiester linkages between the 5'-hydroxyl of the sugar in one nucleotide and the 3'-hydroxyl of another. Again, we can imagine that these linkages were formed by the elimination of water (Eq. 5-2). The structures of a pair of short polynucleotide strands in DNA are shown in Fig. 5-2. That of a segment of double-helical DNA is shown in Fig. 5-3 and that of a transfer RNA in Figs. 5-30 and 5-31.

Despite the fact that Levene had deduced the correct structure for polynucleotides, he was thrown off the trail of a deeper understanding by the roughly equal amounts of the four bases found in either DNA or RNA. He assumed that nucleic acids must be regular repeating polymers for which there was no obvious biological function. It was not until 1944 that there

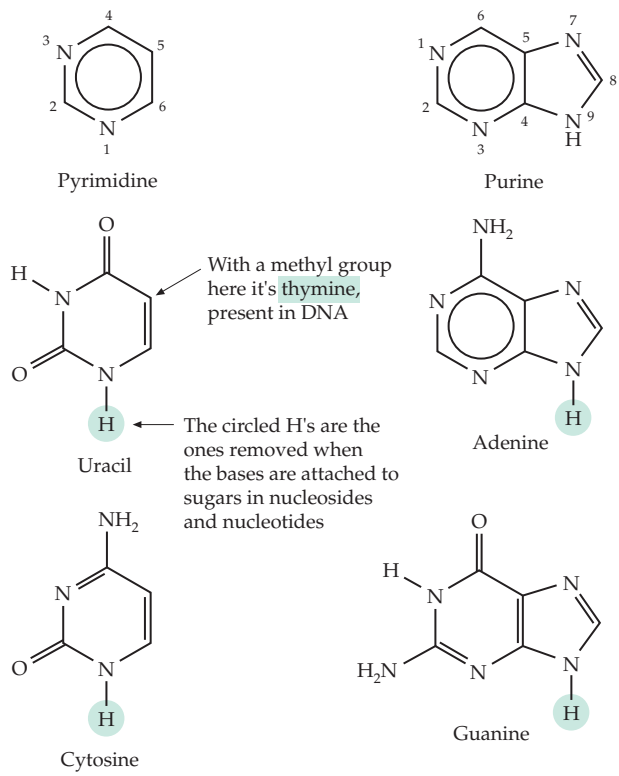
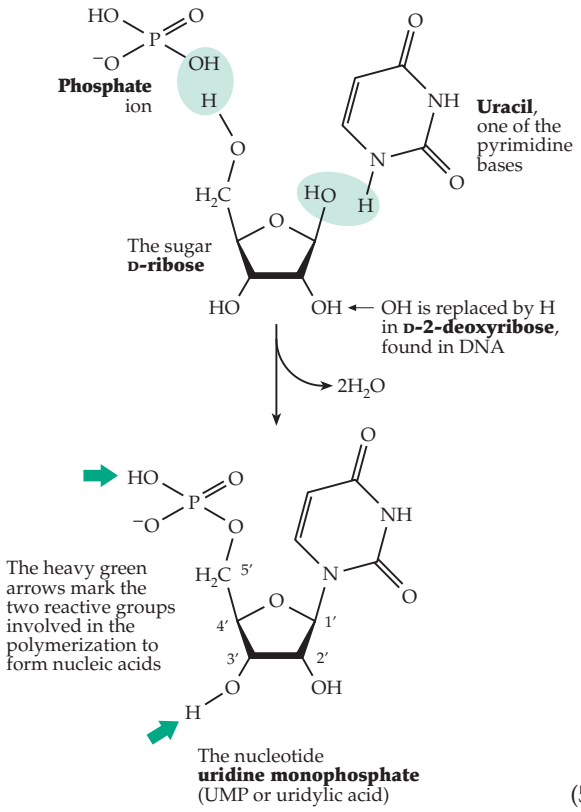
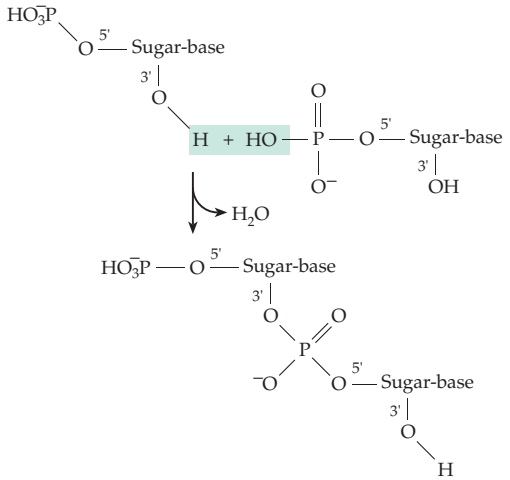



**Figure 5-1** Structures of the major pyrimidine and purine bases of DNA and RNA.




(5-1)




(5-2)

was concrete evidence that DNA carried genetic information (see Chapter 26). However, it was James Watson and Francis Crick's recognition of the double-helical structure of DNA[4-9] in 1953 and the mechanism of replication that this structure implied that captured the imagination of biologists and chemists alike and paved the way for the present-day explosion of knowledge of DNA, RNA, and of the encoded proteins.

Watson and Crick proposed that DNA is a double helix of two antiparallel polynucleotide chains (Figs. 5-2 and 5-3). The structure was deduced from model building together with knowledge of the X-ray diffraction data of Maurice F. Wilkins and Rosalind Franklin[9a] on artificially formed DNA fibers. An additional key piece of information was the discovery by Erwin Chargaff that *in all double-stranded DNA the content of adenine equals that of thymine and the content of guanine equals that of cytosine.*

The most significant feature of the proposed structure was the pairing of bases in the center through hydrogen bonding. The pairs and triplets of hydrogen bonds (Fig. 5-2) could form in the manner shown only if adenine (A) was paired with thymine (T) and cytosine (C) with guanine (G) at every point in the entire DNA structure. Thus, *the nucleotide sequence in one chain is complementary to but not identical to that in the other chain.* It was apparent almost immediately that the sequence of bases in a DNA chain must convey the encoded genetic information. The complementarity of the two strands suggested a simple mechanism for replication of genes during cell divisions. The two strands could separate and a complementary strand could be synthesized along each strand to give two molecules of the DNA, one for each of the two cells.

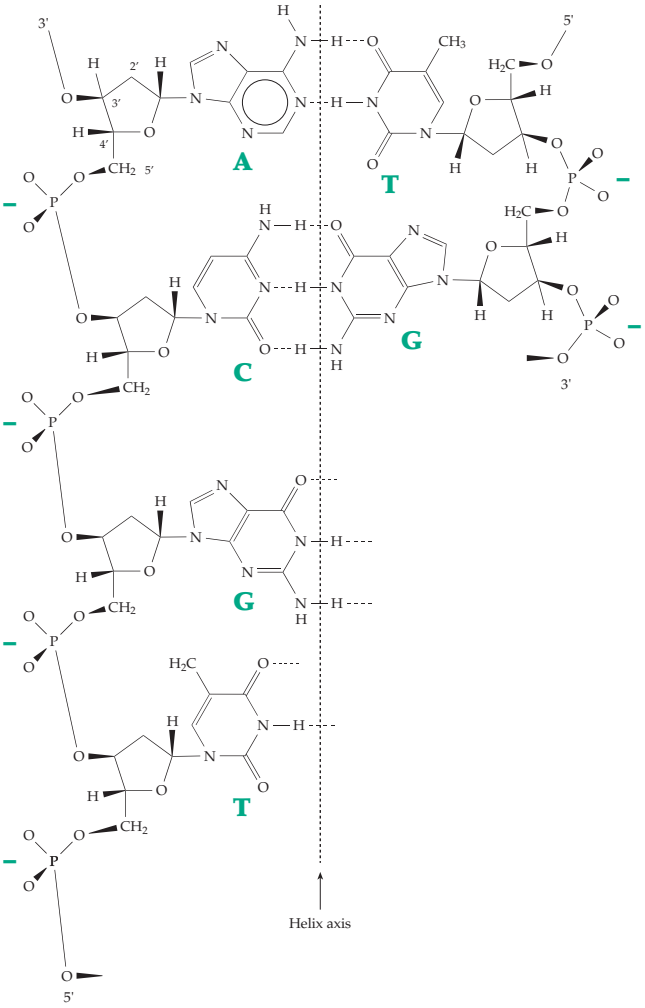



**Figure 5-2** A distorted (flattened) view of the Watson–Crick structure of DNA showing the hydrogen-bonded base pairs.

The essential correctness of the concept has been proved.

Two extremely important developments came in the 1970s: (1) Methods were found for cutting and rejoining DNA fragments and for **cloning** them in bacteria and (2) ways were devised for rapid determination of **nucleotide sequences**. The application of these techniques is now providing startling advances in biology and medicine. In 1970 we knew virtually nothing about the sequences of nucleotides in genes but today, we know the sequences for many thousands of genes of all types. By the 1980s sequences had been established for hemoglobin,[11] γ-globulins,[12] collagen,[13] and for many enzymes. An example is the gene sequence for mitochondrial aspartate aminotransferase (Fig. 5-4).[14] Its coding regions consist of 1299 pairs of mononucleotides, the bases being paired as in Fig. 5-2 in double helical form. Complete DNA sequences are known for numerous viruses including the 9740-base pair (bp) DNA provirus form of the RNA virus that causes AIDS[15] and for bacterial viruses such as T7 bacteriophage (39,936 bp).[16] Also determined in the 1980s were sequences of human mitochondrial DNA (16,598 bp),[17] and of chloroplast DNA from the tobacco plant (155,844 bp).[18]

In 1995 the first complete sequences of bacterial genomes were obtained (Table 1-3).[19,20] These were followed by sequences of many other bacterial genomes,[21] including the 4.2 x $10^6$ bp *E. coli* genome (Table 1-3). Sequencing of the 16 chromosomes of the 12.07 Mbp genome of yeast[22,23] containing ~6300 genes was completed in 1996[24] and that of the 97 Mbp genome of the nematode *Caenorhabditis elegans* in 1998. By 2000 the sequence of the 180 Mbp genome of the fruit fly *Drosophila melanogaster* was largely completed, and most sequences of the 3 x $10^9$ bp human genome were known.[24a]

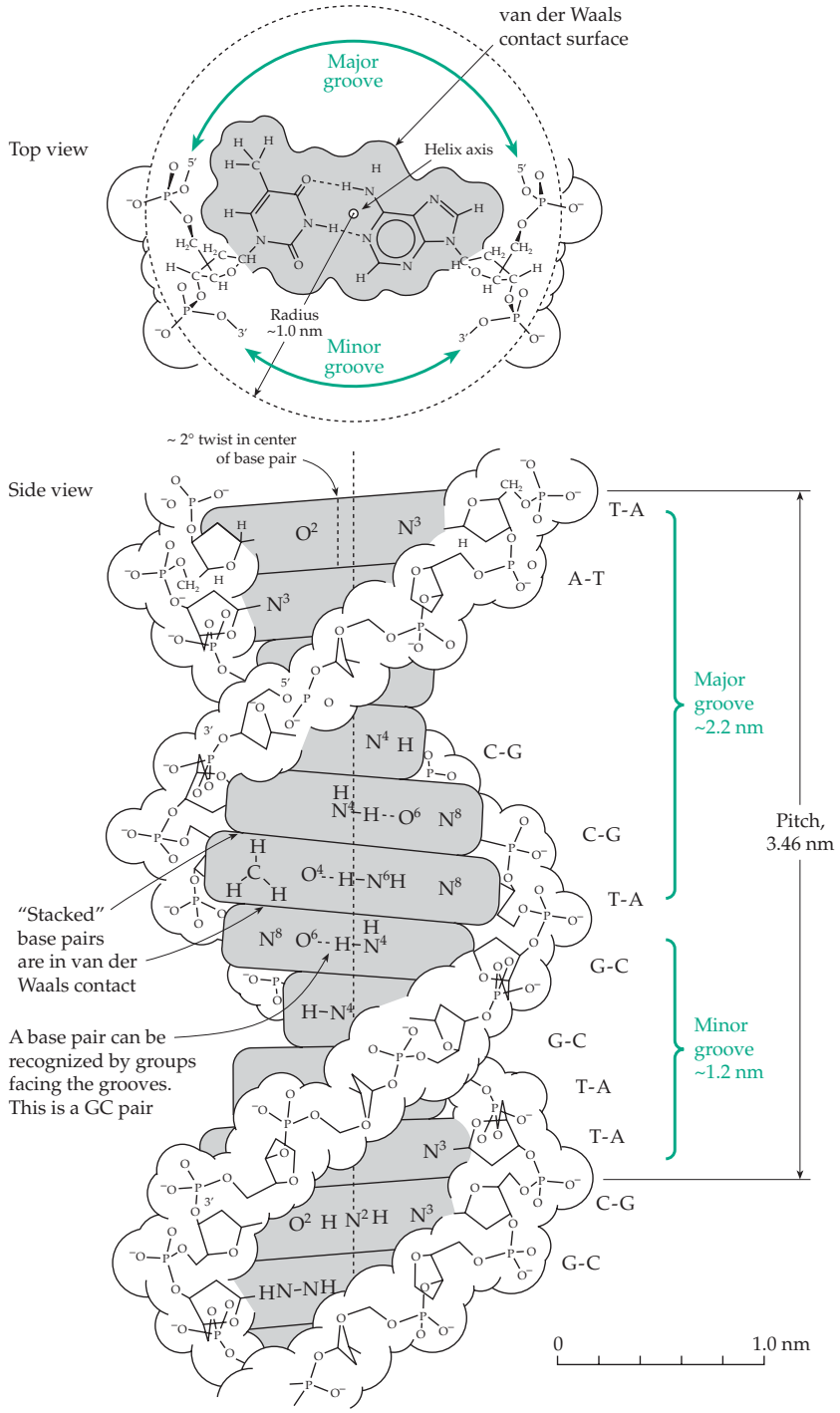



**Figure 5-3** The double-helical structure of DNA. The structure shown is that of the B form and is based on coordinates of Arnott and Hukins.[10] The major and minor grooves, discussed on p. 213, are marked.

Cloning of DNA has not only provided an essential step in sequence determination but also has given birth to a new industry devoted to producing proteins from genes cloned in bacteria, yeast, or other cells. Human insulin and the antiviral protein interferon were two of the first proteins produced in this way. Methods now in use permit us to introduce at will alterations at any point in a DNA sequence. We are able to locate and study the genes responsible for many genetic defects.

```
                                  -200         .         .         .         .       -150       .         .         .
                              5'- CATGACCTCCGTGGAGGCGTCGTCGTTCTACCCCCGACCGCAAGGGGCCGCTGACATGATTCGCTCCAGCGCAGGGCCGCT
        .       -100         .         .         .         .       -50         .         .         .         .       +1
            ↓         ↓         ↓     ↓                         ↓                                                      Me
CTAGCCACGCCCAGGGAAGTCACTGTCCTCACCTTTTAGGAGCCCGCGCCTCGGTTCCAGCGGACGCTTCCCCAGATCTCGGCTCTACCACCATCCACTGCCGTCTTACCGCCCACC AT

t Ala Leu Leu His Ser Ser Arg Ile Leu Ser Gly Met Ala Ala Ala Phe His Pro Gly Leu Ala Ala Ala Ala Ser Ala Arg Ala Se
G GCC CTC CTG CAC TCC AGC CGC ATC CTC TCC GGG ATG GCT GCT GCC TTT CAC CCT GGC CTA GCT GCT GCA GCC TCT GCC AGA GCC AG GTG
                                                                                  A
                                                                                INTRON 1                                r Ser
AGCCGAGGGATACAGAGATGCAGCGGCACCGGGCCTGCCCTGCCAGCCGAAGTGTGGACCCTGTGAA----(10.2 kb)----ATGGCTGGGGTAACCTATTTCTCATTTCAG C TCC

Trp Trp Thr His Val Glu Met Gly Pro Pro Asp Pro Ile Leu Gly Val Thr Glu Ala Phe Lys Arg Asp Thr Asn Ser Lys Lys Met Asn
TGG TGG ACC CAT GTT GAA ATG GGA CCT CCA GAT CCC ATC CTG GGC GTT ACC GAA GCC TTC AAG AGA GAT ACC AAC AGC AAG AAG ATG AAC

Leu Gly Val Gly Ala Tyr Arg Asp Asp Asn Gly Lys Pro Tyr Val Leu Pro Ser Val Arg Lys A
CTG GGA GTT GGT GCC TAC CGG GAT GAT AAC GGA AAA CCT TAC GTG CTC CCC AGT GTC CGG AAG G GTGAGCTTGGCACTCGTCTCCTGCCAGCTAGGAT

          INTRON 2                                    la Glu Ala Gln Ile Ala Ala Lys Asn Leu Asp Lys Glu Tyr Leu Pro Ile
GTGGAA----(1.4 kb)----AAAAACTAAGACTTATGATTTTCTTGTCTAG CA GAG GCC CAG ATT GCT GCA AAA AAT TTG GAC AAA GAA TAC CTG CCC ATT

Gly Gly Leu Ala Glu Phe Cys Lys Ala Ser Ala Glu Leu Ala Leu Gly Glu Asn Asn Glu Val Asn Glu Val Leu Lys Ser Gly Arg
GGG GGA CTG GCT GAA TTC TGT AAG GCT TCT GCA GAA CTG GCC CTG GGC GAG AAC AAT GAA GTG AAT GAA GTG TTG AAA AGC GGC CGG GTAA

                                INTRON 3                                                  Phe Val Thr Val Gln Thr Ile Ser Gly
GCCAGCGGAGTCGGGCTTGAGCTTGATA----(1.2 kb)----TGTAAAGGGTAGAGAGAGTGACTCTGTGTATCCCCTGCAG TTC GTC ACT GTG CAG ACC ATT TCC GGG

Thr Gly Ala Leu Arg Val Gly Ala Ser Phe Leu                                          INTRON 4
ACT GGA GCC TTA AGG GTC GGG GCC AGT TTT CTG GTCAGTGGAACTCTTTCAAGAATGAATCTTTTGGGGTGGC----(2.2 kb)----CTTTCTTCATTTTCTCATTC

                     Gln Arg Phe Phe Lys Phe Ser Arg Asp Val Phe Leu Pro Lys Pro Ser Trp Gly Asn His Thr Pro Ile Phe Arg
CCTTTTATCCCGACTTTTAG CAA AGG TTT TTT AAG TTC AGC CGA GAT GTC TTT CTG CCC AAA CCA TCC TGG GGA AAT CAC ACG CCC ATC TTC AGG

Asp Ala Gly Met Gln Leu Gln Gly Tyr Arg Tyr Tyr Asp Pro Lys Thr Cys Gly Phe Asp Phe Ser Gly Ala Leu Glu Asp Ile Ser
GAT GCC GGC ATG CAG CTC CAA GGT TAT CGC TAC TAT GAC CCC AAG ACT TGC GGT TTT GAC TTC TCC GGA GCC CTA GAA GAC ATA TCA GTAA

                              INTRON 5                                            Lys Ile Pro Glu Gln Ser Val Leu Leu Leu
GTGTGGCTTTCCAGGCCCGACTTCTG----(0.2 kb)----GAAGCTGCACAGCCAAAAATCTCGATGTTTCTCCTTAG AAA ATC CCA GAG CAG AGT GTC CTC CTC CTG

His Ala Cys Ala His Asn Pro Thr Gly Val Asp Pro Arg Pro Glu Gln Trp Lys Glu Ile Ala Ser Val Val Lys
CAT GCC TGC GCT CAC AAC CCC ACC GGC GTG GAC CCG CGT CCC GAG CAG TGG AAG GAG ATA GCG TCC GTG GTG AAG GTGAGGAGGATGAAGCGTCA

                       INTRON 6                                     Lys Lys Asn Leu Phe Ala Phe Phe Asp Met Ala Tyr Gln Gly
GGAGCTGGTTGCTTAACC----(0.2 kb)----TCAAACACTGGTCTTCTCATTCCTTCCCAG AAA AAG AAT CTC TTC GCA TTC TTT GAC ATG GCC TAC CAA GGC

Phe Ala Ser Gly Asp Gly Asp Lys Asp Ala Trp Ala Val Arg His Phe Ile Glu Gln Gly Ile Asn Val Cys Leu Cys Gln Ser Tyr Ala
TTT GCC AGC GGT GAT GGT GAT AAG GAT GCC TGG GCC GTG CGG CAC TTC ATC GAG CAG GGC ATC AAT GTC TGC CTC TGC CAA TCG TAT GCC

Lys Asn Met Gly Leu Tyr G                                   INTRON 7                                                ly Glu
AAG AAC ATG GGC CTG TAC G GTAAGCCAGAGGGCCCAATATAAGGTGTGTT----(1.2 kb)----GTTTCAGCCTGCTAGAAGACAGTATCCTCTGTCTTTTCAG GT GAG

Arg Val Gly Ala Phe Thr Val Val Cys Lys Asp Ala Glu Glu Ala Lys Arg Val Glu Ser Gln Leu Lys Ile Leu Ile Arg Pro Leu Tyr
CGT GTG GGA GCC TTC ACG GTG GTC TGC AAA GAT GCA GAA GAA GCC AAA AGG GTG GAG TCA CAG CTG AAG ATC TTG ATC CGT CCC CTG TAT

Ser Asn Pro Pro Leu Asn Gly Ala Arg Ile Ala Ala Thr Ile Leu Thr Ser Pro Asp Leu Arg Lys Gln Tr
TCC AAC CCA CCT CTC AAT GGG GCC CGG ATC GCA GCA ACC ATC CTG ACT TCT CCA GAC TTC CGG AAG CAA TG GTAACGATTACTAGCTGTATACCGT

                     INTRON 8                                       p Leu Gln Glu Val Lys Gly Met Ala Asp Arg Ile Ile Ser Met
GACTACAGCTCCATGA----(2.4 kb)----GAGCACGGCATCCCTCTGCTTTCCTCACAG G TTG CAA GAG GTG AAA GGC ATG GCT GAC CGC ATC ATC AGC ATG

Arg Thr Gln Leu Val Ser Asn Leu Lys Lys Glu Gly Ser Ser His Asn Trp Gln His Ile Thr Asp Gln Ile Gly Met Phe Cys Phe Thr
AGG ACC CAG CTG GTC TCC AAC CTG AAG AAA GAG GGC TCT TCC CAC AAC TGG CAG CAC ATC ACC GAC CAG ATC GGC ATG TTC TGT TTC ACC

Gly Leu Lys Pro Glu Gln                                     INTRON 9                                              Val Glu
GGC CTA AAG CCA GAG CAG GTGAGTGGGCTCTGGTTCCCCACAGCCAACCCCCGC----(2.0 kb)----ACCAGAAGTGGTTATCTAGCTCTTTTCTTTCTGCAG GTA GAG

Arg Leu Thr Lys Glu Phe Ser Val Tyr Met Thr Lys Asp Gly Arg Ile Ser Val Ala Gly Val Thr Ser Gly Asn Val Gly Tyr Leu Ala
CGG CTG ACC AAG GAG TTC TCG GTC TAC ATG ACA AAG GAT GGC CGA ATC TCC GTG GCA GGG GTC ACC TCT GGC AAT GTG GGC TAC CTT GCC

His Ala Ile His Gln Val Thr Lys   *
CAT GCC ATT CAC CAG GTC ACC AAG TAA TTCCCAAGTGCAAAGGAACAGAGACCACTTTCCCGACAGCCTTTGCGCTCGTGAGCGTCACGTGCAGGGTGAGGGAGGGTGGGT
                                      G
GGTGGTGAATAGATCCGGTTTCCAACCACGGTGCATAACTCAGCGACTGAACTCGCTCCTCGGAAGAGAGGTAGGGCAGAGGCTTCCTCGGCTGGTATCTGGAACTTCGTCGGCTCTAAA

CCAAACTCCCTCATCCTTTTGTCTCCAGCTTTTCTGAAAGTTTACACACGCAAGAAAAATCACAGCACCAAACACCTGTCAGCCATGGCACTGGAATAGG|←   620 bases   →|
                                                                                   ↓
CACCATCTGCTCTAATCATGTAGACGTACTGCCGCCTGGTTTCTCTGTTACAATAAAATTACTATAGACCCAGCCACTGTCTGCTTTTCTACTTACTGAGAGGGAAGGGGTAAGAGTGGA -3'
```

**Figure 5-4** The nucleotide sequence of the gene for mitochondrial aspartate aminotransferase from the mouse. From Tsuzuki *et al.*[14] The gene encodes a 433-residue protein requiring 1299 nucleotide pairs (1.3 kb). However, 29 residues are cut off from the N terminus to form the mature mitochondrial protein. In addition, the gene is split by nine introns which vary in length from 0.2–10.2 kb. The sequences at the ends of the introns are shown. There is also an "upstream" region (of which 200 nucleotides are shown) at the 5'-end of the gene. It contains two binding sites for **transcription factor Sp1** (boxed). At the 3' end the 993 additional nucleotides contain signals (boxed) surrounding the **polyadenylation site** (green arrow) for 3′ processing and termination of transcription. The mature messenger RNA is about 2400 nucleotides (2.4 kb) in length but the gene, with introns, occupies about 25 kb. The +1 marks the position of the first nucleotide of the initiation codon ATG (encoding methionine) and the asterisk (*) the termination codon TAA. These are AUG and UAA in the mRNA. Other codons are indicated by amino acid abbreviations.

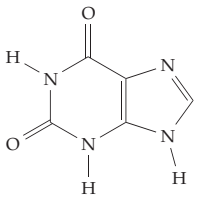
The first successes in using virus-like particles to carry new pieces of DNA into the cells to help correct these defects have been reported. Our ability to breed new varieties of plants and microorganisms has been enormously enhanced. We can foresee the production of artificially designed enzymes to conduct many industrial chemical processes. These are among the many reasons for the excitement today in the fields of nucleic acid chemistry and molecular genetics.

## A. Structure and Chemistry of Nucleotides

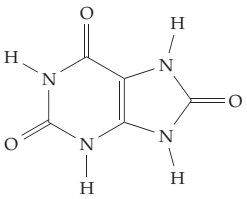
### 1. Names and Abbreviations

The purine and pyrimidine ring compounds found in nucleic acids are known as "bases," even though some of them have almost no basic character. **Nucleosides** are the *N*-glycosyl derivatives of the bases with ribose or 2-deoxyribose. The **nucleotides** are phosphate esters of nucleosides. Similar names are applied to related compounds such as adenosine triphosphate (ATP) that are not present in DNA or RNA. The names of the principal nucleotides from which the nucleic acids are formed are given in Table 5-1. The novice may find these confusing! Even worse than the names in the table is **hypoxanthine** (Hyp), which is derived from adenine by replacement of its $-NH_2$ group with –OH and tautomerization:

The nucleoside formed from hypoxanthine and ribose is known as **inosine** (Ino or I) and the corresponding nucleotide as **inosinic acid**. Further substitution at C-2 of –H by –OH and tautomerization yields **xanthine** (Xan). Its nucleoside is xanthosine (Xao, X). A similar hydroxylation at C-7 converts xanthine to **uric acid**, an important human urinary excretion product derived from nucleic acid bases.



Xanthine



Uric acid

**TABLE 5-1**
**Names of Pyrimidine and Purine Bases, Nucleosides, and 5'-Nucleotides[a]**

**A Nucleotide units of RNA (abbreviations in parentheses)[b]**

| | | | | |
|---|---|---|---|---|
| Base: | Uracil (Ura) | Cytosine (Cyt) | Adenine (Ade) | Guanine (Gua) |
| Nucleoside: | Uridine (Urd or U) | Cytidine (Cyd or C) | Adenosine (Ado or A) | Guanosine (Guo or G) |
| 5'-Nucleotide: | Uridine 5'-phosphate or 5'-uridylic acid (Urd-5'-*P* or UMP) | Cytidine 5'-phosphate or 5'-cytidylic acid (Cyd-5'-*P* or CMP) | Adenosine 5'-phosphate or 5'-adenylic acid (Ado-5'-*P* or AMP) | Guanosine 5'-phosphate or 5'-guanylic acid (Guo-5'-*P* or GMP) |

**B Nucleotide units of DNA**

These contain 2-deoxyribose and the nucleosides and nucleotides are called deoxyadenosine (dAdo or dA), deoxyadenosine 5'-phosphate (dAMP), etc.

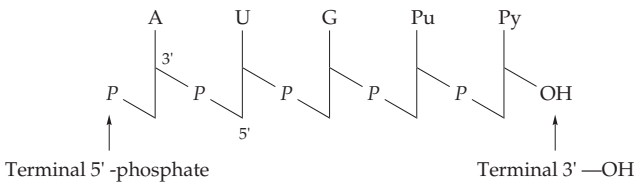
DNA contains thymine (Thy) rather than uracil. The deoxyribose derivatives are thymidine (dThd or dT) and thymidine 5'-phosphate. The ribose derivatives of thymine are the nucleoside ribosylthymidine (Thd) and ribosylthymidine 5'-phosphate (Thd-5'-*P*).

**C Abbreviations used in describing polynucleotide sequences**

| | |
|---|---|
| U,T,C,A,G | Uracil, thymine, cytosine, adenine, guanine |
| Y or Pyr | Pyrimidine (T or C) |
| R or Pur | Purine (A or G) |
| M | Amino base (A or C) |
| K | Keto base (G or T) |
| S | Strongly pairing (G or C) |
| W | Weakly pairing (A or T) |
| H | Not G (any other base) |
| B | Not A |
| V | Not T or U |
| D | Not C |
| N | Any base |

[a] From "Biochemical Nomenclature"Liébecq, C., ed.[25]
[b] Isomers of the 5'-nucleotides, in which the phosphate is attached to the oxygen on C-3', are the 3'-nucleotides. Care must be taken to avoid ambiguity. The simple abbreviations UMP, CMP, AMP, and GMP always refer to the 5'-nucleotides.

Nucleic acid structures are abbreviated in several ways. For example, the sugar rings may be portrayed by vertical lines. The abbreviations A, C, U, T, and G for the individual bases or Pu (purine) and Py (pyrimidine) are placed at the upper ends of the lines, and slanted lines with *P* in the centers represent the 3'–5' phosphodiester linkages in a polynucleotide.

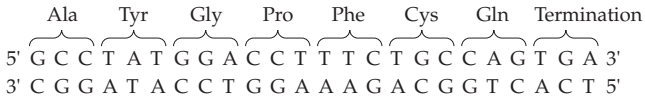



The same structure can be further abbreviated

| | | | |
|---|---|---|---|
| | 5' end | pApUpGpRpY | 3' end |
| or | 5' end | A U G R Y | 3' end |

Here purine is abbreviated R and pyrimidine Y. By convention the 5' end of a polynucleotide is ordinarily placed to the left in these formulas. Lengths of nucleic acid chains are usually given as a number of bases or kilobases (kb). For double-stranded DNA (dsDNA) the length is given as base pairs (bp), kilobase pairs (kbp), or megabase pairs (Mbp). However, in most places, including this book, the abbreviations kb and Mb will be used for a length of DNA whether single or double stranded.

For double-stranded DNA, one strand, usually the **coding strand,** from which the amino acid sequence can be read using the code in Tables 5-4 or 5-5 (Section C), has the 5' end at the left while the complementary strand has the 3' end at the left, e.g.,

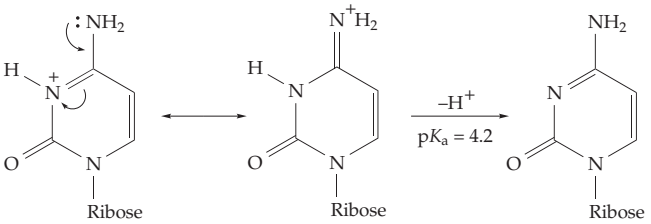



## 2. Acid–Base Chemistry and Tautomerism

The ionized phosphate groups of the polymer "backbone" give nucleic acid molecules a high negative charge. For this reason DNA in cells is usually associated with basic proteins such as the **histones** or **protamines** (in spermatozoa), with polycations of amines such as **spermidine** ($H_3{}^+NCH_2CH_2CH_2CH_2NH_2{}^+CH_2CH_2CH_2NH_3{}^+$), or with alkaline earth cations such as $Mg^{2+}$. If the pH of a solution containing double-helical DNA is either lowered to ~3 or raised to ~12, the two strands unravel and can be separated.

Over the entire range of pH the alternating sugar-phosphate backbone of the polymeric chains remains negatively charged. However, depending on the pH, the bases can be protonated or deprotonated with a resultant breaking of the hydrogen bonds that hold the pairs of bases together.[26,27]

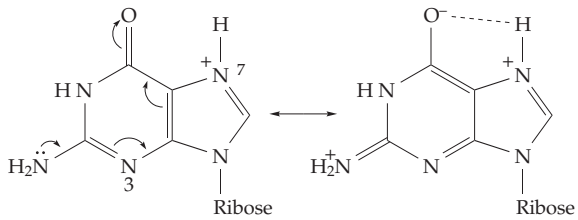
Pyrimidines and purines, which contain the $-NH_2$ group, are weakly basic. The cationic protonated conjugate acid forms of cytidine, adenosine, and guanosine have p$K_a$ values of 4.2, 3.5, and 2.7, respectively. Similar values are observed for the 5'-nucleotides. In these compounds it is not the $-NH_2$ group that binds the proton but an adjacent nitrogen atom in the ring (Eq. 5-3).



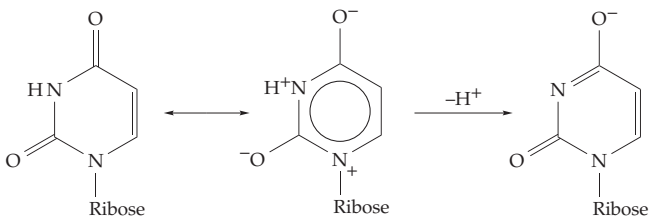

(5-3)

We can understand this if we recognize that the bases have substantial aromatic character.[26,28] In aniline (aminobenzene) electrons are withdrawn from the amino group into the aromatic ring with a strong decrease in basicity of this $-NH_2$ group (the p$K_a$ is 4.6). Similarly, electrons are withdrawn from the $NH_2$ groups of cytosine, adenosine, and guanine into the pyrimidine and purine rings as is indicated by the small curved arrows on the left-hand structure of Eq. 5-3. The effect is even stronger than in aniline, largely because of the presence of the nitrogen atoms in the rings. In cytosine it is primarily N-3 that serves as the electron acceptor. As a consequence this nitrogen becomes more basic than the $-NH_2$ group and is the major site of protonation. However, as is indicated in Eq. 5-3, the positive charge on the cation is shared by resonance with the exocyclic amino group.

Adenosine is similar to cytosine in its acid–base chemistry; N-1, adjacent to the $-NH_2$ group, is the principal site of protonation. A tautomer of the cation protonated at N-3 is formed in smaller amounts. Guanosine is electronically more complex, being protonated mainly at N-7 and to a lesser extent at N-3[29]. This can be understood in terms of electronic interaction with the adjacent oxygen as indicated in the resonance structure to the right in the following diagram:



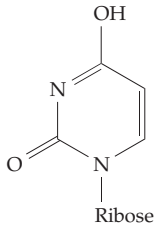
Under basic conditions the proton on N-3 of uridine or thymine or on N-1 of guanosine can dissociate with a p$K_a$ of ~9.2. These "bases" are actually weak acids!




(5-4)

Look at the "Kekule" resonance structure shown in the center of Eq. 5-4. The two negative charges flanking the dissociable proton are sufficient to hold it firmly to the ring at low pH. The proton is half dissociated only when the pH is raised to 9.2. At still higher values of pH (p$K_a$ ~12.4) a proton dissociates from a ribose hydroxyl group.

The tautomerism of pyrimidines was discussed in Chapter 2, Section A,6. The tautomeric forms shown in Fig. 5-1 predominate. However, it is possible that minor tautomers such as the following are sometimes preferentially bound into active sites of enzymes where the dielectric




A minor tautomer of uridine

constant may be low and where the geometrical arrangement of amino acid functional groups may favor protonation on oxygen rather than nitrogen.[30,31]

## 3. Absorption of Ultraviolet Light

Nucleic acids strongly absorb ultraviolet light of wavelengths below about 300 nm, with an absorption maximum at ~260 nm and a stronger one below 200 nm. This property is biologically important because of the resultant induction of mutations, a natural result of exposure to sunlight. In the laboratory the same property is useful in identification and quantitative analysis of nucleotides. The ultraviolet spectra of four nucleosides are shown in Fig. 5-5. Notice the changes that accompany protonation or deprotonation of the ring. The absorption bands, which are related to those of benzene[26,32] (see Chapter 23), provide another indication of the partial aromatic character of the bases. Chapter 23 also provides information on photochemical reactions of the pyrimidines.

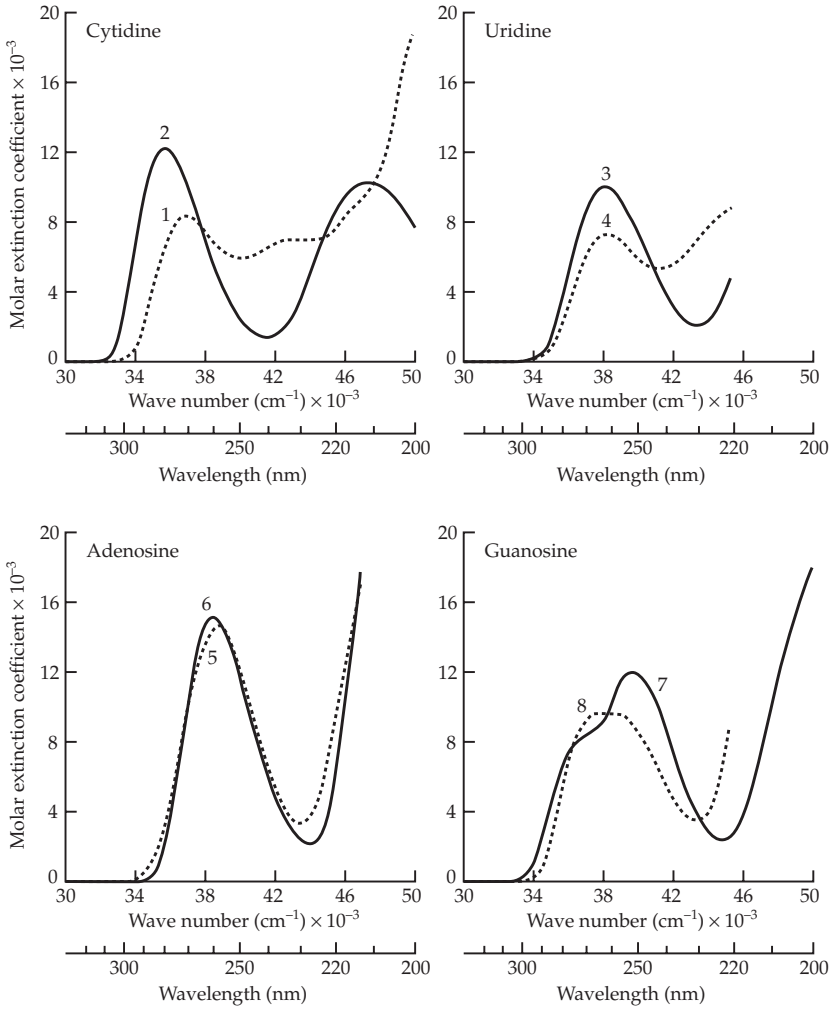



**Figure 5-5** Near ultraviolet absorption spectra of cytidine, uridine, adenosine, and guanosine. 1. Monoprotonated form of cytidine (for which p$K_a$ = 4.2). 2. Neutral form (pH ~ 7) of cytidine. 3. Neutral form of uridine (for which p$K_a$ = 9.2). 4. Monoanionic form of uridine. 5. Monoprotonated form of adenosine (p$K_a$ = 3.5). 6. Neutral form of adenosine. 7. Neutral form of guanosine (p$K_a$ = 9.2). 8. Monoanion of guanosine.

## BOX 5-A THE ALKALI METAL IONS

The many negative charges along a nucleic acid backbone interact with all of the cations in a cell. This box discusses some of these ions with emphasis on the group IA metal ions. Although sodium and potassium occur in similar amounts in the crust of the earth, living cells all accumulate potassium ions almost to the exclusion of sodium.[a-c] Sodium ions may be required only by certain marine organisms and by multicellular animals that regulate their internal body fluids. Most nonmarine plants have no demonstrable need for sodium.

The tendency to accumulate $K^+$ is even more remarkable since seawater is ~0.46 M in $Na^+$ and only 0.01 M in $K^+$. Other alkali metals occur in even smaller amounts, e.g., 0.026 mM $Li^+$, 0.001 mM $Rb^+$, and a trace of $Cs^+$. Soil water is ~0.1 mM in $K^+$ and 0.65 mM in $Na^+$. Again, strong discrimination in favor of potassium is observed in uptake by plants.

Intracellular concentrations of $K^+$ range from 200 mM in *E. coli* and 150 mM in mammalian muscle to ~30 mM in freshwater invertebrates such as clams, hydra, and some protozoa. While $K^+$ cannot be replaced by $Na^+$, a partial replacement by $Rb^+$ and to a lesser extent by $Cs^+$ is usually possible. In many microorganisms rubidium can almost completely replace potassium, and even a rat can survive for a *short* time with almost complete substitution of $K^+$ by $Rb^+$. Protons replace most $K^+$ in brown algae.[d] The human nutritional requirement for potassium is high, amounting to ~2 g/day. Present populations may suffer a chronic deficiency of potassium as a result of food processing and boiling of vegetables.[e]

Sodium is also essential to higher animals, and rats die on a sodium-free diet. The sodium content of cells varies among species, but it is usually no more than 0.1–0.2 times that of $K^+$. A measurement of the [$Na^+$] within heart cells gave a concentration of ~9 mM, which was increased by a factor of ~2.5 in a low $Ca^{2+}$ insulin-containing medium.[f] In this measurement the NMR resonance of the abundant external $Na^+$ was shifted by use of a paramagnetic reagent (e.g., a dysprosium (III) complex), that remained outside the cell. The signal from the internal $Na^+$ was then seen clearly. In blood, the relationship between $Na^+$ and $K^+$ concentrations is reversed from that within cells. Human plasma is 0.15 M in $Na^+$ and 0.005 M in $K^+$. Curiously, the taste for salt in the diet appears to be largely an acquired one.[g]

It is not immediately obvious why $K^+$ is the preferred counterion within tissues, but a fundamental reason may lie in the differences in hydration between $Na^+$ and $K^+$ (Chapter 6). On the other hand, the relationship of these ions to the excitability of membranes (Chapter 30) may be of paramount importance, even in bacteria. The concentration differences in the two ions across membranes represent a readily available source of Gibbs energy for a variety of membrane-associated activities. Cells actively pump $Na^+$ out and $K^+$ into cells (Chapter 8).

Many intracellular enzymes require $K^+$ for activity.[b,c] These include those promoting phosphorylation of carboxyl groups or enolate anions and elimination reactions yielding enols as well as some enzymes dependent upon the coenzyme pyridoxal phosphate.[h–j] In all of these enzymes $NH_4^+$, $Rb^+$, or $Tl^+$ can usually replace $K^+$. This permits study of the binding site for $Tl^+$ by the very sensitive $^{205}Tl$ NMR spectroscopy.[k] The discovery that $K^+$ is preferentially bound in some tetraplex DNA structures (see Fig. 5-8) further emphasizes the significant difference in biological properties of the alkali metal ions. Various synthetic macrocyclic compounds are also able to selectively bind specific alkali metal ions.[l]

The following tabulation gives concentrations not only of $K^+$ and $Na^+$ but also of the other principal ionic constituents in human blood plasma and within cells of skeletal muscle.[m] Units are mmol/kg $H_2O$.

| Ion | Blood plasma | Skeletal muscle (intracellular) |
|---|---|---|
| $Na^+$ | 150 | 14 |
| $K^+$ | 5 | 150 |
| $Mg^{2+}$ | 0.9 | 8 |
| $Ca^{2+}$ | 2.5 | 1 |
| $Cl^-$ | 105 | 16 |
| $HCO_3^-$ | 27 | 10 |
| Proteins$^-$ | 17* | 50* |
| Other anions† | 6 | 146 |

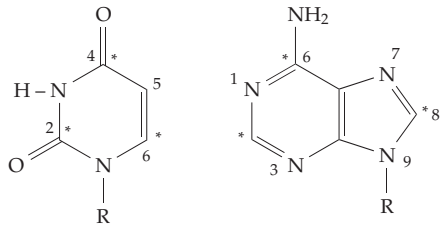
* Milliequivalents/kg $H_2O$
† Phosphates and other nonprotein anions.


[a] Kernan, R. P. (1965) *Cell K*, Butterworth, London
[b] Suelter, C. H. (1974) in *Metal Ions in Biological Systems*, Vol. 3 (Sigel, H., ed), pp. 201–251, Dekker, New York
[c] Suelter, C. H. (1970) *Science* **168**, 789–795
[d] Steinbach, H. B. (1962) *Comp. Biochem. Physiol.* **4**, 677–720
[e] Weber, C. E. (1970) *J. Theor. Biol.* **29**, 327–328
[f] Wittenberg, B. A., and Gupta, R. K. (1985) *J. Biol. Chem.* **260**, 2031–2034
[g] Kaunitz, H. (1956) *Nature (London)* **178**, 1141–1144
[h] Toney, M. D., Hohenester, E., Cowan, S. W., and Jansonius, J. N. (1993) *Science* **261**, 756–759
[i] Antson, A. A., Demidkina, T. V., Gollnick, P., Dauter, Z., Von Tersch, R. L., Long, J., Berezhnoy, S. N., Phillips, R. S., Harutyunyan, E. H., and Wilson, K. S. (1993) *Biochemistry* **32**, 4195–4206
[j] Metzler, C. M., Viswanath, R., and Metzler, D. E. (1991) *J. Biol. Chem.* **266**, 9374–9381
[k] Markham, G. D. (1986) *J. Biol. Chem.* **261**, 1507–1509
[l] Christensen, J. J., Hill, J. O., and Izatt, R. M. (1971) *Science* **174**, 459–467
[m] Composite data from Muntwyler, E. (1968) *Water and Electrolyte Metabolism and Acid–Base Balance*, p. 14. Mosby, St. Louis, Missouri; White, A., Handler, P., and Smith, E. L. (1973) *Principles of Biochemistry*, 5th ed., p. 802. McGraw-Hill, New York; Long, C. (1961) *Biochemist's Handbook*, p. 670. Van Nostrand, Princeton, New Jersey. Reported ranges for some constituents are very wide.

## 4. Chemical Reactions of the Bases

The purines and pyrimidines are relatively stable compounds with considerable aromatic character. Nevertheless, they react with many different reagents and, under some relatively mild conditions, can be completely degraded to smaller molecules. The chemistry of these reactions is complex and is made more so by the fact that a reaction at one site on the ring may enhance the reactivity at other sites. The reactions of nucleic acids are largely the same as those of the individual nucleosides or nucleotides, the rates of reaction are often influenced by the position in the polynucleotide chain and by whether the nucleic acid is single or double stranded. The reactions of nucleosides and nucleotides are best understood in terms of the electronic properties of the various positions in the bases.[26,33] Most of the chemical reactions are nucleophilic addition or displacement reactions of types that are discussed in Chapters 12 and 13.

Positions 2, 4, and 6 of pyrimidine bases are deficient in electrons and are therefore able to react with nucleophilic reagents. The 6 position is especially reactive toward additions, while the 2 position is the least reactive. The corresponding electron-deficient positions in the purine bases are 2, 6, and 8. These positions, which are marked by asterisks on the following structures, have electrophilic character in all of the commonly occurring pyrimidines and purines.



All of the oxygen and nitrogen atoms in the pyrimidines, as well as the 5 position of the ring, have nucleophilic character and can therefore react with electrophilic centers of various reagents. A number of specific reactions that have been found useful to biochemists are described in Section H,3.

## 5. Base Pairs, Triplets, and Quartets

The purine and pyrimidine bases are the "side chains" of the nucleic acids. The polar groups that are

$$>C=O,\quad >NH,\ \text{and}\quad -NH_2$$

present in the bases can form hydrogen bonds to other nucleic acid chains, e.g., in the base pairs of the DNA double helix, and also to proteins. Figure 5-6 shows the shapes and the hydrogen bonding groups available in the bases. The number of both electron donor groups and proton donors available for hydrogen bonding is large and more than one mode of base pairing is possible.

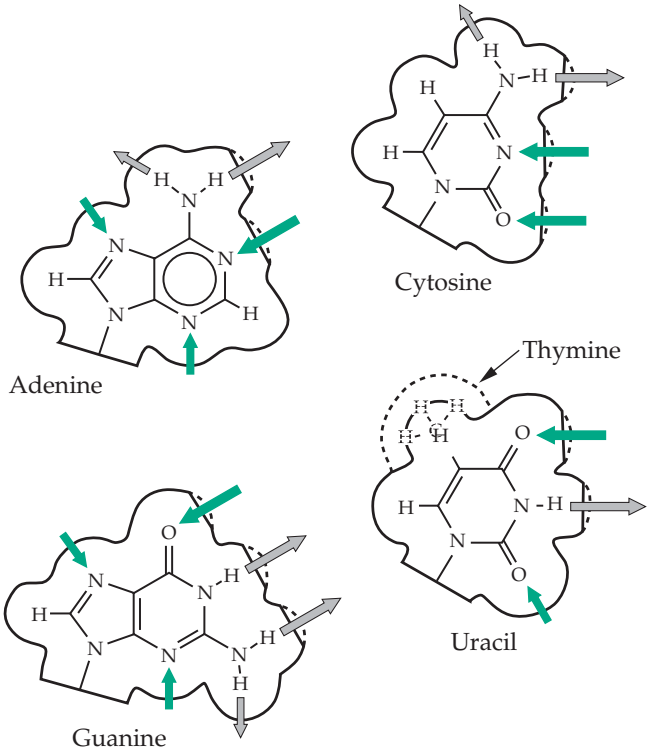
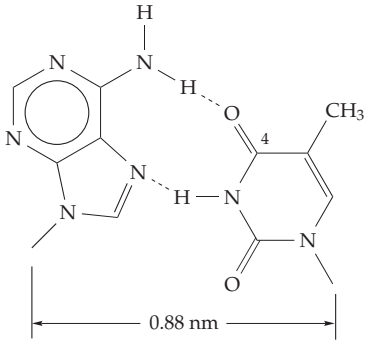



**Figure 5-6** Outlines of the purine and pyrimidine bases of nucleic acids showing van der Waals contact surfaces and some of the possible directions in which hydrogen bonds may be formed. Large arrows indicate the hydrogen bonds present in the Watson–Crick base pairs. Smaller arrows indicate other hydrogen bonding possibilities. The directions of the green arrows are from a suitable hydrogen atom in the base toward an electron pair that serves as a hydrogen acceptor. This direction is *opposite* to that in the first edition of this book to reflect current usage.

The base pairs proposed by Watson and Crick are shown in Fig. 5-2 and again in Fig. 5-7. While X-ray diffraction studies indicate that it is these pairs that usually exist in DNA, other possibilities must be considered. For example, Hoogsteen proposed an alternative A-T pairing using the 6-$NH_2$ and N-7 of adenine.[34] Here the distance spanned by the base pair, between the C-1' sugar carbons, is 0.88 nm, less than the 1.08 nm of the Watson–Crick pairs. Duplexes of certain substituted poly (A) and poly (U) chains contain only Hoogsteen base pairs[35] and numerous X-ray structure determinations have established that Hoogsteen pairs




Hoogsteen A*T base pair

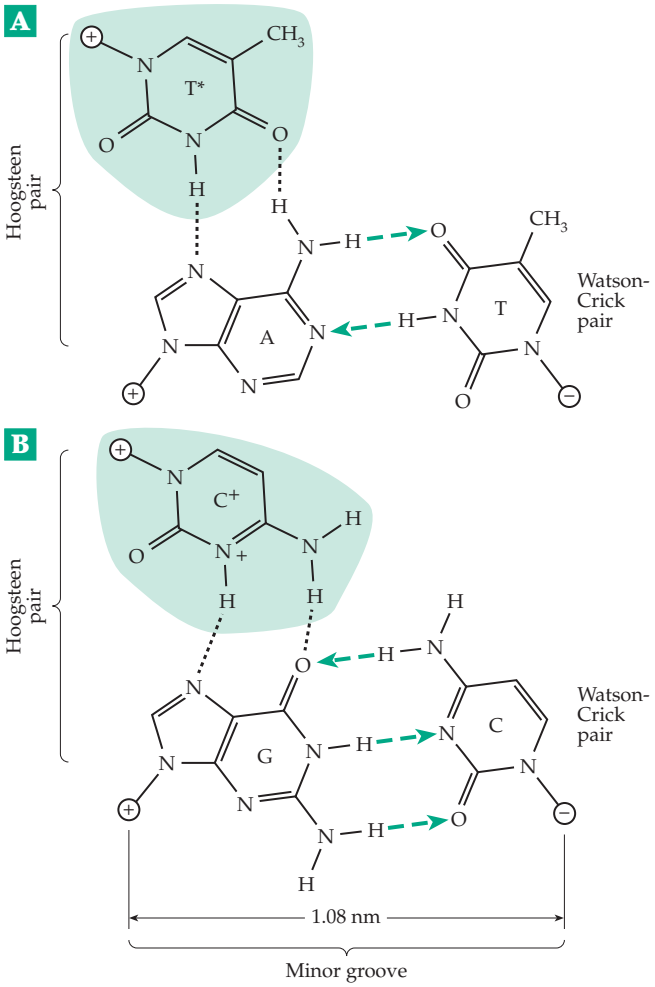



**Figure 5-7** Two base triplets that form in triple-stranded DNA and involve both Watson–Crick and Hoogsteen base pairing. (A) The triplet T*A•T, where the T (marked T*) of the third strand is hydrogen bonded as a Hoogsteen pair (·····) to an adenine of a Watson–Crick AT pair (whose hydrogen bonds are indicated (– – –). (B) The triplet $C^+$*G•C, where $C^+$ is cytosine in its N-1 protonated (low pH) form. The Watson–Crick strands are antiparallel, as indicated by the ⊕ and ⊖ signs. The third strand may have either orientation, but when it contains largely pyrimidines it is parallel to a purine-rich strand. An example is shown in Fig. 5-24.

do occur in true nucleic acids.[36] Figure 5-7 shows the structures of two Hoogsteen pairs. In each case the purine component has formed a Watson–Crick pair with a third base to give a **base triplet**. The first of these triplets may be designated T*A•T, where * represents the Hoogsteen hydrogen bonding and • represents the Watson–Crick bonding. The second triplet in Fig. 5-7, $C^+$*G•C, can form only with the N-1 protonated form (low pH form) of cytosine.

**Reversed** Watson–Crick or Hoogsteen AU or AT pairs are formed if the 2-carbonyl rather than the 4-carbonyl of the U or T makes a hydrogen bond with the amino group of adenine.[27] Because of the resulting arrangement of the ribose rings the base pairs cannot fit into the ordinary Watson–Crick double-stranded DNA structure. A reversed Watson–Crick pair can also be formed between G and C but with only two hydrogen bonds, while a reversed Hoogsteen pair can form only if a minor tautomer of cytosine is used.

Hoogsteen pairs were first observed in nature in transfer RNA molecules (Fig. 5-31). These molecules contain mostly Watson–Crick base pairs but there are also two reversed Hoogsteen pairs. One of them, between U8 and A14, is invariant in all tRNAs studied. Hoogsteen pairing also occurs in four-stranded DNA, which has important biological functions. A **G quartet** from a DNA tetraplex held together by Hoogsteen base pairs is shown in Fig. 5-8.

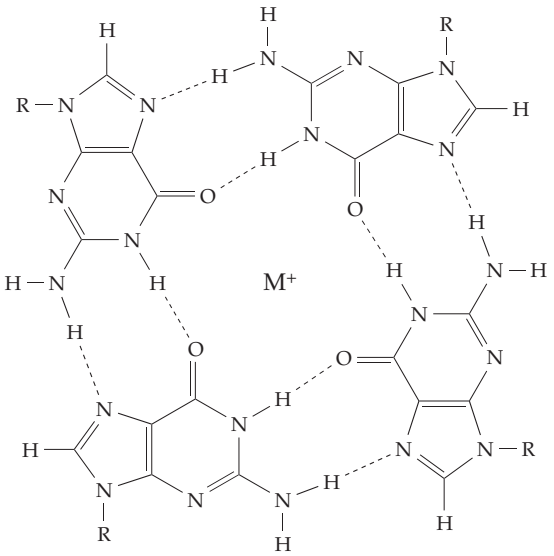
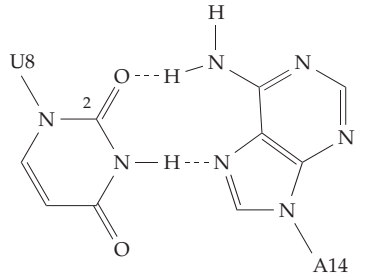
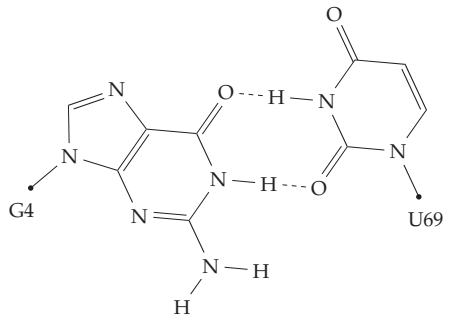



**Figure 5-8** A guanine quartet held together by Hoogsteen base pairing. This structure is found in the telomeres at the ends of linear chromosomes. Four segments of DNA, each of which may be part of a single folded strand, (see p. 227) give structures in which four or more of these G quartets are stacked one above the other. Monovalent ions, usually $K^+$ or $Na^+$, are bound in the center, although not always in the plane of the bases. See Fang and Cech[38] and Gellert *et al.*[39]




A reversed Hoogsteen base pair present in tRNA molecules

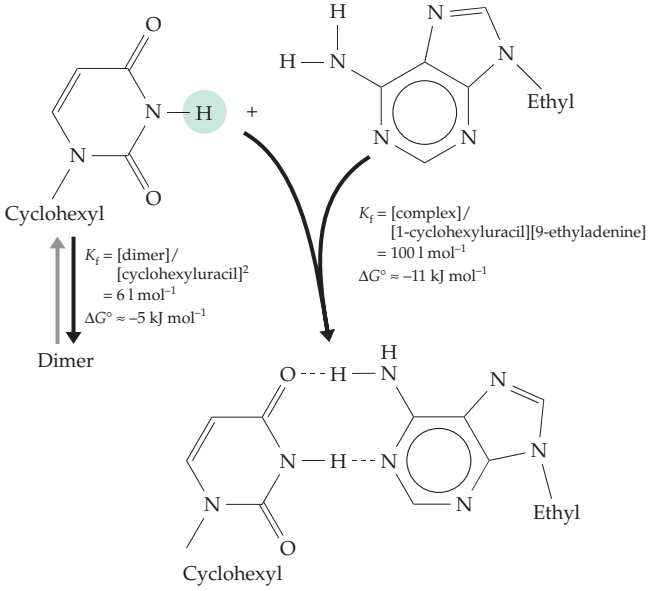
Another pairing that occurs in tRNAs allows guanine to pair with uracil, e.g., G4 with U69. This was originally proposed to account for codon–anticodon interactions betweentRNA molecules and messenger RNA (Chapter 29). It is commonly called **wobble pairing** because the uracil must wobble away from its orientation in the normal Watson–Crick pair.[27,37]




A "wobble" base pair

***Strengths of base pairs.*** How strong are the bonds between pairs of bases in DNA? The question is hard to answer because of the strong interaction of the molecules with polar solvents through hydrogen bonding and hydrophobic effects. Some insight has come from studies of the association of bases in nonpolar solvents. Thus, 1-cyclohexyluracil forms a dimer involving either hydrogen bonding or stacking, but the association is weak with the Gibbs energy of formation $\Delta G_f° = 5$ kJ mol$^{-1}$. When the same compound was mixed with 9-ethyladenine, a base-paired complex formed between the two compounds with a formation constant over ten-fold greater than that for the dimer (Eq. 5-5).[40] When the circled hydrogen atom in Eq. 5-5 was replaced by $-CH_3$, which blocked the pairing, $K_f$ fell below 1 kJ/mol ($\Delta G° > 0$). The difference in $\Delta G_f$ in the two cases was only 7 kJ mol$^{-1}$. Many other estimates have been made.[40a,40b] The small energies summed over the many base pairs present in the DNA molecule help provide stability to the structure.

***Stacking of bases.*** The purines and pyrimidines of nucleic acids, as well as many other compounds with flat ring structures and containing both polar and nonpolar regions, are sparingly soluble in either water




Complex (base pair) (5-5)

or organic solvents. Molecules of these substances prefer neither type of solvent but adhere tightly to each other in solid crystals. Both experimental measurements and theoretical computations[41] suggest that hydrogen bonding is the predominant force in the pairing of bases in a vacuum or in nonpolar solvents. However, in water stacking becomes important.[42] In a fully extended polynucleotide chain consecutive bases are 0.7 nm apart, twice the van der Waals thickness of a pyrimidine or purine ring, but in double-helical DNA of the B type (Figs. 5-3, 5-12) the distance between consecutive base pairs is only 0.34 nm. They are touching.

One effect of stacking is a decrease in the expected intensity of light absorption. The molar extinction coefficient of a solution of double-helical DNA or RNA is always less by up to 20–30% than that predicted from the spectra of the individual nucleosides (Fig. 5-5). This **hypochromic effect** is considered further in Chapter 23.

Because both hydrogen bonding and stacking are involved, the thermodynamics of base pairing in nucleic acids is complicated.[43] The hydrophobic parts of exposed bases tend to induce an ordering of the surrounding water molecules and therefore a decrease in their entropy. However, hydrogen bonding of the polar groups of the bases to the solvent causes a decrease in water structure. This is greater than the increase in structure around the hydrophobic regions and the stacking of bases leads to a net decrease in entropy. The entropy change $\Delta S$ for addition of a base pair to the end of a double-stranded RNA helix in a hairpin loop such as that displayed in Fig. 5-9 ranges from −0.05 to −0.15 kJ/degree per base pair.[44,45]

The enthalpy change $\Delta H$ tends to be small and positive for association of alkyl groups in water and nearly zero for association of aromatic hydrocarbons

(Chapter 2). However, $\Delta H$ is distinctly negative for association of heterocyclic bases. This has also been attributed to a decrease in the ordering of solvent around the bases as a result of exclusion of water. Attraction or repulsion of partial charges on the polar groups comprising the purine and pyrimidine bases may also be an important factor.[43,46–48] For addition of a base pair to an RNA helix, the change in enthalpy, $\Delta H$, varies from about –24 to –60 kJ/mol.[44,45]

Since $\Delta G = \Delta H - T\Delta S$, the net result is a negative value of $\Delta G$, a "hydrophobic effect" that favors association of bases. Substantial efforts have been made to estimate quantitatively the Gibbs energies of formation of helical regions of RNA molecules in hairpin stem-loops such as that of Fig. 5-9.[44,45,49–51] Table 5-2 shows the observed increments in $\Delta G_f°$ of such a helix upon addition of one base pair at the end of an existing helix. Addition of an AU pair supplies only –4 to –5 kJ to $\Delta G_f°$. The exact amount depends upon whether an A or a U is at the 5' end in the existing helix. If an AU pair is added to the helix terminating in CG or GC, about –9 kJ/mol is added. Larger increases in $-\Delta G_f°$ result from addition of GC pairs, which contain three hydrogen bonds between the bases versus the two in AU pairs.[27,41,52]

UG pairs provide a very small amount of stabilization to an RNA double helix, while the presence of unpaired bases has a destabilizing effect. The most stable hairpin loops contain four or five bases. Depending upon whether the loop is "closed" by CG or AU, the helix is destabilized by 20–30 kJ/mol. "Bulge loops," which protrude from one side of a helix, have a smaller destabilizing effect. An example of the way in which Table 5-2 can be used to estimate the energies of formation of a loop in a straight-chain RNA is illustrated in Fig. 5-9. Similar analysis of base pairing in DNA can also be done.[53–55]

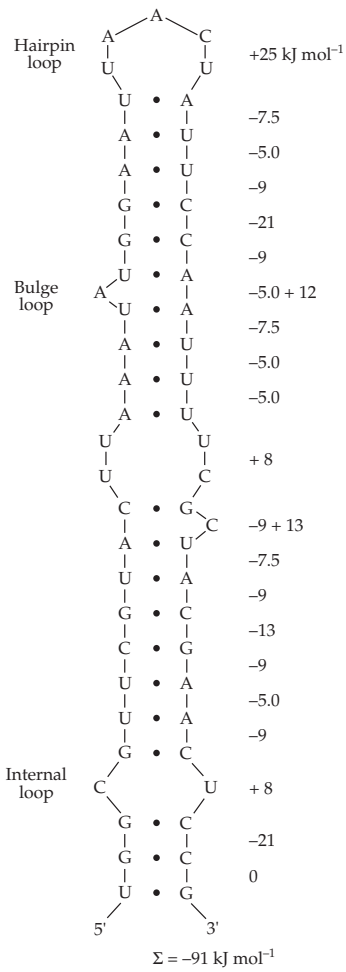



**Figure 5-9** The contribution of base-paired regions and loops to the Gibbs energy of a possible secondary structure for a 55 base fragment from R17 virus. The stem-loop structure shown here is part of a larger one considered by Tinoco *et al.*[58]

**TABLE 5-2**
**Gibbs Energies of Formation $\Delta G_f°$ at 25°C for Addition of One Base Pair to an Existing RNA Helix[a,b]**

| Base pair at end of existing helix | Base pair added | $\Delta G_f°$ (kcal mol$^{-1}$ ± 10%) | $\Delta G_f°$ (kJ mol$^{-1}$ ± 10%) |
|---|---|---|---|
| A•U | A•U | –1.2 | –5.0 |
| U[c]•A | A•U | –1.8 | –7.5 |
| C•G or G[d]•C | A•U | –2.2 | –9 |
| C•G | G•C | –3.2 | –13 |
| G•C | C•G or G•C | –5.0 | –21 |
| G•U | U•G | –0.3 | –1 |
| Hairpin loops of 4 or 5 bases | | | |
| Closed by GC | | +5 | +21 |
| Closed by AU | | +7 | +29 |
| Bulge loops | | | |
| 1 base | | +3 | +12 |
| 4–7 bases | | +5 | +21 |

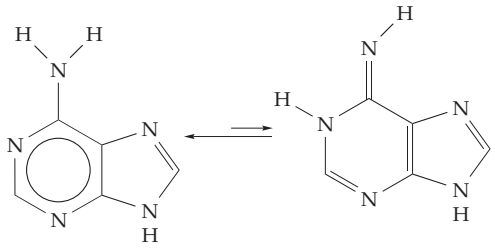
[a] Table modified from that of Tinoco *et. al.*[58]

[b] All base pairs in the table are oriented as follows: 5'—A→3' / 3'←U—5'

[c] $\Delta G_f°$ is the same for U•A added to an A•U end.

[d] $\Delta G_f°$ is the same for C•G or G•C added to an A•U end.

***Tautomerism and base pairing.*** Tautomerism has an interesting relationship to the formation of the pairs and triplets of hydrogen bonds in DNA or RNA. Each base exists predominately as one preferred




(5-6)

tautomer, but at any moment a very small fraction is present as less stable tautomers. Watson and Crick suggested that this fact may be responsible for the occurrence of some mutations.[56,57] Thus, tautomer B of Eq. 5-6 would not be able to pair with thymine, its proper pairing partner, but could pair with cytosine.

Similarly, tautomer B of uracil in Eq. 2-4 could pair with guanine instead of its proper partner adenine. If a similar event occurred to an AT pair during gene replication an incorrect copy of the gene, differing in a single "code letter," would be formed. However, because the tautomerism is affected so strongly by the environment (Chapter 2), the extent to which it may cause mispairing while replication enzymes act is uncertain.[59] Both bases in a pair could be tautomerized by synchronous transfer of protons in two parallel hydrogen bonds. However, theoretical calculations predict a high energy barrier to this process.[60] Proton transfer, which can also be induced by light, has been studied on a femtosecond scale.[61] After photochemical transfer of one proton a second is transferred within a few picoseconds.

## 6. Conformations of Nucleotides

The furanose ring of ribose or deoxyribose is flexible and can be interconverted smoothly among an infinite number of envelope (E) and skew or twist (T) conformations. See Chapter 4, Section A,2. However, there are limits set by steric and anomeric effects.[27,62–64] Conformations are often described as in Fig. 5-10 by stating which atom in an envelope conformation lies mostly out of the plane of the other four atoms. If this atom lies above the ring, i.e., toward the base, the ring

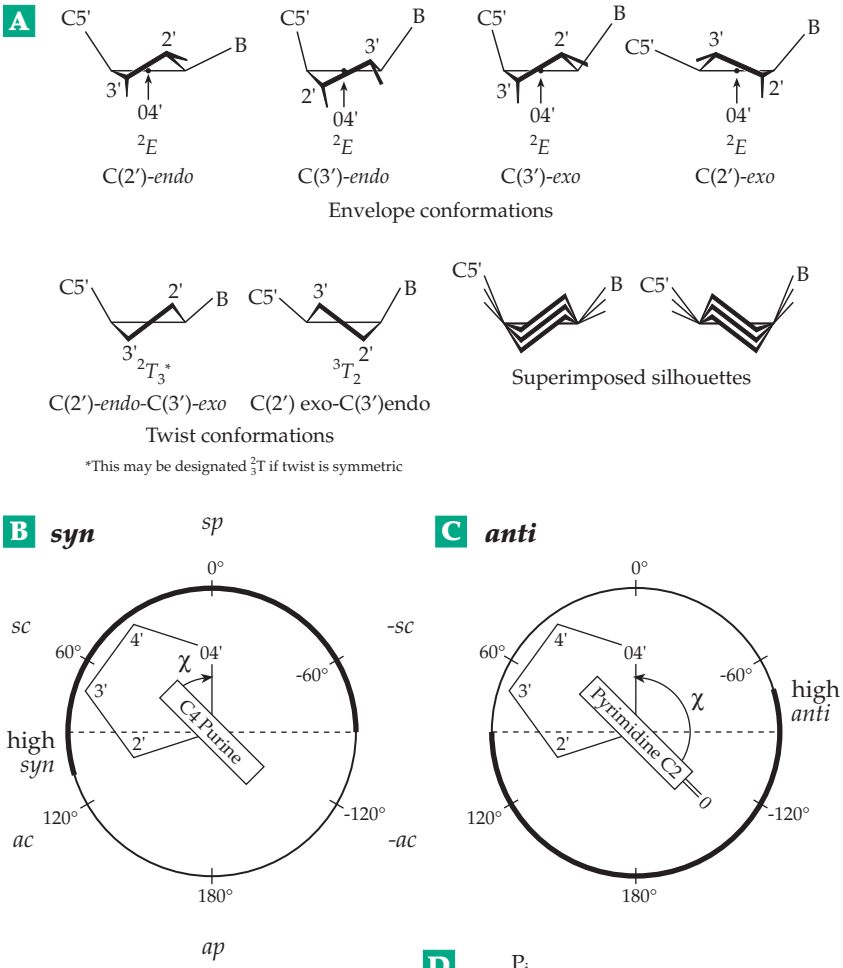
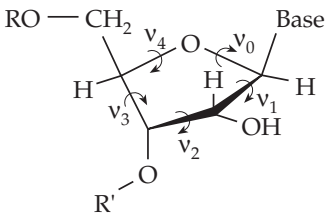



**Figure 5-10** Conformational properties of nucleosides. (A) Representations of several conformations of a ribose or deoxyribose ring in a nucleoside. (B) View down the N–C axis joining a purine base to the sugar in a nucleoside in a *syn* conformation. The atom marked C4 is in the six-membered ring, which protrudes further over the sugar ring. The angle χ is measured from 0° to ±180°. *Syn* conformations are those for which χ falls in the region of the heavy semicircle. (C) An *anti* conformation of a pyrimidine nucleoside. Other *anti* conformations fall in the region of the heavy semicircle. (D) Labeling of the conformational angles of the main chain in a polynucleotide.[25,27,37,37a] See also Fig. 2-2.

conformation is known as ***endo***; when below the ring, it is known as ***exo***. The C(2')-*endo* ($^2$E) and C(3')-*endo* ($^3$E) conformations are most commonly approximated in nucleotides and nucleic acids.[65] The C(3')-*exo* conformation is designated $E_3$, the twist conformation C(2')-*endo*-C(3')-*exo* as $^2T_3$, etc. A conformation can be specified more precisely by the five torsion angles $\nu_0$ to $\nu_4$ (which have also been designated $\tau_0$ to $\tau_4$). All of the envelope and twist conformers can be interconverted readily. The interconversions can be imagined to occur in a systematic way by **pseudorotation**, a rotation of the pucker around the sugar ring. In a commonly used convention a pseudorotation phase angle *P* ranges from 0° to 360° as the pucker moves *twice* around the ribose ring to restore the original conformation. *P* is taken as 0° for the symmetric $^3T_2$ twist, −18° for $E_2$, +18° for $^3$E, +54°




for $^4$E, +198° for $E_3$, etc. The five torsion angles $\nu_j$, for j = 0 – 4 as defined in the foregoing diagram, are related to *P* as follows:

$$\nu_j = \nu_{max} \cdot \cos[P + j \cdot \phi] \qquad (5\text{-}7)$$

where $\phi = 720°/5 = 144°$. The maximum torsion angle, $\nu_{max}$, is about 40°. See Saenger[27] for details.

Conformational alterations of ribose and deoxyribose rings can occur within polynucleotides and are of biochemical importance. An interesting consequence of changes in ring conformation is that the distance between the C(5') and N atoms attached to the sugar ring of a nucleoside may vary by as much as 0.05 nm (see Fig. 5-10A). The orientation of a base with respect to the sugar is specified by the angle $\chi$. In one convention (Fig. 5-10B,C) its zero value is taken as that in which the N(1)–C(2) bond of a pyrimidine or the N(9)–C(4) bond of a purine is *cis* to the C(1')–O(4') bond, [sometimes called the C(1')–O(1') bond]. However, other definitions have been used.[27] Typical values of $\chi$ for nucleotides and nucleosides vary between −75° and −165°. In these ***anti*** conformations the CO and NH groups in the 2 and 3 positions of the pyrimidine ring (or in positions 1, 2, and 6 of the purine ring) are *away* from the sugar ring, while in the ***syn*** conformations they lie over the ring. *Anti* conformations are more often present than *syn* in nucleic acids. A different view of *syn* and *anti* pairs of nucleotides is shown in Fig. 5-11.

An additional five torsion angles are needed to specify the backbone conformation of a polynucleotide. According to the convention adopted by the IUB the six angles are desig-

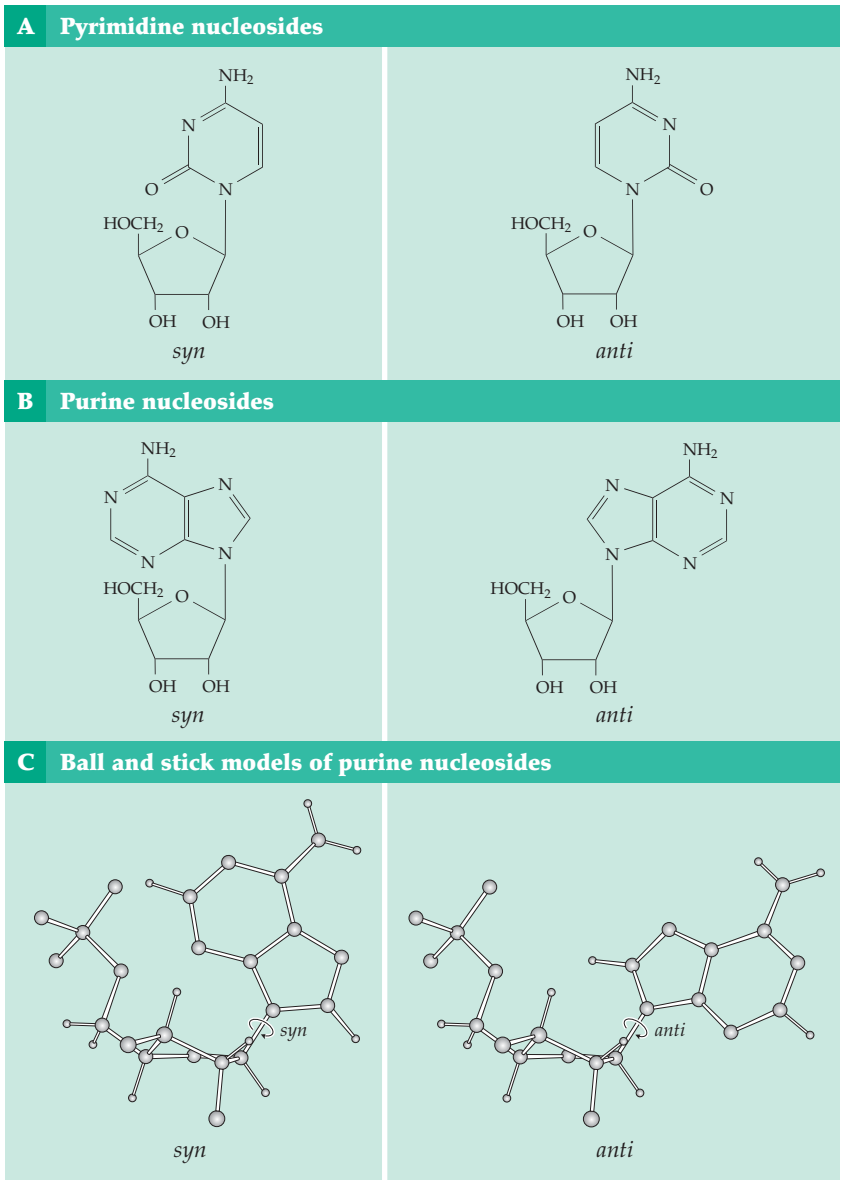



**Figure 5-11** Views of a nucleoside in *syn* and *anti* conformations.[66] (C) Courtesy of Dr. Muttaiya Sundaralingam.

nated α, β, γ, δ, ε, and ζ, as is indicated in Fig. 5-10D.[25,27] In an older but much used convention,[66] starting at any phosphorus atom the angles ω, ϕ, and ψ specify the next three torsion angles as one moves toward the 3' end of the chain, while ω′, ϕ′, and ψ′ specify the angles lying toward the 5' end (these are shown in parentheses in Fig. 5-10D). Notice that $\delta(\psi') = \nu_3 + 120°$.

## B. Double Helices

### 1. The B Form of DNA

This form of DNA, whose structure is depicted in Figs. 5-3 and 5-12, is stable at high humidity and is thought to approximate that of most DNA in cells.[27,37,67] If we look directly at the axis of the double helix and perpendicular to one of the base pairs, and ignore the fact that the base pair is asymmetric, we see that the nucleotide unit in one chain is related to the nucleotide unit lying across from it in the opposite plane by a two-fold axis of rotation (**dyad axis**). This symmetry element, which arises from the antiparallel arrangement of the chains, makes the DNA molecules from the outside look nearly identical whether viewed from one end or the other – and whether viewed as a model by the human eye or through contact with an enzyme which might act on the molecule. Actually, *the two chains are not identical*, and the genetic information can be read off from the functional groups exposed in two **grooves** in the surface of the helix (Figs. 5-3 and 5-7). The broader groove in the B form, which is referred to as the **major groove,** is about 0.85 nm deep and 1.1–1.2 nm wide when allowance is made for the van der Waals radii of the atoms. In some other forms of DNA the major groove is narrow, but it can always be identified by the larger of the arcs that can be drawn between the two N– C bonds of the nucleosidic linkages in a nucleotide pair. The **minor groove** or narrow groove, which is defined by the smaller arc between the two N–C bonds, is ~ 0.75 nm deep and 0.6 nm wide in B-DNA.

The diameter of the double helix of B-DNA, measured between phosphorus atoms, is just 2.0 nm. The rise per turn, the **pitch**, is 3.4 nm. There are about ten base pairs per turn (9.7 and 10.6 in two different crystal forms).[68,69] Thus, the rise per base pair is 0.34 nm, just the van der Waals thickness of an aromatic ring (Table 2-1). It is clear that the bases are stacked in the center of the helix. A 1000-bp (1-kb) gene would be a segment of DNA rod about 340 nm long, about 1/40 the length of the molecule in the electron micrograph of Fig. 5-13.

As is appropriate for the cell's master blueprint, DNA in the double helix is stable. Factors contributing to this stability are (1) the pairs and triplets of hydrogen bonds between the bases; (2) the van der Waals attraction between the flat bases which stack together; (3) the fact that on the outside of the molecule are many oxygen atoms, some negatively charged, which are able to form strong hydrogen bonds with water, with small ions, or with proteins that surround the DNA; and (4) the ability to form superhelices (see Section C,3). Nevertheless, the long DNA chains present in our chromosomes are frequently broken and an elaborate system of repair enzymes is needed to preserve the reliability of this master code for the cell.[73,74]

Today it is generally accepted that most DNA exists in nature as a double helix resembling the B form. However, doubts were expressed as recently as 1980. The reason for the uncertainty lay in the fact that X-ray data from the stretched "paracrystalline" DNA fibers used in earlier studies are much less precise than those from true crystals.[10] The X-ray data for B-DNA fibers could also be interpreted in terms of an alternative "side-by-side" structure which would permit easier separation of strands during replication.[75] However, numerous high-resolution X-ray structure determinations on single crystals of synthetic DNA fragments have confirmed the double-helical structure and the presence of Watson–Crick base pairs.[68,69,76–84] Similarly, fragments related to the double helices of RNA have been crystallized and the structures determined to atomic resolution.[85–87] The right-handed helical structure of DNA in solution has also been confirmed by independent methods based on electron microscopy,[88] scanning probe microscopy,[89,90] fluorescence resonance energy transfer,[91] and computation.[92]

Dickerson and associates discovered[78,79] that B-DNA has a "spine" of water in the minor groove of regions rich in A-T base pairs. Two water molecules per base pair are hydrogen bonded to form a long chain. Half of the water molecules in the chain also form hydrogen bonds to oxygen and nitrogen atoms that are exposed in the minor groove while the others hydrogen-bond to a second chain of water molecules, forming a ribbon of hydration (Fig. 5-14A).[81,93-95] Additional water binds to polar groups in the major groove. The hydration pattern is largely *local*, i.e., each base has characteristic hydration sites (Fig. 5-14B).[94] Some of the hydration sites may be occupied partially by the nomovalent ions, $Na^+$ or $K^+$, depending upon the medium. Bound divalent metal ions are also sometimes seen in X-ray structures.[81,83a] However, most cations are thought to remain mobile.

### 2. Other Double-Helical Forms of DNA

The B form of fibrous DNA is stable under conditions of high (~ 93%) humidity but at 75% humidity it is converted into **A-DNA** in which the base pairs are inclined to the helix axis by about 13° and in which the ribose rings are primarily C3'-*endo* rather than C2'-*endo*.

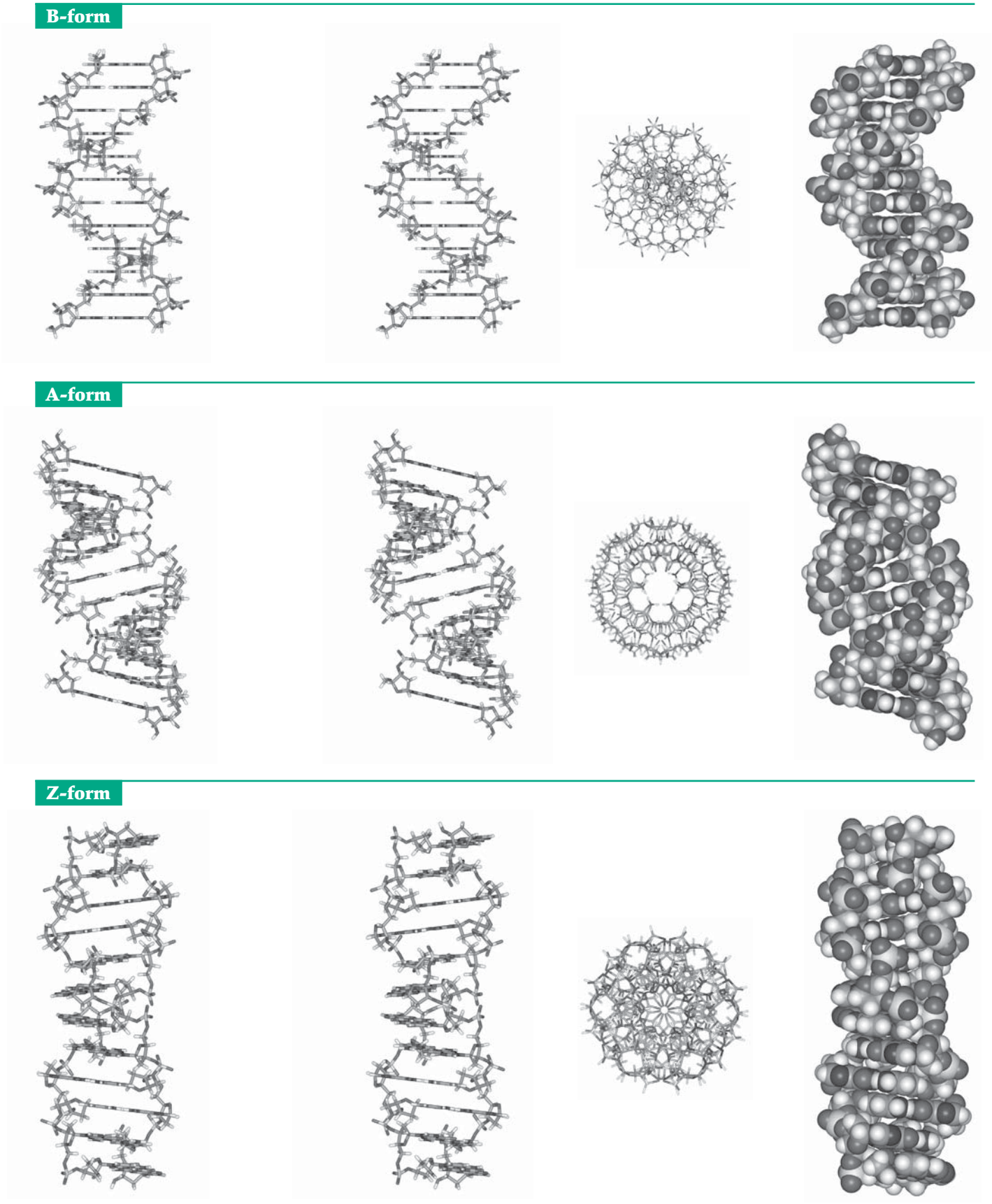



**Figure 5-12** (Left) Stereoscopic skeleton models of the B, A, and Z forms of double helical DNA. See Schlick.[70] (Center) End views (Right) Space-filling models of the same three DNA forms: B, A, and Z. Courtesy of Tamar Schlick.

There are 10.9 base pairs per turn[93,96–98] (Fig 5-12; Table 5-3). In the A form the major groove is very deep (~ 1.35 nm) and narrower (~ 0.27 nm) than in the B form and extensively hydrated.[99] The minor groove is wider (~ 1.1 nm) and shallower (~ 0.28 nm) than in the B form. The crystal structure for a form intermediate between A and B has also been reported.[99a] Paracrystalline forms of DNA known as B′, C, and D have also been observed[100,101] and others have been proposed.[102,103]

The most interesting additional form, **Z-DNA**, was discovered by X-ray studies of the alternating oligodeoxyribonucleotides d(CpGp-CpGpCpG) and d(CpGpCpG).[104] The helix of Z-DNA is *left-handed*.[71,72,105–112] The repeating unit consists of two Watson–Crick base pairs, the backbone following a zigzag pattern (Fig. 5-12). There are 12 base pairs per turn. The cytosine groups have the usual *anti* conformation but the guanosine groups are *syn*. Some of the ribose rings have the C2'-*endo* conformation characteristic of B-DNA but some are C3'-*endo*. The alternating CpG sequence is not essential for Z-DNA formation. However, alternating *syn* and *anti* conformations are important and purines assume the *syn* conformation more readily than do pyrimidines. The major groove of Z-DNA is shallow (Fig. 5-12), allowing it to accomodate bulky substituents at C8 of purines or C5 of pyrimidines. Such substituents favor Z-DNA.

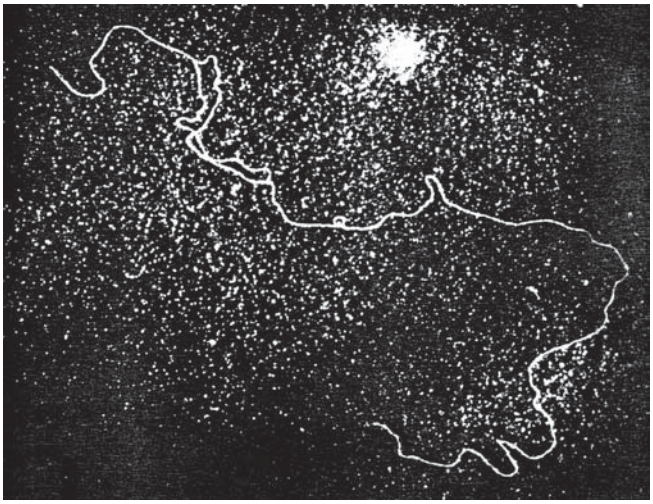


**Figure 5-13** Electron micrograph of a DNA molecule (from a bacterial virus bacteriophage T7) undergoing replication. The viral DNA is a long (~14 µm) duplex rod containing about 40,000 base pairs. In this view of a replicating molecule an internal "eye" in which DNA has been duplicated is present. The DNA synthesis was initiated at a special site (origin) about 17% of the total length from one end of the duplex. The DNA was stained with uranyl acetate and viewed by dark field electron microscopy. Micrograph courtesy J. Wolfson and D. Dressler.

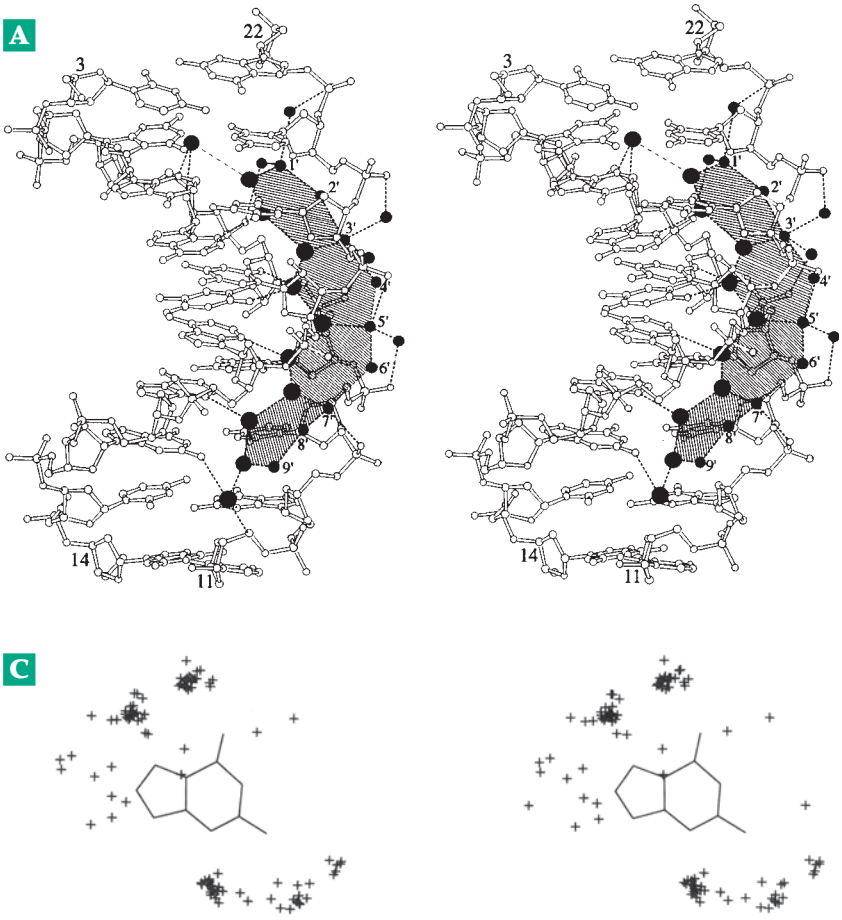



**Figure 5-14** (A) Stereoscopic drawing showing two layers of water molecules that form a "spine" or "ribbon" of hydration in the minor groove of B-DNA. The inner layer is shown as larger filled circles; water molecules of the outer layer are depicted with smaller dots and are numbered. Hydrogen bonds are shown as dashed lines. (B) Electron density map. (A) and (B) from Tereshko *et al.*[95] (C) Stereoscopic representation of the superimposed electron densities of 101 water molecules observed to hydrate 14 guanine rings found in 14 B-DNA molecules for which high-resolution X-ray structures were available. Positions of 101 water molecules within 0.34 nm from any atom of the 42 guanines are plotted. From Schneider and Berman.[94]

The B form is the most hydrated and most stable form of DNA under conditions of high humidity but even in solution it can be converted to A-DNA and Z-DNA by a high concentration of NaCl.[113] This is presumably because the salt dehydrates the DNA. Saenger and coworkers pointed out that the oxygen atoms of successive phosphate groups in the polynucleotide backbone of B-DNA are at least 0.66 nm apart, too far apart to be bridged by a water molecule.[114] However, the phosphates are individually hydrated. On the other hand, in A-DNA and Z-DNA the oxygen atoms on successive phosphates are as close as 0.53 and 0.44 nm, respectively. This allows one $H_2O$ molecule to bridge between two phosphates, stabilizing these forms in environments of low humidity. This may be one factor that affects the B to A and Z transitions.[115–117] In the narrow Z-DNA helix repulsion between negative charges on the phospho groups of opposite strands is strong. By shielding these charges high salt concentrations also help to stabilize Z-DNA.

Since the Z form of DNA is favored in regions rich in G-C pairs,[118,119] it is reasonable to expect that it may occur in nature. Antibodies have been prepared which bind specifically to regions of DNA in the Z form and have been used to identify many such regions.[111,120] Genetic studies in *E. coli* have also provided strong evidence that left-handed DNA sequences are formed *in vivo* in that bacterium.[121,122] Z-DNA-forming sequences have also been found in the *Halobacterium* genome.[119] Segments of Z-DNA may occur in control regions called **enhancers** (Chapter 28)[121–124] and Z-DNA may also form behind RNA polymerase molecules that are moving along a gene while synthesizing mRNA. The associated negative supercoiling of the DNA could cause it to assume the Z form. This region of Z-DNA can, in turn, be a site for interaction with specific proteins.

In the Watson–Crick structure the two strands are *antiparallel*, an essential for replication. However, stable segments of double-stranded DNA with *parallel* strands can also be formed and may occur in specialized regions of the genome.[125–128]

**TABLE 5-3**
**Helix Parameters for Three Types of DNA[a]**

| Parameter | Form of DNA or RNA (A-form) | | |
|---|---|---|---|
| | B | A | Z |
| Helical twist, degrees | 28–42 | 16–44 | GC −51 ± 2 |
| mean | 36 ± 4 | 33 ± 6 | CG −8.5 ± 1 |
| Base pairs per turn | 10.0 (9.7–10.6) | 11-12 | 12 |
| Helix rise per base pair, nm | 0.34 ± .04 | 0.29 ± .04 | GC 0.35 ± .02<br>CG 0.41 ± .02 |
| Base inclination, degrees | −2.0 ± 5 | 13 ± 2 | 8.8 ± .7 |
| Propeller twist, degrees | 12 ± 5 | 15 ± 6 | 4.4 ± 3 |
| Base roll degrees | −1.0 ± 5 | 6 ± 5 | 3.4 ± 2 |
| Predominant conformation of deoxyribose | C2′-*endo* | C3′-*endo* | C C2′-*endo*<br>G C3′-*endo* |
| Depth of grooves (nm) | | | |
| Major | 0.85 | 1.35 | very shallow |
| Minor | ~0.75 | ~0.28 | very deep |
| Width of grooves (nm) | | | |
| Major | 1.1–1.2 | 0.27 | broad |
| Minor | 0.6 | ~1.1 | narrow |

[a] See R. E. Dickerson[78,78a]; based on single-crystal X-ray analysis.

## 3. The Conformational Flexibility of Double Helices

Local variations in the sequence of nucleotides affect the conformation of a DNA molecule and it is clear that the helix is not uniformly coiled throughout the entire length.[80,103,124,129–134] While most helix segments probably have a right-handed twist others may be left-handed. Most DNA is probably in the B form but there are segments in the A form. These may arise from formation of hybrid duplexes with RNA, which assume the A conformation and are also favored by certain base sequences. Rules for predicting the DNA conformation from the nucleotide sequence have been proposed.[116,135] In the simplest case[135] we consider each pair of adjacent nucleotides in the double helix. There are 16 possible pairs in one chain. These can be designated as in the following examples: (AA,TT), (CG,CG), and (AG,CT). The first two letters within the parentheses represent the sequence (from 5' to 3') in one chain while the second pair of letters represent the sequence (again from 5' to 3') in the complementary chain. The rules state that (AA,TT) or (TT,AA) repeated in a sequence will stabilize the B form of DNA. Repetition of (CC,GG) or (GG,CC) will favor conversion to the A form. Repetitions of (CG,CG) favor the Z form, especially if an alternating sequence of purines and pyrimidines is present throughout the (G + C)-rich region.

Even within the regular B structure the

torsion angles χ and δ (Fig. 5-10) are variable. Their changes are highly correlated.[136,136a] As χ ranges from – 140° to – 90°, ζ varies between 80° and 160°. Other pairs of torsion angles are also correlated. Besides allowing for changes between B, A, and Z conformations, this flexibility of the DNA helix together with cooperativity with adjacent base pairs may allow transmission of conformational effects for some distances along a DNA helix.[137,138] Supercoiling, discussed in Section C,3, also affects the helical conformation.

DNA can be stretched into yet another form or forms. Application of a force of 65 – 70 piconewtons (pN) stretches B-DNA by 70%. This "overstretched DNA" may also be important biologically.[130,139–141a]

***Rotational and translational movements of bases.*** In considering the conformational flexibility of a polynucleotide it is useful to define the parameters associated with movement of base pairs or individual bases. The possible movements are indicated in Fig. 5-15 and in Table 5-3. A **long axis** is established for each base pair in a structure for which X-ray data are available. Making use of these axes, we can move along the helix and measure the **angle of twist** from one base pair to the next.[142–145] While the mean value of the twist for B-DNA is 36°, values of 28–42° have been observed in structures of oligonucleotides determined by X-ray analysis. The **base tilt**, which is nearly zero for B-DNA, is less variable. Within each base pair there is also a **propeller twist** around the long axis.[146] It averages about 12° for B-DNA. In addition, the whole base pair may **roll**, i.e., be rotated around its long axis by several degrees. These motions relieve steric interferences in the center of the helix, for example, those that would arise between purines in adjacent base pairs if the same helix parameters were imposed on each base pair.[142,147] A more detailed analysis reveals the large range of motions, mostly small, which are described in Fig. 5-15.[145,148] The mathematics needed to deal with computer-based modeling of polynucleotide structures has been developed.[80,149,150] Its application showed that there is a strong correlation between twist and roll. Typical average values of base tilt, propeller twist, and roll are given in Table 5-3 for B-, A-, and Z-DNA.

***Bends and bulges.*** If we overlook the ridges and grooves on their surfaces the DNA structures shown in Figs. 5-3 and 5-12 are straight rods. However, real DNA rods are crooked and may contain distinct bends. The

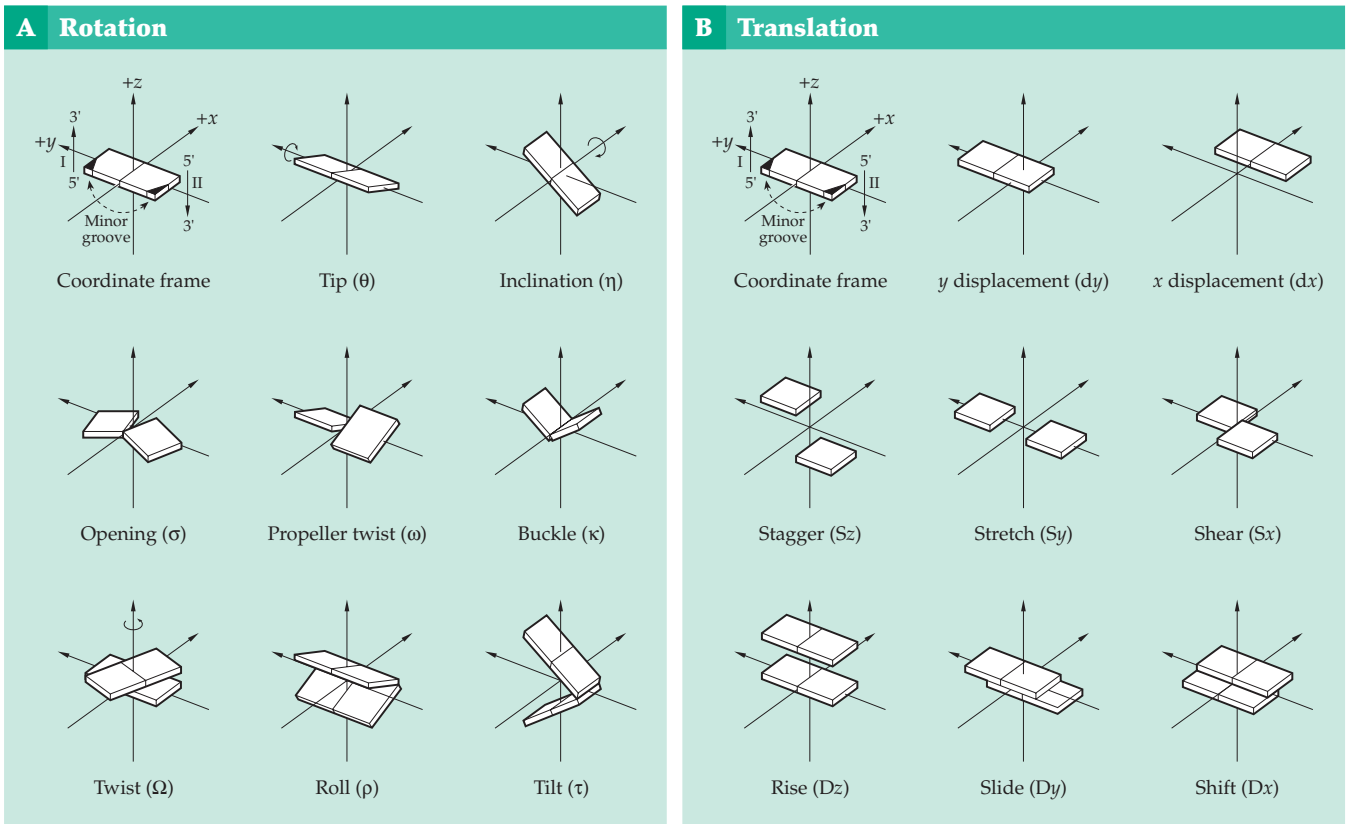



**Figure 5-15** (A) Drawing illustrating various rotational movements of bases in polynucleotides. *Upper two rows*: rotations of two bases of a pair. *Lower row*: rotations involving two successive base pairs. (B) Translational movements. *Upper two rows*: involving two bases of a pair. *Lower row*: two successive base pairs. From Diekmann.[145]

existence of **bent DNA** in nature was first discovered from study of fragments enzymatically cut from kinetoplast DNA of tropical parasites.[151,152] The bent fragments moved anomolously slowly during electrophoresis in polyacrylamide gels of small pore size. Evidently, the bent shape impedes movement through the pores. Since the initial discovery, regions of bent DNA have been found in origins of replication[152–154] and other specialized sites. Bending of DNA because of local sequence variations may also be important to the positioning of nucleosomes (Fig. 27-3) on DNA.[155]

Several causes for DNA bending have been identified.[124,138,142,156–164] For example, a sharp 26° bend is expected at the junction between B- and A-DNA segments.[158] The presence of a thymine photodimer (Eq. 23-26) may cause a 30° bend.[165] Some naturally occurring bent DNAs contain repeated (A+T)-rich sequences. This suggested that the sequence (AA,TT), containing adjacent thymine methyl groups, can be thought of as a wedge, consisting of both roll and tilt components. Such a wedge repeated at intervals averaging 10.5 bp along the helix might cause the helix to bend.[129,144] However, computer modeling experiments suggest that thymine methyl groups do not distort the DNA helix. The bending may be a result of the greater tendency for AT pairs to roll and to enhance the van der Waals and electrostatic attractions between the 2'-O of cytosine and the 2'-$NH_2$ of guanine in the CG base pairs of bent DNA.[157,166] Bending can occur in both (A+T)- and (G+C)-rich regions and is often induced by binding to proteins or protein complexes that act in replication, transcription, and recombination.[167]

Various errors are made during replication or recombination of DNA. Incorporation of an incorrect nucleotide will cause a mismatched base pair in which proper hydrogen bonds cannot be formed. Most of the mispaired bases that result from mistakes in replication are removed by repair processes (Chapter 27). Those that remain can often assume alternative pairings that distort the helix only slightly. For example, a G-T wobble pair fits readily into an oligodeoxyribonucleotide double helix. Even a bulky G-A base pair causes little perturbation of the helix.[168] Incorporation of an extra nucleotide into one strand of the DNA will create a bulge in the helix.[169] NMR spectroscopic studies on bulged oligonucleotides have shown that the extra base can be stacked into the helix causing a sharp bend.[170,171] However, some oligonucleotides with mismatched bases crystallize as straight helices with the extra nucleotide looped out.[171–173] Mismatched base pairs tend to destabilize helices.[174]

***Interactions with ions.*** Because each linking phospho group carries a negative charge (two charges per base pair, Fig. 5-2) the behavior of polynucleotides is strongly affected by cations of all kinds. The predominant small counterions within cells are $K^+$ and $Mg^{2+}$. They are attracted to the negative charges on the polynucleotide backbone and, although they remain mobile, they tend to occupy a restricted volume.[175–177] Some may bind in well-defined locations as in Fig. 5-8. Because of the presence of these positive ions the interactions of nucleic acids with cationic groups of proteins are strongly affected by the salt concentration.

Organic cations compete with the simple counterions $K^+$ and $Mg^{2+}$. Among these, the polyamines are predominant.[178] Crystal structures have revealed that spermine binds across the deep grooves of tRNA and the major groove of B-DNA. Spermine also binds into the deep groove of A-DNA interacting by hydrogen bonding with bases in GTG sequence in both strands.[179] It binds tightly to CG-rich sequences in the minor groove of Z-DNA, where it tightens the structure and shortens the helix.[110] At higher concentrations of DNA, as occur in cell nuclei, spermidine induces the conversion of the DNA into liquid crystalline phases.[180]

Heavier metal ions and metal complexes can find sites on nitrogen atoms of the nucleic acid bases. Examples are the platinum complex **cisplatin** and the DNA-cleaving antibiotic **neocarzinostatin** (Box 5-B). Can metals interact with the π electrons of stacked DNA bases? A surprising result has been reported for intercalating complexes of ruthenium (Ru) and rhodium (Rh). Apparent transfer of electrons between Ru (II) and Rh (III) over distances in excess of 4.0 nm, presumably through the stacked bases, has been observed,[181] as has electron transfer from other ions.[181a] Stacked bases are apparently semiconductors.[182]

## C. The Topology and Dynamics of Nucleic Acids

### 1. Rings, Catenanes, and Knots

While a DNA molecule may exist as a straight rod, the two ends are often covalently joined. Thus, the chromosomes of *E. coli* and of other bacteria are single closed circles. Circular DNA molecules are also found in mitochondria, chloroplasts, and many viruses. Further complexity arises from the fact that the circles of DNA are sometimes interlocked in chainlike fashion (**catenated**). An unusual example of this phenomenon is the presence of thousands of small catenated DNA circles in the single mitochondrion of a trypanosome (Fig. 5-16).[183] Sometimes circular DNA is **knotted** as in Fig. 5-17.[184–186] Knots and catenanes often appear as intermediate forms during replication and recombination, especially involving circular DNA.[187,188]

Methods have been devised for synthesis of even very complex DNA knots.[185,186] Let's look briefly at the topology of knots. The three simple knots shown here have a chirality beyond that of the nucleotide

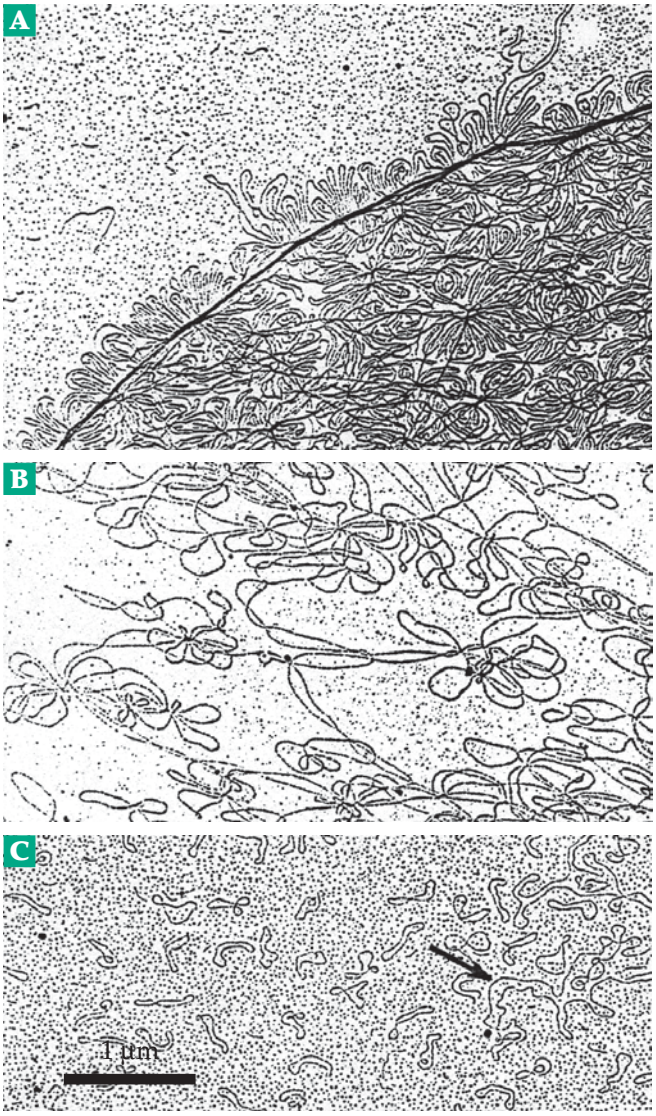
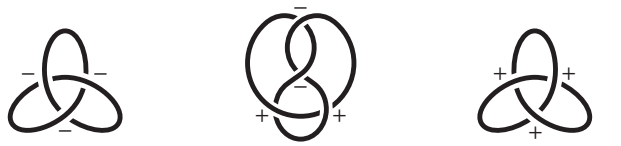



**Figure 5-16** (A) Electron micrograph of the network of catenated DNA circles in the mitochondrion of the trypanosome *Crithidia fasciculata*. (B) and (C) The same network after treatment with a **topoisomerase** from bacteriophage T4 that catalyzes a decatenation to form individual covalently closed circles (Chapter 27). Five times as much enzyme was added in (C) as in (B). Two sizes of circles are present. Most are "minicircles", each containing about 2300 bp but a smaller number of larger ~35-kb "maxicircles" are also present. One of these is marked by the arrow. From Marini, Miller, and Englund.[183]

units.[189,190] This can be expressed by indicating the sign of each node in the knot:



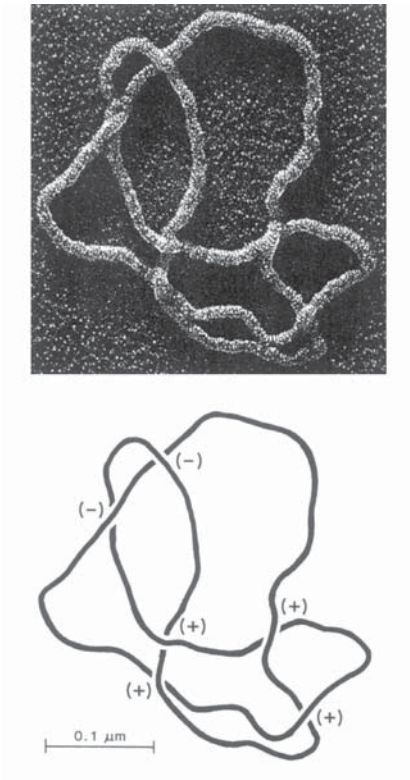
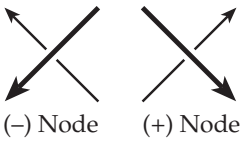




**Figure 5-17** Electron micrograph of a six-noded knot made by the Tn3 resolvase which is involved in movement of the Tn3 transposon (Chapter 27) from one location to another within the genome. Putative six-noded knot DNA was isolated by electroelution from an agarose gel. The knots, which are nicked in one strand, were denatured to allow the nicked strand to slide away and leave a ssDNA knot. This was coated with *E. coli* recA protein (Fig. 27-24) to greatly thicken the strand and to permit the sign of each node (designated in the tracing) to be seen. From Wasserman *et al.*[184]

The node is negative if the crossing is like that in a left-handed supercoil and positive if like that in a right-handed supercoil:[191,192]




## 2. Supercoiled DNA

Double-stranded DNA in solutions of low salt content usually assume the B-DNA conformation with 10.4–10.5 base pairs per turn. If the two ends are joined the resulting covalently closed circular DNA will be "relaxed." However, there are topoisomerases that act on this form of DNA by cutting both strands, holding the ends, and twisting the two chains (Chapter 27).

The energy source for the process is provided by cleavage of ATP. **DNA gyrase** untwists relaxed circular dsDNA one turn at a time and reseals the cut ends. A **reverse gyrase** from certain bacteria twists the relaxed DNA more tightly. In both cases the change causes the DNA to form **superhelical turns**.[67,193–197] These may be either **solenoidal** or **plectonemically interwound** (as a twisted thread; Fig. 5-18).

The geometric and topological properties of closed supercoiled DNA molecues may be described by three quantities: The **linking number** (***Lk*** also called the winding number, α), the **twist** (***Tw)***, and the **writhe** (***Wr***). If a segment of double helical DNA were laid on a flat surface and the ends were joined to form a relaxed circle both *Lk* and *Tw* would equal the number of helical turns in the DNA. The writhe *Wr* would be zero. The linking number is a topological property. It has an integral value which is unchanged if the DNA molecule is distorted. It can be changed only for DNA with open ends. When the ends are joined the linking (winding) number is constant unless one or both chains are cleaved. However, twist and writhe are geometric properties, which can change according to Eq. 5-8 while the *Lk* remains constant. The twist is related to the number of helical turns while the writhe is related to the number of superhelical turns.[27,67,198–200]

$$Lk = Tw + Wr \tag{5-8}$$

However, *Tw* and *Wr* do not usually have integral values. It is hard to define the relaxed state for a circular DNA. For example, changing the ionic composition will alter *Tw* and *Wr* according to equation 5-8. Both *Lk* and *Tw* are taken as positive for a right-handed **toroidal** (solenoidal) supercoil or a left-handed interwound twist.

For relaxed B-DNA, *Lk* is equal to the total number of base pairs in the circle divided by 10.5, $Wr = 0$, and $Tw = Lk$. Both *Lk* and *Tw* are positive in the right-handed B- and A-DNA forms, but *Tw* is negative for Z-DNA. In closed circular DNA the value of *Wr* is usually negative, the secondary structure being a fully formed Watson–Crick helix but with right-handed interwound superhelical turns or left-handed toroidal superhelical turns. The helix is said to be **underwound** ($Lk < Tw$).

Some of the topological properties of double-stranded DNA can be demonstrated by twisting together two pieces of flexible rubber tubing whose ends can be joined with short rods to form a closed circle.[196] Twist the tubing in a right-handed fashion as tightly as possible without causing supercoiling ($Lk = Tw$; $Wr = 0$). If the ends are now joined the circle will be relaxed. Now twist one turn tighter before joining the ends. A right-handed toroidal supercoil will be formed ($\Delta Wr = 1$; $\Delta Lk = 1 + Tw$; *Tw* is the same as before). If twisting is continued until several supercoils appear before the ends are joined, two interconvertible forms result. One form has right-handed toroidal supercoils and the other left-handed interwound supercoils as in Fig. 5-19. If relaxed circular DNA is unwound by one turn by cutting and resealing one chain, a single right-handed interwound supercoil will be formed ($\Delta Wr = -1$, $Lk = Tw - 1$). On the other hand, if the two chains in a closed relaxed helix are pried apart, as happens during intercalation (Section 3), *Lk* must remain constant, *Tw* will decrease, and *Wr* will increase with appearance of left-handed supercoiling.

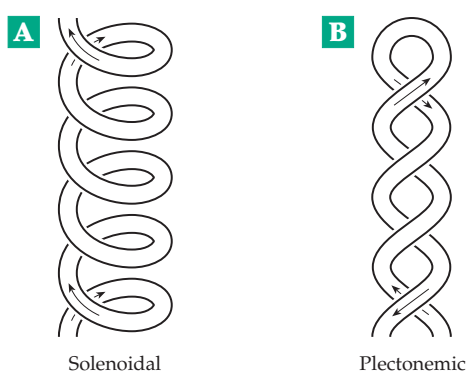



**Figure 5-18** Two forms of supercoiling of a DNA duplex: (A) Solenoidal. (B) Plectonemically interwound. These are both negatively supercoiled, as can be deduced from the arrows at the nodes, but the solenoid is left-handed and the plectonemic form right-handed. From Wasserman and Cozzarelli.[187]

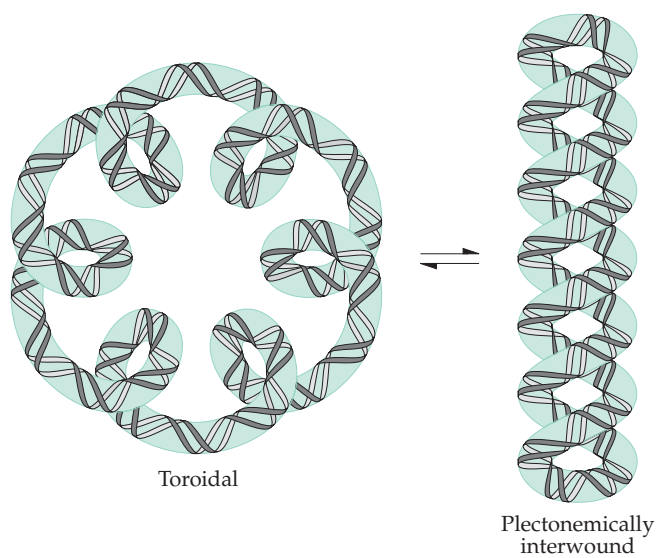



**Figure 5-19** Topological equivalence of toroidal (solenoidal) and plectonemically interwound forms of a circular DNA. These two forms have a constant value of the linking number *Lk* (or α), the twist *Tw*, and writhing number *Wr*.

During replication of DNA (Chapter 27) pairing of some bases associated with the replication apparatus is prevented. Upon release of the constraint the newly replicated DNA forms base pairs and becomes supercoiled.

Since *Lk* is constant in circularly closed DNA a change of one turn of B-DNA ($Tw = 1$) into a turn of Z-DNA ($Tw = -1$) will cause the writhing number *Wr* to change by −2. Conversely, if the writhing number is forced to change by −2, a turn of Z-DNA may develop somewhere in a suitable (G + C)-rich region of the DNA.

Some of the information about supercoiled DNA can be summarized as follows:

| | |
|---|---|
| Relaxed circular DNA | $Lk = Tw$, $Wr = 0$ |
| Underwound circular DNA | $Lk < Tw$; *Wr* is negative; right-handed interwound supercoils or left-handed toroidal supercoils may be present. Alternatively $Lk = Tw$ and left-handed Z-DNA regions may appear. |
| DNA with intercalated molecules | DNA is partially untwisted; *Tw* is lowered; *Wr* is increased with decrease in number of negative supercoils. As *Tw* approaches *Lk* the DNA becomes relaxed. |

The **superhelix density** of a DNA molecule is often expressed as $\sigma = Wr/Tw \approx$ number of superhelical turns per 10 bp.[31,198,199a,199b,201] In most naturally occurring circular DNA molecules σ is negative, a typical value being −0.05 (~5 negative superhelical turns per 1000 bp). The presence of superhelices in circular DNA molecules can be recognized readily by its effect upon the sedimentation constant of the DNA. Naturally occurring supercoiled DNA from polyoma virus sediments rapidly but after nicking of one of the strands of the double helix by brief exposure to a DNA-hydrolyzing enzyme the resulting relaxed form of the molecule sediments more slowly. Supercoiling lowers the viscosity of solutions of DNA and increases the electrophoretic mobility (Fig. 5-20) and may also be recognized by electron microscopy.

Naturally occurring or artificially prepared supercoiled DNA molecules can often be separated by electrophoresis into about ten forms, each differing from the other by one supercoiled turn and by $\Delta Lk = \pm 1$ (Fig. 5-20). The relative amounts of these **topological isomers** form an approximately Gaussian distribution. The isomers apparently arise as a result of thermal fluctuations in the degree of supercoiling at the time that the circles were enzymatically closed.[202]

Why is DNA in cells supercoiled? One effect of supercoiling is to contract the very long, slender double helices into more compact forms. In eukaryotic cells much of the DNA exists in **nucleosomes**. Each bead-like nucleosome consists of a core of eight subunits of proteins called **histones** around which an ~140 bp length of DNA is coiled into two negative, left-handed toroidal superhelical turns (Fig. 5-21).[203,204] There is some spacer DNA between nucleosomes. Otherwise a nucleosome providing two superhelical turns per 200 bp would produce a superhelix density of −0.1, twice that observed. A detailed analysis of the geometric properties of DNA wrapped around nucleosomes or other protein particles has been developed.[199–199c] Nucleosome formation is thought to protect DNA and to keep it in a more compact state than when it is fully active and involved in transcription. Nucleosomes, discussed further in Chapter 27, are also important in the regulation of transcription. Even though nucleosomes as such are absent, bacterial DNA also has a superhelix density of −0.05, apparently a result of interaction with other proteins such as the histonelike HU.[205] Interaction with smaller molecules can also affect supercoiling.[206]

The presence of naturally supercoiled DNA and a variety of topoisomerases suggests that the control of DNA supercoiling is biologically important. In fact,

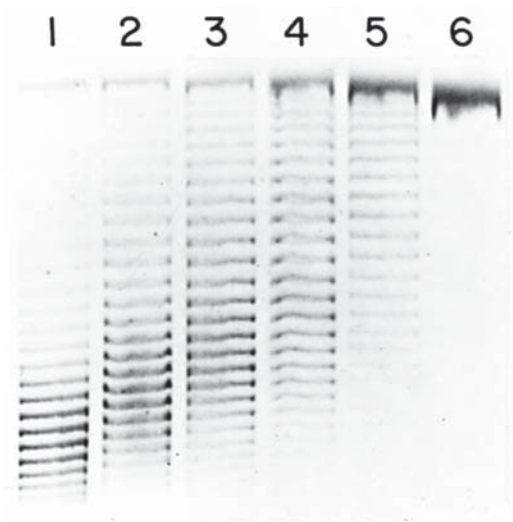



**Figure 5-20** Electrophoresis of DNA from the SV40 virus with varying numbers of superhelical turns. Molecules of the native DNA (lane 1) move rapidly toward the anode as a series of bands, each differing from its neighbors by one superhelical turn. The average number of superhelical turns is about 25. Incubation with a topoisomerase from human cells causes a stepwise removal of the superhelical turns by a cutting and resealing of one DNA strand. The DNA incubated with this enzyme at 0°C for periods of 1, 3, 6, 10, and 30 min (lanes 2–6) is gradually converted to a form with an average of zero supercoils. For details see Keller.[202,212] Electrophoresis was carried out in an 0.5% agarose –1.9% polyacrylamide slab gel (17 x 18 x 0.3 cm). The bands of DNA were visualized by staining with the fluorescent intercalating dye ethidium bromide. Photograph courtesy of Walter Keller.

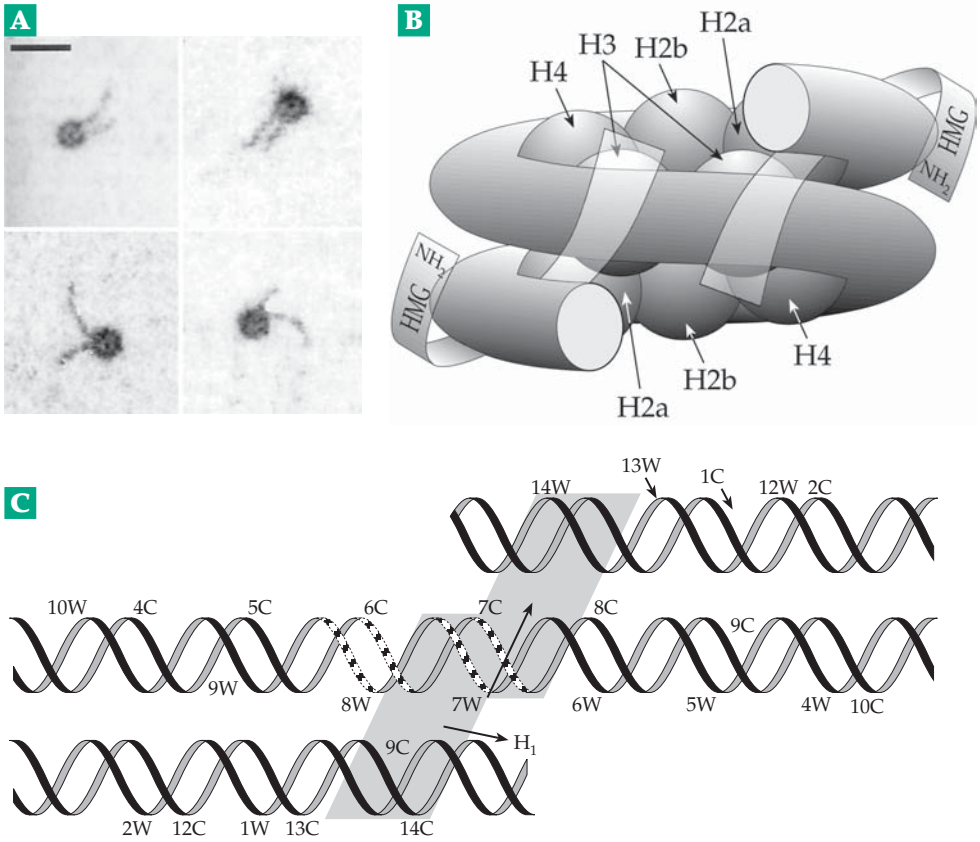



**Figure 5-21** Nucleosomes. (A) Electron micrographs of individual nucleosomes reconstituted from 256-bp DNA fragments and separated proteins. From Hamiche *et al.*[213] Courtesy of Ariel Prunell. (B) Model of a nucleosome core. The 1.75-turn (145-bp) DNA superhelix winds around the histone octomer which consists of two subunits apiece of histones H2A, H2B, H3, and H4. In addition, two elongated molecules of proteins HMG-14 or HMG-17 are indicated (see also Chapter 27). (C) Schematic radial projection of the double-helical DNA showing areas protected from cleavage by hydroxyl radicals (see Fig. 5-50) by the bound proteins. The shaded areas are those protected by HMGs. The zigzag lines near the dyad axis indicate the most prominent regions of protection. (B) and (C) are from Alfonso *et al.*[214]

topoisomerases play essential roles in both replication of DNA and in transcription of genes. Supercoiling requires energy and topoisomerases that induce supercoiling must provide energy, e.g., by cleavage of ATP. Conversely, supercoiled DNA can be a source of energy for biological processes. It has been estimated that the $\Delta H$, $\Delta S$, and $\Delta G$ per mole for formation of a single superhelical turn are 35.9 kJ, 68 J/ °K, and 14.6 kJ, respectively.[207] A reduction in superhelix density is accompanied by a decrease in Gibbs energy and can therefore be coupled to other processes that have positive values of $\Delta G$. An example is conversion of (G + C)-rich regions into the Z form of DNA, which is favored by negative supercoiling.[207,208] Supercoiling affects binding of various proteins to DNA[197,209,210] as well as intercalation (discussed in the next section) and formation of cruciform structures (Section D,3).

The folding of DNA into compact forms, such as that in chromosomes, is also influenced by supercoiling. In the absence of nucleosomes supercoiled DNA may assume the plectonemic form (Figs. 5-18, 5-19). Segments of the resulting rods may aggregate side-by-side in a liquid crystalline state.[200,211] The relatively high cation concentration within cells favors this transformation.[215–217] Polyamines such as spermidine are especially effective in promoting aggregation of DNA, in formation of Z-DNA,[110] and possibly in facilitating cooperative processes that require that two DNA molecules interact with each other.[218,219]

## 3. Intercalation

Flat, aromatic, hydrophobic rings are often able to insert themselves between the base pairs of a DNA duplex. Such **intercalation** is observed for many antibiotics, drugs, dyes, and environmental pollutants. Among them are proflavine, ethidium bromide, actinomycin (Box 28-A), hycanthone (Fig. 5-22), and daunomycin (Fig. 5-23). Hycanthone, employed in the treatment of schistosomiasis, is one of the most widely used drugs in the world. Since intercalating agents can be mutagenic, such drugs are not without their hazards.

Intercalation is often used to estimate the amount of negative supercoiling of DNA molecules. Varying amounts of the intercalating agent are added, and the sedimentation constant or other hydrodynamic property of the DNA is observed. As increasing intercalation occurs, the secondary turns of DNA are unwound (the value of *Tw* in Eq. 5-8 decreases). Each intercalated ring causes an unwinding of the helix of ~26°. Since for a closed covalent duplex the value of *Lk* in Eq. 5-8 is constant, the decrease in *Tw* caused by increased intercalation leads to an increase in the value of *Wr*,

which is usually negative for natural DNA. When sufficient intercalation has occurred to raise *Wr* to zero, a minimum sedimentation rate is observed. Addition of further intercalating agent causes a positive supercoiling.

When the "replicative form" of DNA of the virus ϕX174 (Chapter 27), a small circular molecule containing ~5000 bp, was treated with proflavine,[221] the binding of 0.06 mol of proflavine per mole of nucleotides reduced *Wr* to zero. From this it was estimated that $\sigma = 0.055$, corresponding to −27 superhelical turns at 25°, pH 6.8, ionic strength ~0.2. Changes in temperature, pH, and ionic environment strongly influence supercoiling. In general σ becomes less negative by ~$3.3 \times 10^{-4}$ per degree of temperature increase.[222] For example, the observed value[221] of σ for ϕX174 DNA was −0.059 at 15°C and −0.040 (−20 superhelical turns) at 75°C at an ionic strength of ~0.2.

The exact ways in which intercalating substances can fit between the base pairs of nucleic acids are being revealed by X-ray diffraction studies of complexes with nucleosides, dinucleotides, and other oligonucleotides.[220,223–225] The structure of a complex in which daunomycin is intercalated between two GC base pairs in DNA is shown as in Fig. 5-23. Daunomycin and other related anthracycline antibiotics also have an amino sugar ring that binds into the minor groove of the DNA, providing both electrostatic stabilization and hydrogen bonding. Substituents on this aliphatic ring also hydrogen bond to DNA bases.[226]

Does intercalation of flat molecules into nucleic acid chains have a biochemical function? Aromatic rings of amino acid side chains in proteins designed to interact with nucleic acids may sometimes intercalate into nucleic acid helices serving a kind of "bookmark" function.[227] Changes in superhelix density caused by such intercalation may be important in the orderly handling of DNA by enzymes within cells.

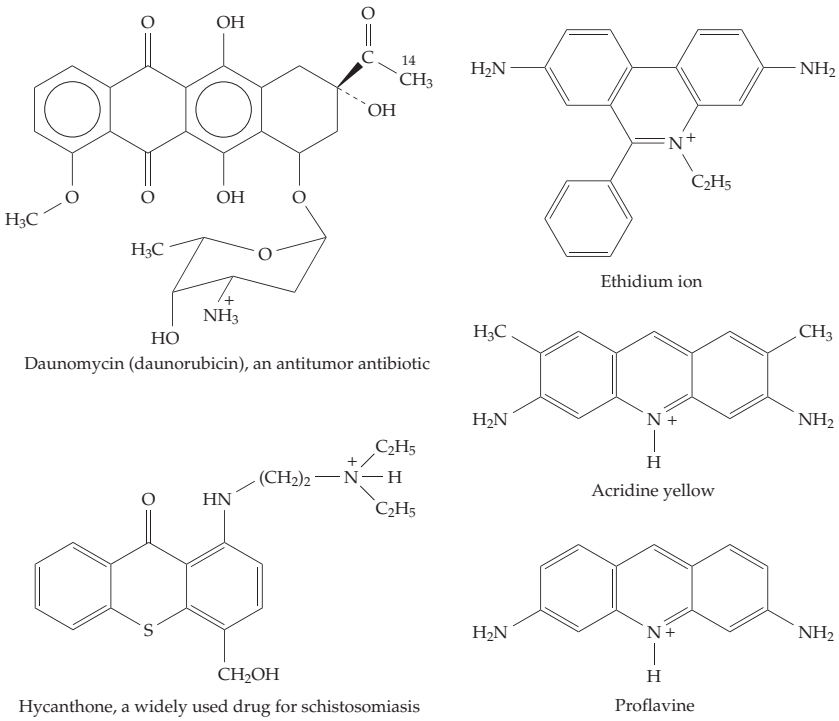



**Figure 5-22** Structures of some substances that tend to "intercalate" into DNA structures. See also Fig. 5-23.

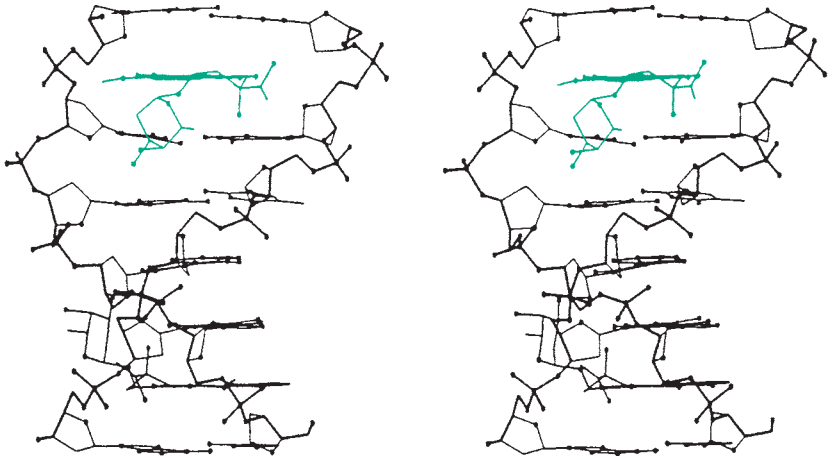


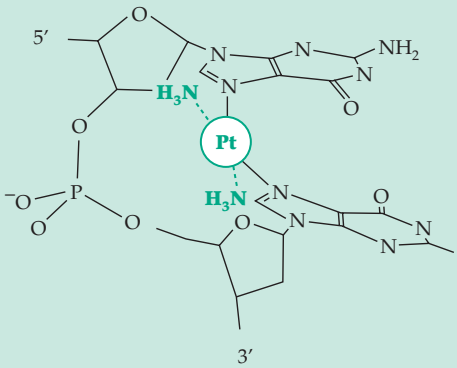
**Figure 5-23** Stereoscopic drawing showing a molecule of daunomycin (Fig. 5-22) intercalated between two base pairs in a molecule of double-helical DNA, d(CGTACG). Nitrogen and oxygen atoms are shown as dots. From Quigley *et al.*[220] Both daunomycin and adriamycin (doxorubicin; 14-hydroxydaunomycin) are important but seriously toxic anticancer drugs.

## BOX 5-B ANTITUMOR DNA DRUGS

Chemotherapy of cancer at present involves simultaneous use of two or more drugs. For example, antifolates (Chapter 15) or nucleoside analogs such as 5-fluorouridine may be used together with a drug that binds directly to DNA and inhibits the replication of cancer cells. In 1963, it was discovered accidentally that platinum ions released from supposedly inert platinum electrodes inhibited the division of *E. coli* cells. This led Rosenberg and associates to test platinum compounds against animal cells.[a] Among many compounds tested *cis*-dichlorodiammineplatinum(II) (*cis*-DDP or **cisplatin**) emerged as an important anticancer drug that is especially effective against testicular and ovarian cancers.[b–d] The *trans* isomer, however, is inactive.

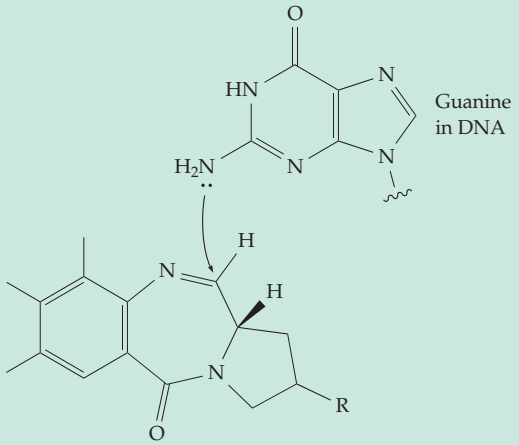
*cis*-DDP (Cisplatin) *trans*-DDP (inactive)

*cis*-DDP binds to adjacent deoxyguanosines within one strand of ds or ssDNA. The stacking of adjacent bases is disrupted as the platinum binds to N-7 nitrogen atoms of the two guanine rings by replacement of the two chloride ions. The product has the following structure,[c,e] with the Pt lying in the minor groove of dsDNA:



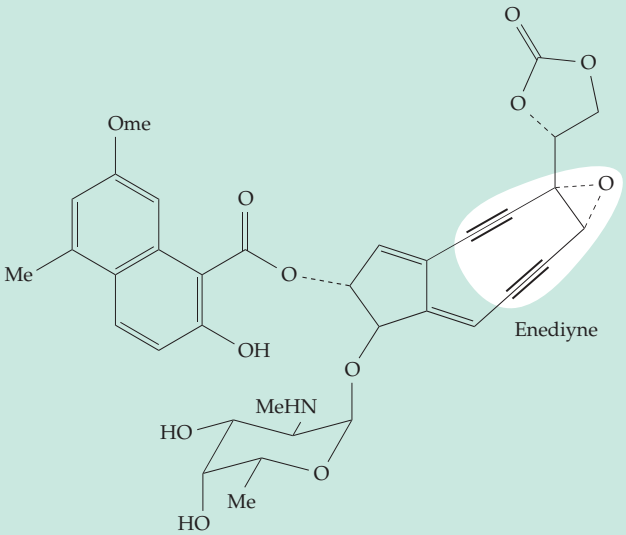

The two guanines are no longer stacked, but the structure is impossible for the *trans* isomer. The cisplatin adducts appear to prevent proper DNA repair and to induce programmed cell death (apoptosis).[f]

Since the 1950s, using a different approach, the U.S. National Cancer Institute, as well as agencies in other countries, has sought to find natural anticancer compounds in plants, fungi, microorganisms, and marine invertebrates.[g] Among these are many antibiotics that intercalate into DNA helices, e.g., **daunomycin** (Figs. 5-22, 5-23), **menogaril**,[h] **triostin A**,[i] and the antitrypanosomal drug **berenil.**[j] Some of these are also alkylating agents that contain double bonds to which such groups as the 2-$NH_2$ of guanine may add:



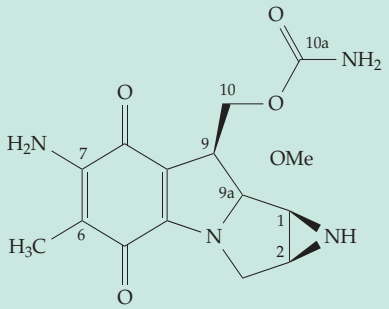

Tomaymycin and related antibiotics[k]

Diol epoxides and cyclic imines such as **mitomycin** also form adducts specifically with guanine 2-$NH_2$ groups:[l] **Neocarzinostatin** is an antitumor protein with a nonprotein "chromophore." After intercalation and binding into the minor groove of bulged DNA, it undergoes "activation" by addition of a thiol group. The enediyne structure undergoes rearrangement with formation of a reactive diradical that attacks the DNA.[m] A family of related antitumor enediynes has also been discovered.[n]

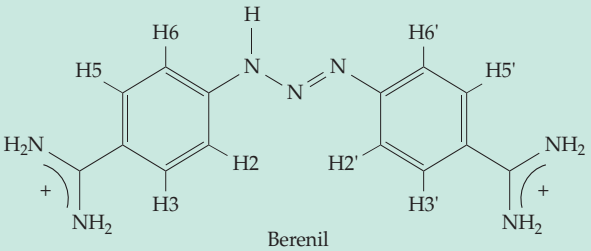



Neocarzinostatin chromophore

**BOX 5-B (continued)**



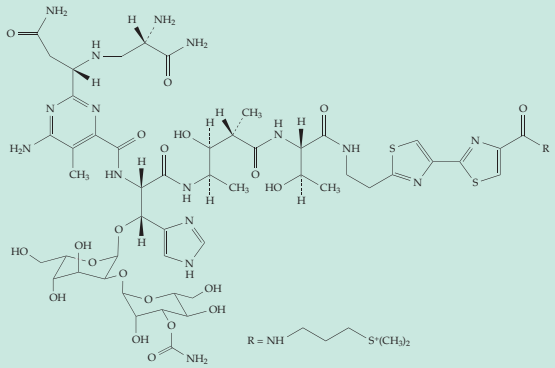

Mitomycin




Berenil

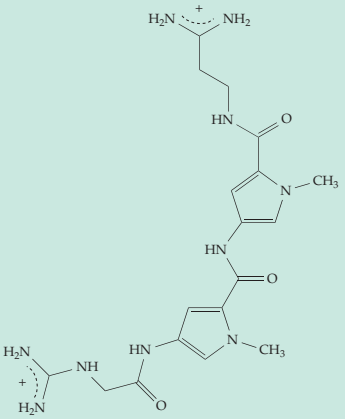
Another group of drugs occupy extended binding sites in the minor grooves of DNA double helices, often with specificity for a particular base sequence. Examples are the antitrypanosomal drug **berenil**,[j] toxic *Streptomyces* antibiotics **netropsin, distamycin**, and related synthetic compounds.[o–q] Netropsin lies within the minor groove in regions with two or more consecutive AT pairs, displacing the spine of hydration as shown in the following stereoscopic drawing. Binding depends upon both electrostatic interactions and formation of specific hydrogen bonds involving the amide groups of the antibiotics.

The antibiotic **bleomycin**, which also binds in the minor groove of B-DNA with some specificity for G-C Sites, forms an iron (II) complex. It can be

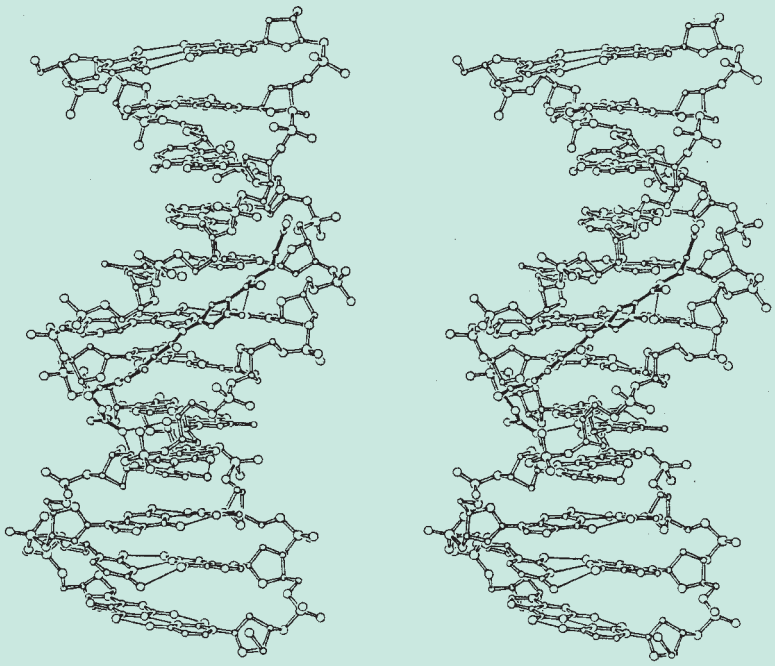



Bleomycin $A_2$

oxygenated to form an Fe(II)–$O_2$ complex (see Chapter 16) which cleaves the DNA chain.[s,t] Synthetic compounds that do the same thing have been made by connecting an EDTA–iron, or other iron chelate complex covalently to the DNA-binding compound.[n,u–x] A goal is to direct drugs to selected target sites in DNA and to induce bond cleavage at those sites in a manner analogous to that observed with restriction endonucleases. Most current chemotherapeutic agents are very toxic. Present research is designed to target these drugs more precisely to specific DNA sequences and to identify target sequences peculiar to cancers. See also **Designed third strands** in main text.



Netropsin



Netropsin lying in the minor groove of B-DNA hydrogen-bonded to bases in the central ATAT tetranucleotide. From Coll *et al.*[r]

**BOX 5-B Continued**


[a] Rosenberg, B., Van Camp, L., and Krigas, T. (1965) *Nature (London)* **205**, 698–699
[b] Zamble, D. B., and Lippard, S. J. (1995) *Trends Biochem. Sci.* **20**, 435–439
[c] Pilch, D. S., Dunham, S. U., Jamieson, E. R., Lippard, S. J., and Breslauer, K. J. (2000) *J. Mol. Biol.* **296**, 803–812
[d] Temple, M. D., McFadyen, W. D., Holmes, R. J., Denny, W. A., and Murray, V. (2000) *Biochemistry* **39**, 5593–5599
[e] Huang, H., Zhu, L., Reid, B. R., Drobny, G. P., and Hopkins, P. B. (1995) *Science* **270**, 1842–1845
[f] Zlatanova, J., Yaneva, J., and Leuba, S. H. (1998) *FASEB J.* **12**, 791–799
[g] Booth, W. (1987) *Science* **237**, 969–970
[h] Chen, H., and Patel, D. J. (1995) *J. Am. Chem. Soc.* **117**, 5901–5913
[i] Wang, A. H.-J., Ughetto, G., Quigley, G. J., Hakoshima, T., van der Marel, G. A., van Boom, J. H., and Rich, A. (1984) *Science* **225**, 1115–1121
[j] Pilch, D. S., Kirolos, M. A., Liu, X., Plum, G. E., and Breslauer, K. J. (1995) *Biochemistry* **34**, 9962–9976
[k] Barkley, M. D., Cheatham, S., Thurston, D. E., and Hurley, L. H. (1986) *Biochemistry* **25**, 3021–3031
[l] Kumar, G. S., Lipman, R., Cummings, J., and Tomasz, M. (1997) *Biochemistry* **36**, 14128–14136
[m] Yang, C. F., Stassinopoulos, A., and Goldberg, I. H. (1995) *Biochemistry* **34**, 2267–2275
[n] Nicolaou, K. C., Dai, W.-M., Tsay, S.-C., Estevez, V. A., and Wrasidlo, W. (1992) *Science* **256**, 1172–1178
[o] Goodsell, D. S., Ng, H. L., Kopka, M. L., Lown, J. W., and Dickerson, R. E. (1995) *Biochemistry* **34**, 16654–16661
[p] Rentzeperis, D., Marky, L. A., Dwyer, T. J., Geierstanger, B. H., Pelton, J. G., and Wemmer, D. E. (1995) *Biochemistry* **34**, 2937–2945
[q] Tanious, F. A., Ding, D., Patrick, D. A., Tidwell, R. R., and Wilson, W. D. (1997) *Biochemistry* **36**, 15315–15325
[r] Coll, M., Aymami, J., van der Marel, G. A., van Boom, J. H., Rich, A., and Wang, A. H.-J. (1989) *Biochemistry* **28**, 310–320
[s] Burger, R. M., Drlica, K., and Birdsall, B. (1994) *J. Biol. Chem.* **269**, 25978–25985
[t] Kane, S. A., Hecht, S. M., Sun, J.-S., Garestier, T., and Hélène, C. (1995) *Biochemistry* **34**, 16715–16724
[u] Veal, J. M., and Rill, R. L. (1988) *Biochemistry* **27**, 1822–1827
[v] Campisi, D., Morii, T., and Barton, J. K. (1994) *Biochemistry* **33**, 4130–4139
[w] Mack, D. P., and Dervan, P. B. (1992) *Biochemistry* **31**, 9399–9405
[x] Han, H., Schepartz, A., Pellegrini, M., and Dervan, P. B. (1994) *Biochemistry* **33**, 9831–9844


## 4. Polynucleotides with Three or Four Strands

Some nucleotide sequences in DNA favor the formation of a regular triple-helical (**triplex**) structure. This is possible when there are long stretches of adjacent pyrimidines having any sequence of C and T in one strand of double-helical DNA. The other strand of the DNA will contain the correct purines for formation of Watson–Crick base pairs. With such a structure it is always possible to add a third strand of a polypyrimidine using Hoogsteen base pairing. Triads formed in this way contain either two T's and one A or two C's and one G, as is shown in Fig. 5-7. In the latter case one of the C's must be protonated to allow formation of the pair of hydrogen bonds.[227a] These triplex structures can be formed only for stretches of DNA containing all pyrimidines in one strand (a **homopyrimidine** strand) and all purines in the other (a **homopurine** strand). A poly (AAU) triplex can also be formed, using the hydrogen bond pattern of the first triplet in Fig. 5-7. Triple-stranded synthetic polynucleotides of these types were prepared by Felsenfeld and others as early as 1957.[228–230] The third strand in a triplex can be either parallel or antiparallel to the homopurine strand. The third strand is either homopurine or a mixture of purine and thymine. The triplets are G•C*C, A*A•T, or T*A•T and are formed by Watson–Crick (•) and **reversed Hoogsteen** (*) pairing.[231] Recently, there has been a renewed interest in DNA triplets because of their occurrence in natural DNA, their possible importance in genetic recombination, and the potential for design of powerful inhibitors of replication and transcription that function via triplex formation.[35,232–238]

***H-DNA.*** Strands of DNA contain many homopyrimidine "tracts" consisting of repeated sequences of pyrimidines, e.g., $d(T\text{-}C)_n$, which may be abbreviated more simply as $(TC)_n$. In this example the complementary strand would contain the two purines in the repeated sequence $(GA)_n$. At low pH, where protonation of the cytosine rings occurs, or in negatively supercoiled DNA, the two strands of the repeating sequence may separate, with the $(TC)_n$ strand folding back to form a triple helix in which the base triplets have the hydrogen bonding pattern of Fig. 5-7. The resulting structure, which is shown in Fig. 5-24, is known as **H-DNA**.[239–241] A variety of related "nodule" and looped forms of DNA can also be formed.[35,242]

***R-DNA.*** A different type of triplex DNA may be formed during genetic recombination. A Watson–Crick duplex is brought together by one or more proteins with a single strand that is, for at least a considerable distance, an exact copy of one of the strands of the duplex. It is within such a triplex that cutting of a strand of the duplex takes place to initiate recombination (see Chapter 27 for a detailed discussion). Can a triplex structure containing two identical chains be formed? Possible base triplets include the following C•G*G triplet which occurs in a crystalline oligonucleotide structure.[243] This triplet contains a variation on Hoogsteen pairing and is related to the first triplet in Fig. 5-7. The other triplets needed for the proposed **R-DNA** recombination intermediates are G•C*C, T•A*A, and A•T*T. While there is keen interest in R-DNA[35,244–247] the formation of a stable intermediate triplex is still uncertain.[35,247]

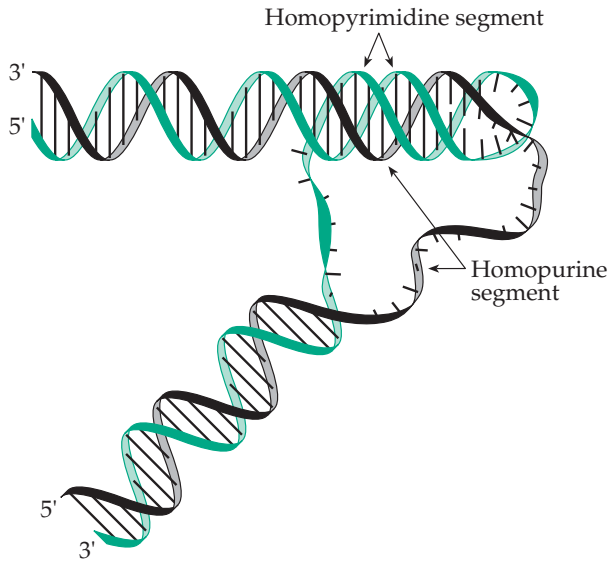
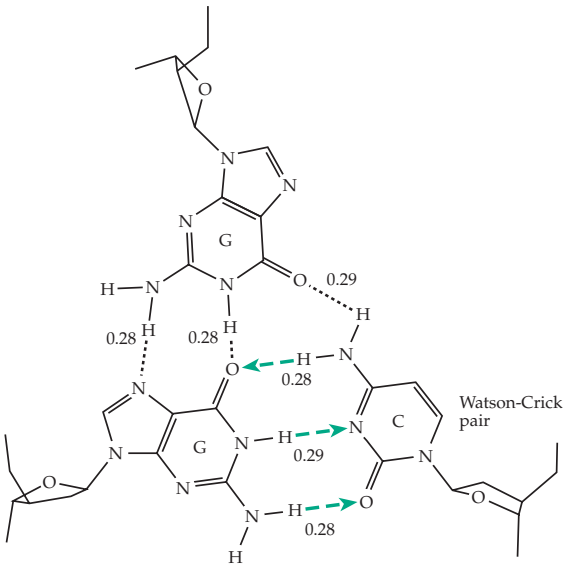
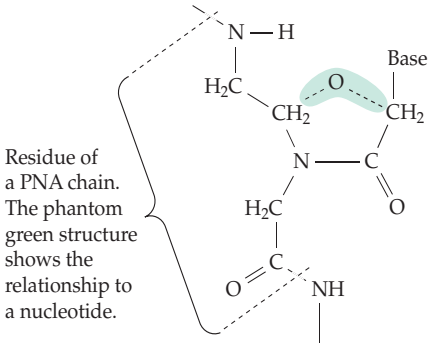



**Figure 5-24** Proposed structure of H-DNA which can be formed when a homopyrimidine segment and its complementary homopurine segment separate as a result of protonation of the cytosine rings or of negative supercoiling stress.[239,240] The triple-helical portion contains base triplets of the kind shown in Fig. 5-7.



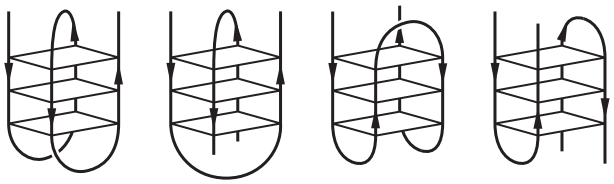

***Designed third strands.*** If triplex DNA segments can form naturally it should be possible to design oligonucleotides that will bind into the major groove of a DNA duplex to form a triplex at a specific location or locations in the genome. Dervan and associates are studying this approach systematically.[248–250] Such oligonucleotides may be modified chemically to provide stronger binding to targeted locations and may prove to be useful therapeutic agents. Among these are synthetic phosphoramidates in which the phosphodiester linkage between nucleosides is replaced by 3'-NH–$PO_3^-$–O-5'. The resulting oligonucleotide is resistant to digestion by phosphodiesterases present in cells. Its geometry favors binding to A-DNA or RNA.[251] Another oligonucleotide mimic, dubbed PNA (peptide nucleic acid), contains monomer units of the following type[252–253b] and is able to form a triple helix:




Yet another approach is to synthesize a "hairpin polyamide" that contains pyrrole, hydroxypyrrole, and imidazole groups in a sequence that favors tight binding to a specific dsDNA sequence.[254] Some possible geometries for triplex DNA are illustrated in Fig. 5-25. These are based on computer-assisted modeling.[238]

***Tetraplex (quadruplex) structures.*** The ends of linear chromosomes, the **telomeres**, have unusual nucleotide sequences repeated hundreds or thousands of times.[255–258] There is usually a **guanine-rich strand** running 5' to 3' toward the end of the chromosome and consisting of sequences such as TTAGGG in vertebrates, TTTTGGGG in the ciliate *Oxytricha*, and $TG_{1-3}$ in *Saccharomyces cerevisiae*. The complementary strand is cytosine rich. The guanine-rich strand is longer than its complement, "overhanging" by about 2 repeat units. The significance for the replication of chromosomes is discussed in Chapter 27.

The thing that has attracted most attention to telomeric DNA is the unusual structure of the G-rich strand that was signaled by the first NMR studies.[259] Subsequent investigation[260,261] revealed the presence of G quartets (Fig. 5-8) which are apparently stacked in folding patterns such as the following:[255,262–268]



These structures are stabilized by the presence of univalent cations, $K^+$ being more effective than $Na^+$,

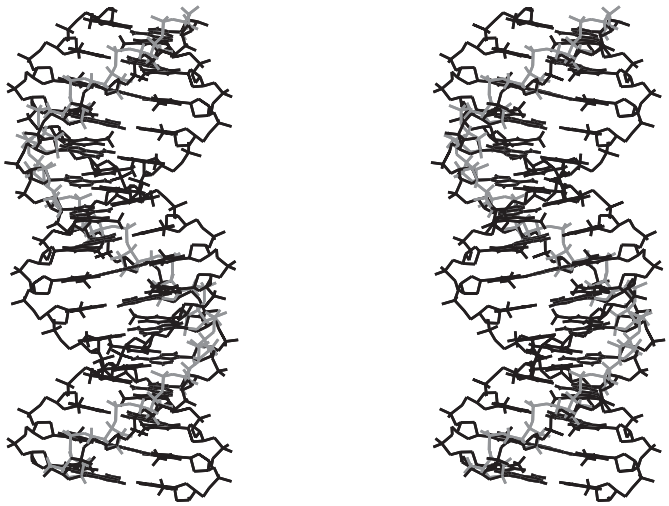
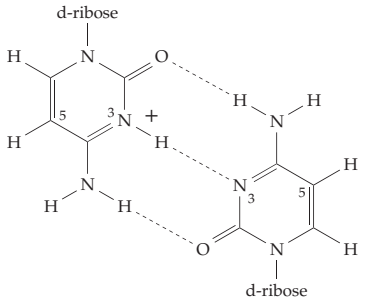


**Figure 5-25** Two superposed stereoscopic diagrams illustrating the lowest energy triple helical conformations of one pyrimidine, one purine, and one Hoogsteen DNA strand (in black). Superposed is an RNA purine strand in gray. See Srinivasan and Olson.[238] Courtesy of Wilma K. Olson.

$Li^+$, $Rb^+$, or $Cs^+$.[260,269–271] A protein isolated from *Tetrahymena* binds specifically to the G4-DNA regions.[272] The presence of 5-methylcytosine ($m^5c$) also stabilizes quadruplex structures.[273] The oligoribonucleotide $UG_4U$ forms $G_4$ quartets so stable that it takes days at 40°C in $D_2O$ for the hydrogen-bonded NH protons forming the quartet to be exchanged for $^2H$.[274]

Tetraplex structures have also been observed for G-rich repeating sequences associated with the human **fragile X syndrome**.[275,276] This is the most common cause of inherited mental retardation and appears to arise as a result of the presence of an excessive number of repeats of the trinucleotide sequence $(CGG)_n$. For normal persons n = 60 or less; for healthy carriers n may be as high as 200 but for sick individuals it may be much higher.[275] The structure in solution, as determined by NMR spectroscopy, is shown in Fig. 5-26. Another variant of four-stranded DNA, which arises from cytosine-rich DNA, contains $C{\bullet}CH^+$ pairs such as the following at low pH.[277–279]




Hemiprotonated $d(CH^+{\bullet}C)$

In sequences such as $d(TC_5)$ and $d(C_3T)$ these $C{\bullet}CH^+$ pairs are intercalated as is shown in Fig. 5-27. This intercalated DNA (**I-DNA**) may provide an alternative conformation for some telomeric sequences.[281,282] An I-DNA motif has also been identified in oligonucleotides from the DNA of human centromeres.[283] The seemingly unusual forms of DNA described in this section may represent only a fraction of the naturally occurring DNA structures of biological significance.

## 5. Junctions

Special structural features may be found at junctions between different types of DNA, e.g., between A-DNA and B-DNA.[284–286] However, the most interesting junctions are *branched*.[287–290] For example, Fig. 5-28 shows a four-way junction in which all of the bases form Watson–Crick pairs. This junction is better known as a **Holliday junction** because it was proposed by Holliday in 1964 as an intermediate in genetic recombination.[291] As shown at the top of Fig. 5-28A the junction is formed from *two homologous DNA duplexes*. These are identical except for the boxed and shaded base pairs. The ends of the first duplex are marked I and II and those of the second III and IV. The Holliday junction appears to arise by cleavage of one strand of each duplex with rejoining of the strands as indicated by the green arrows. Rotation gives the untwisted Holliday junction structure

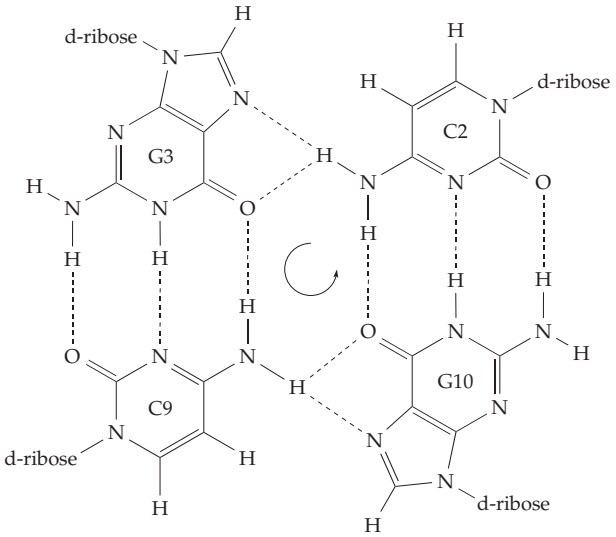



**Figure 5-26** Structure of a G•C∗G•C∗ tetrad present in a quadruplex structure formed by the oligonucleotide d(GCGCTTTGCGC) in $Na^+$-containing solution. See Kettani *et al.*[276]

shown. The postulated three-dimensional structure of the junction is indicated in Fig. 5-28B.[292–294]

An important characteristic of Holliday junctions formed from homologous duplexes is that they can move by a process called **branch migration**.[295] Because of the twofold symmetry of the branched structure the hydrogen bonds of one base pair can be broken while those of a new base pair are formed, the branch moving as shown in Fig. 5-28. Notice that, in this example, the nonhomologous (boxed) base pairs TA and GC have become *mispaired* as TG and AC after branch migration. More significantly, the junction may be cut by a **resolvase** at the points marked

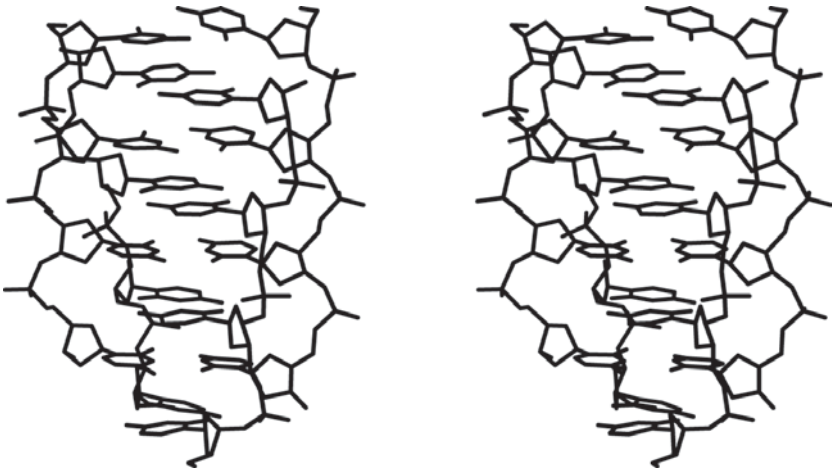
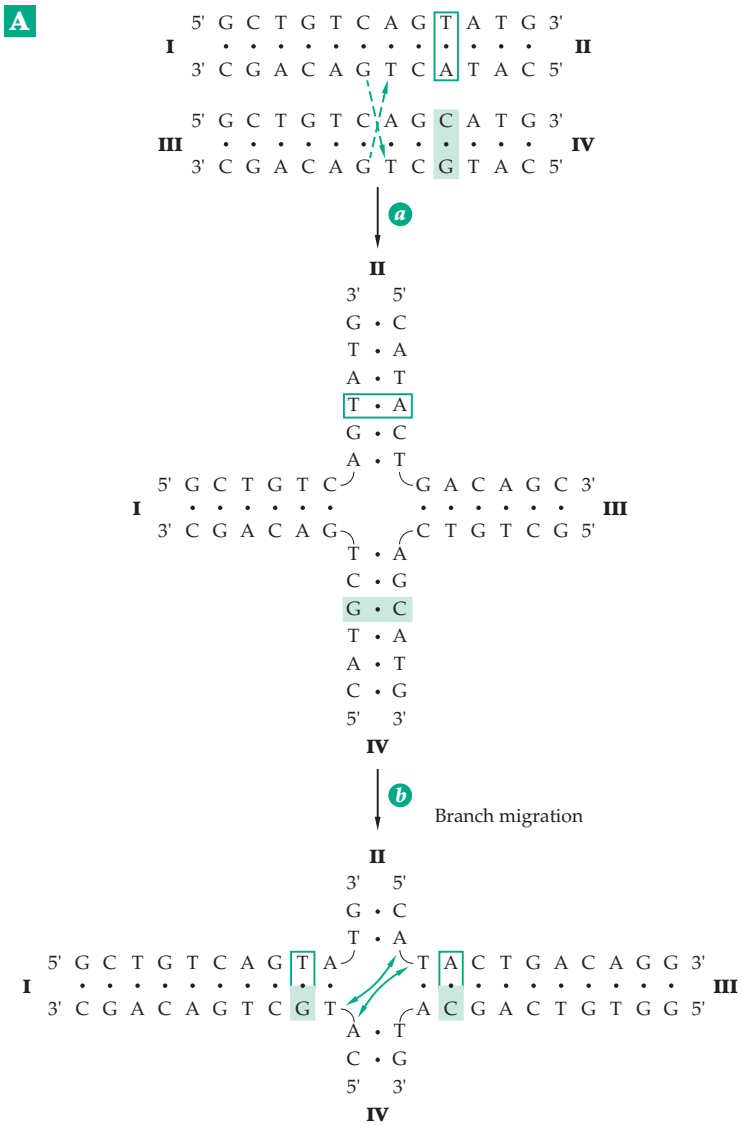


**Figure 5-27** Stereoscopic view of a four-stranded intercalated DNA or I-DNA formed from d($C_4$). Two parallel duplexes with C•$CH^+$ pairs are intercalated into each other. From Chen *et al.*[280]



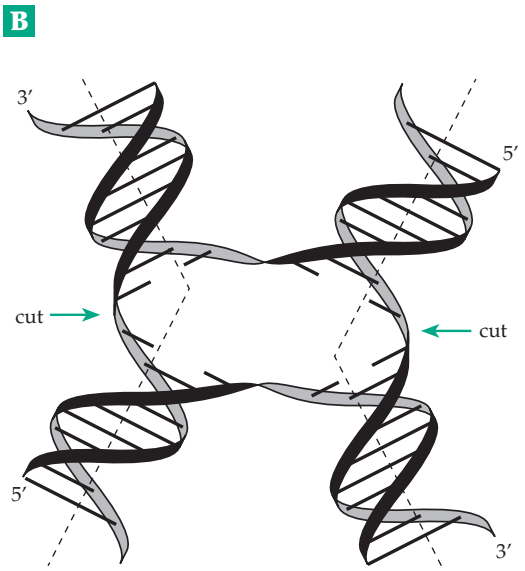




**Figure 5-28** (A) Abbreviated reaction sequence for formation of a four-way Holliday junction between two homologous DNA duplexes. In step *a* strands are cut and rejoined with movement of the strands to a roughly antiparallel orientation. The resulting structure is thought to resemble that shown below the four-stranded representation. In step *b* branch migration takes place, separating the nonhomologous base pairs TA and CG and causing mismatched pairs which will be subject to repair. (B) Proposed three-dimensional structure (after drawing by Bennett and West).[292]

by the green double-headed arrow. If the strand break is then resealed, and the DNA strands are replicated, the boxed base pair in duplex I–II will have been transferred into a strand replicated from III–IV and genetic recombination will have been accomplished. Branch migration can occur over much longer distances than are indicated in this figure, so the alteration transferred may be far from the site of the initial cleavage and whole genes or groups of genes can be transferred. Recombination is considered in more detail in Chapter 27.

The Holliday junctions formed during recombination are mobile, but synthetic **immobile Holliday junctions** can be synthesized by using nonhomologous base sequences or by locking the junctions.[288] This has permitted careful physical study of these and other more elaborate synthetic junctions. With suitable choices of base sequences for the oligonucleotides from which they are made, such junctions will assemble spontaneously. Double-stranded DNA with immobile junctions is a very suitable construction material on a "nanochemical" scale. It has been assembled into knots, rings, cubes, and more complex polyhedra.[296–297b]

## D. Ribonucleic Acids (RNA)

The best known forms of RNA are: (1) the long chains of messenger RNA (mRNA), which carry genetic messages copied from DNA to the ribosomes where proteins are made; (2) the much shorter transfer RNAs (tRNAs) which participate in reading the genetic code, correctly placing each amino acid in its sequence in the proteins; and (3) ribosomal RNAs (rRNA), which provide both structural material and a catalytic center for peptide bond formation. In addition there are numerous small RNAs that function in the splicing and editing[298–300] of mRNA, processing of tRNA precursors, methylation of ribosomal RNA,[301] transfer of proteins across membranes, and replication of DNA.[302] The genomes of many viruses consist of RNA. There are doubtless additional as yet undiscovered types of RNA.

Unlike DNA, which exists largely as double helices, the single chains of RNA can fold into complex forms containing many bulges and loops of the sort depicted in Fig. 5-9.[300,303,304] These loops are closed by double-stranded **stems** which have the A conformation. The B conformation is impossible because of the presence of the 2'-hydroxyl groups on the ribose rings in RNA. Even one ribonucleotide in a 10-nucleotide oligomer prevents formation of the B structure.[55,83] The 2'-OH groups not only keep the RNA in the A form but also engage in hydrogen-bond formation. Hydrogen bonds may form between the 2'-OH and the oxygen atom in the next ribose ring in the 3' → 5' direction. The –OH groups also hydrogen bond to water molecules which form a network within the minor groove.[305,306] These bound water molecules, in turn, can bond to associated protein and to other atoms of the complex loops found in RNA molecules. The 2'-OH groups also act as ligands for divalent metal ions in some tRNAs and in some RNA catalytic sites. Transient **hybrid DNA–RNA double helices** also exist within cells and they too usually have the overall shape of A-DNA.[55,307–309] However, the minor groove is intermediate in width between that expected for the A and B forms.

### 1. RNA Loops and Turns

Like polypeptides, polynucleotide chains have preferred ways of bending or turning. The loops at the ends of the hairpin turns of RNA molecules sometimes consist of a trinucleotide such as UUU,[304] but are usually larger. In tRNA there are typically seven bases that do not participate in regular Watson–Crick pairing (see Figs. 5-30, 5-31). The tetranucleotide **UUCG** is frequently present, and the sequence 5'-GGAC**UUCG**GUCC forms an unusually stable hairpin.[85,310] Other tetranucleotides, such as UGAA,[311] CCCG (also found in DNA loops),[312] GCAA,[313] and GAAA,[314] occur often. The latter are members of a larger group of loop structures with the consensus sequence GNRA, where N is any nucleotide and R is a purine.[315,315a] These sequences are very common in highly folded structures of ribosomal RNAs. Until recently high-resolution X-ray structures were available for only a few tRNAs and ribozymes. To help remedy this deficiency the structures of a great variety of oligonucleotide stem/loop (hairpin) structures are being determined, most by NMR spectroscopy.[304,316,317] The sharp turns in the loops involve mostly rotation about the two torsion angles around the phosphorus atom of the third, from the 5' end, of the seven nucleotides. Base bulges on stems (Fig. 5-9) not only introduce kinks and bends in RNA stems[169] but also provide well-defined hydrogen-bonded binding sites for proteins. Numerous branched three-way and more complex junctions provide other important motifs in folded RNA.[318]

Among the new RNA structures are those of RNA–antisense RNA pairs in "kissing" hairpin complexes.[87,319] Another interesting complex folding pattern in RNA is the **pseudoknot**, a structural feature that has been identified in many RNA sequences.[85,320–327] A pseudoknot can be formed if nucleotide sequences favorable to formation of two short RNA stems are overlapped as shown in Fig. 5-29. After stem 2 in this drawing is formed (step *a*) additional base pairing can lead to formation of stem 1 (step *b*). The base pairs of the two stems can stack coaxially to form the pseudoknot (step *c*).

For the formation of base-paired stems the RNA must contain antiparallel sequences that allow Watson

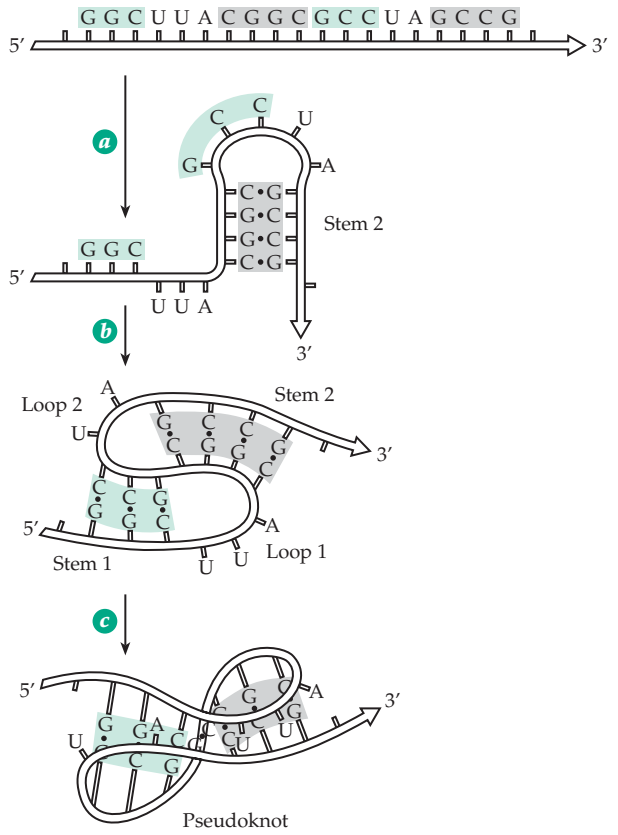


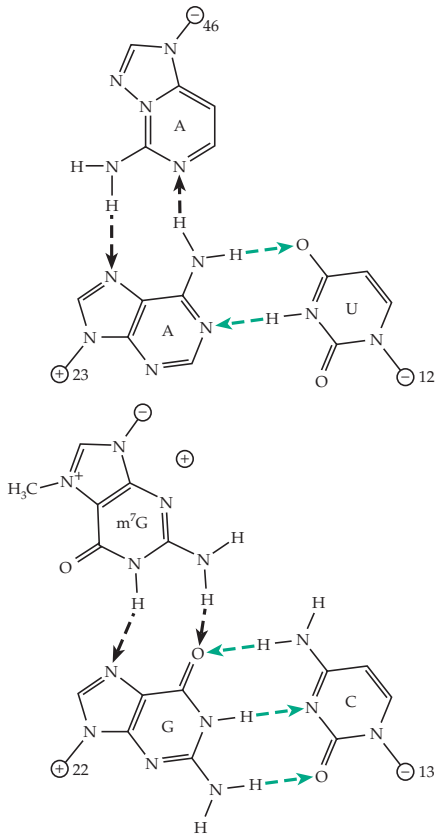

**Figure 5-29** Formation of a pseudoknot in an RNA chain. After Puglisi *et al.*[325]

–Crick or wobble (GU) base pairing in the stems. This requires certain relationships in the sequences of the DNA in the genes that encode these molecules as discussed in Section E,3. Because of the base-pairing requirements, some of the bases in the stems protrude as bulges and fail to form pairs. Often, it is possible to find more than one reasonable structure, each having some bases unpaired.[328]

## 2. Transfer RNAs

In all tRNAs the bases can be paired to form "cloverleaf" structures with three hairpin loops and sometimes a fourth as is indicated in Fig. 5-30.[329–331] This structure can be folded into the L-shape shown in Fig. 5-31. The structure of a phenylalanine-carrying tRNA of yeast, the first tRNA whose structure was determined to atomic resolution by X-ray diffraction, is shown.[170,332–334] An aspartic acid-specific tRNA from yeast,[335] and an *E. coli* chain-initiating tRNA, which places *N*-formylmethionine into the N-terminal position of proteins,[336,337] have similar structures. These molecules are irregular bodies as complex in conformation as globular proteins. Numerous NMR studies show that the basic structure is conserved in all tRNAs. However, animal mitochondrial tRNAs often lack some of the usual stem-loop "arms" as well as the invariant nucleotides in the dihydrouridine and TψC loops (Fig. 5-30).[307,338,339] At the bottom of the structure as shown in Figs. 5-30 and 5-31 is the **anticodon**, a triplet of bases having the correct structures to permit pairing with the three bases of the codon specifying a particular amino acid (see Table 5-5), in this case phenylalanine.

While tRNAs consist largely of loops and stems containing Watson–Crick base pairs, they also contain Hoogsteen pairs, wobble pairs, and triplets such as the following.




The first of these contains a Watson–Crick A•U pair with a second A bound to form a reversed Hoogsteen A∗A pair. The second contains a G•C Watson–Crick pair with a Hoogsteen N-7-guanine∗G ($m^7G$∗G) pair. Among the complex base associations present in tRNAs are some that also involve hydrogen bonding to the 2'-hydroxyl groups of ribose rings and to at least one of the phosphate groups. There are over 100 internal hydrogen bonds, a large proportion of which are relatively invariant among the known tRNAs.

## 3. Properties and Structures of Some Other RNA Molecules

A few specialized RNA molecules are listed in

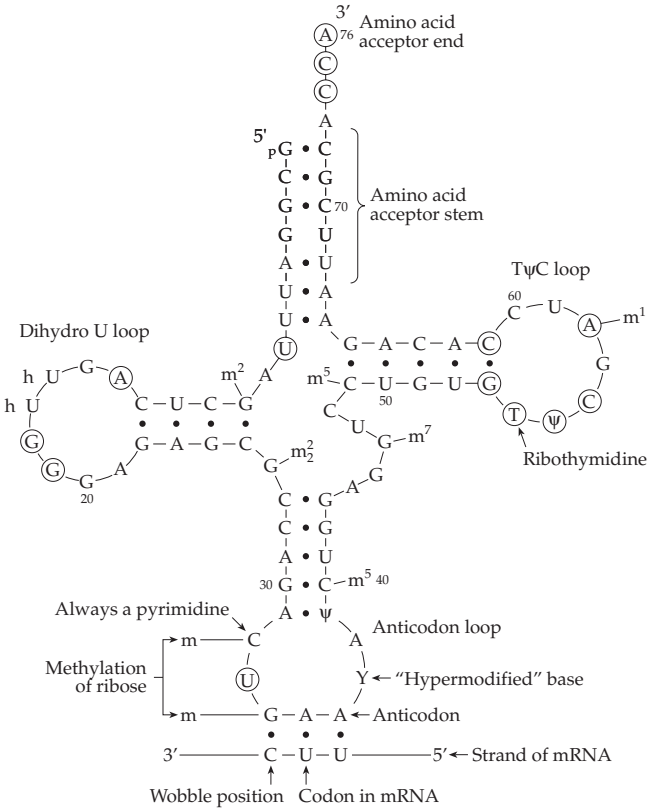



**Figure 5-30** Schematic cloverleaf structure of a phenylalanine-specific transfer RNA ($tRNA^{Phe}$) of yeast. The dots represent pairs or triplets of hydrogen bonds. Nucleosides common to almost all tRNA molecules are circled. Other features common to most tRNA molecules are also marked. The manner in which the anticodon may be matched to a codon of mRNA is indicated at the bottom.

Table 5-4. There are many others.[302] In addition, there are thousands of different mRNAs within a cell. The most abundant RNA molecules in cells are those of the ribosomes. Ribosomes consist of two elaborate RNA-protein subunits, a large subunit with sedimentation constant ~ 30S in bacteria and ~ 40S in eukaryotes. The small subunit contains 16S or 18S RNA and the large subunit 23S or 28S as well as smaller 5S and 5.8S RNAs (Table 5-4; Table 28-1). A proposed three-dimensional structure[341,342] of a bacterial 16S ribosomal RNA and the corresponding ribosomal subunit with its 21 proteins are shown in Fig. 5-32. It might seem impossible that the folding pattern of the RNA was deduced correctly before an X-ray structure was available. However, a **phylogenetic approach**, the comparison of nucleotide sequences among several species, suggested that the stem structures of rRNAs are highly conserved (see discussion in Chapter 29). This fact, together with a variety of other chemical

## TABLE 5-4
## Some Specialized RNA Molecules

| Kind of RNA | | Number of nucleotides |
|---|---|---|
| Transfer RNAs (Figs. 5-30, 5-31) | | 60 – 85 |
| Ribosomal RNAs | | |
| 5S | | ~ 120 |
| 5.8S | (rat) | 158 |
| 16S | (*E. coli*; Fig. 5-32) | 1542 |
| 18S | (rat) | 1874 |
| 23S | (*E. coli*) | 2904 |
| 28S | (rat) | 4718 |
| Telomerase guide RNA[a-d] | | 159 (*Tetrahymena*) |
| M1 RNA of Ribonuclease P[e,f] | | 350 – 410 |
| | | 377 (*E. coli*) |
| *Tetrahymena* Intron ribozyme (Fig. 12-26)[g] | | 413 |
| Viroid hammerhead ribozyme (Fig. 12-27)[h] | | ~ 55 |
| Signal recognition particle 7S RNA[i] | | 295 (human) |
| Small nuclear RNAs[j,k] | | 65 – 1200 |
| RNA-editing guide RNA[l] | | ~ 60 |
| Thermotolerance factor (G8 RNA)[m] | | ~ 300 |
| Viroid RNA (Fig. 28-19) | | 240 – 380 |
| Virus MS2 genome (Chapter 29) | | 3569 |

[a] Greider, C. W., and Blackburn, E. H. (1989) *Nature (London)* **337**, 331–337
[b] Bhattacharyya, A., and Blackburn, E. H. (1994) *EMBO J.* **13**, 5721–5731
[c] Singer, M. S., and Gottschling, D. E. (1994) *Science* **266**, 404–409
[d] Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A. A., Chiu, C.-P., Adams, R. R., Chang, E., Allsopp, R. C., Yu, J., Le, S., West, M. D., Harley, C. B., Andrews, W. H., Greider, C. W., and Villeponteau, B. (1995) *Science* **269**, 1236–1241
[e] Stark, B. C., Kole, R., Bowman, E. J., and Altman, S. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3717–3721
[f] Mattsson, J. G., Svärd, S. G., and Kirsebom, L. A. (1994) *J. Mol. Biol.* **241**, 1–6
[g] Cech, T. R. (1987) *Science* **236**, 1532–1539
[h] Hertel, K. J., Herschlag, D., and Uhlenbeck, O. C. (1994) *Biochemistry* **33**, 3374–3385
[i] Li, W.-Y., Reddy, R., Henning, D., Epstein, P., and Busch, H. (1982) *J. Biol. Chem.* **257**, 5136–5142
[j] Maxwell, E. S., and Fournier, M. J. (1995) *Ann. Rev. Biochem.* **35**, 897–934
[k] Nicoloso, M., Qu, L.-H., Michot, B., and Bachellerie, J.-P. (1996) *J. Mol. Biol.* **260**, 178–195
[l] Kable, M. L., Seiwert, S. D., Heidmann, S., and Stuart, K. (1996) *Science* **273**, 1189–1195
[m] Fung, P. A., Gaertig, J., Gorovsky, M. A., and Hallberg, R. L. (1995) *Science* **268**, 1036–1039

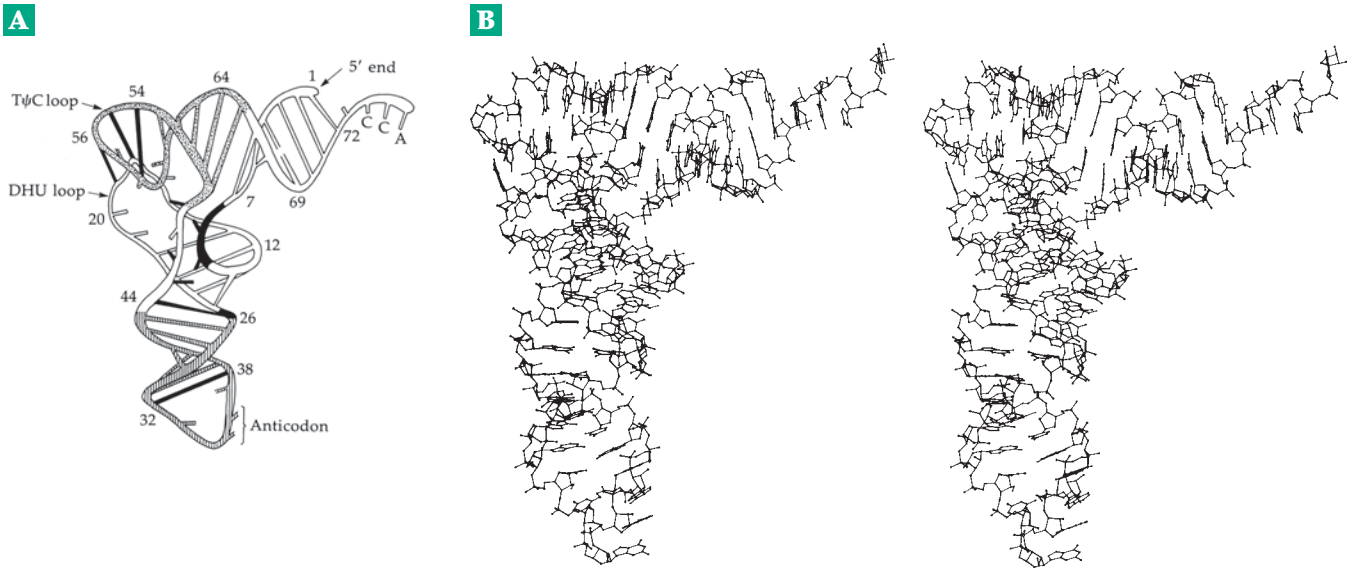



**Figure 5-31** The three-dimensional structure of a phenylalanine-specific transfer RNA of yeast. (A) Perspective diagram of folding of polynucleotide chain. The ribose phosphate backbone is drawn as a continuous cylinder with bars to indicate hydrogen-bonded base pairs. The positions of single bases are indicated by rods which are intentionally shortened. The TψC arm is heavily stippled, and the anticodon arm is marked by vertical lines. Tertiary structure interactions are illustrated by black rods. Redrawn from Quigley and Rich.[340] (B) Stereoscopic view of the structure of yeast $tRNA^{Phe}$ as revealed by X-ray crystallography. The acceptor stem with the protruding ACCA sequence at the 3' end is to the right. The anticodon GAA is at the bottom right side of the drawing. The guanine ring is clearly visible at the very bottom. The middle adenine of the anticodon is seen exactly edge-on as is the "hypermodified" base Y (see Fig. 5-33) which lies just above the anticodon. Its side chain is visible at the back of the drawing. Preceding the anticodon on the left (5' side) are two unpaired bases (C and U). The 2'-hydroxyl groups of the cytosine and of the guanosine in the anticodon are methylated. Moving far up the anticodon stem one can see two groups of base triplets which utilize both Watson–Crick and Hoogsteen types of base pairing. Drawing courtesy of Alexander Rich.

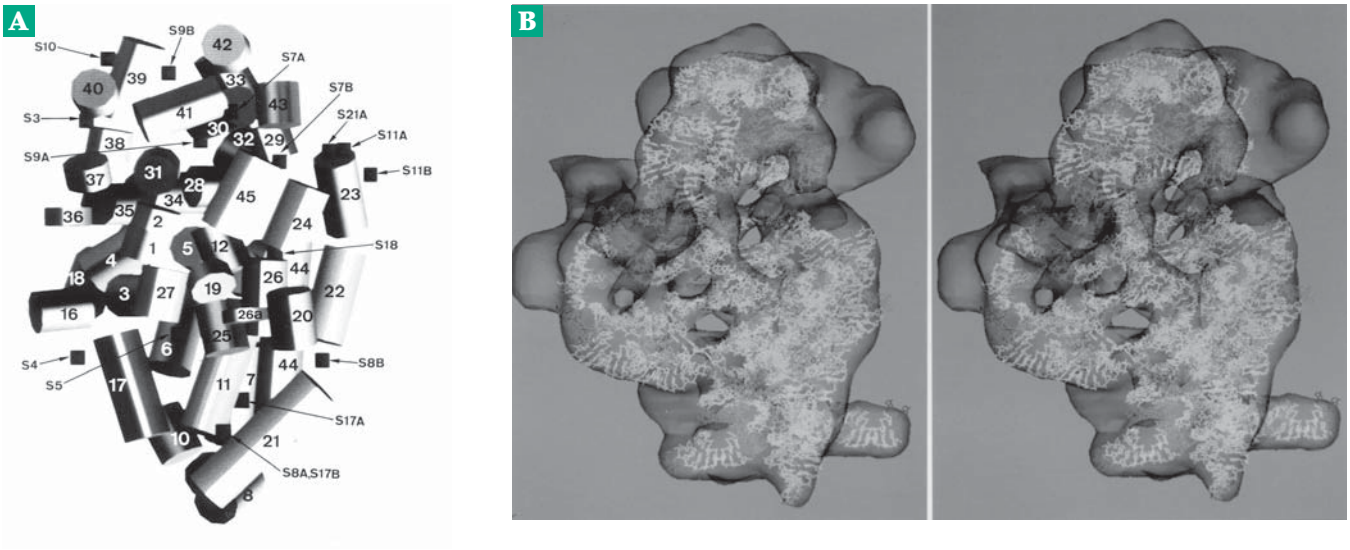



**Figure 5-32** (A) A three-dimensional computer graphics model proposed by Brimacombe *et al.*[341] for the single chain of *E. coli* 16S ribosomal RNA. The helices are depicted as cylinders, which are all connected. The small dark squares denote the positions of artificially formed RNA–protein crosslinks, marked with the appropriate protein number. For proteins exhibiting more than one crosslink site (e.g., S17), the sites are denoted A or B, in each case A being the site nearer to the 5' terminus of the 16S RNA. (B) Stereoscopic view of tentative atomic model of 16S RNA in the 30S ribosomal subunit. The viewing direction is different from that in (A). From Mueller and Brimacombe.[342] Courtesy of Richard Brimacombe.

evidence,[341,343,344] allowed the prediction of the structure shown as well as many characteristics of the rRNA of the large subunit.[345] Culminating decades of effort[346–350] (Chapter 29), complete structures of bacterial ribosomes were established by 2000[342,351–354] and the peptidyl transferase center was identified as a ribozyme[355–357] (discussed in Chapter 12).

## 4. Modified Nucleotides

The picture of DNA or RNA as chains of only four kinds of nucleotides is not quite accurate. DNA contains a significant number of methylated bases among which are **5-methylcytosine** (5mC or $m^5C$) and 6-methyladenine (6mA). The former is regularly present in the nuclear DNA of higher animals and plants. In mammals from two to seven percent of the cytosine is methylated. It is likely that these methylated bases mark special points in the genetic blueprint. As is discussed in Chapter 27, methylation, which is accomplished after the synthesis of the polynucleotide, may block the expression of certain genes.[358] This appears to happen when one of the two X-chromosomes becomes inactivated in cells of females. Methylation may also be used to **imprint** certain genes, i.e. to mark them as coming from a specific one of the two parents.[359]

Another function of methylation is to protect DNA from attack by enzymes (restriction endonucleases) formed in response to invading viruses (Chapter 26). Some viruses, notably the bacteriophage of the T-even series that attack *E. coli* (Box 7-C), have developed their own protective devices. They contain **5-hydroxymethylcytosine** (HOMeC) in place of cytosine. The extra hydroxyl groups provided in this fashion often carry one or two glucose units in glycosidic linkage.[360] A bacteriophage attacking *Bacillus subtilis* substitutes hydroxymethyluracil for uracil and 5-dihydroxypentyluracil for thymine, and phage W14 of *Pseudomonas acidovorans* substitutes the 5-methyl of thymine with $-CH_2-NH-(CH_2)_4-NH_3^+$.[360,361]

The modifications carried out on RNA molecules are more varied and more extensive than those of DNA. Sixty or more modification reactions are known for tRNA, with the number and the extent of modification depending upon the species. Structures of some of the modified bases are indicated in Fig. 5-33. Uridine can be methylated either on the base or on the 2'-hydroxyl of the sugar. Methylation at the 5 position of uridine yields **ribothymidine**. Cytidine can be modified in the same positions. Reduction of the 5,6 double bond of uridine gives **dihydrouridine** (hU). Replacement of the oxygen at position 4 by sulfur gives **4-thiouridine** ($^4sU$). Positions in the guanosine structure that can be methylated are also indicated in Fig. 5-33. The symbol m is commonly used to designate methylation in nucleic acid bases; $m_2$ indicates dimethylation, e.g., 6,6-dimethyladenine is abbreviated $m_2^6A$.

A remarkable transformation is that of uridine into pseudouridine (ψ).

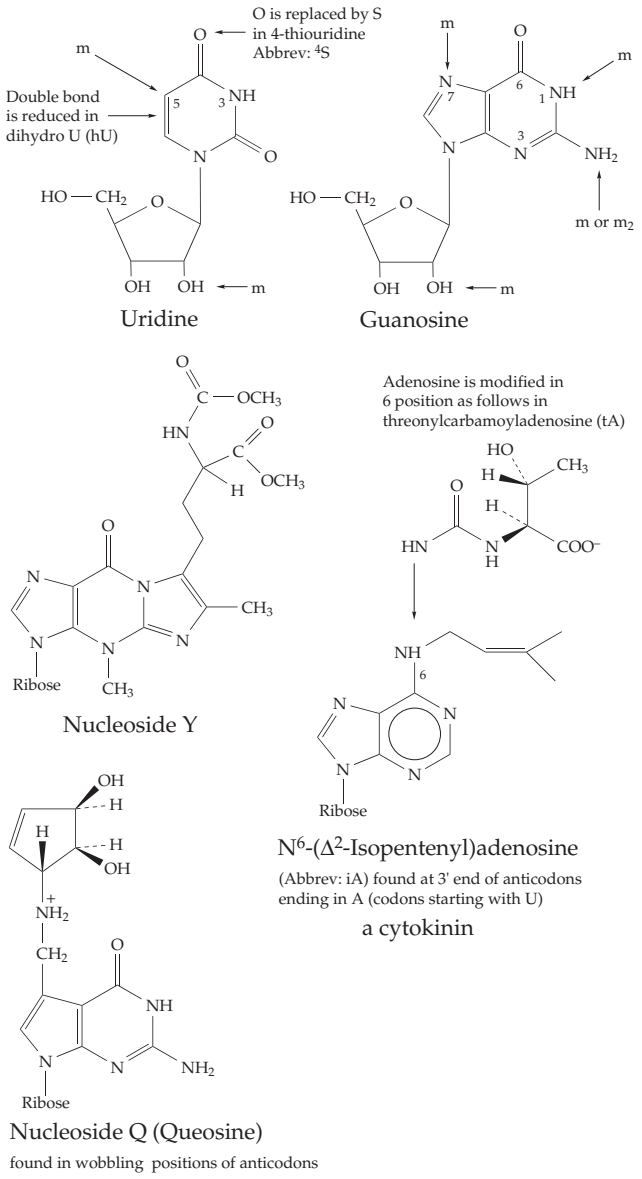
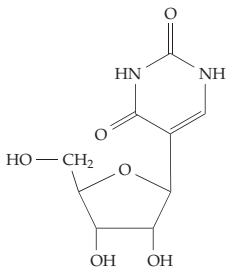



**Figure 5-33** Structures of some nucleosides containing modified bases and found in tRNA molecules. Positions where methylation may occur are designated m. Some other abbreviations are e (ethyl), ac (acetyl), and f (formyl).[25]



## BOX 5-C THE RNA WORLD

The discovery in the 1980s that RNA molecules often have catalytic properties and may serve as true enzymes (ribozymes; Chapter 12) stimulated new thinking about evolution. Although RNA catalysts are not as fast as the best enzymes they are able to catalyze a wide variety of different reactions. Could it be that in the early evolution of organisms RNA provided both the genetic material and catalysts? The "RNA world" would have been independent of both DNA and protein.[a,b] Later DNA could have been developed as a more stable coding molecule and proteins could have evolved as more efficient catalysts. Plausible reactions by which both cytosine and uracil could have arisen in drying ponds on early Earth have been demonstrated.[c]

A major objection to the RNA world is the lack of stability of ribose and the inability to demonstrate the nonenzymatic synthesis of ribose in significant amounts. Even if ribose were present, it would be largely in the pyranose ring forms. Initial formation of the 5-phosphate would be required to allow formation of a nucleotide with a furanose ring. These and other obstacles to the RNA world have led to the suggestion that some *other* genetic material preceded RNA and DNA.[d,e] One possibility is a peptide-like RNA analog.[e] A simple coding system could also have been used, e.g. one based on only two bases, such as C and G, instead of four.[f,g]

Perhaps it is more probable that formation of proteins *and* the present coding system evolved simultaneously? The major metabolic cycles (Chapter 10) could also have developed at the same time. RNAs could not have been the *first catalysts*. Hydrogen ions, hydroxyl ions, ammonium, cyanide, and other simple ions as well as amines and peptides could all have played a role in prebiotic chemistry. Another speculation suggests an "iron-sulfur world" in which organic materials would be formed on mineral surfaces through reactions involving reduction of bicarbonate by iron sulfide and $H_2S$.[h]

If the RNA world did exist, has it left us with any real clues? Benner *et al.* suggest that modern metabolism is a palimpsest of the RNA world, a parchment that has been inscribed two or more times, with previous texts imperfectly erased and therefore still partially legible.[i] If so, can we find a way to read the text of the ancient RNAs?

---

[a] Gesteland, R. F., Cech, T. R., and Atkins, J. F., eds. (1999) *The RNA World*, 2nd ed. Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
[b] Orgel, L. E. (1994) *Sci. Am.* **271** (Oct), 77 – 83
[c] Robertson, M. P., and Miller, S. L. (1995) *Nature (London)* **375**, 772 – 774
[d] Larralde, R., Robertson, M. P., and Miller, S. L. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8158 – 8160
[e] Böhler, C., Nielsen, P. E., and Orgel, L. E. (1995) *Nature (London)* **376**, 578 – 581
[f] Sievers, D., and von Kiedrowski, G. (1994) *Nature (London)* **369**, 221 – 224
[g] Piccirilli, J. A. (1995) *Nature (London)* **376**, 548 – 549
[h] Maden, E. H. (1995) *Trends Biochem. Sci.* **20**, 337 – 341
[i] Benner, S. A., Ellington, A. D., and Tauer, A. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7054 – 7058

Pseudouridine is formed by enzymatic rearrangement of uridine in the original transcript (Eq. 28-3). It can form a base pair with adenine in the same manner as does uracil. Pseudouridine is found not only in tRNA but also in several places in both large and small ribosomal RNA subunits. For example, it is present at position 516 in the *E. coli* 16S RNA,[364] at a specific position in the 23S RNA, and at many more locations in eukaryotic rRNA.

The bases called "Y" and "Q" are highly modified guanines (Fig. 5-33). Q is found at the 5' end of some anticodons in the "wobble position" (see Fig. 5-30). Two **hypermodified** adenosines are also shown in Fig. 5-33. The $N^6$-isopentenyladenosine is found at the 3' end of the anticodons that pair with codons starting with U. This compound is also a plant hormone, a **cytokinin** (Chapter 30). Another highly modified purine, threonylcarbamoyladenine, occurs adjacent to the end of anticodons pairing with codons starting with A. The function of these hypermodified bases is uncertain, but they appear to promote proper binding to ribosomes. The modifications are often not absolutely essential for function.

Another source of modified bases in both DNA and RNA is spontaneous or "accidental" alteration. Nucleic acids encounter many highly reactive and mutagenic materials including hydroxyl radicals, formed from $O_2$, and are able to convert guanine rings into 7,8-dihydro-8-oxoguanine.[362] Other reactive and carcinogenic compounds can form adducts with nucleic acid bases.[363] See Eq. 5-18 and also Chapter 27.

### 5. RNA Aptamers

Ellington and Szostak[365] synthesized a random "pool" of ~$10^{15}$ different oligodeoxyribonucleotides, each ~100 nucleotides in length. They "amplified" these using the polymerase chain reaction (PCR; Section H,6) and prepared a mixture of the corresponding RNAs by *in vitro* transcription. From the ~$10^{13}$ different sequences still present they selected individual

oligonucleotides by affinity chromatography on columns that contained well-defined immobilized ligands such as organic dyes. They called the selected RNAs **aptamers**. Their approach is being used to find RNA sequences that bind to such ligands as ATP, FMN,[366] the bronchodilator theophylline,[367] aminoglycoside antibiotics,[368] arginine,[369] etc.[370] Many of the selected aptamers bind their ligands very tightly and studying them may shed light on interactions of RNA with proteins and on the catalytic activities of RNA, which are discussed in Chapter 12.

## E. The Genetic Code

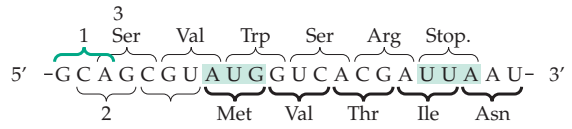
The general nature of the genetic code was suggested by the structure of DNA. Both DNA and proteins are linear polymers. Thus, it was logical to suppose that the sequence of the bases in DNA codes for the sequence of amino acids. There are only four bases in DNA but 20 different amino acids in proteins at the time of their synthesis. It is obvious that each amino acid must be specified by some combination of more than one base. While 16 pairs of bases are possible, this is still too few to specify 20 different amino acids. Therefore, it appeared that at least a triplet group of three nucleotides would be required to code for one amino acid.[371] Sixty-four ($4^3$) such triplet **codons** exist, as is indicated in Tables 5-5 and 5-6.

Simplicity argues that the genetic blueprint specifying amino acid sequences in proteins should consist of consecutive, nonoverlapping triplets. This assumption turned out to be correct, as is illustrated by the DNA sequence for a gene shown in Fig. 5-5. In addition to the codons that determine the sequence of amino acids in the protein, there are **stop codons** that tell the ribosomal machinery when to terminate the polypeptide chain. One methionine codon serves as an **initiation codon** that marks the beginning of a polypeptide sequence. One of the valine codons sometimes functions in the same way.

How does a cell read the code? This question is dealt with in detail in Chapters 28 and 29. A key step is the positioning of each amino acid on the ribosome in proper sequence. This is accomplished by the pairing of codons of messenger RNA with the anticodons of the appropriate transfer RNA molecules as is indicated at the bottom of Fig. 5-30. Each tRNA carries the appropriate "activated" amino acid at its 3' end ready to be inserted into the growing peptide.

### 1. The "Reading Frames"

It is immediately obvious that there are three ways of reading the genetic code in mRNA depending upon which nucleotide is used to start each codon. For example, in the following mRNA sequence either codons GCA, CAG, or AGC could be selected as first.




These codons define the three reading frames or phases in which the code may be read. Here, the term *frame*

**TABLE 5-5**
**The Genetic Code[a]**

| Amino acid | Codons | Total number of codons |
|---|---|---|
| Alanine | GCX | 4 |
| Arginine | CGX, AGA, AGG | 6 |
| Asparagine | AAU, AAC | 2 |
| Aspartic acid | GAU, GAC | 2 |
| Cysteine | UGU, UGC | 2 |
| Glutamic acid | GAA, GAG | 2 |
| Glutamine | CAA, CAG | 2 |
| Glycine | GGX | 4 |
| Histidine | CAU, CAG | 2 |
| Isoleucine | AAU, AUC, AUA | 3 |
| Leucine | UUA, UUG, CUX | 6 |
| Lysine | AAA, AUG | 2 |
| Methionine (also initiation codon) | AUG | 1 |
| Phenylalanine | UUU, UUC | 2 |
| Proline | CCX | 4 |
| Serine | UCX, AGU, AGC | 6 |
| Threonine | ACX | 4 |
| Tryptophan | UGG | 1 |
| Tyrosine | UAU, UAC | 2 |
| Valine (GUG is sometimes an initiation codon) | GUX | 4 |
| Termination | UAA (*ochre*)<br>UAG (*amber*)<br>UGA | 3 |
| Total | | 64 |

[a] The codons for each amino acid are given in terms of the sequence of bases in messenger RNA. From left to right, the sequence is from the 5' end to the 3' end. The symbol X stands for any one of the four RNA bases. Thus, each codon symbol containing X represents a group of four codons.

does not designate a single codon, although frame does designate a single exposure in a motion picture film. **Reading frame** designates which of the three possible sets of codons we are using. In the foregoing sequence the codons in reading frames 2 and 3 are labeled. Reading frame 2 contains the initiation codon AUG (shaded) which could mark the beginning of an encoded protein sequence. Reading frame 3 contains a termination (stop) codon which, when the mRNA transcript is read by ribosomes, will terminate polypeptide synthesis. It may be in the position shown but not have any real function. However, it could represent the end of the coding sequence that is marked if genes for the two proteins overlap a little at the ends, a situation that actually occurs in nature.

A reading frame in a specified part of a DNA sequence is said to be **"open"** if there is an initiation codon preceded by suitable regulatory signals (an **operator** region). This means that it *could* encode a protein. The reading frame of a sequence is open until the next termination codon. Recently another usage has appeared. Many writers refer to an **open reading frame** as a segment of DNA in which any one of the three reading frames is open. Another complexity in the reading of genetic messages arises because splicing of RNA may sometimes cause a shift in the reading frame. For example, a mRNA being transcribed from the sequence in reading frame 2 in the foregoing example may skip over a nucleotide part of the time to form an RNA in which the first part is encoded by reading frame 2 and the second by reading frame 3 (a + 1 frameshift). Alternatively, a nucleotide could be read twice with a – 1 frameshift with the sequence of reading frame 1 for the latter part of the mRNA. Frameshifts can also occur during protein synthesis as the mRNA is being read.

In the present example we have examined the sequence in mRNA. In the DNA there are two strands. One is the **coding strand** (also called the nontranscribing or nontranscribed strand), which has a sequence that corresponds to that in the mRNA and the one that is given in Fig. 5-4. The second antiparallel and complementary strand can be called the **template strand** or the noncoding, transcribing, or transcribed strand.[372] The mRNA that is formed is sometimes referred to as a **sense strand**. The complementary mRNA, which corresponds in sequence to the noncoding strand of DNA, is usually called **antisense RNA**.

**TABLE 5-6**
**The Sixty-Four Codons of the Genetic Code**

| 5′–OH Terminal base | Middle base | | | | 3′–OH Terminal base |
|---|---|---|---|---|---|
| | U(T) | C | A | G | |
| U(T) | Phe | Ser | Tyr | Cys | U(T) |
| | Phe | Ser | Tyr | Cys | C |
| | Leu | Ser | Term[c] | Term[d] | A |
| | Leu | Ser | Term | Trp | G |
| C | Leu | Pro | His | Arg | U(T) |
| | Leu | Pro | His | Arg | C |
| | Leu[a] | Pro | Gln | Arg | A |
| | Leu | Pro | Gln | Arg | G |
| A | Ile | Thr | Asn | Ser | U(T) |
| | Ile | Thr | Asn | Ser | C |
| | Ile | Thr | Lys | Arg | A |
| | Met[b] | Thr | Lys | Arg | G |
| G | Val | Ala | Asp | Gly | U(T) |
| | Val | Ala | Asp | Gly | C |
| | Val | Ala | Glu | Gly | A |
| | Val[b] | Ala | Glu | Gly | G |

[a] The codon CUA (CTA) encodes threonine and the codon AUA (ATA) methionine in mammalian mitochondria.
[b] Initiation codons. The methionine codon AUG is the most common starting point for translation of a genetic message but GUG can also serve. In such cases it codes for methionine rather than valine.
[c] The "termination codon" UAA (TAA) encodes glutamine in *Tetrahymena*.
[d] The termination codon UGA (TGA) encodes tryptophan in mitochondria and selenocysteine in some contexts in nuclear genes.

## 2. Variations

Is the genetic code "universal" or does it vary from one organism to another? Studies with bacteria, viruses, and higher organisms including humans have convinced us that the code is basically the same for all oganisms. However, there are some variations. For example, in mitochondria of both humans and yeast the codon TGA is not a termination codon but represents tryptophan. In mammalian mitochondria CTA represents threonine rather than leucine, and ATA encodes methionine instead of isoleucine. These differences in the code are related to the fact that mitochondria contain their own piece of DNA. It encodes not only several proteins but also tRNA molecules whose anticodon structures are altered to accommodate the changed meanings of the codons of the mitochondrial DNA and mRNA.[373,374]

Variations in the code for cytoplasmic proteins have been found. In *Tetrahymena* and other ciliates the codon TAA represents glutamine rather than being a termination codon.[375] A few proteins, including some in the human body, contain **selenocysteine**, the selenium-containing analog of cysteine. Selenocysteine is encoded by termination codon TGA. See Chapter 29 for details. However, even though TGA is occasionally used in this way, it serves as a termination codon for most proteins within the same cells.[376] Thus, the *context* in which the codon TGA occurs determines how it is read by the ribosomal machinery.

## 3. Palindromes and Other Hidden Messages

The sentence "Madam, I'm Adam" reads the same either forward or backward. Such sentences, known as **palindromes**,[377] are infrequent in the English language. However, most DNA contains many palindromes, sequences of base pairs that read the same in forward and reverse directions. Consider, for example, the gene that specifies the sequence of nucleotides for the tRNA molecules of Fig. 5-30. It is a double-stranded DNA segment in which one strand has a sequence identical to that of Fig. 5-30 except for the substitution of T for U and for ψ (pseudouridine) and for the lack of methylation and other base modifications. The second strand is the exact complement. Figure 5-34 shows the part of this gene (residues 49–76) that corresponds to the 3' end of the tRNA molecule. This DNA segment could exist in a second conformation having a loop on each side of the molecule (Fig. 5-34). The stems of the two loops in this **cruciform** conformation are identical and symmetrically disposed around the center of the molecule. If we overlook the seven nucleotides in the center of the loop, the message in the stems reads the same in both directions, as is indicated by the green arrows.

Palindromes are often imperfect as is the one shown in Fig. 5-34. Here the two stems in the cruciform structure are related by an exact twofold rotational symmetry but the loops at the ends of the stems are not. Unpaired bases may bulge at various points in double-stranded stems of longer palindromes. These imperfect palindromes in the DNA are responsible for much of the tertiary structure of the various kinds of RNA. The tertiary structure, in turn, often determines the interaction of the RNA with enzymes and other proteins.

Special properties may be observed for palindromes containing homopurine tracts in one strand, and therefore homopyrimidine tracts in the other. If two identical palindromes of this type occur close together it is possible that the pyrimidine-containing strand of one can join with a hairpin loop of the other to form a triplex base structure. A related triplex structure may be formed when inverted repeat sequences occur within a homopyrimidine tract in one chain of DNA (Fig. 5-34B).[378] These have been called **mirror-repeats** or **H-palindromes** to distinguish them from true palindromes.[237,379] Each base triplex structure contains one set of Watson–Crick hydrogen bonds and one set of Hoogsteen hydrogen bonds. The triplexes of H-DNA are all either TAT or CGC$^+$, where one C is protonated (see also Fig. 5-24).

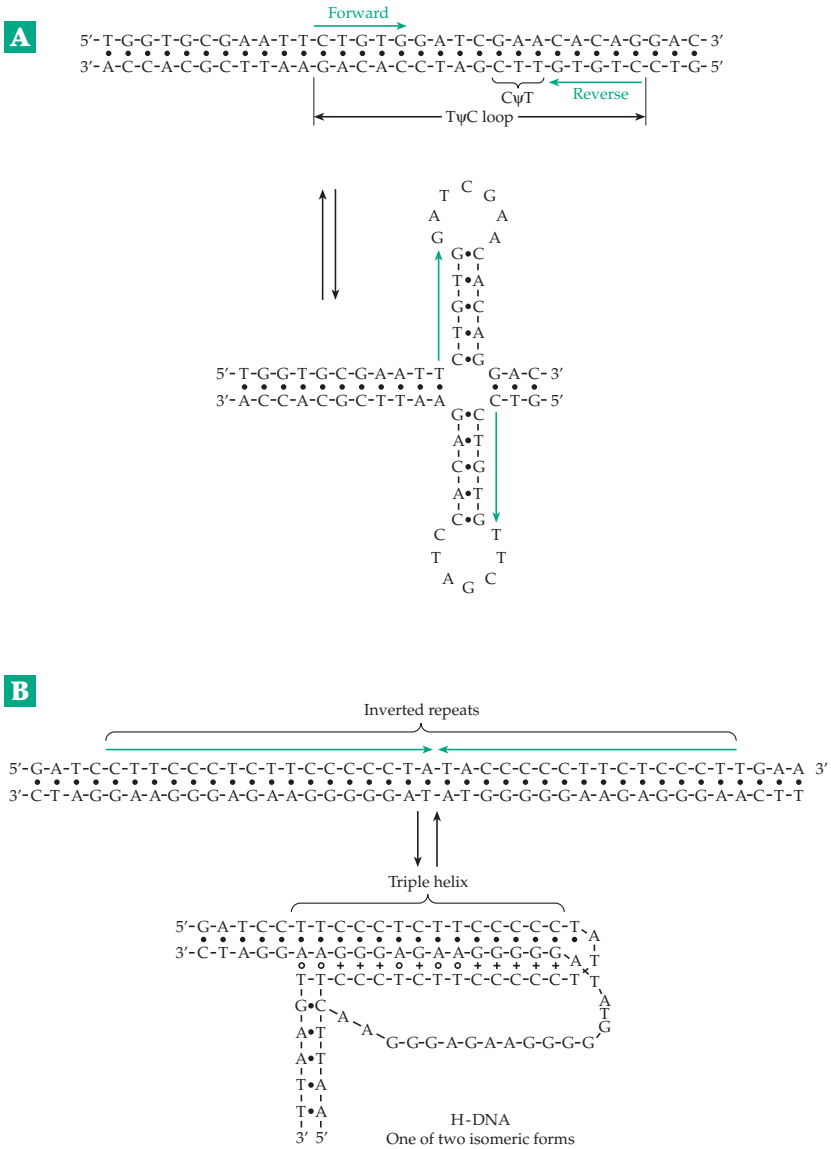



**Figure 5-34** (A) Two conformations of a segment of the yeast phenylalanine tRNA gene. The segment shown codes for the 3' end of the tRNA molecule shown in Fig. 5-30, including the TψC loop. (B) Formation of H-DNA (Fig. 5-24) proposed for a sequence in plasmid pGG32. The major element of the structure is the triplex, which is formed from the Watson–Crick duplex (•) associated with the homopyrimidine loop through Hoogsteen base pairing (∘, +). One of the two possible "isomeric" forms is shown. See Mirkin *et al.*[378]

Segments of H-DNA might block transcription, and proteins that bind to the triplex H-DNA may be involved in transcriptional regulation.[379]

Until rather recently there had been little to indicate that DNA actually assumes cruciform conformations in cells. However, strong experimental evidence suggests that some cruciform structures do form naturally.[380] Their formation from palindromic DNA [like the formation of Z-DNA from (G + C)-rich sequences] is a way of relieving torsional strain induced by supercoiling. Whether or not cruciform structures occur frequently within cells, there is no doubt that palindromic sequences are of great importance in the interaction of nucleic acids with symmetric dimeric and tetrameric protein molecules such as the gene repressor protein shown in Fig. 5-35.[381–383]

DNA contains numerous other protein binding sites which are not palindromes but whose sequences represent additional encoded information. The RNA transcripts likewise contain sequences that direct the catalytic machinery involved in splicing, that bind to ribosomal proteins, that control rates of transcription, and that cause termination of transcription.

### 4. The Base Composition of DNA and RNA

The nucleotide composition of DNA is surprisingly variable. The sum of the percentage cytosine plus the percentage guanine (C + G) for bacteria varies from 22 to 74%. That for *E. coli* is 51.7%. Among eukaryotic organisms, the range is somewhat narrower (28 to 58%; for humans, 39.7%). The fact that bacterial DNA molecules are more varied than those of higher organisms is not surprising. The prokaryotes have evolved for just as many more million of years as have we. Perhaps because of their simpler structure and rapid rate of division, nature has done more experimentation with genetic changes in bacteria than in people.

Comparisons of the C + G content of organisms have been used as a basis for establishing genetic relatedness.[384] However, since thymine is especially susceptible to photochemical alteration by ultraviolet light, bacteria with a high (C + G) content may have evolved in environments subject to strong sunlight or high temperatures, whereas those with a low (C + G) content have developed in more protected locations.[385,386]

## F. Interaction of Nucleic Acids with Proteins

Most functions of DNA and RNA are dependent on proteins. Cells contain a vast array of polypeptides that bind to nucleic acids in many specific ways.[387–390] We have seen (Fig. 5-21) that the histones hold supercoiled DNA in the form of nucleosomes[391] in eukaryotic nuclei and similar proteins bind to and protect the double-stranded DNA of bacteria.[205,392] Sperm cells "package" a large amount of DNA into a small space with the help of small arginine-rich proteins called **protamines.**[393] Cells always have some single-stranded DNA segments as well as single-stranded DNA binding (**SSB**) proteins. Among the latter are proteins from *E. coli*[394,395] and from viruses.[396–399] Specialized proteins bind to DNA sequences in telomeres[38,272] and centromeres.[400] A large number of proteins interact with RNA in ribosomes, spliceosomes, and other complexes.

A host of enzymes, which are described elsewhere in the book, act on DNA and RNA. They include hydrolytic nucleases, methyltransferases, polymerases, topoisomerases, and enzymes involved in repair of damaged DNA and in modifications of either DNA or RNA. While most of these enzymes are apparently proteins, a surprising number are **ribozymes**, which consist of RNA or are RNA–protein complexes in which the RNA has catalytic activity.

### 1. The Helix–Turn–Helix Motif

Much current interest in DNA–protein interactions is focused on regulatory processes. In prokaryotes the initiation of transcription of a large fraction of the genes is blocked by the binding of proteins known as **repressors**. While their structures are varied, one large group of repressors have DNA-binding domains with a similar helix–turn–helix architectural motif. They bind with high affinity to specific control regions which contain palindromic DNA sequences such as the following one, which defines a binding site for the *E. coli* **trp** (tryptophan) **repressor**:

```
5′-T-G-T-A-C-T-A-G-T-T-A-A-C-T-A-G-T-A-C-3′
   • • • • • • • • • • • • • • • • • • • •
3′-C-A-T-G-A-T-C-A-A-T-T-G-A-T-C-A-T-G-T-5′
```

These repressor proteins form dimers joined through a rigid central domain with a pair of arms, each containing two helices that form the helix–turn–helix motif. One helix of each pair fits into the major groove of the DNA helix and forms a "reading head" that carries a specific arrangement of amino acid side chains able to locate the symmetric nucleotide sequences flanking the twofold axis of the palindrome. In the case of the *trp* repressor, these are the ACTAGT hexanucleotides (marked by the arrows in the preceding structure). The interaction is depicted in Fig. 5-35.[401–405] The repressor has a high affinity for the DNA only if one molecule of L-tryptophan is bound to each subunit at a specific site near the DNA helix. Binding of the tryptophan causes a conformational change, which is pictured in Fig. 5-35. If tryptophan is absent the

repressor binds to the palindromic DNA only weakly and transcription of genes needed for tryptophan biosynthesis occurs freely. However, if tryptophan accumulates within the cell it binds to the repressor molecules causing them to bind firmly to DNA and prevent transcription. There are at least three of these palindromic sequences in *E. coli*, each one regulating a set of genes (**operons**) involved in tryptophan synthesis. In contrast to the effect of tryptophan, the tryptophan analog indole-3-propionate, which lacks the amino group of tryptophan, *derepresses* the same operons. It also binds to the *trp* repressor, but with the indole ring flipped over by 180° so that its carboxylate group contacts the phosphate groups of the DNA repelling them through both electrostatic and steric effects.[406,407] The picture in Fig. 5-35 is an oversimplification. In fact, at some binding sites (operator sites) more than one dimeric repressor binds in a tandem fashion.[405]

The helix–turn–helix motif is also found in many other proteins. One of these is the bacterial **lac** (lactose) **repressor** which controls the *lac* operon and for which

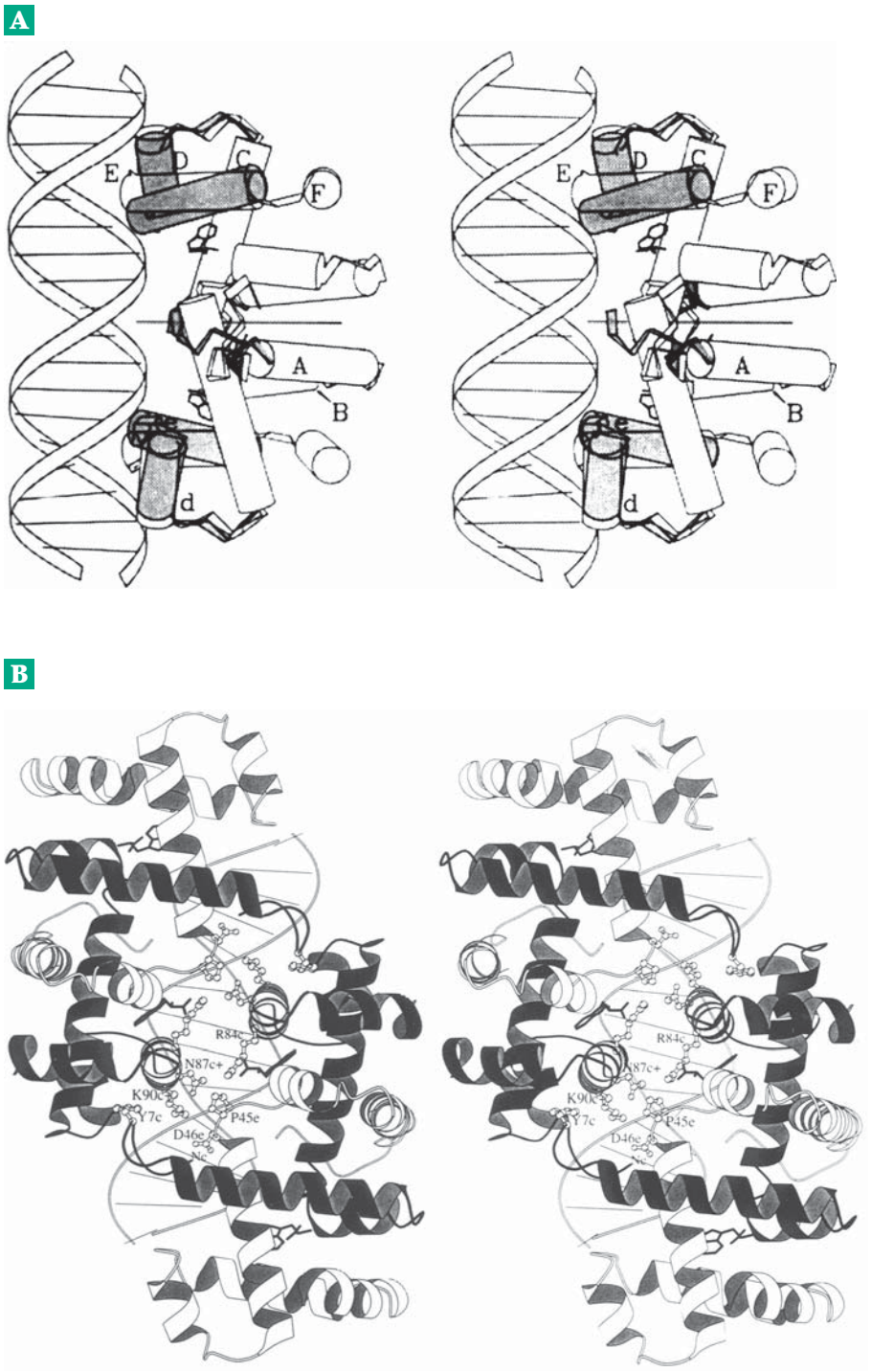



**Figure 5-35** Stereoscopic drawings illustrating the binding of a dimeric molecule of the Trp repressor protein to a palindromic sequence in DNA. (A) Schematic view showing structures of the aporepressor (partly shaded gray) and the holorepressor with bound tryptophan (unshaded) are superimposed. Cylinders represent the α helices in (B). From Zhang *et al.*[402] (B) MolScript ribbon diagram with a few side chains that interact with the DNA shown. Two tandemly bound dimeric repressor molecules are shown. Two bound molecules of tryptophan are visible in each dimer. The DNA is drawn as a double helix with lines representing the base pairs. From Lawson and Carey.[405]

the terms operator, promoter, repressor, and operon were first introduced (Chapter 28).[408–410] Some bacterial viruses, such as **phage lambda (λ)** of *E. coli*, encode repressors that allow the virus to reside in the bacteria without immediately destroying them. The λ repressor and other closely related proteins also utilize the helix–turn–helix motif[411–415] as do some proteins that *activate* transcription (see Chapter 28).

## 2. Other DNA-Binding Motifs

Nobody knows how many different DNA-binding structures may be discovered. However, most of those that are designed to recognize specific DNA sequences have some part that fits into the major groove of B-DNA.[416] Usually the DNA structure must be in a specific form: B-, A- or Z-, but it may sometimes be bent or distorted. The DNA recognition motifs in the proteins may consist of helices, β strands, or loops.

***Leucine zipper proteins.*** Several transcription factors have the leucine zipper structure, which has been described in Chapter 2 and was illustrated there by the structure of transcription factor Max (Fig. 2-21).[417] Related structures include those of the transcription activators c-Jun and c-Fos[418] and of the yeast transcription factor GCN4.[419,420] The latter is a dimeric protein in which the C-terminal halves of the two monomers form the helices of the leucine zipper. The helices, which are continuous for over 60 residues, fan out to interact with the DNA double helix as in Fig. 2-21. A 19-residue basic domain of each protein helix crosses the major groove of the DNA with side chain groups interacting with the DNA as is illustrated in Fig. 5-36.

***Zinc fingers.*** Another large group of transcription factors contain a bound zinc ion in a "finger" motif. The $Zn^{2+}$ is held by two cysteine $-S^-$ groups present on a loop of β structure and by two imidazole groups present on an α helix (Fig. 5-37).[421–424] Amino acid sequences found in many proteins that regulate transcription tend to form zinc fingers and to interact with DNA. These were recognized first in the transcription factor TFIIIA from *Xenopus laevis* in 1985.[421,425,426] This protein contains a zinc finger motif repeated nine times. X-ray structures, one of which is shown in Fig. 5-38, are known for proteins with three[423] and five[427] zinc fingers. (See Chapter 28.)

***Beta ribbons.*** An antiparallel double-stranded β ribbon can fit into the major groove of DNA and form

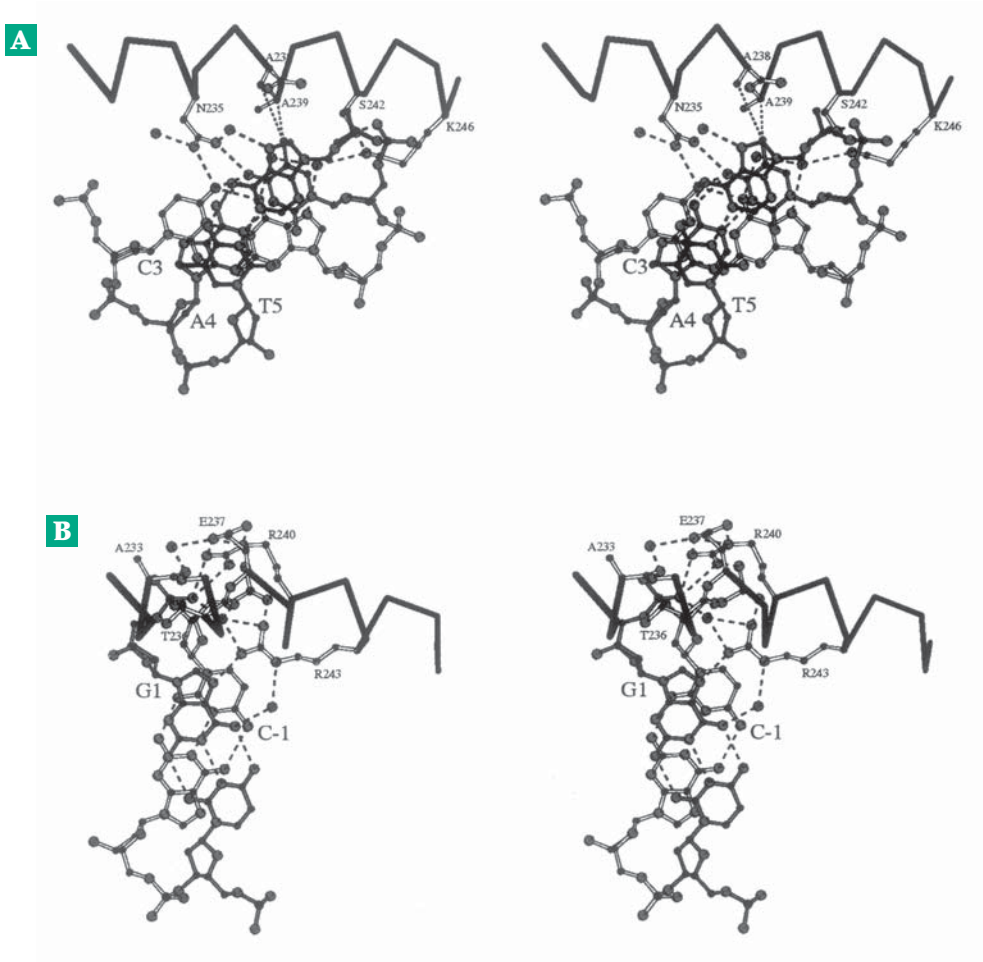



**Figure 5-36** Stereoscopic diagrams showing some of the interactions between an N-terminal helical domain of the yeast transcription factor GCN4-bZIP, a leucine zipper protein, and a specific palindromic DNA binding site:

```
     –5          –1 1          5
5′  –A T G A C G T C A T
```

(The bases are numbered outward from the central C and G.) The small solid spheres are water molecules. Notice the water mediated interactions of the basic arginine and lysine side chains with the nucleic acid bases and also the interaction of R240 and R243 (in B) with a backbone phosphate. The overall structure of the protein is similar to that of another leucine zipper shown in Fig. 2-21. From Keller *et al.*[419] Drawings courtesy of Timothy J. Richmond.

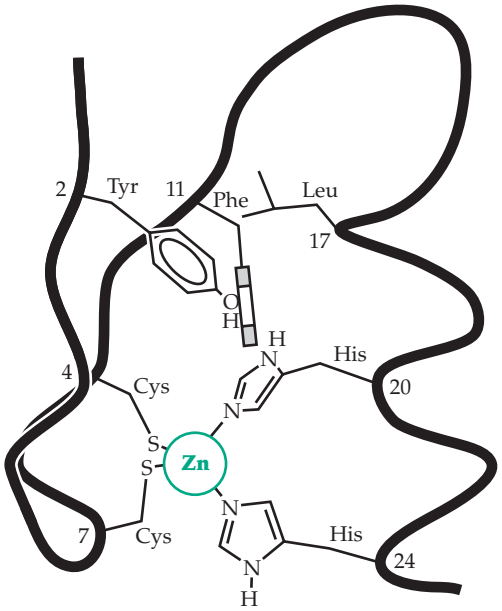


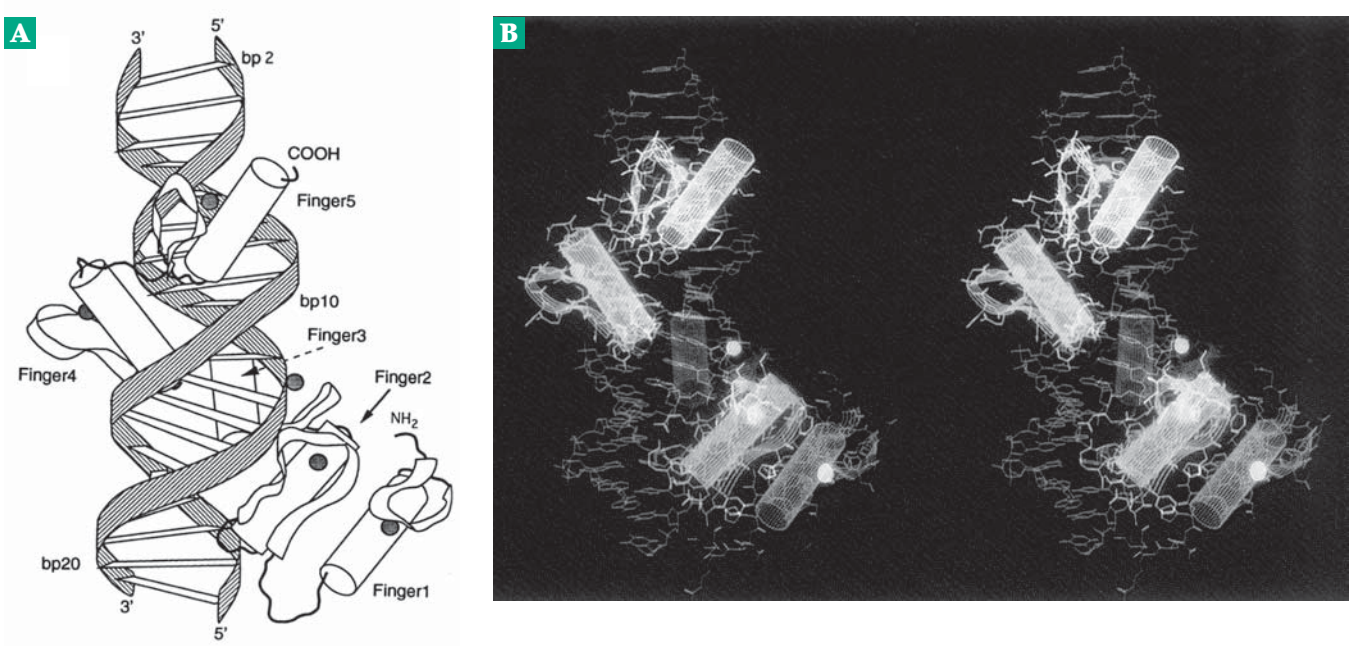

**Figure 5-37** Three-dimensional structure of a zinc finger. This is formed by the binding of $Zn^{2+}$ to the following sequence in a protein:

$$\phi \;\; X \;\; C \;\; X_{2-4} \;\; C \;\; X \;\; X \;\; X \;\; F \;\; X_5 \;\; L \;\; X \;\; X \;\; H \;\; X_{3,4} \;\; H$$

$$\xrightarrow{\quad} \text{turn} \xrightarrow{\quad} \text{turn} \xrightarrow{\quad}$$

$$\beta \qquad \beta \qquad \alpha \text{ helix}$$

Here $\phi$ is a hydrophobic amino acid and X may be any amino acid. After Krizek *et al.*[428]



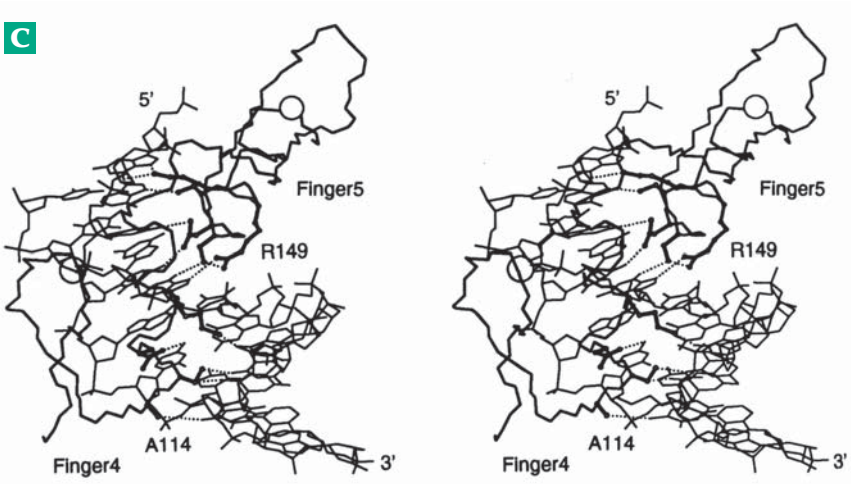




**Figure 5-38** (A) Sketch of a complex of a five-zinc finger protein bound to a high-affinity site in DNA. This protein is a human oncogene called GL1. Its gene is amplified in many tumors, but it is similar to many other transcription factors. The cylinders and ribbons highlight the $\alpha$ helices and $\beta$ sheets. Cobalt ions, which replaced the $Zn^{2+}$ in the crystals, are shown as spheres. (B) Stereoscopic view of the complex in a similar orientation. (C) Stereoscopic view emphasizing the contacts of side chains from fingers 5 and 6 with the DNA. The view is similar to that in (A) and (B) but the structure has been tilted back. Drawings courtesy of Nikola Pavletich.[427]

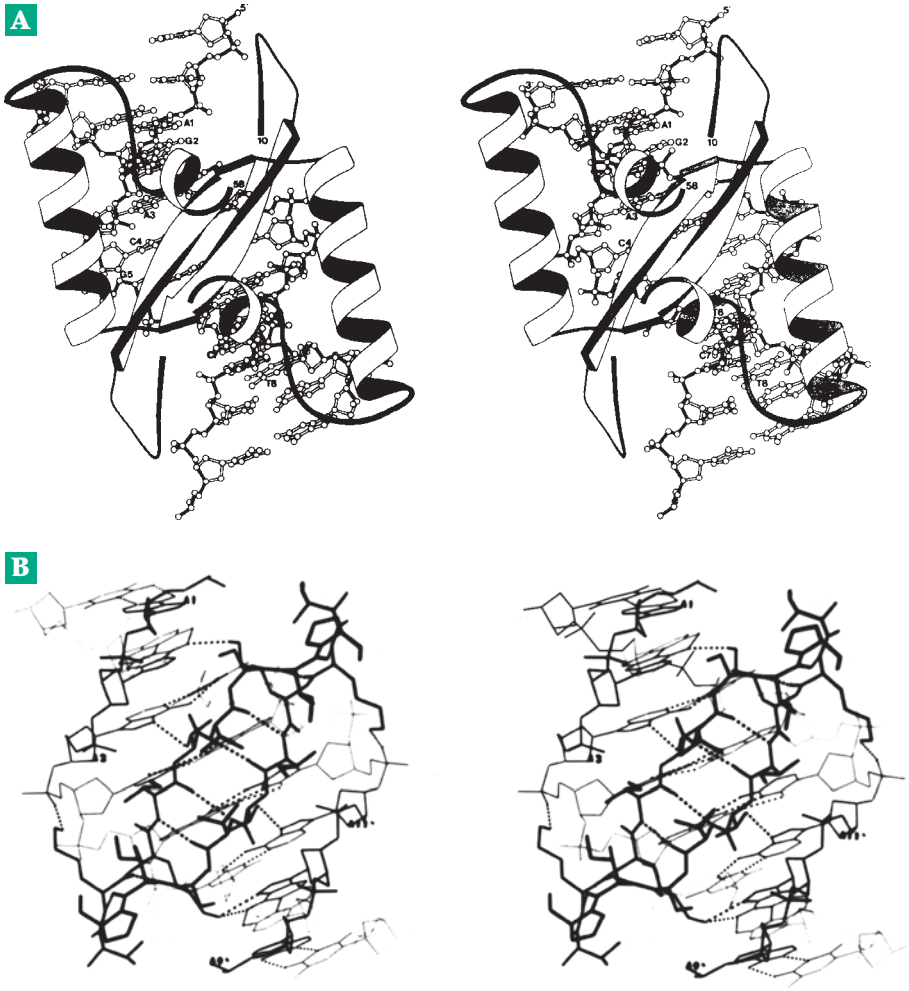



**Figure 5-39** (A) Stereoscopic ribbon representation of the *E. coli* methionine repressor–operator complex. Two subunits form a dimer with a double-stranded antiparallel β ribbon that fits into the major groove of the DNA in the B form. One strand is shaded more darkly than the other. (B) View of the β ribbon and its interactions with the DNA. Notice the direct hydrogen bonds from amino acid side chains of lysine and threonine residues to bases in the specific palindromic sequence AGACGTCT. From Somers and Phillips.[429] Drawings courtesy of S. E. V. Phillips.

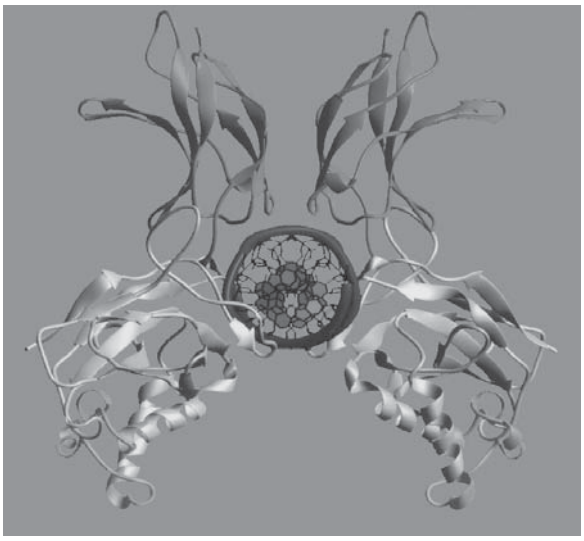


**Figure 5-40** Structure of a protein known as transcription factor NF-κB bound to its DNA target. Each subunit of the dimeric protein contains two β barrel domains. The loops at the ends of the barrels interact with the DNA in the center. From Müller *et al.*[433] Courtesy of Stephen C. Harrison.

hydrogen bonds and other interactions with the DNA. This is the basis for specific recognition of the operator sequence for the *E. coli* **met** (methionine) **repressor** as is illustrated in Fig. 5-39.[429] The same binding motif is employed by some other repressors[430,431] and also by the abundant bacterial DNA-binding protein **HU**.[432] It has also been utilized in designing the previously mentioned hairpin polyamide DNA-binding compounds.

***The winged helix family.*** A group of large protein transcription factors contain an N-terminal DNA-binding domain with the striking winged helix motif shown in Fig. 5-40.[433–436] It occurs in proteins from a wide range of organisms from yeast to human.

## 3. RNA-Binding Proteins

Among the many proteins that bind to RNA molecules[437–439] are the aminoacyl-tRNA synthetases, a variety of other well known enzymes,[440] the ribosomal proteins discussed in Chapter 29, and various proteins with dual functions of catalysis and regulation of

translation. A widely found RNA-binding **ribonucleoprotein domain** (RNP domain), occurs in hundreds of proteins including many of the RNA processing proteins considered in Chapter 28.[437,441,442] This 70–90 residue αβ module binds an RNA strand against a β sheet surface. Another ~ 70-residue protein motif binds ds RNA.[443]

## G. Viruses and Plasmids

Attacking every living thing from the smallest mycoplasma to human beings the nucleoprotein particles known as viruses have no metabolism of their own. However, they "come alive" when the nucleic acid that they contain enters a living cell. Viruses are significant to us not only because of the serious disease problems that result from their activities but also as tools in the study of molecular biology. A mature virus particle or **virion** consists of one or more nucleic acid molecules in a protein coat or **capsid**, usually of helical or icosahedral form. The capsid is made up of morphological subunits called **capsomers**. They can sometimes be seen clearly with the electron microscope. The capsomers in turn are usually composed of a number of smaller protein subunits. Some of the larger viruses are surrounded by membranous envelopes. Others, such as the T-even **bacteriophage** which attack *E. coli*, have extraordinarily complex structures (Box 7-C).

Most viruses contain a genome of either double-stranded DNA or single-stranded RNA, but some small viruses have single-stranded DNA and others have double-stranded RNA. The number of nucleotides in a virus genome may vary from a few thousand to several hundred thousand and the number of genes from 3 to over 200. Sometimes the nucleic acid molecules within the virion are circular, but in other cases they are linear. Table 5-7 lists a few of the known types of viruses as well as some individual viruses.[72,444] The size of the genome, in kilobases (kb), or kilobase pairs (kbp) is indicated. The number of genes in a virus is often somewhat more than one per kbp of DNA. A vast amount of information is available about these infectious particles. Only a few reference sources are cited here.[72,444–446] The architectures of some helical and icosahedral protein coats that surround genomic DNA or RNA are described further in Chapter 7.

### 1. Viruses with Single-Stranded DNA

The very small **helical bacteriophages** of the Ff family, such as fd, f1, and M13 (see Fig. 7-7) resemble thin bacterial pili but each virus particle contains a molecule of single-stranded circular ~6400 nucleotide DNA of $M_r$ ~2 x $10^6$ which encodes only ten proteins.[447–449] Bacteriophage ϕX174 (Fig. 5-41), an **icosahedral** DNA-containing virus 25 nm in diameter, is only three times as thick as the thinnest cell membrane. Its DNA contains just 5386 nucleotides.[450–452] The similar bacteriophage G4 contains 5577.[453,454]

It is remarkable that such tiny viruses are able to seize control of the metabolic machinery of the cell and turn it all in the direction of synthesis of more virus particles. There are only 11 known proteins encoded by the genes of ϕX or G4. Three genes encode the three kinds of protein subunits of the virus coat. Sixty copies of each are needed, as are 12 copies of a "pilot" protein.[451] Eight of the genes are spaced closely together, occupying most of the DNA. The other three genes are embedded within some of the first eight but in different reading frames. In one short region of the G4 chromosome the same nucleotides are part of three different genes, using all three possible reading frames. In addition to their own genes, these small viruses make use of many components of the cell that they infect. A large group of animal viruses, the **parvoviruses**, are similar in size and architecture to bacteriophage ϕX174.

Canine parvovirus, first identified in 1978, is now endemic.[455,456] Childhood **fifth disease** is also caused by a parvovirus.[456,457] When these single-stranded DNA viruses infect cells a double-stranded **replicative form** of DNA arises by synthesis of the complementary **negative strand** alongside the original **positive** DNA strand. Many copies of the replicative form are then synthesized. The negative strands of the replicative forms serve as templates for synthesis of numerous new positive strands that are incorporated into the progeny viruses. The whole process may take only 20 minutes. Some parvoviruses are unable to reproduce unless the cell is also infected by a larger adenovirus.

While most plant viruses contain dsRNA, the **geminiviruses**,[458] which cause a number of plant diseases, contain single-stranded DNA. The virus particle consists of a fused pair of incomplete icosahedra, evidently containing a single ~2500-bp DNA strand. Replication may require coinfection with two virus particles of differing sequence.

### 2. Viruses with Double-Stranded DNA

One of the smallest viruses containing double-stranded DNA is the 3180-nucleotide human **hepatitis B virus**. It infects millions of people throughout the world causing chronic hepatitis and often liver cancer.[459–461] The circular DNA is surrounded by an envelope consisting of proteins, carbohydrate, and a lipid bilayer. Many icosahedral viruses also contain dsDNA. Among them are the **papovaviruses**, some of which cause warts and others malignant tumors. Much studied by biochemists is the 5386-nucleotide **simian virus 40 (SV40)**, a monkey virus capable of inducing tumors in other species.[462,463] Closely related

## TABLE 5-7
## Characteristics of Some Individual Viruses and Groups of Viruses

| Type of genome[a] and group or individual virus name | Shape[b] | Diameter[c] (nm) | Masses in daltons x $10^{-6}$ | | Thousands of bases or base pairs (kilobases) |
|---|---|---|---|---|---|
| | | | Total | DNA or RNA | |
| **DNA, single-stranded** | | | | | |
| Bacteriophages *fd*, *f1*, *M13* | H | ~6 x ~880 (length) | 17.6 | 2.1 | 6.4 |
| Bacteriophages ϕ174, G4 | I | 25 | 6.2 | 1.8 | 5.4–5.6 |
| Parvoviruses | I | 18–25 | | 1.8 | 5.5 |
| Geminiviruses | I (fused pairs) | | | | |
| **DNA, double-stranded** | | | | | |
| Hepatitis B virus | I (enveloped) | | | 2.1 | 3.18 |
| Papoviruses | I | 45–55 | | | |
| SV40 (monkey) | I | | 17.6 | 3.5 | 5.22 |
| Polyoma (mouse) | I | 45 | 23.6 | 3.3 | 4.96 |
| BK virus (human) | I | | | 3.4 | 4.96 |
| Papilloma (human wart) | I | 56 | | 5.3 | 8.0 |
| Bacteriophage ϕ29 | T | | | 12 | 19.3 |
| Adenoviruses | I | 70 | | 20–30 | 30–45 |
| Bacteriophage Mu | | | | 25 | 38 |
| Bacteriophage T7 | T(short) | | | | 39.9 |
| Bacteriophage P22 | T | | | 28.5 | 43.2 |
| Bacteriophage λ | T | | | 32 | 48.6 |
| T-even bacteriophage | T | 100 x 80 (head) | 215 | 130 | 166 |
| Baculoviruses of insects | I | 70–130 | | | |
| Herpesviruses | I | 100 | | | |
| | core | 78 | ~1000 | 80–120 | 80–140 |
| | envelope | 150–200 | | | |
| Pox viruses, e.g., | | | | | |
| Smallpox, vaccinia (cowpox) | C | 160 x 250 | ~4000 | 150–240 | 240–300 |
| Cauliflower mosaic virus | I | | | | 8.0 |
| **RNA, single-stranded** | | | | | |
| Unsheathed | | | | | |
| Potato spindle tuber viroid | | | | 0.116 | 0.30 |
| Hepatitis delta virus | | | | | 1.7 |
| Sheathed, *plus-strand* | | | | | |
| Tobacco necrosis satellite | I | 18 | 1.7 | 0.4 | 1.20 |
| Small bacteriophages | | | | | |
| R17, MS2, Qβ | I | 23–26 | 3.6–4.0 | 1.2–1.5 | 3.5–4.5 |
| Picornaviruses | I | | 8.4 | 2.6 | 7.9 |
| Polioviruses | I | 27 | 6.4 | 2 | 6.1 |
| Rhinoviruses | I | 27–30 | 7–8 | 2.2–2.8 | 6.7–8.5 |
| Turnip yellow mosaic virus | I | 28 | 5.0–6.0 | 2.0 | 6.1 |
| Tobacco mosaic virus | H | 18 x 300 | 40 | 2.2 | 6.7 |
| Togaviruses | I | 20–40 | | 4 | 11 |
| *Negative-strand* viruses | | | | | |
| Influenza virus | I | 80–100 | 200 | 2.0 | 6.1 |
| Bullet-shaped viruses | | | | | |
| Rhabdoviruses | C | 20 x 130 | | | |
| Retroviruses | I | 80–100 | | 7–10 | 20–30 |
| **RNA, double-stranded** | | | | | |
| Reoviruses | I | 55–60 | | 11–12 | 16–18 |
| Mobillivirus | I (enveloped) | 38 | | | |
| Rotavirus | Wheel-shaped | 70 | | | |
| *Leishmania* RNA virus | | | | | 5.28 |

[a] Complete nucleotide sequences are known for most of these viruses.
[b] Shapes are indicated as I, icosahedral; H, helical; T, a tailed phage; C, complex.
[c] The second dimension given for some helical and complex viruses is the length (nm).

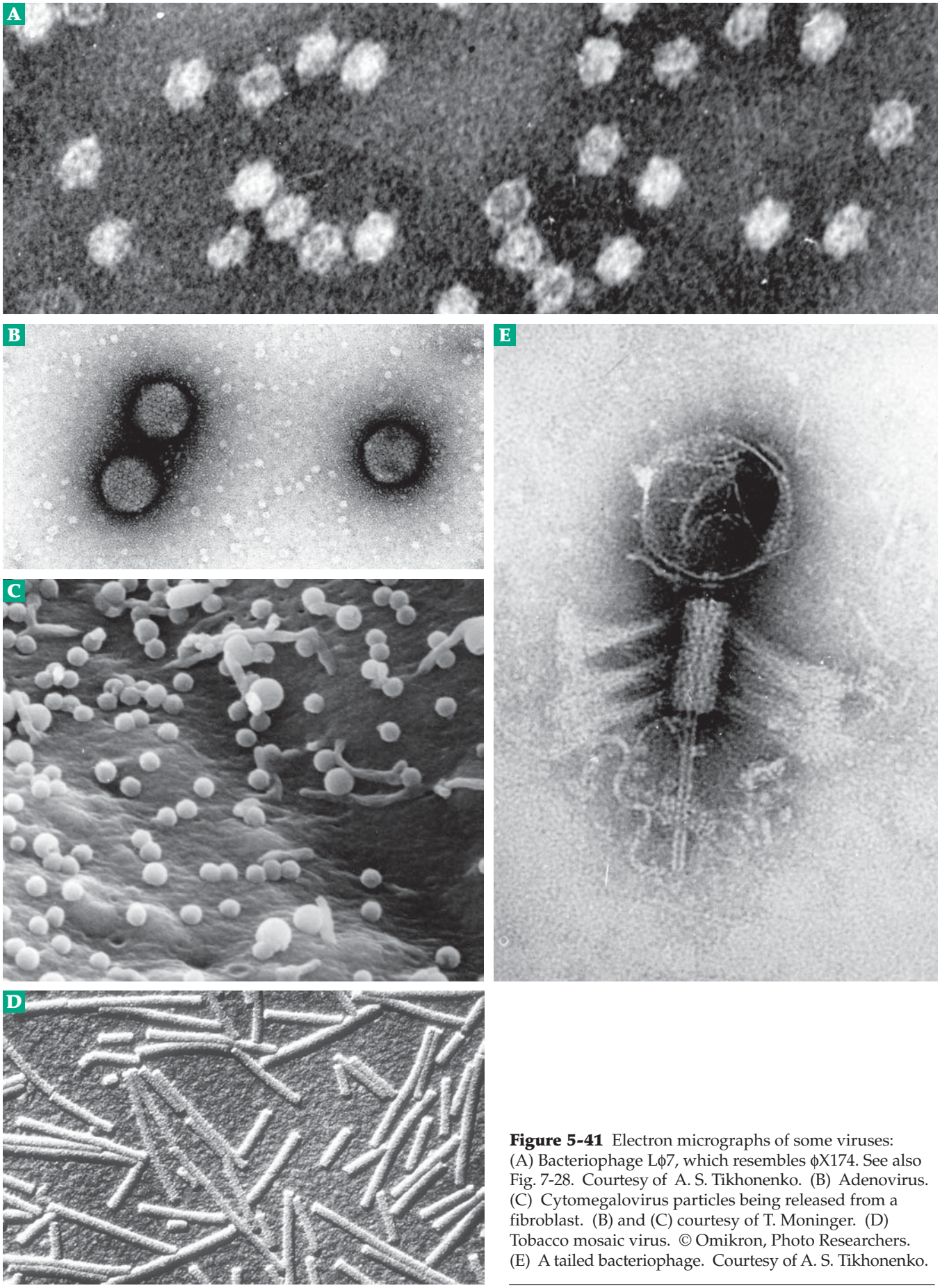



**Figure 5-41** Electron micrographs of some viruses: (A) Bacteriophage Lϕ7, which resembles ϕX174. See also Fig. 7-28. Courtesy of A. S. Tikhonenko. (B) Adenovirus. (C) Cytomegalovirus particles being released from a fibroblast. (B) and (C) courtesy of T. Moninger. (D) Tobacco mosaic virus. © Omikron, Photo Researchers. (E) A tailed bacteriophage. Courtesy of A. S. Tikhonenko.

is the **polyomavirus** of the mouse[464] and the **BK virus** of humans,[465] also suspected of causing cancer. The **papillomaviruses** cause warts and perhaps cancer[466] and the larger (70 nm diameter) **adenoviruses** (Fig. 5-41),[467,468] cause respiratory infections. Some 32 types infect humans. Many of the details of eukaryotic transcription were first studied using the adenovirus.[469] The very large **herpesviruses** are enveloped by a lipid-containing membrane.[470–472] Among them are herpes simplex viruses, which infect human mucous membranes and varicella-zoster virus, which causes chicken pox and shingles. Other herpesviruses include cytomegaloviruses (Fig. 5-41), another common human pathogen, and the Epstein–Barr virus, which causes mononucleosis and is suspected of causing cancer.[473] Another herpesvirus has been associated with multiple sclerosis.[474] The very large icosahedral **baculoviruses** cause polyhedroses in insects. One that infects the fly *Tipula* measures 130 nm in diameter. **Poxviruses** are also large and complex.[475] The tailed bacteriophages (Box 7-D) range in size from the small ~29-kDa phage P22 to the very complex T-even phage. Some plant viruses also contain dsDNA. One of these, the **cauliflower mosaic virus**,[476,477] is transmitted by aphids. It has proved useful as a gene-transfer vehicle for genetic engineering.

## 3. Viruses Containing RNA

Several plant diseases including the **potato spindle tuber disease** are caused by **viroids**, molecules of single-stranded RNA only 240–380 nucleotides in length with a folded structure.[478–481] Such an RNA could code for a protein containing only about 100 amino acids. However, the known 359-nucleotide sequence of the potato spindle tuber virus contains no AUG initiation codon. It seems impossible that the virus carries any gene for a protein. Conserved features of viroid sequences suggest a close relationship to the intervening DNA sequences known as type I introns (see Fig 28-19). Whatever its genetic message, a plant viroid causes the plant cell to replicate many copies of the viroid molecule, which may then be transmitted to other plants by aphids, or on the surface of tools, by humans. A larger 1678-nucleotide viroid-like RNA (**hepatitis delta virus)** has been identified in human patients with severe chronic hepatitis and who were also infected with hepatitis B virus. The ssRNA in this virus does encode at least one protein.[482–484]

One of the smallest of the encapsulated RNA-containing viruses is the **satellite tobacco necrosis** virus. It replicates only when the plant is also infected with the larger tobacco necrosis virus. The satellite virus, whose three-dimensional structure is known from X-ray diffraction studies,[485] contains a 1200-nucleotide strand of RNA which encodes a 195-residue protein. Sixty copies of the latter are assembled around the RNA in an icosahedral array (Fig. 7-14) to form the virion. The structure of the similar satellite tobacco mosaic virus has also been described in detail.[486,487]

RNAs of the small bacteriophages f2, R17, MS2 and Qβ contain 3500 – 4500 nucleotides and are enclosed in icosahedral shells made up of 90 identical subunits.[488–490] Initially only three genes were evident but a fourth small gene in a different reading frame has since been found in at least one of them.[491] The RNA molecules in these and other **positive-strand** viruses serve as mRNA within the host cells. Another group of small RNA viruses are the **picornaviruses** (picoRNA, meaning very little RNA). Many of these icosahedral viruses of 15–30 nm diameter attack humans. Among them are the **enteroviruses** including the **polioviruses**,[492,493] the **hepatitis A virus**,[494] the **coxsackieviruses**, and some of the **echoviruses**. A second class of picornaviruses include **rhinoviruses** which cause the common cold. More than 100 types are known.[495–497] The **foot–and–mouth disease** virus,[498–500] which attacks the cloven-footed animals, and the **Mengo virus**,[501] which can cause a fatal encephalitis in mice, are also picornaviruses. The three-dimensional structures of polio virus, human rhinoviruses, Mengo virus, and many other eicosahedral viruses are known. Their architecture is discussed in Chapter 7.

The **togaviruses**, which are a little larger than the picornaviruses, have an icosahedral core surrounded by a lipid membrane. Yellow fever and rubella (German measles) are both caused by togaviruses. Other togaviruses, such as **Sindbis virus**[502] and **Semliki Forest virus**,[503] have become important in biological research.

A large number of icosahedral RNA viruses of diameter 28–30 nm (Fig. 7-14) attack plants, causing diseases such as tomato bushy stunt,[504] southern bean mosaic,[505] or turnip yellow mosaic. Best known of the helical RNA viruses is the **tobacco mosaic virus** (Figs. 5-41, 7-8).[506-507a] Its genome contains 6395 nucleotides as linear ssRNA. Many strains are known. Related viruses cause cucumber green mottle[508] and other plant diseases.

Large viruses of 80 – 100 nm diameter bearing 8–10 spikes at the vertices of the icosahedra cause influenza,[509,510] mumps, measles, and related diseases. The internal structure must be complex. Only 1% of the virus is RNA, and that consists of several relatively small pieces. These are **negative strand viruses** whose RNA is of the opposite polarity to the mRNA. The latter must be formed by transcription from the negative strand. The viruses carry their own RNA polymerase for this purpose. Of even more complex structure are the bullet-shaped **rhabdoviruses** which cause rabies and vesicular stomatitis.[511] The diameter of these viruses is 65–90 nm and the length 120–500

nm. The internal structure includes a helical arrangement of nucleoprotein. They also are negative strand viruses.

Among the **retroviruses**[512–514] are types B and C **oncoviruses** which induce malignant tumors in mammals and birds[515] and the **human immunodeficiency virus** (HIV), the apparent causative agent of AIDS. Their RNA functions in a surprising way. Each virion contains a **reverse transcriptase**, an enzyme that transcribes copies of circular dsDNA copies from the one or two mRNA-like molecules that make up the virus genome. Following action of the reverse transcriptase, one of the transcribed DNA circles becomes covalently spliced into the host's own cellular DNA. There it remains permanently as a **provirus**. RNA molecules transcribed from the provirus serve as mRNA for virus-encoded proteins and also as the genomes for new virus particles.

Double-stranded RNA is unusual in nature but constitutes the genome of the **reoviruses**.[516] The RNA of these viruses fragments into segments upon infection. One member of the group is thought to be the cause of acute diarrhea of infants.[517]

## 4. Viruses without Nucleic Acid?

The cause of the slow, fatal neurological disease of sheep known as **scrapie** has been a mystery for many years. Similar human diseases include **kuru** and **Creutzfeldt-Jakob** disease.[518–520] Scrapie can be transmitted by injection and this has permitted isolation of the apparent infective agent, a 27- to 30-kDa hydrophobic protein particle[521] which is devoid of DNA or RNA. Prusiner[521] suggested the name **prion** (proteinaceous infectious particle) for the scrapie agent. However, mRNA for the prion is present in normal as well as infected brains, and protein produced in mouse cells from cloned prion genes did not cause scrapie infections. Therefore, there was doubt about the causative agent for the disease. The prion concept is now generally accepted and is considered further in Chapter 29. There are still some who are looking for a nucleic acid component.[519,522–524]

## 5. Life Cycles

Viruses have many modes of life. They enter cells in various ways. Some enter through coated pits from which they are taken into lysosomes via endocytosis. Others are literally injected into the cells (See Box 7-C). Within cells some viruses are assembled in the nucleus, some in the cytoplasm, and some in membranes. The typical life cycle of a virus leads to rapid formation of large numbers of progeny. Within 20 minutes after entrance into a bacterial cell, a bacteriophage can induce the formation of 100–200 new bacteriophage particles. One of the bacteriophage genes encodes a protein that is also synthesized by the host and which induces lysis of the cell membrane and destruction of the cell. Many animal viruses destroy cells in a similar fashion.

**Temperate bacteriophage**, the best known being phage λ, have a very different life cycle. Their DNA usually becomes integrated at a specific point into the genome of the bacterium (Chapter 27). Only rarely is an infected cell lysed. The retroviruses that attack mammals and birds have a similar characteristic. Their DNA is also integrated into the host genome. Some viruses that usually produce lysis of cells, e.g., SV40, adenoviruses, herpes viruses, and hepatitis B virus, can occasionally be integrated into the DNA of the host. If such integration occurs in the middle of a gene, that gene will be mutated. This is one way in which such viruses may induce cancers.

One of the most important results of integration of viral DNA into the host genome is that the integrated genes are replicated as part of the genome and are transmitted from one generation to the next. Among these are the cancer-causing **viral oncogenes** (*v-onc*), which are discussed in Box 11-D and in Chapter 11, Section H. While viruses are important causes of cancer in some animals, relatively few human cancers are thought to result directly from the action of viruses. However, the Epstein–Barr virus, which causes mononucleosis, can sometimes be integrated into epithelial cells of nasal regions and can evidently cause cancer. The same virus appears to be responsible for Burkitt's lymphoma, a common cancer in certain areas in Africa.[525]

## 6. Plasmids and Transposable Genetic Elements

In addition to their chromosomal DNA, bacteria often carry extra small pieces of DNA as permanent parts of their genome. These **plasmids** (sometimes called **episomes**), which are about the size of the DNA of viruses, replicate independently of the host chromosomes. Each bacterial cell usually contains more than a single copy of the plasmid. For example, the "colicinogenic" plasmid **ColE1**, that infects *E. coli* is a circular piece of DNA of molecular mass $4.2 \times 10^6$ Da. Over 20 copies are normally found per cell but in the presence of a suitable concentration of the drug chloramphenicol the number may rise to 1000–2000.

Plasmids carry a variety of genes which are often useful to bacteria. Some proteins encoded by plasmid genes confer drug resistance to a bacterium. Some are antibiotics. For example, a protein encoded by a gene in plasmid ColE1 is toxic to other strains of *E. coli.* Some plasmids carry genes for enzymes needed for the oxidation of hydrocarbons. Some plasmids contain

genes for the **restriction endonucleases** which have become essential to present-day molecular biology and genetic engineering (Section H, 2).

As with some viruses, the DNA of many plasmids can become integrated into the genome of the host. An example is provided by the large 62-kDa plasmids known as **sex factors**. They contain genes encoding the protein subunits of the sex pili (Chapter 7) and can become integrated into the bacterial chromosome. Bacteria containing integrated sex factors are "male" and are able to transfer genes not only of the sex factor but also of virtually the entire bacterial genome into other susceptible bacterial cells. This provides bacteria with the means for sexual reproduction. The transfer of DNA between the bacteria may occur via the sex pili (see Chapter 26). In this respect the sex factors are similar to viruses such as M13 that also appear to gain entrance to bacteria via sex pili.[72]

Integrated viruses are also related to **transposable genetic elements** (transposons). These are segments of DNA that allow genes to move from place to place within the chromosomes (Chapter 27).

## H. Methods of Study

Many of the methods discussed in Chapter 3 are directly applicable to nucleic acids. A few additional methods will be considered in this section.[525a]

### 1. Isolation and Separation of Nucleic Acids

RNA is often extracted from lysed cells or tissues, separated ribosomes, mitochondria, plastids, or nuclei by warming with aqueous phenol and a detergent such as sodium dodecyl sulfate (SDS). Proteins are denatured by this treatment and are dissolved by the phenol, while RNA remains in the lighter aqueous layer. Depending on the conditions DNA may either remain in the aqueous layer or be removed.[526–528] Various precipitation and extraction procedures may be used to separate the RNA in the aqueous layer from polysaccharides, from DNA (if present), and from their components.[526,528,529] DNA may be extracted from cells or nuclei as a nucleic acid–protein complex using 1 M NaCl. The protein can then be denatured with an organic solvent, by detergents, or by phenol. It is desirable to digest proteins away with a nonspecific protein-hydrolyzing enzyme such as proteinase K.[528] After removal of proteins DNA is often precipitated with cold ethanol.

During isolation of RNA, bentonite (a type of clay) or other inhibitors of ribonuclease are often added. For the same reason, chelating agents that complex metal ions needed for the action of deoxyribonucleases are used to protect DNA. Care is necessary to avoid shearing of the very long, narrow strands of DNA. Even rapid pipetting of solutions will cause such breakage.

Extracted RNA molecules may be separated from each other by centrifugation in a sucrose gradient (Chapter 3).[528] Fragments of DNA are purified in the same way or by equilibrium centrifugation in **CsCl gradients**.[37] Concentration gradients in the dense salt solution are stable, and the sharpness of banding of particles is ensured by use of a high centrifugal field. Single-stranded DNA may be separated from double-stranded DNA, and DNAs of differing G + C content can be separated. The latter separation is based on differences in buoyant densities ρ in CsCl which are approximately shown in Eq. 5-9.

$$\rho = 1.660 + 0.098\ (\text{mole fraction C + G}) \qquad (5\text{-}9)$$

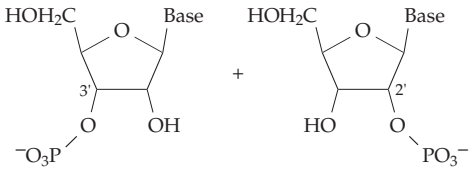
One of the most important methods for separating either RNA or DNA mixtures is zone electrophoresis through polyacrylamide or agarose gels. The separated bands may be visualized by scanning in ultraviolet light or by fluorescence of intercalated dyes such as ethidium bromide (Figs. 5-20, 5-22). This method is being displaced to some extent by HPLC using DEAE type ion exchange columns[530,531] for small lengths of DNA including plasmids. The procedure called **pulsed field electrophoresis** makes it possible to isolate very large pieces of DNA, up to several million base pairs in length.[532–534] The separation is carried out in agarose gels, through which the long DNA rods must move in a snakelike fashion.[534] The current is delivered in a pulse and then, after a period of a second to several minutes, a second pulse in a different direction, usually at 90° to the first. The procedure is repeated many times. The size of the DNA seems to affect the time required to reorient the molecules and to start moving in the second direction. Intact DNA from small chromosomes can be separated (Fig. 5-42). To prevent breakage of the DNA by shearing, intact cells are suspended in liquid agarose and allowed to gel into a block about 2 x 5 x 10 mm in size. The block is treated with enzymes and detergents to lyse the cells and to remove all protein and RNA.[534] The block, containing the residual DNA molecules, is then embedded in the electrophoresis gel. Other methods of DNA separation include chromatography on hydroxyl-apatite and gel filtration.

### 2. Hydrolysis and Analysis

Both DNA and RNA are easily broken down by acid-catalyzed hydrolysis. Thus, heating at 100°C for one hour in 12 M $HClO_4$ is sufficient to hydrolyze nucleic acids to their constituent bases. However, for analysis of RNA it is better to heat in 1 N HCl for 1 h at

100°C. The products are adenine, guanine, cytidine-5'-phosphate, and uridine-5'-phosphate.[535] As is suggested by this distribution of products, the glycosylamine linkages to purines are more labile than those to pyrimidines. The linkages are also less stable in DNA than in RNA. A procedure based on these differences and useful in sequencing by the Maxam-Gilbert method, is to leave DNA overnight in the cold at pH 2 to cleave off all of the purine bases. The resulting polymer is known as an **apurinic** acid.

In alkaline solutions RNA is hydrolyzed to a mixture of 2'- and 3'-nucleotides.




The mechanism involves participation of the free 2'-OH of the ribose groups and formation of cyclic 2', 3'-phosphates and is similar to that of pancreatic ribonuclease (Chapter 12). Because deoxyribose lacks the free 2'-OH, the phosphodiester linkages in DNA are quite stable in base.

Hydrolytic cleavages of nucleic acids by the enzymes known as **nucleases** are of great practical value. Pancreatic ribonuclease, an **endonuclease**, cuts a chain adjacent to a pyrimidine in nearly random fashion, leaving phospho groups attached to the 3' position in the nucleotide products (Fig. 5-43). **Exonucleases** cleave from the ends of chains. For example, the phosphodiesterase of snake venom cleaves from the 3' end, which must have a free 3'-OH group, to give 5'-nucleotides. On the other hand, the phosphodiesterase from spleen has the opposite polarity, cleaving chains from the 5' end to give 3'-nucleotides. Similar variations in specificity are found among enzymes that cleave DNA. For example, pancreatic DNAase I, which cleaves preferentially between adjacent purines and pyrimidines, yields 5' mononucleotides whereas DNAase II gives 3'-mononucleotides. Various hydrolytic cleavage reactions of polynucleotides are summarized in Fig. 5-43.

The most striking specificity in DNA hydrolysis is displayed by the **restriction endonucleases** which are discussed further in Chapter 26. These fussy catalysts cleave only at points within or close to a defined sequence of several nucleotides in double-stranded DNA. For example, the enzyme EcoR I cuts only at the following palindromic sequence:

```
        ↓
5′ – G A A T T C – 3′
     • • • • • •
3′ – C T T A A G – 5′
            ↑
```

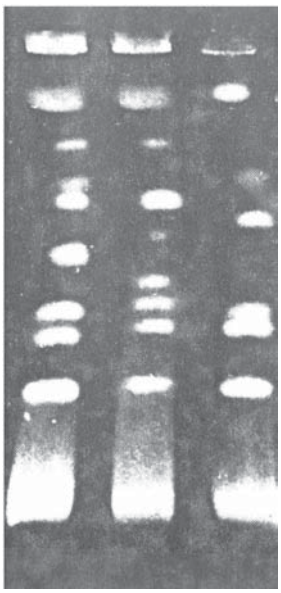
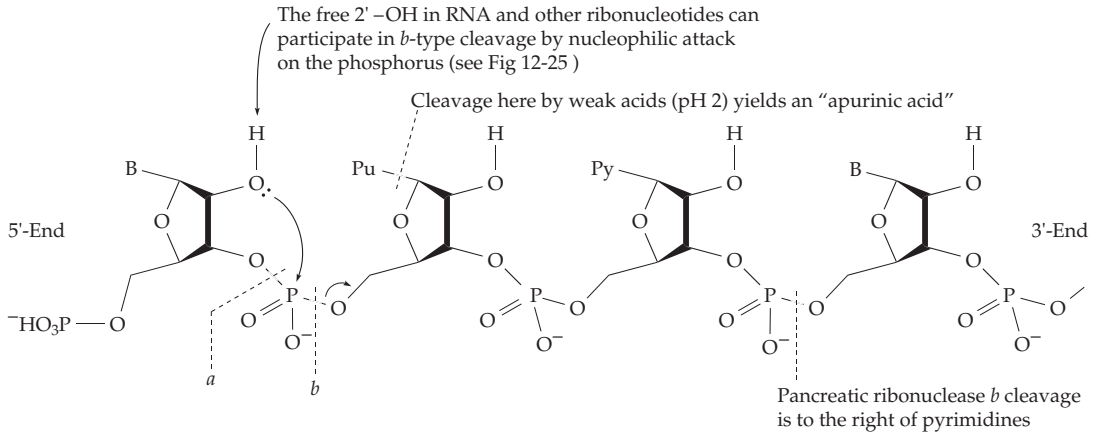


**Figure 5-42** Intact DNA from the chromosomes of three strains of the malaria parasite *Plasmodium falciparum*, ranging from 750 Kb to 5 Mb, separated by pulsed-field gel electrophoresis. Courtesy of C. Smith and T. E. Wellems. Reproduced by permission of Amersham Pharmacia Biotech Inc.

The cuts in the two strands are made at the points indicated by the arrows. This one endonuclease will cut almost any DNA into long pieces averaging about 5000 base pairs each. These pieces can in turn be cleaved by other restriction endonucleases to form smaller fragments. Since there are about 2400 of these enzymes known, with 188 different specificities,[536] it is possible to cut any piece of DNA down to a size of 100–500 base pairs, ideal for sequencing.[537–539] Each fragment has known sequences at the two ends. Some restriction enzymes cleave outside their specific recognition sequence (see Table 26-2). Some recognize 16-nucleotide palindromes and cut at rare sites.

It is sometimes desirable to cut a large DNA molecule at only a few points. One approach is to protect most sites of a restriction enzyme's action by methylating them (see Chapter 26) while protecting the desired cleavage site, for example by a repressor protein[540] or by a PNA molecule (p. 227) of specific sequence designed to "clamp" the site chosen for protection.[541] Ribozymes (Chapter 12) have been engineered to be as specific or more specific than endonucleases.[542,543] Other new approaches are being developed.[544]

The base composition of either RNA or DNA can be determined after hydrolysis catalyzed by 98% formic acid at 175°C for 30 min or by 12 M perchloric acid at 100°C for 1 h.[545] The bases can then be separated by ion exchange chromatography on a sulfonated polystyrene resin. RNA can be hydrolyzed to a mixture of nucleoside 2′- and 3′-phosphates by 0.3 M NaOH at




A. Cleavage at point *a* is catalyzed
   1. Throughout the molecule by endonucleases
      Pancreatic deoxyribonuclease I
   2. Only at the 3' end by exonucleases
      Venom diesterase, nonspecific, attacks DNA and RNA. A free 3'-OH is essential

B. Cleavage at point *b* is catalyzed
   1. Randomly throughout the molecule by endonucleases and by bases (nonenzymatically)
      Pancreatic ribonuclease cleaves only to the right of a pyrimidine-containing nucleotide
      Ribonuclease T1 of *Aspergillus oryzae* cleaves to the right of a guanine-containing residue (3'-guanylate) in ssRNA
      Ribonuclease T2 of *Aspergillus oryzae* cleaves to the right of an adenine-containing residue (3'-adenylate) in ssRNA
      Pancreatic deoxyribonuclease (DNase) II
      Micrococcal DNase
   2. Only at the 5' end by exonucleases
      Bovine spleen phosphodiesterase hydrolyzes both polyribo- and polydeoxyribonucleotides

**Figure 5-43** Some hydrolytic cleavage reactions of polynucleotides. Reactions of both RNA and DNA are included.

37°C for 16 h and DNA can be hydrolyzed to nucleotides enzymatically. The negatively charged nucleotides can then be separated by ion exchange chromatography on a quaternary base-type resin (Chapter 3). Periodate cleavage (Eq. 4-12) and reduction of the resulting dialdehydes by [$^3$H]$NaBH_4$ to trialcohols allows introduction of a radioactive label (Fig. 5-44). Alternatively, the dialdehydes can be reductively alkylated by an amine plus $NaCNBH_3$.[546]

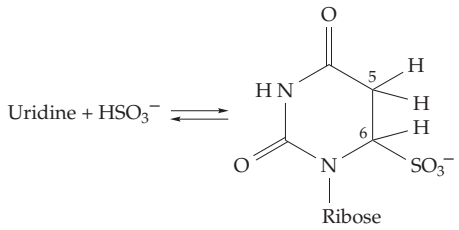
The total content of RNA + DNA in tissues may be estimated from the phosphorus content or by color reactions of the sugars.[37,545] These reactions depend upon dehydration to furfural or deoxyfurfural by concentrated sulfuric acid or HCl (Eq. 4-4). Furfural formed from RNA reacts with orcinol (3,5-dihydroxytoluene) and ferric chloride to produce a green color useful in colorimetric estimation of RNA. A similar reaction of DNA with diphenylamine yields a blue color.

Quantitative determination of over 90 free nucleotide compounds found within cells can be accomplished by thin layer chromatographic procedures on as few as $10^6$ bacterial cells (~2 μg) labeled by growth in a $^{32}P_i$-containing medium.[547]

## 3. Characteristic Reactions of the Bases and Backbone

***Reactions of nucleophiles.*** A number of nucleophilic reagents add reversibly at the 6 position of pyrimidines. Thus, bisulfite adds to uridine (Eq. 5-10).[528] Hydroxylamine ($HONH_2$) adds in a similar fashion to give a compound with –HNOH in the 6 positions.[528] Sodium borohydride ($NaBH_4$), which can be viewed as a donor of a hydride ion ($H^-$), reduces uridine to the 5,6-dihydro derivative. This presumably occurs by attack of the hydride ion at position 6 in a manner analogous to the reaction of bisulfite in Eq. 5-10.

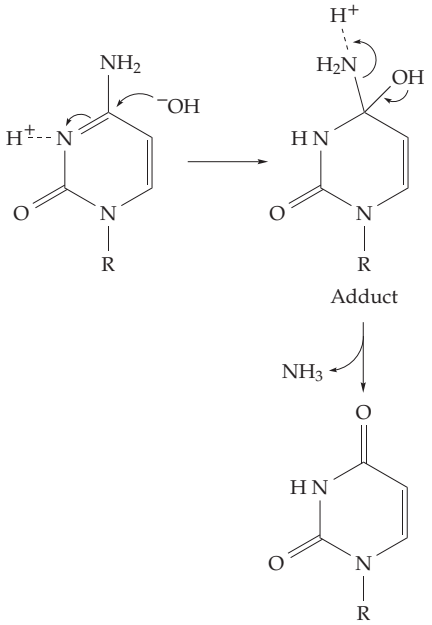
Cytidine reacts in the same way, but the bisulfite addition compound is unstable. These C5–C6 adducts of cytidine all have a greatly enhanced reactivity at C4,



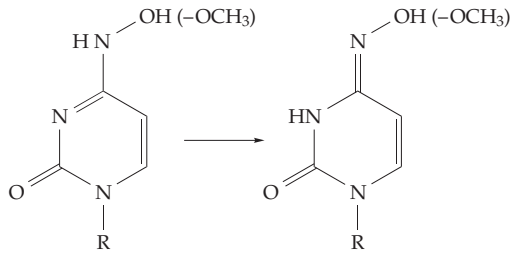

(5-10)

presumably because of the lessened aromatic character of the ring. Cytidine is slowly deaminated by base, presumably as a result of attack by hydroxyl ion on the electrophilic center at C4 and subsequent elimination of $NH_3$ (Eq. 5-11). The reaction is catalyzed by buffer salts and by bisulfite and hydroxylamine. Catalysis probably occurs, at least in part, as a result of addition of these nucleophiles at the 6 position to form compounds with increased nucleophilic reactivity at C4.[548] Hydroxylamine and methoxyamine ($NH_2OCH_3$) participate in reactions parallel to that of the hydroxyl ion in Eq. 5-11. Products contain –NHOH or –NH–$OCH_3$ in place of –$NH_2$ but tautomerize to the more stable forms shown in Eq. 5-12. Similar substitution reactions occur with other amines.[33] The C6 adduct with hydrazine can undergo ring cleavage (Eq. 5-13). The initial product then undergoes β elimination, leaving ribosylurea or deoxyribosylurea. The same reaction can be carried out on intact strands of DNA and is widely used in determination of nucleotide sequences.

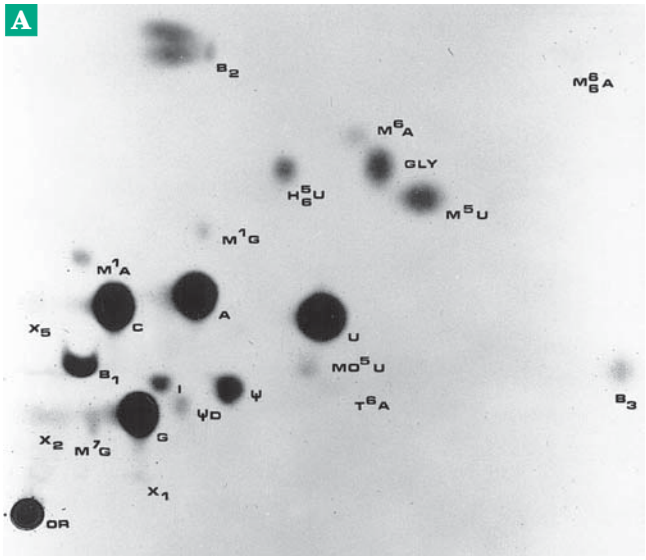
The conversion of 5-hydroxymethylcytosine to the




(5-11)




(5-12)



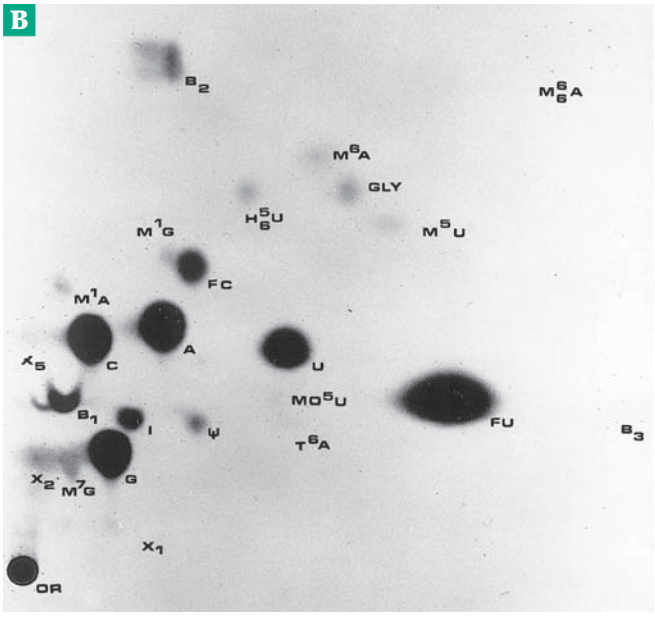
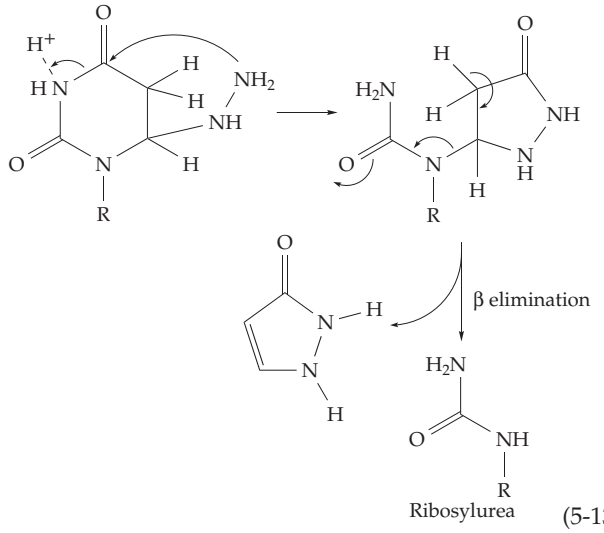




**Figure 5-44** Fluorographs of 2′3′-[$^3$H] nucleoside trialcohols from *Bacillus subtilis* grown in the absence (A) and presence (B) of 5-fluorouracil. About 2.3 nmol of nucleosides from each sample was chromatographed and exposed to X-ray film for 90 h. at –80 °C. Or (origin) $B_1$, $B_2$, and $B_3$ contain unidentified material present in a reaction mixture lacking RNA. Abbreviations used: FU, 5-fluorouridine; FC, 5-fluorocytidine; U, uridine; C, cytidine; G, guanosine; A, adenosine; I, inosine; m′G, 1-methylguanosine; $m^7G$, 7-methylguanosine; m′A, 1-methyladenosine; $m^6A$, 6-methyladenosine; $m^6{}_6A$, 6,6-dimethyladenosine; $t^6A$, N-[9-(β-D-ribofuranosyl) purin-6-yl carbamoyl] threonine; $H^5{}_6U$, 5,6-dihydrouridine; ψ, pseudouridine; $ψ_D$, decomposition product of ψ; $m^5U$, 5-methyluridine (ribosylthymine); $mo^5U$, 5-methoxyuridine; N′, a nucleoside trialcohol obtained by reduction of a nucleoside dialdehyde with [$^3$H]$NaBH_4$; FU-5 and FU-20 samples correspond to tRNAs from cells grown at that final concentration of 5-fluorouracil in μg / ml. Courtesy of Ivan Kaiser.

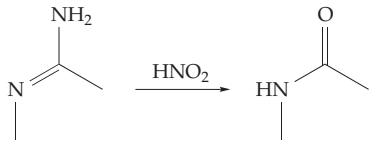



(5-13)

5-methylenesulfonate (5-$CH_2SO_3^-$) by reaction with bisulfite should also be mentioned.[549] This is a nucleophilic displacement on the electron-deficient methylene group of this base.

***Reactions with electrophilic reagents.*** Reactions of nucleic acids with the simplest electrophile, the proton, have been considered in Section A2. Somewhat similar are the reactions by which metal ions bind at many sites on both the bases and the phosphate groups of the backbone.[550]

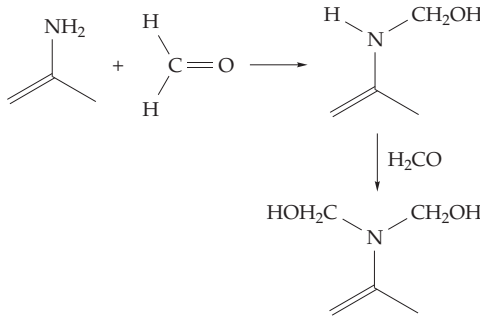
An important reaction is the deamination of amines by dilute **nitrous acid**. This reagent, by a complex mechanism, converts the amino groups of cytidine, adenosine, and guanosine to hydroxyl groups; hydroxy compounds tautomerize to the corresponding amides (Eq. 5-14).




(5-14)

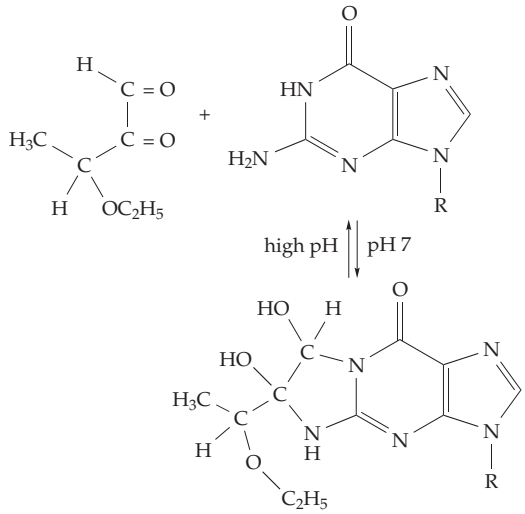
Cytidine reacts more rapidly than does adenosine which in turn reacts more rapidly than guanosine. The reaction converts cytosine into uracil and adenine into hypoxanthine. The changes are mutagenic because during replication the modified bases of the DNA pair differently than do the original bases. Guanine is converted to xanthine but this is not likely to be highly mutagenic. Nitrous acid can also convert uridine to 5-nitrouridine.

The amino groups of the bases react reversibly with aldehydes but to a lesser extent than do the more strongly basic amino groups of the amino acids. Formaldehyde forms adducts containing either one or two molecules of the aldehyde (Eq. 5-15).




(5-15)

These are reversible reactions. A more nearly irreversible crosslinking can occur by elimination of water between one of these products and a nucleophilic group in another base. A dicarbonyl reagent that is widely used because of its specificity toward guanine is **kethoxal** (Eq. 5-16).




(5-16)

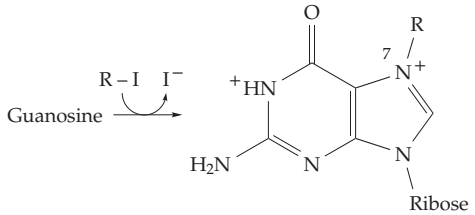
Formation of the cyclic product is a consequence of the presence of the adjacent amino and NH groups in the guanine ring.

Pyrimidines undergo **halogenation** at position 5 (Eq. 5-17), while guanine reacts at position 8. Adenine is quite unreactive. Elemental halogens or a variety of other halogenating reagents may be used. Of special

$$Br_2 + \text{cytidine} \rightarrow \text{5-Br-cytidine} + H^+ + Br^- \quad (5\text{-}17)$$

value is iodination with $^{131}I$ or $^{125}I$, by which a high level of radioactivity may be introduced into nucleic acids.

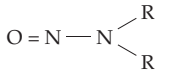
**Alkylation** reactions are not only of use in structural studies but also provide the basis for the action of a large class of mutagenic compounds.[528] Treatment of a nucleoside, nucleotide, or nucleic acid with an alkyl iodide or a dialkylsulfate converts residues of guanosine to an $N^7$-alkyl-guanosine (Eq. 5-18).




(5-18)

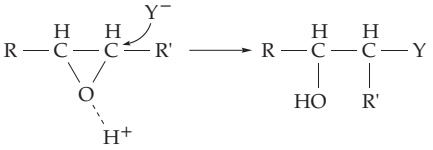
Reaction occurs at other nitrogen atoms as well as the oxygen atom of the base and of the ribose ring to a lesser extent. Adenine is alkylated preferentially at N-1 and cytosine and thymine at the corresponding position (N-3) almost exclusively. Uridine and thymidine react very slowly. Adenine is also alkylated at N-3, N-7 and at the exocyclic N-6.

Other alkylating reagents include the powerful mutagens dialkylnitrosamines and alkylnitrosoureas.



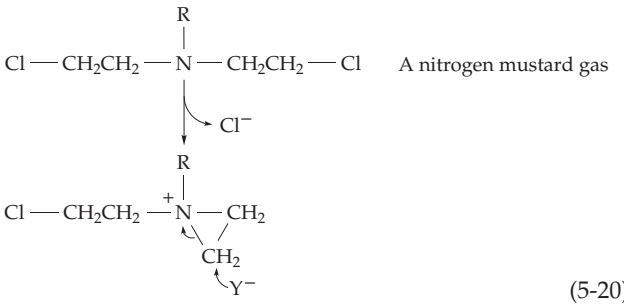

Dialkylnitrosamines

Epoxides alkylate by a nucleophilic displacement reaction that opens the ring (Eq. 5-19).



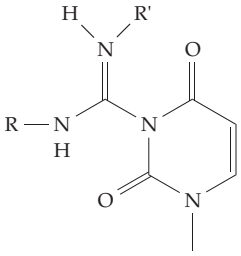

(5-19)

The nitrogen and sulfur mustards undergo internal ring closure to an iminium ion (Eq. 5-20) which can then open by attack of a nucleophilic atom of the nucleic acid.



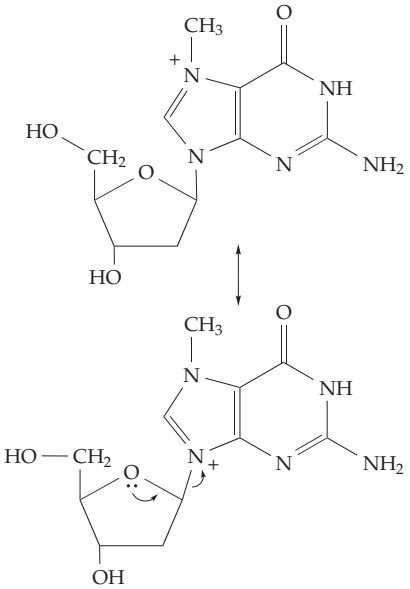

(5-20)

Other alkylating agents react through nucleophilic addition to a carbon–carbon double bond. Thus, acrylonitrile reacts with the nitrogen or oxygen atoms of nucleic acids in the same manner as does the SH group in Eq. 3-25. The water-soluble carbodiimides react as in the first step of Eq. 3-10 to form adducts of the following type:



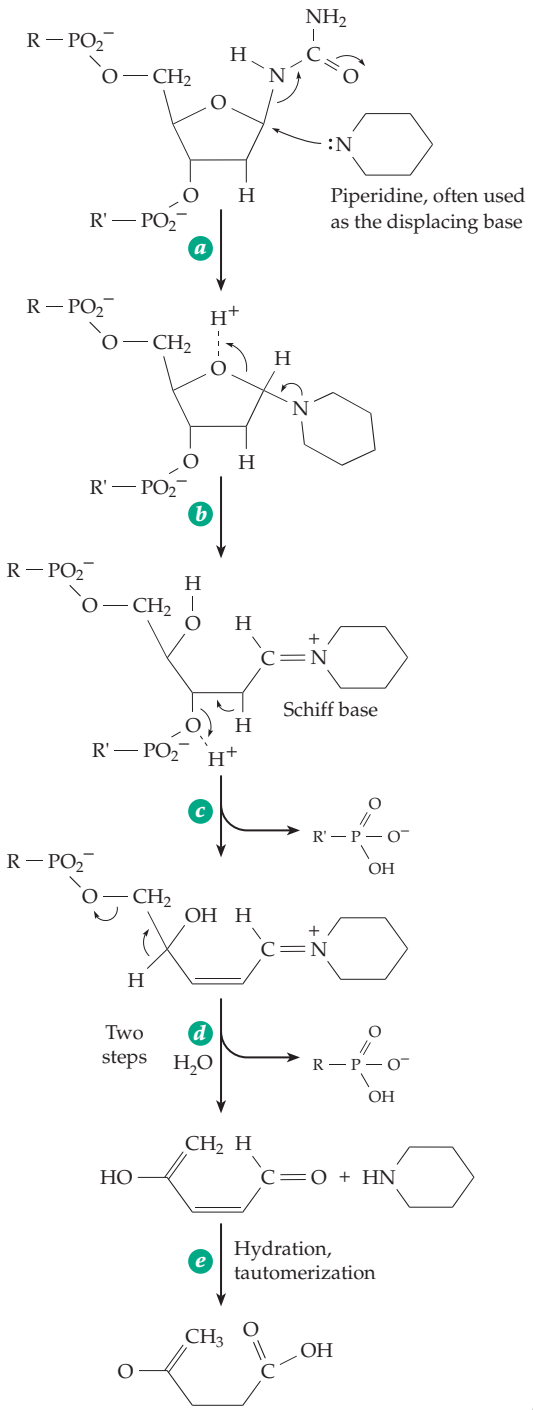

There are many other alkylating agents which often display widely varying reactivity and specificity toward particular nucleic acid bases and particular nucleotide sequences.

A striking effect of alkylation of guanine in nucleic acids is the labilization of the *N*-glycosyl linkage to the ribose or deoxyribose. This effect can be understood in terms of the induction by resonance of a partial positive charge on the nitrogen of the glycosyl linkage.




As is indicated by the small arrows on the right-hand structure, the positive charge assists in an elimination reaction that produces an oxycarbocation. The latter can then react with a hydroxyl ion from water.

***Reactions causing cleavage of the sugar-phosphate backbone.*** Treatment of DNA with 16–18 M hydrazine (Eq. 5-13) leads to the destruction of the pyrimidine rings. The reaction can be made somewhat specific for cytosine by carrying it out in the presence of a high concentration of chloride.[551] The remaining polymer, an apyrimidinic acid, contains residues of ribosylurea. These undergo an amino-catalyzed displacement and a β elimination sequence that cleaves the polynucleotide chain (Eq. 5-21). Hydration of the aldehyde (Eq. 13-1) and several




(5-21)

tautomerization steps are involved in step *e* of this equation. This reaction is very useful in sequence determination (Section 6). Notice that "tracts" of purine nucleotides remain intact after this treatment. A similar base-catalyzed reaction sequence can be used to displace $N^7$-methylguanine and to cleave the polynucleotide. Ethylnitrosourea, in its reaction with purines, is useful as a structural probe of RNA.

## 4. Melting, Hybridization, and Polynucleotide Probes

Like proteins, nucleic acids can undergo denaturation. The strands of the double helix of DNA are separated and the double-stranded regions of RNA molecules "melt." Denaturation can be accomplished by addition of acids, bases, and alcohols or by removal of stabilizing counter ions such as $Mg^{2+}$. The product is a random coil and denaturation can be described as a helix → coil transition. Denaturation of nucleic acids by heat, like that of proteins, is cooperative (Chapter 7, Section A,3) and can be described by a characteristic **melting temperature**.

A plot of the optical absorbance at 260 nm (the wavelength of maximum light absorption by nucleic acids) versus temperature is known as a **melting curve** (Fig. 5-45). The absorbance is lower, by up to 40%, for native than for denatured nucleic acids. This **hypochromic effect** (Chapter 23) is a result of the interaction between the closely stacked bases in the helices of the native molecules. The melting temperature $T_m$ is taken as the midpoint of the increase in absorbance (Fig. 5-45). As the percentage of G + C increases, the nucleic acid becomes more stable toward denaturation because of the three hydrogen bonds in each GC pair. $T_m$ increases almost linearly with increases in the G + C content. In the "standard" citrate buffer (0.15 M NaCl + 0.015 M sodium citrate, pH 7.0) Eq. 5-22 holds. The exact numerical relationship depends strongly upon the ionic composition and pH of the medium.[37,72,552,553]

$$\%(G + C) = 2.44(T_m - 69.3);\ T_m \text{ in °C} \qquad (5\text{-}22)$$

The curves in Fig. 5-45 appear simple, but using newer apparatus and plotting the first derivative of the melting curve yields a complex pattern that depends on the sequence of bases.[555]

Complete denaturation of DNA leads to separation of the two complementary strands. If a solution of denatured DNA is cooled quickly, the denatured strands remain separated. However, if the temperature is held for some time just below $T_m$ (a process known as **annealing**), the native double-stranded structure can be reformed. An important tool for studying DNA has been the measurement of the **kinetics of reassociation** of separated strands of relatively short DNA fragments.[72,556,557]

Because it depends upon the concentration of two separated strands, reassociation obeys second-order kinetics (Chapter 9) and Eq. 5-23, which is readily derived by integrating Eq. 9-8 for [A] = [B] = C from time 0 to *t*:

$$C / C_0 = 1 / (1 + k\,C_0 t) \qquad (5\text{-}23)$$

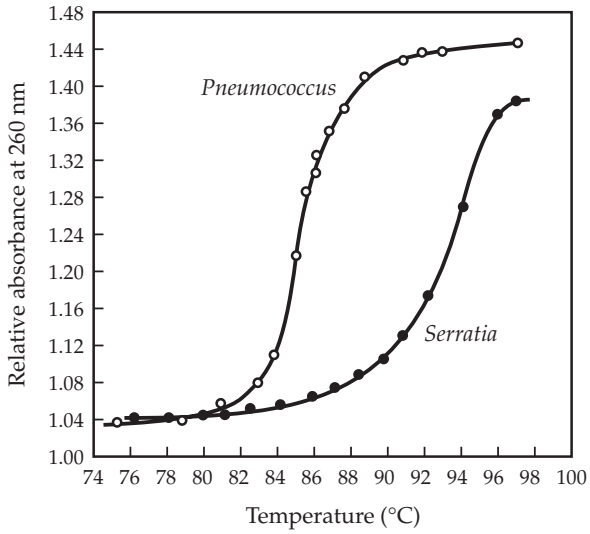



**Figure 5-45** A melting curve for DNA molecules from two different sources. From Davidson.[554]

The initial concentration of denatured DNA, $C_0$, is related in this way to the concentration C of DNA remaining dissociated at time $t$. A plot of the fraction of molecules remaining single-stranded versus the logarithm of $C_0t$ (Fig. 5-46A) is a convenient way of displaying data. As indicated in Fig. 5-46B, the value of $C_0t$ increases in direct proportion to the length of the DNA chain in the genome, but it is very much decreased if the sequence of bases is highly repetitive [poly(T) and poly(A)]. The slope of the plot at the midpoint gives an indication of the heterogeneity of the DNA fragments in a solution.

Denatured DNA fragments can sometimes reassociate with DNA from a different source to form **hybrid duplexes**. Such double helices, in which one strand comes from one strain of an organism and the other strand from a genetic variant of the same organism or from a different species, are known as **heteroduplex**. Some mutations consist of **deletions** or **additions** of one or a substantial number of bases to a DNA chain. Heteroduplexes prepared from DNA of such mutants hybridized with that from a nonmutant strain have normal hydrogen-bonded Watson–Crick base pairs for the most part. However, they may have single-stranded loops in regions where long deletions or additions prevent complementary base pairing.

Hybridization measurements have been used in many studies of **homology** of nucleic acids from different species. A nucleic acid is cut (e.g., by sonic oscillation) into pieces of moderate length (~ 1000 nucleotides) and is denatured. The denatured DNA fragments are mixed with denatured DNA of another species. Nucleotide sequences that are closely similar between species tend to hybridize, whereas sequences that are drastically different between two species do not (Fig. 5-46). One way to do such an experiment is to immobilize the long-chain denatured DNA from the one organism by embedding it in an agar gel[558] or by absorbing it onto a nitrocellulose filter.[559,560] The DNA fragments from the second organism are passed through a column containing "beads" of the DNA-containing agar or through the filter with adsorbed DNA. Pairing of fragments with complementary sequences occurs

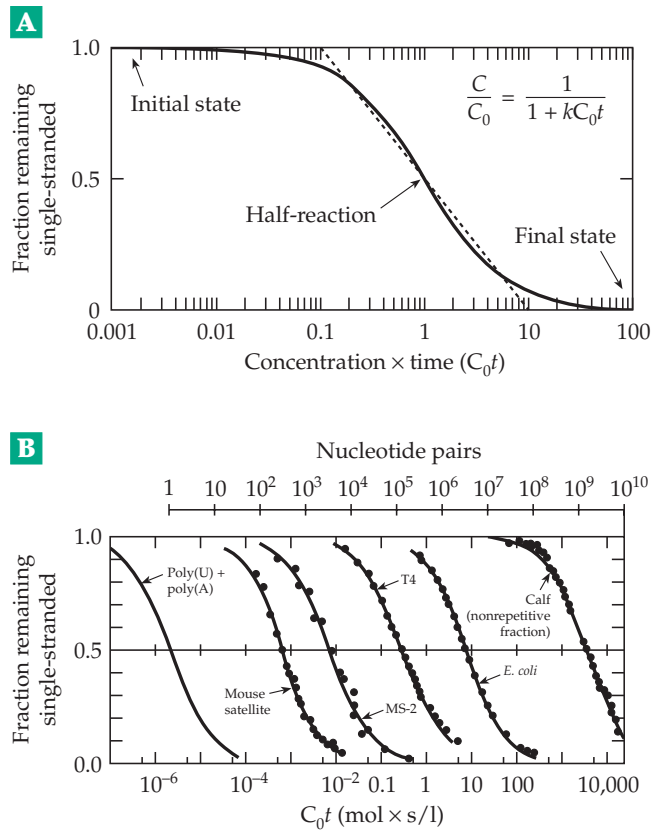


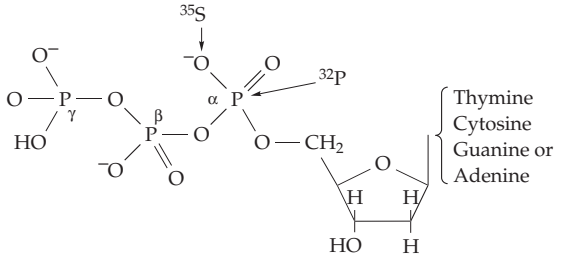

**Figure 5-46** Reassociation curves for DNA from Britten and Kohne.[556,561] (A) Time course of an ideal, second-order reaction to illustrate the features of the log $C_0t$ plot. The equation represents the fraction of DNA which remains single-stranded at any time after the initiation of the reaction. For this example, $k$ is taken to be 1.0, and the fraction remaining single-stranded is plotted against the product of total concentration and time on a logarithmic scale. (B) Reassociation of double-stranded nucleic acids from various sources. The genome size is indicated by the arrows near the upper nomographic scale. Over a factor of $10^9$, this value is proportional to the $C_0t$ required for half-reaction. The DNA was sheared, and the other nucleic acids are reported to have approximately the same fragment size (about 400 nucleotides, single-stranded). Correction has been made to give the rate that would be observed at 0.18 M sodium-ion concentration. The temperature in each case was optimal, i.e., ~30°C below the melting temperature $T_m$. The extent of reassociation was established by measuring optical rotation (calf thymus DNA), ribonuclease resistance (MS-2), or hypochromicity.

and such paired fragments are retained while strands that do not pair pass on through the column (or filter).

Both DNA hybrids and **DNA–RNA hybrid duplexes** are very important to present day genetic research.[560,562] Molecules of mRNA that represent transcripts of a particular gene will hybridize only with one of the two separated strands of DNA for that gene.

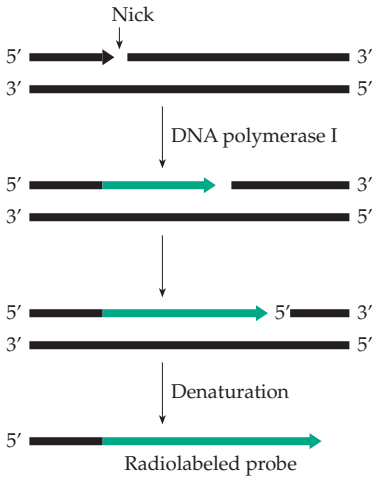
A major use of hybridization is to locate a gene or other DNA or RNA sequence by means of a **synthetic probe**.[563] This is a small piece of DNA or RNA which is labeled in some way, e.g., with a radioisotope such as $^{3}H$, $^{32}P$, or $^{125}I$. Alternatively, the probe may carry a highly fluorescent dye or a "tag" that can be recognized by a specific antibody.[564,565] An example of the latter is the use of the vitamin biotin and the specific binding protein **avidin** (see Box 14-B).[566] Related procedures employ labeling with **digoxigenin** and often employ chemiluminescent detection.[567–569] Several methods for preparation of probes are in use. Some are enzymatic but the direct chemical synthesis of oligonucleotide probes is probably used the most.

One of the first methods devised for making a highly radioactive DNA probe is called **nick translation**.[570] A piece of dsDNA, e.g., a "restriction" fragment cut from a larger piece of DNA by restriction endonucleases, is selected. A small amount of pancreatic DNase I is added. It creates "nicks" in which one strand has been cut and some nucleotides have been removed leaving a gap. Now the DNA is incubated with DNA polymerase I (pol I) and a mixture of the four mononucleotide triphosphates, the precursors of biological synthesis of DNA (Chapter 27). Usually a $^{32}P$ or $^{35}S$ label with high radioactivity is present in one of the nucleotide triphosphates as indicated in the following structure. The polymerase fills the gap, adding nucleotides to the exposed 3' end of the nicked chain. *E. coli* pol I has a second enzymatic activity which allows it to digest a polynucleotide chain from




The four nucleotide triphosphate precursors of DNA

the 5' end. Thus, as synthesis proceeds at the 3' side the nick is "translated" as shown in (Eq. 5-24).




(5-24)

Probes may also consist of DNA copied from mRNA. This is known as **cDNA** and is also widely used to determine indirectly the sequences of mRNA molecules. Messenger RNA may be isolated from the total cellular RNA by affinity chromatography on bound poly (dT) or poly (U). These materials selectively hold RNA with the poly (A) tails characteristic of most eukaryotic mRNA (see Chapter 28). Another source of mRNA is polyribosomes (polysomes), which are "reading" mRNA and actively making proteins. An antibody to the protein for which mRNA is desired will often bind to the protein chains being synthesized and precipitate the polysomes. The mRNA can be recovered and used as a template for cDNA.

Synthesis of cDNA, usually in radiolabeled form is accomplished with **reverse transcriptase**, the enzyme from retroviruses that synthesize a DNA–RNA hybrid from ssRNA.[570–572] A short oligo (dT) primer is usually hybridized to the 3' poly (A) tail to initiate synthesis. Reverse transcriptase also has ribonuclease (RNase H) activity and will digest away the RNA. If desired, synthesis of the second strand can be carried out by a DNA polymerase to give a complete DNA duplex. Many gene sequences have been deduced from cDNA copies.

Often the most practical approach to obtaining a DNA probe is synthesis of a mixture of short oligonucleotides, often in radioactive form as described in the next section. The "redundancy" in the genetic code, i.e., the existence of two or more codons for most amino acids, presents a problem in designing an oligonucleotide probe based on amino acid sequence information. Examination of Table 5-5 suggests part of the solution. Whereas only Met and Trp have single unambiguous codons nine amino acids have only two codons each. We should try to find an amino sequence that contains Met and Trp and as many of the nine others as possible. We should avoid sequences that contain Ser, Leu, or Arg because each has six codons. We can then make a mixture of oligonucleotides, using

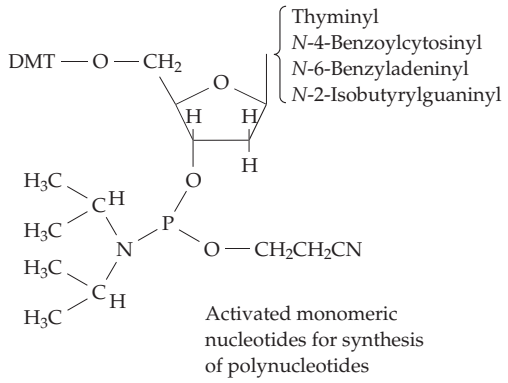
the various coding possibilities. Mixtures of as many as 1024 ($2^{10}$) oligonucleotides have been used. One of these may bind tightly and specifically to the desired DNA segment. Instead of such a complex mixture it may be more useful to incorporate a modified base at the most ambiguous positions. For example, inosine, which occurs in the wobble position in anticodons (Fig. 5-30), can pair with A, C, or T.[573] Substitution of 2-aminoadenosine can cause a probe to bind more tightly because a third hydrogen bond will be present in each AT pair.[574]

Another important procedure is labeling ends of polynucleotides. Most often the 5' end is labeled with a radioisotope or by covalent attachment of a fluorescent dye. For example, a **polynucleotide kinase** can be used to transfer a radioactive γ-phospho group from ATP to the 5' end of a polynucleotide that has a free 5'-OH group.

## 5. Synthesis of Oligonucleotides and Polynucleotides

Efficient solid-phase methods of synthesis analogous to those for polypeptides (Fig. 3-15) have been devised. The pioneering work was done by H. G. Khorana, who made the first synthetic gene[575] and later synthesized a gene for the visual pigment rhodopsin (Chapter 23). Several synthetic approaches have been developed.[576] Currently the most popular method involves the use of phosphite esters. Most nucleophilic groups of the monomers are derivatized with removable blocking groups. For example, N-4 of cytosine and N-6 of adenosine may carry benzoyl groups. The 5'–OH of each nucleotide is blocked by a di-*p*-anisylphenylphenylmethyl (also called dimethoxytrityl, DMTO) group. The 3'–OH is converted to one of a number of activated derivatives such as the following *N*, *N*-diisopropylamino phosphines.[575,577–579]

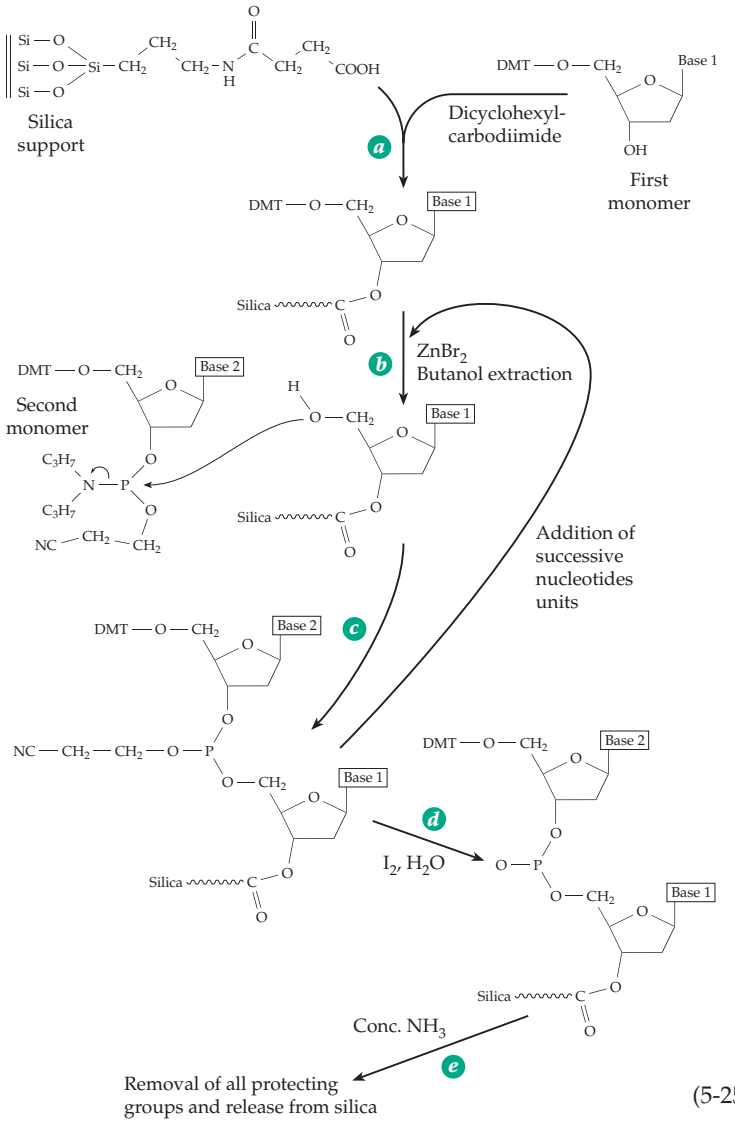
Solid-phase synthesis is usually done on a silica support with a covalently attached succinamide as shown in Eq. 5-25. The first nucleotide at the 3' end of the chain to be synthesized is attached by an ester linkage to the bound succinamide (step *a*, Eq. 5-25). The 5'-protecting group is removed in step *b* and the 5'-OH reacts with the activated phosphine of the second nucleotide (step *c*, Eq. 5-25). Steps *b* and *c* are then repeated as often as necessary to complete the chain. The finished polynucleotide can be removed from the solid support, the cyanoethyl groups removed



Activated monomeric nucleotides for synthesis of polynucleotides

from the phosphorus atoms by β elimination and all of the other blocking groups removed by treatment with concentrated $NH_3$. The whole procedure has been automated.[580–582]

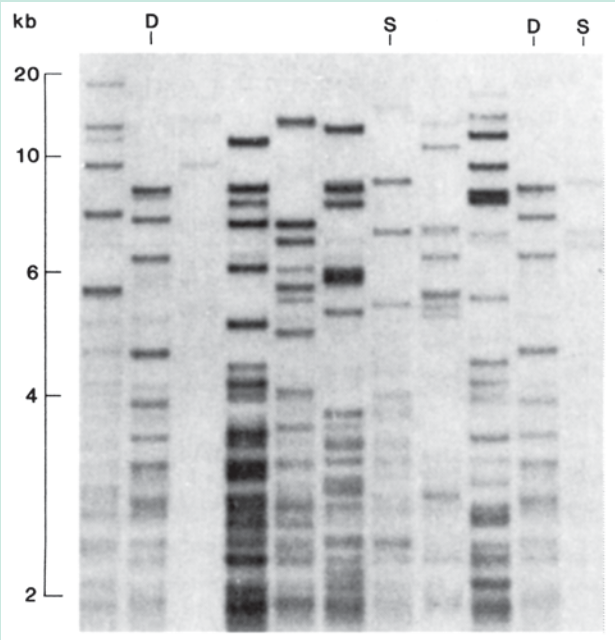
If a large piece of DNA is needed several oligonucleotides can be joined end to end enzymatically (Eq.



(5-25)

## BOX 5-D DNA FINGERPRINTING

Jeffreys *et al.*[a–c] digested human DNA to completion with *Hin* f I and *Sau*3A restriction endonucleases. Certain fragments, which originated from "minisatellite" bands of repetitive DNA showed a very high degree of **polymorphism** among the population. Many different fragments sharing these repeated sequences were formed in the restriction digest. If a suitable labeled probe was used, it hybridized with as many as 80 different bands.[a,d–f] The resulting pattern appeared, like a fingerprint, to be different for every individual, as is shown in the accompanying photo. Unlike a fingerprint the DNA pattern also contains information that often allows deductions about parentage.




DNA "fingerprints" made from one or two drops of blood from ten different individuals. DNA was isolated, digested to completion with restriction endonuclease *Hin* f I, and subjected to electrophoresis in a 20-cm-long agarose gel until all DNA fragments smaller than 1.5 kb in length had passed off the gel. The DNA was then transferred to a nitrocellulose filter by Southern's method and hybridized with a $^{32}$P-labeled single-stranded DNA probe prepared from cloned human minisatellite DNA. The probe used had the "consensus" composition (AGAGGTGGGCAGGTGG). Within the 29 tandem repeats in this 0.46-kb probe there are various sequences close to the one shown. Filters were then autoradiographed for four days. Two duplicate samples (marked D) were taken from the same individual and two others (marked S) from two sisters. A number of bands in common are evident. From Jeffreys *et al.*[c]

The technique has come into widespread use in forensic analysis, with DNA typing being possible from a single hair.[g] In a famous early case it was used to allow an immigrant child to be reunited with his mother[h] and it is being used regularly to protect innocent persons accused of rape or murder.[i] It is also widely used to provide evidence of guilt. However, the very small chance of a close match between unrelated persons prevents the use of DNA typing alone as proof of guilt. Because DNA samples are often "amplified" by PCR (Section H,6), there is also a possibility of contamination and forensic use of DNA typing is still controversial.[i–m]

However, DNA typing continues to be improved[f,n,o] and to be applied in a great variety of ways. For example, the skeletal remains of a murder victim were identified by DNA fingerprints after being buried for eight years.[p] DNA typing is also used to study mating habits of birds,[q] the genetic variability of populations of whales sampled by biopsy,[r] etc.


[a] Jeffreys, A. J., Wilson, V., and Thein, S. L. (1985) *Nature (London)* **314**, 67–73
[b] Lewin, R. (1986) *Science* **233**, 521–522
[c] Jeffreys, A. J., Wilson, V., and Thein, S. L. (1985) *Nature (London)* **316**, 76–79
[d] Vassart, G., Georges, M., Monsieur, R., Brocas, H., Lequarre, A. S., and Christophe, D. (1987) *Science* **235**, 683–684
[e] Huang, L.-S., and Breslow, J. L. (1987) *J. Biol. Chem.* **262**, 8952–8955
[f] Kirby, L. T. (1990) *DNA Fingerprinting*, Stockton Press, New York
[g] Higuchi, R., von Beroldingen, C. H., Sensabaugh, G. F., and Erlich, H. A. (1988) *Nature (London)* **332**, 543–546
[h] Jeffreys, A. J., Brookfield, J. F. Y., and Semeonoff, R. (1985) *Nature (London)* **317**, 818 – 819
[i] Balding, D. J., and Donnelly, P. (1994) *Nature (London)* **368**, 285–286
[j] Neufeld, P. J., and Colman, N. (1990) *Sci. Am.* **262**(May), 46–53
[k] Lewontin, R. C., and Hartl, D. L. (1991) *Science* **254**, 1745–1750
[l] Lander, E. S., and Budowle, B. (1994) *Nature (London)* **371**, 735–738
[m] Lewontin, R. C. (1994) *Nature (London)* **372**, 398
[n] Uitterlinden, A. G., Slagboom, P. E., Knook, D. L., and Vijg, J. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2742–2746
[o] Jeffreys, A. J., MacLeod, A., Tamaki, K., Neil, D. L., and Monckton, D. G. (1991) *Nature (London)* **354**, 204–209
[p] Hagelberg, E., Gray, I. C., and Jeffreys, A. J. (1991) *Nature (London)* **352**, 427 – 429
[q] Burke, T., and Bruford, M. W. (1987) *Nature (London)* **327**, 149–152
[r] Hoelzel, A. R., and Amos, W. (1988) *Nature (London)* **333**, 305

27-5). For example, a functional 17-bp gene for the 53-residue human epidermal growth factor was synthesized by joining ten oligonucleotides of lengths 11–59 bp.[575] DNA is often synthesized enzymatically using methods described in Chapter 26. Cloned sequences of synthetic DNA can also be transcribed to produce **polyribonucleotides** of any desired sequence.[583] New nonenzymatic methods for RNA synthesis have also been devised.[583–587]

## 6. The Polymerase Chain Reaction (PCR)

This important technique was first described in 1971–1974 by Khorana and associates[588,589] but was not used until it was rediscovered in 1983 by Mullis.[590–593] It was quickly developed[591,592,594–596] and has played a major role in biochemistry ever since. It continues to be applied in numerous ways.[589,597–599a]

The PCR technique provides a way of "amplifying" a small number of DNA molecules, i.e., to produce many copies. This is often done by cloning but PCR offers a quick and easy way to obtain millions of copies of a desired relatively short segment of DNA. Standard PCR can be used for up to about 5000-nucleotide pieces. More recently modified procedures have allowed 35-kb segments to be amplified.[600]

The basic PCR procedure is initiated by hybridizing two oligonucleotide primers onto opposite strands of denatured DNA, one at each end of the section chosen for amplification (Fig. 5-47). A DNA polymerase is then used to convert each of the separated strands into a duplex. The mixture of products is then heated to denature the two new duplexes. After cooling, the primers, which are present in great excess, hybridize to all four strands. In a second cycle of polymerase action these are all converted to duplexes, etc. After 20 cycles millions of copies will be made. At first, copies with tails extending beyond the limits specified by the oligonucleotide primers will be formed. However, it is easy to see that after a few cycles, most molecules will be of just the desired length. A heat-stable polymerase from *Thermus aquaticus* (*Taq* polymerase) is used so that the enzyme is not denatured by the repeated cycles of heating and cooling, which are conducted automatically by a simple apparatus.

The polymerase chain reaction is being used to speed up prenatal diagnosis of genetic diseases, to detect viral infections, for tissue typing needed for organ transplantation, in forensic procedures, and in the study of the DNA of ancient tissues such as those of frozen wooly mammoths.[601–603] If suitable restriction enzyme sites are present in the primers, the amplified DNA can be cloned readily.[604] The $3.3 \times 10^{-9}$ fmol of a DNA sequence found in a diploid chromosome pair in a single cell can be amplified in 50 cycles to 5–500 fmol, enough to study by hybridization with radioactive probes.[605] A large sample of a few pg of DNA can be amplified in 20 cycles to micrograms. By placing sequencing primers within the amplified segments, it is possible to generate DNA that can be sequenced directly using the dideoxy sequencing technique (Chapter 5) without cloning.[606] The PCR technique has also been used to amplify cDNA molecules formed from RNA transcripts present in very low abundance. One of the problems with the PCR is that priming may occur by DNA fragments other than the added primers. Contamination must be scrupulously avoided. Another problem is that errors are introduced into DNA during amplification by PCR. Perhaps 1 in 200 of the copies will contain an incorrect base.[607] If such a molecule is cloned the error will be perpetuated. Good practice requires that more than one clone is selected and sequenced to allow such errors to be avoided.

## 7. Sequence Determination

Satisfactory (but slow) methods for determining sequences of RNA molecules have been known for over 30 years. The procedures are somewhat parallel to those used in sequencing proteins. However, no similar method could be devised for DNA. Little progress was made until rather recently when new approaches led to extremely rapid procedures for sequencing DNA. As a consequence, it is now much easier to learn the sequences of genes than it is to sequence the proteins which they encode!

***Preparing the DNA.*** The first step is to obtain a sample of enough identical DNA molecules to permit sequence analysis. This in itself may be a complex undertaking. Perhaps we want to know the sequence of one particular gene in the 3,500,000 kilobase pairs of DNA present in a single human cell. How can this gene be found and the DNA be obtained for analysis? Three techniques have been essential: cutting the DNA with restriction endonucleases, hybridization, and cloning. More recently PCR and related methods[608] have simplified the sample preparation. DNA can often be amplified using primers that contain sequences that will later serve as **sequencing primers**.

***Restriction maps and Southern blots.*** Although it doesn't require the synthesis of a primer, the Maxam–Gilbert procedure usually demands that a "restriction map" of the DNA be prepared to help keep track of the fragments being sequenced.[609] See also Chapter 26. Figure 5-48[610–612] shows the restriction map of the mitochondrial DNA gene, *oxi3* from yeast. This gene, which encodes one of the subunits of cytochrome oxidase (Chapter 18), consists of 9979 base pairs. It was cloned in a suitable plasmid after which the restriction map (Fig. 5-48) was prepared by cutting with 18 different

restriction enzymes.[611] The protein subunit contains 510 residues and therefore requires a coding capacity in the DNA of 1530 base pairs. This is only 16% of the total length of the gene, the majority of whose DNA is found in the three large introns.

One way to select a desired segment of DNA from a digest of chromosomal DNA is to sort out the "restriction fragments" by gel electrophoresis.[613] The DNA from the gel can be transferred to a nitrocellulose sheet while retaining the separation pattern using a method devised by Southern.[560,614,615] In this **Southern blot** technique, solvent flows from a pool beneath the gel up through the gel and the nitrocellulose sheet into paper towels. The DNA is trapped on the nitrocellulose in the same pattern observed in the electropherogram. A suitably labeled probe such as cDNA with incorporated $^{32}P$ is flowed repeatedly across the nitrocellulose sheet under conditions that favor formation of hybrids. Only the DNA complementary to the cDNA probe will retain the label. This DNA can then be located with the help of a autoradiogram. It is important that single-stranded DNA be used. If double-stranded restriction fragments are separated on the electropherogram they must be denatured while in place in the gel before hybridization is attempted.

Once the desired piece of DNA has been identified it is usually necessary to increase its amount. The conventional approach is to incorporate the DNA fragment into a plasmid and clone by the methods described in Chapter 26. The selected DNA can usually be cut cleanly from the plasmid used for cloning with the same restriction endonuclease originally used in

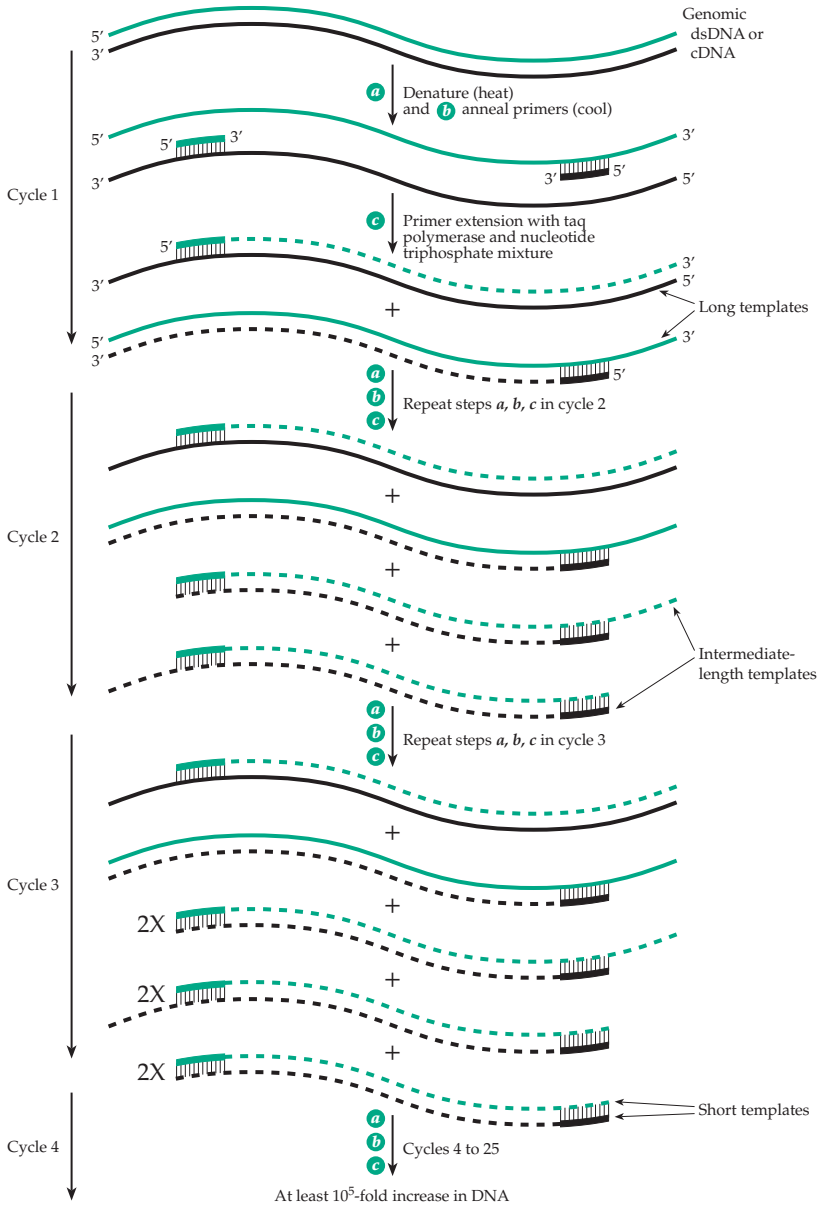


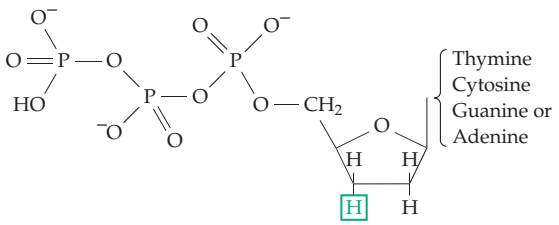

**Figure 5-47** Amplification of DNA using the polymerase chain reaction (PCR). Double-stranded DNA is denatured by heating to 90–99° C (step *a*) and oligonucleotide primers complementary to short 12–18 nucleotide sequences at the two ends of the piece of DNA to be amplified are annealed to the separated strands by cooling to 40–75° C (step *b*). The two DNA strands serve as templates for synthesis of new complementary strands using a heat-stable DNA polymerase and a mixture of the four nucleotide triphosphates. Nucleotide units are added to the 3' ends of the primers, with the new chains growing in the 5' → 3' direction (step *c*). Steps *a*, *b*, and *c* are then repeated as many as 30 times using a thermal cycler device that periodically raises and lowers the temperature with a cycle time of a few minutes. The polymerase is unharmed by the heating and is reused in each cycle. An excess of the primer and of the nucleotide triphosphates sufficient for all of the cycles is present initially. In the early cycles new long and intermediate length templates are created. However, the number of short templates increases exponentially and the final product consists predominately of the short selected DNA segment (short templates).

fragmenting the DNA. An alternative procedure is to clone a mixture of DNA fragments and then sort colonies of bacteria containing the cloned fragments using DNA–RNA hybridization.[616] Bacteria from the selected colonies are then propagated to produce large amounts of the plasmid DNA. Alternatively, PCR can be used directly on the selected DNA fragment. This is more often the preferred choice.[604,606]

***The Sanger dideoxy method.*** The rapid sequencing methods all depend upon the fact that single-stranded DNA fragments under denaturing conditions migrate on electrophoresis in polyacrylamide gels strictly according to their length. Thus, if a mixture contains all lengths of radiolabeled polynucleotides from very short oligonucleotides to fragments containing 200 or 300 bases, the polynucleotides will all appear, one above the other, as a series of bands that can be visualized by radioautography. The first of these methods was published by Sanger and Coulson[617] in 1975 and was followed in 1977 by the method which is now used.[618–621]

A sample of double-stranded DNA is denatured. One of the resulting single strands is used as a template to direct the synthesis of a complementary strand of radioactive DNA using a suitable DNA polymerase. The "Klenow fragment" of *E. coli*, DNA polymerase I, reverse transcriptase from a retrovirus, bacteriophage T7 DNA polymerase, *Taq* polymerase, and specially engineered enzymes produced from cloned genes have all been used.

Before the sequencing begins it is necessary to prepare a short **primer** that is complementary to a sequence at one end of the DNA strand to be sequenced. This may be prepared enzymatically,[622,623] or by nonenzymatic synthesis. The short primer is annealed to the end of the DNA and the resulting molecule is incubated with a DNA polymerase and a mixture of the four mononucleotide triphosphates, one of which is radiolabeled in this position. Four reaction mixtures are prepared. Each mixture contains all four nucleoside triphosphates and also one of four different **chain-terminating inhibitors**, the most popular of which are the 2', 3'–dideoxyribonucleoside triphosphates:




2',3'-Dideoxyribonucleotide triphosphates

These inhibitors are added in a ratio of about 100:1 with the natural substrates. In this ratio they are incorporated into the growing DNA chain about once in 200 times on the average. However, in the various growing DNA chains they are incorporated at different points ranging from the very first nucleotide to the last. Since the incorporated dideoxy monomer lacks the 3-hydroxyl group needed for polymer formation, chain growth is terminated abruptly. Synthesized polynucleotides are denatured and subjected to electrophoresis in four adjacent lanes. The resulting patterns contain bands corresponding to all of the successive oligonucleotides but not all in the same lane. A given lane will contain only the bands of the oligonucleotides terminated by the particular inhibitor used. The other bands will be found in the other three lanes. Each band will have been terminated by the inhibitor employed in that lane. The nucleotide sequence can be read directly from the banding pattern as is shown in Fig. 5-49. Arabinosyl nucleotide triphosphates have also been used as chain-terminating inhibitors. A sequence determined by the Sanger method is usually checked by also sequencing the complementary strand.

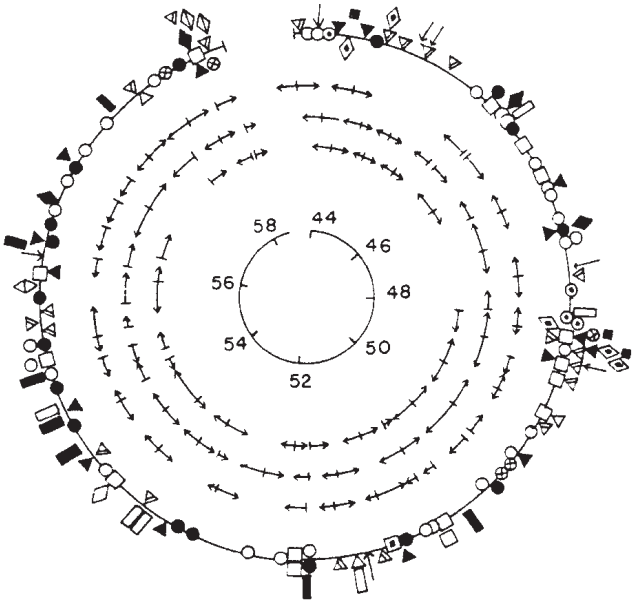



**Figure 5-48** A physical map of the *oxi3* locus of yeast mitochondrial DNA. The restriction fragments used for DNA sequencing are indicated by the *arrows*. The extent to which the sequences were read is represented by the lengths of the arrows. The map units are shown in the *inner circle*. The following symbols, together with the names of the restriction enzymes (Chapter 26), are used for the restriction sites:

| | | |
|---|---|---|
| ▲ *Hinf* I | ○ *Alu* I | ⊙ *Hha* I |
| △ *Hpa* II | ◇ *Pvu* II | ◬ *Rsa* I |
| ▯ *Hae* III | ◈ *Hinc* II | ▮ *Hph* I |
| ◆ *Taq* I | ⊖ *Hind* III | ⊡ *Blg* II |
| ● *Mbo* I | ⊗ *Eco* RI | ◇ *Bam* HI |
| □ *Mbo* II | ↓ *Eco* RII | ■ *Hpa* I |

From Bonitz *et al.*[611]

A major factor in the success of the dideoxy sequencing method has been the development of cloning techniques that provide ssDNA in a form ready for use. In any cloning procedure the DNA that is to be sequenced has been covalently ligated to the end of a DNA strand of the **cloning vehicle**, a modified plasmid or virus (Chapter 26). Sequencing is often done on DNA cloned in a modified ssDNA bacterial virus such as M13. Although dsDNA is ligated to the ds replicating form of viral DNA, the virus particles produced when the virus is propagated in *E. coli* cells are single stranded. Cleavage of the viral DNA with its incorporated "passenger DNA" with an appropriate restriction endonuclease releases the passenger DNA (to be sequenced) with a short piece of DNA from the cloning vehicle attached at the 3' end. Since the sequence of this small piece of the cloning vehicle is known, a suitable primer of length ~12–18 nucleotides can be synthesized (or purchased) and annealed to the DNA. This serves as the primer and allows the sequence to be read from the 5' end to the 3' end of the synthesized complementary strand. Double-stranded DNA attached to vehicles such as the pUC plasmids can also be sequenced directly if the DNA is denatured by alkali treatment. After neutralization and precipitation an appropriate primer is annealed to one or the other of the two strands.[624]

Since its introduction, many modifications and improvements have been developed. The sequencing gels have been improved. The use of $^{35}S$ labeling has given sharper autoradiographs.[625] Alternatively, a silver stain can be used with unlabeled primers.[626] GC-rich DNA sequences are often difficult to sequence, probably because even in the denaturing polyacrylamide gels used for sequencing they tend to form hairpin loops, perhaps as a result of formation of Hoogsteen base pairs (Fig. 5-7). Formation of these loops results in uneven spacing between the adjacent bands in the sequencing gel, so-called "compression artifacts." Use of a 7-deaza-dGTP in place of dGTP in the sequencing reaction ameliorates this problem.[627] Sensitivity can be improved by use of "cycle sequencing" in which a heat-stable polymerase such as *Taq* polymerase is used, and after heating the same template DNA is used repeatedly to give a higher yield of labeled fragments.[628,629]

About 200 – 400 bases can be successfully sequenced manually in a single run. By prolonging the time of electrophoresis in a second run, the sequence can be extended considerably. By using the just obtained sequence information, it is possible to select a new start point 200 or more nucleotides further along the template chain and to synthesize an oligonucleotide primer to anneal to the template at this point. In this way it is possible to "walk" along the template adding additional sequences at each step. Another procedure is to delete by mutation various segments of the cloned DNA above the sequence that binds the primer. This

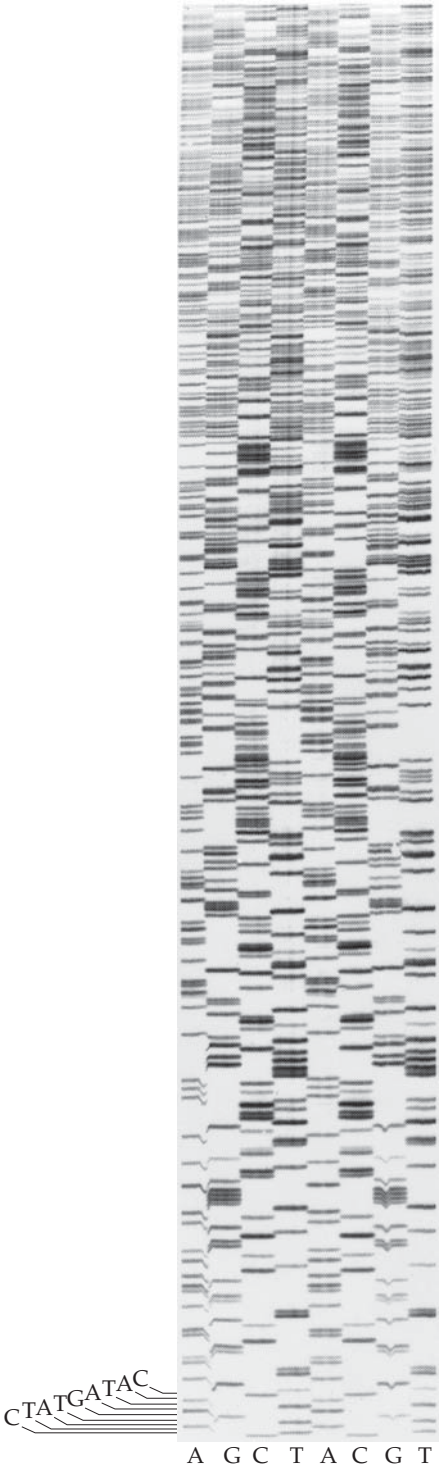



**Figure 5-49** A DNA sequencing gel obtained using a segment of DNA from salmon sperm selected by suitable oligonucleotide primers, amplified by PCR, and sequenced with a $^{35}S$ label in the primer. Four samples were used, one with each of the four dideoxy chain terminators (A, G, C, T, A, C, G, T from left to right). After electrophoresis the shorter fragments are at the lower end of the gel. The sequence of the strand complementary to the template strand whose sequence is being determined is read from the bottom of the gel. Here it starts CTATGATAC. Reproduced by permission of Amersham Pharmacia Biotech, Limited.

permits analysis of the whole cloned fragment via an overlapping set of sequences from the deletion mutants.[630,631]

Many DNA sequences continue to be determined manually by the well-developed long gel procedures as illustrated in Fig. 5-49. However, sequencing whole genomes has depended upon the development of high-speed automated procedures.[632] Instead of radiolabeling, fluorescent dyes may be joined to the primer to allow detection of the chain fragments produced during sequencing. Automatic sequencers use four dyes that fluoresce with different colors.[633,634] A different dye is used for each of the reaction mixtures. Then the four samples are mixed together and the DNA fragments are subjected to electrophoresis in a single lane. A laser beam excites the fluorescence, scanning several lanes with different samples as the electrophoresis progresses. A photomultiplier tube records the fluorescence intensity of each band through a series of four filters in a rotating wheel. This allows automatic recognition of the four different colors of fluorescence and therefore of the nucleic acid base present in each position in the sequence. Improved strategies for "primer walking"[635] and for "shotgun sequencing"[19,632,636] have been devised. In the shotgun strategy, whole bacterial genomes have been cut by restriction enzymes into large numbers of overlapping fragments which have been separated and sequenced.[637] A computer program is used to analyze and assemble the sequences into a complete genomic sequence. An example is provided by the genome of the *Methanococcus jannaschii*. Its large circular chromosome contains 1,664,976 bp and there are two additional pieces to the genome, one containing 58,407 bp and the other 16,550 bp. The sequences were deduced from 36,718 individual sequencing runs on high-speed automatic sequencers. For each run, on average, 481 bp could be read.[636] To sequence the human genome faster methods are needed.[638] Capillary electrophoresis with a single laser beam scanning the output of 24 capillaries has been demonstrated.[639,640] Extremely rapid sequencing of oligonucleotides up to 100 bp in length can be accomplished by mass spectrometry.[641] This may be an important technique for diagnosis of genetic defects (Chapter 26).

***The method of Maxam and Gilbert.*** The nonenzymatic method devised by Maxam and Gilbert[642–644] can be used to sequence either ss or dsDNA. Before the sequencing is begun, a radioactive label is incorporated, usually at the 5' end. This is often done by cleaving off any phosphate groups present on the 5' end with alkaline phosphatase and then transferring a new radioactive phospho group with the assistance of the enzyme polynucleotide kinase and radioactive $\gamma$-$^{32}$p-labeled ATP. If dsDNA is used the strands are separated so that each has a label only at one end.

The key step in sequencing by the Maxam–Gilbert procedure is to cleave chemically the DNA at random locations using reagents that have some specificity for particular bases. The cleavage process involves three distinct steps: (1) chemical modification, as specific as is possible for the chosen base; (2) displacement of the modified base from the sugar; and (3) elimination and chain cleavage using amine catalysis (Eq. 5-21). Two consecutive steps can often be combined. There are several versions of the method; one involves dimethyl sulfate as the specific reagent for guanine. It forms $N^7$-methylguanosine (see Eq. 5-18) which upon heating with the strong base piperidine at 90°C undergoes addition of hydroxyl ion with ring opening and displacement of the modified base according to Eq. 5-26. The product, a glycosylamine of piperidine, is in equilibrium with a Schiff base which can undergo chain cleavage as in Eq. 5-21.

A second sample of DNA is treated with a piperidine-formate buffer of pH 2 in the cold. This promotes the acid depurination of both guanine and adenosine. A third sample is treated with hydrazine, with both cytidine and thymidine being cleaved to ribosylurea according to Eq. 5-13. Again, this is followed by displacement and $\beta$ elimination (Eq. 5-21) catalyzed by piperidine. The fourth sample is also treated with hydrazine but in the presence of a high concentration of NaCl which inhibits the reaction with thymidine, by lowering the p$K_a$ of the thymine, and allows the cleavage to be more nearly specific for cytidine. Each of the reactions is conducted in such a way that on the average only one cleavage event occurs per molecule of DNA. Since the cleavages occur at many different points, a family of nested radioactively labeled oligonucleotides, one from each original molecule, is produced. When these are sorted by polyacrylamide gel electrophoresis, the pattern of the oligonucleotides in the four adjacent channels allows the nucleotide sequence to be read directly from the autoradiogram.

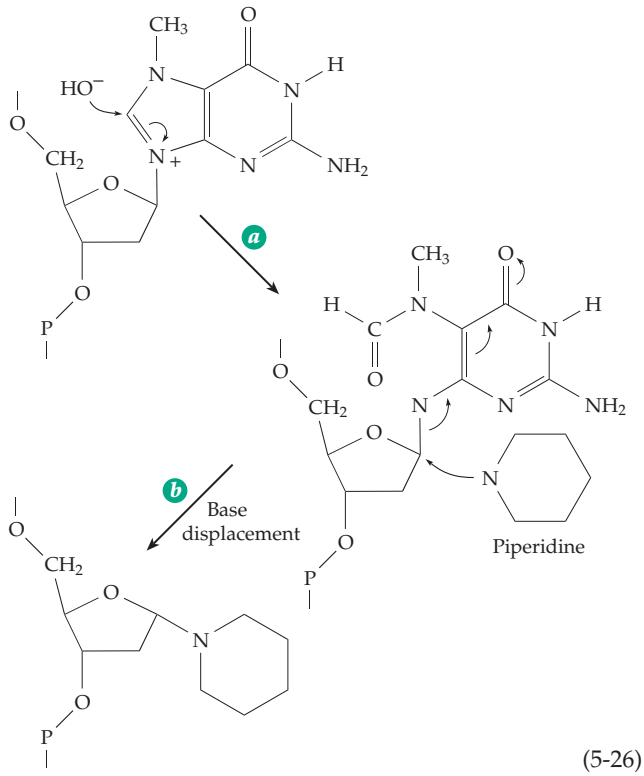
The Maxam–Gilbert method doesn't require synthesis of a primer and it sometimes works well for sequences that are difficult to obtain with the Sanger–Coulson procedure. The two methods may both be used to provide additional certainty about a sequence. The Maxam–Gilbert method is very convenient for sequencing small oligonucleotides which often react poorly with the polymerase used for the chain termination method. The method usually requires that a restriction map be prepared.

***Sequencing RNA.*** The first known RNA sequence, that of an alanine tRNA, was determined by Holley and associates in 1965. The RNA was subjected to partial hydrolysis with pancreatic ribonuclease and ribonuclease $T_1$ (Fig. 5-43). The small oligonucleotide fragments were separated by ion exchange chromatography under denaturing conditions (7 M urea) and were then characterized individually.[645] The availability

of additional enzymes such as ribonuclease $U_2$, the *B cereus* ribonuclease and ribonuclease Phy M of *Physarum* (Fig. 5-43) and the use of radiolabeling and of two-dimensional fingerprinting of digests have made the procedures more versatile.[646] A valine tRNA was sequenced independently by Bayev[647] and Campbell.[648]

Since the development of the rapid methods for sequencing DNA, many mRNA sequences have been determined by using reverse transcriptase to make a cDNA strand complementary to the RNA. The cDNA is then sequenced.[649] Rapid sequencing methods parallel to those used for DNA have also been devised.[650–652]

***Nearest neighbor analysis.*** A technique developed by Kornberg and associates before the availability of sequencing methods is the **nearest neighbor sequence analysis**. Using a single radioactive $^{32}P$-containing nucleoside triphosphate together with the three other unlabeled nucleoside triphosphates, a primer chain of DNA is elongated from the 3' end along a ssDNA template chain using a DNA polymerase. The incorporation of $^{32}P$ from the α position of the nucleotide triphosphate occurs in the bridge phosphates that connect the nucleotide originally carrying the $^{32}P$ to the 3' position of the neighboring nucleotide. Cleavage of the $^{32}P$-containing product of the reaction with a mixture of micrococcal DNase and spleen phosphodiesterase, which catalyze *b*-type cleavage (Fig. 5-43), gives fragments in which the $^{32}P$ will now be attached to what was the 5′ nearest neighbor to the radioactive nucleotide in the DNA.[72,653] Measurement of the radioactivity in each of the 3' nucleotides of thymine, cytosine, adenine, and guanine gives the frequencies of the adjacent pairs, TA, CA, AA, and GA. Using the other radioactive nucleoside triphosphates one at a time in separate experiments, all of the nearest neighbor frequencies can be obtained. From such an experiment it was possible to deduce that the strands in the double helix were oriented in an antiparallel fashion, as predicted by Watson and Crick. If the strands had been parallel, different nearest neighbor frequencies would have been observed.




(5-26)

***Understanding sequences.*** Sequences of over 20 million nucleotides from hundreds of organisms had been determined by 1988 and the number is doubling each 2–3 years.[654] Sequencing the human genome has required rates of millions of bases per day. With the massive amount of data already available it has become of great interest to compare sequences of genes, whether they encode similar or dissimilar proteins, to make comparisons between species, and to search for sequences that bind specific proteins or that encode particular regulatory signals. Relationships of common evolutionary origin or homology as well as other sequence similarities are often sought.

To handle the mass of existing data, powerful computer programs have been developed and various graphical procedures have also been developed to help the human mind comprehend the results.[654,655] One important problem is to define and locate what are called **consensus sequences**. The problem is best illustrated by examples.[654] The cleavage site for the *Eco*RI restriction endonuclease is **GAATTC**. There is no ambiguity. In a DNA of random sequence this would be expected to occur by chance in about $(1/4)^6$ nucleotides (4 kb). On the other hand, the *Hin*II restriction endonuclease cleaves within the consensus sequence **GTYRAC** where Y = C or T and R = A or G. It would be expected to occur by chance in about $1/4^5$ nucleotides. Many binding sites for RNA polymerase, the so called **promoters** (Chapter 28) contain the consensus sequence **TAtAaT**, at position -10, ahead of the 5' end of the sequence that is transcribed into mRNA. The lower case t and a used here imply that other nucleotides may often replace T or A at these positions. There are many promoters and over 70% of those described have this consensus sequence. All have the less restricted sequence **TAxxxT**, where x may be any nucleotide. Our definition of consensus sequence is somewhat arbitrary. Now consider the problem of locating a –35 site whose consensus sequence is **TTGACA** but which may, for different genes, be shifted backward or forward by a nucleotide or two. This is a consensus sequence. Therefore, in many cases one or more substitutions in the sequence will have been made. The result is that the sequence of nucleotides in

which the consensus sequence is to be found is likely to appear entirely random. Sophisticated computer programs are helpful in locating it.[654]

## 8. Protein–DNA Interactions

The most detailed information about interactions of proteins with DNA is coming from X-ray crystallographic studies. Examples are seen in Figs. 5-35 to 5-40. Several other methods have also been very useful. Much has been learned from the effects of mutations in DNA-binding proteins or in regions of DNA to which a protein binds. Binding of proteins to DNA can also be recognized by its effects on the mobility of DNA during gel electrophoresis.[656,657] Chemical[658] or laser-induced crosslinking can show that within a complex a specific residue in a protein is adjacent to a certain sequence in the DNA.

The technique of **protection mapping** or **"footprinting"** is widely used to determine which nucleotides in a sequence are covered by a bound protein.[659] A reagent which attacks and cleaves DNA nearly randomly is used. DNase I was first introduced for this purpose[659] and has been used widely. An important finding is that certain sites that are readily cleaved (**hypersensitive sites**) are frequently located in chromatin undergoing transcription. Footprinting has also been accomplished with other nucleases, with dimethylsulfate (which acts on A and C), with carbodiimides (which act on U and G), and with the Maxam–Gilbert guanine-specific cleavage (Eq. 5-26). One of the most popular methods employs cleavage by hydroxyl radicals.[660–663] **Photofootprinting** depends upon decreased or increased sensitivity to ultraviolet light at sites bound by proteins.[214,662,664] In footprinting experiments the DNA to be studied is radioactively labeled at one end of one strand. In the absence of the protecting protein, denaturation and electrophoresis of the cleaved fragments yields a nearly random "ladder" of DNA fragments. In the presence of the binding protein some cleavage products will be missing from the ladder. The bound protein leaves a "footprint" (Fig. 5-50A,C).

Related methods are being applied to the determination of the secondary structure of RNA molecules[665,666] and to the study of interactions with proteins. For example, treatment with dimethyl sulfate under appropriate conditions methylates bases that are not paired, giving largely 1-methyladenosine and 3-methylcytidine.[667]

## 9. Nuclear Magnetic Resonance

Much of the initial effort to study polynucleotides by NMR spectroscopy was directed toward transfer RNAs, only a few of which have been crystallized in a form suitable for X-ray diffraction. Study of the other tRNAs by NMR techniques has established that all of the tRNAs have a similar architecture and that the structures observed in the crystals are preserved in solution.[668] Figure 5-51 shows the low-field end of the NMR spectrum of a valine-specific tRNA from *E. coli.* The spectrum is run in $H_2O$ rather than $D_2O$ so that exchangeable hydrogens in the hydrogen bonds of the Watson–Crick base pairs can be observed.[669] The protons giving rise to the downfield resonances are primarily those attached to nitrogen atoms of the rings and in hydrogen-bonded positions. These protons are shielded by adjacent electron-donating groups and by their attachment to the semiaromatic rings of the bases. The NMR signals are further shifted downfield to varying degrees depending upon whether or not the proton being observed is attached to a base that is stacked with other bases. The size of the shift also depends upon which neighboring bases are present. The proton on N-3 of AU base pairs is deshielded more than the proton on the N-1 of GC base pairs. Therefore, the AU protons appear further downfield than the GC protons. The stronger ring current in A than in C enhances this separation.

All of the resonances in Fig. 5-51 have been assigned to particular bases. This was done in part by varying the temperature, changing the magnesium ion concentration, and predicting shifts caused by ring currents in adjacent bases making use of the X-ray crystal structures. NMR spectra of hairpin helical fragments also provided essential information. From integration of the areas under the peaks it was concluded that the 20 resonances seen below −11 ppm represent 27 protons. Twenty of these are in Watson-Crick base pairs and correspond to those expected from the X-ray structure. Six more belong to protons involved in tertiary interactions, such as base pair triplets or non-Watson–Crick pairs. One of these, labeled "G" in the figure, is in the dihydrouridine stem and involves the ring proton of N-1 of $m^7$G46 which is hydrogen bonded to N-7 of G22 in the major groove of the RNA. G22 is located at the beginning of the "extra loop."

Measurements of the NOE of nearby protons in both small RNA molecules[668,669] and DNA oligonucleotides[670,671] provided much additional information. Figures 5-51B and C show NOESY spectra of the $tRNA^{Val}$ and the way in which weak cross-peaks between the H-bonded imino protons in adjacent base pairs (Fig. 5-51B) can be used to establish connectivities.[669] Beginning with resonance B, it is possible to establish the sequence of the NH groups giving rise to these resonances as OBUGJNT. Using other data as well, it was concluded that these represent the seven base pairs of the acceptor stem (see Fig. 5-30), with resonance C representing the first GC pair, resonance B the second, etc. Resonance A was identified as coming from the base triplet containing a Hoogsteen

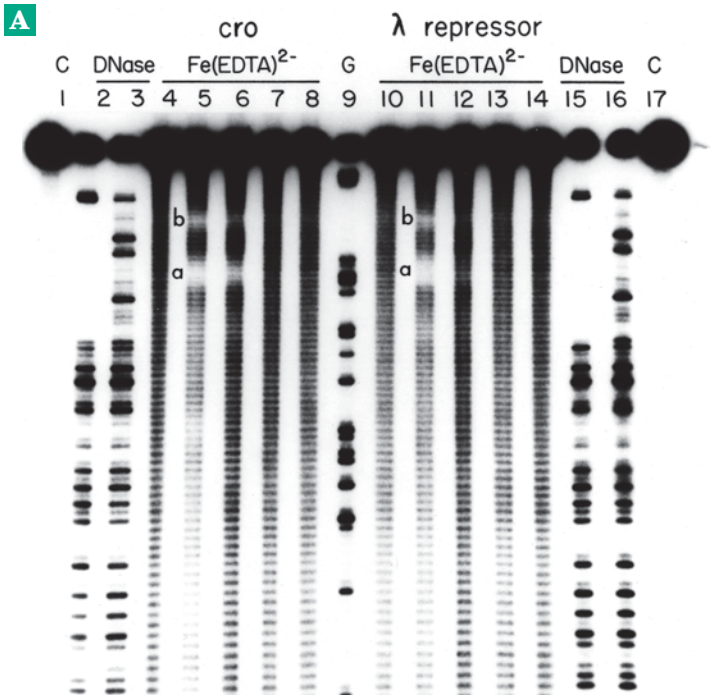
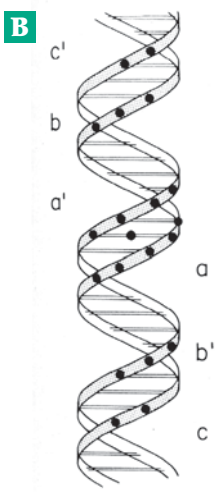
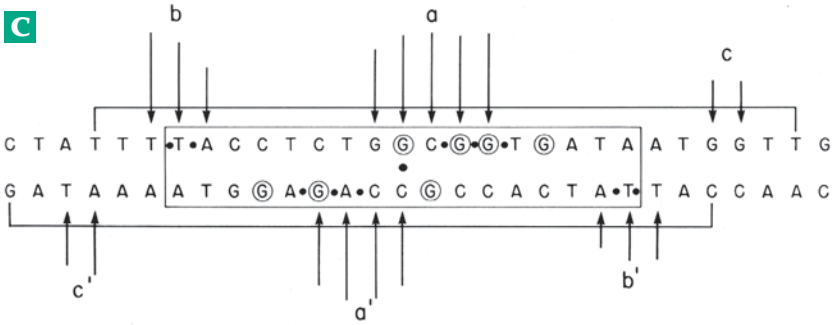



**Figure 5-50** (A) Hydroxyl radical footprints of phage λ repressor on both strands of the 120-bp restriction fragment of *E. coli* DNA carrying the $O_R1$ operator sequence. "Bottom" and "top" refer to the DNA strands as drawn in (C), and as diagrammed in (B). Labeling of the 5' end of the *Bgl* II restriction fragment afforded data for the bottom strand and labeling of the 3' *Bgl* II end afforded data for the top strand. Lanes 1 and 12, untreated DNA, labeled on the bottom and top strands, respectively; lanes 2 and 10, products of DNase I digestion of DNA labeled on the bottom and top strands, respectively, with no repressor present; lanes 3 and 11, products of DNase I digestion of DNA labeled on the bottom and top strands, respectively, complexed with λ repressor (675 nM); lanes 4 and 9, products of Maxam–Gilbert guanine-specific sequencing reactions performed on DNA labeled on the bottom and top strands, respectively; lanes 5 and 6, products of hydroxyl radical cutting of DNA labeled on the bottom strand, complexed with λ repressor. Lane 5, 90-nM λ repressor and lane 6, 675-nM λ repressor. Lanes 7 and 8, products of hydroxyl radical cutting of DNA labeled on the top strand, complexed with λ repressor. Lane 7, 675 nM λ repressor. Lane 8, 90 nM λ repressor. The labels a, b, c, a′, b′, and c′ mark the hydroxyl radical footprints. Courtesy of Thomas Tullius.[661]




base pair of 4-thiouracil at position 8 with A14 (see Fig. 5-7). Its connectivity to the sequence KCEO is also outlined in Fig. 5-51B. However, it could not be established without additional data which also helped to identify the sequence O-K as residues 10–13 of the dihydrouracil stem. Peak O is a multiproton peak representing not only GC 10 but also UA 7. In general, the GC protons are at the higher field side of the spectrum and the AU protons at the lower side. However, AU 7 is shifted to an anomolously high position. Cross-peaks between imino protons of uracil and the nearby C2 protons of adenine in Watson–Crick AU base pairs or C8 protons of Hoogsteen AU pairs can be observed in the 6.5 – 9 ppm region as shown in Fig. 5-35B. This region also contains information about other protons bound to the nuclei acid bases.

Similar techniques are being used for the study of DNA.[672] The presence of a second hydrogen in the 2' position of the deoxyribose rings of DNA adds several H-H distances (Fig. 5-52) that can be measured in addition to those seen in RNAs. Characteristic differences are seen in the NOESY plots of A, B, and Z forms of DNA.[670,671,673,674] Although detailed structural information has been obtained for short segments of DNA, spectra of larger oligonucleotides are impossible to analyze with two-dimensional methods because of extensive overlap of resonances.[665] The difficulty is already apparent in the 17 base pair DNA segment for which a one-dimensional spectrum as well as COSY and NOESY spectra are shown in Fig. 5-53.

Some help with the complexity can be obtained by incorporation of $^{13}$C-enriched methyl groups into

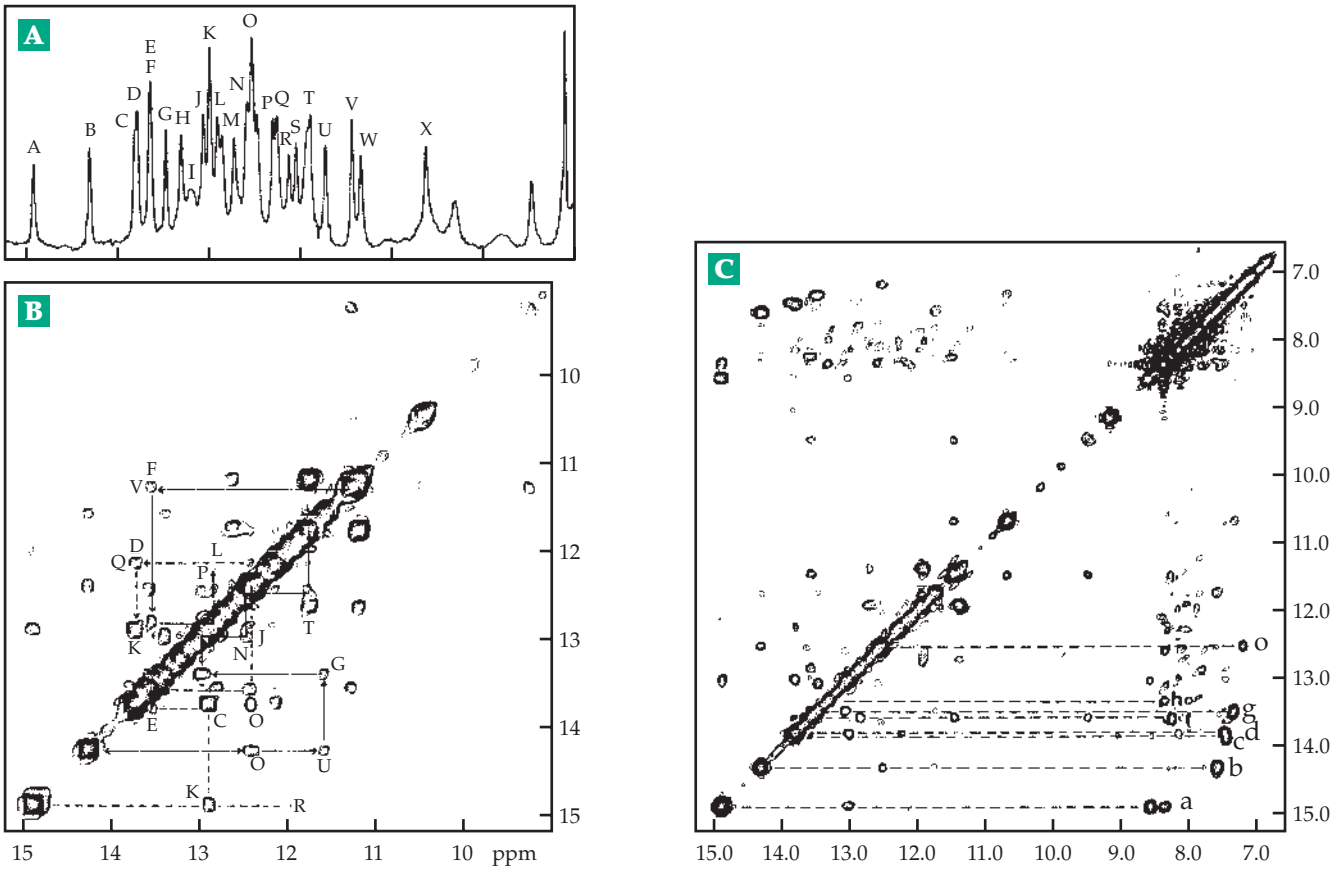



**Figure 5-51** (A) The low-field region of the one-dimensional $^1$H NMR spectrum of *E. coli* $tRNA_1^{Val}$ at 27°C in $H_2O$. Resonances are identified by letters A – X. (B) NOESY spectrum of the same tRNA under similar conditions showing the imino-imino NOEs. In the lower right sector the connectivity traces of the acceptor helix and dihydrouridine helix are shown as solid and dotted lines, respectively. In the NOESY sample the two protons in peak EF are partially resolved whereas the two protons in peak T have coalesced. (C) NOESY spectrum of *E. coli* $tRNA_1^{Val}$ at 32°C showing the imino and aromatic proton regions. AU-type imino protons have been connected horizontally by a dotted line to the cross-peak of their proximal C2-H or C8-H in the 7 to 9 ppm region, which has been labeled with the corresponding lower-case letter. From Hare *et al.*[669] Courtesy of Brian Reid.

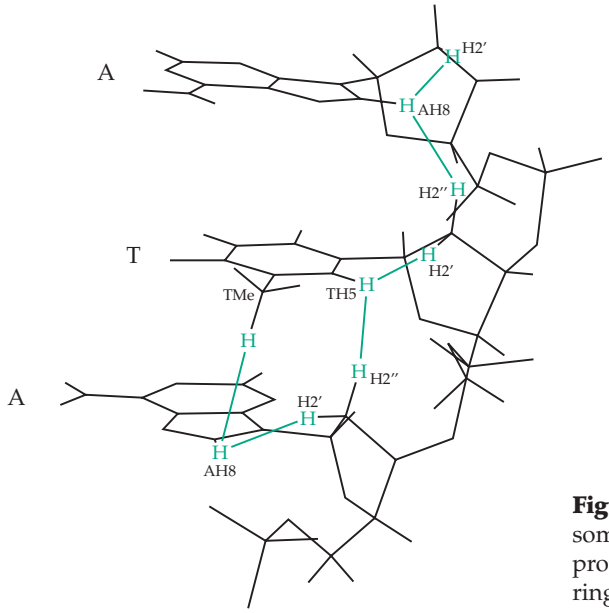



**Figure 5-52** Segment of a DNA chain in the B conformation illustrating some intrachain NOEs that may be observed. The close juxtaposition of proton pairs is provided by the H2'-*endo* ($^2$E) conformation of the sugar rings with *anti* base conformation (Fig. 5-11). After Cohen.[675]

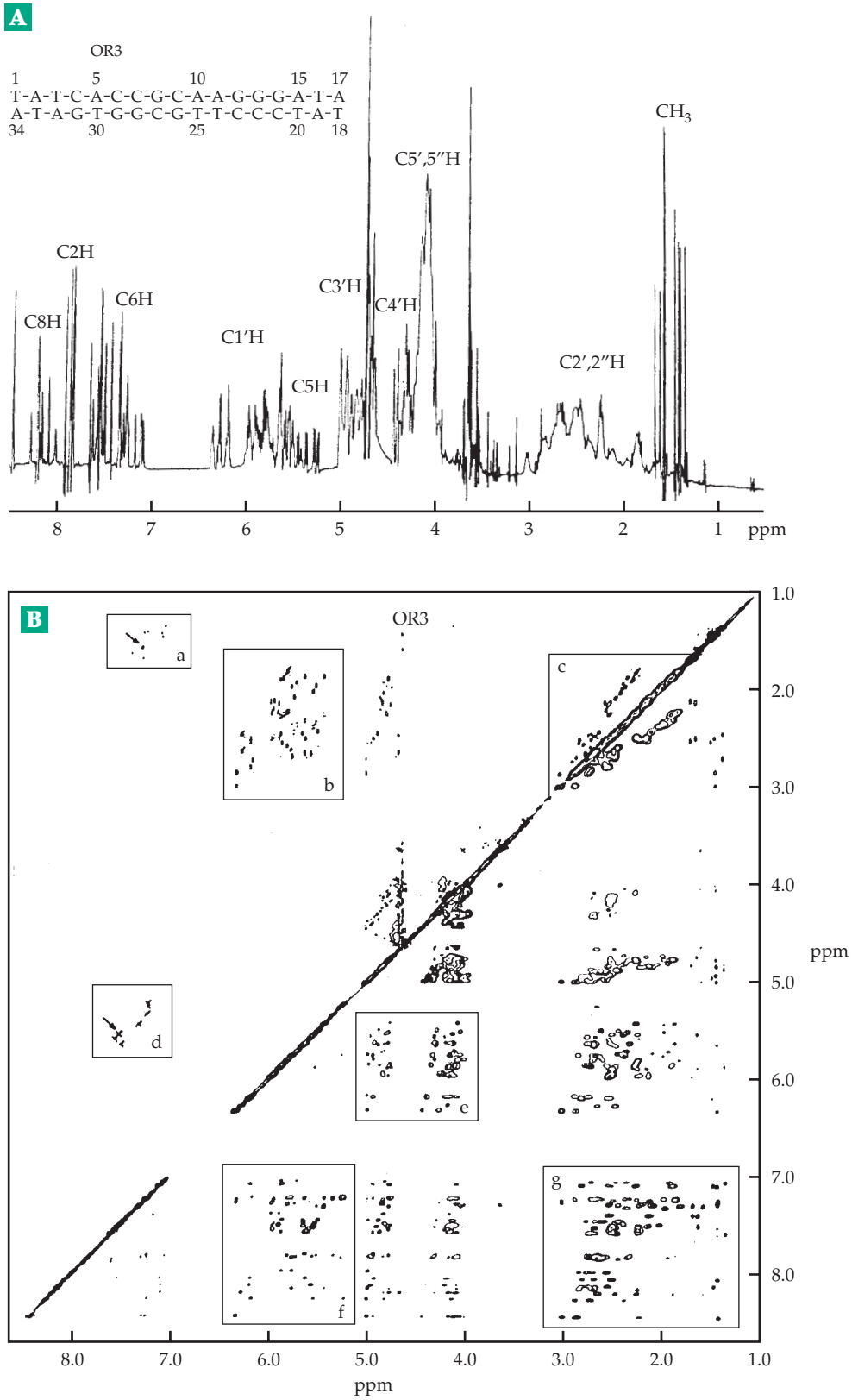



**Figure 5-53** (A) $^1$H NMR spectrum of a 17 base-pair DNA segment from the operator sequence OR3 from bacteriophage λ in $D_2O$ at 37°C. (B) Combined COSY above the diagonal and NOESY (below the diagonal) spectra. C5H and C6H J coupling is established from cross-peaks in box d for cytosines and in box a for thymines. Two unresolved cross-peaks give rise to the more intense spots marked by arrows. Box b contains cross-peaks from scalar coupling of the two H2' protons to the H1' protons of the deoxyribose rings. Most of the aromatic proton resonances could be assigned using the NOE cross-peaks in box f. For further details see Wemmer *et al.*[676] See also Bax and Lerner.[672] Courtesy of B. Reid.

polynucleotides. One way to do this is to grow yeast on a medium containing methionine with $^{13}C$ in its methyl group. The most intense peaks in $^{13}C$ NMR spectra of tRNA molecules from this yeast will represent methyl groups in modified bases in the tRNA. Incorporated $^{13}C$ also permits study of internal motion within tRNA or other oligonucleotides by NMR methods.[677] However, as with protein NMR spectroscopy the major recent advances have come from systematic incorporation of both $^{13}C$ and $^{15}N$ into nucleic acids and the development of three- and four-dimensional NMR methods.[678–684b] Also important are methods for replacing some hydrogen atoms with deuterium to simplify spectra.[685,686]

The very sensitive $^{19}F$ nucleus can be introduced into tRNAs by incorporation of 5-fluorouracil in place of uracil[687] (Fig. 5-54A,B). Phosphorus 31 NMR spectra (Fig. 5-54C) can provide information about conformations of the chain.[688]

A few recent NMR investigations of polynucleotides include studies of triple-helical DNA,[689] Holliday junctions,[290] double-stranded oligonucleotides containing adducts of carcinogens,[690,691] of hairpin loops with sheared A•A and G•G pairs,[692] and of proton exchange in both imino and amino groups.[693]

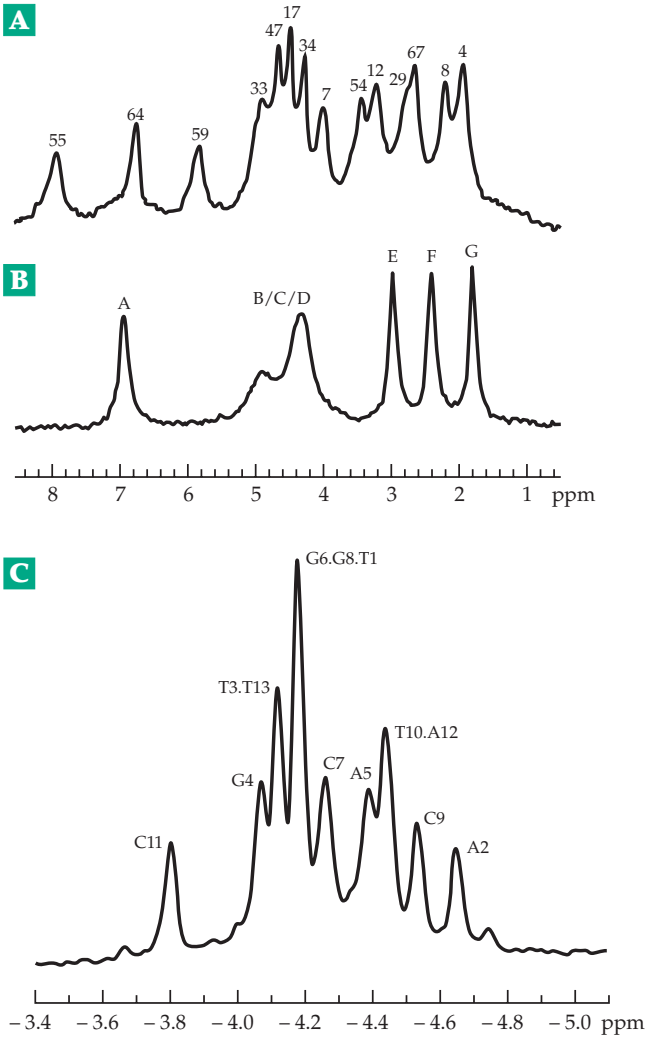



**Figure 5-54** (A) An $^{19}F$ NMR spectrum of the 76-residue *E. coli* $tRNA^{Val}$ containing 5-fluorouracil in 14 positions. Recorded at 47°C. The numbers above the resonances indicate the position in the sequence. (The sequence is not identical to that for the yeast tRNA shown in Fig. 5-30.) Modified from Chu *et al.*[694] Courtesy of Jack Horowitz. (B) A similar spectrum for a 35-residue "minihelix" that contains the acceptor stem of the $tRNA^{Val}$ and seven fluorouracils. The broad peaks B, D, and E are shifted far upfield by reaction with bisulfite (Eq. 5-11) suggesting that they are not hydrogen bonded and are present in the loop of the stem–loop structure. Peaks A, E, F, and G correspond to resonances 64, 7, 67, and 4, respectively, in (A) and represent fluorouracil in the stem structure. From Chu *et al.*[694] Courtesy of Jack Horowitz. (C) A $^{31}P$ NMR spectrum of a synthetic 14 base-pair DNA segment related to the *E. coli lac* operator. The palindromic sequence is TCTGAGCGCTCAGA. The numbers refer to the positions from the 5' end. From Schroeder *et al.*[688]

## References


1. Portugal, F. H., and Cohen, J. S. (1977) *A Century of DNA: A History of the Discovery of the Structure and Function of the Genetic Substance*, MIT Press, Cambridge, Massachusetts
2. Levene, P. A., and Tipson, R. S. (1935) *J. Biol. Chem.* **109**, 623–630
3. Schlenk, F. (1988) *Trends Biochem. Sci.* **13**, 67–69
4. Watson, J. D., and Crick, F. H. C. (1953) *Nature (London)* **171**, 737–738
5. Watson, J. D. (1968) *The Double Helix*, Atheneum, New York
6. Crick, F. (1988) *What a Mad Pursuit: A Personal View of Scientific Discovery*, Basic Books, New York
7. Olby, R. (1974) *The Path to the Double Helix*, Macmillan, London
8. Wilson, H. R. (1988) *Trends Biochem. Sci.* **13**, 275–278
9. Chambers, D. A., Reid, K. B. M., and Cohen, R. L. (1994) *FASEB J.* **8**, 1219–1226

9a. Piper, A. (1998) *Trends Biochem. Sci.* **23**, 151–154

10. Arnott, S., and Hukins, D. W. L. (1973) *J. Mol. Biol.* **81**, 93–105
11. Proudfoot, N. J., Shander, M. H. M., Manley, J. L., Gefter, M. L., and Maniatis, T. (1980) *Science* **209**, 1329–1336
12. Tucker, P. W., Liu, C.-P., Mushinski, J. F., and Blattner, F. R. (1980) *Science* **209**, 1353–1360
13. Yamada, Y., Mudryj, M., Sullivan, M., and de Crombrugghe, B. (1983) *J. Biol. Chem.* **258**, 2758–2761
14. Tsuzuki, T., Obaru, K., Setoyama, C., and Shimada, K. (1987) *J. Mol. Biol.* **198**, 21–31
15. Ratner, L., and 18 other authors (1985) *Nature (London)* **313**, 277–284
16. Dunn, J. J., and Studier, F. W. (1983) *J. Mol. Biol.* **166**, 477– 535
17. Peden, K. W. C. (1983) *Gene* **22**, 277–280

## References


18. Anderson, S., Bankier, A. T., Barrell, B. G., deBruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R., and Young, I. G. (1981) *Nature (London)* **290**, 457–470
19. Fraser, C. M., and 28 other authors (1995) *Science* **270**, 397–403
20. Fleischmann, R. D., and 39 other authors (1995) *Science* **269**, 496–512
20a. Doolittle, R. F. (1998) *Nature (London)* **392**, 339–342
21. Nowak, R. (1995) *Science* **267**, 172–174
22. Dujon, B., and 107 other authors (1994) *Nature (London)* **369**, 371–378
23. Johnston, M., and 34 other authors (1994) *Science* **265**, 2077–2082
24. Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. (1996) *Science* **274**, 546–557
24a. Macilwain, C. (2000) *Nature (London)* **405**, 983–984
25. Liébecq, C., ed. (1992) *Biochemical Nomenclature*, Portland Press, London and Chapel Hill, North Carolina (for the International Union of Biochemistry and Molecular Biology)
26. Shabarova, Z., and Bogdanov, A. (1994) *Advanced Organic Chemistry of Nucleic Acids*, VCH Publ., Weinheim
27. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer, New York
28. Hurst, D. T. (1980) *An Introduction to the Chemistry and Biochemistry of Pyrimidines, Purines and Pteridines*, Wiley, New York
29. Markowski, V., Sullivan, G. R., and Roberts, J. D. (1977) *J. Am. Chem. Soc.* **99**, 714–718
30. Katritzky, A. R., and Karelson, M. (1991) *J. Am. Chem. Soc.* **113**, 1561–1566
31. Szczesniak, M., Leszczynski, J., and Person, W. B. (1992) *J. Am. Chem. Soc.* **114**, 2731–2733
32. Scanlan, M. J., and Hillier, I. H. (1984) *J. Am. Chem. Soc.* **106**, 3737–3745
33. Brown, D. M. (1974) in *Basic Principles in Nucleic Acid Chemistry*, Vol. II (Tso, P. O. P., ed), pp. 2–90, Academic Press, New York
34. Hoogsteen, K. (1963) *Acta Crystallogr.* **16**, 907–916
35. Frank-Kamenetskii, M. D., and Mirkin, S. M. (1995) *Ann. Rev. Biochem.* **64**, 65–95
36. Portugal, J. (1989) *Trends Biochem. Sci.* **14**, 127–130
37. Adams, R. L. P., Knowler, J. T., and Leader, D. P. (1992) *The Biochemistry of the Nucleic Acids*, 11th ed., Chapman & Hall, London
37a. Saenger, W. (1973) *Angew. Chem. Int. Ed. Engl.* **12**, 591–601
38. Fang, G., and Cech, T. R. (1993) *Biochemistry* **32**, 11646–11657
39. Gellert, M., Lipsett, M. N., and Davies, D. R. (1962) *Proc. Natl. Acad. Sci. U.S.A.* **48**, 2013–2018
40. Kennell, D. E. (1971) *Prog. Nucleic Acid Res. Mol. Biol.* **11**, 259–301
40a. Stofer, E., Chipot, C., and Lavery, R. (1999) *J. Am. Chem. Soc.* **121**, 9503–9508
40b. Peyret, N., Seneviratne, P. A., Allawi, H. T., and SantaLucia, J., Jr. (1999) *Biochemistry* **38**, 3468–3477
41. Pranata, J., Wierschke, S. G., and Jorgensen, W. L. (1991) *J. Am. Chem. Soc.* **113**, 2810–2819
42. Dang, L. X., and Kollman, P. A. (1990) *J. Am. Chem. Soc.* **112**, 503–507
43. Newcomb, L. F., and Gellman, S. H. (1994) *J. Am. Chem. Soc.* **116**, 4993–4994
44. Borer, P. N., Dengler, B., Tinoco, I., Jr., and Uhlenbeck, O. C. (1974) *J. Mol. Biol.* **86**, 843–853
45. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9373–9377
46. Sarai, A., Mazur, J., Nussinov, R., and Jernigan, R. L. (1988) *Biochemistry* **27**, 8498–8502
47. Hunter, C. A. (1993) *J. Mol. Biol.* **230**, 1025–1054
48. Friedman, R. A., and Honig, B. (1995) *Biophys. J.* **69**, 1528–1535
49. DePrisco Albergo, D., Marky, L. A., Breslauer, K. J., and Turner, D. H. (1981) *Biochemistry* **20**, 1409–1413
50. Privalov, P. L., and Filimonov, V. V. (1978) *J. Mol. Biol.* **122**, 447–464
51. Privalov, P. L., and Filimonov, V. V. (1978) *J. Mol. Biol.* **122**, 465–470
52. Petersen, S. B., and Led, J. J. (1981) *J. Am. Chem. Soc.* **103**, 5308–5313
53. Petruska, J., and Goodman, M. F. (1995) *J. Biol. Chem.* **270**, 746–750
54. SantaLucia, J., Jr., Allawi, H. T., and Seneviratne, P. A. (1996) *Biochemistry* **35**, 3555–3562
55. Egli, M., Usman, N., and Rich, A. (1993) *Biochemistry* **32**, 3221–3237
56. Watson, J. D., and Crick, F. H. C. (1953) *Nature (London)* **171**, 964–967
57. Topal, M. D., and Fresco, J. R. (1976) *Nature (London)* **263**, 285–289
58. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., and Gralla, J. (1973) *Nature New Biol.* **246**, 40–41
59. Fazakerley, G. V., Gdaniec, Z., and Sowers, L. C. (1993) *J. Mol. Biol.* **230**, 6–10
60. Florián, J., Hrouda, V., and Hobza, P. (1994) *J. Am. Chem. Soc.* **116**, 1457–1460
61. Douhal, A., Kim, S. K., and Zewail, A. H. (1995) *Nature (London)* **378**, 260–263
62. Gabb, H. A., and Harvey, S. C. (1993) *J. Am. Chem. Soc.* **115**, 4218–4227
63. Ellervik, U., and Magnusson, G. (1994) *J. Am. Chem. Soc.* **116**, 2340–2347
64. Thibaudeau, C., Plavec, J., and Chattopadhyaya, J. (1994) *J. Am. Chem. Soc.* **116**, 8033–8037
65. Santos, R. A., Tang, P., and Harbison, G. S. (1989) *Biochemistry* **28**, 9372–9378
66. Sundaralingam, M. (1969) *Biopolymers* **7**, 821–860
67. Vologodskii, A. (1992) *Topology and Physics of Circular DNA*, CRC Press, Boca Raton, Florida
68. Baikalov, I., Grzeskowiak, K., Yanagi, K., Quintana, J., and Dickerson, R. E. (1993) *J. Mol. Biol.* **231**, 768–784
69. Lipanov, A., Kopka, M. L., Kaczor-Grzeskowiak, M., Quintana, J., and Dickerson, R. E. (1993) *Biochemistry* **32**, 1373–1389
70. Schlick, T., Hingerty, B. E., Peskin, C. S., Overton, M. L., and Broyde, S. (1991) *Theoretical Biochemistry and Molecular Physics*, Vol. 1 (Beveridge, D. L. and Lavery, R., eds), pp. 39–58, Adenine Press, Guilderland, New York
71. Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., and Rich, A. (1979) *Nature (London)* **282**, 680–686
72. Kornberg, A., and Baker, T. A. (1992) *DNA Replication*, 2nd ed., Freeman, New York
73. Kaufmann, W. K., and Paules, R. S. (1996) *FASEB J.* **10**, 238–247
74. Lindahl, T. (1993) *Nature (London)* **362**, 709–715
75. Rodley, G. A., and Bates, R. H. T. (1980) *Trends Biochem. Sci.* **5**, 231–233
76. Kennard, O., and Salisbury, S. A. (1993) *J. Biol. Chem.* **268**, 10701–10704
77. Drew, H. R., and Dickerson, R. E. (1981) *J. Mol. Biol.* **151**, 535–556
78. Dickerson, R. E. (1983) *Sci. Am.* **249**(Dec), 94–111
78a. Dickerson, R. E. (1997) in *International Tables for Macromolecular Crystallography* (Rossmann, M. G., and Arnold, E., eds), (Chapter 23, Part c)
79. Shui, X., McFail-Isom, L., Hu, G. G., and Williams, L. D. (1998) *Biochemistry* **37**, 8341–8355
80. Gorin, A. A., Zhurkin, V. B., and Olson, W. K. (1995) *J. Mol. Biol.* **247**, 34–48
81. Tereshko, V., Minasov, G., and Egli, M. (1999) *J. Am. Chem. Soc.* **121**, 470–471
82. Yuan, H., Quintana, J., and Dickerson, R. E. (1992) *Biochemistry* **31**, 8009–8021
83. Ban, C., Ramakrishnan, B., and Sundaralingam, M. (1994) *J. Mol. Biol.* **236**, 275–285
83a. Kielkopf, C. L., Ding, S., Kuhn, P., and Rees, D. C. (2000) *J. Mol. Biol.* **296**, 787–801
84. Johansson, E., Parkinson, G., and Neidle, S. (2000) *J. Mol. Biol.* **300**, 551–561
85. Shen, L. X., Cai, Z., and Tinoco, I., Jr. (1995) *FASEB J.* **9**, 1023–1033
86. Rosenberg, J. M., Seeman, N. C., Day, R. O., and Rich, A. (1976) *J. Mol. Biol.* **104**, 145–167
87. Marino, J. P., Gregorian, R. S., Jr., Csankovszki, G., and Crothers, D. M. (1995) *Science* **268**, 1448–1454
88. Iwamoto, S., and Hsu, M.-T. (1983) *Nature (London)* **305**, 70–72
89. Morel, G., ed. (1995) *Visualization of Nucleic Acids*, CRC Press, Boca Raton, Florida
90. Arscott, P. G., Lee, G., Bloomfield, V. A., and Evans, D. F. (1989) *Nature (London)* **339**, 484–486
91. Jares-Erijman, E. A., and Jovin, T. M. (1996) *J. Mol. Biol.* **257**, 597–617
92. Duan, Y., Wilkosz, P., Crowley, M., and Rosenberg, J. M. (1997) *J. Mol. Biol.* **272**, 553–572
93. Auffinger, P., and Westhof, E. (2000) *J. Mol. Biol.* **300**, 1113–1131
93a. Soler-López, M., Malinina, L., Liu, J., Huynh-Dinh, T., and Subirana, J. A. (1999) *J. Biol. Chem.* **274**, 23683–23686
94. Schneider, B., and Berman, H. M. (1995) *Biophys. J.* **69**, 2661–2669
95. Tereshko, V., Minasov, G., and Egli, M. (1999) *J. Am. Chem. Soc.* **121**, 3590–3595
96. Franklin, R. E., and Gosling, R. G. (1953) *Nature (London)* **172**, 156–157
97. Ramakrishnan, B., and Sundaralingam, M. (1995) *Biophys. J.* **69**, 553–558
98. Ramakrishnan, B., and Sundaralingam, M. (1993) *J. Mol. Biol.* **231**, 431–444
99. Eisenstein, M., and Shakked, Z. (1995) *J. Mol. Biol.* **248**, 662–678
99a. Ng, H.-L., Kopka, M. L., and Dickerson, R. E. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2035–2039
100. Selsing, E., Wells, R. D., Alden, C. J., and Arnott, S. (1979) *J. Biol. Chem.* **254**, 5417–5422
101. Mahendrasingam, A., Forsyth, V. T., Hussain, R., Greenall, R. J., Pigram, W. J., and Fuller, W. (1986) *Science* **233**, 195–197
102. Hopkins, R. C. (1981) *Science* **211**, 289–291
103. Gupta, G., Bansal, M., and Sasisekharan, V. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6486–6490
104. Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W., and Ratliff, R. L. (1980) *Nature (London)* **283**, 743–745
105. Herbert, A., and Rich, A. (1996) *J. Biol. Chem.* **271**, 11595–11598
106. Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A., and Wang, A. H.-J. (1989) *J. Biol. Chem.* **264**, 7921–7935
107. Zhou, G., and Ho, P. S. (1990) *Biochemistry* **29**, 7229–7236

## References


108. Parkinson, G. N., Arvanitis, G. M., Lessinger, L., Ginell, S. L., Jones, R., Gaffney, B., and Berman, H. M. (1995) *Biochemistry* **34**, 15487–15495
109. Sadasivan, C., and Gautham, N. (1995) *J. Mol. Biol.* **248**, 918– 930
110. Bancroft, D., Williams, L. D., Rich, A., and Egli, M. (1994) *Biochemistry* **33**, 1073–1086
111. Rich, A., Nordheim, A., and Wang, A. H.-J. (1984) *Ann. Rev. Biochem.* **53**, 791–846
112. Moore, M. H., Van Meervelt, L., Salisbury, S. A., Lin, P. K. T., and Brown, D. M. (1995) *J. Mol. Biol.* **251**, 665–673
113. Wang, Y., Thomas, G. A., and Peticolas, W. L. (1987) *Biochemistry* **26**, 5178–5186
114. Saenger, W., Hunter, W. N., and Kennard, O. (1986) *Nature (London)* **324**, 385–388
115. Frank-Kamenetskii, M. (1986) *Nature (London)* **324**, 305
116. Basham, B., Schroth, G. P., and Ho, P. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6464–6468
117. Misra, V. K., and Honig, B. (1996) *Biochemistry* **35**, 1115–1124
118. McLean, M. J., Blaho, J. A., Kilpatrick, M. W., and Wells, R. D. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5884–5888
119. Kim, J.-m, and DasSarma, S. (1996) *J. Biol. Chem.* **271**, 19724–19731
120. Nordheim, A., Pardue, M. L., Lafer, E. M., Möller, A., Stollar, B. D., and Rich, A. (1981) *Nature (London)* **294**, 417–422
121. Jaworski, A., Hsieh, W.-T., Blaho, J. A., Larson, J. E., and Wells, R. D. (1987) *Science* **238**, 773–777
122. Zacharias, W., O'Conner, T. R., and Larson, J. E. (1988) *Biochemistry* **27**, 2970–2978
123. Nordheim, A., and Rich, A. (1983) *Nature (London)* **303**, 674–679
124. Wells, R. D. (1988) *J. Biol. Chem.* **263**, 1095–1098
125. Rentzeperis, D., Kupke, D. W., and Marky, L. A. (1994) *Biochemistry* **33**, 9588–9591
126. Liu, K., Miles, H. T., Frazier, J., and Sasisekharan, V. (1993) *Biochemistry* **32**, 11802–11809
127. Zhou, N., Germann, M. W., van de Sande, J. H., Pattabiraman, N., and Vogel, H. J. (1993) *Biochemistry* **32**, 646–656
128. Otto, C., Thomas, G. A., Rippe, K., Jovin, T. M., and Peticolas, W. L. (1991) *Biochemistry* **30**, 3062–3069
129. Trifonov, E. N. (1991) *Trends Biochem. Sci.* **16**, 467–470
130. Rich, A. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13999–14000
131. Hageman, P. J. (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 265–286
132. Poncin, M., Hartmann, B., and Lavery, R. (1992) *J. Mol. Biol.* **226**, 775–794
133. MacKerell, A. D., Jr., Wiórkiewicz-Kuczera, J., and Karplus, M. (1995) *J. Am. Chem. Soc.* **117**, 11946–11975
134. Kosikov, K. M., Gorin, A. A., Zhurkin, V. B., and Olson, W. K. (1999) *J. Mol. Biol.* **289**, 1301–1326
135. Peticolas, W. L., Wang, Y., and Thomas, G. A. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2579–2583
136. Dickerson, R. E., and Drew, H. R. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7318–7322
136a. Packer, M. J., and Hunter, C. A. (1998) *J. Mol. Biol.* **280**, 407–420
137. Hogan, M., Dattagupta, N., and Crothers, D. M. (1979) *Nature (London)* **278**, 521–524
138. Mitra, C. K., Sarma, M. H., and Sarma, R. H. (1981) *J. Am. Chem. Soc.* **103**, 6727–6737
139. Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J.-L., Chatenay, D., and Caron, F. (1996) *Science* **271**, 792–794
140. Smith, S. B., Cui, Y., and Bustamante, C. (1996) *Science* **271**, 795–799
141. Strick, T. R., Allemand, J. F., Bensimon, D., Bensimon, A., and Croquette, V. (1996) *Science* **271**, 1835–1837
141a. Allemand, J. F., Bensimon, D., Lavery, R., and Croquette, V. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14152–14157
142. Dickerson, R. E. (1983) *J. Mol. Biol.* **166**, 419–441
143. Tung, C.-S., and Harvey, S. C. (1986) *J. Biol. Chem.* **261**, 3700–3709
144. Ulanovsky, L. E., and Trifonov, E. N. (1987) *Nature (London)* **326**, 720–722
145. Diekmann, S. (1989) *EMBO J.* **8**, 1–4
146. El Hassan, M. A., and Calladine, C. R. (1996) *J. Mol. Biol.* **259**, 95–103
147. Calladine, C. R. (1982) *J. Mol. Biol.* **161**, 343–352
148. Hunter, C. A., and Lu, X.-J. (1997) *J. Mol. Biol.* **265**, 603–619
149. Babcock, M. S., Pednault, E. P. D., and Olson, W. K. (1994) *J. Mol. Biol.* **273**, 125–156
150. Lu, X.-J., El Hassan, M. A., and Hunter, C. A. (1997) *J. Mol. Biol.* **273**, 668–680
151. Kitchin, P. A., Klein, V. A., Ryan, K. A., Gann, K. L., Rauch, C. A., Kang, D. S., Wells, R. D., and England, P. T. (1986) *J. Biol. Chem.* **261**, 11302–11309
152. Koo, H.-S., Wu, H.-M., and Crothers, D. M. (1986) *Nature (London)* **320**, 501–506
153. Zahn, K., and Blattner, F. B. (1987) *Science* **236**, 416–422
154. Snyder, M., Buchman, A. R., and Davis, R. W. (1986) *Nature (London)* **324**, 87–89
155. Travers, A. A. (1987) *Trends Biochem. Sci.* **12**, 108–112
156. Grzeskowiak, K., Goodsell, D. S., Kaczor-Grzeskowiak, M., Cascio, D., and Dickerson, R. E. (1993) *Biochemistry* **32**, 8923–8931
157. Allewell, N. (1988) *Trends Biochem. Sci.* **13**, 193–195
158. Haran, T. E., Kahn, J. D., and Crothers, D. M. (1994) *J. Mol. Biol.* **244**, 135–143
159. Strauss, J. K., and Maher, L. J., III. (1994) *Science* **266**, 1829–1834
160. Olson, W. K., Marky, N. L., Jernigan, R. L., and Zhurkin, V. B. (1993) *J. Mol. Biol.* **232**, 530–554
161. Price, M. A., and Tullius, T. D. (1993) *Biochemistry* **32**, 127–136
162. DiGabriele, A. D., and Steitz, T. A. (1993) *J. Mol. Biol.* **231**, 1024–1039
163. Chuprina, V. P., Fedoroff, O. Y., and Reid, B. R. (1991) *Biochemistry* **30**, 561–568
164. Crothers, D. M. (1994) *Science* **266**, 1819–1820
165. Husain, I., Griffith, J., and Sancar, A. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2558–2562
166. Srinivasan, A. R., Torres, R., Clark, W., and Olson, W. K. (1987) *J. Biomol. Struct. Dyn.* **5**, 459–496
167. Lutter, L. C., Halvorson, H. R., and Calladine, C. R. (1996) *J. Mol. Biol.* **261**, 620–633
168. Privé, G. G., Heinemann, V., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., and Dickerson, R. E. (1987) *Science* **238**, 498–504
169. Lilley, D. M. J. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7140–7142
170. Woodson, S. A., and Crothers, D. M. (1988) *Biochemistry* **27**, 3130–3141
171. Miller, M., Harrison, R. W., Wlodawer, A., Appella, E., and Sussman, J. L. (1988) *Nature (London)* **334**, 85–86
172. Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D., and Kennard, O. (1987) *J. Biol. Chem.* **262**, 9962–9970
173. Lane, A., Martin, S. R., Ebel, S., and Brown, T. (1992) *Biochemistry* **31**, 12087–12095
174. Li, Y., and Agrawal, S. (1995) *Biochemistry* **34**, 10056–10062
175. Manning, G. S. (1978) *Q. Rev. Biophys.* **11**, 179–246
176. Mascotti, D. P., and Lohman, T. M. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3142–3146
177. Li, A. Z., Huang, H., Re, X., Qi, L. J., and Marx, K. A. (1998) *Biophys. J.* **74**, 964–973
178. Frydman, L., Rossomando, P. C., Frydman, V., Fernandez, C. O., Frydman, B., and Samejima, K. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9186–9190
179. Jain, S., Zon, G., and Sundaralingam, M. (1989) *Biochemistry* **28**, 2360–2364
180. Pelta, J., Durand, D., Doucet, J., and Livolant, F. (1996) *Biophys. J.* **71**, 48–63
181. Rajski, S. R., Kumar, S., Roberts, R. J., and Barton, J. K. (1999) *J. Am. Chem. Soc.* **121**, 5615–5616
181a. Aich, P., Labiuk, S. L., Tari, L. W., Dalbaere, L. J. T., Roesler, W. J., Falk, K. J., Steer, R. P., and Lee, J. S. (1999) *J. Mol. Biol.* **294**, 477–485
182. Fink, H.-W., and Schönenberger, C. (1999) *Nature (London)* **398**, 407–410
183. Marini, J. C., Miller, K. G., and Englund, P. T. (1980) *J. Biol. Chem.* **255**, 4976–4979
184. Wasserman, S. A., Dungan, J. M., and Cozzarelli, N. R. (1985) *Science* **229**, 171–174
185. Du, S. M., Wang, H., Tse-Dinh, Y.-C., and Seeman, N. C. (1995) *Biochemistry* **34**, 673–682
186. Du, S. M., Stollar, B. D., and Seeman, N. C. (1995) *J. Am. Chem. Soc.* **117**, 1194–1200
187. Wasserman, S. A., and Cozzarelli, N. R. (1986) *Science* **232**, 951–960
188. White, J. H., and Cozzarelli, N. R. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3322–3326
189. Cipra, B. (1992) *Science* **255**, 403
190. Adams, C. C. (1994) *The Knot Book*, Freeman, New York
191. Krasnow, M. A., Stasiak, A., Spengler, S. J., Dean, F., Koller, T., and Cozzarelli, N. R. (1983) *Nature (London)* **304**, 559–560
192. Du, S. M., and Seeman, N. C. (1992) *J. Am. Chem. Soc.* **114**, 9652–9655
193. Wang, J. C. (1980) *Trends Biochem. Sci.* **5**, 219–221
194. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry, Part 3*, Freeman, San Francisco, California (pp. 1265–1290)
195. Lebowitz, J. (1990) *Trends Biochem. Sci.* **15**, 202–207
196. Bates, A. D., and Maxwell, A. (1993) *DNA Topology*, Oxford Univ. Press, London
197. Yang, Y., Westcott, T. P., Pedersen, S. C., Tobias, I., and Olson, W. K. (1995) *Trends Biochem. Sci.* **20**, 313–319
198. Vinograd, J., Lebowitz, J., and Watson, R. (1968) *J. Mol. Biol.* **33**, 173–197
199. White, J. H., Cozzarelli, N. R., and Bauer, W. R. (1988) *Science* **241**, 323–327
199a. Vologodskii, A. V., and Cozzarelli, N. R. (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23**, 609–643
199b. Cozzarelli, N. R., Boles, T. C., and White, J. H. (1990) in *DNA Topology and its Biological Effects*, pp. 139–184, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
199c. Podtelezhnikov, A. A., Cozzarelli, N. R., and Vologodskii, A. V. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12974–12979
200. Reich, Z., Levin-Zaidman, S., Gutman, S. B., Arad, T., and Minsky, A. (1994) *Biochemistry* **33**, 14177–14184
201. Eriksson, M. A. L., Härd, T., and Nilsson, L. (1995) *Biophys. J.* **68**, 402–426
202. Keller, W. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2550–2554
203. Kornberg, R. D., and Klug, A. (1981) *Sci. Am.* **244**(Feb), 52–64
204. Wang, B.-C., Rose, J., Arents, G., and Moudrianakis, E. N. (1994) *J. Mol. Biol.* **236**, 179–188

# References


205. Malik, M., Bensaid, A., Rouviere-Yaniv, J., and Drlica, K. (1996) *J. Mol. Biol.* **256**, 66–76
206. Serrano, M., Salas, M., and Hermoso, J. M. (1993) *Trends Biochem. Sci.* **18**, 202–206
207. Seidl, A., and Hinz, H.-J. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1312–1316
208. Rahmouni, A. R., and Wells, R. D. (1989) *Science* **246**, 358–363
209. Spolar, R. S., and Record, M. T., Jr. (1994) *Science* **263**, 777–784
210. White, J. H., Gallo, R. M., and Bauer, W. R. (1992) *Trends Biochem. Sci.* **17**, 7–12
211. van Holde, K., and Zlatanova, J. (1995) *J. Biol. Chem.* **270**, 8373–8376
212. Keller, W. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4876–4880
213. Hamiche, A., Schultz, P., Ramakrishnan, V., Oudet, P., and Prunell, A. (1996) *J. Mol. Biol.* **257**, 30–42
214. Alfonso, P. J., Crippa, M. P., Hayes, J. J., and Bustin, M. (1994) *J. Mol. Biol.* **236**, 189–198
215. Bednar, J., Furrer, P., Stasiak, A., Dubochet, J., Egelman, E. H., and Bates, A. D. (1994) *J. Mol. Biol.* **235**, 825–847
216. Kumar, A. (1995) *Biochemistry* **34**, 12921–12925
217. Stigter, D. (1995) *Biophys. J.* **69**, 380–388
218. Baeza, I., Gariglio, P., Rangel, L. M., Chavez, P., Cervantes, L., Arguello, C., Wong, C., and Montañez, C. (1987) *Biochemistry* **26**, 6387–6392
219. Schmid, M. B. (1988) *Trends Biochem. Sci.* **13**, 131–135
220. Quigley, G. J., Wang, A. H.-J., Ughetto, G., van der Marel, G., van Boom, J. H., and Rich, A. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7204–7208
221. Campbell, A. M., and Jolly, D. J. (1973) *Biochem. J.* **133**, 209–226
222. Depew, R. E., and Wang, J. C. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4275–4279
223. Westof, E., and Sundaralingam, M. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1852–1856
224. Sobell, H. M. (1974) *Sci. Am.* **231**(Aug), 82–91
225. Chen, H., and Patel, D. J. (1995) *J. Am. Chem. Soc.* **117**, 5901–5913
226. Chaires, J. B., Satyanarayana, S., Suh, D., Fokt, I., Przewloka, T., and Priebe, W. (1996) *Biochemistry* **35**, 2047–2053
227. Rajeswari, M. R., Bose, H. S., Kukreti, S., Gupta, A., Chauhan, V. S., and Roy, K. B. (1992) *Biochemistry* **31**, 6237–6241
227a. Leitner, D., Schröder, W., and Weisz, K. (2000) *Biochemistry* **39**, 5886–5892
228. Felsenfeld, G., Davies, D., and Rich, A. (1957) *J. Am. Chem. Soc.* **79**, 2023–2024
229. Felsenfeld, G., and Miles, H. T. (1967) *Ann. Rev. Biochem.* **36**, 407–448
230. Arnott, S., and Selsing, E. (1974) *J. Mol. Biol.* **88**, 509–521
231. Wang, E., Koshlap, K. M., Gillespie, P., Dervan, P. B., and Feigon, J. (1996) *J. Mol. Biol.* **257**, 1052–1069
232. Radhakrishnan, I., and Patel, D. J. (1994) *Biochemistry* **33**, 11405–11416
233. Radhakrishnan, I., and Patel, D. J. (1994) *J. Mol. Biol.* **241**, 600–619
234. Rhee, S., Han, Z.-j, Liu, K., Miles, H. T., and Davies, D. R. (1999) *Biochemistry* **38**, 16810–16815
235. Roberts, R. W., and Crothers, D. M. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4320–4325
236. van Dongen, M. J. P., Heus, H. A., Wymenga, S. S., van der Marel, G. A., van Boom, J. H., and Hilbers, C. W. (1996) *Biochemistry* **35**, 1733–1739
237. Klysik, J. (1995) *J. Mol. Biol.* **245**, 499–507
238. Srinivasan, A. R., and Olson, W. K. (1998) *J. Am. Chem. Soc.* **120**, 484–491
239. Hanvey, J. C., Shimizu, M., and Wells, R. D. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6292–6296
240. Htun, H., and Dahlberg, J. E. (1989) *Science* **243**, 1571–1576
241. Mirkin, S. M., and Frank-Kamenetskii, M. D. (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23**, 541–576
242. Singh, S., Patel, P. K., and Hosur, R. V. (1997) *Biochemistry* **36**, 13214–13222
243. Van Meervelt, L., Vlieghe, D., Dautant, A., Gallois, B., Précigoux, G., and Kennard, O. (1995) *Nature (London)* **374**, 742–744
244. Reddy, G., Jwang, B., Rao, B. J., and Radding, C. M. (1994) *Biochemistry* **33**, 11486–11492
245. Kim, M. G., Zhurkin, V. B., Jernigan, R. L., and Camerini-Otero, R. D. (1995) *J. Mol. Biol.* **247**, 874–889
246. Cox, M. M. (1995) *J. Biol. Chem.* **270**, 26021–26024
247. Jain, S. K., Cox, M. M., and Inman, R. B. (1995) *J. Biol. Chem.* **270**, 4943–4949
248. Best, G. C., and Dervan, P. B. (1995) *J. Am. Chem. Soc.* **117**, 1187–1193
249. Greenberg, W. A., and Dervan, P. B. (1995) *J. Am. Chem. Soc.* **117**, 5016–5022
250. Hashem, G. M., Pham, L. P., Vaughan, M. R., and Gray, D. M. (1998) *Biochemistry* **37**, 61–72
251. Gryaznov, S. M., Lloyd, D. H., Chen, J.-K., Schultz, R. G., DeDionisio, L. A., Ratmeyer, L., and Wilson, W. D. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5798–5802
252. Betts, L., Josey, J. A., Veal, J. M., and Jordan, S. R. (1995) *Science* **270**, 1838–1841
253. Tomac, S., Sarkar, M., Ratilainen, T., Wittung, P., Nielsen, P. E., Nordén, B., and Gräslund, A. (1996) *J. Am. Chem. Soc.* **118**, 5544–5552
253a. Ray, A., and Nordén, B. (2000) *FASEB J.* **14**, 1041–1060
253b. Izvolsky, K. I., Demidov, V. V., Nielsen, P. E., and Frank-Kamenetskii, M. D. (2000) *Biochemistry* **39**, 10908–10913
254. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. (1998) *Nature (London)* **391**, 468–471
255. Marsh, T. C., and Henderson, E. (1994) *Biochemistry* **33**, 10718–10724
256. Blackburn, E. H. (1991) *Nature (London)* **350**, 569–573
257. Blackburn, E. H. (1991) *Trends Biochem. Sci.* **16**, 378–381
258. Moyzis, R. K. (1991) *Sci. Am.* **265**(Aug), 48–55
259. Henderson, E., Hardin, C. C., Wolk, S. K., Tinoco, I., Jr., and Blackburn, E. (1987) *Cell* **51**, 899–908
260. Laughlan, G., Murchie, A. I. H., Norman, D. G., Moore, M. H., Moody, P. C. E., Lilley, D. M. J., and Luisi, B. (1994) *Science* **265**, 520–527
261. Kang, C. H., Zhang, X., Ratliff, R., Moyzis, R., and Rich, A. (1992) *Nature (London)* **356**, 126–131
262. Smith, F. W., and Feigon, J. (1993) *Biochemistry* **32**, 8682–8692
263. Balagurumoorthy, P., and Brahmachari, S. K. (1994) *J. Biol. Chem.* **269**, 21858–21869
264. Williamson, J. R. (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23**, 703–730
265. Frank-Kamenetskii, M. (1992) *Nature (London)* **356**, 105
266. Wang, K. Y., Swaminathan, S., and Bolton, P. H. (1994) *Biochemistry* **33**, 7517–7527
267. Wang, Y., and Patel, D. J. (1995) *J. Mol. Biol.* **251**, 76–94
268. Laporte, L., and Thomas, G. J., Jr. (1998) *Biochemistry* **37**, 1327–1335
269. Ross, W. S., and Hardin, C. C. (1994) *J. Am. Chem. Soc.* **116**, 6070–6080
270. Sen, D., and Gilbert, W. (1990) *Nature (London)* **344**, 410–414
271. Miura, T., Benevides, J. M., and Thomas, G. J., Jr. (1995) *J. Mol. Biol.* **248**, 233–238
272. Schierer, T., and Henderson, E. (1994) *Biochemistry* **33**, 2240–2246
273. Hardin, C. C., Corregan, M., Brown, B. A., II, and Frederick, L. N. (1993) *Biochemistry* **32**, 5870–5880
274. Cheong, C., and Moore, P. B. (1992) *Biochemistry* **31**, 8406–8414
275. Chen, F.-M. (1995) *J. Biol. Chem.* **270**, 23090–23096
276. Kettani, A., Kumar, R. A., and Patel, D. J. (1995) *J. Mol. Biol.* **254**, 638–656
277. Kang, C. H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R., and Rich, A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11636–11640
278. Gehring, K., Leroy, J.-L., and Guéron, M. (1993) *Nature (London)* **363**, 561–565
279. Lacroix, L., Mergny, J.-L., Leroy, J.-L., and Hélène, C. (1996) *Biochemistry* **35**, 8715–8722
280. Chen, L., Cai, L., Zhang, X., and Rich, A. (1994) *Biochemistry* **33**, 13540–13546
281. Kang, C. H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R., and Rich, A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3874–3878
282. Benevides, J. M., Kang, C., and Thomas, G. J., Jr. (1996) *Biochemistry* **35**, 5747–5755
283. Gallego, J., Chou, S.-H., and Reid, B. R. (1997) *J. Mol. Biol.* **273**, 840–856
284. Selsing, E., Wells, R. D., Alden, C. J., and Arnott, S. (1979) *J. Biol. Chem.* **254**, 5417–5422
285. Nishizaki, T., Iwai, S., Ohkubo, T., Kojima, C., Nakamura, H., Kyogoku, Y., and Ohtsuka, E. (1996) *Biochemistry* **35**, 4016–4025
286. Salazar, M., Champoux, J. J., and Reid, B. R. (1993) *Biochemistry* **32**, 739–744
287. Seeman, N. C., and Kallenbach, N. R. (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23**, 53–86
288. Zhang, S., Fu, T.-J., and Seeman, N. C. (1993) *Biochemistry* **32**, 8062–8067
289. Pikkemaat, J. A., van den Elst, H., van Boom, J. H., and Altona, C. (1994) *Biochemistry* **33**, 14896–14907
290. Carlström, G., and Chazin, W. J. (1996) *Biochemistry* **35**, 3534–3544
291. Holliday, R. (1964) *Genet. Res. Camb.* **5**, 282–304
292. Bennett, R. J., and West, S. C. (1995) *J. Mol. Biol.* **252**, 213–226
293. Grainger, R. J., Murchie, A. I. H., and Lilley, D. M. J. (1998) *Biochemistry* **37**, 23–32
294. Lilley, D. M. J. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9513–9515
295. Panyutin, I. G., Biswas, I., and Hsieh, P. (1995) *EMBO J.* **14**, 1819–1826
296. Chen, J., and Seeman, N. C. (1991) *Nature (London)* **350**, 631–633
297. Zhang, Y., and Seeman, N. C. (1994) *J. Am. Chem. Soc.* **116**, 1661–1669
297a. Liu, F., Sha, R., and Seeman, N. C. (1999) *J. Am. Chem. Soc.* **121**, 917–922
297b. Fahlman, R. P., and Sen, D. (1999) *J. Am. Chem. Soc.* **121**, 11079–11085
298. Feagin, J. E. (1990) *J. Biol. Chem.* **265**, 19373–19376
299. Stuart, K. (1991) *Trends Biochem. Sci.* **16**, 68–72
300. Simons, R. W., and Grunberg-Manago, M., eds. (1997) *RNA Structure and Function*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
301. Nicoloso, M., Qu, L.-H., Michot, B., and Bachellerie, J.-P. (1996) *J. Mol. Biol.* **260**, 178–195
302. Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. (1995) *Ann. Rev. Biochem.* **64**, 763–797
303. Draper, D. E. (1996) *Trends Biochem. Sci.* **21**, 145–149
304. Sich, C., Ohlenschläger, O., Ramachandran, R., Görlach, M., and Brown, L. R. (1997) *Biochemistry* **36**, 13989–14002

## References


305. Egli, M., Portmann, S., and Usman, N. (1996) *Biochemistry* **35**, 8489–8494
306. Auffinger, P., and Westhof, E. (1997) *J. Mol. Biol.* **274**, 54–63
307. Salazar, M., Fedoroff, O. Y., Miller, J. M., Ribeiro, N. S., and Reid, B. R. (1993) *Biochemistry* **32**, 4207–4215
308. Fedoroff, O. Y., Salazar, M., and Reid, B. R. (1993) *J. Mol. Biol.* **233**, 509–523
309. Ratmeyer, L., Vinayak, R., Zhong, Y. Y., Zon, G., and Wilson, W. D. (1994) *Biochemistry* **33**, 5298–5304
310. Varani, G., Cheong, C., and Tinoco, I., Jr. (1991) *Biochemistry* **30**, 3280–3289
311. Butcher, S. E., Dieckmann, T., and Feigon, J. (1997) *J. Mol. Biol.* **268**, 348–358
312. van Dongen, M. J. P., Wijmenga, S. S., van der Marel, G. A., van Boom, J. H., and Hilbers, C. W. (1996) *J. Mol. Biol.* **263**, 715–729
313. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1992) *Science* **256**, 217–219
314. Murphy, F. L., and Cech, T. R. (1994) *J. Mol. Biol.* **236**, 49–63
315. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. M., and Pardi, A. (1996) *J. Mol. Biol.* **264**, 968–980
315a. Gutell, R. R., Cannone, J. J., Konings, D., and Gautheret, D. (2000) *J. Mol. Biol.* **300**, 791–803
316. Fountain, M. A., Serra, M. J., Krugh, T. R., and Turner, D. H. (1996) *Biochemistry* **35**, 6539–6548
317. Huang, S., Wang, Y.-X., and Draper, D. E. (1996) *J. Mol. Biol.* **258**, 308–321
318. Nowakowski, J., and Tinoco, I., Jr. (1996) *Biochemistry* **35**, 2577–2585
319. Chang, K.-Y., and Tinoco, I., Jr. (1997) *J. Mol. Biol.* **269**, 52–66
320. Rosendahl, G., Hansen, L. H., and Douthwaite, S. (1995) *J. Mol. Biol.* **249**, 59–68
321. Pleij, C. W. A. (1990) *Trends Biochem. Sci.* **15**, 143–147
322. ten Dam, E., Pleij, K., and Draper, D. (1992) *Biochemistry* **31**, 11665–11676
323. Shen, L. X., and Tinoco, I., Jr. (1995) *J. Mol. Biol.* **247**, 963–978
324. Pinard, R., Payant, C., and Brakier-Gingras, L. (1995) *Biochemistry* **34**, 9611–9616
325. Puglisi, J. D., Wyatt, J. R., and Tinoco, I., Jr. (1988) *Nature (London)* **331**, 283–286
326. Qiu, H., Kaluarachchi, K., Du, Z., Hoffman, D. W., and Giedroc, D. P. (1996) *Biochemistry* **35**, 4176–4186
326a. Kolk, M. H., van der Graaf, M., Wijmenga, S. S., Pleij, C. W. A., Heus, H. A., and Hilbers, C. W. (1998) *Science* **280**, 434–438
327. Shi, P.-Y., Brinton, M. A., Veal, J. M., Zhong, Y. Y., and Wilson, W. D. (1996) *Biochemistry* **35**, 4222–4230
328. Hubbard, J. M., and Hearst, J. E. (1991) *Biochemistry* **30**, 5458–5465
329. Soll, D., and Abelson, J., eds. (1979) *Transfer RNA*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
330. Zimmermann, R. A. (1996) *Science* **271**, 1240–1241
331. Steinberg, S., Gautheret, D., and Cedergren, R. (1994) *J. Mol. Biol.* **236**, 982–989
332. Holbrook, S. R., Sussman, J. L., Warrant, R. W., and Kim, S.-H. (1978) *J. Mol. Biol.* **123**, 631–660
333. Rich, A., and Kim, S. H. (1978) *Sci. Am.* **238** (Jan), 52–62
334. Rich, A. (1978) *Trends Biochem. Sci.* **3**, 34–37
335. Moras, D., Comarmond, M. B., Fischer, J., Weiss, R., Thierry, J. C., Ebel, J. P., and Giegé, R. (1980) *Nature (London)* **288**, 669–674
336. Woo, N. H., Roe, B. A., and Rich, A. (1980) *Nature (London)* **286**, 346–351
337. Wakao, H., Romby, P., Westhof, E., Laalami, S., Grunberg-Manago, M., Ebel, J.-P., Ehresmann, C., and Ehresmann, B. (1989) *J. Biol. Chem.* **264**, 20363–20371
338. Hou, Y.-M., and Schimmel, P. (1992) *Biochemistry* **31**, 4157–4160
339. Leehey, M. A., Squassoni, C. A., Friederich, M. W., Mills, J. B., and Hagerman, P. J. (1995) *Biochemistry* **34**, 16235–16239
340. Quigley, G. J., and Rich, A. (1976) *Science* **194**, 796–806
341. Brimacombe, R., Atmadja, J., Stiege, W., and Schüler, D. (1988) *J. Mol. Biol.* **199**, 115–136
342. Mueller, F., and Brimacombe, R. (1997) *J. Mol. Biol.* **271**, 524–544
343. Turner, D. H., and Sugimoto, N. (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 167–192
344. Brimacombe, R. (1988) *Biochemistry* **27**, 4207–4213
345. Schnare, M. N., Damberger, S. H., Gray, M. W., and Gutell, R. R. (1996) *J. Mol. Biol.* **256**, 701–719
346. Yonath, A. (1984) *Trends Biochem. Sci.* **9**, 227–230
347. Pennisi, E. (1999) *Science* **285**, 2048–2051
348. Svergun, D. I., Koch, M. H. J., Pedersen, J. S., and Serdyuk, I. N. (1994) *J. Mol. Biol.* **240**, 78–86
349. Vladimirov, S. N., Druzina, Z., Wang, R., and Cooperman, B. S. (2000) *Biochemistry* **39**, 183–193
350. Gregory, S. T., and Dahlberg, A. E. (1999) *J. Mol. Biol.* **285**, 1475–1483
351. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science* **289**, 905–920
352. Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., and Noller, H. F. (1999) *Science* **285**, 2095–2104
353. Culver, G. M., Cate, J. H., Yusupova, G. Z., Yusupov, M. M., and Noller, H. F. (1999) *Science* **285**, 2133–2135
354. Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000) *Nature (London)* **407**, 340–348
355. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) *Science* **289**, 920–930
356. Muth, G. W., Ortoleva-Donnelly, L., and Strobel, S. A. (2000) *Science* **289**, 947–950
357. Cech, T. R. (2000) *Science* **289**, 878–879
358. Pennisi, E. (1996) *Science* **273**, 574–575
359. O'Gara, M., Klimasauskas, S., Roberts, R. J., and Cheng, X. (1996) *J. Mol. Biol.* **261**, 634–645
360. Gommers-Ampt, J. H., and Borst, P. (1995) *FASEB J.* **9**, 1034–1042
361. Maltman, K. L., Neuhard, J., and Warren, R. A. J. (1981) *Biochemistry* **20**, 3586–3591
362. Lipscomb, L. A., Peek, M. E., Morningstar, M. L., Verghis, S. M., Miller, E. M., Rich, A., Essigmann, J. M., and Williams, L. D. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 719–723
363. Leonard, G. A., McAuley-Hecht, K. E., Gibson, N. J., Brown, T., Watson, W. P., and Hunter, W. N. (1994) *Biochemistry* **33**, 4755–4761
364. Wrzesinski, J., Bakin, A., Nurse, K., Lane, B. G., and Ofengand, J. (1995) *Biochemistry* **34**, 8904–8913
365. Ellington, A. D., and Szostak, J. W. (1990) *Nature (London)* **346**, 818–822
366. Fan, P., Suri, A. K., Fiala, R., Live, D., and Patel, D. J. (1996) *J. Mol. Biol.* **258**, 480–500
367. Jenison, R. D., Gill, S. C., Pardi, A., and Polisky, B. (1994) *Science* **263**, 1425–1429
368. Hamasaki, K., Killian, J., Cho, J., and Rando, R. R. (1998) *Biochemistry* **37**, 656–663
369. Connell, G. J., Illangesekare, M., and Yarus, M. (1993) *Biochemistry* **32**, 5497–5502
370. Patel, D. J., Suri, A. K., Jiang, F., Jiang, L., Fan, P., Kumar, R. A., and Nonin, S. (1997) *J. Mol. Biol.* **272**, 645–664
371. Crick, F. H. C. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 3–9
372. Cornish-Bowden, A. (1996) *Trends Biochem. Sci.* **21**, 155
373. Hall, B. D. (1979) *Nature (London)* **282**, 129–130
374. Barrell, B. G., Bankier, A. T., and Drouin, J. (1979) *Nature (London)* **282**, 189–194
375. Horowitz, S., and Gorovsky, M. A. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2452–2455
376. Chambers, I., and Harrison, P. R. (1987) *Trends Biochem. Sci.* **12**, 255–256
377. Gardner, M. (1970) *Sci. Am.* **223**(Aug), 110–112
378. Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A., and Frank-Kamenetskii, M. D. (1987) *Nature (London)* **330**, 495–497
379. Guieysse, A.-L., Praseuth, D., and Hélène, C. (1997) *J. Mol. Biol.* **267**, 289–298
380. Panayotatos, N., and Fontaine, A. (1987) *J. Biol. Chem.* **262**, 11364–11368
381. Pabo, C. O., and Sauer, R. T. (1984) *Ann. Rev. Biochem.* **53**, 293–321
382. Berg, O. G., and von Hippel, P. H. (1988) *Trends Biochem. Sci.* **13**, 207–211
383. Marmorstein, R. Q., Joachimiak, A., Sprinzl, M., and Sigler, P. B. (1987) *J. Biol. Chem.* **262**, 4922–4927
384. Muto, A., and Osawa, S. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 166–169
385. Singer, C. E., and Ames, B. N. (1979) *Science* **170**, 822–826
386. Singer, C. E., and Ames, B. N. (1972) *Science* **175**, 1391
387. Harrison, S. C. (1991) *Nature (London)* **353**, 715–719
388. Churchill, M. E. A., and Travers, A. A. (1991) *Trends Biochem. Sci.* **16**, 92–97
389. Travers, A. (1993) *DNA-Protein Interactions*, Chapman & Hall, New York
390. Jones, S., van Heyningen, P., Berman, H. M., and Thornton, J. M. (1999) *J. Mol. Biol.* **287**, 877–896
391. Arents, G., and Moudrianakis, E. N. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11170–11174
392. Starich, M. R., Sandman, K., Reeve, J. N., and Summers, M. F. (1996) *J. Mol. Biol.* **255**, 187–203
393. Hud, N. V., Milanovich, F. P., and Balhorn, R. (1994) *Biochemistry* **33**, 7528–7535
394. Lohman, T. M., and Ferrari, M. E. (1994) *Ann. Rev. Biochem.* **63**, 527–570
395. Overman, L. B., and Lohman, T. M. (1994) *J. Mol. Biol.* **236**, 165–178
396. Folmer, R. H. A., Nilges, M., Folkers, P. J. M., Konings, R. N. H., and Hilbers, C. W. (1994) *J. Mol. Biol.* **240**, 341–357
397. Olah, G. A., Gray, D. M., Gray, C. W., Kergil, D. L., Sosnick, T. R., Mark, B. L., Vaughan, M. R., and Trewhella, J. (1995) *J. Mol. Biol.* **249**, 576–594
398. Kanellopoulos, P. N., Tsernoglou, D., van der Vliet, P. C., and Tucker, P. A. (1996) *J. Mol. Biol.* **257**, 1–8
399. Bogdarina, I., Fox, D. G., and Kneale, G. G. (1998) *J. Mol. Biol.* **275**, 443–452
400. Pluta, A. F., Mackay, A. M., Ainsztein, A. M., Goldberg, I. G., and Earnshaw, W. C. (1995) *Science* **270**, 1591–1594
401. Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L., and Sigler, P. B. (1985) *Nature (London)* **317**, 782–786
402. Zhang, R.-G, Joachimiak, A., Lawson, C. L., Schevitz, R. W., Otwinowski, Z., and Sigler, P. B. (1987) *Nature (London)* **327**, 591–597

# References


403. Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) *Nature (London)* **335**, 321–329
404. Shakked, Z., Guzikevich– Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A., and Sigler, P. B. (1994) *Nature (London)* **368**, 469–473
405. Lawson, C. L., and Carey, J. (1993) *Nature (London)* **366**, 178–182
406. Lawson, C. L., and Sigler, P. B. (1988) *Nature (London)* **333**, 869–871
407. Guenot, J., Fletterick, R. J., and Kollman, P. A. (1994) *Protein Sci.* **3**, 1276–1285
408. Rastinejad, F., Artz, P., and Lu, P. (1993) *J. Mol. Biol.* **233**, 389–399
409. Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S., and Miller, J. H. (1994) *J. Mol. Biol.* **240**, 421–433
410. Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) *Science* **271**, 1247–1254
411. Brennan, R. G., and Matthews, B. W. (1989) *J. Biol. Chem.* **264**, 1903–1906
412. Jordan, S. R., and Pabo, C. O. (1988) *Science* **242**, 893–899
413. Wolberger, C., Dong, Y., Ptashne, M., and Harrison, S. C. (1988) *Nature (London)* **335**, 789–795
414. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., and Harrison, S. C. (1988) *Science* **242**, 899–907
415. Bell, A. C., and Koudelka, G. B. (1993) *J. Mol. Biol.* **234**, 542–553
416. Mandel-Gutfreund, Y., Schueler, O., and Margalit, H. (1995) *J. Mol. Biol.* **253**, 370–382
417. Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) *Nature (London)* **363**, 38–44
418. Junius, F. K., O'Donoghue, S. I., Nilges, M., Weiss, A. S., and King, G. F. (1996) *J. Biol. Chem.* **271**, 13663–13667
419. Keller, W., König, P., and Richmond, T. J. (1995) *J. Mol. Biol.* **254**, 657–667
420. Harbury, P. B., Kim, P. S., and Alber, T. (1994) *Nature (London)* **371**, 80–83
421. Klug, A., and Schwabe, J. W. R. (1995) *FASEB J.* **9**, 597–604
422. Rhodes, D., and Klug, A. (1993) *Sci. Am.* **268** (Feb), 56–65
423. Pavletich, N. P., and Pabo, C. O. (1991) *Science* **252**, 809–817
424. Mackay, J. P., and Crossley, M. (1998) *Trends Biochem. Sci.* **23**, 1–4
425. Miller, J., McLachlan, A. D., and Klug, A. (1985) *EMBO J.* **4**, 1609–1614
426. Pieler, T., and Theunissen, O. (1993) *Trends Biochem. Sci.* **18**, 226–230
427. Pavletich, N. P., and Pabo, C. O. (1993) *Science* **261**, 1701–1707
428. Krizek, B. A., Amann, B. T., Kilfoil, V. J., Merkle, D. L., and Berg, J. M. (1991) *J. Am. Chem. Soc.* **113**, 4518–4523
429. Somers, W. S., and Phillips, S. E. V. (1992) *Nature (London)* **359**, 387–393
430. Phillips, S. E. V. (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23**, 671–701
431. Breg, J. N., van Opheusden, H. J., Burgering, M. J. M., Boelens, R., and Kaptein, R. (1990) *Nature (London)* **346**, 586–589
432. Bonnefoy, E., Takahashi, M., and Yaniv, J. R. (1994) *J. Mol. Biol.* **242**, 116–129
433. Müller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L., and Harrison, S. C. (1995) *Nature (London)* **373**, 311–317
434. Ghosh, G., Van Duyne, G., Ghosh, S., and Sigler, P. B. (1995) *Nature (London)* **373**, 303–310
435. Baltimore, D., and Beg, A. A. (1995) *Nature (London)* **373**, 287–288
436. Kaufmann, E., Müller, D., and Knöchel, W. (1995) *J. Mol. Biol.* **248**, 239–254
437. Burd, C. G., and Dreyfuss, G. (1994) *Science* **265**, 615–621
438. Nagai, K., and Mattaj, I. W., eds. (1994) *RNA-Protein Interactions*, IRL Press, Oxford
439. Draper, D. E. (1995) *Ann. Rev. Biochem.* **64**, 593–620
440. Hentze, M. W. (1994) *Trends Biochem. Sci.* **19**, 101–103
441. Allain, F. H.-T., Gubser, C. C., Howe, P. W. A., Nagai, K., Neuhaus, D., and Varani, G. (1996) *Nature (London)* **380**, 646–650
442. Kenan, D. J., Query, C. C., and Keene, J. D. (1991) *Trends Biochem. Sci.* **16**, 214–220
443. Fierro-Monti, I., and Mathews, M. B. (2000) *Trends Biochem. Sci.* **25**, 241–246
444. Voyles, B. A. (1993) *The Biology of Viruses*, Mosby, St. Louis, Missouri
445. Fields, B. N., and Knipe, D. M., eds. (1990) *Fields Virology*, 2nd ed., Raven Press, New York
446. Cann, A. (1993) *Principles of Molecular Virology*, Academic Press, San Diego, California
447. Endemann, H., and Model, P. (1995) *J. Mol. Biol.* **250**, 496–506
448. Williams, K. A., Glibowicka, M., Li, Z., Li, H., Khan, A. R., Chen, Y. M. Y., Wang, J., Marvin, D. A., and Deber, C. M. (1995) *J. Mol. Biol.* **252**, 6–14
449. Wen, Z. Q., Overman, S. A., and Thomas, G. J., Jr. (1997) *Biochemistry* **36**, 7810–7820
450. Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A., III, Slocombe, P. M., and Smith, M. (1977) *Nature (London)* **265**, 687–695
451. McKenna, R., Ilag, L. L., and Rossmann, M. G. (1994) *J. Mol. Biol.* **237**, 517–543
452. McKenna, R., Xia, D., Willingmann, P., Ilag, L. L., Krishnaswamy, S., Rossmann, M. G., Olson, N. H., Baker, T. S., and Incardona, N. L. (1992) *Nature (London)* **355**, 137–143
453. Godson, G. N. (1978) *Trends Biochem. Sci.* **3**, 249–253
454. McKenna, R., Bowman, B. R., Ilag, L. L., Rossmann, M. G., and Fane, B. A. (1996) *J. Mol. Biol.* **256**, 736–750
455. Wu, H., and Rossmann, M. G. (1993) *J. Mol. Biol.* **233**, 231–244
456. Tsao, J., Chapman, M. S., Agbandje, M., Keller, W., Smith, K., Wu, H., Luo, M., Smith, T. J., Rossmann, M. G., Compans, R. W., and Parrish, C. R. (1991) *Science* **251**, 1456–1464
457. Chipman, P. R., Agbandje-McKenna, M., Kajigaya, S., Brown, K. E., Young, N. S., Baker, T. S., and Rossmann, M. G. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7502–7506
458. Howarth, A. J., and Goodman, R. M. (1982) *Trends Biochem. Sci.* **7**, 180–182
459. Tiollais, P., Pourcel, C., and Dejean, A. (1985) *Nature (London)* **317**, 489–495
460. Tiollais, P., and Buendia, M.-A. (1991) *Sci. Am.* **264**(Apr), 116–123
461. Böttcher, B., Wynne, S. A., and Crowther, R. A. (1997) *Nature (London)* **386**, 88–91
462. Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L., and Weissman, S. M. (1978) *Science* **200**, 494–502
463. Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L., and Harrison, S. C. (1991) *Nature (London)* **354**, 278–284
464. Griffith, J. P., Griffith, D. L., Rayment, I., Murakami, W. T., and Caspar, D. L. D. (1992) *Nature (London)* **355**, 652–654
465. Yang, R. C. A., and Wu, R. (1979) *Science* **206**, 456–462
466. Howley, P. M. (1986) *N. Engl. J. Med.* **315**, 1089–1090
467. Hess, M., Cuzange, A., Ruigrok, R. W. H., Chroboczek, J., and Jacrot, B. (1995) *J. Mol. Biol.* **252**, 379–385
468. Athappilly, F. K., Murali, R., Rux, J. J., Cai, Z., and Burnett, R. M. (1994) *J. Mol. Biol.* **242**, 430–455
469. Witkowski, J. A. (1988) *Trends Biochem. Sci.* **13**, 110–113
470. Sugden, B. (1991) *Trends Biochem. Sci.* **16**, 45–46
471. Zhou, Z. H., Prasad, B. V. V., Jakana, J., Rixon, F. J., and Chiu, W. (1994) *J. Mol. Biol.* **242**, 456–469
472. Trus, B. L., Booy, F. P., Newcomb, W. W., Brown, J. C., Homa, F. L., Thomsen, D. R., and Steven, A. C. (1996) *J. Mol. Biol.* **263**, 447–462
473. Henle, W., Henle, G., and Lennette, E. T. (1974) *Sci. Am.* **241**(Jul), 48–59
474. Berardelli, P. (1997) *Science* **278**, 1710
475. Senkevich, T. G., Bugert, J. J., Sisler, J. R., Koonin, E. V., Darai, G., and Moss, B. (1996) *Science* **273**, 813–816
476. Hohn, T., Hohn, B., and Pfeiffer, P. (1985) *Trends Biochem. Sci.* **10**, 205–209
477. Rosa, P., Mantovani, S., Rosboch, R., and Huttner, W. B. (1992) *J. Biol. Chem.* **267**, 12227–12232
478. Diener, T. O. (1991) *FASEB J.* **5**, 2808–2813
479. Diener, T. V. (1984) *Trends Biochem. Sci.* **9**, 133–136
480. Diener, T. O. (1979) *Viroids and Viroid Diseases*, Wiley, New York
481. Martínez-Soriano, J. P., Galindo-Alonso, J., Maroon, C. J. M., Yucel, I., Smith, D. R., and Diener, T. O. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9397–9401
482. Polson, A. G., Bass, B. L., and Casey, J. L. (1996) *Nature (London)* **380**, 454–456
483. Wang, K.-S., Choo, Q.-L., Weiner, A. J., Ou, J.-H., Najarian, R. C., Thayer, R. M., Mullenbach, G. T., Denniston, K. J., Gerin, J. L., and Houghton, M. (1986) *Nature (London)* **323**, 508–514
484. Lai, M. M. C. (1995) *Ann. Rev. Biochem.* **64**, 259–286
485. Unge, T., Liljas, L., Strandberg, B., Vaara, I., Kannan, K. K., Fridborg, K., Nordman, C. E., and Lentz, P. J., Jr. (1980) *Nature (London)* **285**, 373–377
486. Larson, S. B., Koszelak, S., Day, J., Greenwood, A., Dodds, J. A., and McPherson, A. (1993) *Nature (London)* **361**, 179–182
487. Larson, S. B., Koszelak, S., Day, J., Greenwood, A., Dodds, J. A., and McPherson, A. (1993) *J. Mol. Biol.* **231**, 375–391
488. Golmohammadi, R., Valegård, K., Fridborg, K., and Liljas, L. (1993) *J. Mol. Biol.* **234**, 620–639
489. LeCuyer, K. A., Behlen, L. S., and Uhlenbeck, O. C. (1995) *Biochemistry* **34**, 10600–10606
490. Stonehouse, N. J., Valegård, K., Golmohammadi, R., van den Worm, S., Walton, C., Stockley, P. G., and Liljas, L. (1996) *J. Mol. Biol.* **256**, 330–339
491. Beremarand, M. N., and Blumenthal, T. (1979) *Cell* **18**, 257–266
492. Hogle, J. M., Chow, M., and Filman, D. J. (1985) *Science* **229**, 1358–1365
493. Hogle, J. M., Chow, M., and Filman, D. J. (1987) *Sci. Am.* **256**(Mar), 42–49
494. Najaran, R., Caput, D., Gee, W., Potter, S. J., Renard, A., Merryweather, J., Van Nest, G., and Dina, D. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2627–2631
495. Rossmann, M. G., Arnold, E., Griffith, J. P., Kamer, G., Luo, M., Smith, T. J., Vriend, G., Rueckert, R. R., Sherry, B., McKinlay, M. A., Diana, G., and Otto, M. (1987) *Trends Biochem. Sci.* **12**, 313–318

## References


496. Olson, N. H., Kolatkar, P. R., Oliveira, M. A., Cheng, R. H., Greve, J. M., McClelland, A., Baker, T. S., and Rossmann, M. G. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 507–511
497. Hadfield, A. T., Oliveira, M. A., Kim, K. H., Minor, I., Kremer, M. J., Heinz, B. A., Shepard, D., Pevear, D. C., Rueckert, R. R., and Rossmann, M. G. (1995) *J. Mol. Biol.* **253**, 61–73
498. Küpper, H., Keller, W., Kurz, C., Forss, S., Schaller, H., Franze, R., Stohmaier, K., Marquardt, O., Zaslavsky, V. G., and Hofschneider, P. H. (1981) *Nature (London)* **289**, 555–559
499. Brown, F. (1981) *Trends Biochem. Sci.* **6**, 325–327
500. Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D., and Brown, F. (1989) *Nature (London)* **337**, 709–716
501. Luo, M., Vriend, G., Kamer, G., Minor, I., Arnold, E., Rossmann, M. G., Boege, U., Scraba, D. G., Duke, G. M., and Palmenberg, A. C. (1987) *Science* **235**, 182–191
502. Choi, H.-K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G., and Wengler, G. (1991) *Nature (London)* **354**, 37–43
503. Simons, K., Garoff, H., and Helenius, A. (1982) *Sci. Am.* **246**(Feb), 58–66
504. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K., and Bricogne, G. (1978) *Nature (London)* **276**, 368–373
505. Erickson, J. W., Silva, A. M., Murthy, M. R. N., Fita, I., and Rossmann, M. G. (1985) *Science* **229**, 625–629
506. Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R., and Klug, A. (1978) *Nature (London)* **276**, 362–368
507. Holmes, K. C. (1980) *Trends Biochem. Sci.* **5**, 4–7
507a. Lauffer, M. A. (1984) *Trends Biochem. Sci.* **9**, 369–371
508. Wang, H., and Stubbs, G. (1994) *J. Mol. Biol.* **239**, 371–384
509. Kaplan, M. M., and Webster, R. G. (1977) *Sci. Am.* **237**(Dec), 88–106
510. von Itzstein, M., and 17 other authors. (1993) *Nature (London)* **363**, 418–423
511. Clarke, D. K., Duarte, E. A., Elena, S. F., Moya, A., Domingo, E., and Holland, J. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4821–4824
512. Varmus, H. (1988) *Science* **240**, 1427–1434
513. Darlix, J.-L., Lapadat-Tapolsky, M., de Rocquigny, H., and Roques, B. P. (1995) *J. Mol. Biol.* **254**, 523–537
514. Löwer, R., Löwer, J., and Kurth, R. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5177–5184
515. Stephenson, J. R., ed. (1980) *Molecular Biology of RNA Tumor Viruses*, Academic Press, New York
516. Compans, R. W., and Bishop, D. H. L., eds. (1983) *Double-Stranded RNA Viruses*, Elsevier, Biomedical, New York
517. Thornton, A., and Zuckerman, A. J. (1975) *Nature (London)* **254**, 557–558
518. Baldwin, M. A., Cohen, F. E., and Prusiner, S. B. (1995) *J. Biol. Chem.* **270**, 19197–19200
519. Manuelidis, L., Sklaviadis, T., Akowitz, A., and Fritch, W. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5124–5128
520. Prusiner, S. B. (1996) *Trends Biochem. Sci.* **21**, 482–487
521. Prusiner, S. B. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13363–13383
522. Weissmann, C. (1999) *J. Biol. Chem.* **274**, 3–6
523. Chesebro, B. (1998) *Science* **286**, 660–662
524. Balter, M. (1999) *Science* **286**, 660–662
525. Temin, H. M. (1978) *Trends Biochem. Sci.* **3**, N80–N83
525a. Sambrook, J., and Russell, D. (2000) *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Lab. Press, Plainview, New York
526. Grossman, L., and Moldave, K., eds. (1967) *Methods in Enzymology*, Vol. 12A, Academic Press, New York (pp. 531–708)
527. Robyt, J. F., and White, B. J. (1996) *Biochemical Techniques, Theory and Practice*, 2nd ed., Waveland Press, Prospect Heights, Illinois
528. Blackburn, G. M., and Gait, M. J., eds. (1996) *Nucleic Acids in Chemistry and Biology*, 2nd ed., Oxford Univ. Press, Oxford
529. Rajeswari, M. R., Montenay-Garestier, T., and Helene, C. (1987) *Biochemistry* **26**, 6825–6831
530. Hecker, R., Colpan, M., and Riesner, D. (1985) *Journal of Chromatography* **326**, 251–261
531. Edwardson, P. A. D., Atkinson, T., Lowe, C. R., and Small, D. A. P. (1986) *Anal. Biochem.* **152**, 215–220
532. Lawrance, S. K., Smith, C. L., Srivastava, R., Cantor, C. R., and Weissman, S. M. (1987) *Science* **235**, 1387–1390
533. Cantor, C. R., Smith, C. L., and Matthew, M. K. (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 287–304
534. Smith, C. L., and Cantor, C. R. (1987) *Trends Biochem. Sci.* **12**, 284–287
535. Kochetkov, N. K., and Budovskii, E. I. (1972) *Organic Chemistry of Nucleic Acids*, Plenum, New York (Part A, pp. 137–147)
536. Bozic, D., Grazulis, S., Siksnys, V., and Huber, R. (1996) *J. Mol. Biol.* **255**, 176–186
537. Smith, H. O. (1979) *Science* **205**, 455–462
538. Air, G. M. (1979) *CRC Crit. Revs. Biochem.* **6**, 1–33
539. Roberts, R. J. (1980) *Methods Enzymol.* **65**, 1–15
540. Koob, M., and Szybalski, W. (1990) *Science* **250**, 271–273
541. Veselkov, A. G., Demidov, V. V., Frank-Kamenetskii, M. D., and Nielsen, P. E. (1996) *Nature (London)* **379**, 214
542. Zaug, A. J., Been, M. D., and Cech, T. R. (1986) *Nature (London)* **324**, 429–433
543. Uhlenbeck, O. C. (1987) *Nature (London)* **328**, 596–600
544. Landgraf, R., Chen, C.-hB., and Sigman, D. S. (1994) *Biochemistry* **33**, 10607–10615
545. Adams, R. L. P., Burdon, R. H., Campbell, A. M., and Smellie, R. M. S. (1976) *Davidson's The Biochemistry of the Nucleic Acids*, 8th ed., Academic Press, New York (pp. 50–82)
546. Rayford, R., Anthony, D. D., Jr., O'Neill, R. E., Jr., and Merrick, W. C. (1985) *J. Biol. Chem.* **260**, 15708–15713
547. Bochner, B. R., and Ames, B. N. (1982) *J. Biol. Chem.* **257**, 9759–9769
548. Chen, H., and Shaw, B. R. (1994) *Biochemistry* **33**, 4121–4129
549. Hayatsu, H., and Shiragami, M. (1979) *Biochemistry* **18**, 632–647
550. Spiro, T. G., ed. (1980) *Nucleic Acid-Metal Ion Interactions*, Wiley, New York
551. Dische, Z. (1955) *Nucleic Acids* **1**, 755
552. Marmur, J., and Doty, P. (1962) *J. Mol. Biol.* **5**, 109–118
553. Korolev, N., Lyubartsev, A. P., and Nordenskiöld, L. (1998) *Biophys. J.* **75**, 3041–3056
554. Davidson, J. N. (1972) *The Biochemistry of Nucleic Acids*, 7th ed., Academic Press, New York (p. 148)
555. Wada, A., Yabuki, S., and Husimi, Y. (1980) *Crit. Revs. Biochem.* **9**, 87–144
556. Britten, R. J., and Kohne, D. E. (1968) *Science* **161**, 529–540
557. Wilson, D. A., and Thomas, C. A., Jr. (1974) *J. Mol. Biol.* **84**, 115–144
558. Bendich, A. J., and Bolton, E. T. (1968) in *Methods in Enzymology*, Vol. 12B (Grossman, L., and Moldave, K., eds), pp. 635–640, Academic Press, New York
559. Gillespie, D. (1968) in *Methods in Enzymology*, Vol. 12B (Grossman, L., and Moldave, K., eds), pp. 641–668, Academic Press, New York
560. Hall, B. D., Haarr, L., and Kleppe, K. (1980) *Trends Biochem. Sci.* **5**, 254–256
561. Britten, R. J., and Kohne, D. E. (1967) *Carnegie Instit. Washington Yearbook*, 65th ed., Carnegie Instit., Washington, D. C. (pp. 78–106)
562. Marmur, J. (1994) *Trends Biochem. Sci.* **19**, 343–346
563. Szabo, P., and Ward, D. C. (1982) *Trends Biochem. Sci.* **7**, 425–427
564. Kessler, C., ed. (1992) *Nonradioactive Labeling and Detection of Biomolecules*, Springer-Verlag, New York
565. Leitch, A. R., Chwarzacher, T., Jackson, D., and Leitch, I. J. (1994) *In Situ Hybridization: a Practical Guide*, Bios, Oxford
566. Kumar, A., Tchen, P., Roullet, F., and Cohen, J. (1988) *Anal. Biochem.* **169**, 376–
567. Isaac, P. G., ed. (1994) *Protocols for Nucleic Acid Analysis by Nonradioactive Probes*, Humana Press, Totowa, New Jersey
568. Gillevet, P. M. (1990) *Nature (London)* **348**, 657–658
569. Negro, F., Pacchioni, D., Shimizu, Y., Miller, R. H., Bussolati, G., Purcell, R. H., and Bonino, F. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2247–2251
570. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
571. Gubler, U., and Hoffman, B. J. (1983) *Gene* **25**, 263–269
572. Watson, C. J., and Jackson, J. F. (1985) in *DNA Cloning Vol. I: A Practical Approach* (Glover, D. M., ed), pp. 79–100, IRL Press, Washington, DC
573. Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y., and Matsubara, K. (1985) *J. Biol. Chem.* **260**, 2605–2608
574. Warner, B. D., Warner, M. E., Karus, G. A., Ku, L., Brown-Shimer, S., and Urdea, M. S. (1983) *DNA* **4**, 401–411
575. Sekiya, T., Takeya, T., Brown, E. L., Belagaje, R., Contreras, R., Fritz, H.-J., Gait, M. J., Lees, R. G., Ryan, M. J., Khorana, H. G., and Norris, K. E. (1979) *J. Biol. Chem.* **254**, 5787–5801
576. Wada, T., Sato, Y., Honda, F., Kawahara, S.-i, and Sekine, M. (1997) *J. Am. Chem. Soc.* **119**, 12710–12721
577. Matteucci, M. D., and Caruthers, M. H. (1981) *J. Am. Chem. Soc.* **103**, 3185–3191
578. Beaucage, S. L., and Caruthers, M. H. (1981) *Tetrahedron Letters* **22**, 1859–1862
579. Air, G. M. (1979) *CRC Crit. Revs. Biochem.* **6**, 1–33
580. Alvarado-Urbina, G., Sathe, G. M., Liu, W.-C., Gillen, M. F., Duck, P. D., Bender, R., and Ogilvie, K. K. (1981) *Science* **214**, 270–274
581. Caruthers, M. H. (1985) *Science* **230**, 281–285
582. Lashkari, D. A., Hunicke-Smith, S. P., Norgren, R. M., Davis, R. W., and Brennan, T. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7912–7915
583. Usman, N., Ogilvie, K. K., Jiang, M.-Y., and Cedergren, R. J. (1987) *J. Am. Chem. Soc.* **109**, 7845–7854
584. Ebe, K., Schöed, M., Rossi, J. J., and Wallace, R. B. (1987) *DNA* **6**, 497–504
585. Usman, N., and Cedergren, R. (1992) *Trends Biochem. Sci.* **17**, 334–339
586. Eaton, B. E., and Pieken, W. A. (1995) *Ann. Rev. Biochem.* **64**, 837–863
587. Rohatgi, R., Bartel, D. P., and Szostak, J. W. (1996) *J. Am. Chem. Soc.* **118**, 3340–3344
588. Kleppe, K., Ohtsuka, E., Kleppe, R., and Khorana, H. G. (1971) *J. Mol. Biol.* **56**, 341–361
589. Panet, A., and Khorana, H. G. (1974) *J. Biol. Chem.* **249**, 5213–5221

# References


590. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N. (1985) *Science* **230**, 1350–1354
591. Mullis, K. B. (1990) *Sci. Am.* **262**(Apr), 56–65
592. Rabinow, P. (1996) *Making PCR: A Story of Biotechnology*, Univ. Chicago Press, Chicago, Illinois
593. Mullis, K., Ferré, F., and Gibbs, R., eds. (1994) *The Polymerase Chain Reaction*, Birkhäuser, Boston, Massachusetts
594. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) *Science* **239**, 487–491
595. Erlich, H. A., Gelfand, D. H., and Saiki, R. K. (1988) *Nature (London)* **331**, 461–462
596. Appenzeller, T. (1990) *Science* **247**, 1030–1032
597. Newton, C., and Graham, A. (1997) *PCR*, 2nd ed., Springer-Verlag, New York
598. Bloch, W. (1991) *Biochemistry* **30**, 2735–2747
599. White, B. A., ed. (1993) *PCR Protocols*, Humana Press, Totowa, New Jersey

599a. Innis, M. A., Gelfand, D. H., and Sninsky, J. J., eds. (1999) *PCR Applications*, Academic Press, San Diego, California

600. Cohen, J. (1994) *Science* **263**, 1564–1565
601. Pääbo, S. (1993) *Sci. Am.* **269**(Nov), 86–92
602. Pääbo, S., Higuchi, R. G., and Wilson, A. C. (1989) *J. Biol. Chem.* **264**, 9709–9712
603. Höss, M., Pääbo, S., and Vereshchagin, N. K. (1994) *Nature (London)* **370**, 333
604. Scharf, S. J., Horn, G. T., and Erlich, H. A. (1986) *Science* **233**, 1076–1078
605. Li, H., Gyllensten, U. B., Cui, X., Saiki, R. K., Erlich, H. A., and Arnheim, N. (1988) *Nature (London)* **335**, 414–417
606. Engelke, D. R., Hoener, P. A., and Collins, F. S. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 544–548
607. Karlovsky, P. (1990) *Trends Biochem. Sci.* **15**, 419
608. Hengen, P. N. (1995) *Trends Biochem. Sci.* **20**, 372–373
609. Wong, G. K.-S., Yu, J., Thayer, E. C., and Olson, M. V. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5225–5230
610. Bernardi, G. (1978) *Nature (London)* **276**, 558–559
611. Bonitz, S. G., Coruzzi, G., Thalenfeld, B. E., Tzagoloff, A., and Macino, G. (1980) *J. Biol. Chem.* **255**, 11927–11941
612. Martin, N. C., Miller, D. L., and Donelson, J. E. (1979) *J. Biol. Chem.* **254**, 11729–11734
613. Fischer, S. G., and Lerman, L. S. (1979) *Methods Enzymol.* **68**, 183–191
614. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517
615. Southern, E. (1979) *Methods Enzymol.* **68**, 152–176
616. Grunstein, M., and Wallis, J. (1979) *Methods Enzymol.* **68**, 379–389
617. Sanger, F., and Coulson, A. R. (1975) *J. Mol. Biol.* **94**, 441–448
618. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
619. Alphey, L. (1997) *DNA Sequencing*, Springer-Verlag, New York
620. Ansorge, W., Voss, H., and Zimmerman, J., eds. (1997) *DNA Sequencing Strategies - Automated and Advanced Approaches*, Wiley, New York
621. Griffin, H. G., and Griffin, A. M., eds. (1993) *DNA Sequencing Protocols*, Humana Press, Totowa, New Jersey
622. Smith, A. J. H. (1980) *Methods Enzymol.* **65**, 560–580
623. Gillam, S., and Smith, M. (1980) *Methods Enzymol.* **65**, 687–701
624. Chen, E. Y., and Seeburg, P. H. (1985) *DNA* **4**, 165–170
625. Ornstein, D., Moen, P. T., and Kashdan, M. A. (1985) *DNA* **4**, 94
626. Bassam, B. J., Caetano-Anolles, G., and Gresshoff, P. M. (1991) *Anal. Biochem.* **196**, 80–83
627. Mizusawa, S., Nishimura, S., and Seela, F. (1986) *Nucleic Acids Res.* **14**, 1319–1324
628. Gerken, T. A., Gupta, R., and Jentoft, N. (1992) *Biochemistry* **31**, 639–648
629. Hengen, P. N. (1996) *Trends Biochem. Sci.* **21**, 33–34
630. Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E., and Surrey, S. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4298–4302
631. Henikoff, S. (1984) *Gene* **28**, 351–359
632. Adams, M. D., Kerlavage, A. R., Kelley, J. M., Gocayne, J. D., Fields, C., Fraser, C. M., and Venter, J. C. (1994) *Nature (London)* **368**, 474–475
633. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., and Hood, L. E. (1986) *Nature (London)* **321**, 674–679
634. Ju, J., Ruan, C., Fuller, C. W., Glazer, A. N., and Mathies, R. A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4347–4351
635. Fu, D.-J., Broude, N. E., Köster, H., Smith, C. L., and Cantor, C. R. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10162–10166
636. Bult, C. J., and 39 other authors. (1996) *Science* **273**, 1058–1073
637. Sutcliffe, J. G. (1995) *Trends Biochem. Sci.* **20**, 87–90
638. Rowen, L., Mahairas, G., and Hood, L. (1997) *Science* **278**, 605–607
639. Mathies, R. A., Huang, X. C., and Quesada, M. A. (1992) *Anal. Chem.* **64**, 2149–2154
640. Yeung, E. S., and Li, Q. (1998) in *High Performance Capillary Electrophoresis* (Khaldi, M. G., ed), Wiley, New York
641. Little, D. P., Aaserud, D. J., Valaskovic, G. A., and McLafferty, F. W. (1996) *J. Am. Chem. Soc.* **118**, 9352–9359
642. Maxam, A. M., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560–564
643. Gilbert, W. (1981) *Science* **214**, 1305–1312
644. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560
645. Brownlee, G. G. (1972) in *Determination of Sequences in RNA* (Work, T. S., and Work, E., eds), North-Holland Publ., Amsterdam
646. Pilly, D., Niemeyer, A., Schmidt, M., and Bargetzi, J. P. (1978) *J. Biol. Chem.* **253**, 437–445
647. Bayev, A. A. (1995) *Comprehensive Biochemistry* **38**, 439–479
648. Campbell, P. N. (1995) *Trends Biochem. Sci.* **20**, 259–260
649. Myers, T. W., and Gelfand, D. H. (1991) *Biochemistry* **30**, 7661–7666
650. Simoncsits, A., Brownlee, G. G., Brown, R. S., Rubin, J. R., and Guilley, H. (1977) *Nature (London)* **269**, 833–836
651. Stanley, J., and Vassilenko, S. (1978) *Nature (London)* **274**, 87–89
652. Kitamura, N., and Wimmer, E. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3196–3200
653. Josse, J., Kaiser, A. D., and Kornberg, A. (1961) *J. Biol. Chem.* **236**, 864–875
654. Stormo, G. D. (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 241–263
655. Goad, W. B. (1986) *Ann. Rev. Biophys. Biophys. Chem.* **15**, 79–95
656. Garner, M. M., and Revzin, A. (1986) *Trends Biochem. Sci.* **11**, 395–396
657. Berger, R., Duncan, M. R., and Berman, B. (1993) *BioTechniques* **15**, 650–652
658. Welsh, J., and Cantor, C. R. (1984) *Trends Biochem. Sci.* **9**, 505–508
659. Galas, D., and Schmitz, A. (1978) *Nucleic Acids Res.* **5**, 3157–3170
660. Tullius, T. D. (1987) *Trends Biochem. Sci.* **12**, 297–300
661. Tullius, T. D., and Dombroski, B. A. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5469–5473
662. Prigodich, R. V., and Martin, C. T. (1990) *Biochemistry* **29**, 8017–8019
663. Flaus, A., Luger, K., Tan, S., and Richmond, T. J. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1370–1375
664. Tornaletti, S., and Pfeifer, G. P. (1995) *J. Mol. Biol.* **249**, 714–728

664a. Chaulk, S. G., Pezacki, J. P., and MacMillan, A. M. (2000) *Biochemistry* **39**, 10448–10453

665. Jaeger, J. A., SantaLucia, J., Jr., and Tinoco, I., Jr. (1993) *Ann. Rev. Biochem.* **62**, 255–287
666. Hüttenhofer, A., and Noller, H. F. (1994) *EMBO J.* **13**, 3892–3901
667. Laughrea, M., and Tam, J. (1992) *Biochemistry* **31**, 12035–12041
668. Reid, B. R. (1981) *Ann. Rev. Biochem.* **50**, 969–996
669. Hare, D. R., Ribeiro, N. S., Wemmer, D. E., and Reid, B. R. (1985) *Biochemistry* **24**, 4300–4306
670. Ikuta, S., Chattopadhyaya, R., Ito, H., Dickerson, R. E., and Kearns, D. R. (1986) *Biochemistry* **25**, 4840–4849
671. Lefèvre, J.-F., Lane, A. N., and Jardetzky, O. (1987) *Biochemistry* **26**, 5076–5090
672. Bax, A., and Lerner, L. (1986) *Science* **232**, 960–967
673. Patel, D. J., Shapiro, L., and Hare, D. (1987) *Ann. Rev. Biophys. Biophys. Chem.* **16**, 423– 454
674. Flynn, P. F., Kintanar, A., Reid, B. R., and Drobny, G. (1988) *Biochemistry* **27**, 1191–1197
675. Cohen, J. S. (1987) *Trends Biochem. Sci.* **12**, 133–135
676. Wemmer, D. E., Chou, S.-H., and Reid, B. R. (1984) *J. Mol. Biol.* **180**, 41–60
677. Schmidt, P. G., Sierzputowska-Gracz, H., and Agris, P. F. (1987) *Biochemistry* **26**, 8529–8534
678. Farmer, B. T., Jr., and Müller, L. (1993) *J. Am. Chem. Soc.* **115**, 11040–11041
679. Nikonowicz, E. P., and Pardi, A. (1992) *Nature (London)* **355**, 184–186
680. Tate, S.-i, Ono, A., and Kainosho, M. (1994) *J. Am. Chem. Soc.* **116**, 5977–5978
681. Marino, J. P., Prestegard, J. H., and Crothers, D. M. (1994) *J. Am. Chem. Soc.* **116**, 2205–2206
682. Moore, P. B. (1995) *Acc. Chem. Res.* **28**, 251–256
683. Pardi, A. (1995) *Nature Struct. Biol.* **1**, 846–849
684. Mer, G., and Chazin, W. J. (1998) *J. Am. Chem. Soc.* **120**, 607–608

684a. Nikonowicz, E. P., Michnicka, M., and DeJong, E. (1998) *J. Am. Chem. Soc.* **120**, 3813–3814

684b. Liu, A., Majumdar, A., Hu, W., Kettani, A., Skripkin, E., and Patel, D. J. (2000) *J. Am. Chem. Soc.* **122**, 3206–3210

685. Tolbert, T. J., and Williamson, J. R. (1996) *J. Am. Chem. Soc.* **118**, 7929–7940
686. Xu, J., Lapham, J., and Crothers, D. M. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 44–48
687. Horowitz, J., Ofengand, J., Daniel, W. E. Jr., and Cohn, M. (1977) *J. Biol. Chem.* **252**, 4418–4420
688. Schroeder, S. A., Roongta, V., Fu, J. M., Jones, C. R., and Gorenstein, D. G. (1989) *Biochemistry* **28**, 8292–8303
689. Radhakrishnan, I., Patel, D. J., and Gao, X. (1992) *Biochemistry* **31**, 2514–2523
690. Schurter, E. J., Sayer, J. M., Oh-hara, T., Yeh, H. J. C., Yagi, H., Luxon, B. A., Jerina, D. M., and Gorenstein, D. G. (1995) *Biochemistry* **34**, 9009–9020
691. Mao, H., Deng, Z., Wang, F., Harris, T. M., and Stone, M. P. (1998) *Biochemistry* **37**, 4374–4387
692. Chou, S.-H., Zhu, L., Gao, Z., Cheng, J.-W., and Reid, B. R. (1996) *J. Mol. Biol.* **264**, 981–1001
693. Kettani, A., Guéron, M., and Leroy, J.-L. (1997) *J. Am. Chem. Soc.* **119**, 1108–1115
694. Chu, Wen-C., Liu, J. C.-H., and Horowitz, J. (1997) *Nucleic Acids Res.* **25**, 3944–3949

## Study Questions

1. Describe the typical distribution pattern of RNA and DNA in bacterial cells and in eukaryotic cells.

2. Draw the structures of the Watson–Crick base pairs guanine–cytosine (GC) and adenine–thymine (AT). Also draw the GU pair, which is not a Watson–Crick pair.

3. Draw the tautomeric structures possible for the cation formed by protonation of 9-methyladenine.

4. What unusual base pairs could arise from a minor tautomer of cytosine or from a minor tautomer of guanine?

5. The minor imino tautomeric form of adenosine occurs infrequently in DNA. Can this cause mutations? Explain; draw structures to illustrate your answer.

6. Will the substitution of hypoxanthine for adenine in DNA result in mutation? Explain.

7. Why is the methylation of DNA to form $O^6$-methyl guanine mutagenic?

8. Draw the structure of a dinucleotide that might be obtained by the partial hydrolysis of RNA. Indicate the following:
   a) The 5' end
   b) The 3' end
   c) The torsion angle $\chi$
   d) The point of cleavage by pancreatic ribonuclease
   e) The point of cleavage by periodic acid
   f) Two points at which the structure might be methylated by modifying enzymes acting on a polynucleotide

9. Draw the structure of guanosine-5'-phosphate in such a way that the configurations of the sugar ring and of the glycosidic linkage are clearly indicated. State whether you have drawn a *syn* or an *anti* conformer. Circle the most acidic proton in the guanine ring and indicate its approximate $pK_a$. Which is the most basic center? What is the approximate $pK_a$ of the conjugate acid?

10. What are the chemical functional groups in DNA? In RNA?

11. Electrophoresis of a mixture of the dinucleotides ApC and ApU at pH 3.5 separates two components. Identify these and explain the order of migration. Be as quantitative as possible.

12. Draw the structure of the predominant form of pGpC as it occurs at pH 3.5.

13. Why is DNA denatured at pH 11?

14. Draw a schematic representation of the polynucleotide portion of a DNA molecule and of an RNA molecule and indicate positions of cleavage by the following treatments:
   a) Mild HCl
   b) More vigorous HCl
   c) Mild NaOH
   d) More vigorous NaOH
   e) Pancreatic RNase
   f) Pancreatic DNase
   g) Splenic DNase
   h) Splenic phosphodiesterase
   i) Snake venom phophodiesterase
   j) Dnase from *Micrococcus*

15. A sample of DNA from a virus was hydrolyzed by acid and was found to have the following base composition (in mol%): adenine, 30; thymine, 39; guanine, 18; cytosine, 13. This differs from that of most DNA preparations. Offer a possible explanation. Sketch the expected temperature-absorbance profile of this DNA. Do you expect much hyperchromicity? Explain your answer.

16. Adenine is found to constitute 16.3% of the nucleic acid bases in a sample of bacterial DNA. What are the percentages of the other three bases?

17. For the following DNA sequence

    3'-CGATACGGCTATGCCATAGGC-5'

    write
    a) the sequence of the complementary DNA strand;
    b) the sequence of the corresponding segment of mRNA formed using the DNA segment above as the template;
    c) the amino acid sequence encoded by this segment.

18. What is the molecular mass of a segment of B-DNA that encodes a 386-residue protein? What is the length in nm? in Å? Do not make allowance for introns.

## Study Questions

19. Complete the following table:

| Name | Monomer | Linkage | Range of molecular masses |
|---|---|---|---|
| Protein | | | |
| Polysaccharide | | | |
| Nucleic acid | | | |
| Teichoic acid | | | |
| Poly-β-hydroxybutyrate | | | |

20. What is meant by the $T_m$ of a DNA sample? How does $T_m$ vary with base composition and what is the explanation of this?

21. Isolated "naked" bacterial DNA, from which proteins have been removed, is supercoiled. DNA in the bacterial chromosome is also supercoiled. When naked DNA is nicked, its supercoiling is abolished. In contrast nicking the chromosomal DNA does not abolish its supercoiling. Explain.

22. A closed circular duplex DNA has a 90 base-pair segment of alternating G and C residues. Upon transfer to a solution containing a high salt concentration, this segment undergoes a transition from the B conformation to the Z conformation.
    a) Explain why the high salt concentration induces a B→Z transformation.
    b) What changes would you expect in (1) the linking number *Lk*, (2) the writhe *Wr*, and (3) the twist *Tw* of the DNA as a result of this transition.

23. Name two or more characteristics of a DNA sequence or of its environment that will favor conversion of B-DNA into Z-DNA.

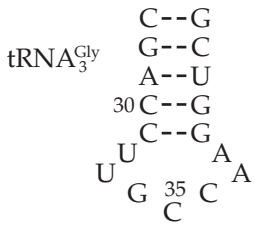
24. Suppose one double helical turn of a superhelical DNA molecule changes from a B conformation to the Z conformation. Calculate the approximate changes in (1) the linking $\Delta Lk$, (2) the writhe $\Delta Wr$, and (3) the twist $\Delta Tw$ of the DNA as a result of this transition. Show your calculations and explain your answers. For this problem assume that the B form of DNA has 10.4 bp per turn. Why is the B→Z transition favored in naturally occurring supercoiled DNA?

25. A circular DNA plasmid of length 1144 bp is supercoiled with a twist (*Tw*) of 110. Assume that the DNA has 10.4 bp per turn in its relaxed state.
    a) What is the linking number *Lk* and the writhe *Wr* in the plasmid?
    b) Is the plasmid negatively or positively supercoiled?
    c) Ethidium bromide is an intercalating agent that inserts between the stacked base pairs, separating the stacks and causing local unwinding that decreases the value of *Tw*. What effect would ethidium bromide have on the migration rate of the plasmid during electrophoresis?
    d) If part of the plasmid were to undergo a transition from B-DNA to Z-DNA, what would be the effect on *Lk*, *Tw*, and *Wr*?

26. You have been given a sample of nucleic acid, describe two ways you could determine whether it is RNA or DNA and two ways to determine whether it is single- or double-stranded.

27. What conclusions can you draw about the nature of the protein binding site on DNA from the observation that methylation of cytosine residues in the protein-recognition sequence inhibits protein binding?

28. The anticodon loop of one of the tRNA Gly molecules from *E. coli* is as follows. Identify the anticodon, reading from 3' to 5'. This tRNA recognizes two different Gly codons. What are they? Write them from 5' to 3'.




The complete tRNA contains 75 nucleotides. Sketch the rest of the molecule in the cloverleaf representation. Label the 5' and 3' ends and the dihydrouridine and TψC loops. What are the last three nucleotides at the 3' end?

29. Are viruses alive? Explain your answer.

Ice and water are in equilibrium at 0°C and atmospheric pressure. When ice melts under these conditions the heat $Q$ absorbed from the surroundings is the **enthalpy change**, $\Delta H$, which equals 6.008 kJ mol$^{-1}$. For a reversible reaction the **entropy change**, $\Delta S$, equals $\Delta H/T$ and the entropy increases when the ice melts by $6.008 \times 10^3$ J / 273.16 = 22.0 JK$^{-1}$. Along the frozen edges of the river the ice and water are not in a true equilibrium but in a steady state. It is an **open system** in which water flows and energy is exchanged with the surroundings. Similarly, the cells of *Vorticella* represent open systems through which water and nutrients flow, and energy is exchanged with the surrounding water. Many chemical reactions within the cells are near equilibrium while the organism maintains its own structure. Photos: Frosty River, Banff Natl. Park, Alberta © Stephen J. Kraseman; *Vorticella* courtesy of Ralph Buchsbaum.

# Contents


281 **A. Thermodynamics**
282 1. The First Law of Thermodynamics
282 2. Enthalpy Changes and Thermochemistry
282 *Processes at constant pressure*
282 *Enthalpies of combustion and physiological fuel values*
284 3. The Second Law of Thermodynamics
284 *The thermodynamic temperature*
285 *Entropies from measurement of heat capacities*
285 4. A Criterion of Spontaneity: The Gibbs Energy
286 5. Practical Thermochemistry
286 *Summing changes in Gibbs energy*
287 *Reactions in solution*
287 *$\Delta G°$ and the equilibrium constant*
288 *Activity coefficients and concentration equilibrium constants*
289 *Changes in equilibria with temperature*
289 6. Thermodynamics and Life Processes
292 **B. Tables of $\Delta G°$ Values for Biochemical Compounds**
292 1. Gibbs Energies of Formation
292 2. Gibbs Energies of Dissociation of Protons
292 3. Group Transfer Potentials
293 4. "Constants" That Vary with pH and Magnesium Ion Concentrations
297 5. A New Standard for Biochemical Thermodynamic?
297 6. Bond Energies and Approximate Methods for Estimation of Thermodynamic Data
297 7. Gibbs Energies of Combustion by $O_2$ and by $NAD^+$
300 **C. Electrode Potentials and Gibbs Energy Changes for Oxidation–Reduction Reactions**
302 **D. The Adenylate System**
302 1. Storage and Utilization of Energy
303 2. Synthesis of ATP
303 3. Creatine Phosphate, an Energy Buffer
303 4. Phosphorus-31 NMR
304 **E. Complex Biochemical Equilibria**
305 1. Effects of pH on Equilbria
305 2. Microscopic Dissociation Constants and Tautomerization
307 3. The Binding of Metal Ions
307 *Factors affecting the strength of binding of a metal in a complex*
311 *How properties of the metal ion affect chelation*
311 *Metal binding sites in cells*
312 *Calcium-binding proteins*

319 **References**
321 **Study Questions**


**Boxes**


295 Box 6-A Measurement of Intracellular pH
299 Box 6-B Magnesium
308 Box 6-C Linear Gibbs Energy Relationships
314 Box 6-D Calcium
317 Box 6-E Metallothioneins


**Tables**


283 Table 6-1 Units of Energy and Work and the Values of Some Physical Constants
283 Table 6-2 Caloric Values of Food Components
285 Table 6-3 Entropies of Selected Substances
290 Table 6-4 Gibbs Energies of Formation and of Oxidation at 25°C for Compounds of Biochemical Interest
293 Table 6-5 Values of p*Ka*, $\Delta G°$, and $\Delta H°$ for Ionization of Acids at 25°C
294 Table 6-6 Gibbs Energies of Hydrolysis at 25°C (in kJ mol$^{-1}$)
298 Table 6-7 Empirical Bond Energies and Resonance Energies
301 Table 6-8 Reduction Potentials of Some Biologically Important Systems
310 Table 6-9 Logarithms of Binding Constants for Some 1:1 Metal Complexes at 25°C
311 Table 6-10 Ionic Radii in Nanometers for Some Metallic and Nonmetallic Ions

# Thermodynamics and Biochemical Equilibria

# 6

We all know from experience the importance of energy to life. We know that we must eat and that hard work not only tires us but also makes us hungry. Our bodies generate heat, an observation that led Lavoisier around 1780 to the conclusion that respiration represented a slow combustion of foods within the body. It soon became clear that respiration must provide the energy for both the mechanical work done by muscles and the chemical synthesis of body constituents. All organisms require energy and the ways in which living things obtain and utilize energy is a major theme of biology.

The discovery of the first and second laws of thermodynamics permitted the development of precise, quantitative relationships between heat, energy, and work. It also allowed **chemical equilibria** to be understood. Modern biochemical literature abounds with references to the thermodynamic quantities **energy** $E$, **enthalpy** $H$, **entropy** $S$, and **Gibbs energy** (also called free energy) $G$. The purposes of this chapter are: (1) to provide a short review of thermodynamic equations, (2) to provide tables of thermodynamic quantities for biochemical substances and to explain the use of these data in the consideration of equilibria in biochemical systems, and (3) to introduce the **adenylate system**, which consists of adenosine triphosphate (ATP), adenosine diphosphate (ADP), adenosine mono-phosphate (AMP), and inorganic phosphate ($P_i$). This system plays a central role in energy metabolism, and (4) to provide a quantitative understanding of the effects of pH and of metal ions on biochemical equilibria. *Many readers will want to go directly to Section D, which deals with the adenylate system and its significance for life.*

## A. Thermodynamics

Thermodynamics is concerned with the quantitative description of heat and energy changes and of chemical equilibria.[1–10] Knowledge of *changes* in thermodynamic quantities, such as $\Delta H$ and $\Delta S$, enables us to predict the equilibrium positions in reactions and whether or not under given circumstances a reaction will or will not take place. Furthermore, the consideration of thermodynamic quantities provides insight into the nature of forces responsible for bonding between molecules, enzymatic catalysis, functioning of DNA and RNA, and many other phenomena.

It is important to realize that while thermodynamic information will tell us whether or not a reaction can take place it says nothing about the rate of the reaction. It will not even say whether a reaction will proceed at all within a given period of time. This has led to the occasional assertion that thermodynamics is not relevant to biochemistry. This is certainly not true; it is important to understand energy relationships in biochemical reactions. At the same time, one should avoid the trap of assuming that thermodynamic calculations appropriate for *equilibrium* situations can always be applied directly to the *steady state* found in a living cell.

Thermodynamics is an exact science and its laws deal with measurable quantities whose values are determined only by the **state** of the **system** under consideration. For example, the system might be the solution in a flask resting in a thermostated bath. To specify its state we would have to say whether it is pure solid, liquid, or gas, or a solution of specified composition and give the temperature and pressure. The flask, the bath, and everything else would be

called **surroundings** or **environment**. The system plus surroundings is sometimes referred to as **the universe**.

## 1. The First Law of Thermodynamics

The first law of thermodynamics asserts the conservation of energy and also the equivalence of **work** and **heat**. Work and heat are both regarded as energy in transit. Heat may be absorbed by a system from the surroundings or evolved by a system and absorbed in the surroundings. Work can be done *by* a system on the surroundings or it can be done *on* a system. The first law postulates that there is an **internal energy** $E$ (also designated $U$), which is dependent only on the present state of the system and in no way is dependent upon the history of the system. The first law states that *E can be changed only by the flow of energy as heat or by work*. In other words, energy can neither be created nor destroyed.

In mathematical form, the first law is given as follows:

$$\Delta E = E\,(\text{products}) - E\,(\text{reactants}) = Q - W \quad (6\text{-}1)$$

Here $Q$ is the heat absorbed by the system from the surroundings and $W$ is the work done by the system on the surroundings. Energy, heat, and work are all measured in the same units. Chemists have traditionally used the **calorie** (cal) or **kilocalorie** (kcal) but are switching to the SI unit, the **joule** (Table 6-1).[10a] Work done by the system may be **mechanical** (e.g., by changing the volume of the surroundings), **electrical** (e.g., by charging of a battery), or **chemical** (e.g., by effecting the synthesis of a polypeptide from amino acids).

## 2. Enthalpy Changes and Thermochemistry

We are most often interested in the *changes* in the thermodynamic functions when a chemical reaction takes place; for example, the heat absorbed by the system within a **bomb calorimeter** where the volume stays constant ($Q_V$) is a direct measure of the change in $E$:

$$Q_V = \Delta E \quad (6\text{-}2)$$

To measure $\Delta E$ for combustion of a biochemical compound, the substance may be placed in a bomb together with gaseous oxygen and the mixture ignited within the calorimeter by an electric spark. In this case, heat will be evolved from the bomb and will pass into the surroundings. $Q_V$ and $\Delta E$ will be negative. The bomb calorimeter is designed to measure $Q_V$ and thereby to give us a way of determining $\Delta E$ for reactions.

***Processes at constant pressure.*** Chemical and biochemical reactions are much more likely to be conducted at constant pressure (usually 1 atm) than they are at constant volume. For this reason, chemists tend to use the **enthalpy *H*** more often than the internal energy $E$.

$$H = E + PV \quad (6\text{-}3)$$

It follows from Eq. 6-3 that *if the pressure is constant*, $\Delta H_P$ is equal to $\Delta E_P + P\,\Delta V$. Since in a process at constant pressure, $P\,\Delta V$ is exactly the pressure–volume work done on the surroundings, the heat absorbed at constant pressure ($Q_P$) is a measure of $\Delta H_P$.

$$Q_P = \Delta E_P + P\,\Delta V = \Delta H_P \quad (6\text{-}4)$$

Since enthalpy changes can be obtained directly from measurement of heat absorption at constant pressure, even small values of $\Delta H$ for chemical and biochemical reactions can be measured using a microcalorimeter.[11,12] Using the technique of pulsed acoustic calorimetry, changes during biochemical processes can be followed on a timescale of fractions of a millisecond. An example is the laser-induced dissociation of a carbon monoxide–myoglobin complex.[13]

The term *enthalpy* was coined to distinguish $H$ from $E$, but we sometimes tend to be careless about language and many discussions of energy in the literature are in fact about enthalpy. The difference is often not significant because if the pressure–volume work is negligible, $E$ and $H$ are the same.

***Enthalpies of combustion and physiological fuel values.*** The **heat of combustion** ($-\Delta H_c$) of an organic substance is usually determined from $\Delta E_c$, which is measured in a bomb calorimeter. Since $\Delta E_V$ and $\Delta E_P$ are nearly identical, it follows that $\Delta H_P = \Delta E_V + P\,\Delta V$. Here $\Delta V$ is the volume change which would have occurred if the reaction were carried out at constant pressure $P$; thus, $\Delta H_P$ can be estimated by calculation. Since $\Delta H$ is desired for combustion to carbon dioxide, water, elemental nitrogen ($N_2$), and sulfur, correction must be made for the amounts of the latter elements converted into oxides. By these procedures, it has been possible to obtain highly accurate values of $\Delta H_c$ both for biochemical compounds and for mixed foodstuffs. In nutrition, $-\Delta H_c$ is sometimes referred to as the **gross energy**. Values are usually expressed in kilocalories (kcal) by chemists but often as **Cal** (with a capital C) in the nutritional literature.

**Caloric values** of foods (physiological fuel values) are enthalpies of combustion but with an opposite sign, ($-\Delta H_c$), and corrected for energy lost in urine (e.g., as urea) and feces. While enthalpies of combustion of foods are all negative, the caloric values are given as positive numbers. Caloric values for proteins are

**TABLE 6-1**
**Units of Energy and Work and the Values of Some Physical Constants**

| | |
|---|---|
| The joule, SI unit of energy | 1 J = 1 kg $m^2$ $s^{-2}$<br>= 1 N m (newton meter)<br>= 1 W s (watt second)<br>= 1 C V (coulomb volt) |
| Thermochemical calorie | 1 cal = 4.184 J |
| Large calorie | 1 Cal = 1 kcal = 4.184 kJ |
| Work required to raise 1 kg 1 m on earth | (at sea level) = 9.807 J |
| Gibbs energy of hydrolysis of 1 mole of ATP | at pH 7, millimolar concentrations = −12.48 kcal = −52.2 kJ |
| Work required to concentrate 1 mole of a substance | 1000-fold, e.g., from $10^{-6}$ to $10^{-3}$ M = 4.09 kcal = 17.1 kJ |
| Avogadro's number, the number of particles in a mole | $N = 6.0220 \times 10^{23}$ |
| Faraday | 1 F = 96,485 C $mol^{-1}$ (coulombs per mole) |
| Coulomb | 1 C = 1 A s (ampere second)<br>= $6.241 \times 10^{18}$ electronic charges |
| The Boltzmann constant | $k_B = 1.3807 \times 10^{-23}$ J $deg^{-1}$ |
| The gas constant, $R = N k_B$ | $R$ = 8.3144 J $deg^{-1}$ $mol^{-1}$<br>= 1.9872 cal $deg^{-1}$ $mol^{-1}$<br>= 0.08206 1 atm $deg^{-1}$ $mol^{-1}$<br>and at 25°C $RT$ = 2.479 kJ $mol^{-1}$ |
| The unit of temperature is the kelvin (K); 0°C = | 273.16 K |
| ln $x$ = 2.3026 log $x$ | |
| One atmosphere (atm) = 101.325 kilopascals (kPa) | |

calculated for the conversion of the nitrogen to urea, the major nitrogenous excretion product in mammals, rather than to elemental nitrogen. Typical values are shown in Table 6-2.

**TABLE 6-2**
**Caloric Values of Food Components**

| Component | Caloric values per gram | |
|---|---|---|
| Carbohydrates | 4.1 kcal | 17 kJ |
| Pure glucose | 3.75 kcal | 15.7 kJ |
| Lipids | 9.3 kcal | 39 kJ |
| Proteins[a] | 4.1 kcal | 17 kJ |

[a] Nitrogen excreted as urea.

From a thermochemical viewpoint, can a human or animal be regarded as just a catalyst for the combustion of foodstuffs? To answer this question, large calorimeters were constructed into which an animal or a human being was placed. If, while in the calorimeter, the subject neither gained nor lost weight, the heat evolved should have been just equal to $-\Delta H$ for combustion of the food consumed to $CO_2$, water, and urea. That this prediction was verifed experimentally does not seem surprising, but at the time that the experiments were first done in the early years of the century there may have been those who doubted that the first law of thermodynamics applied to mammals.

In practice, animal calorimetry is quite complicated because of the inherent difficulty of accurate heat measurements, uncertainties about the amount of food stored, and the necessity of corrections for $\Delta H_c$ of the waste products. However, the measurement of energy metabolism has been of considerable importance in nutrition and medicine. Indirect methods of calorimetery have been developed for use in measuring the **basal metabolic rate** of humans. For a good discussion see White *et al.*[14]

The basal metabolic rate is the rate of heat evolution in the resting, postabsorptive state, in which the subject has not eaten recently. In this condition, stored foods provide the energy and are oxidized at a relatively constant rate. The basal metabolic rate tends to be *proportional to the surface area*; which can be approximated (in units of $m^{-2}$) as 1/60 [height (cm) × mass (kg)].[15] For a young adult female, the basal metabolic rate is typically ~154 kJ $h^{-1}$ $m^{-2}$ and for a young adult male ~172 kJ $h^{-1}$ $m^{-2}$. This is ~320 – 360 kJ $h^{-1}$ for a 70-kg person. Note that 360 kJ $h^{-1}$ is the same as the power output of a 100-watt lightbulb. While there is considerable variation among individuals, basal metabolic rates far below or above normal may indicate a pathological condition such as an insufficiency or oversupply of the thyroid hormone thyroxine. Metabolic rates fall somewhat below the basal value during sleep and are much higher than basal during hard exercise. A human may attain rates as high as 2500 kJ (600 kcal) per hour. At a basal rate of 320 kJ (76 kcal) per hour, a person requires 7680 kJ (1835 kcal) each 24 h to supply his or her basal needs, plus additional energy during periods of muscular exercise. Routine light exercise as in the office or during housework increases metabolism to about double the basal rate. Although the caloric values in Table 6-2 are reliable for prediction of metabolic energy needs, they must be adjusted to predict the efficiency of utilization for growth. In one study[16] a group of rats deposited 28% of the available energy from sucrose as body protein and fats, but fats were deposited with an efficiency of 36%.

## 3. The Second Law of Thermodynamics

Why does heat flow from a warm body into a cold one? Why doesn't it ever flow in the reverse direction? We can see that differences in temperature control the direction of flow of heat, but this observation raises still another question: What *is* temperature? Reflection on these questions, and on the interconversion of heat and work, led to the discovery of the second law of thermodynamics and to the definition of a new thermodynamic function, the **entropy *S***.

Consider the melting of ice. This is a phase transition that usually takes place at constant temperature and pressure. At a temperature just above 0°C ice melts completely, but at a temperature just below 0°C it does not melt at all. At 0°C we have an equilibrium. In the language of thermodynamics, the melting of ice at 0°C is a **reversible reaction**. What criterion could be used to predict this behavior for water? For many familiar phenomena, e.g., combustion, a spontaneous reaction is accompanied by the evolution of a large amount of heat, i.e., $\Delta H$ is negative. However, when ice melts it *absorbs* heat. The $\Delta H$ of fusion amounts to 6.008 kJ mol$^{-1}$ at 0°C and is nearly the same just below 0°C, where the ice does not melt, and just above 0°C, where the ice melts completely. In the latter case, the melting of ice is a spontaneous reaction for which $\Delta H$ is positive. It is clear from such facts that *the sign of the enthalpy change does not serve as a criterion of spontaneity*.

A correct understanding of the ice–water transition came when it was recognized that when ice melts not only does $H$ increase by 6.008 kJ mol$^{-1}$, as the molecules acquire additional internal energy of translation, vibration, and rotation, but also *the molecules become more disordered*. Although historically entropy was introduced in a different context, it is now recognized to be a measure of "microscopic disorder." When ice melts, the entropy $S$ increases because the structure becomes less ordered.

The second law of thermodynamics is stated in many different ways, but the usual mathematical formulation asserts that for the universe (or for an isolated system)

$$\Delta S \text{ (system + surroundings)} = 0 \quad \text{for reversible processes}$$

$$\Delta S > 0 \quad \text{for real (nonreversible) processes} \tag{6-5}$$

The second law is sometimes stated in another way: *The entropy of the universe always increases*.

The second law also defines both $S$ and the thermodynamic temperature scale as follows:

$$dS_{\text{reversible}} = q/T \tag{6-6}$$

Here $q$ is an infinitesimal quantity of heat absorbed from the surroundings by the system and $T$ is measured in kelvins (K). For a *reversible phase transition* such as the melting of ice at constant pressure and temperature, the change in entropy of the $H_2O$ is just $\Delta H/T$.

$$\Delta S)_{P,T,\text{reversible}} = Q/T = \Delta H/T \tag{6-7}$$

Entropy is measured in units of joules per kelvin (or °C) or calories per K, the latter sometimes being abbreviated as e. u. (entropy units). Since the melting of ice at 0°C is a reversible process, the second law asserts that the entropy of the surroundings decreases by the same amount that the entropy of the water increases. The value of $T\,\Delta S$ is numerically equal to the heat of fusion, 6.008 kJ mol$^{-1}$ in the case of water at 0°C. Thus, the entropy increase in the ice as it melts at 0°C is 6.008 x $10^3$ J/273.16 K = 22.0 J K$^{-1}$.

***The thermodynamic temperature.*** The definition of thermodynamic temperature in kelvins (Eq. 6-8) also follows from Eq. 6-6. See textbooks of thermodynamics for further treatment.

$$T = (\partial E/\partial S)_V = (\partial H/\partial S)_P \tag{6-8}$$

The entropy of a substance can be given a precise mathematical formulation involving the degree of molecular disorder (Eq. 6-9).

$$S = k_B \ln \Omega \tag{6-9}$$

Here $k_B$ is the **Boltzmann constant** (see Table 6-1) and $\Omega$ is given precisely as the number of microscopic states (different arrangements of the particles) of the system corresponding to a given macroscopic state, i.e., to a given temperature, pressure, and quantity. It increases as volume or temperature is increased and in going from solid to liquid to gaseous states. Equation 6-9 is not part of classical thermodynamics (which deals only with macroscopic systems, i.e., with large collections of molecules). However, using the methods of statistical thermodynamics,[17] this equation can be used to predict the entropies of gases.

The *racemization of an amino acid* provides a biochemical example that can be related directly to Eq. 6-9. A solution of an L-amino acid will be efficiently changed into the racemic mixture of 50% D and 50% L by the action of an enzyme (a **racemase**) with no uptake or evolution of heat. Thus, $\Delta H = 0$ and the only change is an entropy change. Let us designate $\Omega$ for the pure isomer as $\Omega'$. Since there are just two choices of configuration for each of the N molecules in 1 mole of the racemate we see that for the racemate

$$\Omega = 2^N \Omega' \tag{6-10}$$

Applying Eq. 6-9 we calculate $\Delta S$ as follows:

$$\Delta S = k_B (\ln 2^N + \ln \Omega') - k_B \ln \Omega'$$
$$= Nk_B \ln 2 = R \ln 2 = 5.76 \text{ J K}^{-1} \text{ mol}^{-1} \quad (6\text{-}11)$$

***Entropies from measurement of heat capacities.*** It follows from Eq. 6-9 that $S = 0$ when $T = 0$ for a perfect crystalline substance in which no molecular disorder exists. The *third law of thermodynamics* asserts that as the thermodynamic temperature $T$ approaches 0 K the entropy $S$ also approaches zero for perfect crystalline substances. From this it follows that at any temperature above 0 K, the entropy is given by Eq. 6-12.

$$S = \int_0^T C_p \, d \ln T \quad (6\text{-}12)$$

In this equation $C_p$ is the heat capacity at constant pressure:

$$C_P = (\partial H/\partial T)_P \quad (6\text{-}13)$$

If $C_p$ is measured at a series of low temperatures down to near zero K, Eq. 6-12 can be used to evaluate the absolute entropy $S$. If phase transitions occur as the temperature is raised, entropy increments given by Eq. 6-7 must be added to the value of $S$ given by Eq. 6-12. For a few compounds, such as water (Chapter 2), molecular disorder is present in the crystalline state even at 0 K. For these substances a term representing the entropy at 0 K must be added to Eq. 6-12.

The entropies of a few substances are given in Table 6-3. Notice how the entropy increases with increasing complexity of structure, with transitions from solid to liquid to gas, and with decreasing hardness of solid substances.

Measurements of $C_p$ versus temperature for solutions of macromolecules or for biological membranes (Fig. 8-9) over a narrower temperature range are also of interest. These can be obtained with a **differential scanning calorimeter**[18,19] or by an indirect procedure.[20] Denaturation of polymers or phase changes in membranes may be observed. Larger values of $C_p$ are observed for open, denatured, or random-coil structures that are usually present at higher temperatures than for tightly folded molecules.

**TABLE 6-3**
**Entropies of Selected Substances[a]**

| Substance | State[b] | Entropy $S$ | |
|---|---|---|---|
| | | cal K$^{-1}$ mol$^{-1}$ | J K$^{-1}$ mol$^{-1}$ |
| C (diamond) | s | 0.55 | 2.3 |
| C (graphite) | s | 1.36 | 5.7 |
| Cu | s | 8.0 | 33 |
| Na | s | 12.2 | 51 |
| $H_2O$ (ice) | s | 9.8 | 41 |
| $H_2O$ | l | 16.7 | 70 |
| $H_2O$ | g (1 atm) | 45.1 | 189 |
| He | g | 30.1 | 126 |
| $H_2$ | g | 31.2 | 131 |
| $N_2$ | g | 45.8 | 192 |
| $CO_2$ | g | 51.1 | 214 |
| Benzene | g | 64.3 | 269 |
| Cyclohexane | g | 71.3 | 298 |

[a] All values are given in entropy units (e.u.) of calories per Kelvin per mole and in joules per Kelvin per mole at 25°C (298.16 K).
[b] Here s stands for solid, l for liquid, and g for gaseous.

## 4. A Criterion of Spontaneity: The Gibbs Energy

We have seen that while many spontaneous processes, e.g., combustion of organic compounds, are accompanied by liberation of heat (negative $\Delta H$), others are accompanied by absorption of heat from the surroundings (positive $\Delta H$). An example of the latter is the melting of ice at a temperature just above 0°C, during which there is a large increase in the entropy of the water. As we have seen, at 0°C at equilibrium $T\,\Delta S$ is just equal to $-\Delta H$ (Eq. 6-7).

The recognition that $\Delta H - T\,\Delta S = 0$ for a system at equilibrium led J. W. Gibbs to realize that the proper thermodynamic function for determining the spontaneity of a reaction is what is now known as the **Gibbs energy** or Gibbs function **G** (Eq. 6-14).

$$G = H - TS \quad (6\text{-}14)$$

In the older literature the Gibbs energy was usually called the **free energy** or Gibbs free energy and was often given the symbol $F$. For a process at constant temperature and pressure the change in $G$ is given by Eq. 6-15 in which all quantities refer to the system.

$$\Delta G_{T,P} = \Delta H - T\,\Delta S \quad (6\text{-}15)$$

For a reversible (equilibrium) process doing only pressure–volume work:

$$\Delta G)_{T,\text{reversible}} = \Delta H - T\,\Delta S = 0 \quad (6\text{-}16)$$

It can also be shown readily that $\Delta G$ *is negative for any spontaneous* (irreversible) *process*. Such a process is called **exergonic**. Likewise, if $\Delta G$ is positive, a given

reaction will *not* proceed spontaneously and is called **endergonic**. The magnitude of the decrease in the Gibbs energy ($-\Delta G$) is a direct measure of the maximum work which could be obtained from a given chemical reaction if that reaction could be coupled in some fashion reversibly to a system able to do work. It represents the maximum amount of electrical work that could be extracted or the maximum amount of muscular work or osmotic work obtainable from a reaction in a biological system. In any real system, the amount of work obtainable is necessarily less than $-\Delta G$ because real processes are irreversible, i.e., entropy is created.

Returning to the older assumption that the magnitude of $\Delta H$ might be an index of work obtainable, we note that $T\,\Delta S$ amounts to only a few kilojoules for most reactions. Therefore, if $\Delta H$ is large, as in the combustion of foodstuffs, it is not greatly different from $\Delta G$ for the same process. Therefore, we can justify use of the caloric value of a food as an approximate measure of the work obtainable from its metabolism in the body.

## 5. Practical Thermochemistry

For thermodynamic data to be useful in chemical calculations, we must agree upon **standard states** for elements and compounds. If we wish to talk about the change in the Gibbs energy that occurs when one or more pure compounds are converted to other pure substances, we must agree upon a state (crystalline, liquid, gaseous, or in solution) and upon a pressure (especially when gases are involved) at which the data apply. The standard pressure is usually 1 atm. Standard states of the elements are the pure *crystalline*, *solid*, or *gaseous materials*, e.g., C (graphite), S (crystalline, rhombic), P (crystalline, white), and $N_2$, $O_2$, and $H_2$ (gaseous). It is also essential to specify the temperature. Thermodynamic data are most often given for 25°C, but there is a standard state for each substance at each temperature.

It is usually impractical to measure the values of $G$ or $H$, but $\Delta G$ and $\Delta H$ for a chemical reaction can be evaluated. Changes in Gibbs energy can be calculated from tables of $\Delta G$ for formation of compounds from the elements (Eq. 6-17). These values of $\Delta G_f$ can be obtained experimentally by measuring $\Delta H$ of combustion for the compound of interest and for $H_2$, elemental carbon, and other elements present in the compound and also by obtaining entropies from heat capacity measurements. Many other tabulated $\Delta G_f$ values have been obtained indirectly utilizing data from equilibrium constants. The resulting **standard Gibbs energies of formation** are given the symbol $\Delta G_f°$. *The values of $\Delta G_f°$ for the elements in their standard states are all exactly zero.*

***Summing changes in Gibbs energy.*** A convenient feature of thermodynamic calculations is that if two or more chemical equations are summed, $\Delta G$ for the resulting overall equation is just the sum of the $\Delta G$'s for the individual equations as illustrated in Eqs. 6-17 to 6-20. The same applies for $\Delta H$ and $\Delta S$.

$$CH_3COOH\ (l) \longrightarrow 2\,C + 2\,H_2 + O_2$$
$$-\Delta G_f° \text{ for acetic acid} = +396.4 \text{ kJ mol}^{-1} \quad (6\text{-}17)$$

$$2\,O_2 + 2\,C \longrightarrow 2\,CO_2$$
$$2\,\Delta G_f° \text{ for } CO_2 = -788.8 \text{ kJ mol}^{-1} \quad (6\text{-}18)$$

$$O_2 + 2\,H_2 \longrightarrow 2\,H_2O$$
$$2\,\Delta G_f° \text{ for } H_2O = -474.4 \text{ kJ mol}^{-1} \quad (6\text{-}19)$$

$$CH_3COOH\ (l) + 2\,O_2 \longrightarrow 2\,CO_2 + 2\,H_2O\ (l)$$
$$\Delta G_c° \text{ for acetic acid} = -866.8 \text{ kJ mol}^{-1} \quad (6\text{-}20)$$

In this example an equation for the decomposition of acetic acid into its elements (Eq. 6-17) has been summed with Eqs. 6-18 and 6-19, which represent the formation of the proper number of molecules of $CO_2$ and $H_2O$ from the elements. The sum of the three equations gives the equation for the combustion of acetic acid to $CO_2$ and water, and the sum of the $\Delta G$ values for the three equations gives $\Delta G$ for combustion of acetic acid. The resulting value of $\Delta G$ is for combustion of pure liquid acetic acid by oxygen at 1 atm to give $CO_2$ at 1 atm and pure liquid water, all reactants and products being in their standard states.

The process described in the preceding paragraph is represented by Eq. 6-21, which is a general equation for calculation of $\Delta G°$ for any reaction from $\Delta G_f°$ of products and reactants.

$$\Delta G° = \Sigma\Delta G_f°\,(\text{products}) - \Sigma\Delta G_f°\,(\text{reactants}) \quad (6\text{-}21)$$

How does the change in Gibbs energy vary if we go from the standard state of a compound to some other state? Consider a change of pressure in a gas. It is easy to show (see any thermodynamics text) that

$$(\partial G/\partial P)_T = V \quad (6\text{-}22)$$

Using Eq. 6-22 together with the perfect gas law, we obtain the relationship (Eq. 6-23) between the Gibbs energy $\overline{G}$ of one mole of a substance at pressure $P$ and the standard Gibbs energy $\overline{G}°$ at pressure $P°$.

$$\overline{G} - \overline{G}° = RT \ln \frac{P}{P°} = RT\ \Delta \ln P \quad (6\text{-}23)$$

Here the bar over the symbol $G$ indicates that the Gibbs energy is for one mole of substance. Since $P°$ is by definition 1 atm, the Gibbs energy change per mole upon changing the pressure from $P°$ to $P$ is just $RT \ln P$.

***Reactions in solution.*** It is customary in books on thermodynamics to develop most of the important thermodynamic equations as applied to a perfect gas, but we will move at this point to a consideration of biochemical substances in solution. Biochemists are usually interested in the behavior of substances dissolved in relatively dilute aqueous solutions but also in cytoplasm, in which some concentrations may be very high. Sometimes the interest may be in nonaqueous solutions. In any case, it is necessary to establish a standard state for the solute. The standard state of a substance in aqueous solution is customarily taken as a strictly hypothetical one **molal** solution (one mole of solute per kilogram of water) *whose properties are those of a solute at infinite dilution*. An equation exactly analogous to Eq. 6-23 can be written relating the Gibbs energy of one mole of dissolved solute $\overline{G}_i$ to the Gibbs energy $\overline{G}_i°$ in the hypothetical standard state of unit activity and to the **activity** $a_i$ of the solute (Eq. 6-24).

$$\overline{G}_i = \overline{G}_i° + RT \ln a_i \tag{6-24}$$

Here the subscript $i$ designates a particular component in a solution which also contains solvent and, perhaps, other components. To be precise, $\overline{G}_i$ is a *partial molar Gibbs energy*, i.e., the changes in total Gibbs energy of a very large volume of solution when one mole of the component is added.

From Eq. 6-24 it follows that the *Gibbs energy change for dilution* from one activity $a_1$ to another $a_2$ is:

$$\Delta\overline{G} \text{ (dilution from } a_1 \text{ to } a_2) = RT \ln (a_2/a_1) \tag{6-25}$$

Equation 6-24 and the equations that follow from it apply to molal activities. However, the concentration can be substituted for activity in very dilute solution where the behavior of the dissolved molecules approximates that of the hypothetical ideal solution for which the standard state is defined. For any real solution, the activity can be expressed as the product of an activity coefficient and the concentration (Eq. 6-26).

$$a = \gamma c \tag{6-26}$$

where $a$ = activity, $\gamma$ = activity coefficient, and $c$ = molal concentration. Thus, to use the tabulations of thermodynamic functions for substances in solution to predict behavior in other than very dilute solution, we must multiply the concentration of every component by the appropriate activity coefficient. For the approximate calculations which are often of interest to biochemists, it is customary to equate concentration with activity. Furthermore, in dilute solutions, the more usual **molar** concentrations (moles per liter) are nearly equal to molal concentrations. The same equations can be used with *mole fractions* rather than molal concentrations.

***ΔG° and the equilibrium constant.*** Consider the following generalized chemical equation (Eq. 6-27) for reaction of $a$ moles of A with $b$ moles of B to give products C and D, etc.

$$a\,\mathrm{A} + b\,\mathrm{B} + \cdots = c\,\mathrm{C} + d\,\mathrm{D} \cdots \tag{6-27}$$

The standard Gibbs energy change $\Delta G°$ for the process is given by Eq. 6-28:

$$\Delta G° = c\,\overline{G}°(\mathrm{C}) + d\,\overline{G}°(\mathrm{D}) + \cdots - a\,\overline{G}°(\mathrm{A}) - b\,\overline{G}°(\mathrm{B}) \cdots \tag{6-28}$$

The symbol $G°$ (A) designates the Gibbs energy of A, etc. The value of $\Delta G$ for any desired concentrations of reactants or products can be related to this $\Delta G°$ by applying to each component Eq. 6-24 with the following result:

$$\Delta G = \Delta G° + RT \ln \frac{a_C{}^c\, a_D{}^d \cdots}{a_A{}^a\, a_B{}^b \cdots} \tag{6-29}$$

Here $a_C$ represents the activity of component C, etc. This useful equation permits us to calculate $\Delta G$ for the low concentrations usually found in biochemical systems. These are more often in the millimolar range or less rather than approaching the hypothetical 1 M of the standard state. Often concentrations are substituted in Eq. 6-29 for activities:

$$\Delta G \approx \Delta G° + RT \ln \frac{[\mathrm{C}]^c[\mathrm{D}]^d}{[\mathrm{A}]^a[\mathrm{B}]^b} \tag{6-30}$$

Equation 6-29 is used in another way by noting that $\Delta G = 0$ when a system is at equilibrium and that at equilibrium the product $a_C{}^c\, a_D{}^d \ldots / a_A{}^a\, a_B{}^b \ldots$ is just the equilibrium constant $K$. It follows that

$$\begin{aligned} \Delta G° &= -RT \ln K = -2.303RT \log K \\ &= -19.145T \log K \text{ J mol}^{-1} \\ &= -5.708 \log K \text{ kJ mol}^{-1} \text{ at } 25°\mathrm{C} \\ &= -1.364 \log K \text{ kcal mol}^{-1} \text{ at } 25°\mathrm{C} \end{aligned} \tag{6-31}$$

Although the units of $\Delta G°$ are kJ mol$^{-1}$, the Gibbs energy change in Eq. 6-31 is that for the reactions of $a$ moles of

A, *b* moles of B, etc., as in the equation used to define *K* (Eq. 6-27 in this instance). It is also important to realize that the log term in Eq. 6-31 must be unitless. Although we usually write ln *K* or log *K*, *K* here represents *K* / *Q*, where *Q* = 1 because it has the same form as *K* but with all components in their standard states. Since the units of *K* and *Q* are the same, log *K* is unitless. Similar considerations apply to expressions of *K* in exponential form.

***Activity coefficients and concentration equilibrium constants.*** Strictly speaking, Eq. 6-31 applies only to thermodynamic equilibrium constants –that is, to constants that employ activities rather than concentrations. The experimental determination of such constants requires measurements of the apparent equilibrium constant or **concentration equilibrium constant**[21] $K_c$ at a series of different concentrations and extrapolation to infinite dilution (Eq. 6-32).

$$K_c = \text{concentration equilibrium constant} = \frac{[C]^c[D]^d}{[A]^a[B]^b} \text{ at equilibrium} \quad (6\text{-}32)$$

Extrapolation of $K_c$ to infinite dilution to give *K* is usually easy because the activity coefficients of most ionic substances vary in a regular manner with **ionic strength** and follow the **Debye–Hückel** equation (Eq. 6-33) in very dilute solutions (ionic strength < 0.01).

$$\log \gamma = -0.509 z_1 z_2 \sqrt{\mu} \quad (6\text{-}33)$$

The integers $z_1$ and $z_2$ are the numbers of charges (valences) for the cation and anion of the salt. The ionic strength (μ, or I) is evaluated as follows:

$$\mu = \tfrac{1}{2} \Sigma_i c_i z_i^2 \quad (6\text{-}34)$$

Here $c_i$ are the molar concentrations of the ions. The summation is carried out over all the ions present. The activity coefficient γ (Eq. 6-33) is the mean activity coefficient for both the cation and anion.

Equation 6-33 suggests that extrapolation of equilibrium constants to infinite dilution is done appropriately by plotting log $K_c$ vs $\sqrt{\mu}$. For example, Fig. 6-1 shows plots of p$K'_a$ for dissociation of $H_2PO_4^-$, $AMP^-$, and $ADP^{2-}$, and $ATP^{3-}$ vs $\sqrt{\mu}$. The variation of p$K'_a$ with $\sqrt{\mu}$ at low concentrations (Eq. 6-35) is derived by application of the Debye-Hückel equation (Eq. 6-33):

$$pK_a' = pK_a - 0.509\,(z_A^2 - z_{HA}^2)\,\sqrt{\mu} \quad (6\text{-}35)$$

Straight lines of slope – 0.509 ($z_A^2 - z_{HA}^2$) are expected. The observed (negative) slopes (Fig. 6-1) are ~1.5 for $H_2PO_4^-$ and $AMP^-$, ~2.5 for $ADP^{2-}$, and ~3.5 for $ATP^{3-}$. The data over the entire range of ionic strength are fitted by empirical relationships of the type of Eq. 6-35a:

$$pK_a' = pK_a - a\sqrt{\mu} + b\mu \quad \text{for } \mu < 0.2 \quad (6\text{-}35a)$$

in which *a* and *b* are empirically determined constants. For example, for $H_2PO_4^-$ *a* = 1.52 and *b* = 1.96. The value of p$K_a$ found was 7.18, about 0.22 greater than the value at μ = 0.2, an ionic strength more commonly used in the laboratory and close to that found in tissues. Note that the difference between the extrapolated p$K_a$ for $ATP^{3-}$ of 7.68 and the observed value of ~ 7.04 at μ = 0.2 is even greater. Serious errors can be introduced into calculations by using extrapolated values for *K* for solutions of appreciable ionic strength. The errors will be maximal for ions of high charge type such as $ATP^{3-}$ and $ATP^{4-}$.

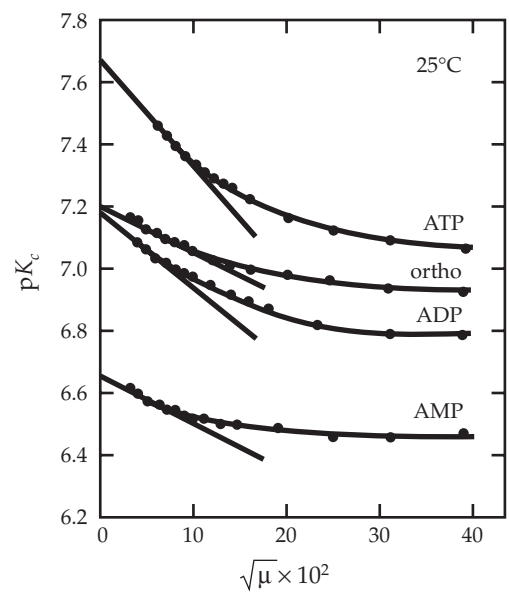



**Figure 6-1** The apparent p$K_a$ values for secondary ionizations of AMP, ADP, ATP, and $H_3PO_4$ (abbreviated ortho) plotted against $\sqrt{\mu}$. Temperature: 25°C. From R. C. Phillips *et al.*[22]

Another problem with equilibrium constants for reactions that use or produce hydrogen ions is that there is no rigorous relationship between pH and $a_{H^+}$ or [$H^+$]. Indeed, the concept of an activity of a single ion has little meaning in thermodynamics. Nevertheless, in the pH range of interest to biochemists, results that are very close to those obtained by more rigorous methods are achieved by assuming that the pH meter

responds to hydrogen ion activity. The almost universal practice of biochemists is to assume the pH meter reading obtained with a glass electrode equal to – log $a_{H^+}$ and to substitute the value of $a_{H^+}$ so obtained for $[H^+]$ in defining the concentration equilibrium constant, $K_c$. It is also customary in most branches of chemistry to use values of equilibrium constants and of Gibbs energy which have *not* been extrapolated to $\mu = 0$. Thus most values of $K$ and $\Delta G$, including many of those in this book, are actually of $K_c$ and $\Delta G_c$. An international Commission on Biothermodynamics[21] recommended that values of $K_c$ be measured with the lowest effective buffer concentration and that the ionic strength be brought to 0.1 with KCl.

***Changes in equilibria with temperature.*** At constant pressure $\Delta G$ varies with absolute temperature as follows:

$$\frac{d(\Delta G/T)}{dT} = -\Delta H/T^2 \tag{6-36}$$

The corresponding variation in $K$ is described by the **van't Hoff equation**:

$$\frac{d \ln K}{dT} = \frac{\Delta H^\circ}{RT^2}$$

or

$$\frac{d \ln K}{d(1/T)} = \frac{-\Delta H^\circ}{R}$$

or

$$\Delta H^\circ(\text{kJ mol}^{-1}) = \frac{-0.01914\ d \log_{10} K}{d(1/T)} \tag{6-37}$$

If $\Delta H^\circ$ can be assumed constant over the temperature range of an experiment, a plot of $\ln K$ vs $1/T$ provides a convenient estimate of $\Delta H^\circ$ (or $\Delta H'$ if $\ln K'$ is plotted). The slope of the line will be $-\Delta H^\circ/R$. Since $\Delta G^\circ$ can be calculated from $K$, the method also permits evaluation of $\Delta S^\circ$ using Eq. 6-15. However, unless great care is taken the method is of low accuracy[23] and it is preferable to establish $\Delta H$ by direct calorimetry. Also, especially for proteins, the assumption that $\Delta H^\circ$ is constant over a significant temperature range may be erroneous.

From observations at only two temperatures, $T_1$, and $T_2$, Eq. 6-37 becomes

$$\Delta H^\circ = R \ln (K_2/K_1)[T_1T_2/(T_2 - T_1)]\ \text{kJ mol}^{-1} \tag{6-37a}$$

## 6. Thermodynamics and Life Processes

Can thermodynamics be applied to living organisms? Classical thermodynamics deals with equilibria, but living beings are never in equilibrium. The laws of thermodynamics are usually described as statistical laws. How can such laws apply to living things, some of which contain all of their genetic information in a single molecule of DNA?[24,25] The ideal, reversible reactions of classical thermodynamics occur at infinitesimal speeds. How can thermodynamics be applied to the very rapid chemical reactions that take place in organisms? One answer is that thermodynamics can be used to decide *whether or not a reaction is possible* under given conditions. Thus, if we know the steady-state concentrations of reactants and products within a cell, we can state whether a reaction will or will not tend to go in a given direction.

We may still ask whether there are generalities comparable to the laws of thermodynamics that apply to the kind of steady state or "dynamic equilibrium" that exists in organisms. Lars Onsager showed that such relationships can be found for conditions that are *near equilibrium*. Ilya Prigogine and associates extended Onsager's findings and showed that under conditions that are *far from equilibrium* the system tends to become unstable and to spontaneously develop new structures, which Prigogine calls **dissipative structures**.[26–28] Vortices in flowing water, tornados and hurricanes are examples of dissipative structures. The maintenance of dissipative structures depends upon a flow of energy and matter through the system. The flow of energy is sufficient to "organize" a system. Thus, if a flask of water is placed on a hot plate, a cycle is established. The water moves via cyclic convection currents that develop as a result of the flow of energy through the system. Morowitz reckoned that the $6 \pm 3 \times 10^{18}$ kJ/year of solar energy that falls on the earth supplies the organizing principle for life.[28] Just as it drives the great cycles within the atmosphere and within the seas, it gives rise to the branching and interconnecting cycles of metabolism. This idea may even make the spontaneous development of the organized systems that we call life from inanimate precursors through evolution seem a little more understandable.

A characteristic of this nonequilibrium or irreversible thermodynamics is that time is explicitly introduced. Furthermore, open systems, in which materials and energy flow into and out of the system, are considered. Clearly, a living organism is an open system not a closed one of classical thermodynamics. Because of the flow of materials concentration gradients are set up and transport phenomena often become of primary importance. Articles and books that provide an introduction to nonequilibrium thermodynamics and to the literature in the field include the following.[10,26,28–34] Whether these methods can be applied in a practical way to metabolic systems has been debated.[35,36]

## TABLE 6-4
## Gibbs Energies of Formation and of Oxidation at 25°C for Compounds of Biochemical Interest[a,b]

| Compound | Formula | $\Delta G_f°$ (kJ mol$^{-1}$) | $\Delta G_c°$ (kJ mol$^{-1}$) | For oxidation by $NAD^+$: $\Delta G°_{ox}$ (kJ mol$^{-1}$) | For oxidation by $NAD^+$: $\Delta G°_{ox}$ (pH 7) (kJ mol$^{-1}$) | Number of electrons |
|---|---|---|---|---|---|---|
| Acetaldehyde | $C_2H_4O$ | –139.7 | –1123.5 | 171.5 | –28.3 | 10 |
| Acetic acid | $C_2H_4O_2$ | –369.4 | –866.8 | 169.2 | 9.3 | 8 |
| Acetate$^-$ | | –369.2 | –894.0 | 142.0 | 22.1 | 8 |
| Acetyl-CoA | | –374.1* | –889.1* | 146.9* | –13.0* | 8 |
| Acetyl-*P* | | –1218.4 | –901.7 | 134.3 | –25.6 | 8 |
| Acetylene[c] | $C_2H_2$ | 209.2 | –1235.2 | 59.8 | –140.0 | 10 |
| Acetoacetate$^-$ | $C_4H_5O_3^-$ | –493.7 | –1795.4 | 276.5 | –3.2 | 16 |
| Acetone | $C_3H_6O$ | –161.2 | –1733.6 | 338.4 | 18.7 | 16 |
| *cis*-Aconitate$^{3-}$ | $C_6H_3O_6^{3-}$ | –920.9 | –2157.0 | 173.9 | –65.8 | 18 |
| L-Alanine | $C_3H_7O_2N$ | –371.3 | –1642.0 | 300.4 | 0.8 | 15 |
| L-Asparagine | $C_4H_8O_3N_2$ | –526.6 | –1999.7 | 331.2 | –28.4 | 18 |
| L-Aspartate$^-$ | $C_4H_6O_4N^-$ | –700.7 | –1707.0 | 235.4 | –24.3 | 15 |
| *n*-Butanol | $C_4H_{10}O$ | –171.8 | –2591.7 | 516.2 | 36.7 | 24 |
| *n*-Butyric acid | $C_4H_8O_2$ | –380.2 | –2146.1 | 443.8 | 44.2 | 20 |
| *n*-Butyrate$^-$ | $C_4H_7O_2^-$ | –352.6 | –2173.7 | 416.2 | 56.6 | 20 |
| Butyryl-CoA | | –357.5* | –2168.8* | 421.1* | 21.5* | 20 |
| Caproate$^-$ | $C_6H_{11}O_2^-$ | –329.7 | –3459.7 | 684.1 | 84.7 | 32 |
| $CO_2$ (g) | | –394.4 | 0.0 | 0.0 | 0.0 | 0 |
| $CO_2$ (aq) | | –386.2 | –8.2 | –8.2 | –8.2 | 0 |
| $HCO_3^-$ | | –587.1 | –44.5 | –44.5 | –4.6 | 0 |
| CO (g) | | –137.3 | –257.1 | 1.9 | –38.1 | 2 |
| Citrate$^{3-}$ | $C_6H_5O_7^{3-}$ | –1166.6 | –2148.4 | 182.5 | –57.3 | 18 |
| Creatine | $C_4H_9O_2N_3$ | –264.3 | –2380.6 | 338.8 | –80.8 | 21 |
| Creatinine | $C_4H_7ON_3$ | –28.9 | –2378.8 | 340.6 | –79.0 | 21 |
| Crotonate$^-$ | $C_4H_5O_2^-$ | –275.7 | –2013.4 | 317.5 | –2.1 | 18 |
| Cysteine | $C_3H_7O_2NS$ | –339.8 | –2178.3 | 541.1 | 121.5 | 21 |
| Cystine | $C_6H_{12}O_4N_2S_2$ | –665.3 | –4133.8 | 1046.0 | 246.9 | 40 |
| Dihydroxyacetone-*P*[d] | | –1293.2 | –1458.4 | 95.5 | –144.2 | 12 |
| Erythrose 4-*P*[d] | | –1439.1 | –1944.1 | 127.8 | –191.9 | 16 |
| Ethanol | $C_2H_6O$ | –181.5 | –1318.8 | 235.1 | –4.6 | 12 |
| Ethylene (g)[c] | $C_2H_4O$ | 68.1 | –1331.3 | 222.7 | –17.1 | 12 |
| Formaldehyde | $CH_2O$ | –130.5 | –501.0 | 16.9 | –63.0 | 4 |
| Formic acid | $CH_2O_2$ | –356.1 | –275.5 | –16.5 | –56.5 | 2 |
| Formate$^-$ | $CHO_2^-$ | –350.6 | –281.0 | –22.0 | –22.0 | 2 |
| Fructose | $C_6H_{12}O_6$ | –915.4 | –2874.1 | 233.8 | –245.7 | 24 |
| Fructose 6-*P*[d] | | –1758.3 | –2888.1 | 219.8 | –259.7 | 24 |
| Fructose di-*P*[d] | | –2600.8 | –2902.5 | 205.4 | –274.1 | 24 |
| Fumaric acid | $C_4H_4O_4$ | –647.1 | –1404.8 | 149.2 | –90.6 | 12 |
| Fumarate$^-$ | $C_4H_3O_4^-$ | –604.2 | –1447.7 | 106.2 | –93.6 | 12 |
| α-D-Galactose | $C_6H_{12}O_6$ | –923.5 | –2865.9 | 242.0 | –237.5 | 24 |
| α-D-Glucose | $C_6H_{12}O_6$ | –917.2 | –2872.2 | 235.6 | –243.8 | 24 |
| Glucose 6-*P* | | –1760.3 | –2886.0 | 221.8 | –257.6 | 24 |
| L-Glutamate$^-$ | $C_5H_8O_4N^-$ | –696.8 | –2342.5 | 376.9 | –2.7 | 21 |
| L-Glutamine | $C_5H_{10}O_3N_2$ | –524.8 | –2633.1 | 474.8 | –4.7 | 24 |
| 3-*P* glycerate$^{-d}$ | | –1515.7 | –1235.9 | 59.0 | –100.8 | 10 |
| 2-*P* glycerate$^{-d}$ | | –1509.9 | –1241.8 | 53.2 | –106.6 | 10 |
| Glyceraldehyde 3-*P*[d] | | –1285.6 | –1466.0 | 87.9 | –151.8 | 12 |
| Glycerol | $C_3H_8O_3$ | –488.5 | –1643.4 | 169.5 | –110.2 | 14 |
| Glycerol-*P* | | –1336.2 | –1652.6 | 160.3 | –119.3 | 14 |
| Glycine | $C_2H_5O_2N$ | –373.5 | –1008.3 | 157.2 | –22.6 | 9 |
| Glycogen | $C_6H_{10}O_5$ | –665.3 | –2887.0 | 220.9 | –258.6 | 24 |
| Glycolate$^-$ | $C_2H_3O_3^-$ | –523.4 | –739.7 | 37.2 | –42.7 | 6 |
| Glyoxylate$^-$ | $C_2HO_3^-$ | –461.1 | –564.9 | –46.9 | –86.9 | 4 |
| $H_2O$ (l) | | –237.2 | 0.0 | 0.0 | 0.0 | 0 |
| $OH^-$ | | –157.3 | –79.9 | –79.9 | –39.9 | 0 |
| $H^+$ | | 0.0 | 0.0 | 0.0 | | 0 |
| $H_2$ (g) | | 0.0 | –237.2 | 21.8 | –18.2 | 2 |
| $H_2O_2$ | | –136.8 | –100.4 | –359.4 | –319.4 | –2 |
| $H_2S$ | | –27.4 | –714.6 | 321.3 | 161.5 | 8 |
| $HS^-$ | | 12.6 | –754.5 | 281.4 | 161.5 | 8 |
| β-Hydroxybutyric acid | $C_4H_8O_3$ | –531.4 | –1994.9 | 336.0 | –23.6 | 18 |
| β-Hydroxybutyrate$^-$ | $C_4H_7O_3^-$ | –506.3 | –2020.0 | 310.9 | –8.8 | 18 |
| Hydroxypyruvate | $C_3H_4O_4$ | –615.9 | –1041.6 | –5.7 | –165.5 | 8 |

**TABLE 6-4 (continued)**

| Compound | Formula | $\Delta G_f^\circ$ (kJ mol$^{-1}$) | $\Delta G_c^\circ$ (kJ mol$^{-1}$) | For oxidation by $NAD^+$: $\Delta G^\circ_{ox}$ (kJ mol$^{-1}$) | For oxidation by $NAD^+$: $\Delta G^\circ_{ox}$ (pH 7) (kJ mol$^{-1}$) | Number of electrons |
|---|---|---|---|---|---|---|
| Hypoxanthine | $C_5H_6O$ | 89.5 | –2773.0 | 334.8 | –144.6 | 24 |
| Isocitrate$^{3-}$ | $C_6H_5O_7^{3-}$ | –1160.0 | –2155.1 | 175.8 | –63.9 | 18 |
| α-Ketoglutarate$^{2-}$ | $C_5H_4O_5^{2-}$ | –798.0 | –1885.5 | 186.4 | –53.3 | 16 |
| Lactate$^-$ | $C_3H_5O_3^-$ | –516.6 | –1378.1 | 175.9 | –23.9 | 12 |
| α-Lactose | $C_{12}H_{22}O_{11}$ | –1515.2 | –5826.5 | 389.3 | –569.7 | 48 |
| L-Leucine | $C_6H_{13}O_2N$ | –356.3 | –3551.7 | 721.6 | 62.3 | 33 |
| Mannitol | $C_6H_{14}O_6$ | –942.6 | –3084.0 | 282.8 | –236.6 | 26 |
| Malate$^{2-}$ | $C_4H_4O_5^{2-}$ | –845.1 | –1444.0 | 109.9 | –49.9 | 12 |
| Methane (g) | $CH_4$ | –50.8 | –818.0 | 218.0 | 58.2 | 8 |
| Methanol | $CH_4O$ | –175.2 | –693.5 | 83.4 | –36.4 | 6 |
| $NH_4^+$ | | –79.5 | –276.3 | 112.2 | 12.3 | 3 |
| $NO_2^-$ | | –34.5 | –84.1 | –472.6 | –372.7 | –3 |
| NO (g) | | 86.7 | –86.7 | –345.7 | –305.7 | –2 |
| $NO_3^-$ | | –110.5 | –8.1 | –655.6 | –515.7 | –5 |
| Oxalate$^{2-}$ | $C_2O_4^{2-}$ | –674.9 | –351.1 | –92.1 | –52.1 | 2 |
| Oxaloacetate$^{2-}$ | $C_4H_2O_5^{2-}$ | –797.2 | –1254.7 | 40.2 | –79.7 | 10 |
| $H_3PO_4$ (aq)[e] | | –1147.3 | 0.0 | 0.0 | 0.0 | 0 |
| $H_2PO_4^-$ (aq)[e] | | –1135.1 | –12.1 | –12.1 | 27.8 | 0 |
| $HPO_4^{2-}$ (aq)[e] | | –1094.1 | –53.1 | –53.1 | 26.8 | 0 |
| *n*-Propanol | $C_3H_8O$ | –175.8 | –1956.1 | 374.8 | 15.2 | 18 |
| Isopropanol | $C_3H_8O$ | –185.9 | –1946.0 | 384.9 | 25.3 | 18 |
| Propionate$^-$ | $C_3H_5O_2^-$ | –360.0 | –1534.7 | 278.2 | 38.5 | 14 |
| Pyruvate$^-$ | $C_3H_3O_3^-$ | –474.5 | –1183.1 | 111.9 | –47.9 | 10 |
| Phosphoenolpyruvate$^{3-}$ | | –1269.5 | –1245.0 | 50.0 | –109.8 | 10 |
| Ribose 5-*P*[d] | | –1599.9 | –2414.9 | 175.0 | –224.6 | 20 |
| Ribulose 5-*P*[d] | | –1597.6 | –2417.1 | 172.8 | –226.8 | 20 |
| Sedoheptulose 7-*P*[d] | | –1913.3 | –3364.6 | 261.2 | –298.2 | 28 |
| Sedoheptulose di-*P*[d] | | –2755.8 | –3379.0 | 246.9 | –312.5 | 28 |
| Sorbitol | $C_6H_{14}O_6$ | –942.7 | –3083.9 | 282.9 | –236.5 | 26 |
| Succinate$^{2-}$ | $C_4H_4O_4^{2-}$ | –690.2 | –1598.9 | 214.1 | 14.3 | 14 |
| Succinyl-CoA | | –686.7* | –1602.4* | 210.6* | –29.2* | 14 |
| Sucrose | $C_{12}H_{22}O_{11}$ | –1551.8 | –5789.9 | 425.9 | –533.1 | 48 |
| $SO_4^{2-}$ | | –742.0 | 0.0 | 0.0 | 79.9 | 0 |
| $SO_3^{2-}$ | | –497.1 | –244.9 | 14.1 | 54.0 | 2 |
| $S_2O_3^{2-}$ | | –513.4 | –733.4 | 302.5 | 222.6 | 8 |
| L-Threonine | $C_4H_9O_3N$ | –514.6 | –2130.3 | 330.1 | –49.5 | 19 |
| L-Tyrosine | $C_9H_{11}O_3N$ | –387.2 | –4466.8 | 842.5 | 23.4 | 41 |
| Urea | $CH_4ON_2$ | –203.8 | –664.9 | 112.0 | –7.8 | 6 |
| Uric acid | $C_5H_4O_3N_4$ | –356.9 | –2089.4 | 241.5 | –118.1 | 18 |
| L-Valine | $C_5H_{11}O_2N$ | –360.0 | –2916.5 | 579.9 | 40.5 | 27 |
| Xanthine | $C_5H_5O_2N_4$ | –139.3 | –2425.6 | 293.8 | –125.7 | 21 |
| D-Xylulose | $C_5H_{10}O_5$ | –748.1 | –2409.8 | 180.1 | –219.5 | 20 |

[a] The quantities tabulated are $\Delta G_f^\circ$, the standard free energy of formation from the elements; $\Delta G_c^\circ$, the standard free energy of combustion; $\Delta G^\circ_{ox}$, the standard free energy of oxidation by $NAD^+$ to products NADH + $H^+$, $CO_2$, $H_2O$, $N_2$, $HPO_4^{2-}$, and $SO_4^{2-}$; $\Delta G'_{ox}$ (pH 7), the apparent standard free energy change at pH 7. All values are in kJ mol$^{-1}$ at 25°C in aqueous solution unless indicated otherwise. If a compound is designated (g) the values are for the gaseous phase at 1 atm pressure. The number of electrons involved in complete oxidation to $CO_2$, $H_2O$, $N_2$, and $H_2SO_4$ is given in the final column. If this number is negative, the compound must be reduced to obtain the products, e.g., $2\ NO_3^- + 10\ e^- + 12\ H^+ \rightarrow N_2 + 6\ H_20$. The data for phosphate esters refer to the compounds with completely dissociated phosphate groups ($-O-PO_3^{2-}$). The values of $\Delta G_f^\circ$ for many of these compounds were calculated as $\Delta G_f^\circ$ (nonphosphorylated compound) – $\Delta G^\circ$ for hydrolysis (to $HPO_4^{2-}$, Table 6-6) $-\Delta G_f^\circ$ for $H_2O$ (one molecule for each phosphate ester formed) + $\Delta G_f^\circ$ for $HPO_4^{2-}$ (from this table). Data from Bassman and Krause[d] were used directly. For acyl-CoA derivatives CoA (–SH) is treated as an "element," i.e., the values of $\Delta G_f^\circ$ given and designated with an asterisk (*) are for formation from the elements plus free CoA. The values of $\Delta G_c^\circ$ and $\Delta G_{ox}$ are for oxidation to the usual products plus CoA. Values of $\Delta G^\circ$ of hydrolysis (Table 6-6) were used in computing $\Delta G_f^\circ$ for each of these compounds from that of the corresponding alcohol or carboxylate anion. Another source containing an extensive table of Gibbs energy values is Wilhoit, R. C. (1969) in *Biochemical Microcalorimetry* (Brown, H. D. ed.), pp. 305–317. Academic Press, New York

[b] The major source is Long, C., ed. (1961) *Biochemists Handbook*, pp. 90–92. Van Nostrand, Reinhold, Princeton, New Jersey. Most of the values in this collection are from Burton, K. (1957) *Ergeb. Physiol., Biol. Chem. Exp. Pharmakol.* **49**, 275–298

[c] From Stull, D.R., Westrum, E. F., Jr., and Sinke, G. C. (1969) *The Chemical Thermodynamics of Organic Compounds*. Wiley, New York

[d] Bassham, J. A. and Krause, G. H. (1969) *Biochim. Biophys. Acta.* **189**, 207–221

[e] Van Wazer, J. R. (1958) *Phosphorus and Its Compounds*, Vol. I, p. 889. Wiley (Interscience), New York

## B. Tables of $\Delta G°$ Values for Biochemical Compounds

### 1. Gibbs Energies of Formation

Table 6-4 gives, in the first column, standard values of Gibbs energies of formation from the elements $\Delta G_f°$ for a variety of pure solids, gases, and liquids as well as values for substances in solution at the hypothetical 1 M activity. As an example, consider the value of $\Delta G_f°$ for pure liquid acetic acid, $-389.1$ kJ mol$^{-1}$. The equation for its formation from the elements is:

$$2\,C\,(s) + 2\,H_2\,(g, 1\,atm) + O_2\,(g, 1\,atm) \longrightarrow C_2H_4O_2\,(l)$$
$$\Delta G_f° = -389.1\ \text{kJ mol}^{-1} \qquad (6\text{-}38)$$

To obtain the Gibbs energy of formation in aqueous solution, we must have solubility data as well as activity coefficients of acetic acid at various concentrations. From these data the change in Gibbs energy for solution of the liquid acetic acid in water to give aqueous acetic acid in the hypothetical 1 molal standard state (Eq. 6-39) can be obtained.

$$\text{Acetic acid (l)} \longrightarrow \text{acetic acid (aq)}$$
$$\Delta G = -7.3\ \text{kJ mol}^{-1} \qquad (6\text{-}39)$$

Summing Eqs. 6-38 and 6-39 we obtain:

$$2\,C + 2\,H_2 + O_2 \longrightarrow \text{acetic acid (aq)}$$
$$\Delta G_f° = -396.4\ \text{kJ mol}^{-1} \qquad (6\text{-}40)$$

In many computations it is convenient to have $\Delta G$ values for single ions, e.g., for acetate$^-$. We can obtain $\Delta G_f°$ of acetate$^-$ (aq) from that of acetic acid (aq) by making use of $\Delta G°$ of dissociation (Eq. 6-41).

$$\text{Acetic acid} \longrightarrow H^+ + \text{acetate}^-$$
$$\Delta G° = -5.708 \log K_a = 5.708\ pK_a$$
$$= +27.2\ \text{kJ mol}^{-1} \text{ at } 25°C \qquad (6\text{-}41)$$

By convention *we define the Gibbs energy of formation of $H^+$ as zero*. Then, by summing Eqs. 6-40 and 6-41 we obtain $\Delta G_f°$ of acetate$^-$ = $-369.2$ kJ mol$^{-1}$.

### 2. Gibbs Energies of Dissociation of Protons

Table 6-5 gives thermodynamic dissociation constants and values of $\Delta G°$ and $\Delta H°$ for a number of acids of interest in biochemistry. Some of these values were used in obtaining the values of $\Delta G_f°$ for the ions of Table 6-4. The data of Table 6-5 can also be used in evaluation of Gibbs energy changes for reactions of ionic forms not given in Table 6-4.

### 3. Group Transfer Potentials

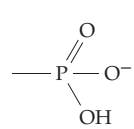
Recall that the equilibria for reactions by which monomers are linked to form biopolymers (whether amides, esters, phosphodiesters, or glycosides) usually favor *hydrolysis* rather than formation (condensation). The equilibrium positions depend on the exact structures. Some linkages are formed easily if monomer concentrations are high enough, but others are never formed in significant concentrations. Likewise, hydrolysis may be partial at equilibrium or it may be 99.9% or more complete.

Let us compare the hydrolysis of the two organic phosphates **adenosine triphosphate** (ATP) and **glucose 6-phosphate** (Eqs. 6-42 and 6-43).

$$HATP^{3-} + H_2O \longrightarrow HADP^{2-} + H_2PO_4^-$$
$$\Delta G° \approx -32.9\ \text{kJ mol}^{-1}$$
$$\Delta H° \approx -22.6\ \text{kJ mol}^{-1} \text{ at } 25°C,\ \mu = 0.25 \qquad (6\text{-}42)$$

$$\alpha\text{-D-Glucose 6-phosphate}^- + H_2O \longrightarrow \text{glucose} + H_2PO_4^-$$
$$\Delta G° = -16.4\ \text{kJ mol}^{-1} \text{ at } 25°C \qquad (6\text{-}43)$$

The decrease in Gibbs energy upon hydrolysis is twice as large for ATP as it is for glucose 6-phosphate. Glucose phosphate is thermodynamically more stable than ATP. It would be easier to form than would ATP by a reversal of the hydrolysis reaction and also easier to form biosynthetically. From the Gibbs energies of hydrolysis it follows that a **phospho group** could be transferred spontaneously from ATP to glucose in the presence of a suitable catalyst but not vice versa.



Phospho (phosphoryl) group

Because it reflects quantitatively the thermodynamic tendency for a group to be transferred to another

**TABLE 6-5**
**Values of p$K_a$, $\Delta G°$, and $\Delta H°$ for Ionization of Acids at 25°C [a,b]**

| Acid | p$K_a$ | $\Delta G°$ (kJ mol$^{-1}$) | $\Delta H°$ (kJ mol$^{-1}$) |
|---|---|---|---|
| Formic acid | 3.75 | 21.4 | 0.04 |
| Acetic acid | 4.76 | 27.2 | −0.1 |
| Propionic acid | 4.87 | 27.8 | −0.6 |
| Lactic acid | 3.97 (35°C) | 23.4 | 2.2 |
| Pyruvic acid | 2.49 | 14.2 | 12.1 |
| $NH_4^+$ | 9.25 | 52.8 | 52.2 |
| $CH_3NH_4^+$ | 10.59 | 60.4 | 55.4 |
| Alanine | | | |
| −COOH | 2.35 | 13.4 | 3.1 |
| $-NH_3^+$ | 9.83 | 56.1 | 45.4 |
| β-Alanine | | | |
| −COOH | 3.55 | 20.3 | 4.5 |
| $-NH_3^+$ (apparent) | 10.19 | 58.2 | |
| L-Alanyl-L-alanine | 3.34 | 19.1 | −0.5 |
| Aspartic acid | | | |
| −COOH | 2.05 | 11.7 | 7.7 |
| −COOH | 3.87 | 22.1 | 4.0 |
| $-NH_3^+$ | 10.60 | 60.5 | 38.8 |
| $H_2CO_3$, p$K_1$ | 6.35[c] | 36.2 | 9.4 |
| p$K_2$ | 10.33 | 59.0 | 15.1 |
| $H_3PO_4$, p$K_1$ | 2.12 | 12.1 | −7.9 |
| p$K_2$ | 7.18[d] | 41.0 | 3.8 |
| (apparent) | 6.78[e] | 38.7 | 3.3 |
| p$K_3$ | 12.40 | 70.8 | 17.6 |
| Glycerol 1-phosphate, p$K_2$ | 6.66 | 38.0 | −3.1 |
| Glucose 6-phosphate | 6.50 | 37.1 | −1.8 |
| Pyrophosphoric acid, $H_4P_2O_7$ | | | |
| p$K_3$ | 6.7 | 38.1 | −1.3 |
| (apparent) | 6.12[e] | 34.9 | 0.5 |
| p$K_4$ | 9.4 | 53.6 | −7.1 |
| (apparent) | 8.95[e] | 51.2 | 1.7 |
| Adenosine | 3.5 | 20.1 | 13.0 |
| AMP | | | |
| p$K_1$ (ring, apparent) | 3.74[e] | 21.3 | 4.2 |
| p$K_2$ (phosphate) | 6.67[d] | 38.1 | 3.6 |
| (apparent) | 6.45[e] | 36.8 | 3.6 |
| ADP | | | |
| p$K_2$ (ring, apparent) | 3.93[e] | 22.4 | 4.2 |
| p$K_3$ (diphosphate) | 7.20[d] | 41.1 | −5.7 |
| (apparent) | 6.83[f] | 39.0 | −5.7 |
| ATP | | | |
| p$K_3$ (ring, apparent) | 4.06 | 23.2 | 0 |
| p$K_4$ (triphosphate) | 7.68[d] | 43.8 | −7.0 |
| (apparent) | 7.06[d] | 40.2 | −7.0 |
| Pyridine | 5.17 | 29.5 | 20.1 |
| Phenol | 9.98 | 56.9 | 23.6 |

[a] These are thermodynamic values (infinite dilution) except for those labeled apparent. The latter apply at an ionic strength of 0.2–0.25.
[b] Most data are from Jencks, W. P. and Regenstein, J. (1976) in *Handbook of Biochemistry and Molecular Biology*, 3rd ed., Vol. I (Fasman, G. D. ed.), pp. 305–351, CRC Press, Cleveland, Ohio.
[c] Here, p$K_1$ is for $K_1 = [H^+][HCO_3^-] / [CO_2] + [H_2CO_3]$. From Forster, R. E., Edsall, J. T., Otis, A. B., and Roughton, F. J. W., eds. (1969) *NASA Spec. Publ.* **188**.
[d] From Phillips, R. C., George, P., and Rutman, R. J. (1963) *Biochemistry* **2**, 501–508.
[e] From Alberty, R. A. (1969) *J. Biol. Chem.* **244**, 3290–3302.
[f] Values used by Alberty, R. A. (1972) *Horizons of Bioenergetics*, pp. 135–147, Academic Press, New York, calculated for 0.2 ionic strength from equations of Phillips, R. C., George, P., and Rutman, R. J. (1966) *J. Am. Chem. Soc.* **88**, 2631–2640.

nucleophile (see Chapter 12), *the Gibbs energy decrease (−ΔG°) upon hydrolysis* is sometimes called the **group transfer potential**. During the hydrolysis of ATP (Eq. 6-42) the phospho group of ATP is transferred to a hydroxyl ion from water with a value $\Delta G° = -32.9$ kJ mol$^{-1}$. The group transfer potential of this phospho group is 32.9 kJ/mol and that of the phospho group of glucose 6-phosphate is only 27.6 kJ mol$^{-1}$. While the choice of water as the reference nucleophile for expression of the group transfer potential is somewhat arbitrary, it is customary. Transfer of groups is important in energy metabolism and in biosynthesis of polymers. Gibbs energies of hydrolysis are given in Table 6-6 for several compounds.

## 4. "Constants" That Vary with pH and Magnesium Ion Concentrations

Equation 6-42 is written for hydrolysis of $HATP^{3-}$ to $HADP^{2-}$ + $H_2PO_4^-$, a stoichiometry that applies well in the pH range around 6. However, at a pH above ~7 most of the ATP is in the form $ATP^{4-}$ and is cleaved to $HPO_4^{2-}$ according to Eq. 6-44.

$$ATP^{4-} + H_2O \rightarrow ADP^{3-} + HPO_4^{2-} + H^+$$
$$\Delta G° \approx +5.4 \text{ kJ mol}^{-1} \text{ at } 25°C, \mu = 0.25$$
$$\Delta H° \approx -19.7 \text{ kJ mol}^{-1} \qquad (6\text{-}44)$$

The value of $\Delta G° = +5.4$ kJ mol$^{-1}$ for this reaction is hardly the large negative number expected for a highly spontaneous reaction. What is the matter? The problem is that $H^+$ is produced and that the standard state of $H^+$ is 1 M, not $10^{-7}$ M. Because of this, biochemists often prefer to use another kind of **apparent dissociation constant** and an **apparent ΔG** such that the standard state of $H^+$ is taken as that of the pH at which the experiments were done, usually pH 7. The symbol $K'$ has often be used and is used in this book to represent the following

pH-dependent equilibrium constant (Eq. 6-45) which will be a constant only at a single pH.

$$K' = \frac{[ADP^{3-}][HPO_4^{2-}]}{[ATP^{4-}]} \qquad (6\text{-}45)$$

If one proton is produced in the reaction as in Eq. 6-44, the following relationship will hold.

$$\Delta G' = \Delta G^\circ - 5.708 \text{ x pH kJ mol}^{-1} \text{ at } 25°C \qquad (6\text{-}46)$$

Note that $\Delta G' = -RT \ln K'$ and that $[H^+]$ does not appear in the expression for $K'$ given by Eq. 6-45. From the value $\Delta G^\circ = +5.4$ kJ mol$^{-1}$ and applying Eq.

**TABLE 6-6**
**Gibbs Energies of Hydrolysis at 25°C (in kJ mol$^{-1}$)[a]**

| Compound | Products | $\Delta G^\circ$ | $\Delta G'$(pH 7) | $\Delta H^\circ$ |
|---|---|---|---|---|
| $ATP^{4-}$ | $ADP^{3-} + HPO_4^{2-} + H^+$ | 5.41[b] | –34.54 | –19.71 |
| $ATP^{4-}$ | $AMP^{2-} + HP_2O_7^{3-} + H^+$ | 2.54[c] | –37.4 | –19.0 |
| $MgATP^{2-}$ | $MgADP^- + HPO_4^{2-} + H^+$ | 16.0[d] | –24.0 | –14.2 |
| $ADP^{3-}$ | $AMP^{2-} + HPO_4^{2-} + H^+$ | 3.67[c] | –36.3 | –13.5 |
| $AMP^{2-}$ | Adenosine + $HPO_4^{2-}$ | –9.6[e] | –9.6 | 0 |
| $ATP^{4-}$ | Adenosine + $HP_3O_{10}^{4-}$ | –36.0[e] | –36.0 | –7.9 |
| $HP_2O_7^{3-}$ | $2\ HPO_4^{2-} + H^+$ | 6.54[c] | –33.4 | –12.6 |
| Acetyl phosphate$^{2-}$ | Acetate$^-$ + $HPO_4^{2-} + H^+$ | –7.7[f] | –47.7 | |
| 1,3-Diphosphoglycerate$^{4-}$ | 3-Phosphoglycerate$^{3-}$ + $HPO_4^{2-} + H^+$ | –14.5[g] | –54.5 | |
| Phosphoenolpyruvate$^{3-}$ | Pyruvate$^-$ + $HPO_4^{2-}$ | –61.9 | –61.9 | –25.1 |
| Carbamoyl phosphate$^{2-}$ | $H_2N-C(=O)-O^- + HPO_4^{2-} + H^+$ | –11.5 | –51.5 | |
| Creatine phosphate$^-$ | Creatine$^+$ + $HPO_4^{2-}$ | –43.1 | –43.1 | |
| Phosphoarginine$^-$ | Arginine + $HPO_4^{2-}$ | | –38.1[h] | ($Mg^{2+}$ present) |
| Glycerol phosphate$^{2-}$ | Glycerol + $HPO_4^{2-}$ | –9.2 | –9.2 | |
| α-D-Glucose 6-phosphate$^{2-}$ | α-D-Glucose + $HPO_4^{2-}$ | –13.8 | –13.8 | –2.5 |
| Glucose 1-phosphate$^{2-}$ | Glucose + $HPO_4^{2-}$ | –20.9 | –20.9 | |
| Maltose (or glycogen) | 2-Glucose | –16.7 | –16.7 | |
| Sucrose | Glucose + fructose | –29.3 | –29.3 | |
| UDP glucose$^{2-}$ | Glucose + $UDP^{3-} + H^+$ | 9.4 | –30.5 | |
| $N^{10}$-Formyltetrahydrofolic acid | Formate$^-$ + $H^+$ + tetrahydrofolic acid | 14.1 | –25.9[h] | |
| Acetic anhydride | 2-Acetate$^-$ + $2H^+$ | 31.1 | –48.9 | |
| Acetyl-CoA | Acetate$^-$ + $H^+$ + CoA | 4.9 | –35.1[i] | |
| Succinyl-CoA$^-$ | Succinate$^{2-}$ + $H^+$ + CoA | –3.5 | –43.5[j] | |
| Ethyl acetate | Ethanol + acetate$^-$ + $H^+$ | 20.2 | –19.7 | |
| Asparagine | Aspartate$^-$ + $NH_4^+$ | –15.1 | –15.1 | |
| Glycine ethyl ester$^+$ (39°C) | Glycine + ethanol + $H^+$ | 4.9 | –35.1 | |
| Valyl-tRNA$^+$ | Valine + tRNA + $H^+$ | 4.9 | –35.1 | |

[a] Unless indicated otherwise, the values are based on tables from Jencks, W. P. (1976) *Handbook of Biochemistry and Molecular Biology*, 3rd ed., Vol I (Fasman,G.D. ed.), pp. 296–304. CRC Press, Cleveland, Ohio. For a reaction producing one proton at pH 7 $\Delta G' = -39.96$ kJ mol$^{-1}$.

[b] Guynn, R.W. and Veech, R.L. (1973) *J. Biol. Chem.* **248**, 6966–6972.

[c] Based on +11.80 kcal mol$^{-1}$ for hydrolysis to $P_2O_7^{4-}$ plus $\Delta G^\circ$ of dissociation of $HP_2O_7^{3-}$ as quoted by Alberty, R. A. (1969) *J.Biol.Chem.* **244**, 3290–3324. However, 1.017 kcal mol$^{-1}$ was added to the value of 11.80 to make it consistent with that for hydrolysis of ATP to ADP. Reevaluation by Frey and Arabshahi (1995) *Biochemistry* **34**, 11307–11310, indicates that $\Delta G'$ (pH 7) for hydrolysis of ATP to AMP + $PP_i$ is ~ 10 kJ mol$^{-1}$ more negative than is shown here.

[d] Alberty, R.A. (1972) *Horizons of Bioenergetics*, Academic Press, New York, pp. 135–147.

[e] George, P. , Witonsky, R.J., Trachtman, M., Wu, C., Dorwart, W., Richman, L., Richman, W., Shurayh, F., and Lentz, B. (1970) *Biochim. Biophys. Acta.* **223**, 1–15.

[f] Based on $\Delta G^\dagger = 3.0$ kcal mol$^{-1}$ for acetyl phosphate + CoA → acetyl-CoA + $P_i$ from Stadtman, E. R. (1973) *The Enzymes*, (Boyer, P.D., ed.), 3rd ed., Vol. 8. pp. 1–49. Academic Press, New York, together with $\Delta G'$ (pH 7) for hydrolysis of acetyl-CoA.

[g] Estimated from $\Delta G^\circ$ for ATP hydrolysis + $\Delta G' = -19.9$ kJ mol$^{-1}$ for the 3-phosphoglycerate kinase reaction: Burton, K. and Krebs, H. A. (1953) *Biophys. J.* **54**, 94–107; and (1955) *Biophys. J.* **59**, 44–46.

[h] Estimated from data at pH 7.7 or 8.0 (tables of Jencks).

[i] Guynn, R. W., Gelberg, H. J., and Veech, R. L. (1973) *J. Biol. Chem.* **248**, 6957–6965 found $\Delta G^\circ = -35.75$ kJ mol$^{-1}$ at 38°C. Without data on $\Delta H$, correction to 25°C is difficult. Burton, K., (1955) *Biophys. J.* **59**, 44–46, gave $\Delta G'$ (pH 7) for $ATP^{4-}$ + acetate$^-$ + CoA → $ADP^{3-}$ + acetyl-CoA + $HPO_4^{2-}$ as approximately zero at 25°C. Guynn *et al.* found –0.56 kJ mol$^{-1}$ at 38°C. This same value (–0.56 kJ) at 25°C was assumed to obtain the figure given here. This is equivalent to assuming $\Delta H^\circ$ of hydrolysis as almost the same for ATP and acetyl-CoA, an unsupported assumption.

[j] Assumed 2 kcal mol$^{-1}$ more negative than that of acetyl-CoA as in tables of Jencks.

6-46, we obtain for the hydrolysis of ATP at 25°C, $\mu = 0.25$:

$$\Delta G'\ (\text{pH } 7) = -34.5 \text{ kJ mol}^{-1} = -8.26 \text{ kcal mol}^{-1} \quad (6\text{-}47)$$

An additional set of standard states is frequently met in the biochemical literature. An equilibrium constant, designated in this book as $K^{\dagger}$, is used to relate the *total concentrations of all ionic forms* of the components present at the pH of the experiment. Thus,

$$K^{\dagger} = \frac{[\text{ADP, all forms}][\text{phosphate, all forms}]}{[\text{ATP, all forms}]} \quad (6\text{-}48)$$

## BOX 6-A MEASUREMENT OF INTRACELLULAR pH

What is the pH within a cell? Is it constant or does it vary with physiological conditions? Do all cells operate at similar pH? The answers to these important questions have been sought using a variety of techniques.[a] Tiny microelectrodes with tips only 1 μm in diameter have been inserted into cells. Indicator dyes have been diffused into cells and either light absorption or fluorescence[b] measured. The distribution of a suitable radiolabeled weak acid or weak base, such as [$^{14}$C]methylamine, that is able to permeate cells can be used to calculate the difference between internal and external pH.[c] The activity of the enzyme carbonic anhydrase, which is very pH sensitive, can be used to monitor the pH of mitochrondia.[d] Since 1973 NMR methods have been popular.[e–l] The chemical shifts of $^{31}$P in inorganic phosphate ($pK_a = 6.9$), ATP, glucose 6-phosphate, and of other metabolites of $^{13}$C in citrate and bicarbonate,[l] give direct estimates of pH. However, caution must be exercised if internal pH values good to ± 0.1 unit are to be obtained.[g] More sensitive measurements over the pH range 1.3 – 9.1 can be made by diffusing one of a series of aminophosphonates into cells and measuring the $^{31}$P chemical shift.[h] Fluorinated probes such as dimethylfluoroalanine ($pK_a = 7.3$) are also useful because of their low toxicity and high sensitivity of the $^{19}$F NMR signal. These alanine derivatives can be diffused into cells as their methyl esters, which are rapidly cleaved within cells, allowing the fluorinated amino acids to accumulate.[i,m]

The pH within cells appears to be tightly controlled although small variations are sometimes observed. Red blood cells, thymocytes, liver, skeletal muscles, and intact hearts all maintain a pH in the range 7.0 – 7.3.[b,h,i] However, the pH can fall to 6.2 within 13 minutes of oxygen deprivation (ischemia) and to 6.1 after exhaustive exercise.[n,o] The $^{31}$P NMR technique permits the monitoring of pH as well as the state of the adenylate system (Section D) in human limbs suffering from circulatory insufficiency.[c]

In higher plants the pH of cytoplasm is 7.4–7.5 but vacuoles are acidic with a pH of 4.5–6.[l] The cytoplasm of maize root tips has a pH of 7.1 but the vacuoles are at a pH of 5.5.[f] The pH of granules of the chromaffin cells of the adrenal cortex, which accumulate high concentrations of ATP and catecholamines, is also low, ~5.7.[p] The bacterium *Streptococcus faecalis* maintains a higher internal pH of ~8.0, even when the pH of the medium varies from 6.5 to 8.0[q] while *E. coli* operates at pH 7.6, the extremes of variation being 7.4 – 7.8 when the external pH changes from 5.5–9.[r]

Changes of internal pH during developmental events such as fertilization of sea urchin eggs (+0.3 unit) have been recorded. However, the significance of pH changes in metabolic regulation remains uncertain.[c]


[a] Kotyk, A., and Slavik, J. (1989) *Intracellular pH and its Measurement*, CRC Press, Boca Raton, Florida
[b] Rogers, J., Hesketh, T. R., Smith, G. A., and Metcalfe, J. C. (1983) *J. Biol. Chem.* **258**, 5994–5997
[c] Nuccitelli, R., and Deamer, D. W., eds. (1982) *Intracellular pH: Its Measurement, Regulation and Utilization in Cellular Functions*, Liss, New York
[d] Dodgson, S. J., Forster, R. E., II, and Storey, B. T. (1982) *J. Biol. Chem.* **257**, 1705–1711
[e] Moon, R. B., and Richards, J. H. (1973) *J. Biol. Chem.* **248**, 7276–7278
[f] Roberts, J. K. M., Ray, P. M., Wade-Jardetsky, N., and Jardetsky, O. (1980) *Nature (London)* **283**, 870–872
[g] Avison, M. J., Hetherington, H. P., and Shulman, R. G. (1986) *Ann. Rev. Biophys. Biophys. Chem.* **15**, 377–402
[h] Pietri, S., Miollan, M., Martel, S., Le Moigne, F., Blaive, B., and Culcasi, M. (2000) *J. Biol. Chem.* **275**, 19505–19512
[i] Taylor, J. S., and Deutsch, C. (1983) *Biophys. J.* **43**, 261–267
[j] Bailey, I. A., Williams, S. R., Radda, G. K., and Gadian, D. G. (1981) *Biochem. J.* **196**, 171–178
[k] Barton, J. K., Den Hollander, J. A., Lee, T. M., MacLaughlin, A., and Shulman, R. G. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2470–2473
[l] Gout, E., Bligny, R., and Douce, R. (1992) *J. Biol. Chem.* **267**, 13903–13909
[m] Deutsch, C. J., and Taylor, J. S. (1987) *Ann. N.Y. Acad. Sci.* **508**, 33
[n] Garlick, P. B., Radda, G. K., and Seeley, P. J. (1979) *Biochem. J.* **184**, 547–554
[o] Pan, J. W., Hamm, J. R., Rothman, D. L., and Shulman, R. G. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7836–7839
[p] Pollard, H. B., Shindo, H., Creutz, C. E., Pazoles, C. J., and Cohen, J. S. (1979) *J. Biol. Chem.* **254**, 1170–1177
[q] Kobayashi, H., Murakami, N., and Unemoto, T. (1982) *J. Biol. Chem.* **257**, 13246–13252
[r] Slonczewski, J. L., Rosen, B. P., Alger, J. R., and MacNab, R. M. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6271–6275

and

$$\Delta G^{\dagger} = -RT \ln K^{\dagger} \tag{6-49}$$

The Gibbs energy change $\Delta G^{\dagger}$ can be related to $\Delta G'$ by considering the relationship of $K^{\dagger}$ to $K'$. For ATP hydrolysis in the pH range of 2–10, $K^{\dagger}$ is given by Eq. 6-50.

$$K^{\dagger} = \frac{K'\left(1 + \frac{[H^+]}{K_{HADP^{2-}}} + \frac{[H^+]^2}{K_{HADP^{2-}}K_{H_2ADP^-}}\right)\left(1 + \frac{[H^+]}{K_{HPO_4^{2-}}}\right)}{\left(1 + \frac{[H^+]}{K_{HATP^{3-}}} + \frac{[H^+]^2}{K_{HATP^{3-}}K_{H_2ATP^{2-}}}\right)} \tag{6-50}$$

In this equation $K_{HADP^{2-}}$, etc., are consecutive dissociation constants as given in Table 6-4. The expressions in parentheses are the **Michaelis pH functions**, which were considered in Chapter 3 (Eqs. 3-4 to 3-6). In Eq. 6-50 they relate the total concentration of each component to the concentration of the most highly dissociated form. Thus, for the pH range 2–10

$$[P_i]_{total} = [HPO_4^{2-}](1 + [H^+] / K_{H_2PO_4^-}) \tag{6-51}$$

Using apparent $pK_a$ values ($\mu = 0.2$) for $H_2PO_4^-$, $HADP^{2-}$, and $HATP^{3-}$ of 6.78, 6.83, and 7.06 (Table 6-5) and taking $\Delta G'$ at pH 7 as –34.5 kJ mol$^{-1}$, we compute $\Delta G^{\dagger} = -35.0$ kJ mol$^{-1}$ at pH 7. The difference between $\Delta G'$ and $\Delta G^{\dagger}$ in this case is small, but it would be larger if the ionic forms in Eq. 6-44 were not the ones predominating at pH 7.

To obtain the Gibbs energy change for a reaction under *other than standard conditions,* Eq. 6-29 must be applied. Thus, at pH 7 and 0.01 M activities of $ADP^{3-}$, $ATP^{4-}$, and $HPO_4^{2-}$, $\Delta G$ for hydrolysis of ATP according to Eq. 6-44 is –34.5 –(2 x 5.71) = –45.9 kJ mol$^{-1}$ = –11.0 kcal mol$^{-1}$. We see that at concentrations existing in cells (usually in the millimolar range) ATP has a substantially higher group transfer potential than under standard conditions.

The reader should bear in mind that there is no accepted standard usage of $K'$ and that $K^{\dagger}$ is just for this book! An international committee[21] has recommended that $K'$ be used with the same meaning as $K^{\dagger}$ in this book and more changes may be coming (see the next section).

To obtain $\Delta G$ at a temperature other than 25°C, we must know $\Delta H$ for the reaction. Using Eq. 6-37a it is easy to show that $\Delta G$ at temperature $T_2$ is related to that at $T_1$ as follows:

$$\Delta G_2 \approx \frac{T_2\,\Delta G_1 - (T_2 - T_1)\,\Delta H}{T_1} \tag{6-52}$$

The enthalpy of hydrolysis of ATP according to Eq. 6-44 is approximately –19.7 kJ mol$^{-1}$. Using this value and applying Eq. 6-52 we can calculate that $\Delta G'$ (pH 7) for the hydrolysis of ATP at 38°C is –35.2 kJ mol$^{-1}$. The value of $\Delta G'$ (25°C, pH 7), used in obtaining this answer, is – 35.54 kJ mol$^{-1}$ which was computed from the value of $\Delta G'$ (pH 7), 38°C of 35.19 kJ mol$^{-1}$ reported by Guynn and Veech[37] using Eq. 6-52. Bear in mind that all of the foregoing Gibbs energy changes are *apparent* values applying to solutions of ionic strength ~0.25.

Both ADP and ATP as well as inorganic pyrophosphate form complexes with metal ions. Since the magnesium complexes are often the predominant forms of ADP and ATP under physiological conditions, we must consider the following Gibbs energy changes. These are apparent values for $\mu = 0.2$ at 25°C.

$$ATP^{4-} + Mg^{2+} \rightarrow Mg\,ATP^{2-}$$
$$\Delta G^{\circ} = -26.3;\ \Delta H^{\circ} = 13.8 \text{ kJ mol}^{-1} \tag{6-53}$$

$$ADP^{3-} + Mg^{2+} \rightarrow Mg\,ADP^{-}$$
$$\Delta G^{\circ} = -19.8;\ \Delta H^{\circ} = 15.1 \text{ kJ mol}^{-1} \tag{6-54}$$

Combining the apparent $\Delta G^{\circ}$ values for Eqs. 6-47, 6-53, and 6-54, we obtain

$$MgATP^{2-} + H_2O \rightarrow MgADP^{-} + HPO_2^{2-} + H^+$$
$$\Delta G' \text{ (pH 7)} = -28.0 \text{ kJ mol}^{-1} \tag{6-55}$$

The stoichiometry of Eq. 6-55 never holds exactly. Some of the $Mg^{2+}$ dissociates from the $MgADP^-$; both protons and $Mg^{2+}$ bind to $H_2PO_4^-$; and $HATP^{3-}$ and $HADP^{2-}$ are present and bind $Mg^{2+}$ weakly.[38] Thus, the observed value of $\Delta G^{\dagger}$ varies with both pH and magnesium concentration as well as with changes in ionic strength. Tables and graphs showing the apparent value of $\Delta G^{\dagger}$ under various conditions have been prepared by Alberty[38] and by Phillips *et al.*[39] An example is shown in Fig. 6-2. From this graph we find that $\Delta G^{\dagger}$ for hydrolysis of ATP at pH 7, 25°C, $\mu = 0.2$, and 1 mM $Mg^{2+}$ (a relatively high intracellular concentration[40]) is –30.35 kJ mol$^{-1}$ (–7.25 kcal mol$^{-1}$).

Figure 6-2 was drawn using the equations of Alberty, but the value of $\Delta G'$ (pH 7) of hydrolysis of ATP = 34.54 kJ mol$^{-1}$ at $[Mg^{2+}] = 0$ based on results of Guynn and Veech[37] was used. The $pK_a$ values and formation constants of $Mg^{2+}$ complexes were those of Alberty.[38] Note that George *et al.*[40] provided formation constants of these complexes at infinite dilution where the values of $\Delta G^{\circ}$ of formation are considerably more negative than those given in Eqs. 6-53 to 6-55.

From the foregoing considerations we see that complexing with $Mg^{2+}$ somewhat decreases the group transfer potential of the phospho group of ATP. Furthermore, changes in the concentration of free $Mg^{2+}$ with time and between different regions of a cell may

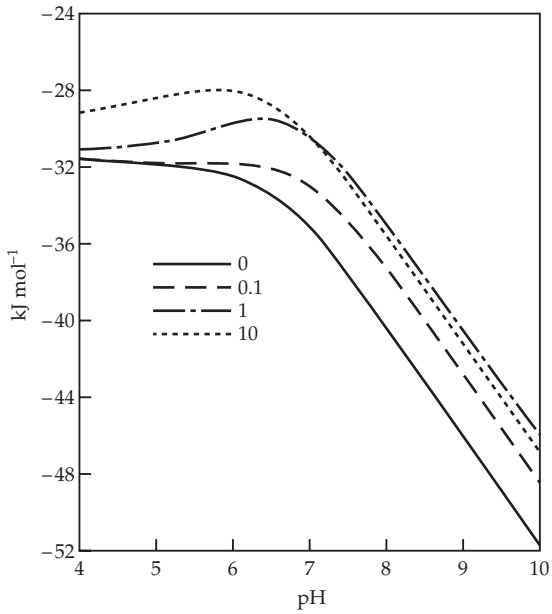



**Figure 6-2** Plots of the apparent Gibbs energy $\Delta G^{\dagger}$ for hydrolysis of ATP as a function of pH at a series of different concentrations of free magnesium ions. Millimolar $[Mg^{2+}]$ is indicated by the numbers by the curves. Computer-drawn graphs courtesy of Carol M. Metzler.

have significant effects.[41] While $Mg^{2+}$ is a principal cation in tissues, it is by no means the only one. Thus, $Ca^{2+}$, $Mn^{2+}$, and even $K^+$ will affect equilibria involving polyphosphates such as ATP.

It is not easy to measure the group transfer potential of ATP and published values vary greatly. George *et al.*[40] reported $\Delta G°$ for Eq. 6-42 as −39.9 kJ mol$^{-1}$ at 25°C, μ = 0.2, and as − 41.25 kJ mol$^{-1}$ at infinite dilution. They regarded these values as good to ±4 kJ mol$^{-1}$. However, most other estimates have been at least 4 kJ mol$^{-1}$ less negative.[42–44] The self-consistent set of thermodynamic data used throughout this book are based in part on the value of $\Delta G$ for hydrolysis of ATP obtained by Guynn and Veech.[37]

## 5. A New Standard for Biochemical Thermodynamics?

Because of the complexities of the equilibria involved in biochemical reactions Alberty and others[45–47] and a Panel on Biochemical Thermodynamics[48] have suggested that tables of thermodynamic properties for biochemical use be tabulated for the following conditions: $T$ = 298.15 K, $P$ = 1 bar (0.1 $MP_a \equiv 0.987$ atm), pH = 7, pMg = 3 ($[Mg^{2+}] = 10^{-3}$ M), and ionic strength I (μ in this chapter) = 0.25 M. There appears to be both advantages and disadvantages to this. The proponents suggest that the symbols $\Delta G'$, $K'$, etc. be used for this new standard. To change all of the numbers in this book, which uses a self-consistent set of thermodynamic quantities, is impractical. However, it is worthwhile to compare the value of Alberty and Goldberg for $\Delta G'°$ for hydrolysis of ATP to ADP + $P_i$ under the proposed biochemical standard conditions with other values given in this chapter, all at 25°C.

$\Delta G'°$ = −32.49 kJ mol$^{-1}$ (new proposed biochemical standard)
$\Delta G^{\dagger}$ (pH 7) −35.0 kJ mol$^{-1}$ (Eq. 6-50)
$\Delta G'$ (pH 7) −34.5 kJ mol$^{-1}$ (for $ATP^{4-} \rightarrow ADP^{3-} + HPO4^{2-}$; Eq. 6-47)
$\Delta G'$ (pH 7) −28.0 kJ mol$^{-1}$ (for $MgATP^{2-} \rightarrow MgADP^{-} + HPO4^{2-}$; Eq. 6-55)

## 6. Bond Energies and Approximate Methods for Estimation of Thermodynamic Data

For approximate estimation of enthalpy changes during reactions, use can be made of empirical bond energies (Table 6-7) which represent the approximate enthalpy changes ($-\Delta H°$) for formation of compounds in a gaseous state from atoms in the gas phase. Other more comprehensive methods of approximation have been developed.[49,50]

## 7. Gibbs Energies of Combustion by $O_2$ and by $NAD^+$

Since oxidation processes are so important in the metabolism of aerobic organisms, it is often convenient to discuss Gibbs energies of combustion. These are easily derived from the Gibbs energies of formation. For example, $\Delta G_c$ for acetate ion (aqueous) may be obtained as follows:

$$\text{Acetate}^- + H^+ \rightarrow 2\,C + 2\,H_2 + O_2$$
$$\Delta G° = +369.2 \text{ kJ mol}^{-1} = -\Delta G_f° \text{ of acetate (from Table 6-3)}$$

$$2\,H_2 + O_2 \rightarrow 2\,H_2O$$
$$\Delta G° = 2 \times -237.2 = -474.4 \text{ kJ mol}^{-1} = 2 \times \Delta G_f° \text{ of } H_2O \qquad (6\text{-}56)$$

$$2\,C + 2\,O_2 \rightarrow 2\,CO_2$$
$$\Delta G° = 2 \times -394.4 = -788.8 \text{ kJ mol}^{-1} = 2 \times \Delta G_f° \text{ of } CO_2$$

---

$$\text{Acetate}^- (1\text{ M}) + H^+ (1\text{ M}) + 2\,O_2 (1\text{ atm}) \rightarrow 2\,CO_2 (1\text{ atm}) + H_2O\,(l)$$
$$\Delta G° = -894.0 \text{ kJ mol}^{-1} = \Delta G_c° \text{ of acetate}^-$$

**TABLE 6-7**
**Empirical Bond Energies and Resonance Energies[a]**

| Energy values for single bonds (kJ mol$^{-1}$) | | Energy values for multiple bonds (kJ mol$^{-1}$) | | Empirical resonance energy values (kJ mol$^{-1}$) | |
|---|---|---|---|---|---|
| H–H | 436 | C=C | 615 | Benzene | 155 |
| C–C | 346 | N=N | 418 | Naphthalene | 314 |
| Si–Si | 177 | O=O | 402 (‘Δ state) | | |
| N–N | 161 | C=N | 615 | Styrene | 155 + 21 |
| O–O | 139 | C=O | 686 (formaldehyde) | Phenol | 155 + 29 |
| S–S | 213 | | 715 (aldehydes) | | |
| C–H | 413 | | 728 (ketones) | Benzaldehyde | 155 + 17 |
| Si–H | 295 | C=S | 477 | Pyridine | 180 |
| N–H | 391 | C≡C | 812 | Pyrrole | 130 |
| P–H | 320 | N≡N | 946 ($N_2$) | Indole | 226 |
| O–H | 463 | C≡N | 866 (HCN) | | |
| S–H | 339 | | 891 (nitriles) | $-C(=O)OH$ | 117 |
| C–Si | 290 | | | | |
| C–N | 292 | | | $-C(=O)OR$ | 100 |
| C–O | 351 | | | | |
| C–S | 259 | | | $-C(=O)NH_2$ | 88 |
| C–F | 441 | | | | |
| C–Cl | 328 | | | | |
| C–Br | 276 | | | Urea | 155 |
| C–I | 240 | | | CO | 439 |
| Si–O | 369 | | | $CO_2$ | 151 |

[a] From Pauling, L. ( 1960) *The Nature of the Chemical Bond*, 3rd ed., pp. 85, 189, and 195–198, Cornell Univ. Press, Ithaca, New York

Table 6-4 lists values of $\Delta G_c°$ as well as those of $\Delta G_f°$. Besides $CO_2$ and $H_2O$, the other products assumed in calculating the values of $\Delta G_c°$ in Table 6-4 are $N_2$ (1atm), $H_3PO_4$ (1 M), and $H_2SO_4$ (1 M).

Much of the oxidation occurring in cells is carried out by the biological oxidizing agent **nicotinamide adenine dinucleotide** (**NAD$^+$**) or by the closely related NADP$^+$ (Chapter 10). It is convenient to tabulate values of $\Delta G°$ for complete oxidation of compounds to $CO_2$ using NAD$^+$ rather than $O_2$ as the oxidizing agent. These values, designated $\Delta G°_{ox}$ and $\Delta G'_{ox}$ (pH 7), are also given in Table 6-3. Notice that these values are relatively small, corresponding to the fact that little energy is made available to cells by oxidation with NAD$^+$; for example (Eq. 6-57):

$$\text{Acetate}^- + 2\,H_2O + 4\,NAD^+ \rightarrow 2\,CO_2 + 4\,NADH + 3\,H^+\,(10^{-7}\,M)$$
$$\Delta G'_{ox}\,(pH\ 7) = +22.1\ kJ\ mol^{-1} \qquad (6\text{-}57)$$

When the reduced NAD$^+$ (designated NADH) so formed is reoxidized in mitochondria (Eq. 6-58), a large amount of energy is made available to cells.

$$4\,NADH + 4\,H^+ + 2\,O_2 \rightarrow 4\,H_2O + 4\,NAD^+$$
$$\Delta G'_{ox}\,(pH\ 7) = -876.1\ kJ\ mol^{-1} \qquad (6\text{-}58)$$

The sum of Eqs. 6-57 and 6-58 is the equation for combustion of acetate by $O_2$ (Eq. 6-59) and the two $\Delta G$ values sum to $\Delta G_c$ for acetate$^-$.

$$\text{Acetate}^- + H^+ + 2\,O_2 \rightarrow 2\,CO_2 + 2\,H_2O$$
$$\Delta G_c° = -894.0\ kJ\ mol^{-1}$$
$$\Delta G_c'\,(pH\ 7) = -854.0\ kJ\ mol^{-1} \qquad (6\text{-}59)$$

The values of $\Delta G_{ox}$ (Table 6-4) not only give an immediate indication of the relative amounts of energy available from oxidation of substrate with NAD$^+$ but also are very convenient in evaluating $\Delta G$ for fermentation reactions. For example, consider the fermentation of glucose to ethanol (Eq. 6-60):

$$\alpha\text{-D-Glucose} \rightarrow 2\,CO_2 + 2\ \text{ethanol} \qquad (6\text{-}60)$$

The Gibbs energy change $\Delta G'$ (pH 7) for fermentation of glucose to ethanol and $CO_2$ can be written immediately from the data of Table 6-4 (Eq. 6-61).

$$\Delta G'\,(pH\ 7) = -243.8 - 2(-4.6) = -234.6\ kJ\ mol^{-1} \qquad (6\text{-}61)$$

The values of $\Delta G_{ox}$ for $H_2O$, $CO_2$, and $H^+$ are always zero and need not be considered. The same computation can be made using the awkwardly large values of $\Delta G_c°$ or using values of $\Delta G_f°$ The latter are also large, and $CO_2$ and water must be considered in the equations. Table 6-4 can be used to obtain values of $\Delta G°$ for many metabolic reactions considered later in the book. Data from any column in the table may be used for this purpose, but for simplicity try the values in the $\Delta G'_{ox}$ column.

Note, however, that for reactions involving oxidation by $O_2$ or by any oxidant, other than NAD$^+$, not appearing in Table 6-4, the following two-step procedure is necessary if the $\Delta G_{ox}$ values are used. From the $\Delta G_{ox}$ values compute $\Delta G°$ or $\Delta G'$ for the reaction under consideration using NAD$^+$ as the oxidant. Then add to this the Gibbs energy of oxidation by $O_2$ (or other oxidant) of the appropriate number of moles of NADH formed. The latter is given for $O_2$ in Table 6-8 and can also be evaluated for a number of other oxidants, such as $Fe^{3+}$ and cytochrome *c*, from the data in Table 6-8.

## BOX 6-B MAGNESIUM

After potassium ion $Mg^{2+}$ is the most abundant cation present in tissues[a,b]. The average adult ingests ~10–12 mmol of magnesium ion daily (~1/4 g). Of this, about one-third is absorbed from the digestive tract. An equivalent amount is excreted in the urine to maintain homeostasis. Sixty percent of the magnesium in the body is found in the bones. The total $Mg^{2+}$ concentration in serum is ~1 mM, while various tissues contain as much as 5–17 mM.[c] The concentration of *free* $Mg^{2+}$ is difficult to measure. A variety of measurements using $^{31}P$ NMR spectroscopy of ATP,[d–f] $^{19}F$ NMR of added fluorine-containing chelators,[g] fluorescent chelators,[h] ion selective electrodes,[c] and other indirect procedures[g,i,j] indicate that the free magnesium concentration is almost the same in extracellular fluids, cytosol, and mitochondria.[h] Most values have been about 0.5 mM, somewhat less in erythrocytes. However, the most recent estimates indicate an intracellular [$Mg^{2+}$] of 0.8–1.1 mM.[c,f]

The additional $Mg^{2+}$ present in cells is bound to proteins, nuclei acids, and soluble compounds such as ATP, ADP, citrate,[k] and other phosphate- and carboxylate-containing substances. The binding is reversible and equilibration is usually rapid. It has been suggested that [$Mg^{2+}$], like [$H^+$], remains relatively constant within cells and that these two ions are in free equilibrium in the blood serum.[l] Nevertheless, there are instances in which at least temporary alterations in concentrations of both free $Mg^{2+}$ and $H^+$ occur.[m] During rapid catabolism of carbohydrates the formation of lactic acid by glycolysis leads to acidification of muscle cells, the pH falling from 7.3 to as low as 6.3. This drop in pH causes a large decrease in the extent of binding of $Mg^{2+}$ to molecules such as ATP and to a transient increase in [$Mg^{2+}$]. Similarly, the release of bisphosphoglycerate from hemoglobin upon oxygenation leads to a decreased concentration of free $Mg^{2+}$ as the latter coordinates with bisphosphoglycerate.[n] The 0.25 mM [$Mg^{2+}$] of aerobic red blood cells rises to 0.67 mM under anaerobic conditions.[d] Such changes in the free $Mg^{2+}$ concentrations will affect many equilibrium[o] and may be of significance in metabolic regulation.

The magnesium ion has a smaller radius than $Ca^{2+}$, a fact that may account for its more ready entry into cells. $Mg^{2+}$ can often be replaced by $Mn^{2+}$ with full activity for enzymes that require it. On the other hand, high concentrations of $Ca^{2+}$ are often antagonistic to $Mg^{2+}$. This antagonism is clearly seen in the effect of the two ions on irritability of protozoa.[p] Both deficiency of magnesium and excess of calcium in the surrounding medium cause increased irritability. Excess magnesium leads to anesthesia. The $Mg^{2+}$ concentration is high in hibernating animals.

Over 300 enzymes are dependent upon $Mg^{2+}$, the largest single group being the phosphotransferases, for which MgATP complex may be regarded as the substrate.[q] Magnesium has a special role in photosynthesis as a component of chlorophyll.

One of the most toxic metals is beryllium. It has been suggested that $Be^{2+}$ competes with $Mg^{2+}$ at many enzyme sites, including those of phosphoglucomutase and of phosphatases. However, as pointed out by Petsko,[r] because of its small size beryllium tends not to form $Be^{2+}$ ions but, rather, covalent complexes such as $BeF_3{\bullet}OH_2^-$, $BeF_4^{2-}$, and MgADP•-$BF_2$(OH). In the latter complex beryllium is covalently linked to the oxygen atom of the β phospho group of ADP to give an analog of MgATP, which inhibits many enzymes. See for example, Fig. 19-16A, in which MgADP•$BeF_x$ occupies the active site of myosin.


a Cowan, J. A., ed. (1995) *The Biological Chemistry of Magnesium*, VCH Publ., New York

b Strata, P., and Carbone, E., eds. (1991) *$Mg^{2+}$ and Excitable Membranes*, Springer-Verlag, Berlin and New York

c McGuigan, J. A. S., Blatter, L. A., and Buri, A. (1991) in *$Mg^{2+}$ and Excitable Membranes* (Strata, P., and Carbone, E., eds), Springer-Verlag, Berlin and New York

d Gupta, R. K., and Moore, R. D. (1980) *J. Biol. Chem.* **255**, 3987–3993

e Garfinkel, L., and Garfinkel, D. (1984) *Biochemistry* **23**, 3547–3552

f Clarke, K., Kashiwaya, Y., King, M. T., Gates, D., Keon, C. A., Cross, H. R., Radda, G. K., and Veech, R. L. (1996) *J. Biol. Chem.* **271**, 21142–21150

g Levy, L. A., Murphy, E., Raju, B., and London, R. E. (1988) *Biochemistry* **27**, 4041–4048

h Jung, D. W., Apel, L., and Brierley, G. P. (1990) *Biochemistry* **29**, 4121–4128

i Corkey, B. E., Duszynski, J., Rich, T. L., Matschinsky, B., and Williamson, J. R. (1986) *J. Biol. Chem.* **261**, 2567–2574

j Magneson, G. R., Puvathingal, J. M., and Ray, W. J., Jr. (1987) *J. Biol. Chem.* **262**, 11140–11148

k Kwack, H., and Veech, R. L. (1992) *Curr. Top. Cell. Regul.* **33**, 185–207

l Veloso, D., Guynn, R. W., Oskarrson, M., and Veech, R. L. (1973) *J. Biol. Chem.* **248**, 4811–4819

m Purich, D. L., and Fromm, H. J. (1972) *Curr. Top. Cell. Regul.* **6**, 131–167

n Bunn, H. F., Ransil, B. J., and Chao, A. (1971) *J. Biol. Chem.* **246**, 5273–5279

o Cornell, N. W. (1979) *J. Biol. Chem.* **254**, 6522–6527

p Meli, J., and Bygrave, F. L. (1972) *Biochem. J.* **128**, 415–420

q Vink, R., McIntosh, T. K., and Faden, A. I. (1991) in *$Mg^{2+}$ and Excitable Membranes* (Strata, P., and Carbone, E., eds), Springer-Verlag, Berlin and New York

r Petsko, G. A. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 538–540

## C. Electrode Potentials and Gibbs Energy Changes for Oxidation–Reduction Reactions

We live under a blanket of the powerful oxidant $O_2$. By cell respiration oxygen is reduced to $H_2O$, which is a very poor reductant. Toward the other end of the scale of oxidizing strength lies the very weak oxidant $H^+$, which some bacteria are able to convert to the strong reductant $H_2$. The $O_2$–$H_2O$ and $H^+$ – $H_2$ couples define two biologically important oxidation–reduction (**redox**) systems. Lying between these two systems are a host of other pairs of metabolically important substances engaged in oxidation–reduction reactions within cells.

There are two common methods for expressing the oxidizing or reducing powers of redox couples in a quantitative way. On the one hand, we can list values of $\Delta G$ for oxidation of the reduced form of a couple to the oxidized form by $O_2$. A compound with a large value of $-\Delta G$ for this oxidation will be a good reductant. An example is $H_2$ for which $\Delta G$ of combustion at pH 7 (Table 6-4) is – 237 kJ / mol. Poor reductants such as $Fe^{2+}$ are characterized by small values of $\Delta G$ of oxidation ($-8.5$ kJ mol$^{-1}$ for $2\ Fe^{2+} \rightarrow 2\ Fe^{3+}$). The Gibbs energies of oxidation of biological hydrogen carriers, discussed in Chapter 15, for the most part fall between those of $H_2$ and $Fe^{2+}$.

A second way of expressing the same information is to give **electrode potentials** (Table 6-8). Electrode potentials are also important in that their direct measurement sometimes provides an experimental approach to the study of oxidation–reduction reactions within cells. To measure an electrode potential it must be possible to reduce the oxidant of the couple by flow of electrons (Eq. 6-62) from an electrode surface, often of specially prepared platinum.

$$A + 2\,H^+ + 2\,e^- \rightarrow AH_2 \qquad (6\text{-}62)$$

Equation 6-62 represents a reversible reaction taking place at a single electrode. A complete electrochemical cell has two electrodes and the reaction occurring is the sum of two half-reactions. The electrode potential of a given half-reaction is obtained from the measured electromotive force of a complete cell in which one half-reaction is that of a standard **reference electrode** of known potential. Figure 6-3 indicates schematically an experimental setup for measurement of an electrode potential. The standard hydrogen electrode consists of platinum over which is bubbled hydrogen gas at one atmosphere pressure. The electrode is immersed in a solution containing hydrogen ions at unit activity ($a_{H^+} = 1$). The potential of such an electrode is conventionally taken as zero. In practice it is more likely that the reference electrode will be a calomel electrode or some other electrode that has been established experimentally as reliable and whose potential is accurately known.

The standard electrode is connected to the experimental electrode compartment by an electrolyte-filled bridge. In the experimental compartment the reaction represented by Eq. 6-62 occurs at the surface of another electrode (often platinum). The voltage difference between the two electrodes is measured with a potentiometer. The difference between the observed voltage and that of the reference electrode gives the electrode potential of the couple under investigation. It is important that the electrode reaction under study be strictly reversible. When the electromotive force (emf) of the experimental cell is balanced with the potentiometer against an external voltage source, no current flows through the cell. However, for a reversible reaction a slight change in the applied voltage will lead to current flow. The flow will be in either of the two directions, depending upon whether the applied voltage is raised or lowered.

Not all redox couples are reversible. This is especially true of organic compounds; for example, it is not possible to determine readily the electrode potential for an aldehyde–alcohol couple. In some cases, e.g., with enzymes, a readily reducible dye with a potential similar to that of the couple being measured can be added. A list of suitable dyes has been described by Dutton.[51] If the dye is able to rapidly exchange electrons with the couple being studied, it is still possible to measure the electrode potential directly. In many cases electrode potentials appearing in tables have been calculated from Gibbs energy data. The student should be able to calculate many of the potentials in Table 6-8 from Gibbs energy data from Table 6-4. If A, $H^+$, and $AH_2$ are all present at unit activity in the experimental cell, the observed potential for the half-reaction will be the **standard electrode potential** $E°$. If the emf of the hypothetical cell with the standard hydrogen electrode is positive when electron flow is in the direction indicated by the arrow in Fig. 6-3, the potential of the couple $A / AH_2$ is also taken as positive (and is often called a **reduction potential**). This is the convention used in establishing Table 6-8, but potentials of exactly the opposite sign (oxidation potentials) are used by some chemists. To avoid confusion in reading it is best to be familiar with values of one or two potentials such as those of the $O_2$–$H_2O$ and $NAD^+$–NADH couples.

When electrons flow in the external circuit the maximum amount of work that can be done per mole of electrochemical reaction ($-\Delta G$) is given by Eq. 6-63

$$\begin{aligned}\Delta G &= nEF = nE \times 96.487 \text{ kJ mol}^{-1}\text{ V}^{-1} \\ &= -\,nE \times 23.061 \text{ kcal mol}^{-1}\text{ V}^{-1}\end{aligned} \qquad (6\text{-}63)$$

where $F$ equals the number of coulombs per mole of electrons (Avogadro's number multiplied by the charge on the electron = 96,487 coulombs). $E$ is measured in volts and represents the difference of the

electrode potentials of the two half-cells. In the case of a cell using the standard hydrogen electrode, $E$ is the electrode potential of the experimental couple. The number of moles of electrons transferred in the reaction equation ($n$) is usually 1 or 2 in biochemical reactions (2 for Eq. 6-62).

Since the reactants and products need not be at unit activity, we must define the observed electrode potential $E$ as a function of $E°$ and the activities (concentrations) of A, $AH_2$, and $H^+$ (Eq. 6-64).

$$E = E° + \frac{RT}{nF} \ln \frac{[A][H^+]^2}{[AH_2]}$$

If $n = 2$

$$E = E° + 0.0296 \log \frac{[A][H^+]^2}{[AH_2]} \quad \text{volts at } 25°C \tag{6-64}$$

In the biochemical literature values of the apparent standard electrode potential at pH 7 ($E°'$) are usually tabulated instead of values of $E°$ (Table 6-8, second column). Note that $E°'$ (pH 7) for the hydrogen electrode is not zero, as it is at pH 0, but −0.414 V. Values of $E$ are related to $E°'$ by Eq. 6-64 with $[H^+]^2$

**TABLE 6-8**
**Reduction Potentials of Some Biologically Important Systems[a,b]**

| Half-reaction | $E°$ (V) | $E°'$ (pH 7) (V) | $-\Delta G'$ (pH 7) (kJ mol$^{-1}$) for oxidation by $O_2$ (per 2 electrons) |
|---|---|---|---|
| $O_2 + 4\,H^+ + 4e^- \rightarrow 2\,H_2O$ | +1.229 | +0.815 | 0.0 |
| $Fe^{3+} + e^- \rightarrow Fe^{2+}$ | 0.771 | 0.771 | 8.5 |
| $NO_3^- + 2\,H^+ + 2e^- \rightarrow NO_2^- + H_2O$0.42176.0 | | | |
| Cytochrome *f* ($Fe^{3+}$) + $e^- \rightarrow$ cytochrome *f* ($Fe^{2+}$) | | 0.365 | 86.8 |
| Fe $(CN)_6^{3-}$ (ferricyanide) + $e^- \rightarrow$ Fe $(CN)_6^{4-}$ | | 0.36 | 87.8 |
| $O_2 + 2H^+ + 2\,e^- \rightarrow H_2O_2$ | 0.709 | 0.295 | 100.3 |
| Cytochrome *a* ($Fe^{3+}$) + $e^- \rightarrow$ Cytochrome *a* ($Fe^{2+}$) | | 0.29 | 101.3 |
| *p*-Quinone + 2 $H^+$ + 2 $e^- \rightarrow$ hydroquinone | 0.699 | 0.285 | 102.3 |
| Cytochrome *c* ($Fe^{3+}$) + $e^- \rightarrow$ cytochrome *c* ($Fe^{2+}$) | | 0.254 | 108.3 |
| Adrenodoxin ($Fe^{3+}$) + $e^- \rightarrow$ adrenodoxin ($Fe^{2+}$) | | 0.15 | 128.3 |
| Cytochrome $b_2$ ($Fe^{3+}$) + $e^- \rightarrow$ cytochrome $b_2$ ($Fe^{2+}$) | | 0.12 | 134.1 |
| Ubiquinone + 2 $H^+$ + 2 $e^- \rightarrow$ ubiquinone $H_2$ | | 0.10 | 138.0 |
| Cytochrome *b* ($Fe^{3+}$) + $e^- \rightarrow$ cytochrome *b* ($Fe^{2+}$) | | 0.075 | 142.8 |
| Dehydroascorbic acid + 2 $H^+$ + 2 $e^- \rightarrow$ ascorbic acid | | 0.058 | 146.1 |
| Fumarate$^{2-}$ + 2 $H^+$ + 2 $e^- \rightarrow$ succinate$^{2-}$ | | 0.031 | 151.3 |
| Methylene blue + 2 $H^+$ + 2 $e^- \rightarrow$ leucomethylene blue (colorless) | | 0.011 | 155.2 |
| Crotonyl-CoA + 2 $H^+$ + 2 $e^- \rightarrow$ butyryl-CoA | | −0.015 | 160.2 |
| Glutathione + 2 $H^+$ + 2 $e^- \rightarrow$ 2-reduced glutathione | | ~ −0.10 | 176.6 |
| Oxaloacetate$^{2-}$ + 2 $H^+$ + 2 $e^- \rightarrow$ malate$^{2-}$ | | −0.166 | 189.3 |
| Pyruvate$^-$ + 2 $H^+$ + 2 $e^- \rightarrow$ lactate$^{1-}$ | | −0.185 | 193.0 |
| Acetaldehyde + 2 $H^+$ + 2 $e^- \rightarrow$ ethanol | | −0.197 | 195.3 |
| Riboflavin + 2 $H^+$ + 2 $e^- \rightarrow$ dihydroriboflavin | | −0.208 | 197.4 |
| Acetoacetyl-CoA + 2 $H^+$ + 2 $e^- \rightarrow$ β-hydroxybutyryl-CoA | | −0.238 (38°C) | 203.2 |
| S + 2 $H^+$ + 2 $e^- \rightarrow H_2S$ | 0.14 | −0.274 | 210.2 |
| Lipoic acid + 2 $H^+$ + 2 $e^- \rightarrow$ dihydrolipoic acid | | −0.29 | 213.2 |
| $NAD^+ + H^+ + 2\,e^- \rightarrow NADH$ | −0.113 | −0.32 | 219.0 |
| $NADP^+ + H^+ + 2\,e^- \rightarrow NADPH$ | | −0.324 | 219.8 |
| Ferredoxin ($Fe^{3+}$) + $e^- \rightarrow$ ferredoxin ($Fe^{2+}$) (*Clostridia*) | | −0.413 | 237.0 |
| 2 $H^+$ + 2 $e^- \rightarrow H_2$ | 0 | −0.414 | 237.2 |
| $CO_2 + H^+ + 2\,e^- \rightarrow$ formate$^-$ | | −0.42 (30°C) | 238.3 |
| Ferredoxin ($Fe^{3+}$) + $e^- \rightarrow$ ferredoxin ($Fe^{2+}$) (spinach) | | −0.432 | 240.6 |

[a] A compound with a more positive potential will oxidize the reduced form of a substance of lower potential with a standard free energy change $\Delta G° = -nF\,\Delta E° = -n\,\Delta E° \times 96.49$ kJ mol$^{-1}$ where $n$ is the number of electrons transferred from reductant to oxidant. The temperature is 25°C unless otherwise indicated. $E°$ refers to a standard state in which the hydrogen ion activity = 1; $E°'$ refers to a standard state of pH 7, but in which all other activites are unity.

[b] The major source is Loach, P. A. (1976) in *Handbook of Biochemistry and Molecular Biology* 3rd ed. Vol. I (Fasman, G. D. ed.), pp. 122–130, CRC Press, Cleveland, Ohio.

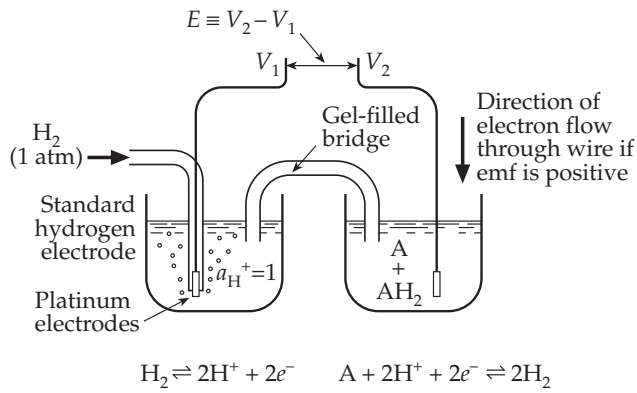
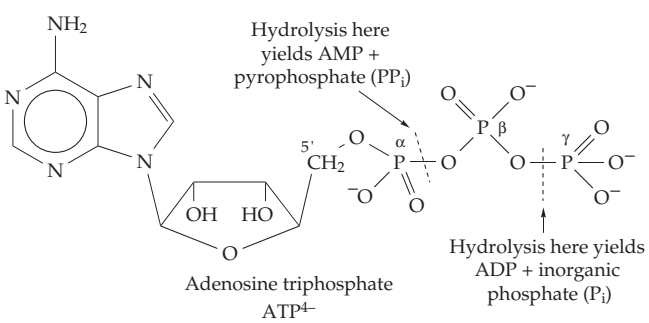



**Figure 6-3** Device for measurement of electrode potentials. The electrode reactions are indicated below each half-cell. The maximum electrical work that can be done by such a cell on its surroundings is $-\Delta G = nEF$, where $E = V_2 - V_1$ as measured by a potentiometer. If A is reduced to $AH_2$ by $H_2$, electrons will flow through an external circuit as indicated. A will be reduced in the right-hand cell. $H_2$ will be oxidized to $H^+$ in the left-hand cell. Protons will flow through the gel bridge from left to right as one of the current carriers in the internal circuit.

deleted from the numerator (since the term in log [$H^+$] is contained in $E$ °′). On the scale of $E$ °′ (pH 7) the potential of the oxygen–water couple is 0.815 V, while that of the $NAD^+$–NADH couple is –0.32 V.

## D. The Adenylate System




Of central importance to the energy metabolism of all cells is the **adenylate system** which consists of adenosine 5'-triphosphate (**ATP**), adenosine 5'-diphosphate (**ADP**), and adenosine 5'-monophosphate (**AMP**) together with **inorganic phosphate** ($P_i$), **pyrophosphate** ($PP_i$), and **magnesium** ions. Remember that $P_i$ refers to the mixture of ionic forms of phosphoric acid present under experimental conditions. Between pH 4 and pH 10 this will be mainly $H_2PO_4^-$ ($pK_a = 6.8$) and $HPO_2^{2-}$. Likewise, the symbols AMP, ADP, and ATP refer to mixtures of ionic forms and $PP_i$ refers to a mixture of the ions of pyrophosphoric acid. Above pH 4.4 only $H_2P_2O_7^{2-}$ ($pK_a = 6.1$), $HP_2O_7^{3-}$ ($pK_a = 9.0$), and $P_2O_7^{4-}$ contribute appreciably to $PP_i$.

### 1. Storage and Utilization of Energy

ATP is a thermodynamically unstable molecule with respect to hydrolysis to either ADP or AMP as is indicated in the foregoing diagram. The standard Gibbs energy of hydrolysis, $\Delta G'$, of $ATP^{4-}$ to $ADP^{3-}$ + $HPO_4^{2-}$ at pH 7 is –34.5 kJ mol$^{-1}$ and that of hydrolysis of $ADP^{3-}$ to $AMP^{2-}$ + $HPO_4^{2-}$ is –36.3 kJ mol$^{-1}$ at 25°C (Table 6-6). The exact value of these changes in Gibbs energy depends on pH and on the concentration of $Mg^{2+}$ as is detailed in Sections B, 4, 5. Rates of reaction of components may also depend upon metal ions. Magnesium ion is especially important and complexes such as $MgATP^{2-}$ are regarded as the true substrates for many ATP-utilizing enzymes.

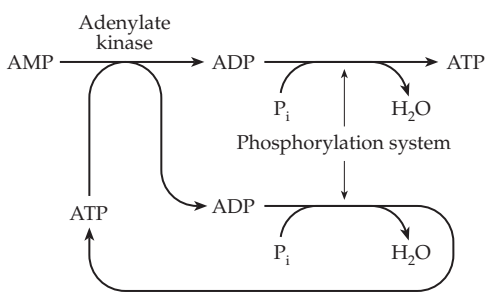
The large *Gibbs energy decreases upon hydrolysis (high group transfer potentials)* enable cells to use ATP and ADP as stores of readily available energy. Energy from the adenylate system is used for many purposes including biosynthesis, transport of ions and molecules across membranes, and for doing mechanical work. The mechanisms by which this energy is utilized are considered later (see Chapters 10, 12, and 17–19). The first step most often involves transfer of either the terminal γ-phospho group of ATP to a site on a different molecule or transfer of the entire AMP portion of the molecule onto another group. Thus the products of cleavage of ATP in these energy-utilizing processes may be either ADP + $P_i$ or AMP + $PP_i$. In the latter case the pyrophosphate is usually hydrolyzed rapidly to two molecules of $P_i$ by pyrophosphatases. This process, too, serves an essential function in the adenylate system, because it removes a product of the initial ATP cleavage reaction, shifting the overall equilibrium of the reaction sequence in the direction of the products.

The adenylate system provides the major store of rapidly available energy but the whole family of **nucleoside triphosphates** that are related to ATP in structure have similar functions. These include guanosine triphosphate (GTP), uridine triphosphate (UTP), cytidine triphosphate (CTP), and deoxyribonucleotide triphosphates (Table 5-1). These compounds are formed by successive transfer to the nucleoside monophosphate (GMP, UMP, CMP, dGMP, etc.) from two different moleclues of ATP of two phospho groups. The resulting compounds are used to provide energy for a variety of specific biosynthetic processes, including synthesis of RNA and DNA. Inorganic **polyphosphates**, linear polymers of orthophosphate ($P_i$), are present in nearly all living forms.[52–54] Like ATP, they can also store

energy and in some organisms substitute for ATP in certain enzymatic reactions.

## 2. Synthesis of ATP

The **phosphate anhydride** (pyrophosphate) linkages of ATP are generated by the joining of ADP and inorganic phosphate by means of special **phosphorylation** reactions. The most important of the latter occur in the photosynthetic membranes of chloroplasts (**photosynthetic phosphorylation**) and in oxygen-utilizing membranes of bacteria and of mitochondria (**oxidative phosphorylation**). Conversion of AMP to ADP is accomplished by transfer of the terminal phospho group from an ATP molecule to AMP in a reaction catalyzed by the extremely active enzyme **adenylate kinase** (Chapter 12) which is found in all cells. The following equations indicate how one of the special phosphorylation reactions must be used twice for the conversion of one molecule of AMP into one molecule of ATP.




(6-65)

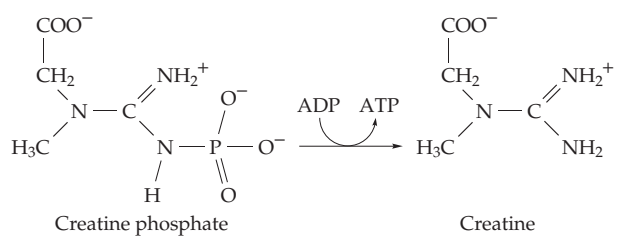
Various measures of the phosphorylating potential of the adenylate system within cells have been proposed. One measure is the product [ATP] / [ADP][$P_i$], which will be called the **phosphorylation state ratio** or $R_p$ in this book. It is also sometimes called the "phosphorylation potential." This ratio directly affects the Gibbs energy of hydrolysis of ATP, as is shown by Eq. 6-29. The value of $R_p$ may be as high as $10^4$ to $10^5$ $M^{-1}$ within cells[55] adding – 22.8 kJ $mol^{-1}$ to $\Delta G$ of hydrolysis of ATP. Another quantity that is sometimes cited is the "energy charge," the mole fraction of adenylic acid "charged" by conversion to ATP. ADP is regarded as "half-charged."[56,57]

$$\text{Energy charge} = \frac{[\text{ATP}] + \frac{1}{2}[\text{ADP}]}{[\text{ATP}] + [\text{ADP}] + [\text{AMP}]} \tag{6-66}$$

The energy charge varies from 0 if only AMP is present to 1.0 if all of the AMP is converted to ATP. Measurements on a variety of cells and tissues show that the energy charge is usually between 0.75 and 0.90. Although it is easy to calculate its numerical value, the energy charge cannot be used in chemical equations and the idea that the energy charge of a cell plays a key role in regulation of metabolism has been challenged.[58]

## 3. Creatine Phosphate, an Energy Buffer

Although ATP provides the immediate source of energy for operating muscle, its concentration is only about 5 mM. However, muscle contains, in addition, a **phosphagen**, an *N*-phospho derivative of a guanidinium compound. In mammalian muscle, the phosphagen is **creatine phosphate**. Related compounds including **arginine phosphate** serve in various invertebrates. The group transfer potential ($-\Delta G°'$ of hydrolysis) for creatine phosphate is 43.1 kJ $mol^{-1}$. Thus, the transfer of a phospho group from creatine phosphate to ADP to form ATP (Eq. 6-67) is spontaneous with $\Delta G' = -8.6$ kJ $mol^{-1}$. Recent values under a variety of conditions have been reported by Taugue and Dobson.[59] Creatine phosphate, which is present at a concentration of 20 mM, provides a reserve of high-energy phospho groups and keeps the adenylate system of muscle buffered at a high value of $R_p$.




$\Delta G' = -8.6$ kJ $mol^{-1}$ (6-67)

## 4. Phosphorus-31 NMR

A spectacular development is the ability to observe components of the adenylate system as well as phosphocreatine and other phosphate esters in living cells by $^{31}P$ NMR.[60–61a] Spectra can be recorded on suspensions of cells[60] or of mitochondria,[62] on individual excised muscles (Fig. 6-4),[63,64] or on perfused organs[65,66] including beating rat hearts and surgically exposed animal organs.[61,61a] Metabolism can be observed in human erythrocytes[67] and even in human limbs, liver, hearts, and brain.[68–70] The method can play a valuable role in understanding human diseases.[63,67,70] As is seen in Fig. 6-4, each phosphorus atom of ATP gives a distinct resonance. The area of the $P_\gamma$ peak provides a direct measure of the ATP concentration, and

phosphocreatine (creatine-*P*) and $P_i$ can be estimated in a similar way. Knowing the concentrations of ATP, $P_i$, and creatine-*P*, the amount of ADP present, usually too low to be estimated by NMR, may be calculated.[61] Barnacle muscle contains, instead of phosphocreatine, a high concentration of phosphoarginine which has also been measured by NMR.[71]

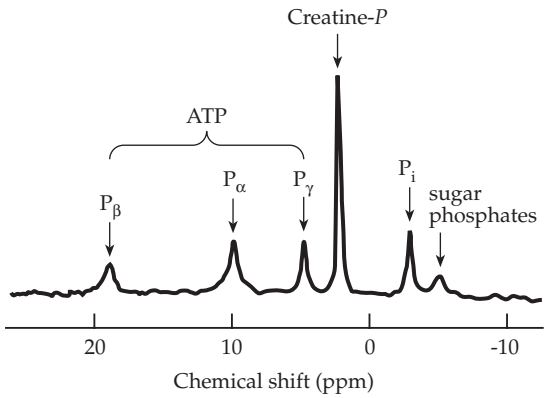



**Figure 6-4** Phosphorus-31 NMR spectrum of an excised rat muscle (*vastus lateralis*) in Ringer solution at 15°C. The spectrum represents the accumulation of 400 scans. From P. J. Seeley *et al.*[64]

Usually no ADP can be seen in NMR spectra, although there may be about 0.5 mM ADP according to chemical analysis of rapidly frozen tissues. It has been concluded that most ADP is bound to proteins. In muscle the proteins myosin and actin (Chapter 20) hold most of the ADP leaving only about 0.02 mM free.[62,72] The same conclusion has been reached on the basis of other evidence.[55]

If a perfused heart in an NMR spectrometer is deprived of oxygen, the level of phosphocreatine drops rapidly and that of $P_i$ increases while that of ATP remains relatively constant until the phosphocreatine is gone. Then it too falls and becomes undetectable after 17 min.[72] Similar changes occur when a tourniquet is placed on a human arm in the NMR spectrometer. Study of the rate and extent of recovery of the adenylate system when oxygen is readmitted is helping to provide a better means of protecting kidneys and other organs during transplantation operations. Diagnosis of circulatory ailments in human limbs by $^{31}P$ NMR may soon be routine. Use of radiofrequency coils placed on the body surface allows monitoring of the adenylate system in the heart, brain, and other tissues deep within the body.[69,73] Changes of concentration in the adenylate system and of phosphocreatine can be monitored very rapidly and evenly throughout the cardiac heartbeat cycle.[74]

## E. Complex Biochemical Equilibria

The binding of small molecules to larger ones is basic to most biological phenomena. Substrates bind to enzymes and hormones bind to receptors. Metal ions bind to ATP, to other small molecules, and to metalloproteins. Hydrogen ions bind to amino acids, peptides, nucleotides, and most macromolecules. In this section we will consider ways of describing mathematically the equilibria involved.

The strength of bonding between two particles can be expressed as a **formation constant** (or **association constant)** $K_f$. Consider the binding of a molecule X to a second molecule P, which may be a protein or some other macromolecule. If there is on the surface of P only one single binding site for X, the process can be described by Eq. 6-68 and the equilibrim constant $K_f$ by Eq. 6-69.

$$X + P \rightarrow PX \tag{6-68}$$

$$K_f = [PX] / [P][X] \tag{6-69}$$

The units for $K_f$ are liters per mole ($M^{-1}$). The constant $K_f$ is a direct measure of the strength of the binding: The higher the constant, the stronger the interaction. This fact can be expressed in an alternative way by giving the standard Gibbs energy change ($\Delta G°$) for the reaction (Eq. 6-70). The more negative $\Delta G_f°$, the stronger the binding.

$$\begin{aligned}\Delta G_f° &= -RT \ln K_f = -2.303RT \log K_f \\ &= -5.708 \log K_f \text{ kJ mol}^{-1} \text{ at } 25°\text{C}\end{aligned} \tag{6-70}$$

To avoid confusion, it is important to realize that the frequently used **dissociation constants** ($K_d$) are reciprocal association constants (Eq. 6-71). The use of

$$K_d = 1 / K_f \tag{6-71}$$

association constants and of dissociation constants is firmly entrenched in different parts of the chemical literature; be sure to keep them straight. Dissociation constants are customarily used to describe acid–base chemistry, while formation constants are more often employed to describe complexes with metal ions or associations of macromolecules (Section C). However, both types of equilibria can be described using either formation constants or dissociation constants.

Logarithms of formation constants, which are proportional to the Gibbs energies of association, are often tabulated. The logarithms of formation constants and $pK_a$ values of dissociation constants are identical (Eq. 6-72) and are a measure of the standard

$$\log K_f = -\log K_d = pK_d \tag{6-72}$$

Gibbs energy decrease in the association reaction. The difference in $\Delta G°$ corresponding to a change of one unit in log $K_f$ or p$K_d$ is –5.7 kJ mol$^{-1}$, a handy number to remember.

## 1. Effects of pH on Equilibria

Compounds that contain several acidic or basic groups can exist in a number of different ionic forms, $H_nA$, $H_{n-1}A$, $H_2A$, HA, A, etc., as is indicated in the following equation, which is identical in form to Eq. 3-3.

$$H_nA \xrightarrow{K_1} H_{n-1}A \longrightarrow \cdots H_2A \xrightarrow{K_{n+1}} HA \xrightarrow{K_n} A \qquad (6\text{-}73)$$

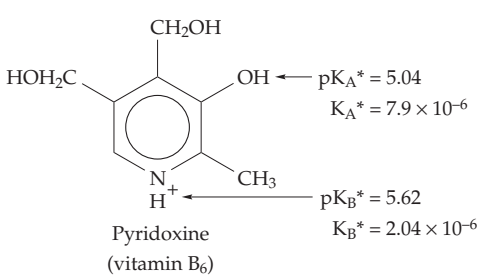
The dissociation constants $K_1$ ---- $K_n$ for a multiprotic acid $H_nA$ are defined as stepwise or **macroscopic constants** (also called molecular constants). For some compounds, e.g. alanine, the p$K_a$ values are far apart (p$K_1$ and p$K_2$ are 2.4 and 9.8, respectively). The macroscopic constants can be assigned specifically, $K_1$ to the carboxyl group and $K_2$ to the protonated amino group. At the isoelectric pH of 6.1 the alanine exists almost entirely as the dipolar ion. However, for compounds in which the macroscopic p$K_a$ values are closer together, they cannot be assigned to specific groups. We will consider some specific examples in the next section.

## 2. Microscopic Dissociation Constants and Tautomerization

A microscopic constant applies to a single site. Consider the dissociation of a simple carboxylic acid:

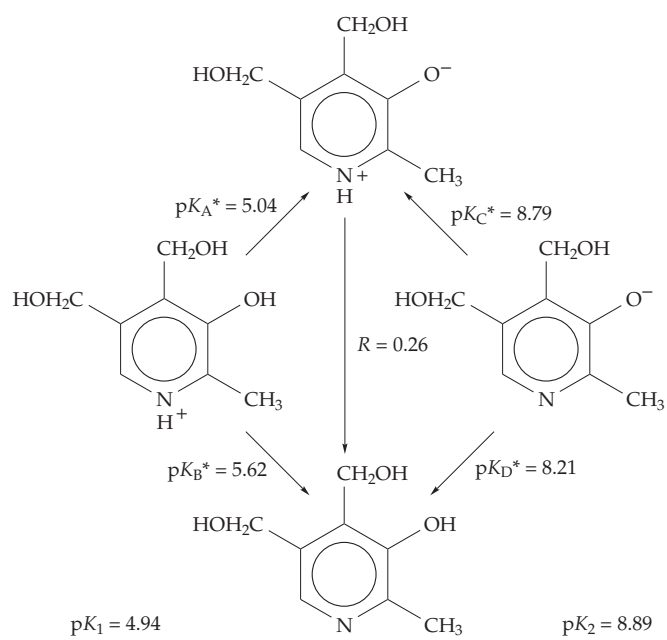
$$R—COOH \longrightarrow R—COO^- + H^+ \qquad (6\text{-}74)$$

The dissociation constant is about 1.7 x 10$^{-5}$ and p$K_a$ = 4.8. Since there is only one proton, the observed dissociation constant is also the microscopic dissociation constant. Now consider the cation of pyridoxine which has two dissociable protons bound to distinctly different sites, the phenolic oxygen and the ring nitrogen.




Either of the two protons might dissociate first as the pH is raised (Eq. 6-75). However, the two microscopic dissociation constants p$K_A$* and p$K_B$* are distinctly different. The result is that at 25°C in the neutral (monoprotonated) form 80% of the molecules carry a proton on the N, while the other 20% are protonated on the less basic – O$^-$. Notice that the subscripts a and b used in this discussion do *not* refer to acidic and basic but to the individual dissociation steps shown in Eq. 6-75. Microscopic constants will always be indicated with asterisks in this discussion.

The two monoprotonated forms of pyridoxine are the tautomeric pair shown in Eq. 6-75 and whose concentrations are related by the **tautomeric ratio**, $R$ = [neutral form]/[dipolar ion], a *pH-independent equilibrium constant* with a value of 0.204/0.796 = 0.26 at 25°C.[75] Evaluation of microscopic constants for dissociation of protons from compounds containing non-identical groups depends upon measurement of the tautomeric ratio, or ratios if more than two binding sites are present. In the case of pyridoxine, a spectrophotometric method was used to estimate $R$.




(6-75)

To calculate microscopic constants from stepwise constants and tautomeric ratios, consider Eq. 6-76 in which $[HP]_A$ and $[HP]_B$ are the concentrations of the two tautomers and $K_1$ is the first **stoichiometric** or **macroscopic** dissociation constant for the diprotonated species $H_2P$.

$$K_1 = \frac{([HP]_A + [HP]_B)[H^+]}{[H_2P]} = K_A^* + K_B^*$$

$$pK_1 = 4.94;\ K_1 = 1.15 \times 10^{-5} = 9.1 \times 10^{-6} + 2.4 \times 10^{-6} \qquad (6\text{-}76)$$

We see that $K_1$ is just the sum of the two microscopic constants $K_A$* and $K_B$* for dissociation of $H_2P$ to the pair of tautomers HP(A) and HP(B). In a similar fashion it can be shown that the second stoichiometric constant $K_2$ is related to the microscopic constants $K_C$* and $K_D$* for dissociation of HP(A) and HP(B) to form P (Eq. 6-77).

$$1 / K_2 = 1 / K_C^* + 1 / K_D^* \qquad (6\text{-}77)$$

Since the tautomeric ratio $R$ equals $[HP]_B / [HP]_A$, Eqs. 6-76 and 6-77 can be rearranged to Eqs. 6-78 to 6-81. These allow the evaluation of all of the microscopic constants from the two stoichiometric constants $K_1$ and $K_2$ plus the tautomeric ratio $R$.

$$pK_A^* = pK_1 + \log (1 + R) \qquad (6\text{-}78)$$

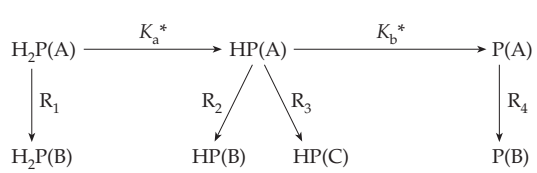
$$pK_B^* = pK_A^* - \log R \qquad (6\text{-}79)$$

$$pK_C^* = pK_2 - \log (1 + R) \qquad (6\text{-}80)$$

$$pK_D^* = pK_C^* + \log R \qquad (6\text{-}81)$$

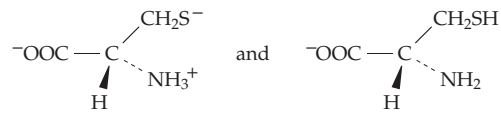
For pyridoxine $pK_1$ and $pK_2$ were determined spectrophotometrically as 4.94 and 8.89. These values, together with that of $R$ given above, were used to estimate the microscopic constants that are given in Eq. 6-74.[75] Notice that the microscopic constants of Eq. 6-74 are not all independent; if any three of the five equilibrium constants are known the other two can be calculated readily. In describing and measuring such equilibria it is desirable to select one pathway of dissociation, e.g., $H_2P \rightarrow HP(A) \rightarrow P$, and to relate the species HP(B) to it via the pH-independent constant $R$.

Often more complex situations arise in which additional tautomers or other forms arise via pH-independent reactions. These can all be related back to the reference ionic species by additional ratios $R$, which may describe equilibria for tautomerization, hydration, isomerization, etc. (Eq. 6-82).[76] In the case illustrated, only one of the ratios, namely $R_2$ or $R_3$, is likely to be a tautomerization constant because, as a rule, $H_2P$ and P will not have tautomers. Equations analogous to Eqs. 6-76 to 6-82 can be written easily to derive $K_C$*, $K_D$* and any other microscopic constants desired from the stoichiometric constants plus the ratios $R_1$ to $R_4$. While it is easy to describe tautomerism by equations such as Eqs. 6-76 and 6-82 *it is often difficult to measure the tautomeric ratios R*.[77] In favorable cases measurements of spectra of one kind or another allow their evaluation. However, because tautomerism may be extremely rapid, NMR spectra will often show only one peak for a proton present in a mixture of tautomers.




(6-82)

As was pointed out in Chapter 2, tautomerization ratios are often affected strongly by changes in solvent. Tautomerism among monoprotonated forms of cysteine, glutathione, and histidine (Eq. 2-6) has received considerable attention by biochemists[77–79] as has tautomerism in binding of protons and other small ligands to proteins.[80,81] For cysteine,[78,79] for which the following species coexist in the alkaline pH range, the distribution of the various ionic species including the two tautomers is shown in Fig. 6-5. A similar situation is met in the small protein thioredoxin (Box 15-C) which has a buried aspartate carboxylate that interacts with a nearby cysteine –SH group,[82] in papain where –SH and imidazole groups interact (Fig. 12-15),[83] and in carbohydrases where two or more carboxyl and carboxylate side chains interact (Chapter 12).[84]



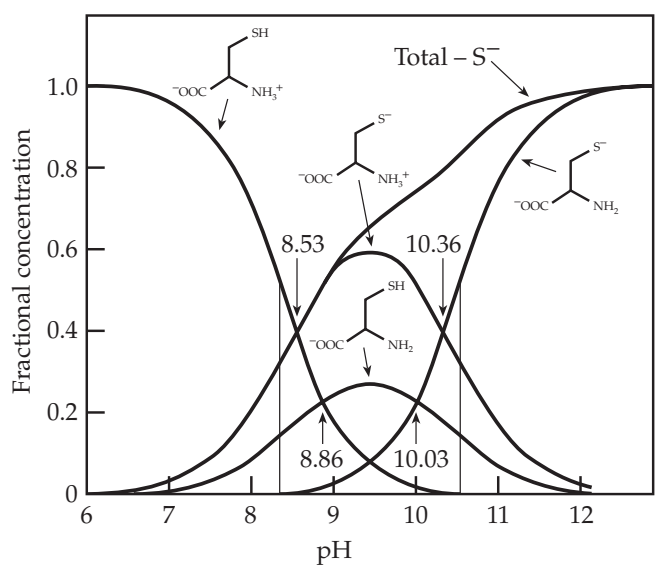
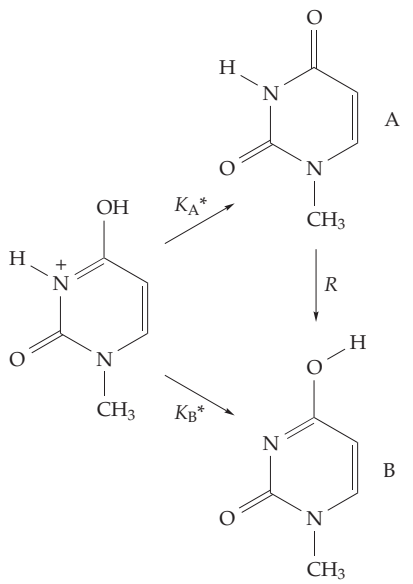



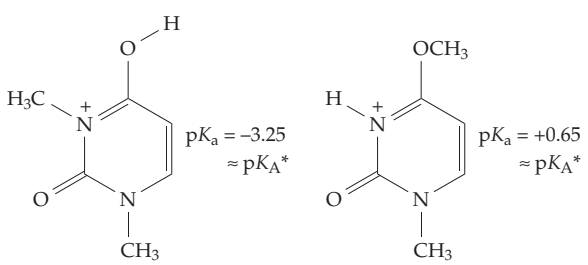

**Figure 6-5** Distribution of various ionic species of cysteine as a function of pH. The function in each ionic and tautomeric form is plotted. Microscopic $pK_a$ values are given numerically and macroscopic $pK_a$'s are indicated by the vertical lines. From Dixon and Tipton.[79]

Because it is usually difficult to measure tautomeric ratios, dissociation constants are frequently assumed identical to those of compounds in which one of the acidic groups has been modified by methylation or esterification or in some other manner to prevent dissociation of a proton. For example, the cation of 1-methyluracil can be dissociated to the two tautomers shown in Eq. 6-83.




(6-83)

The apparent p$K_a$ values were estimated for dissociation of the following two dimethylated cations which resemble that of Eq. 6-83.




It can reasonably be assumed that these p$K_a$ values approximate p$K_A$* and p$K_B$* as indicated. Thus, applying Eq. 6-85, log $R \approx -3.25 - 0.65 + 3.90$. This result indicates that tautomer A of Eq. 6-89 is overwhelmingly predominant in water.[85,86] It also suggests that, within experimental error, the p$K_a$ for 1-methyluracil will equal p$K_A$*, namely, –3.25. In fact, it is close to this (–3.40).[85]

## 3. The Binding of Metal Ions

Equilibria in the formation of complex ions with metals are treated exactly as is the binding of small molecules and ions to macromolecules.[87–89] Stepwise constants are defined for the formation of complexes containing one, two, or more ligands L bound to a central metal ion M. The binding constants $K_f$'s are usually referred to as β's as in Eq. 6-84.

$$\beta_1 = K_{f1} \quad \beta_n = K_{f1} K_{f2} \ldots K_{fn} \tag{6-84}$$

Many important questions can be asked about the binding of metal ions within living cells. For example, What fraction of a given metal ion is free and what fraction is bound to organic molecules? To what ligands is a metal bound? Since many metal ions are toxic in excess, it is clear that homeostatic mechanisms must exist. How do such mechanisms sense the free metal ion activity within cells? How does the body get rid of unwanted metal ions? Answers to all these questions depend upon the quantitative differences in the binding of metal ions to the variety of potential binding sites found within a cell.

Table 6-9 gives formation constants for 1:1 complexes of several metal ions and a number of inorganic as well as organic ligands.[89] Only the values of log $K_1$ are given when a series of stepwise constants have been established. However, in many cases two or more ligands can bind to the same metal ion. Thus for cupric ion and ammonia there are four constants.

$$Cu^{2+} + NH_3 \quad \log K_i = 4.0, 3.3, 2.7, 2.0$$

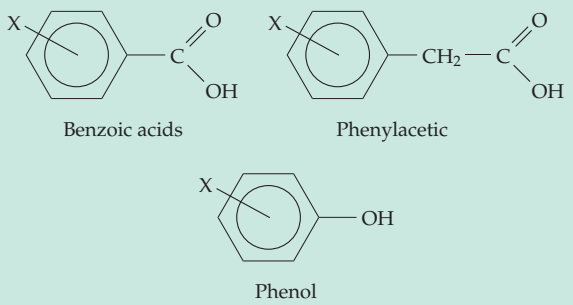
They are all separated by more than the "statistical distance," which in this case is less than the 0.6 logarithmic units for two equivalent binding sites (see Chapter 7, Section A,2). Thus, the second ligand binds less tightly than the first and **anticooperativity** in binding of successive ligands is observed. Most metal ions will also bind two or three successive amino acids. In the case of copper, whose preferred coordination number is four, two ligands may be bound. Again, a distinct anticooperativity is evident in the spread of the two constants.

$$Cu^{2+} + \text{alanine} \quad \log K_i = 8.1, 6.8$$

***Factors affecting the strength of binding of a metal in a complex.*** More basic ligands tend to bind metal ions more tightly just as they do protons. However, the strength of bonding to metal ions to a group is more nearly proportional to the **nucleophilic character** (Chapter 12) which is only partly determined by basicity to protons.

## BOX 6-C LINEAR GIBBS ENERGY RELATIONSHIPS

Organic functional groups exert characteristic electronic effects upon other groups to which they are attached. The quantitative expression of such effects can sometimes be correlated by linear Gibbs energy relationships. The best known of these is the **Hammett equation**, which deals with the transmission of electronic effects across a benzene or other aromatic ring. Consider the acid dissociation constants of three classes of compounds:




The values of p$K_a$ given in the following table have been observed for the parent compounds and for the *meta*-chloro- and the *meta*-nitro-substituted compounds.

**Values of p$K_a$ for Unsubstituted and Substituted Benzoic Acids, Phenylacetic Acids, and Phenols**

| Meta substituent | Benzoic acids* | Phenylacetic acids | Phenols |
|---|---|---|---|
| — H (parent compound) | 4.202 | 4.31 | 9.92 |
| — Cl | 3.837 | 4.13 | 9.02 |
| — $NO_2$ | 3.460 | 3.97 | 8.39 |

* For an independent set of p$K_a$ values for substituted benzoic acids see Bolton and Fleming.[a] The values of σ calculated from them are slightly different from those given here.

Since p$K_a$ is the negative logarithm of a dissociation constant, it follows from Eq. 6-30 that values of p$K_a$ are directly proportional to values of $\Delta G°$ for dissociation of protons. In the Hammett treatment differences in p$K_a$ values, rather than differences in $\Delta G°$, are considered. When a hydrogen atom in the meta position of benzoic acid is replaced by the electron-withdrawing Cl or $NO_2$, p$K_a$ is lowered, i.e., the basicity of the conjugate base of the acid is decreased. The decrease in p$K_a$ amounts to −0.365 for *m*-chlorobenzoic acid and −0.742 for *m*-nitrobenzoic acid. The Hammett treatment asserts that these changes in p$K_a$ are a measure of the electron-withdrawing power of the meta substituent.[b] Thus, the nitro group is about twice as strong as the chloro group in this respect. The numerical values of these changes in the dissociation constant of benzoic acid define the **substituent constants** σ, which are used in the Hammett equation and are given in the following table. In this equation we use the symbol p$K_0$ to represent the p$K_a$ of the unsubstituted parent compound and p$K$ to represent the p$K_a$ of the substituted molecule.

$$\text{For substituted benzoic acids: } pK_0 - pK = \sigma$$

The decreases in the p$K$ of phenylacetic acid occasioned by replacement of the *meta*-hydrogen with Cl or $NO_2$ are −0.18 and −0.34, respectively, substantially less than for the benzoic acids. On the other hand, for the phenols the differences amount to −0.90 and −1.53, much *greater* than for the benzoic acids. The Hammett equation asserts that for reactions such as the dissociation of protons from phenylacetic acids or from phenols, the changes in $\Delta G$ occasioned by meta substitutions are proportional to the σ values, i.e., to the changes in $\Delta G$ for the standard reaction – dissociation of a proton from benzoic acid.[b–d]

**Substituent Constants for Use in the Hammett and Taft Equations**[b-d]

| Substituent | $\sigma_m$ | $\sigma_p$ | $\sigma_p^-$ | $\sigma_p^+$ | σ* (Taft) |
|---|---|---|---|---|---|
| $-O^-$ | −0.71 | −0.52 | | | |
| $-NH_2$ | −0.16 | −0.66 | | −1.11 | 0.62 |
| −OH | 0.121 | −0.37 | | −0.85 | 1.34 |
| $-OCH_3$ | 0.115 | −0.27 | | −0.78 | 1.81 |
| $-CH_3$ | −0.069 | −0.17 | | −0.31 | 0.00 |
| $-NH-COCH_3$ | 0.21 | −0.01 | | −0.25 | |
| −H | 0 | 0 | | 0 | 0.49 |
| $-CH_2OH$ | 0.08 | 0.08 | | | 0.56 |
| $-COO^-$ | −0.10 | 0.00 | | 0.11 | −1.06 |
| $-SO_3^-$ | −0.05 | 0.09 | | 0.12 | |
| −SH | 0.25 | 0.15 | | 0.019 | 1.68 |
| $-CH_2Cl$ | | 0.18 | | | 1.05 |
| $-CONH_2$ | 0.28 | 0.36 | | | |
| −F | 0.337 | 0.06 | 0.02 | −0.07 | |
| −I | 0.352 | 0.18 | | 0.135 | |
| −Cl | 0.373 | 0.227 | | 0.114 | |
| −CHO | 0.36 | 0.22 | 0.37 | | |
| $-COCH_3$ | 0.376 | 0.502 | 0.85 | | |
| −COOH | 0.37 | 0.45 | | 0.42 | 2.08 |
| $-COOCH_3$ | 0.39 | 0.31 | | 0.49 | |
| $-SO_2NH_2$ | 0.55 | 0.62 | | 0.61 | |
| −CN | 0.56 | 0.66 | 0.89 | 0.66 | |
| −C≡CH | | | | | 2.18 |
| $-CF_3$ | 0.43 | 0.54 | | 0.61 | 2.61 |
| $-CCl_3$ | | | | | 2.65 |
| $-NO_2$ | 0.710 | 0.778 | 1.25 | 0.790 | 4.0 |
| $-NH_3^+$ | 1.13 | 1.70 | | | 3.76 |
| N (pyridine) | 0.73 | 0.83 | | | |
| $NH^+$ (pyridine) | 2.18 | 2.42 | | 4.0 | |

**BOX 6-C (continued)**

$$\log (K / K_0) = pK_0 - pK = \rho\sigma$$

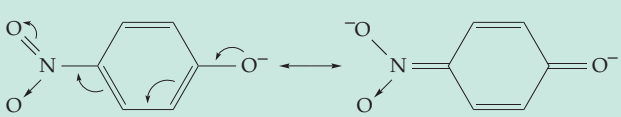
The Hammett equation

The proportionality constant ρ, which also appears in the equation, is a measure of the *sensitivity* of the reaction to the presence of electron-withdrawing or electron-donating substituents in the ring. For benzoic acid, ρ is taken as 1.00. Using the data from the accompanying table together with many other data, an average value of ρ = 0.49 has been found for phenylacetic acids. Likewise, ρ = 2.23 for phenols, and ρ = 5.7 for dissociation of protons from substituted pyridinium ions. The sensitivity to substituent changes is highest (highest ρ) in the latter case where proton dissociation is directly from an atom in the ring and is lowest when the basic center is removed farthest from the ring (phenylacetic acid).

Through knowledge of the value of ρ for a given reaction, it is possible to predict the effect of a substituent on $pK_a$ using the tabulated values of σ. In many cases the effects are additive for multisubstituted compounds.

$$pK_a = pK_0 - \rho\Sigma\sigma$$

Since substituents in *o*, *m*, and *p* positions have quantitatively different influences, different substituent constants σ are defined for each position. Moreover, since special complications arise from ortho interactions, it is customary to tabulate σ values only for meta and para positions. These are designated $\sigma_m$ and $\sigma_p$. Apparent σ values for *ortho* substituents are also available.[d] An additional complication is that certain reactions are unusually sensitive to para substituents that are able to interact by resonance directly across the ring. An example is the acid dissociation of phenols. While $\sigma_p$ for the nitro group is ordinarily 0.778, a correct prediction of the effect of the *p*-nitro group on dissociation of phenol is given only if $\sigma_p$ is taken as 1.25. This higher σ value is designated $\sigma_p^-$. The resonance in the phenolate anion giving rise to this enhanced effect of the nitro group may be indicated as follows:



(Here the curved arrows represent the direction of flow of electrons needed to convert from the one resonance structure to the other.) For similar reasons, some reactions require the use of special $\sigma_p^+$ constants for strongly electron-donating substituents such as OH which are able to interact across the ring by resonance. Thus $\sigma_p$ for the OH group is −0.37, while $\sigma_p^+$ is −0.85.[e] For the methoxyl group (—$OCH_3$) $\sigma_p = -0.27$, $\sigma_p^+ = -0.78$, and $\sigma_m = +0.12$.

The use of different kinds of substituent constants complicates the application of the Hammett equation and over 20 different sets of σ values have been proposed. A simplification is the representation of substituent constants as linear combinations of two terms, one representing "field" or "inductive" effects and the other resonance effects.[e,f]

Many other linear Gibbs energy relationships have been proposed; for example, the acid strengths of aliphatic compounds can be correlated using the "Taft polar substituent constants" σ*.

$$\log (K/K_0) = \rho^* \sigma^*$$

For example, the following give good approximations of $pK_a$ values.[d]

for R—COOH $pK_a = 4.66 - 1.62\,\sigma^*$
for R—$CH_2$—COOH $pK_a = 5.16 - 0.73\,\sigma^*$

While the examples chosen here concern only dissociation of protons, the Hammett equation has a much broader application. Equilibria for other types of reactions can be treated. Furthermore, since rates of reactions are related to Gibbs energies of activation, many rate constants can be correlated. For these purposes the Hammett equation can be written in the more general form in which *k* may be either an equilibrium constant or a rate constant.[b] The subscript *j* denotes the reaction under consideration and *i* the substituent influencing the reaction.

$$\log k_{ij} - \log k_{0j} = \rho_i \sigma_j$$

An example of a linear Gibbs energy relationship that is widely used in discussing mechanisms of enzymatic reactions is the Brönsted plot (Eqs. 9-90 and 9-91).

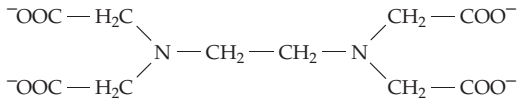

[a] Bolton, P. D., Fleming, K. A., and Hall, F. M. (1972) *J. Am. Chem. Soc.* **94**, 1033–1034
[b] Hammett, L. P. (1970) *Physical Organic Chemistry*, 2nd ed., p. 356, McGraw-Hill, New York
[c] Wells, P. R. (1963) *Chem. Rev.* **63**, 171–219
[d] Barlin, G. B., and Perrin, D. D. (1966) *Q. Rev., Chem. Soc.* **20**, 75–101
[e] Swain, C. G., and Lupton, E. C., Jr. (1968) *J. Am. Chem. Soc.* **90**, 4328–4337
[f] Hansen, L. D., and Hepler, L. G. (1972) *Can. J. Chem.* **50**, 1030–1035

**TABLE 6-9**
**Logarithms of Binding Constants for Some 1:1 Metal Complexes at 25°C[a]**

| Ligand | $H^+$ | $Mg^{2+}$ | $Ca^{2+}$ | $Mn^{2+}$ | $Cu^{2+}$ | $Zn^{2+}$ |
|---|---|---|---|---|---|---|
| Hydroxide, $OH^-$ | 14.0 | 2.5 | 1.4 | | 6.5 | 4.4 |
| Acetate$^-$ | 4.7 | ~0.65 | 0.5 | ~1.0 | 2.0 | 1.5 |
| Lactate$^-$ | 3.8 | ~1.0 | ~1.2 | 1.3 | 3.0 | 2.2 |
| Succinate$^{2-}$ | 5.2 | 1.2 | 1.2 | | 3.3 | 2 |
| $NH_3$ | 9.3 | ~0 | –0.2 | 0.8 | 4.0 | 2.4 |
| Ethylenediamine | 10.2 | 0.4 | | 2.8 | 10.8 | 6.0 |
| Glycine$^-$ | 9.6 | 2.2 | 1.4* | 2.8 | 8.2 | 5.0 |
| Glycine amide | 8.1 | | | ~1.5 | 5.3 | 3.3 |
| Alanine$^-$ | 9.7 | 2.0* | 1.2* | 3.0* | 8.1 | 4.6 |
| Aspartate$^{2-}$ | 9.6 | 2.4 | 1.6 | | 8.6 | 5.8 |
| Glycylglycine$^-$ | 8.1 | | 1.2* | 2.2* | 5.5 | 3.4 |
| Pyridine | 5.2 | | | 0.1 | 2.5 | 1 |
| Imidazole | 7.5 | | | 1.6 | 4.6 | 2.6 |
| Histidine | 9.1 | | | 3.3 | 10.2 | 6.6 |
| Adenine | 9.8 | | | | 8.9 | 6.4 |
| Citrate$^{3-}$ | 5.6 | 3.2 | 4.8 | 3.5 | ~4 | 4.7 |
| EDTA$^{4-}$ [b] | 10.2 | 8.8 | 10.6 | 13.8 | 18.7 | 16.4 |
| EGTA$^{4-}$ [c] | 9.4 | 5.3 | 10.9 | 12.2 | 17.6 | 12.6 |
| ATP$^{4-}$ | 6.5 | 4.2 | 3.8 | 4.8 | 6.1 | 4.9 |

[a] All values are for log $K_1$. Included is the highest p$K_a$ for protons. Data for amino acids are from Martell, A.E. and Smith, R.M. (1974, 1975) *Critical Stability Constants*, 3 vols., Plenum, New York. Others are from Sillén, L.G. and Martell, A.E. (1964) *Stability Constants of Metal–Ion Complexes*, Spec. Publ. No. 17, Chemical Society, London. Most constants for amino acids are for ionic strength 0.1. Some (designated by asterisks) are for zero ionic strength. The values shown for other ligands have been selected from a large number reported without examination of the original literature.

[b] Ethylenediaminetetraacetic acid (EDTA), a chelating agent widely used by biochemists for

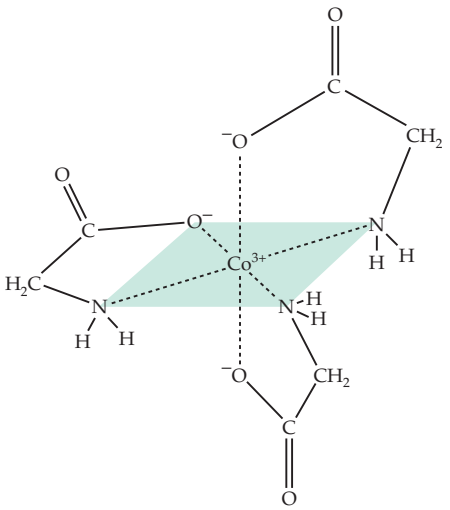


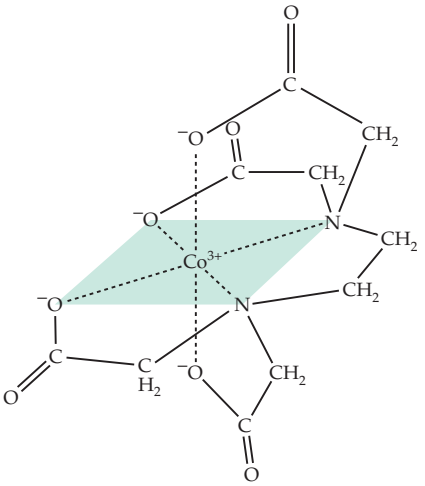
preventing unwanted reactions of metal ions. The high formation constants ensure that most metal ions remain bound to the EDTA.

[c] EGTA is similar to EDTA but has the group $-CH_2-CH_2-O-CH_2-CH_2-O-CH_2-CH_2-$ joining the two nitrogen atoms in place of $-CH_2-CH_2-$ of EDTA. Note that EGTA has a higher selectivity for $Ca^{2+}$ compared to $Mg^{2+}$ than does EDTA.

The pH of the medium always has a strong effect on metal binding. Competition with protons means that metal complexes tend to be of weak stability at low pH. Anions of carboxylic acids are completely protonated below a pH of ~4 and a metal can combine only by displacing a proton. However, at pH 7 or higher, there is no competition from protons. On the other hand, in the case of ethylenediamine, whose p$K_a$ values are 10.2 and 7.5 (Table 6-9), protons are very strong competitors at pH 7, even with a strongly complexing metal ion such as $Cu^{2+}$. At high pH there may be competition between the ligand and hydroxyl ion. At pH 7 about one-half of $Cu^{2+}$ dissolved in water is complexed as $CuOH^+$. Aluminum forms soluble complexes $Al(OH)^{2+}$, $Al(OH)_2^+$, and $Al(OH)_4^+$ (Box 12-F).

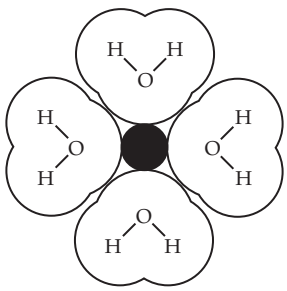
One of the most important factors in determining the affinity of organic molecules for metal ions is the **chelate effect**. The term chelate (pronounced "keel-ate") is from a Greek word meaning crab's claw. It refers to the greatly enhanced binding of metal ions resulting from the presence of two or more complexing groups in the same organic molecule. The chelate effect has been exploited by nature in the design of many important metal-binding molecules, including porphyrins (Fig. 16-5), chlorophyll (Fig. 23-20), the siderophores (Fig. 16-1), and metal-binding proteins. Structures of two **chelate complexes** are shown here. Notice from Table 6-9 that many simple compounds, such as the α-amino acids and citric acid, often form strong chelate complexes with metal ions.





***How properties of the metal ion affect chelation.*** The **charge**, the **ionic radius** (Table 6-10), the **degree of hydration**, and the **geometry of orbitals** used in covalent bonding between metal and chelating groups all affect the formation constants of a complex. Multicharged ions form stronger complexes than do monovalent ions, which have a lower charge density.

Among ions of a given charge type (e.g., $Na^+$ vs $K^+$; $Mg^{2+}$ vs $Ca^{2+}$), the smaller ions are more strongly hydrated than are the larger ions in which the charge is dispersed over a greater surface area. Most cations, except for the large ones, have a primary hydration sphere containing about six molecules of water. Four molecules of water can be placed around the ion in one plane as shown in the following drawing of water molecules coordinated to $Mg^{2+}$.



One additional water molecule can be bound above and another below to provide six molecules in an array of **octahedral** geometry. An alternative tetrahedral arrangement of four molecules of water around the ion has been suggested for $Li^+$ and $Na^+$.[90] In either case additional solvent molecules are held in a looser secondary sphere. For instance, electrochemical transference experiments indicate a total of ~16 molecules of water around $Na^+$ and ~10 around $K^+$.

To form a chelate complex a metal ion must usually lose most of its hydration sphere. For this reason, the larger, less hydrated metal ions often bind more strongly than do the smaller, more hydrated ones. For example, $Ca^{2+}$ binds to EDTA more tightly than does $Mg^{2+}$ (Table 6-9). However, the reverse order may be observed with negatively charged ligands such as $OH^-$ in which the charge in highly concentrated. The same is true for $ATP^{4-}$, which binds $Mg^{2+}$ more tightly than $Ca^{2+}$ (Table 6-9; Section B,5).

Differences in both the charge density and the hydration of ions determine the **Hofmeister series** (lyotropic series).[91]

$$Ba^{2+} > Ca^{2+} > Mg^{2+} > Li^+ > Na^+ > K^+ > Cs^+ > NH_4^+$$

This was originally defined as the order of effectiveness in precipitating colloids or protein molecules. The ions to the left are less hydrated than those to the right. A similar series can be defined for anions. The following, a well-known sequence of the stabilities of complexes of metals of the first transition series, applies to many different types of complexes including those of the α-amino acids, as is shown in Fig. 6-11.

$$Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+}$$

Simple electrostatic theory based upon differences in the ionization potentials or electronegativity of the ions would predict a gradual monotonic increase in chelate stability from manganese to zinc. In fact, with nitrogen-containing ligands $Cu^{2+}$ usually forms by far the strongest complexes. Cobalt, nickel, and iron ions also show an enhanced tendency to bind to nitrogen-containing ligands. The explanation is thought to lie in the ability of the transition metals to supply *d* orbitals which can participate in covalent bond formation by accepting electrons from the ligands. Iron, copper, and cobalt are often located in the centers of nitrogen-containing structures such as the heme of our blood (iron, Fig. 16-5) and vitamin $B_{12}$ (Box 16-B).

***Metal binding sites in cells.*** Functional groups that participate in metal binding include negatively charged carboxylate $-COO^-$, thiolate $-S^-$, phenolate $-O^-$, and phosphate$^-$ as well as uncharged amino, imidazole, $-OH$, and the polarizable carbonyl groups of peptide and amide side chains. To which of these ligands will specific ions tend to bind? The alkali metal ions $Na^+$ and $K^+$ are mostly free within cells

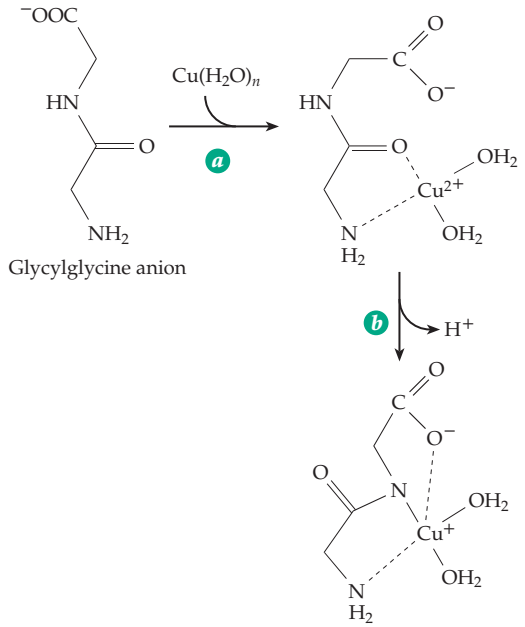
**TABLE 6-10**
**Ionic Radii in Nanometers for Some Metallic and Nonmetallic Ions[a]**

| | | | | | |
|---|---|---|---|---|---|
| | | $Mn^{2+}$ | 0.080 | | |
| $Li^+$ | 0.060 | $Fe^{2+}$ | 0.076 | $H^-$ | 0.21 |
| $Na^+$ | 0.095 | $Co^{2+}$ | 0.074 | $F^-$ | 0.136 |
| $K^+$ | 0.133 | $Ni^{2+}$ | 0.069 | $Br^-$ | 0.195 |
| $Rb^+$ | 0.148 | $Cu^{2+}$ | 0.072[b] | $I^-$ | 0.216 |
| | | $Zn^{2+}$ | 0.074 | | |
| | | $Cd^{2+}$ | 0.097 | | |
| $Be^{2+}$ | 0.031 | | | | |
| $Mg^{2+}$ | 0.065 | $Al^{3+}$ | 0.050 | | |
| $Ca^{2+}$ | 0.099 | $Fe^{3+}$ | 0.064 | | |
| $Sr^{2+}$ | 0.113 | $Mo^{4+}$ | 0.070 | | |
| $Ba^{2+}$ | 0.135 | $Mo^{6+}$ | 0.062 | | |

[a] Radii are calculated according to the method of Pauling and are taken from Cotton, F.A. and Wilkinson, G. (1972) *Advanced Inorganic Chemistry*, 3rd ed. Wiley (Interscience), New York .

[b] From Ahrens, L. H. as given by Sienko, M. J. and Plane, R. A. (1963) *Physical Inorganic Chemistry*, pp. 68–69. Benjamin, New York.

(Box 5-A) but are in part bound to defined sites in proteins. Both $Ca^{2+}$ and $Mg^{2+}$ tend to remain partially free and complexed with the numerous phosphate and carboxylate ions present in cells. However, they may find very specific tight binding sites such as that of $Mg^{2+}$ in chlorophyll (Fig. 23-20). The heavier metal ions, including those of zinc, copper, iron, and other transition metals, usually bind to nitrogen or sulfur atoms.[92] For example, small peptides react with $Cu^{2+}$ to form chelate complexes in which the peptide carbonyl oxygen binds to the metal (Eq. 6-85, step *a*).[93,94] By losing a proton the peptide NH can sometimes also function as a metal ligand (Eq. 6-85, step *b*).




(6-85)

In many **metalloproteins** the metals are found in **prosthetic groups** such as the porphyrin of heme proteins and the molybdopterin of molybdenum containing enzymes (Fig. 16-31). Very often clusters of carboxylate, imidazole, and other groups are used to create binding sites. In carboxypeptidase A (Fig. 12-16) two imidazole groups and one carboxylate of a glutamate site chain hold $Zn^{2+}$. In carbonic anhydrases, three imidazoles hold a zinc ion (Fig. 13-1), while in one kind of superoxide dismutase both $Zn^{2+}$ and $Cu^{2+}$ are bound in adjacent locations with six imidazoles and one carboxylate group participating in the binding. In contrast, $Zn^{2+}$ and $Cd^{2+}$ are bound in metallothioneins by clusters of thiol groups from cysteine side chains (Box 6-E). While $Fe^{2+}$ is bound by from four to six nitrogen atoms in heme proteins, it is also found attached to oxygen atoms of tyrosine side chains, as is shown for a transferrin in Fig. 16-2. One imidazole, one carboxylate group, and a bicarbonate ion also bind to the iron. This site is also quite satisfactory for $Al^{3+}$ (Box 12-F), which tends to occupy a fraction of the transferrin sites in blood, although it binds much more weakly than does Fe. Several other binding sites for transition metals are pictured in Chapter 16.

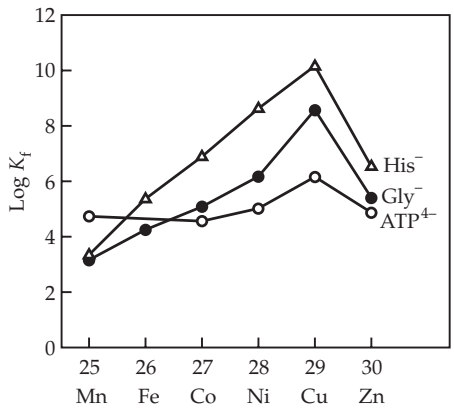



**Figure 6-6** Logarithms of formation constants of metal complexes of histidine, glycine, and ATP plotted against the atomic number of elements from manganese to zinc.

***Calcium-binding proteins.*** Much information about the functions of transition metal ions is given in Chapter 16. Here and in Box 6-D we consider calcium ions which, because of their broad distribution and range of functions, will be discussed in virtually every chapter of this book. The concentration of $Ca^{2+}$ varies greatly in different parts of a cell and also with time (Box 6-D). Many **calcium-binding proteins** participate in mediating the physiological effects of these changes and also in buffering the calcium ion concentration. These include a large family of **helix–loop–helix** or **EF-hand proteins**.[95–97] The first of these to be discovered was **parvalbumin**, a 108-residue protein from carp muscle.[96,98-99a] The structure of the pair of metal-binding sites of the protein is shown in Fig. 6-7.

Each consists of two helices that are almost perpendicular and connected by a loop that forms the $Ca^{2+}$-binding site. In the site at the left side in Fig. 6-7A the $Ca^{2+}$ is bound by four carboxylate groups from aspartate and glutamate side chains, a hydroxyl group of serine, and the residue 57 carbonyl oxygen of the peptide backbone.[99b] The same peptide group is hydrogen bonded to a carbonyl group of another segment of peptide chain near the second $Ca^{2+}$ site (to the right in Fig. 6-7A). This site contains four carboxylate ions, one of which coordinates the $Ca^{2+}$ with both oxygen atoms, and another peptide carbonyl group. By attaching itself to several different side chain groups, the metal ion can induce a substantial change in conformation from that present in the calcium-free protein.[97]

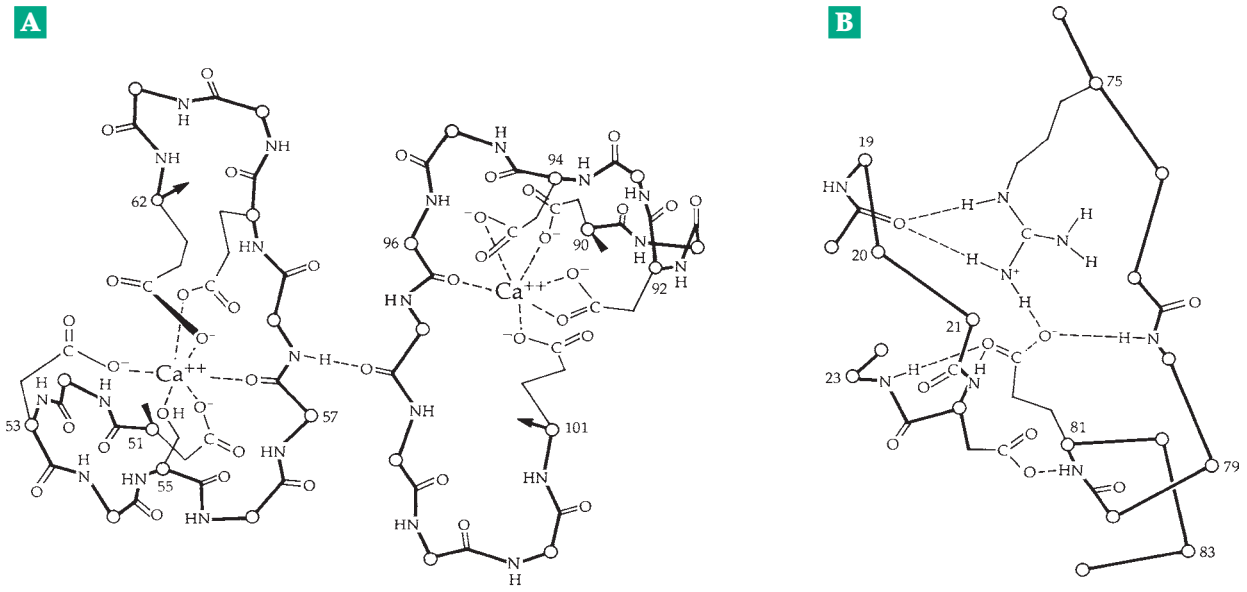



**Figure 6-7** (A) Part of the 108-residue peptide chain of the calcium-binding protein of carp muscle. The two calcium-binding loops are shown together with a hydrogen bond between them. (B) A view of the intricate network of hydrogen bonds linking two segments of the peptide chain in the interior of the molecule. Note especially the bonding of the guanidinium group from arginine-75 to the carboxylate of glutamic acid 81 and to the peptide carbonyl of residue 18. Note that the carboxylate group also interacts with two different peptide NH groups. From Kretsinger and Nockolds.[98]

Kretzinger, who discovered the structure of parvalbumin, named the $Ca^{2+}$-binding helix–loop–helix motif an **EF-hand** because it is formed by helices E and F and resembles a hand with pointer finger extended along the E helix and the thumb in the direction of the F helix, the flexed middle finger forming the $Ca^{2+}$-binding loop. Helices C and D also form a hand. The resulting pair of hands can be visualized better at the top of the calmodulin structure in Fig. 4-8. The EF-hand motif has been identified in over 1000 proteins.[100]

Parvalbumins, which are also found in other vertebrates, are high-affinity $Ca^{2+}$-buffers.[99] Additional calcium buffers with EF-hand structures are the vitamin D-induced **calbindins**. One 9-kDa calbindin is found in mammalian intestinal tissue and in skin. It has two helix–loop–helix $Ca^{2+}$-binding sites of differing affinity[101,102] that presumably function in the absorption of calcium. A 28-kDa vitamin D-dependent protein from chicken intestine contains six similar $Ca^{2+}$-binding loops.[97,103]

Another group of calcium-buffering and storage proteins with remarkable $Ca^{2+}$-binding properties are the 40- to 45-kDa **calsequestrins**, which are found in the lumen of the ER (sarcoplasmic reticulum) of skeletal muscle. Calsequestrins are not typical EF-hand proteins but have a high content of glutamate and aspartate. They bind ~50 $Ca^{2+}$ per molecule of protein with $K_d$ ~ 1 mM.[104,105] Similar proteins called **calreticulins** are found in most non-muscle cells.[106,107]

A very large number of EF-hand proteins have signaling functions. The best known of these is the 148-residue **calmodulin**, which regulates many enzymes and cellular processes (Box 6-D; also Chapter 11).[108,109] The protein, which is present in all eukaryotes, has a conserved sequence that forms two pairs of helix–loop–helix $Ca^{2+}$-binding sites that are separated by a long helix (Fig. 6-8).[108,110,111] Two of the sites bind $Ca^{2+}$ tightly and cooperatively,[110,112] with $K_d$ values in the micromolar range. Calmodulin from almost all species contains the modified amino acid ε-*N*-trimethyllysine at position 115. However, octopus calmodulin has ordinary lysine at this position and seems to function well.[108] Calmodulin's controlling functions result from $Ca^{2+}$-induced conformational changes that modify its affinity for other proteins whose activity may be increased or decreased by bound calmodulin.[109,113] A protein with a similar dumbell shape and structure is **troponin C** of skeletal muscles.[114,115] Troponin C binds to a complex of proteins that assemble on the thin actin filaments of muscle fibers and control contraction in response to changes in the calcium ion concentration (Chapter 19).[116] Other proteins that contain EF-hand motifs and are therefore responsive to $Ca^{2+}$ include **spectrin** of cell membranes,[117] **clathrin light chains** from coated vesicles,[118,119] the extracellular **osteonectin** of bones and teeth,[120] and a birch pollen antigen.[121] Another group of 17 or more small **S100 EF-hand** proteins play a variety of other roles.[122–123a] One of these, which has a high affinity for $Zn^{2+}$, has been named **psoriasin** because of its 5-fold or greater

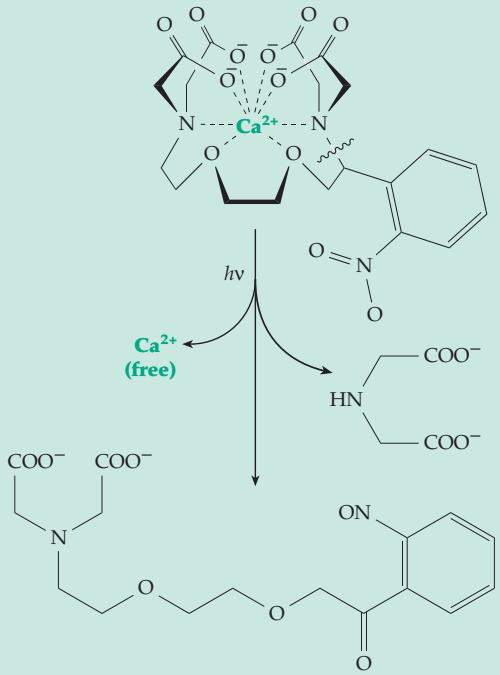
## BOX 6-D CALCIUM

The essentiality of calcium ions to living things was recognized in the last century by S. Ringer, who showed that ~1 μM $Ca^{2+}$ was needed to maintain the beat in a perfused frog heart. Later, calcium was shown essential for repair of ruptures in the cell membrane of the protozoan *Stentor* and for the motion of amebas. The animals quickly died in its absence. The role in the frog heart was traced to transmission of the nerve impulse from the nerve to the heart muscle. More recently it has been recognized that $Ca^{2+}$ is required for blood clotting and is involved in triggering many responses by cells. Calcium ions are also an integral part of many enzymes and have structural roles in proteins, carbohydrate gels, and biological membranes.

Like $Na^+$, the calcium ion is actively excluded from cells. Indeed, 99% of the calcium in the human body is present in the bones.[a–d] The blood serum concentration of $Ca^{2+}$ is ~3 mM, of which ~1.5 mM is free. The rest is chelated by proteins, carbohydrates, and other materials. Within cells the concentration of free $Ca^{2+}$ is < 1 μM and typically ~0.05–0.2 μM for unexcited cells.[d–f] However, the total intracellular $Ca^{2+}$ is considerably higher and may be in excess of 1 mM. Approximate total concentrations are: red blood cells, 20 μM; liver, 1.6 mM; and heart, 4 mM. A gradient in $[Ca^{2+}]$ of $10^3$ or more is maintained across membranes by the calcium ion pump (Chapter 8). The action of this pump is counteracted by a very slow diffusion of the external $Ca^{2+}$ back through the membrane via an $Na^+$–$Ca^{2+}$ exchange into the cells.[g]

Free $Ca^{2+}$ lacks spectroscopic properties suitable for its direct measurement at the low concentrations present in cells. However, it can be measured indirectly by the use of chelating agents that are relatively specific for $Ca^{2+}$ and which have a measurable property that changes upon calcium-binding. The fluorescent photoprotein **aequorin** (Chapter 23), which may be injected into cells or synthesized within cells from transferred genes is often employed.[h,i] Various synthetic $Ca^{2+}$ indicators also fluoresce brilliantly upon chelation.[j–n] Others contain fluorine or a suitably placed atom of $^{13}C$ which changes its NMR chemical shift upon chelation with $Ca^{2+}$.[o] These compounds may be carried into cells in the form of esters which pass through membranes but are then hydrolyzed leaving the indicators trapped in the cytoplasm.[j] Chelate compounds that bind $Ca^{2+}$ within cells and release it rapidly upon irradiation with ultraviolet light have also been developed.[p,q]

Consistent with their role in signaling, calcium ions are unevenly distributed within cells. Mitochondria, endoplasmic reticulum, Golgi, and nuclei may all take up calcium ions. Cytoplasmic $Ca^{2+}$




Calcium chelate compound that releases free calcium ion within cells upon irradiation with ultraviolet light.[p,q]

may sometimes be sequestered in microvesicles (**calciosomes**) or in intracellular granules[c,r,s] in which the $[Ca^{2+}]$ is ~0.5–1 mM but may reach 5–10 mM.[f] Many regulatory mechanisms exist. For example, **calcitonin** and **parathyroid hormone** interact with **vitamin D** and its metabolites in the small intestine, bones, and kidneys to control the deposition of calcium in bones, a topic considered in Box 22-C.

A characteristic function of $Ca^{2+}$ in living things is **activation** of various metabolic processes. This occurs when a sudden change in permeability of the plasma membrane or in the membranes of the endoplasmic reticulum (ER) allows $Ca^{2+}$ to diffuse into the cytoplasm. For example, during the contraction of muscle, the $Ca^{2+}$ concentration rises from ~0.1 to ~10 μM as a result of release from storage in the calciosomes of the ER. The calcium ions bind to **troponin C** initiating muscle contraction (Chapter 19).[t] The ER membranes of muscle are rich in a $Ca^{2+}$ pump protein[u,v] (Fig. 8-26), and in a series of calcium-binding proteins such as **calsequestrin** (see text).[f] Their combined action soon restores the $[Ca^{2+}]$ to the original low value.

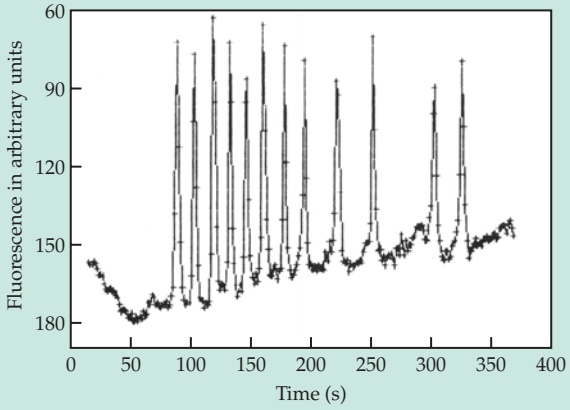
Skeletal muscle is activated by nerve impulses which induce $Ca^{2+}$ release through the action of a **voltage sensor**, a protein also known as the **dihydropyridine receptor**,[w,x] together with a **calcium release channel** known as the **ryanodine**

## BOX 6-D (continued)

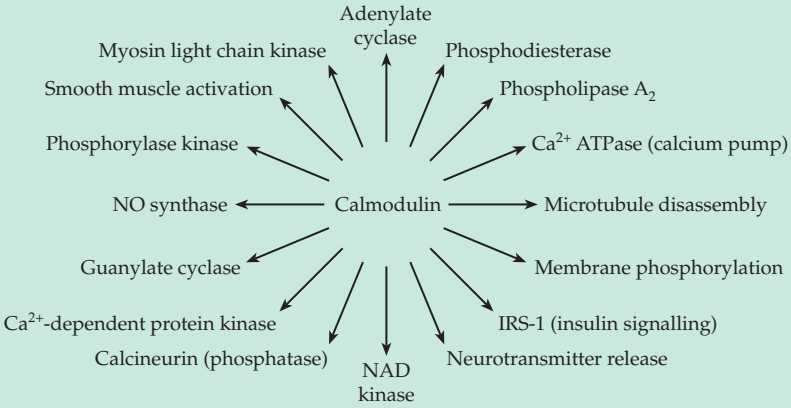
**receptor**.[w,y–aa] Inositol trisphosphate (Fig. 11-9), cyclic ADP-ribose, as well as pH changes[bb] are involved in controlling these channels,[cc,dd] a topic discussed in Chapter 11, E. The release of $Ca^{2+}$ can be visualized using high-speed digital imaging microscopy and fluorescent [$Ca^{2+}$] indicators.[ee,ff] Release of $Ca^{2+}$ stored in sea urchin eggs is induced by cyclic ADP-ribose and by nicotinic acid adenine dinucleotide phosphate ($NAADP^+$, Eq. 15-16).[dd,gg]

In a similar manner, when a nerve impulse reaches a neuron ending (synapse) calcium ions are released into the cytoplasm and provide the trigger that causes stored neurotransmitters to be dumped into the narrow synaptic cleft that separates the endings of two communicating neurons.[y,hh] The released neurotransmitter initiates an impulse in the "postsynaptic neuron" usually again with an inflow of $Ca^{2+}$ (see Chapter 30).[e] Hormones and various other compounds often stimulate the flow of calcium ions into cells.[ii] There $Ca^{2+}$ regulates enzymes[jj,kk] (Chapter 11), microtubules, clathrin of coated vesicles (Chapter 8), $K^+$ channels in nerve membranes, and events within mitochondria,[ll,mm] in the ER,[op] and in the nucleus.[nn,oo] A substantial fraction of these responses are mediated by **calcium-binding regulatory proteins**. Among the most prominent of these is **calmodulin** (Fig. 6-8), which, upon binding of $Ca^{2+}$ activates a host of metabolic processes[pp] as indicated in the following scheme, which is modified from that of Cheung.[qq]

In many cases metabolic control by $Ca^{2+}$ is modulated by phosphorylation and dephosphorylation or other covalent modification of proteins. Such modification may alter the affinity for $Ca^{2+}$, allowing the latter to either bind and induce a conformational change or remain unbound,[rr] without a change in [$Ca^{2+}$]. Nevertheless, the free $Ca^{2+}$ concentration within cells can change greatly and very rapidly. For example, the oscillatory change in intracellular [$Ca^{2+}$] shown above was observed in pancreatic insulin-secreting β cells responding to stimulation by the agonist carbamoylcholine. The free [$Ca^{2+}$] was evaluated from fluorescence measurements using the $Ca^{2+}$ indicator dye fura 2 (From Prentki *et al*[ss]). Oscillations in [$Ca^{2+}$] have been observed under many circumstances.[ff,oo,tt–ww] These are of particular interest because of the possible relationship to the initiation of oscillatory nerve conduction (Chapter 30).[uu] Released $Ca^{2+}$ often appears to move across cells in waves and sometimes to be released as "puffs" or "sparks."[ee,ff,oo]




Calcium oscillations observed with six cultured pancreatic β cells after a single infusion of 0.2 mM carbamoylcholine. The fluorescence intensity of the $Ca^{2+}$ indicator dye fura 2, with excitation at 380 nm, was recorded versus time. From Pretki *et al.*[ss]





[a] Bianchi, C. P. (1968) *Cell Ca*$^{++}$, Appleton, New York
[b] Means, A. R., ed. (1994) *Calcium Regulation of Cellular Function*, Lippincott-Raven, Hagerstown, Maryland
[c] Volpe, P., Krause, K.-H., Hashimoto, S., Zorzato, F., Pozzan, T., Meldolesi, J., and Lew, D. P. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1091–1095
[d] Carafoli, E. (1987) *Ann. Rev. Biochem.* **56**, 395–433
[e] Ghosh, A., and Greenberg, M. E. (1995) *Science* **268**, 239–247
[f] Meldolesi, J., and Pozzan, T. (1998) *Trends Biochem. Sci.* **23**, 10–14
[g] Khananshvili, D. (1991) *J. Biol. Chem.* **266**, 13764–13769


(continued)

**BOX 6-D (continued)**


[h] Rizzuto, R., Simpson, A. W. M., Brini, M., and Pozzan, T. (1992) *Nature (London)* **358**, 325–327
[i] Knight, M. R., Campbell, A. K., Smith, S. M., and Trewavas, A. J. (1991) *Nature (London)* **352**, 524–526
[j] Tsien, R. Y., Pozzan, T., and Rink, T. J. (1984) *Trends Biochem. Sci.* **9**, 263–266
[k] Davies, P. L., and Hew, C. L. (1990) *FASEB J.* **4**, 2460–2468
[l] Ganz, M. B., Rasmussen, J., Bollag, W. B., and Rasmussen, H. (1990) *FASEB J.* **4**, 1638–1644
[m] Lloyd, Q. P., Kuhn, M. A., and Gay, C. V. (1995) *J. Biol. Chem.* **270**, 22445–22451
[n] Etter, E., Minta, A., Poenie, M., and Fay, F. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5368–5373
[o] Robitaille, P.-M. L., and Jiang, Z. (1992) *Biochemistry* **31**, 12585–12591
[p] Ellis-Davies, G. C. R., and Kaplan, J. H. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 187–191
[q] Ellis-Davies, G. C. R., Kaplan, J. H., and Barsotti, R. J. (1996) *Biophys. J.* **70**, 1006–1016
[r] Burgoyne, R. D., and Cheek, T. R. (1991) *Trends Biochem. Sci.* **16**, 319–320
[s] Golovina, V. A., and Blaustein, M. P. (1997) *Science* **275**, 1643–1648
[t] Gagné, S. M., Li, M. X., and Sykes, B. D. (1997) *Biochemistry* **36**, 4386–4392
[u] MacLennan, D. H., Rice, W. J., and Green, N. M. (1997) *J. Biol. Chem.* **272**, 28815–28818
[v] Yonekura, K., Stokes, D. L., Sasabe, H., and Toyoshima, C. (1997) *Biophys. J.* **72**, 997–1005
[w] Yano, M., El-Hayek, R., and Ikemoto, N. (1995) *J. Biol. Chem.* **270**, 3017–3021
[x] Mitterdorfer, J., Sinnegger, M. J., Grabner, M., Striessnig, J., and Glossmann, H. (1995) *Biochemistry* **34**, 9350–9355
[y] McPherson, P. S., and Campbell, K. P. (1993) *J. Biol. Chem.* **268**, 13765–13768
[z] Sakube, Y., Ando, H., and Kagawa, H. (1997) *J. Mol. Biol.* **267**, 849–864
[aa] Zorzato, F., Fujii, J., Otsu, K., Phillips, M., Green, N. M., Lai, F. A., Meissner, G., and MacLennan, D. H. (1990) *J. Biol. Chem.* **265**, 2244–2256
[bb] Donoso, P., Beltrán, M., and Hidalgo, C. (1996) *Biochemistry* **35**, 13419–13425
[cc] Lee, M. G., Xu, X., Zeng, W., Diaz, J., Wojcikiewicz, R. J. H., Kuo, T. H., Wuytack, F., Racymaekers, L., and Muallem, S. (1997) *J. Biol. Chem.* **272**, 15765–15770
[dd] Graeff, R. M., Franco, L., De Flora, A., and Lee, H. C. (1998) *J. Biol. Chem.* **273**, 118–125
[ee] Isenberg, G., Etter, E. F., Wendt-Gallitelli, M.-F., Schiefer, A., Carrington, W. A., Tuft, R. A., and Fay, F. S. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5413–5418
[ff] Horne, J. H., and Meyer, T. (1997) *Science* **276**, 1690–1693
[gg] Lee, H. C., and Aarhus, R. (1997) *J. Biol. Chem.* **272**, 20378–20383
[hh] Miller, R. J. (1992) *J. Biol. Chem.* **267**, 1403–1406
[ii] Quitterer, U., Schröder, C., Müller-Esterl, W., and Rehm, H. (1995) *J. Biol. Chem.* **270**, 1992–1999
[jj] Suzuki, K. (1987) *Trends Biochem. Sci.* **12**, 103–105
[kk] Martin, B. L., and Graves, D. J. (1986) *J. Biol. Chem.* **261**, 14545–14550
[ll] Moudy, A. M., Handran, S. D., Goldberg, M. P., Ruffin, N., Karl, I., Kranz-Eble, P., DeVivo, D. C., and Rothman, S. M. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 729–733
[mm] Kasumi, T., Tsumuraya, Y., Brewer, C. F., Kersters-Hilderson, H., Claeyssens, M., and Hehre, E. J. (1987) *Biochemistry* **26**, 3010–3016
[nn] Santella, L., and Carafoli, E. (1997) *FASEB J.* **11**, 1091–1109
[oo] Lipp, P., Thomas, D., Berridge, M. J., and Bootman, M. D. (1997) *EMBO J.* **16**, 7166–7173
[op] Corbett, E. F., and Michalak, M. (2000) *Trends Biochem. Sci.* **25**, 307–311
[pp] Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B., and Bax, A. (1992) *Science* **256**, 632–638
[qq] Cheung, W. Y. (1980) *Science* **207**, 19–27
[rr] Geisow, M. J. (1987) *Trends Biochem. Sci.* **12**, 120–121
[ss] Prentki, M., Glennon, M. C., Thomas, A. P., Morris, R. L., Matschinsky, F. M., and Corkey, B. E. (1988) *J. Biol. Chem.* **263**, 11044–11047
[tt] Tang, Y., and Othmer, H. G. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7869–7873
[uu] Putney, J. W., Jr. (1998) *Science* **279**, 191–192
[vv] Berridge, M. J. (1997) *Nature (London)* **386**, 759–760
[ww] Hajnóczky, G., and Thomas, A. P. (1997) *EMBO J.* **16**, 3533–3543


increase above the normal level in keratinocytes of patients with the skin disease psoriasis.[123b]

Calcium ions are often involved in holding negatively charged groups together, for example, in the binding of proteins to phospholipid membranes. Among these membrane-associated proteins are the vitamin K-dependent proteins, all of which contain several residues of the chelating amino acid γ-carboxyglutamate (Chapter 15) at their calcium-binding sites.[124,125] Many of these are involved in the clotting of blood (Chapter 12). Some of the membrane-binding proteins called **annexins** (Chapter 8) are also $Ca^{2+}$-dependent ion channels.[126,127] Other $Ca^{2+}$-requiring lipid-binding proteins include the lipocortins, calpactins, and calelectrins.[128] Cadherins bind $Ca^{2+}$ and help provide cohesion between cells.[129] Many proteins contain bound $Ca^{2+}$ in precisely defined sites where it plays a structural role. These include the galactose-binding protein of bacterial transport and chemotaxis (Fig. 4-18)[130] and α-lactalbumin of milk. Although α-lactalbumin does not contain the helix–loop–helix pattern, its $Ca^{2+}$-binding site does consist of three carboxylate groups and two backbone carbonyl oxygen atoms.[131] The α-amylases, thermolysin, and staphylococcal nuclease (Chapter 12), and the lectin favin (Fig. 2-15) all contain bound $Ca^{2+}$. Calcium ions also bind to anionic groups in carbohydrates, e.g., to the sulfate groups in carageenin gels (Chapter 4) where they provide structural stability.

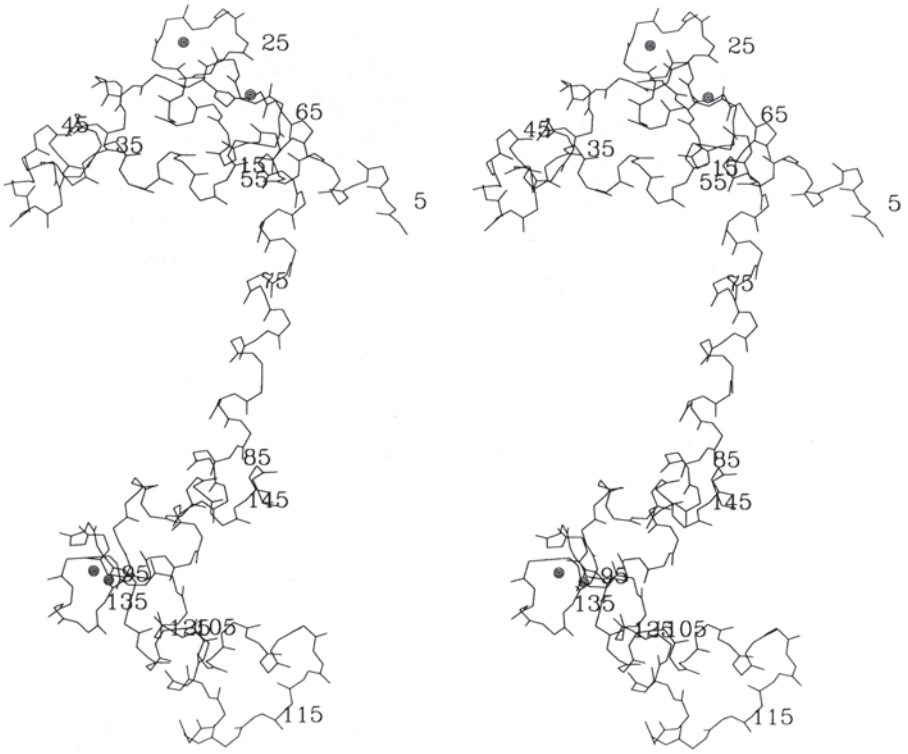
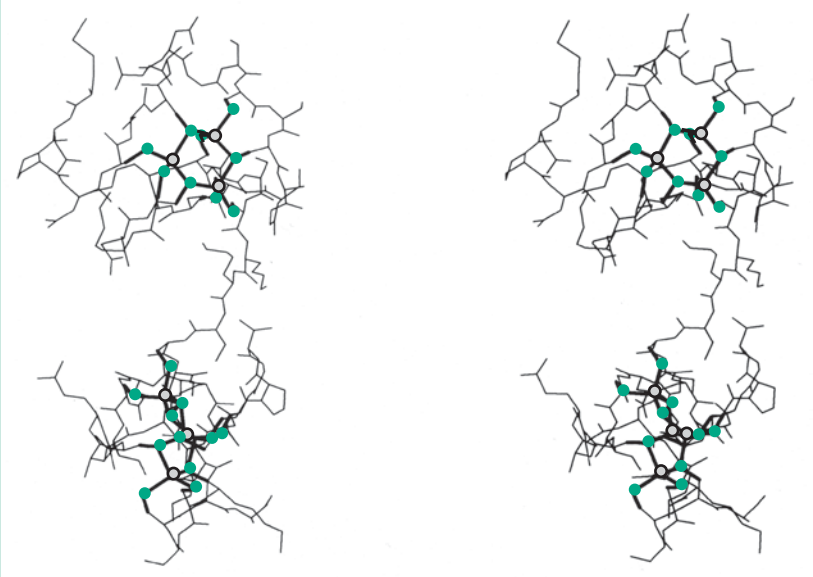



**Figure 6-8** Stereoscopic backbone trace of a 148-residue recombinant calmodulin. The two helix–turn–helix (EF-hand) loops and their bound $Ca^{2+}$ (as concentric circles) are at the top and the two near the C terminus are at the bottom. The long central helix, seen in this crystal structure, may undergo conformational changes during the functioning of this $Ca^{2+}$-sensing molecule.[132] From Chattopadhyaya *et al.*[111] Courtesy of F. A. Quiocho.

## BOX 6-E METALLOTHIONEINS

If animals ingest excessive amounts of Zn(II), Cd (II), Hg(II), or Cu(I) their livers and kidneys accumulate these metals as complexes of proteins called metallothioneins.[a–e] In mammals at least three related genes encode these metal-binding proteins. The best known, metallothionein II, has a highly conserved 61-residue sequence containing 20 cysteine residues and no aromatic residues. The protein is organized into two domains, each able to bind a cluster of metal ions via thiolate side chains. The three-dimensional structure of rat liver metallothionein containing five $Cd^{2+}$ and two $Zn^{2+}$ ions is shown in the accompanying stereoscopic diagram.[f] The 61 alpha carbons, the beta carbon and sulfur atoms (green) of cysteine residues and the bound metal ions are indicated.



Metallothionein containing bound $Cd^{2+}$ and $Zn^{2+}$. From Robbins *et al.*[f]

The N-terminal domain (residues 1–29) contains one $Cd^{2+}$ and two $Zn^{2}$ and nine cysteine sulfurs which bind the metal ions. Three of the sulfur atoms form bridges between pairs of metals. The second cluster contains four $Cd^{2+}$ held by 11 cysteine sulfur atoms, five of which bridge between pairs of metals. All of the metal ions are tetrahedrally coordinated. The polypeptide chains of metallothioneins consist predominantly of beta turns.[g] Important techniques in the study of these proteins include $^{113}Cd$ NMR,[h,i]

## BOX 6-E METALLOTHIONEINS (continued)

spectroscopic methods,[j,k] and X-ray absorption.[l]

Transcription of metallothionein genes is induced by metal ions, and toxic metals such as Cd and Hg accumulate as metallothionein complexes, suggesting that one function is to protect against metal toxicity.[m] However, synthesis is also induced by glucocorticoid hormones,[n] and accumulation of a high concentration of copper and zinc in fetal metallothionein suggests a role in storage of these essential metals.[c] Another metal that binds to metallothioneins is Au(I),[o] which is widely used in thiolate salts as a chemotherapeutic agent for rheumatoid arthritis.

Metallothioneins are also found in insects, lower invertebrates, and even in bacteria.[i,p] Nevertheless, there are other metal-binding proteins.[q,r] For example, albacore tuna contain a 66-kDa glycoprotein that contains eight mole percent histidine and binds three $Zn^{2+}$, each by a cluster of three His.[s] Plants and fungi contain **phytochelins**, peptides consisting largely of repeated γ-glutamylcysteine units.[a,t,u] These appear to protect plants against toxicity of cadmium in the same manner as do the metallothioneins in our bodies.


[a] Mehra, R. K., Garey, J. R., Butt, T. R., Gray, W. R., and Winge, D. R. (1989) *J. Biol. Chem.* **264**, 19747–19753

[b] Kägi, J. H. R., and Kojima, Y., eds. (1987) *Metallothionein II*, Birkhäuser, Basel, Berlin

[c] Fischer, E. H., and Davie, E. W. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3333–3334

[d] Kägi, J. H. R., and Schäffer, A. (1988) *Biochemistry* **27**, 8509–8515

[e] Hamer, D. H. (1986) *Ann. Rev. Biochem.* **55**, 913–951

[f] Robbins, A. H., McRee, D. E., Williamson, M., Collett, S. A., Xuong, N. H., Furey, W. F., Wang, B. C., and Stout, C. D. (1991) *J. Mol. Biol.* **221**, 1269–1293

[g] Pande, J., Pande, C., Gilg, D., Vasák, M., Callender, R., and Kägi, J. H. R. (1986) *Biochemistry* **25**, 5526–5532

[h] Cismowski, M. J., Narula, S. S., Armitage, I. M., Chernaik, M. L., and Huang, P. C. (1991) *J. Biol. Chem.* **266**, 24390–24397

[i] Vasák, M., Hawkes, G. E., Nicholson, J. K., and Sadler, P. J. (1985) *Biochemistry* **24**, 740–747

[j] Willner, H., Vasák, M., and Kägi, J. H. R. (1987) *Biochemistry* **26**, 6287–6292

[k] Lu, W., and Stillman, M. J. (1993) *J. Am. Chem. Soc.* **115**, 3291–3299

[l] George, G. N., Byrd, J., and Winge, D. R. (1988) *J. Biol. Chem.* **263**, 8199–8203

[m] Sadhu, C., and Gehamu, L. (1988) *J. Biol. Chem.* **263**, 2679–2684

[n] Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. M., and Beato, M. (1984) *Nature (London)* **308**, 513–519

[o] Laib, J. E., Shaw, C. F., III, Petering, D. H., Eidsness, M. K., Elder, R. C., and Garvey, J. S. (1985) *Biochemistry* **24**, 1977–1986

[p] Freedman, J. H., Slice, L. W., Dixon, D., Fire, A., and Rubin, C. S. (1993) *J. Biol. Chem.* **268**, 2554–2564

[q] Waalkes, M. P., and Peratoni, A. (1986) *J. Biol. Chem.* **261**, 13097–13103

[r] Willuhn, J., Schmitt-Wrede, H. P., Greven, H., and Wunderlich, F. (1994) *J. Biol. Chem.* **269**, 24688–24691

[s] Dyke, B., Hegenauer, J., Saltman, P., and Laurs, R. M. (1987) *Biochemistry* **26**, 3228–3234

[t] Reese, R. N., and Winge, D. R. (1988) *J. Biol. Chem.* **263**, 12832–12835

[u] Clemens, S., Kim, E. J., Neumann, D., and Schroeder, J. I. (1999) *EMBO J.* **18**, 3325–3333

# References


1. Lewis, G. N., and Randall, M. (1961) *Thermodynamics and the Free Energy of Chemical Substances*, 2nd ed., McGraw-Hill, New York
2. Mahan, B. H. (1963) *Elementary Chemical Thermodynamics*, Benjamin, New York
3. Everdell, M. H. (1965) *Introduction to Chemical Thermodynamics*, Norton, New York
4. Dickerson, R. E. (1969) *Molecular Thermodynamics*, Benjamin, New York
5. Jones, M. N., ed. (1988) *Biochemical Thermodynamics*, 2nd ed., Elsevier, Amsterdam
6. Hinz, H.-J., ed. (1986) *Thermodynamic Data for Biochemistry and Biotechnology*, Springer-Verlag, Berlin
7. Gutfreund, H., and Edsall, J. T. (1983) *Biothermo-dynamics*, Wiley, New York
8. Harold, F. M. (1986) *The Vital Force: A Study of Bioenergetics*, Freeman, San Francisco, California
9. Garby, L., and Larsen, P. S. (1995) *Bioenergetics; Its Thermodynamic Foundations*, Cambridge Univ. Press, London and New York
10. Jou, D., and Llebot, J. E., eds. (1990) *Introduction to the Thermodynamics of Biological Processes*, Prentice Hall, Englewood Cliffs, New Jersey

10a. Wadsö, I. (1985) *Eur. J. Biochem.* **153**, 429–434

11. Sturtevant, J. M. (1971) in *Techniques of Chemistry*, Vol. I, Part V (Weissberger, A., ed), pp. 347–425, Wiley, New York
12. Brown, H. D., ed. (1969) *Biochemical Microcalorimetry*, Academic Press, New York
13. Peters, K. S., and Snyder, G. J. (1988) *Science* **241**, 1053–1057
14. White, A., Handler, P., and Smith, E. L. (1968) *Principles of Biochemistry*, 4th ed., McGraw-Hill, New York (pp. 291–301)
15. Mosteller, R. D. (1988) *N. Engl. J. Med.* **318**, 1130
16. Donata, K., and Hegsted, D. M. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4866–4870
17. Ben-Naim, A. (1992) *Statistical Thermodynamics for Chemists and Biochemists*, Plenum, New York
18. Donovan, J. W. (1984) *Trends Biochem. Sci.* **9**, 340–344
19. Tanaka, A., Flanagan, J., and Sturtevant, J. M. (1993) *Protein Sci.* **2**, 567–576
20. Pace, C. N., and Laurents, D. V. (1989) *Biochemistry* **28**, 2520–2525
21. Interunion Commission on Biothermodynamics. (1976) *J. Biol. Chem.* **251**, 6879–6885
22. Phillips, R. C., George, P., and Rutman, R. J. (1963) *Biochemistry* **2**, 501–508
23. Liu, Y., and Sturtevant, J. M. (1995) *Protein Sci.* **4**, 2559–2561
24. Halling, P. J. (1989) *Trends Biochem. Sci.* **14**, 317–318
25. Örstan, A. (1990) *Trends Biochem. Sci.* **15**, 137–138
26. Nicolis, G., and Prigogine, I. (1977) *Self-organization in Non-equilibrium Systems: From Dissipative Structure to Order Through Fluctuations*, Wiley-Interscience, New York
27. Procaccia, I., and Ross, J. (1977) *Science* **198**, 716–717
28. Morowitz, H. J. (1968) *Energy Flow in Biology*, Academic Press, New York
29. Caplan, S. R., and Essig, A. (1983) *Bioenergetics and Linear Nonequilibrium Thermodynamics/ The Steady State*, Harvard Univ. Press, Cambridge, Massachusetts
30. Peacocke, A. R. (1983) *An Introduction to the Physical Chemistry of Biological Organization*, Clarendon Press, Oxford
31. Blumenfeld, L. A. (1983) *Physics of Bioenergetic Processes*, Springer-Verlag, Berlin
32. Katchalsky, A., and Curran, P. F. (1965) *Non-equilibrium Thermodynamics in Biophysics*, Harvard Univ. Press, Cambridge, Massachusetts
33. Prigogine, I. (1967) *Introduction to Thermodynamics of Irreversible Processes*, 3rd ed., Wiley, New York
34. Coveney, P. V. (1988) *Nature (London)* **333**, 409–415
35. Wilson, D. F. (1982) *Trends Biochem. Sci.* **7**, 275–278
36. Westerhoff, H. V. (1982) *Trends Biochem. Sci.* **7**, 275–279
37. Guynn, R. W., and Veech, R. L. (1973) *J. Biol. Chem.* **248**, 6966–6972
38. Alberty, R. A. (1969) *J. Biol. Chem.* **244**, 3290–3302
39. Phillips, R. C., George, P., and Rutman, R. J. (1969) *J. Biol. Chem.* **244**, 3330–3342
40. George, P., Phillips, R. C., and Rutman, R. J. (1963) *Biochemistry* **2**, 508–512
41. Purich, D. L., and Fromm, H. J. (1972) *Curr. Top. Cell. Regul.* **6**, 131–167
42. Jencks, W. P. (1968) in *Handbook of Biochemistry* (Sober, H. A., ed), pp. J-148, CRC, Cleveland, Ohio
43. Bassham, J. A., and Krause, G. H. (1969) *Biochim. Biophys. Acta.* **189**, 207–221
44. Rosing, J., and Slater, E. C. (1972) *Biochim. Biophys. Acta.* **267**, 275–290
45. Alberty, R. A. (1992) *Biophysical Chem.* **42**, 117–131
46. Alberty, R. A., and Cornish-Bowden, A. (1993) *Trends Biochem. Sci.* **18**, 288–291
47. Alberty, R. A., and Goldberg, R. N. (1992) *Biochemistry* **31**, 10610–10615
48. Alberty, R. A. (1994) *Pure & Appl. Chem.* **66**, 1641–1666
49. Stokes, G. B. (1988) *Trends Biochem. Sci.* **13**, 422–424
50. Mavrovouniotis, M. L. (1991) *J. Biol. Chem.* **266**, 14440–14445
51. Dutton, P. L. (1971) *Biochim. Biophys. Acta.* **226**, 63–80
52. Pepin, C. A., and Wood, H. G. (1986) *J. Biol. Chem.* **261**, 4476–4480
53. Wood, H. G., and Clark, J. E. (1988) *Ann. Rev. Biochem.* **57**, 235–260
54. Crooke, E., Akiyama, M., Rao, N. N., and Kornberg, A. (1994) *J. Biol. Chem.* **269**, 6290–6295
55. Veech, R. L., Lawson, J. W. R., Cornell, N. W., and Krebs, H. A. (1979) *J. Biol. Chem.* **254**, 6538–6547
56. Swedes, J. S., Sedo, R. J., and Atkinson, D. E. (1975) *J. Biol. Chem.* **250**, 6930–6938
57. Chapman, A. G., Fall, L., and Atkinson, D. E. (1971) *J. Bacteriol.* **108**, 1072–1086
58. Purich, D. L., and Fromm, H. J. (1973) *J. Biol. Chem.* **248**, 461–466
59. Teague, W. E., Jr., and Dobson, G. P. (1992) *J. Biol. Chem.* **267**, 14084–14093
60. Cohen, S. M., Ogawa, S., Rottenberg, H., Glynn, P., Yamane, T., Brown, T. R., and Shulman, R. G. (1978) *Nature (London)* **273**, 554–556

60a. Ugurbil, K., Kingsley-Hickman, P. B., Sako, E. Y., Zimmer, S., Mohanakrishnan, P., Robitaille, P. M. L., Thoma, W. J., Johnson, A., Foker, J. E., and From, A. H. L. (1987) *Ann. N.Y. Acad. Sci.* **508**, 265–286

61. Brosnan, M. J., Chen, L., Van Dyke, T. A., and Koretsky, A. P. (1990) *J. Biol. Chem.* **265**, 20849–20855

61a. Saupe, K. W., Spindler, M., Hopkins, J. C. A., Shen, W., and Ingwall, J. S. (2000) *J. Biol. Chem.* **275**, 19742–19746

62. Hutson, S. M., Williams, G. D., Berkich, D. A., LaNoue, K. F., and Briggs, R. W. (1992) *Biochemistry* **31**, 1322–1330
63. Burt, C. T., Glonek, T., and Bárány, M. (1977) *Science* **195**, 145–149
64. Seeley, P. J., Sehr, P. A., Gadian, D. G., Garlick, P. B., and Radda, G. K. (1977) in *NMR in Biology* (Dwek, R. A., Campbell, I. D., Richards, R. E., and Williams, R. J. P., eds), pp. 247–275, Academic Press, London
65. Masson, S., and Quistorff, B. (1992) *Biochemistry* **31**, 7488–7493
66. Jeffrey, F. M., Storey, C. J., Nunnally, R. L., and Malloy, C. R. (1989) *Biochemistry* **28**, 5323–5326
67. Kagimoto, T., Higaki, T., Nagata, K., Morino, Y., and Takatsuki, K. (1989) *NMR in Biomedicine* **2**, 93–97
68. Park, J. H., Brown, R. L., Park, C. R., Cohn, M., and Chance, B. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8780–8784
69. Radda, G. K. (1992) *FASEB J.* **6**, 3032–3038
70. Radda, G. K. (1986) *Science* **233**, 640–645
71. Burt, C. T. (1985) *Trends Biochem. Sci.* **10**, 404–406
72. Grove, T. H., Ackerman, J. J. H., Radda, G. K., and Bore, P. J. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 299–302
73. Bottomley, P. A. (1985) *Science* **229**, 769–772
74. Fossel, E. T., Morgan, H. E., and Ingwall, J. S. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3654–3658
75. Metzler, D. E., Harris, C. M., Johnson, R. J., Siano, D. B., and Thomson, J. A. (1973) *Biochemistry* **12**, 5377–5392
76. Johnson, R. J., and Metzler, D. E. (1970) *Methods Enzymol.* **18**, 433–471
77. Dixon, H. B. F. (1992) *Essays in Biochemistry* **27**, 161–176
78. Vander Jagt, D. L., Hansen, L. D., Lewis, E. A., and Han, L. B. (1972) *Arch. Biochem. Biophys.* **153**, 55–61
79. Dixon, H. B. F., and Tipton, K. F. (1973) *Biochem. J.* **133**, 837–842
80. Klotz, I. M., and Hunston, D. L. (1975) *J. Biol. Chem.* **250**, 3001–3009
81. Weber, G. (1975) *Adv. Prot. Chem.* **29**, 1–83
82. Chivers, P. T., Prehoda, K. E., Volkman, B. F., Kim, B.-M., Markley, J. L., and Raines, R. T. (1997) *Biochemistry* **36**, 14985–14991
83. Johnson, F. A., Lewis, S. D., and Shafer, J. A. (1981) *Biochemistry* **20**, 44–48
84. McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Körner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996) *Biochemistry* **35**, 9958–9966
85. Katritzky, A. R., and Waring, A. (1962) *J. Chem. Soc.*, 1540–1548
86. Katritzky, A. R., and Karelson, M. (1991) *J. Am. Chem. Soc.* **113**, 1561–1566
87. Bjerrum, J. (1941) *Metal Amine Formation in Aqueous Solution*, Haase & Son, Copenhagen
88. Eichhorn, G. L., ed. (1973) *Inorganic Biochemistry*, 2 vols, Elsevier, Amsterdam
89. Martell, A. E., and Smith, R. M. (1975) *Critical Stability Constants*, Plenum, New York
90. Michaelian, K. H., and Moskovits, M. (1978) *Nature (London)* **273**, 135–136
91. Lewsin, S. (1974) *Displacement of Water and its Control in Biochemical Reactions*, Academic Press, London
92. Karlin, S., Zhu, Z.-Y., and Karlin, K. D. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14225–14230
93. Freeman, H. C., Healy, M. J., and Scudder, M. L. (1977) *J. Biol. Chem.* **252**, 8840–8847
94. Torrado, A., Walkup, G. K., and Imperiali, B. (1998) *J. Am. Chem. Soc.* **120**, 609–610
95. Nakayama, S., and Kretsinger, R. H. (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23**, 473–507

## References


96. Kretsinger, R. H. (1976) *Ann. Rev. Biochem.* **45**, 239–266
96a. Pawlowski, K., Bierzynski, A., and Godzik, A. (1996) *J. Mol. Biol.* **258**, 349–366
97. Ikura, M. (1996) *Trends Biochem. Sci.* **21**, 14–17
98. Kretsinger, R. H., and Nockolds, C. E. (1973) *J. Biol. Chem.* **248**, 3313–3326
99. McPhalen, C. A., Sielecki, A. R., Santarsiero, B. D., and James, M. N. G. (1994) *J. Mol. Biol.* **235**, 718–732
99a. Declercq, J.-P., Evrard, C., Lamzin, V., and Parello, J. (1999) *Protein Sci.* **8**, 2194 – 2204
99b. Biekofsky, R. R., Martin, S. R., Browne, J. P., Bayley, P. M., and Feeney, J. (1998) *Biochemistry* **37**, 7617 – 7629
100. Drake, S. K., Zimmer, M. A., Miller, C. L., and Falke, J. J. (1997) *Biochemistry* **36**, 9917–9926
101. Akke, M., Forsén, S., and Chazin, W. J. (1995) *J. Mol. Biol.* **252**, 102–121
102. Denisov, V. P., and Halle, B. (1995) *J. Am. Chem. Soc.* **117**, 8456–8465
103. Heizmann, C. W., and Hunziker, W. (1991) *Trends Biochem. Sci.* **16**, 98–103
104. Ikemoto, N., Ronjat, M., Mészáros, L. G., and Koshita, M. (1989) *Biochemistry* **28**, 6764–6771
105. Meldolesi, J., and Pozzan, T. (1998) *Trends Biochem. Sci.* **23**, 10–14
106. Randolph, J. T., McClure, K. F., and Danishefsky, S. J. (1995) *J. Am. Chem. Soc.* **117**, 5712–5719
107. Fliegel, L., Burns, K., MacLennan, D. H., Reithmeier, R. A. F., and Michalak, M. (1989) *J. Biol. Chem.* **264**, 21522–21528
108. Cohen, P., and Klee, C. B., eds. (1988) *Calmodulin*, Elsevier, Amsterdam
109. James, P., Vorherr, T., and Carafoli, E. (1995) *Trends Biochem. Sci.* **20**, 38–42
110. Lafitte, D., Capony, J. P., Grassy, G., Haiech, J., and Calas, B. (1995) *Biochemistry* **34**, 13825–13832
111. Chattopadhyaya, R., Meador, W. E., Means, A. R., and Quiocho, F. A. (1992) *J. Mol. Biol.* **228**, 1177–1192
112. Pedigo, S., and Shea, M. A. (1995) *Biochemistry* **34**, 10676–10689
113. da Silva, A. C. R., and Reinach, F. C. (1991) *Trends Biochem. Sci.* **16**, 53–57
114. Gagné, S. M., Li, M. X., and Sykes, B. D. (1997) *Biochemistry* **36**, 4386–4392
115. Trigo-Gonzalez, G., Awang, G., Racher, K., Neden, K., and Borgford, T. (1993) *Biochemistry* **32**, 9826–9831
116. Spyracopoulos, L., Li, M. X., Sia, S. K., Gagné, S. M., Chandra, M., Solaro, R. J., and Sykes, B. D. (1997) *Biochemistry* **36**, 12138–12146
117. Travé, G., Lacombe, P.-J., Pfuhl, M., Saraste, M., and Pastore, A. (1995) *EMBO J.* **14**, 4922–4931
118. Näthke, I., Hill, B. L., Parham, P., and Brodsky, F. M. (1990) *J. Biol. Chem.* **265**, 18621–18627
119. Pley, U. M., Hill, B. L., Alibert, C., Brodsky, F. M., and Parham, P. (1995) *J. Biol. Chem.* **270**, 2395–2402
120. Engel, J., Taylor, W., Paulsson, M., Sage, H., and Hogan, B. (1987) *Biochemistry* **26**, 6958–6965
121. Engel, E., Richter, K., Obermeyer, G., Briza, P., Kungl, A. J., Simon, B., Auer, M., Ebner, C., Rheinberger, H.-J., Breitenbach, M., and Ferreira, F. (1997) *J. Biol. Chem.* **272**, 28630–28637
122. Schäfer, B. W., and Heizmann, C. W. (1996) *Trends Biochem. Sci.* **21**, 134–140
123. Drohat, A. C., Baldisseri, D. M., Rustandi, R. R., and Weber, D. J. (1998) *Biochemistry* **37**, 2729 – 2740
123a. Gribenko, A. V., and Makhatadze, G. I. (1998) *J. Mol. Biol.* **283**, 679 – 694
123b. Brodersen, D. E., Nyborg, J., and Kjeldgaard, M. (1999) *Biochemistry* **38**, 1695–1704
124. Christiansen, W. T., Tulinsky, A., and Castellino, F. J. (1994) *Biochemistry* **33**, 14993–15000
125. Schwalbe, R. A., Ryan, J., Stern, D. M., Kisiel, W., Dahlbäck, B., and Nelsestuen, G. L. (1989) *J. Biol. Chem.* **264**, 20288–20296
126. Demange, P., Voges, D., Benz, J., Liemann, S., Göttig, P., Berendes, R., Burger, A., and Huber, R. (1994) *Trends Biochem. Sci.* **19**, 272–276
127. Saurel, O., Cézanne, L., Milon, A., Tocanne, J.-F., and Demange, P. (1998) *Biochemistry* **37**, 1403–1410
128. Moss, S. E., and Crumpton, M. J. (1990) *Trends Biochem. Sci.* **15**, 11–12
129. Shapiro, L., Fannon, A. M., Kwong, P. D., Thompson, A., Lehmann, M. S., Grübel, G., Legrand, J.-F., Als-Nielsen, J., Colman, D. R., and Hendrickson, W. A. (1995) *Nature (London)* **374**, 327–336
130. Vyas, N. K., Vyas, M. N., and Quiocho, F. A. (1988) *Science* **242**, 1290–1295
131. Stuart, D. I., Acharya, K. R., Walker, N. P. C., Smith, S. G., Lewis, M., and Phillips, D. C. (1986) *Nature (London)* **324**, 84–87
132. Williams, R. J. P. (1991) *Trends Biochem. Sci.* **16**, 206

## Study Questions

1. a) From $\Delta G°$ for hydrolysis of sucrose (Table 6-6) calculate the equilibrium constant

$$K = [\text{glucose}][\text{fructose}] / [\text{sucrose}]$$

at 25°C. Call this hydrolysis reaction 1.

b) Is the sucrose in a 1 M solution stable? Explain.

c) If acid is added to a 1 M sucrose solution to catalyze its hydrolysis, what will be the final sucrose concentration at equilibrium? (Assume that concentrations equal activities for the purpose of these calculations.)

d) Reaction 2 is the hydrolysis of α-D-glucose 1-phosphate to glucose and inorganic phosphate ($P_i$). Using $\Delta G°$ for this reaction (Table 6-6) calculate the equilibrium constant.

e) Sucrose phosphorylase from the bacterium *Pseudomonas saccharophila* catalyzes the following reaction (reaction 3):

$$\text{Sucrose} + P_1 \rightarrow \alpha\text{-D-glucose 1-phosphate} + \text{fructose}$$

Calculate the equilibrium constant and the standard Gibbs energy change at 25°C for reaction 3 from the equilibrium constants obtained above for reactions 1 and 2. Show that $\Delta G°$ for reaction 3 = $\Delta G°$ of reaction 1 - $\Delta G°$ of reaction 2.

f) Could the bacterium carry out reaction 3 in the following two consecutive steps? Explain.

$$\text{Sucrose} \rightarrow \text{glucose} + \text{fructose}$$
$$\text{Glucose} + P_i \rightarrow \text{glucose 1-phosphate}$$

2. For each of the following reactions, state whether the equilibrium constant will be between 0.1 and 10 (i.e., about one), greater than 100, or less that 0.01. Assume that the pH is constant at 7.0.

a) $2\ ADP^{3-} \rightarrow ATP^{4-} + AMP^{2-}$

b) $ATP^{3-}$ + glucose → glucose 6-phosphate$^{2-}$ + $ADP^{2-} + H^+$

c) $ADP^{2-} + HPO_4^{2-} + H^+ \rightarrow ATP^{3-}$

d) Glucose 6-phosphate$^{2-}$ → fructose 6-phosphate$^{2-}$

e) Phosphoenolpyruvate$^{3-}$ + glucose → glucose 1-phosphate$^{2-}$ + pyruvate$^-$

3. The combustion of 1 mol of solid urea to liquid water and gaseous carbon dioxide and nitrogen ($N_2$) in a bomb colorimeter at 25°C (constant volume) liberated 666 kJ of heat energy. Calculate $\Delta H$, the change in heat content (enthalpy), for this reaction.

4. Using data of Table 6-4 calculate $\Delta G'$ (pH 7) for the following reactions:

a) Glucose → 2 lactate$^-$ + 2 $H^+$

b) $2\ NH_4^+ + HCO_3^- \rightarrow$ urea + $2\ H_2O + H^+$

c) 2-oxoglutarate$^{2-}$ + $1/2\ O_2$ + CoA + $H^+$ → succinyl-CoA + $H_2O + CO_2$

5. What is the ionic strength of a 0.2 M solution of NaCl? of 0.2 M $Na_2SO_4$?

6. The [ATP] / [ADP] ratio in an actively respiring yeast cell is about 10. What would be the intracellular [3-phosphoglycerate] / [1,3-bisphosphoglycerate] ratio have to be to make the phosphoglycerate kinase reaction (Fig. 9-7, reaction 7) proceed toward 1,3-bisphosphoglycerate synthesis at 25°C, pH 7?

7. a) Using data from Table 6-8 determine the equilibrium constant for the reaction between malate and methylene blue, assuming all reactants present initially at the same concentration. Indicate clearly the direction of the reaction for which the Gibbs energy change is written.

b) Calculate the percentage of the reduced (leuco) form of methylene blue present at pH 7 and 25°C in a system for which the measured electrode potential is 0.065 V.

8. $NAD^+$ is a coenzyme for both pyruvate dehydrogenase and ethanol dehydrogenase. Using the values of $E_o'$ from Table 6-8 calculate the Gibbs energy change and the equilibrium constant for the reaction.

Lactate$^-$ + acetaldehyde → pyruvate$^-$ + ethanol

9. Consider the oxidation of acetate at 25°C:

$$CH_3COO^-\ (0.1\ M) + 2\ O_2\ (g, 0.2\ atm) + H^+\ (10^{-7}\ M) \rightarrow 2\ H_2O\ (l) + 2\ CO_2\ (g)$$

a) What is the equilibrium pressure of $CO_2$ if the reaction is not coupled to any other reaction?

b) What is the equilibrium pressure of $CO_2$ if the reaction is coupled to the formation of 0.01 M ATP from 0.02 M ADP and 0.01 M $HPO_4^{2-}$ in the citric acid cycle?

c) What do the above calculations tell you about the prospects of gaining 100% efficiency of energy storage in ATP from the citric acid cycle?

d) If the actual pressure of $CO_2$ is 0.01 atm, what is the efficiency of energy storage under the conditions in (b)?

## Study Questions

10. The equilibrium constant for the following reaction, which is catalyzed by creatine kinase, has been determined by chemical analysis. The data are given below. [S. A. Kuby and E. A. Noltman, in *The Enzymes*, 2nd ed. (P. D. Boyer, H. Lardy, and K. Myrbäck, eds), Vol. VI, pp. 515–602. Academic Press, New York, 1962]

$$ATP^{4-} + \text{creatine} \rightarrow ADP^{3-} + \text{creatine phosphate}^{2-} + H^+$$

| $t$ (°C) | $K$ |
|---|---|
| 20 | $6.30 \times 10^{-9}$ |
| 30 | $5.71 \times 10^{-9}$ |
| 38 | $5.47 \times 10^{-9}$ |

a) What are $\Delta G°$, $\Delta H°$, and $\Delta S°$ for the reaction at 25°C?
b) What are $\Delta G'$, $\Delta H'$, and $\Delta S'$ (pH 7) for the reaction at 25°C?
c) What are $\Delta G'$, $\Delta H'$, and $\Delta S'$ (pH 7) for the hydrolysis of creatine phosphate at 25°C?

11. The following reaction was carried out in a calorimeter at 25°C in 0.1 M phosphate buffer at pH 7.4 in the presence of a particulate suspension containing the mitochondrial electron transport system [M. Poe, H. Gutfreund, and R. W. Estabrook, *ABB* **122**, 204–211 (1967)]:

$$NADH + H^+ + 1/2\ O_2\ (\text{aq., sat.}) \rightarrow NAD^+ + H_2O$$

The oxygen consumption was monitored continuously with an oxygen electrode. The temperature was monitored simultaneously with a thermocouple immersed in the solution. At the start of the reaction 96 µmol NADH was added to 29.0 ml buffer containing $O_2$. A nearly zero-order reaction was observed with the rate of $O_2$ consumption of 6.87 µmol/min and the rate of temperature rise of 0.01171 K/min. The heat capacity of the calorimeter and contents was 254.6 J/K. What is $\Delta H$ for the above reaction? NOTE: The $H^+$ is supplied by the phosphate buffer, which has a $\Delta H$ of dissociation of 5.4 kJ mol$^{-1}$.

12. Enthalpy and Gibbs energy changes for the following reaction at 25°C are given in Table 6-6.

$$ATP^{4-}\ (1\ M) + H_2O \rightarrow ADP^{3-}\ (1\ M) + H^+\ (10^{-7}\ M) + HPO_4^{2-}\ (1\ M)$$

a) How much heat is evolved at constant temperature and pressure if the reaction takes place in a test tube without doing any work other than $p\,\Delta V$ work?
b) How much heat is evolved or absorbed by the foregoing reaction if it is coupled with 100% efficiency to an endergonic reaction?
c) What efficiency of coupling to an endergonic reaction is required in order that the foregoing reaction neither evolve nor absorb heat?

13. Microorganisms use a great variety of fermentation reactions for obtaining energy. Could the following reaction be used for such a purpose? Explain the reasons for your answer.

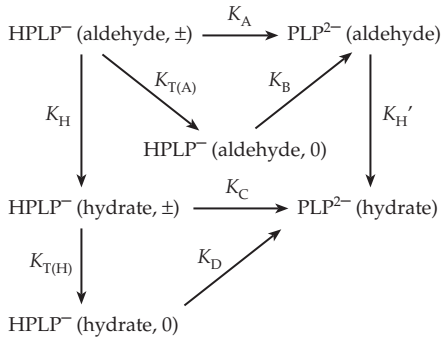
$$C_4H_6O_4\ (\text{succinic acid}) + H_2O \rightarrow C_3H_8O_3\ (\text{glycerol}) + CO_2$$

14. a) Calculate the work done in kJ and in kcal by a 70 kg person in climbing up stairs three stories (13 m).
b) Calculate how much ATP (in mmol and in grams) will be needed for the climb if muscles can use the ATP with 50% efficiency and if the phosphorylation state $R_p$ = [ATP] / [ADP] [$P_i$] is $10^3$ M$^{-1}$. The standard value of $\Delta G^{\circ\prime}$ (pH 7) for hydrolysis of MgATP to MgADP and $P_i$ may be taken as −31 kJ/mol at 37°C.
c) How much of this ATP could be provided by transfer of phospho groups from the stored creatine phosphate (Cr-*P*) to ADP in muscle? Assume that [Cr-*P*] = 20 mM and may fall to 10 mM during the climb. $\Delta G^{\circ\prime}$ (pH 7) for hydrolysis of creatine phosphate is about −43 kJ/mol. If the creatine kinase reaction attains equilibrium what will the value of $R_p$ be for the adenylate system?

15. Acid–base titration gave the following three p$K_a$ values for cysteine: 1.70, 8.36, and 10.53. Spectrophotometric data allowed H. B. F. Dixon and K. F Tipton (1973, *Biochem J.* **133**, 837–842) to estimate the following ratio of tautomeric species at pH 9.4 [$^-$OOC – C($NH_3^+$)– $CH_2S^-$] / [$^-$OOC – C($NH_2$) – $CH_2SH$] = 2.12, as is also shown in Fig. 6-5. Verify and assign the four microscopic p$K_a$ values.

## Study Questions

16. The coenzyme pyridoxal phosphate (PLP; structure on p. 740) has p$K_a$ values of 3.62, 6.10, and 8.33, as determined by acid–base titration or by spectrophotometric titration. The p$K_a$ of 6.10 belongs primarily to the phosphate group and is nearly independent of the others. However, the p$K_a$ values of 3.62 and 8.33 are shared by the protonated ring nitrogen and the phenolic –OH group. PLP exists as an equilibrium mixture of aldehyde together with its covalent hydrate (see Eq. 13-1). The equilibrium constants for hydrate formation ($K_h$ = [hydrate] / [aldehyde]) are independent of pH but differ for each ionization state of the ring. Consider only the equilibria in the pH range 7–12.



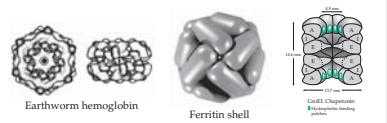

The UV-visible absorption spectrum of the monoprotonated form of PLP was divided mathematically into individual bands for the aldehyde with dipolar ionic ring (±), the aldehyde tautomer with an uncharged ring (0) and the hydrate of the dipolar ion. The following fractions were estimated (Harris *et al.*, 1976, *Biochem. Biophys. Acta.* **521**, 181–194.

| | |
|---|---|
| Aldehyde (±) | 52% |
| Aldehyde (0) | 28% |
| Hydrate (±) | 16% |
| Hydrate (0) | 4%, estimated indirectly |

From 300 MHz $^1$H NMR spectra areas for the following hydrogen atoms were estimated by Robitaille.

| | pH 7 | pH 12 |
|---|---|---|
| 4'-H aldehyde | .86 | .900 |
| 4'-H hydrate | .14 | .012 |
| 6-H, aldehyde | .96 | 1.059 |
| 6-H, hydrate | .17 | .014 |
| 2'-$CH_3$, aldehyde | 2.49 | 2.96 |
| 2'-$CH_3$, hydrate | .51 | .038 |

Evaluate the tautomerization constants $K_{T(A)}$ and $K_{T(H)}$, the hydration constants $K_H$ and $K_{H'}$, and the microscopic p$K_a$ values p$K_A$, p$K_B$, p$K_C$, and p$K_D$ in the foregoing scheme. Assume that the hydration ratios $K_h$ and $K_{h'}$ are the same in $H_2O$ (spectrophotometric data) and in $D_2O$ (NMR data).




Some ways in which protein subunits associate. (Left) The 3.66 MDa hemoglobin of the earthworm *Lumbricus terrestris* contains 144 globin subunits organized as 12 cylindrical disulfide-linked dodecamers. Two 6-dodecamer layers, each a ring with 6-fold cyclic symmetry, lie back-to-back. This reconstructed particle also contains three types of linker proteins in the center region. From Lamy *et al.* (2000) *J. Mol. Biol.* **298**, 638. (Center) The iron storage protein ferritin is formed from 24 19- to 21-kDa 4-helix-bundle subunits with cubic symmetry. As many as 4500 atoms of iron, as hydrated iron oxide, may be stored in the internal cavity. See Fig. 7-13. From Trikha *et al.* (1995) *J. Mol. Biol.* **248**, 954. Courtesy of Elizabeth Theil. (Right) The 2 MDa molecular chaperone GroEL consists of two back-to-back 7-subunit rings, each subunit formed from domains E, I and A. A 7-subunit cap of the smaller GroES may cover either end to form a compartment in which polypeptides fold. See Box 7-A.

# Contents

325 **A. Describing Binding Equilibria**
325 1. Analyzing Data
327 2. Multiple Binding Sites on a Single Molecule
328 *Microscopic binding constants and statistical effects*
329 *Electrostatic repulsion: anticooperativity*
330 3. Cooperative Processes
332 **B. Complementarity and the Packing of Macromolecules**
332 1. Rings and Helices
332 *Molecules with cyclic symmetry*
334 *Helical structures*
334 *Filamentous bacteriophages*
334 *A rod-shaped plant virus*
335 *Bacterial pili*
336 *The thin filaments of muscle*
337 2. Oligomers with Twofold (Dyad) Axes
337 *Paired interactions*
338 *Dihedral symmetry*
342 *Oligomers with cubic symmetry (polyhedra)*
344 *Asymmetry and quasi-equivalence in oligomers*
345 *Quasi-equivalence in virus coats*
348 *Regulatory subunits and multienzyme complexes*
349 **C. Cooperative Changes in Conformation**
349 1. Unequal Binding of Substrate and "Induced Fit"
350 2. Binding Equilibria for a Dimerizing Protein
350 *The Monod–Wyman–Changeux (MWC) model*
351 *The induced fit model*
352 *One conformational state dissociated*
352 3. Higher Oligomers
353 **D. The Oxygen-Carrying Proteins**
353 1. Myoglobin and Hemoglobin
353 *The binding of oxygen*
355 *Structural changes accompanying oxygen binding*
357 *The Bohr effect and allosteric regulators*
358 *Carbon monoxide, cyanide, and nitric oxide*
359 2. Abnormal Human Hemoglobins
362 3. Comparative Biochemistry of Hemoglobin
362 **E. Self-Assembly of Macromolecular Structures**
362 1. Bacteriophages
367 2. "Kringles" and Other Recognition Domains and Motifs
368 **F. The Cytoskeleton**
369 1. Intermediate Filaments
369 2. Microfilaments
370 3. Microtubules

373 **References**
377 **Study Questions**

**Boxes**

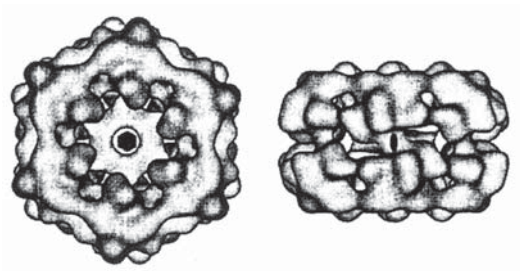
339 Box 7-A Life and Death for Proteins: Chaperonins and Proteasomes
360 Box 7-B Sickle Cell Disease, Malaria, and Blood Substitutes
363 Box 7-C The T-Even Bacteriophages
371 Box 7-D Mitosis, Tetraploid Plants, and Anticancer Drugs

**Tables**

329 Table 7-1 Binding Constants of Protons to Dianions of Dicarboxylic Acids
355 Table 7-2 Thermodynamic Functions for Oxygenation of Hemoglobin
367 Table 7-3 A Few Well-Known Structural Domains

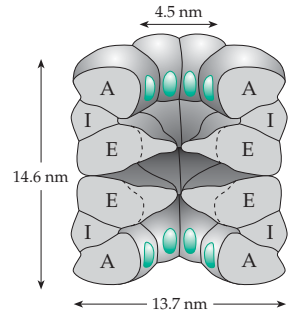
# How Macromolecules Associate

# 7



Earthworm hemoglobin

Ferritin shell




GroEL Chaperonin
Hydrophobic binding patches

The complicated shapes and internal structures of cells are determined to a large extent by the way in which proteins and other macromolecules are bonded one to another. In addition, intimate association of macromolecules is essential to such biological processes as the motion of flagella, the contraction of muscle, the action of antibodies, the transmission of nerve impulses, the replication of DNA, and the synthesis of proteins. Equally important is the binding of small molecules to large ones. In this chapter we will first examine methods of measuring binding with an emphasis on protons and small molecules. Then we will consider the ways in which macromolecules stick together as well as the role of conformational changes within macromolecules.

## A. Describing Binding Equilibria

In previous discussions of pH we have dealt with dissociation constants, but in this section we will use formation constants $K_f$, where $K_f = 1/K_d$. Measurement of the strength of association of molecules is an everyday aspect of modern biochemical research. It may be important to know how strongly a hormone binds to a receptor in a cell membrane or how well a feedback inhibitor binds to an enzyme to determine whether the interaction is significant physiologically. The binding of $O_2$ to hemoglobin and other oxygen carriers is vitally important, but the description of these oxygenation reactions is mathematically complex. This is especially so because we must consider effects of pH changes and of changing concentrations of allosteric effectors on the binding equilibria.

In considering such equilibria we must first examine the individual interactions of different domains of a protein, one with another. These can be described by association constants or, alternatively, by the Gibbs energy changes for the association reaction.[1,2] The average kinetic energy of motion of a molecule in solution is about $3/2k_BT$, where $k_BT$ is Boltzmann's constant. For one mole the kinetic energy is $3/2RT$ or 3.7 kJ (0.89 kcal) mol$^{-1}$ at 25°C. Thus, if $K_f = 10\ M^{-1}$ ($\Delta G° = -5.7$ kJ mol$^{-1}$ or $-1.36$ kcal mol$^{-1}$) the binding energy is only slightly in excess of the thermal energy of the molecules and the complex is weakly bound. In this instance, if X and P are both present in $10^{-4}$ molar concentrations (typical enough for biochemical systems), only 0.1% of the molecules will exist as the complex ([complex] = $K_f$[X][P]). If the formation constant is higher by a factor of 1000, i.e., $K_f = 10^4\ M^{-1}$ ($\Delta G° = -22.8$ kJ mol$^{-1}$), 38% of the molecules will exist as the complex; while if $K_f = 10^7\ M^{-1}$ (extremely strong binding, $\Delta G° = -40$ kJ or $-9.55$ kcal mol$^{-1}$), 97% of the molecules will be complexed.

### 1. Analyzing Data

The extent of binding of a molecule X to another molecule P (Eqs. 7-1, 7-2) is measured by varying the concentrations of X and P and observing changes in the concentration of the complex [PX]. The first

$$X + P \rightleftarrows PX \tag{7-1}$$

$$K_f = [PX]/[P][X]) \tag{7-2}$$

prerequisite is to find a measurable property that is different for the complex than for either of the free components. For example, the complex may be colored and the components colorless. More commonly, the complex simply has a different light absorbance (A) at

a certain wavelength than do the components. Likewise, the circular dichroism or the chemical shift of a peak in the NMR spectrum may change. If P is an enzyme, only the complex PX will undergo decomposition to products. Sometimes (but not always) the rate of breakdown of PX (the enzyme–substrate complex) to form products will be relatively slow compared to the rate at which the equilibrium between X, P, and PX is established. In this case the concentration of complex PX will be proportional to the observed rate of formation of product.

Whatever change of property is measured, its value will increase with increasing concentrations of X if the total concentration of the macromolecule P is kept constant. In the usual experimental design, the molar concentration of P is small and it is possible to increase the concentration of X to quite large values. When this is done, it is usually observed that at high enough values of [X] almost all of the P is converted to PX, and the change being measured (e.g., $\Delta A$ for increased light absorption) no longer increases. This effect is known as **saturation** and is observed in most binding studies and also in many physiological phenomena.

The property being measured ($\Delta A$) reaches a maximum value $\Delta A_{max}$ at saturation and when all of compound P has been converted to PX. The ratio of [PX] to the total concentration of all forms of P present $[P]_t$ is known as the **saturation fraction** and is often given the symbol Y. If P has more than one binding site for X, Y is defined as the fraction of the total binding sites occupied. If $n$ is the number of sites per molecule, the total number of sites is $n[P]$. The value of Y is often taken as $\Delta A/\Delta A_{max}$, an equality that holds for multisite macromolecules only if the change in $A$ is the same for each successive molecule of X added. This is not always true, but when it is Eq. 7-3 is followed.

$$\sum_i i\,\frac{[PX_i]}{n[P]_t} = Y = \frac{\Delta A}{\Delta A_{max}} \qquad (7\text{-}3)$$

Here $i$ represents the number of ligands X bound to P and may vary from 0 to $n$. When $n = 1$ the saturation fraction Y and $\Delta A$ are related to the concentration of free unbound X and the formation constant as follows:

$$Y = \frac{K_f[X]}{1 + K_f[X]} \qquad \Delta A = \frac{\Delta A_{max}\,K_f[X]}{1 + K_f[X]} \qquad (7\text{-}4)$$

A plot of Y or $\Delta A$ against [X] is shown in Fig. 7-1. This kind of plot is sometimes called an **adsorption isotherm** because it describes binding only at a constant temperature. Notice, from both Fig. 7-1 and Eq. 7-4, that Y reaches a value of 0.5 when [X] is just equal to $1/K_f$ (or to $K_d$). Note also that as [X] increases saturation is reached slowly and that even at the point representing the highest concentration of X ($8/K_f$ in Fig. 7-1) saturation is less than 90%. Since in the usual experimental situation, we do not know Y but only $\Delta A$, it is difficult to estimate the limiting value $\Delta A_{max}$ from a plot of this type unless $K_f$ is very high. However, we need to know $\Delta A_{max}$ to evaluate $K_f$. For this reason, plots like that of Fig. 7-1 are seldom used, this one being included mainly to illustrate a point of nomenclature. The curve shown in Fig. 7-1 is a rectangular hyperbola, and the type of saturation curve shown is frequently referred to as **hyperbolic**. This is in contrast to certain other binding curves (Section 3) which, when plotted in this way, are **sigmoidal** (S-shaped).

A better type of plot is often that of Y against log [X] (Fig. 7-2). It has the following features. (1) The curve is symmetric about the midpoint at log [X] = log $K_f$. (2) No matter how high or low the concentration range used in the experiments, it is easy to choose a scale that puts all the points on the same sheet of paper. (3) Spacing between points tends to be more uniform than in a plot against [X]; e.g., compare Figs. 7-1 and 7-2 for which the experimental points represent the same data and for which values of [X] for successive points are each twofold greater than the preceding one. (4) The same logarithmic scale can be used for all compounds, no matter how strong or weak the binding, and the same shape curve is obtained for all 1:1 complexes. The midpoint slope, $dY/d\log[X]$, is 0.576; the change in log [X] in going from 10 to 90% saturation is 1.81. The curve is familiar to most chemists because it is frequently used for pH titration curves in which pH substitutes for −log [X]. To represent a complex with tighter binding, the curve is simply moved to the left, and for weaker binding, it is moved to the right.

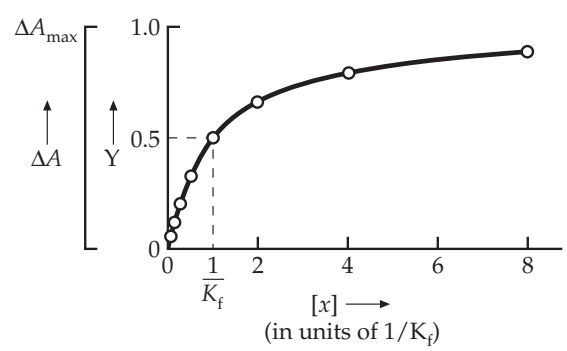



**Figure 7-1** An adsorption isotherm, a plot of the saturation fraction Y or of some change in a measured property $\Delta A$ vs [X], the concentration of a substance that binds reversibly to a macromolecule. The curve is hyperbolic and $[X] = 1/K_f$ when Y = 0.5.

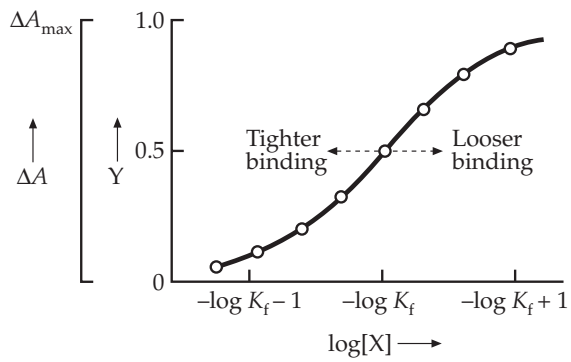



**Figure 7-2** A saturation curve plotted on a logarithmic scale for [X]. The data points are the same as those used in Fig. 7-1.

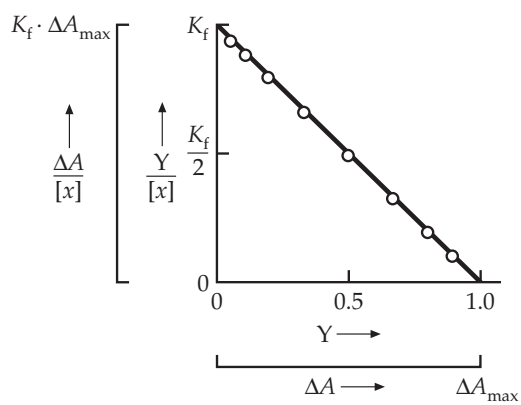



**Figure 7-3** A Scatchard plot of the same data shown in Figs. 7-1 and 7-2. This is the best of the linear plots for studying binding.

Saturation data are often plotted in yet another form known as the **Scatchard plot** (Fig. 7-3). The value of $\Delta A/[X]$ (or of $Y/[X]$) is plotted against $\Delta A$ (or Y) and a straight line is fitted to the points, preferably using the "method of least squares." The intercept on the $x$ axis and the slope of the fitted line give values of $\Delta A_{max}/K_f$ and $K_f$, respectively, as indicated by Eq. 7-5, which follows directly from Eq. 7-4.

$$Y/[X] = K_f - YK_f$$
$$\Delta A/[X] = \Delta A_{max} K_f - \Delta A K_f \qquad (7\text{-}5)$$

The Scatchard plot is the best of the various linear transformations of the saturation equation and is preferred to "double reciprocal plots" analogous to that shown in Fig. 9-3.

Scatchard's original equation was formulated to deal with the binding of two or more ligands to a single macromolecule.[3–5] If we let $[X]_b$ represent the concentration of bound X and $[P]_t$ the total molar concentration of the protein or other macromolecule and if there is only one binding site on the protein, $[X]_b/[P]_t$ will equal Y. However, if there are $n$ independent binding sites that have the same binding constant $K_f$ Eq. 7-5a will hold.

$$\frac{[X]_b}{[P]_t} \cdot \frac{1}{[X]} = nK_f - \frac{[X]_b}{[P]_t} K_f \qquad (7\text{-}5a)$$

If $[X]_b/[P]_t[X]$ is plotted against $[X]_b/[P]_t$ the resulting linear plot will have an intercept of $K_f$ on the $y$ axis and $n$ on the $x$ axis. Thus, $n$ is directly apparent, whereas in Eq. 7-5 it is incorporated into Y. A problem arises if, as discussed in the next section, the multiple binding sites are not independent but interact. Curved Scatchard plots result and attempts to extract more than one binding constant can lead to very large errors. Before measuring saturation curves, the student should read additional articles or books on the subject.[2,6–10]

## 2. Multiple Binding Sites on a Single Molecule

A macromolecule may often be able to bind several molecules of a second compound X. Consider the case in which the macromolecule P binds successively one molecule of X, then a second, and a third, up to a total of $n$. We define the stepwise formation constants, $K_1, K_2, \ldots, K_n$, as follows:

$$P + X \xrightarrow{K_1} PX$$
$$PX + X \xrightarrow{K_2} PX_2 \ldots$$
$$PX_{n-1} + X \xrightarrow{K_n} PX_n \qquad (7\text{-}6)$$

Remember that these are reversible reactions even though unidirectional arrows are used. The general expression for the $i$th stepwise formation constant is given by Eq. 7-7.

$$K_i = \frac{[PX_i]}{[PX_{i-1}][X]} \qquad (7\text{-}7)$$

Remember that Y is the fraction of total binding sites saturated. The number of moles of X bound per mole of P is $n$Y and is obtained by summing the concentra-

tions [PX] + 2[$PX_2$] + . . . and dividing the sum of all the forms of P:

For two binding sites ($n = 2$)

$$2Y = \frac{[PX] + 2[PX_2]}{[P] + [PX] + [PX_2]} \tag{7-8}$$

For the general case

$$nY = \left( \sum_{i=1}^{n} i\,[PX_i] \right) \Big/ \left( [P] + \sum_{i=1}^{n} [PX_i] \right) \tag{7-9}$$

The summations are over all of the integral values of $i$ from 1 to $n$. Now, by expressing each concentration, [$PX_i$], in terms of the concentrations [X] and [P] of *free* X and P, together with the stepwise formation constants, we obtain Eq. 7-10.

For $n = 2$

$$2Y = \frac{K_1[X] + 2K_1K_2[X]^2}{1 + K_1[X] + K_1K_2[X]^2} \tag{7-10}$$

A similar equation can be written for the general case. Note that the concentration of P does not appear in Eq. 7-10 and that Y is a function only of [X] and the stepwise formation constants. Such equations define the isotherms for binding of two or more molecules of X to P. From an experimental plot of Y (or of $\Delta A$) vs [X] or log [X], it is possible in favorable cases to determine the stepwise constants $K_1, K_2, ..., K_n$. However, this becomes quite complicated. To simplify Eq. 7-10 and the corresponding equation for the general case, we can group the constants together and designate the products of constants ($K_1$, $K_1K_2$, $K_1K_2K_3$, etc.) as $\psi_1, \psi_2, . . ., \psi_n$. Our equations are now as follows:

For $n = 2$

$$2Y = \frac{\psi_1[X] + 2\psi_2[X]^2}{1 + \psi_1[X] + \psi_2[X]^2} \tag{7-11}$$

For the general case

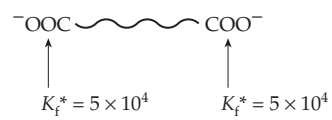
$$nY = \sum_{i=1}^{n} i\psi_i[X]^i \Big/ \left( 1 + \sum_{i=1}^{n} \psi_i[X]^i \right) \tag{7-12}$$

From experimental data, it is usually easiest to first determine the $\psi$'s (there are $n$ of them), and then to calculate from the $\psi$'s the stepwise constants. For example:

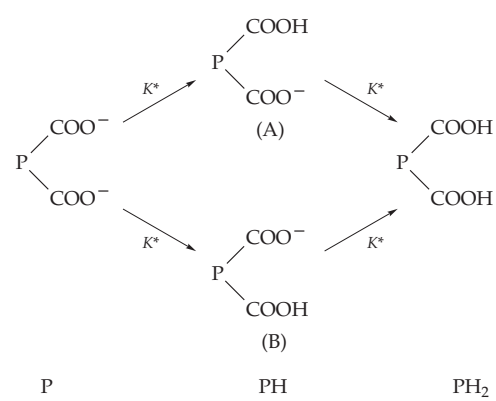
$$K_1 = \psi_1 \qquad K_2 = \psi_2/K_1\text{, etc.} \tag{7-13}$$

While Eq. 7-12, known as the **Adair equation**,[11] might seem to provide a complete description of the binding process, it usually does not. In many cases, there is more than one kind of binding site on a macromolecule and Eq. 7-12 tells us nothing about the distribution of the ligand X among different sites in complex PX. To consider this problem we must examine the *microscopic binding constants*.

***Microscopic binding constants and statistical effects.*** As discussed in Chapter 6, Section E,2, microscopic binding constants represent the constants for binding to specific individual sites. Now, consider a straight-chain dicarboxylic acid which has two *identical* binding sites for protons. If the chain connecting the two carboxylate anions is long enough, the carboxylate groups will be far enough apart that they do not influence each other through electrostatic interaction.




Each group will have a microscopic binding constant ($K_f^*$) of 5 x $10^4$ $M^{-1}$. The constant $K_f^*$ can also be called an **intrinsic binding constant**, because it is characteristic of a carboxylate group that is free of interactions with other groups. Intuition tells us (correctly) that, in its binding of protons, a solution of this dicarboxylic acid dianion will behave exactly like a solution of the monovalent anion R–$COO^-$ at twice the concentration. A single intrinsic binding constant suffices to describe both binding sites. It may seem surprising then that the **stepwise formation constants** (also called **stoichiometric** or **macroscopic** formation constants) $K_1$ and $K_2$ differ: $K_1$ = 10 x $10^4$ $M^{-1}$ and $K_2$ = 2.5 x $10^4$ $M^{-1}$. This fact reflects the so-called statistical effect. Either of the two carboxylate groups in the molecule can bind a proton in the first step to give two indistinguishable molecules, PH:




(7-14)

If we label the two forms of PH as A and B (Eq. 7-14) and consider that each one of them is independently in equilibrium with P through formation constant $K_f$*, we obtain Eq. 7-15 (which may be compared with Eqs. 6-75 and 6-76, which are written for *dissociation* constants).

$$K_1 = \frac{[PH]_A + [PH]_B}{[P][H^+]} = 2K^* \quad \text{and} \quad K_2 = K^*/2 \tag{7-15}$$

This result is related to probability and arises for the same reason that if you reach into a barrel containing 50% white balls and 50% black balls, you will pull out one of each just twice as often as you will pull out a pair of white or a pair of black. In the general case of $n$ equivalent binding sites, the microscopic formation constants $K_i$* are related[12,13] to the stepwise constants $K_i$ as follows:

$$K_i = \frac{(n + 1 - i)}{i} K_i^* \tag{7-16}$$

It is also easy to show,[14] using Eqs. 7-12 and 7-16 that for *n completely equivalent and independent* binding sites Eqs. 7-17 and 7-18 hold:

$$Y = \frac{K^*[X](1 + K^*[X])^{n-1}}{(1 + K^*[X])^n} \tag{7-17}$$

or

$$Y = \frac{K^*[X]}{1 + K^*[X]} \tag{7-18}$$

In this case the microscopic association constants are all identical and represent a single **intrinsic** constant applicable to all of the sites. In fact, Eq. 7-18 is identical to that for association of a single proton (or other ligand) with a single binding site, satisfying our intuitive notion that a set of $n$ completely independent binding sites should behave just like a solution of an $n$-fold more concentrated compound with a single binding site. Thus, our arithmetic has led us to a conclusion that was already obvious. However, it is rarely true that binding sites on a single macromolecule are completely independent; there is almost always *interaction* between them, and the equations that we have derived for evaluation of stepwise and intrinsic constants cannot be applied without modification.

***Electrostatic repulsion: anticooperativity.*** As we have seen, a hypothetical acid with an infinite distance between the carboxylate groups and log $K_f$* = 4.8 would have two macroscopic binding constants separated by the statistical distance (log 4 = 0.6). Compare these values with the observed binding constants for protons with the dianions of acids containing 7, 4, 2, and 1 $CH_2$ groups given in Table 7-1. For the longest chain (that of azelaic acid) the log $K_f$ values are not very different from those of the hypothetical long-chain acid. However, as the groups come closer together, the first binding constant is increased markedly because of the additional electrostatic attraction and the second is decreased. The spread between the two log $K_f$ values increases from 0.6 to as much as 2.9 as a result of interaction between the binding sites.

**TABLE 7-1**
**Binding Constants of Protons to Dianions of Dicarboxylic Acids[a]**

| Acid dianion | No. of $CH_2$ groups | log $K_1$ ($pK_2$) | log $K_2$ (association) ($pK_1$) (dissociation) |
|---|---|---|---|
| Hypothetical dianion with log $K^*$ = 4.8 | ∞ | 5.1 | 4.5 |
| Azelaic | 7 | 5.41 | 4.55 |
| Adipic | 4 | 5.41 | 4.42 |
| Succinic | 2 | 5.48 | 4.19 |
| Malonic | 1 | 5.69 | 2.83 |

[a] From R. P. Bell, (1973) *The Proton in Chemistry*, 2nd ed., p. 96. Cornell Univ. Press, Ithaca, New York

In malonic and succinic acids the first proton bound can be shared by both carboxyl groups through formation of a hydrogen bond. (See discussion in Chapter 9, Section D.) Additional factors operate in oxalic acid where the carboxyl groups are connected directly and for which $pK_a$ values ($pK_a$ = log $K_f$) are 4.19 and 1.23. In all of these examples the binding of the first proton makes it harder to bind a second proton. Such negative interaction or **anticooperativity** between binding sites is very common and always leads to a spread of the formation constants and a broadening of the curve of Y vs log [X]. This is shown graphically on the right side of Fig. 7-4 where the binding curves for protons with acetate ion and with succinic acid dianion are compared. Notice that binding of protons *increases* as log $[H^+]$ increases, giving the curves an unfamiliar appearance when compared with the more familiar curve of *dissociation vs* pH.

Can we predict the $pK_a$ values in Table 7-1? With an appropriate dielectric constant chosen Eq. 2-8 can be applied. The difference between the two successive log $K_f$ values reduced by 0.6 (the statistical factor) is a

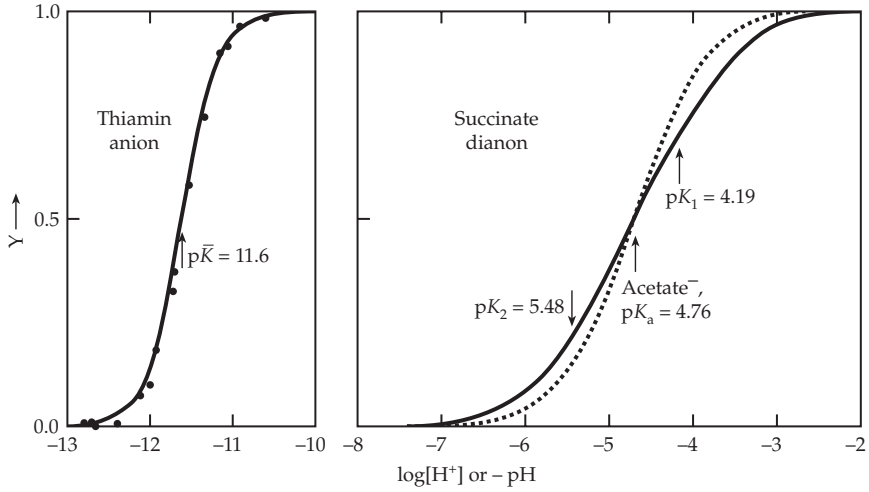
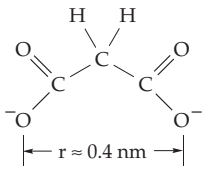



**Figure 7-4** Binding of protons to the thiamin anion, the succinate dianion, and the acetate anion. Acetate (dashed line) binds a single proton with a normal width binding curve. Succinate dianion binds two protons with anticooperativity, hence a broadening of the curve. The thiamin anion (yellow form, see Eq. 7-19) binds two protons with complete cooperativity and a steep binding curve.

measure of the electrostatic effect. For malonic acid this $\Delta pK_a$ is 2.25 and for succinic acid it is 0.69 (Table 7-1). In 1923, N. Bjerrum proposed that the value of $\Delta pK_a$ could be equated directly with the work needed to bring the two negative charges together to a distance representing the charge separation in the malonate dianion.




Thus, applying Eq. 6-31, $\Delta G = 5.708\ \Delta pK_a$ kJ mol$^{-1}$ = 12.84 kJ mol$^{-1}$ for malonate. Equating this with W in Eq. 2-8 and assuming a dielectric constant of 78.5 (that of water), the distance of charge separation $r$ is calculated to be 0.138 nm. This is much too small. The computation was improved by Westheimer and Kirkwood, who assumed a dielectric constant of 2.0 *within* the molecule. By approximating the molecule as an ellipsoid of revolution, they were able to make reasonably accurate calculations of electrostatic effects on $pK_a$ values.[15] Thus, for malonic acid Westheimer and Shookhoff[16] predicted $r$ = 0.41 nm for malonic acid dianion. Recently more sophisticated calculations[17] have been used to predict $pK_a$ values for the compounds in Table 7-1 and others.[18]

Electrostatic theory has also been used successfully to interpret titration curves of proteins in which the net negative or positive charge distributed over the surface of the protein varies continuously from high pH to low as more protons are added.[19]

Electrostatic effects can be transmitted extremely effectively through aromatic ring systems, a fact that explains some of the significance of heterocyclic aromatic systems in biochemical molecules. Consider the microscopic binding constant of the phenolate anion of pyridoxine as influenced by the state of protonation of the ring nitrogen. These are shown in Eq. 6-75, where $pK_a^* = 4.94$ and $pK_d^* = 8.20$ define the binding constants for protonation of a phenolate ion when the ring nitrogen is protonated or unprotonated, respectively. We see that $\Delta\ pK_a = 3.26$, even greater than that of malonic acid.

## 3. Cooperative Processes

Can it ever happen that interaction between groups leads to a *decrease* from the statistical separation between values of the stepwise constants instead of to an increase? At first glance, the answer seems to be no. A decreased separation would imply that the intrinsic binding constant for the second proton bound is higher than that for the first, but common sense tells us that the first proton ought to bind at the site with the highest binding constant. However, look at the experimental binding curve of protons with the anion of thiamin shown in Fig. 7-4. Instead of being broadened from the curve of acetate, it is just half as wide. The explanation depends upon some rather amusing chemistry of thiamine. Under suitable conditions, this vitamin can be crystallized as a yellow sodium salt, the structure of whose anion is shown in Eq. 7-19. Weak binding of a proton to one of the nitrogens as shown in Eq. 7-19 creates an electron deficiency at the adjacent carbon and the –S⁻anion adds to the C=N group, closing the ring to an unstable tricyclic form of thiamin. This tricyclic form can be observed in methanol and can be crystallized. It is unstable in water because the central ring can open, with the electrons flowing as indicated by the small arrows to create a strongly basic site on the same nitrogen. A second proton combines at this basic nitrogen with a high binding constant to form a cation.

*The key to the reversed order of strength of the binding*

*constants lies in the molecular rearrangements intervening between the two binding steps.*[20,21] In this particular case, we cannot measure the successive binding constants $K_1$ and $K_2$ directly from the titration curve because $K_2$ is almost two orders of magnitude larger than $K_1$. Consequently, the binding curve shown in Fig. 7-4 is (within the experimental error) twice as steep at the center (the slope is 2 x 0.576) as that for acetate ion and is accurately represented in Eq. 7-20. Comparison of Eq. 7-20 with Eq. 7-10 shows how the latter has been simplified because no significant concentration of the form PX is present in the cooperative case.

$$Y = \frac{\bar{K}^2[X]^2}{1 + \bar{K}^2[X]^2} \qquad (7\text{-}20)$$

This is only part of the story about the acid–base chemistry of thiamin. For the rest, see Chapter 14, Section D,1.

The binding of protons by the thiamin anion is an example of a **cooperative process**, so named because binding of the first proton makes binding of the second easier. Although relatively rare among small molecules, cooperative processes are very common and important in biochemistry.[22,23] A cooperative binding curve is sometimes referred to as **sigmoidal** because the plot of Y against [X] (the binding isotherm) is S shaped. The maximum possible cooperativity is observed when the binding of the first ligand enhances the affinity of all other sites so much that no species other than P and $PX_n$ are present in significant concentration.

It is easy to show that, for $n$ binding sites with such *completely cooperative binding*, the saturation fraction is:

$$Y = \frac{\bar{K}^n[X]^n}{1 + \bar{K}^n[X]^n} \qquad (7\text{-}21)$$

where $\bar{K} = (K_1 \ldots K_n)^{1/n}$. The midpoint slope in the binding curve (Y vs log [X]) is $0.576n$ and the change, $\Delta\log[X]$, between Y = 0.1 and Y = 0.9 is $1.81/n$.

Equation 7-21 can be rewritten as

$$Y/(1 - Y) = \bar{K}^n[X]^n \qquad (7\text{-}22)$$

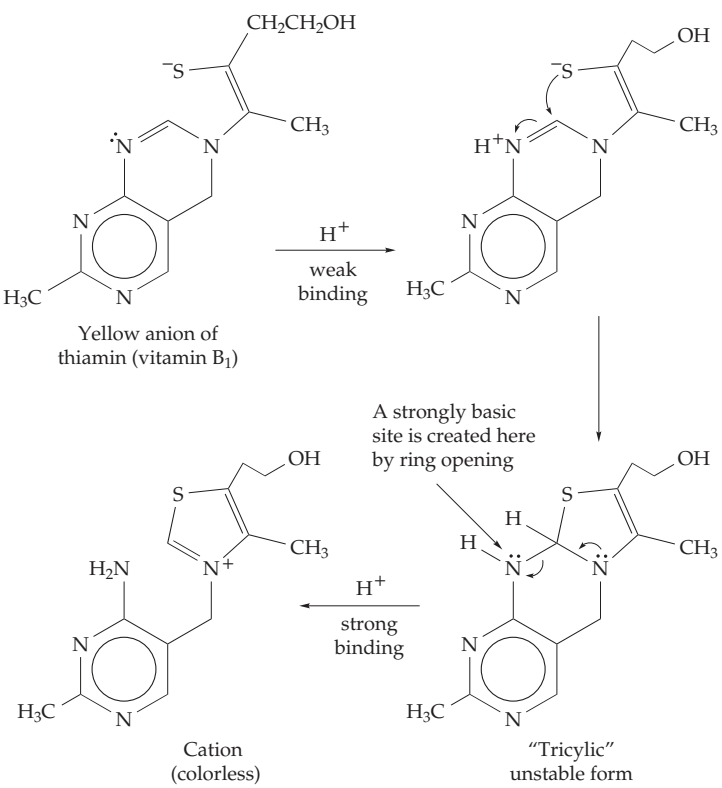
Taking logarithms (Eq. 7-22)

$$\log[Y/(1 - Y)] = n \log \bar{K} + n \log[X] \qquad (7\text{-}23)$$

A plot of log [Y/(1 – Y)] vs log [X] is known as a **Hill plot**. According to Eq. 7-22, it is linear with a slope of $n$. Remember that this equation was derived for an ideal case of completely cooperative binding at $n$ sites.[24] However, Hill plots are often used to plot experimental data for systems in which cooperativity is incomplete. Thus, the experimentally measured slope of a Hill plot ($n_{Hill}$) is not an integer and is usually less than $n$, the number of binding sites. A comparison of $n_{Hill}$ with $n$ is often used as a measure of the degree of cooperativity: $n_{Hill}/n = 1.00$ for complete cooperativity but is less than one if cooperativity, is incomplete. An example of a very high degree of cooperativity is provided by the hexameric enzyme glutamate dehydrogenase, whose saturation curve for substrate displays $n_{Hill}$ approaching six.[25] It is not necessary to make a Hill plot to get $n_{Hill}$. From the usual binding curve of Y (or $\Delta A$) vs log [X] the midpoint slope can be measured with satisfactory precision. Alternatively, the difference, $\Delta\log[X]$, between 0.1 and 0.9 saturation can be evaluated and $n_{Hill}$ calculated from Eq. 7-24.

$$n_{Hill} = \frac{\text{midpoint slope}}{0.576} = \frac{1.81}{\Delta \log[X]} \qquad (7\text{-}24)$$

Binding curves sometimes show more than one step; in such cases Hill plots are not linear and no simple measure of cooperativity can be defined.




(7-19)

A second example of cooperativity is provided by the reversible denaturation of coiled peptide chains. Some proteins can be brought to a pH of 4 by addition of acid but without protonation of buried groups with intrinsic $pK_a$ values greater than four. When a little more acid is added, some less basic group is protonated, permitting the protein to unfold and to expose the more basic hidden groups. Thus, cooperative proton binding is observed. As in the case of thiamin the cooperativity depends upon the occurrence of a conformational change in the molecule linked to protonation of a particular group.

The reversible transformation between an α helix and a random coil conformation is also cooperative. In this case, once a helix is started, additional turns form rapidly and the molecule is completely converted into the helix. Likewise, once it unfolds it tends to unfold completely. Melting of DNA (Chapter 5) or, indeed, of any crystal is cooperative.[8] The stacking of nucleotides alongside a template polynucleotide can also be cooperative. For example, the binding of an adenylate residue to two strands of polyuridylic acid leads to cooperative formation of a triple-helical complex (Chapter 5, Section C,4). Here the stacking interactions make helix growth energetically easier than initiation of new helical regions.[26]

## B. Complementarity and the Packing of Macromolecules

Because the forces acting between them are weak, two molecules will cling together tightly only if there is a close fit between their surfaces. For a firm bond to be formed many atoms must be in contact and the two molecular surfaces must be *complementary* one to the other. If a "knob," such as a $-CH_3$ group, is present on one surface, there must be an appropriate hollow in the complementary surface. A positive charge in one surface is likely to be opposite a negative charge in the other. A proton donor group can form a hydrogen bond only if it is opposite a group with unshared electrons; nonpolar (hydrophobic) groups must be opposite each other if hydrophobic interaction is to occur. An important principle is that *two molecules with complementary surfaces tend to join together and interact, whereas molecules without complementary surfaces do not interact*. Watson called this "selective stickiness."[27] Selective stickiness permits the **self-assembly** of biological macromolecules having surfaces of complementary shape into fibers, tubes, membranes, and polyhedra. It also provides the means for specific pairing of purine and pyrimidine bases during the replication of DNA and during the synthesis of RNA and of proteins.

Complementarity of surfaces is equally important to the chemical reactions of cells. Each of these reactions is catalyzed by an enzyme, which contains reactive chemical groupings in the right places and in the right orientations to interact with and promote a chemical change in another molecule, the **substrate**. Specific catalysis is one of the most basic characteristics of living things. Enzymatic catalysis provides the basis not only for the reactions of metabolism but also for the movement of muscle fibers, the flowing of the cytoplasm in the ameba, and virtually all other biological responses. To understand these phenomena requires an examination of the structures of the macromolecules involved and of the ways in which they can fit together.

Just as the amino acids, sugars, and nucleotides are the building blocks for formation of proteins, polysaccharides, and nucleic acids, these three kinds of macromolecule are the units from which larger subcellular structures are assembled. Fibers, microtubules, virus "coats," and small symmetric groups of **subunits** in **oligomeric proteins** all result from the packing of macromolecules in well-defined ways, something that is often called **quaternary structure**.

### 1. Rings and Helices

Consider first the aggregation of identical protein subunits. While many protein molecules are nearly spherical, they are nevertheless asymmetric. In the drawings that follow the asymmetry is exaggerated, but the principles illustrated are valid. One easily observed lesson from nature is that even though living things are made up of asymmetric materials, a great deal of symmetry is evident.[28] At the molecular level the symmetry of crystalline arrays of atoms or molecules is described mathematically by the elements of symmetry present in **space groups** (p. 133). There are 230 of these but only 65 accommodate asymmetric objects (Chapter 3).[29] Two of the natural ways for identical asymmetric subunits to interact lead to rings and helices, respectively.

***Molecules with cyclic symmetry.*** Consider a subunit (**protomer**) of the shape shown in Eq. 7-25 and containing a region *a* that is complementary to the surface *j* on another part of the same molecule. Two such protomers will tend to stick together to form a dimer, region *a* of one protomer sticking to region *j* of the other. The dimer will still contain a free region *a* at one end and a region *j* at the other which are not involved in bonding. Other protomers can stick to these free ends. In some instances long chains can be formed. However, if the geometry is just right, a third subunit can fit in to form a closed ring (a trimer). Depending on the geometry of the subunits, the ring can be even smaller (a dimer) or it can be larger (a tetramer, pentamer, etc.). The bonding involved is

between two different regions (*a* and *j*) of a subunit and is sometimes described as **heterologous**.[30] To obtain a closed ring of subunits, the angle between the bonding groups *a* and *j* must be correct or the ring cannot be completed.

A ring formed using exclusively heterologous interactions possesses **cyclic symmetry**. The trimer in Eq. 7-25 has a **threefold axis**: Each subunit can be superimposed on the next by rotation through 360°/3. The oligomer is said to have $C_3$ symmetry. Many real proteins, including all of those with 3, 5, or another uneven number of identical protomers, appear to be formed of subunits arranged with cyclic symmetry. An example is the cholera toxin from *Vibrio cholerae*, which forms a pentamer with an outer ring of subunits with $C_5$ symmetry (Fig. 7-5).

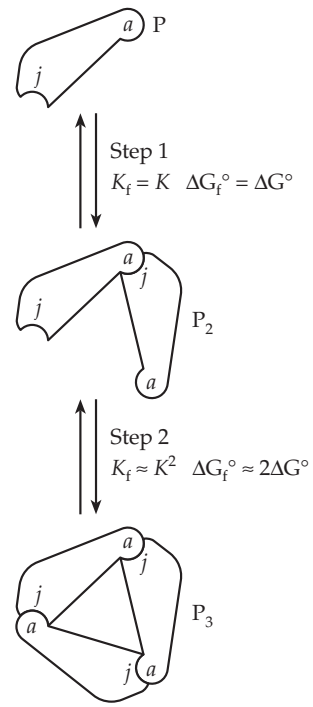
Now consider the quantitative aspects of heterologous interactions with ring formation. Let $K_f$ be the formation constant and $\Delta G°$ the Gibbs energy change for the reaction of the *j* end of protomer P with the *a* end of a second protomer to form the dimer $P_2$ (Eq. 7-25).

In the second step (Eq. 7-25) a third protomer combines. It forms *two* new *aj* interactions. If we assume for this step that $\Delta G_f°$ is 2 $\Delta G°$, $K_f$ will be equal to $K^2$. The overall association constant for formation of a trimer from three protomers will be given by Eq. 7-26.

$$3\,P \rightarrow P_3$$
$$K_f = K^3$$
$$\Delta G_f° = 3\,\Delta G° \qquad (7\text{-}26)$$

This will be true only if $\Delta G°$ for formation of both new *aj* bonds in the trimers is exactly the same as that for formation of the *aj* bond in the dimer. The reader may wish to criticize this assumption[30a,b,c] and to suggest conditions that might lead to overestimation or underestimation of $K_f$ for the trimer as calculated previously.

Now consider a hypothetical example: Protomer P is continuously synthesized by a cell. At the same time some subunits are degraded to a nonaggregating form via a second metabolic reaction. The two reactions are balanced so that [P] is always present at a steady state




(7-25)

value of $10^{-5}$ M. Suppose that a value for a single *aj* interaction of $K = 10^4$ (and $\Delta G° = -22.8$ kJ mol$^{-1}$) governs aggregation to form dimers and trimeric rings. What concentration of dimers and of trimer rings will be present in the cell in equilibrium with the $10^{-5}$ concentration of P. Using Eq. 7-25 we see that the concentration of dimers $[P_2]$ is $10^4 \times (10^{-5})^2 = 10^{-6}$ M. (Note that the amount of material in this concentration of dimer is equivalent to $2 \times 10^{-6}$ M of the monomer units.) The concentration of rings $[P_3]$ is $(10^4)^3 \times (10^{-5})^3 = 10^{-3}$ M, equivalent to $3 \times 10^{-3}$ M of the monomer units. Thus, of the *total* P present in the cell ($10^{-5} + 0.2 \times 10^{-5} + 300 \times 10^{-5}$ M), 99.6% is associated with trimers, 0.33% is still monomers, and only 0.07% exists as dimers. Thus, the formation of two heterologous bonds simultaneously to complete a ring imparts a high degree of cooperativity to the association reaction of Eq. 7-25. We will find in a cell mostly either rings or monomer but little dimer.

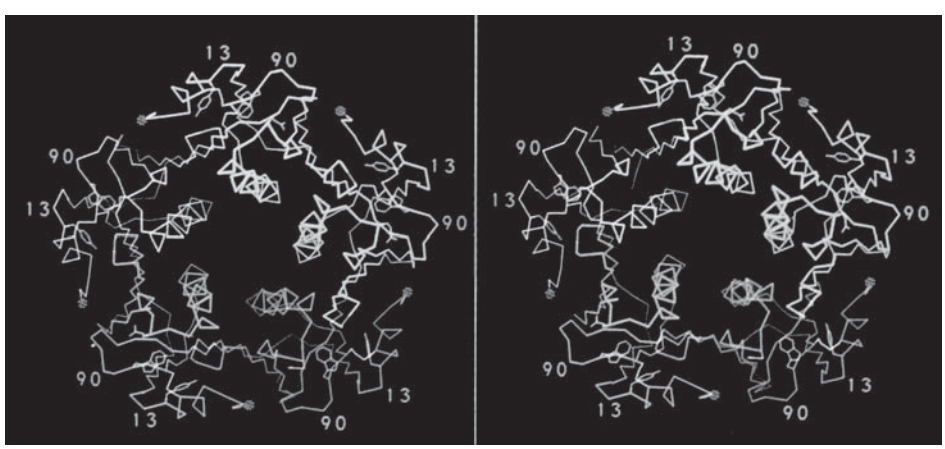



**Figure 7-5** Stereoscopic view of the $B_5$ pentamer of cholera toxin B. The pentamer, known as **choleragenoid**, has a central hole of ~1.5 nm diameter into which a helix from the A subunit is inserted. As viewed here, the front surface of the pentamer has binding sites for the oligosaccharide chains of ganglioside $GM_1$, which serves as the toxin receptor. The back side binds the A subunit. See also Box 11-A. From Zhang *et al.*[31]

Now consider what will happen to the little rings within the cell if the process that removes P to a non-associating form suddenly becomes more active so that [P] falls to $10^{-6}$ M. If $K$ is still $10^4$, what will be the percentages of P, $P_2$, and $P_3$ at equilibrium? Here we note a characteristic of cooperative processes: A higher than first power dependence on a concentration.

***Helical structures.*** If the angle at the interface *aj* is slightly different, instead of a closed ring, we obtain a helix as shown in Fig. 7-6A. The helix may have an integral number of subunits per turn or it may have a nonintegral number, as in the figure. The same type of heterologous interaction *aj* is involved in joining each subunit to the preceding one, but in addition other interactions occur. If the surfaces involved in these additional interactions are complementary and the geometry is correct, groups from two different parts of the molecule (e.g., *b* and *k*) may fit together to form another heterologous bond. Still a third heterologous interaction *cl* may be formed between two other parts of the subunit surfaces. If interactions *aj*, *bk*, and *cl* are strong (i.e., if the surfaces are highly complementary over large areas), extremely strong microtubular structures may be formed, such as those in the flagella of eukaryotic organisms (Fig. 1-8). If the interactions are weaker, labile microfilaments and microtubules, such as are often observed to form and dissociate within cells, may arise.

The geometry of subunits within a helix is often advantageously displayed by imagining that the surface of the structure can be unfolded to give a radial projection (Fig. 7-6B). Here subunits corresponding to those in the helix in Fig. 7-6A are laid out on a plane obtained by slitting the cylinder representing the surface of the helix and laying it out flat. In the example shown, the number of subunits per turn is about 4.8 but it can be an integral number. The interactions *bk* between subunits along the direction of the fiber axis may sometimes be stronger than those (*aj*) between adjacent subunits around the spiral. In such cases the microtubule becomes frayed at the ends through breaking of the *aj* interactions. This phenomenon can be observed under the electron microscope for the microtubules from flagella of eukaryotic organisms. Figures 7-7 to 7-10 show four helical structures from the molecular domain. They are a filamentous bacteriophage, a plant virus, a bacterial pilus, and an actin microfibril. Each is composed largely of a single kind of protomer. A larger and more complex helical structure, the microtubule, is shown in Fig. 7-34.

***Filamentous bacteriophages.*** Bacteriophages of the Ff family include the fd, f1, and M13 strains.[31a,32–36] Phage M13 is widely used in cloning genes and for many other purposes (Chapter 26). The genome is a circular, single-stranded DNA of ~6400 nucleotides

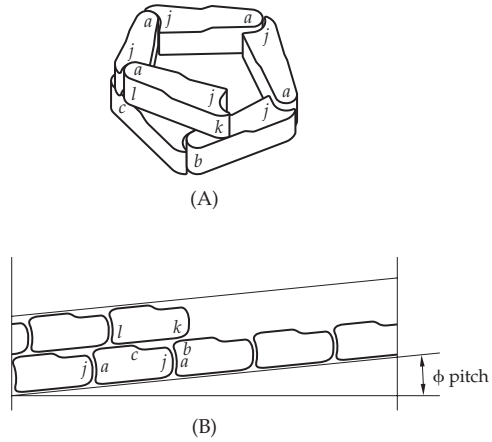



**Figure 7-6** (A) Heterologous bonding of subunits to form a helix. (B) Radial projection of subunits arranged as in helix A. Different bonding regions of the subunit are designated *a*, *b*, *c*, *j*, *k*, and *l*.

which is held in an elongated double-stranded form by a helical sheath of about 2700 subunits of a 50-residue protein. The rod is about 6 nm in diameter and 880 nm long and it is capped by two specialized proteins at one end and a different pair of proteins at the other end. The five coat proteins are encoded by five of the 11 genes present in these little viruses.[37,38]

Each coat subunit in the Ff viruses is coiled into an α-helical rod of 7 nm length. These are arranged in the virus in a right-handed helical pattern with a pitch of 1.5 nm and with 4.4 subunits per turn (Fig. 7-7). The protein rods are inclined to the helix axis and extend inward. This arrangement permits a "knobs-in-holes" hydrophobic bonding between subunits. The helix of pitch 1.5 nm is the **primary** or **one-start helix**. However, in every regular helical structure we can also trace a two-start helix, a three-start helix, etc. In this instance the five-start helix is easiest to see.

The protein coat of these viruses provides an elongated cylindrical cavity to protect the circular, single-stranded DNA molecule that is the genome. Although there are two antiparallel strands of DNA, a regular base-paired structure is impossible and the DNA is probably not present in a highly ordered form.[38a] There are about 2.4 nucleotides in the DNA per protein subunit. However, there are related viruses with ratios as low as one nucleotide per protein subunit and containing more highly extended DNA.[34,39]

***A rod-shaped plant virus.*** The tobacco mosaic virus (Figs. 5-41, 7-8) is a 300-nm-long rod constructed from 2140 identical wedge-shaped subunits whose detailed molecular structure is known.[40] Each 158-residue subunit contains five helices and a small β sheet. A single strand of RNA containing 6395 nucle-

otides (~ 3 per protein subunit) lies coiled in a groove where it interacts with side chains from two of the helices (Fig. 7-8B).[41–44a] The virus is assembled by the binding of a region of the RNA 800 – 1000 nucleotides from the 3' end to a two-turn helix of subunits that appears to form spontaneously. Additional subunits then add at each end, binding to the RNA as well as the adjacent protein subunits.[42,44,45] A relative with a very similar structure is cucumber green mottle mosaic virus.[46]

***Bacterial pili.*** The adhesion pili, or fimbriae,[48] of bacteria are also helical arrays of subunits. The P pili of *E. coli* are encoded by a cluster of 11 genes in the *pap* (pilus associated with pyelonephritis) cluster. They are needed to allow the bacteria to colonize the human urinary tract. The bulk of the ~1-µm-long pilus is made up of about 1000 subunits of a 185-residue protein encoded by gene *PapA*. They form a right-handed helix of ~7 nm diameter with 3.28 subunits per turn and a pitch of ~2.5 nm (Fig. 7-9A).[49–51] The rod is anchored to the bacterial outer membrane by a protein encoded by gene *PapH*, while subunits encoded by *PapE* and *PapF* fasten the adhesin protein (PapG) to the tips of the pili.[49,52,53,53a] The adhesin binds to the Galα1→4Gal ends of glycolipids in

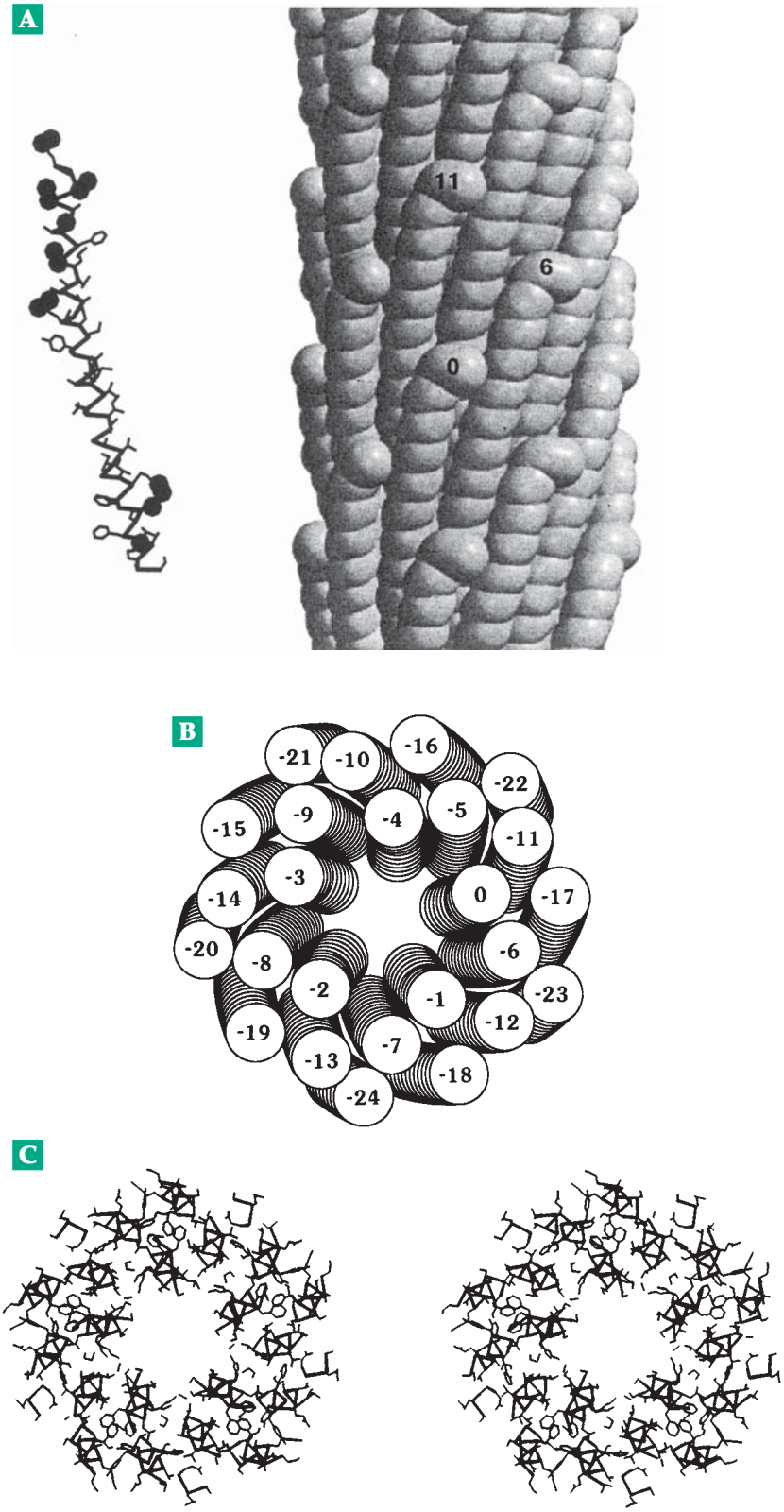



**Figure 7-7** Structure of the virus fd protein sheath. (A) Left. A single coat subunit, with its N terminus towards the top, as if moved from the left side of the sheath. The dark circles represent charged atoms of Asp, Glu, and Lys side chains. The backbone of the protein is a $C^{\alpha}$ diagram. The positively charged atoms near the C terminus line the inner surface of the sheath neutralizing the negative charge of the DNA core. Right. Each subunit is represented by a helical tube through successive $C^{\alpha}$ atoms. Three nearest neighbors, indexed as 0, 6, and 11, are indicated. The axial slab shown represents ~1% of the total length of the virion. From Marvin.[31a] (B) A 2.0 nm section through the virus coat with the helices shown as curved cylinders. The view is down the axis from the N-terminal ends of the rods. The rods extend upward and outward. The rods with indices 0 to –4 start at the same level, forming a five-start helical array. The rods with more negative indices start at lower levels and are therefore further out when they are cut in this section. (C) The same view but with "wire models" of the atomic structure of the rods. From Marvin *et al.*[32]

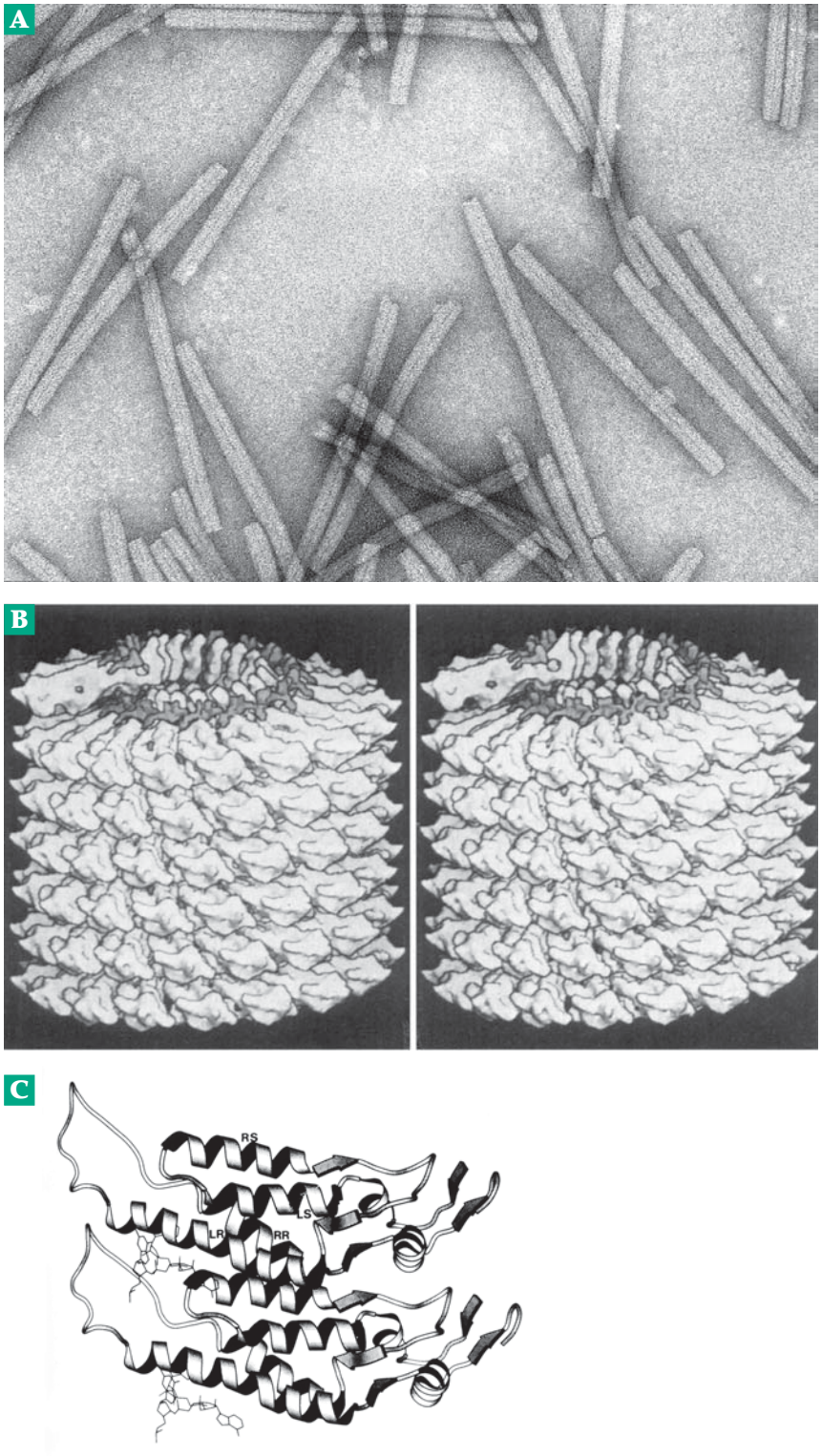



**Figure 7-8** (A) Electron micrograph of the rod-shaped particles of tobacco mosaic virus. © Omikron, Photo Researchers. See also Butler and Klug.[42] (B) A stereoscopic computer graphics image of a segment of the 300 nm long tobacco mosaic virus. The diameter of the rod is 18 nm, the pitch of the helix is 2.3 nm, and there are 16 1\3 subunits per turn. The coat is formed from ~2140 identical 17.5-kDa subunits. The 6395-nucleotide genomic RNA is represented by the dark chain exposed at the top of the segment. The resolution is 0.4 nm. From Namba, Caspar, and Stubbs.[47] (C) A MolScript ribbon drawing of two stacked subunits. From Wang and Stubbs.[46]

the kidney.[51,54] The PapE, F, and G subunits form a thin ~2.5-nm-thick by 15 nm "fibrillum" which is attached by an adapter protein encoded by gene *PapK*. A special chaperonin (*PapD* gene) is also required for pilus assembly in *E. coli*[53a] as well as other bacteria.[55] Another *E. coli* pilus adheres to mannose oligosaccharides.[53b]

Similar pili of *Neisseria gonnorheae* are used by that bacterium. The three-dimensional structure of the 158-residue pilin subunit is that of a globular subunit with an 8.5 nm α-helical spine at one end.[56,57] A proposed model of the intact pilus shaft is shown in Fig. 7-9B,C. Notice the similarity of the packing of the α-helical spines in the center to the packing arrangement in the bacteriophage coat in Fig. 7-7. Similar features may be present in the P pilus rod shown in Fig. 7-9A. However, there is uncertainty about the packing arrangement. The *E. coli* type 1 pilus subunits contain immunoglobulin folds that are completed by donation of an N-terminal strand from a neighboring subunit.[53a] In thin fimbriae of *Salmonella* extended, parallel β helices may be formed (see Fig. 2-17)[53e,53f]

Other types of pili are also well-known.[53c,d] F pili or conjugative pili are essential for sexual transfer of DNA between bacterial cells (Chapter 26). $F^+$ strains of *E. coli* form hollow pili of 8.5 nm diameter with a 2.0-nm central hole.[58,59] Their 90-residue subunits apparently form rotationally symmetric pentamers which stack to form the pili.[59] These pili are essential to establishing the initial contact between conjugating bacterial cells.

***The thin filaments of muscle.*** An essential component of skeletal muscle (discussed further in Chapter 19) is filamentous **actin** (F-actin). It is composed of 375-residue globular subunits of a single type and with a highly conserved sequence.[60,61] It is found not only in muscle but also in other cells where it is a component of the cytoskeleton. The actin microfilament has the geometry of a left-

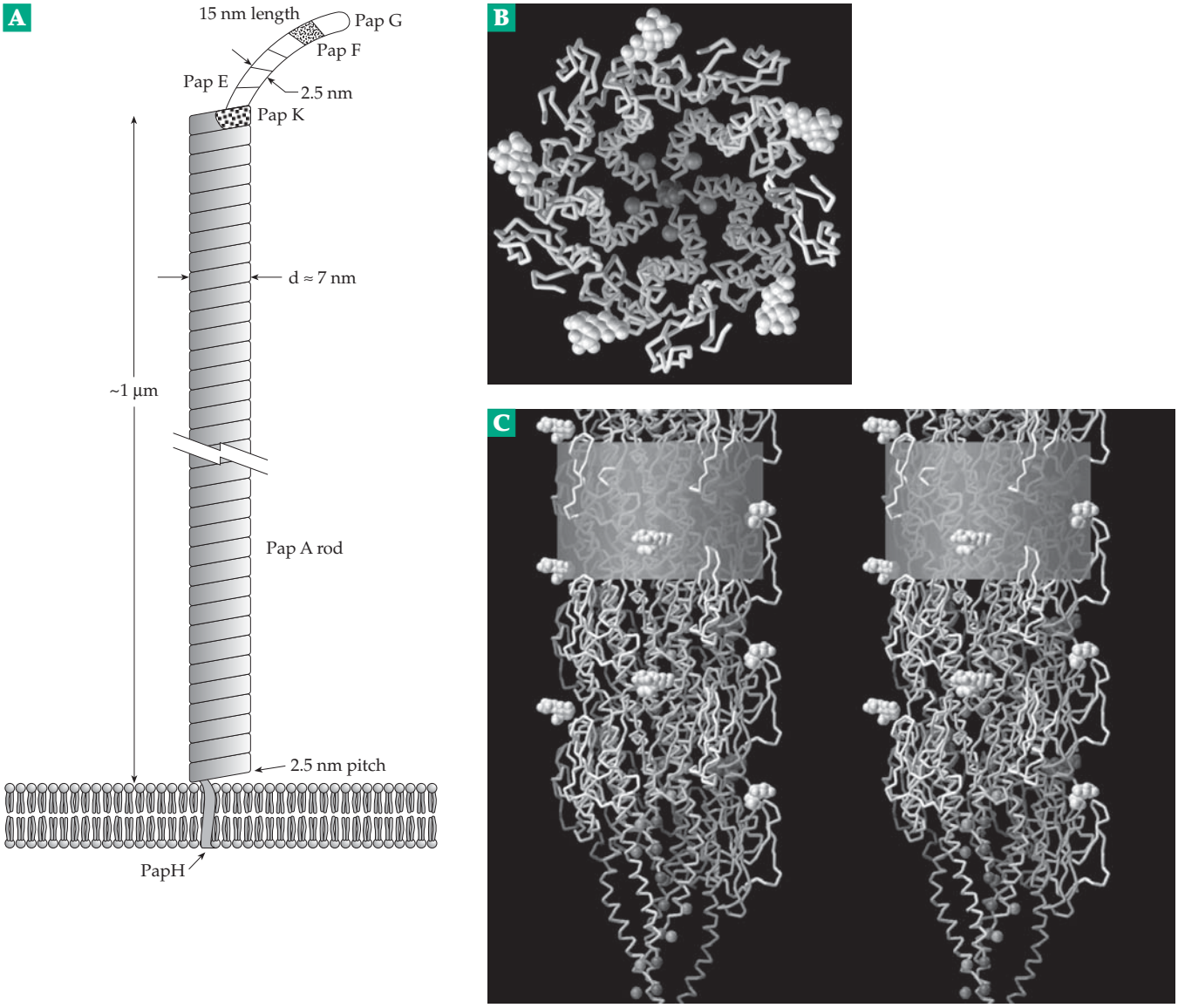



**Figure 7-9** (A) Schematic diagram of a bacterial P pilus. The ~1-µm-long helical rod is anchored to the outer cell membrane by protein Pap H. The adhesion Pap G binds to galactosyl glycolipids of the host. (B, C) The structure of pilus fiber from *Neisseria gonnorrheae* modeled from the atomic structure of the 54-residue pilin subunit. The exact structure of the fiber is uncertain, but the model generated here by trying various possible helical packings matches the dimensions obtained from fiber diffraction patterns and electron microscopic images. (B) Cross section. (C) Stereoscopic view. The experimental dimensions of 4.1-nm pitch and 6.0-nm diameters are shown by the "transparent" ring in (B). From Parge *et al.*[57]

handed one-start or **primary helix** with a pitch of only 0.54 nm and with approximately two subunits per turn (Fig. 7-10).[62,63]
It can also be described as a right-handed two-start helix in which *two chains* of subunits coil around one another with a long pitch (Fig. 7-10).

## 2. Oligomers with Twofold (Dyad) Axes

***Paired interactions.*** If two subunits are held together with interactions *aj* and are related by a twofold axis of rotation as shown in Fig. 7-11, we obtain an **isologous dimer**. Each point such as *a* in one subunit is related to the same point in the other subunit by reflection through the axis of rotation. In the center, along the twofold axis, points *c* and *c′* are *directly opposite the same points* in the other subunit. Figure 7-11

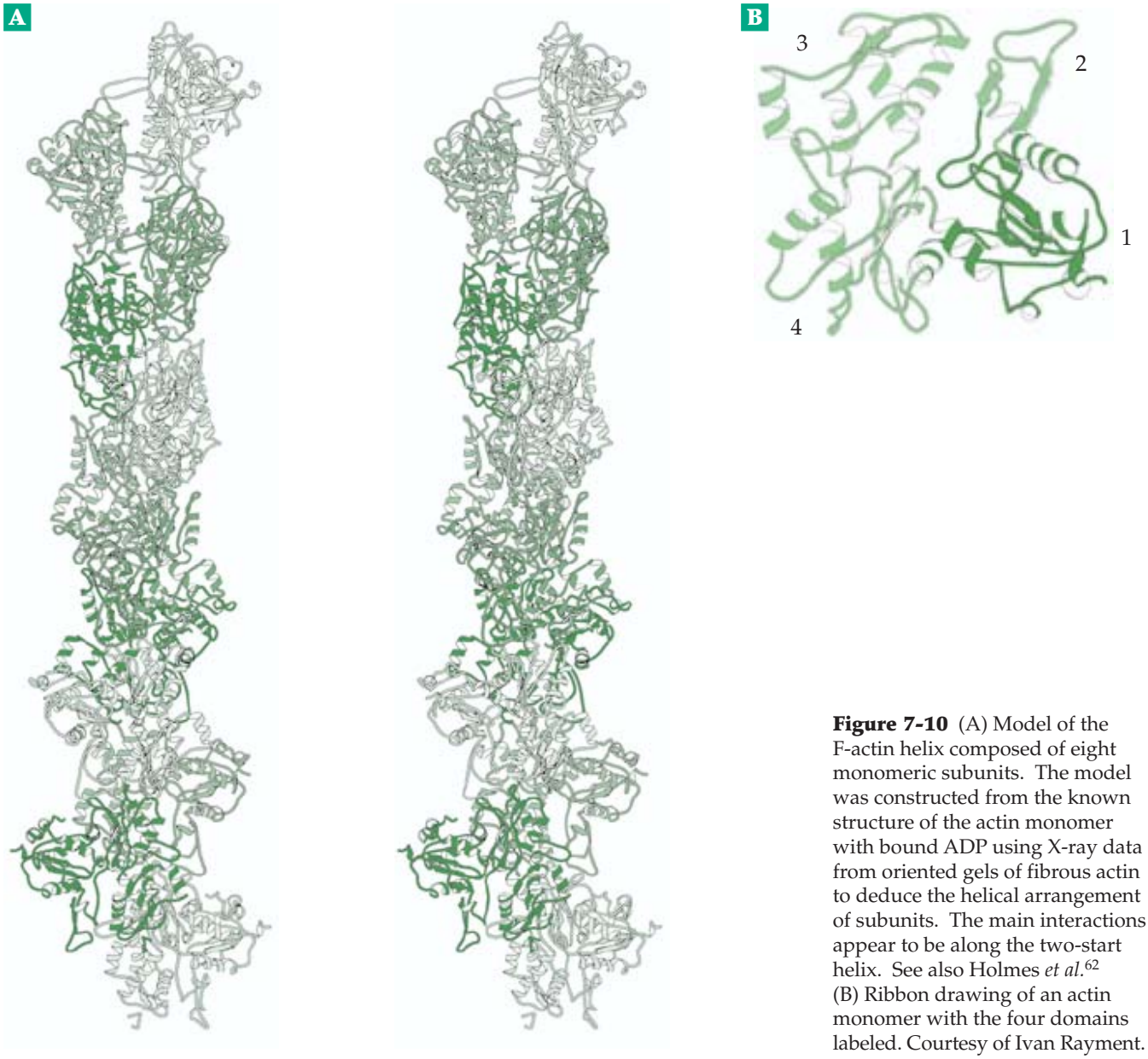



**Figure 7-10** (A) Model of the F-actin helix composed of eight monomeric subunits. The model was constructed from the known structure of the actin monomer with bound ADP using X-ray data from oriented gels of fibrous actin to deduce the helical arrangement of subunits. The main interactions appear to be along the two-start helix. See also Holmes *et al.*[62] (B) Ribbon drawing of an actin monomer with the four domains labeled. Courtesy of Ivan Rayment.

is drawn with a hole in the center so that groups *c* and *c′* do not actually touch, and it is the paired interactions such as *aj* of groups not adjacent to the axis that contribute most to the bonding. However, a real protein dimer may or may not have such a hole. The pair of identical interactions in an isologous dimer may be referred to as a single **isologous bond**. Such a bond always contains the paired interactions between complementary groups (*aj*) and has pairs of identical groups along the axis. However, because they are identical those groups usually cannot interact in a specific complementary manner.

Isologous bonding is very important in oligomeric enzymes, and it has been suggested that isologous interactions evolved early. Initially there may not have been much complementarity in the bonding but two "hydrophobic spots" on the surface of the subunits came together in a nonspecific association.[64] Later in evolution the more specific paired interactions could have been added.

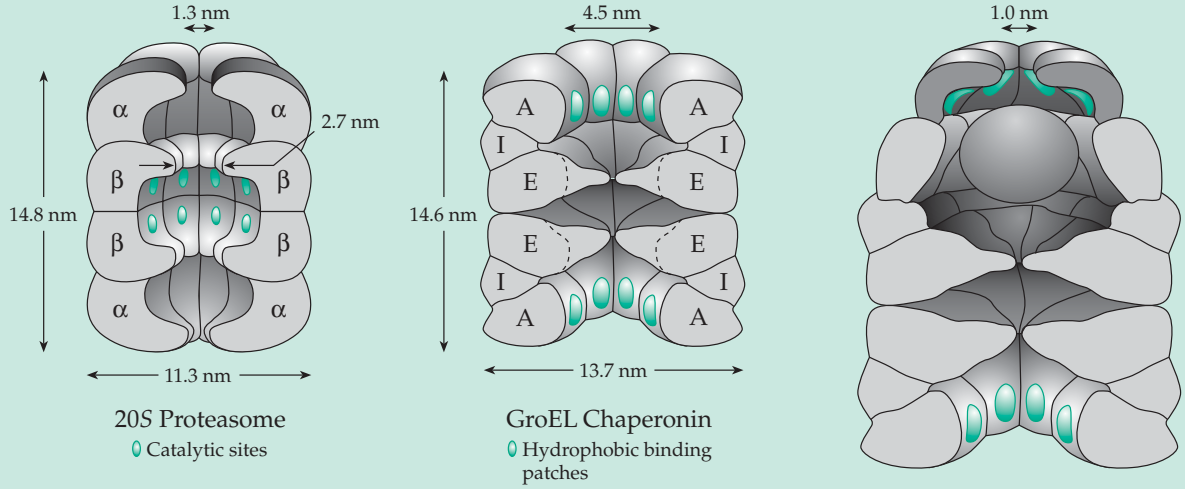
***Dihedral symmetry.*** Isologous dimers can serve as subunits in the formation of larger closed oligomers and helices; for example, an isologous pair of the sort shown in Fig. 7-11A can be flipped over onto the top of another similar pair as shown in Figs. 7-11B and 7-11C. Again, if the proper complementary surfaces exist, bonds can form as shown (*bk* in Fig. 7-11B and *bk* and *cl* in Fig. 7-11C). Both structures in Figs. 7-11B and 7-11C possess dihedral ($D_2$) symmetry.[65] In addition to the twofold axis of rotation lying perpendicular to the two rings, there are two other twofold axes of rotation as

## BOX 7-A LIFE AND DEATH FOR PROTEINS: CHAPERONINS AND PROTEASOMES

In 1968, a tiny cylindrical particle, which appeared to be a stack of 11-nm rings, was observed by electron microscopy of an extract of erythrocytes.[a,b] Later, a similar particle was found in both the nucleus and the cytoplasm of other cells of many organisms. The particles were soon recognized as a new type of protein-hydrolyzing enzyme, a large 700-kDa particle consisting of 20–30 subunits of several different types which came to be known as the **multicatalytic protease** or **20S proteasome**.[c,d] Electron microscopy and X-ray diffraction showed that the particle is formed from four stacked rings, each of which consists of seven subunits whose molecular masses range from 21 to 31 kDa.[e–j]

Proteasomes are strikingly similar in architecture, though not in peptide sequences, to another particle found in both bacteria and eukaryotes: a molecular "chaperone" or **chaperonin**. The chaperonins, of which there are several types, protect proteins while they fold or undergo translocation within cells.[k] One of the best studied members is the *E. coli* protein **GroEL**, which is also composed of double rings of 14 subunits with seven-fold rotational symmetry and with two of these assemblies associated back-to-back with dihedral symmetry.[l] The dimensions of GroEL and 20S proteasomes are nearly the same. However, GroEL has only two rings of ~60-kDa subunits, more than twice the size of proteosomal subunits. The accompanying sketch illustrates this fact and also the basic structural similarity of 20S proteosomes with GroEL. The αβ pairs of the proteosome, correspond to single subunits of the chaperonin, but these subunits have three distinct domains–apical, intermediate, and equatorial (labeled A, I, and E , respectively, in the drawing).[m] After a protein, whether correctly, incorrectly, or only partially folded, enters a cavity in GroEL, a second protein **GroES** of smaller size (~10 kDa) but with seven-fold symmetry binds to one end of the chaperonin.[p] Seven molecules of ATP also bind to sites on the GroEL ring to which GroES binds (the *cis* ring). The binding of the ATP and GroES evidently induces a major conformational change in the GroEL subunits[l,m,n,q] which causes the binding cavity to expand to over twice the original volume. This change (see drawing) also causes hydrophobic surfaces of the cavity to become buried and hydrophilic side chains to be exposed. The cavity surface was initially largely hydrophobic and able to bind many proteins nonspecifically, but upon expansion it becomes hydrophilic and less likely to bind. This releases the encased protein to complete its folding or to partially unfold and refold without interference from other proteins.

While a protein is adjusting its folding in the *cis* compartment another protein molecule may become trapped in the *trans* compartment. After some time the bound ATP molecules are hydrolyzed. As in the contraction of muscle, which is discussed in Chapter 19, the loss of inorganic phosphate ($P_i$) and ADP from the active site can be accompanied by movement. In the chapenonin this involves a conformational switch so that the ES heptamer is released and the conformation of the *trans* ring of EL is switched to that of the initial *cis* ring and vice versa. The new *cis* ring is ready to receive an ES cap and the new *trans* ring can release the folded protein.[r] A variety of experimental approaches are being used in an




From Weissman, Sigler and Horwich[g]. Illustrations by K. Sutliff.
Right: From Mayhew and Hartl[o]

Expanded GroEL-GroES complex with enclosed folding polypeptide[o]

## BOX 7-A LIFE AND DEATH FOR PROTEINS: CHAPERONINS AND PROTEASOMES (continued)

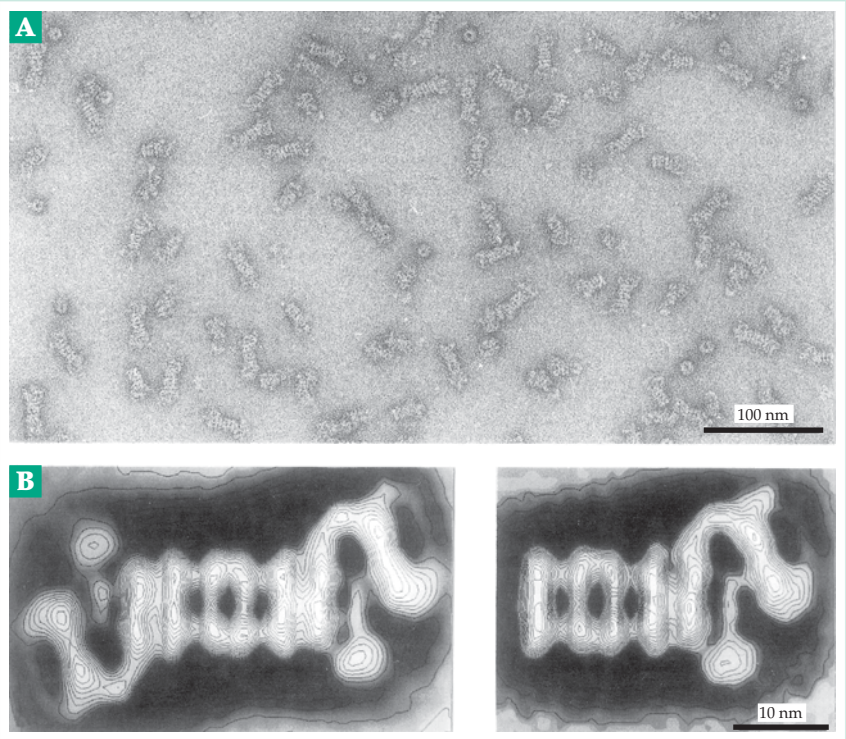
effort to further understand the action of GroEL.[s–x] The chaperonin may function repeatedly before a protein becomes properly folded.[t]

While chaperonins assist proteins to fold correctly proteasomes destroy unfolded chains by partial hydrolysis, cutting the chains into a random assortment of pieces from 3 to 30 residues in length with an average length of ~ 8 residues.[y] Proteasomes destroy not only unfolded and improperly folded proteins but also proteins marked for destruction by the ubiquitin system described in Box 10-C. It has been hard to locate true proteosomes in most bacteria. However, they do contain protease particles with similar characteristics[z–bb] and archaeons, such as *Thermoplasma acidophilum,* have proteasomes similar to those of eukaryotes.[cc]

The *Thermoplasma* proteasome contains only two kinds of subunits, α and β, which have similar amino acid sequences. These form $\alpha_7$ and $\beta_7$ rings which associate in $\alpha_7\beta_7$ pairs with two of these double rings stacked back-to-back with dihedral D7 symmetry: $\alpha_7\beta_7\beta_7\alpha_7$. The crystal structure has been determined for this 20S proteasome from *T. acidophilum*[g,dd] and for the corresponding proteasome from yeast (*Saccharomyces cerevisiae*).[f,ee] The accompanying drawings illustrate top and side views of the *T. acidophilum* proteasome. The particle contains three internal cavities. The outer two are formed between the $\alpha_7$ and $\beta_7$ rings and the inner is formed between the two $\beta_7$ rings. A channel only 1.3 nm in diameter permits the entrance of peptide chains into the compartments.

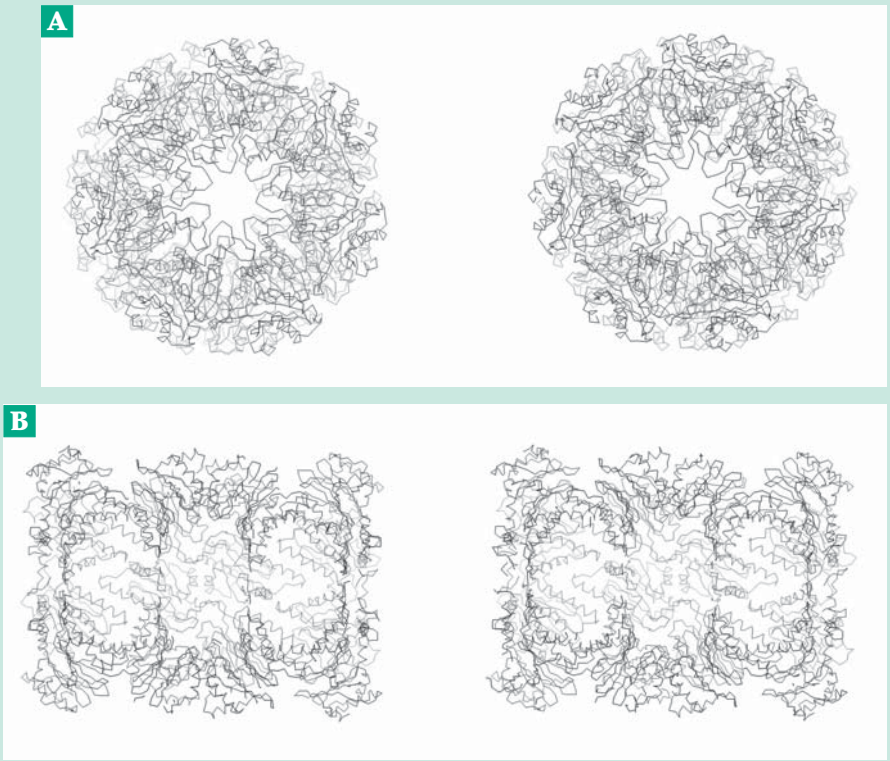
The active sites of the enzymes[ff] are located in the β subunits in the central cavity.[dd] While the yeast and human proteasomes are similar to those of *Thermoplasma,* the β subunits consist of seven different protein-hydrolyzing enzymes whose catalytic activities and mechanisms are considered in Chapter 12. There are also seven different α subunits, all of whose sequences are known.[i,gg] To make the story more complex, additional subunits, some of which catalyze ATP hydrolysis, form a 600- to 700-kDa cap which adds to one or both ends of a 20S proteosome to give a larger **26S proteasome**.[b,d,e,w] These larger proteasomes carry out an ATP-dependent cleavage of proteins selected for degradation by the ubiquitin system (Box 10-C; Chapter 12). Some of the short peptide segments formed by proteasomes may leave cells and participate in intercellular communication. For example, pieces of antigenic peptides are used by cells of the immune system for "antigen presentation" (Chapter 31),[hh] an important process by which the immune system recognizes which cells are "self" and which are foreign or malignant and must be killed.

The structure of the caps on the 26S proteasome ends is complex. At least 20 different regulatory subunits have been identified.[ii,jj]




(A) Electron micrograph of 26S proteasomes from *Xenopus* oocytes negatively stained with 2% uranyl acetate. (B) Image of the 26S proteasome (left) and a 20S proteasome with only one end cap. These views were obtained by correlation averaging of 527 individual images of the 26S proteasome and 395 images of the single-ended form. From Peters *et al.*[e] Courtesy of Wolfgang Baumeister.

## BOX 7-A (continued)




(A) Top view of the 20S proteasome as an α-carbon plot showing the seven-fold symmetry. The α subunits are in front of the β subunits. (B) Side view showing the proteasome cut open along its seven-fold axis. From Löwe *et al.*[dd] Courtesy of Robert Huber.


[a] Harris, J. R. (1968) *Biochim. Biophys. Acta.* **150**, 534–537
[b] Peters, J.-M. (1994) *Trends Biochem. Sci.* **19**, 377–382
[c] Bosch, G., Baumeister, W., and Essen, L.-O. (2000) *J. Mol. Biol.* **301**, 19–25
[d] Ferrell, K., Wilkinson, C. R. M., Dubiel, W., and Gordon, C. (2000) *Trends Biochem. Sci.* **25**, 83–88
[e] Peters, J.-M., Cejka, Z., Harris, J. R., Kleinschmidt, J. A., and Baumeister, W. (1993) *J. Mol. Biol.* **234**, 932–937
[f] Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H. D., and Huber, R. (1997) *Nature (London)* **386**, 463–471
[g] Weissman, J. S., Sigler, P. B., and Horwich, A. L. (1995) *Science* **268**, 523–524
[h] DeMartino, G. N., and Slaughter, C. A. (1999) *J. Biol. Chem.* **274**, 22123–22126
[i] Gerards, W. L. H., de Jong, W. W., Bloemendal, H., and Boelens, W. (1998) *J. Mol. Biol.* **275**, 113–121
[j] Schmidtke, G., Schmidt, M., and Kloetzel, P.-M. (1997) *J. Mol. Biol.* **268**, 95–106
[k] Netzer, W. J., and Hartl, F. U. (1998) *Trends Biochem. Sci.* **23**, 68–73
[l] Sigler, P. B., Xu, Z., Rye, H. S., Burston, S. G., Fenton, W. A., and Horwich, A. L. (1998) *Ann. Rev. Biochem.* **67**, 581–608
[m] Kawata, Y., Kawagoe, M., Hongo, K., Miyazaki, T., Higurashi, T., Mizobata, T., and Nagai, J. (1999) *Biochemistry* **38**, 15731–15740
[n] Betancourt, M. R., and Thirumalai, D. (1999) *J. Mol. Biol.* **287**, 627–644
[o] Mayhew, M., and Hartl, F. U. (1996) *Science* **271**, 161–162
[p] Hunt, J. F., Weaver, A. J., Landry, S. J., Gierasch, L., and Deisenhofer, J. (1996) *Nature (London)* **379**, 37–45
[q] Cliff, M. J., Kad, N. M., Hay, N., Lund, P. A., Webb, M. R., Burston, S. G., and Clarke, A. R. (1999) *J. Mol. Biol.* **293**, 667–684
[r] Rye, H. S., Burston, S. G., Fenton, W. A., Beechem, J. M., Xu, Z., Sigler, P. B., and Horwich, A. L. (1997) *Nature (London)* **388**, 792–798
[s] Chatellier, J., Buckle, A. M., and Fersht, A. R. (1999) *J. Mol. Biol.* **292**, 163–172
[t] Nieba-Axmann, S. E., Ottiger, M., Wüthrich, K., and Plückthun, A. (1997) *J. Mol. Biol.* **271**, 803–818
[u] Gervasoni, P., Gehrig, P., and Plückthun, A. (1998) *J. Mol. Biol.* **275**, 663–675
[v] Torella, C., Mattingly, J. R., Jr., Artigues, A., Iriarte, A., and Martinez-Carrion, M. (1998) *J. Biol. Chem.* **273**, 3915–3925
[w] Horwich, A. L., Weber-Ban, E. U., and Finley, D. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11033–11040
[x] Buckle, A. M., Zahn, R., and Fersht, A. R. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3571–3575
[y] Kisselev, A. F., Akopian, T. N., and Goldberg, A. L. (1998) *J. Biol. Chem.* **273**, 1982–1989
[z] Kessel, M., Maurizi, M. R., Kim, B., Kocsis, E., Trus, B. L., Singh, S. K., and Steven, A. C. (1995) *J. Mol. Biol.* **250**, 587–594
[aa] Shin, D. H., Lee, C. S., Chung, C. H., and Suh, S. W. (1996) *J. Mol. Biol.* **262**, 71–76
[bb] Tamura, T., Tamura, N., Cejka, Z., Hegerl, R., Lottspeich, F., and Baumeister, W. (1996) *Science* **274**, 1385–1389
[cc] Maupin-Furlow, J. A., and Ferry, J. G. (1995) *J. Biol. Chem.* **270**, 28617–28622
[dd] Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995) *Science* **268**, 533–539
[ee] Stuart, D. I., and Jones, E. Y. (1997) *Nature (London)* **386**, 437–438
[ff] Voges, D., Zwickl, P., and Baumeister, W. (1999) *Ann. Rev. Biochem.* **68**, 1015–1068
[gg] Schmidt, M., and Kloetzel, P.-M. (1997) *FASEB J.* **11**, 1235–1243
[hh] Goldberg, A. L., and Rock, K. L. (1992) *Nature (London)* **357**, 375–379
[ii] Adams, G. M., Falke, S., Goldberg, A. L., Slaughter, C. A., DeMartino, G. N., and Gogol, E. P. (1997) *J. Mol. Biol.* **273**, 646–657
[jj] Knowlton, J. R., Johnston, S. C., Whitby, F. G., Realini, C., Zhang, Z., Rechsteiner, M., and Hill, C. P. (1997) *Nature (London)* **390**, 639–643

indicated in the drawings. Again, the interactions are paired and isologous; of many possible contacts two *bk* interactions and two *cl* interactions are marked for each pair of subunits in Fig. 7-11. There are a total of six pairs of these interactions, one between each combination of two subunits. This may be a little more difficult to see in Fig. 7-11B than in Fig. 7-11C because in the former the subunits are arranged in a more or less square configuration. Nevertheless, a pair of interactions between the left-hand subunit in the top ring and the subunit in the lower ring at the right does exist, even if it is only electrostatic and at a distance. An example of a tetrameric enzyme with perfect dihedral symmetry of the type shown in Fig. 7-11B is **lactate dehydrogenase** (Chapter 15). The plant agglutinin **concanavalin A** has a quaternary structure resembling that in Fig. 7-11C.

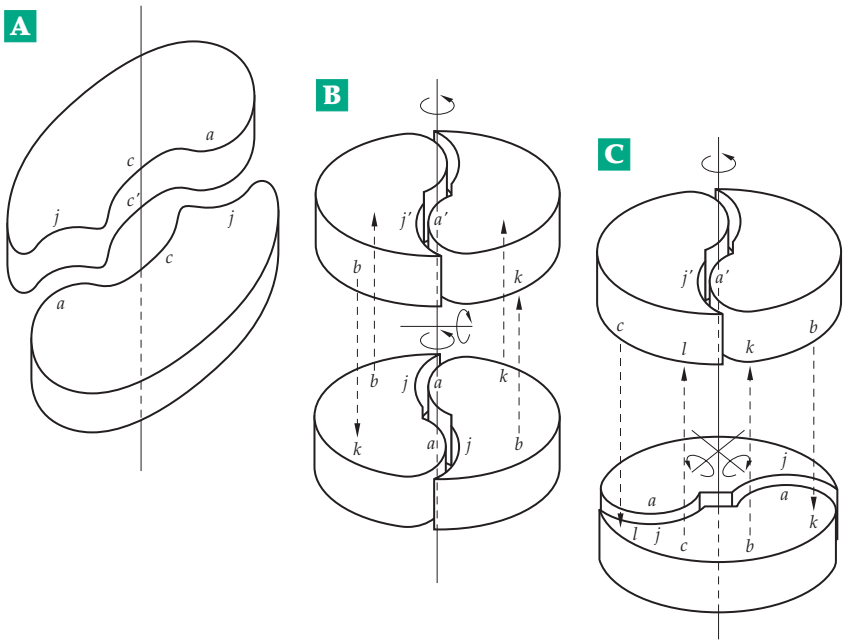



**Figure 7-11** (A) Isologous bonding between pairs of subunits; (B) an "isologous square" arrangement of subunits; (C) an apparently "tetrahedral" arrangement of subunits. Note the three twofold axes.

Square arrays of four subunits can be formed using either heterologous or isologous interactions. Both types of bonding can occur in larger aggregates. For example, two trimers such as that shown in Eq. 7-25 can associate to a hexamer having dihedral ($D_3$) symmetry; a heterologous "square tetramer" can dimerize to give a dihedral ($D_4$) octamer.[65] The enzymes **ornithine decarboxylase** (Fig. 7-12)[66] and **glutamine synthetase** (Chapter 24)[67] each consist of double rings of six subunits each. The upper ring is flipped over onto the lower giving dihedral symmetry ($D_6$) with one 6-fold axis and six 2-fold axes at right angles to it.

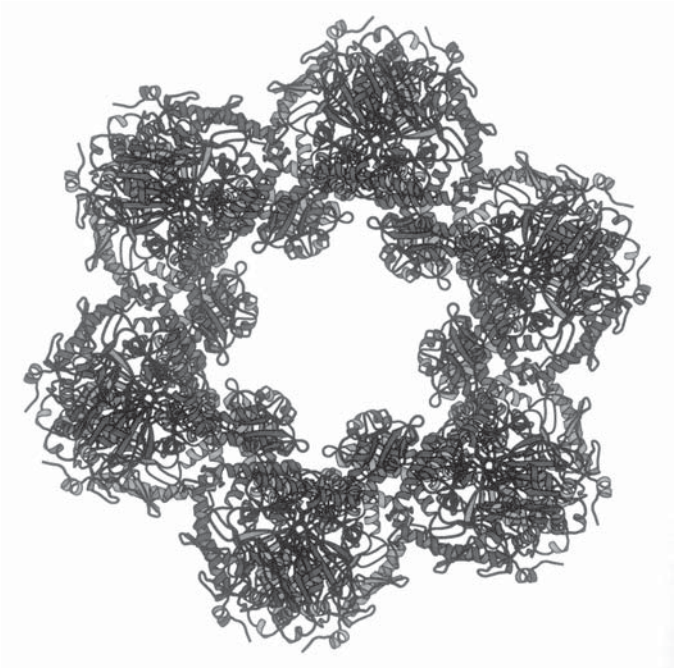


**Figure 7-12** A ribbon representation of the ornithine decarboxylase dodecamer. Six dimers of the 730-residue subunits are related by $C_6$ crystallographic symmetry. MolScript drawing from Momany *et al.*[66] Courtesy of Marvin Hackert.

### *Oligomers with cubic symmetry (polyhedra).*

Symmetrical arrangements containing more than one axis of rotation of order higher than 2-fold are said to have cubic symmetry. The **tetrahedron** is the simplest example. It contains four 3-fold axes which pass through the vertices and the centers of the faces and three 2-fold axes which pass through the midpoints of the six edges. *Since protein subunits are always asymmetric, a tetrameric protein cannot possess cubic symmetry*. As we have already seen, tetrameric enzymes have dihedral symmetry. However, a heterologous trimer with 3-fold symmetry can form a face of a tetrahedron containing a total of 12 asymmetric subunits. Twenty-four subunits can interact to form a **cube**. Three 4-fold axes pass through the centers of the faces, four 3-fold axes pass through the vertices, and six 2-fold axes pass through the edges (see Figs. 7-13 and 16-3).

The largest structure of cubic symmetry that can

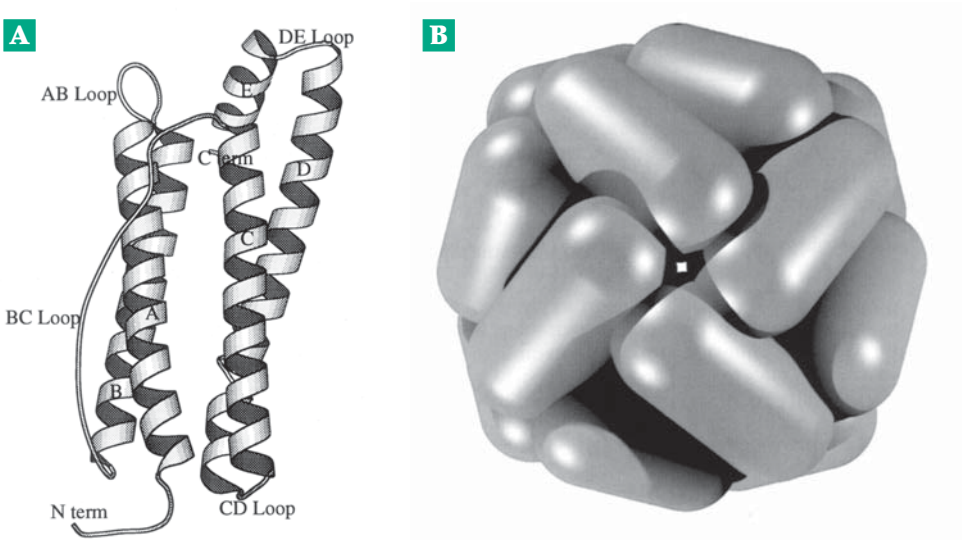



**Figure 7-13** (A) MolScript ribbon drawing of a subunit of the iron oxide storage proteins L-ferritin from amphibian red cells. This 4-helix bundle is represented by cylinders of 1.3 nm diameter in the oligomer. (B) Helices A and C of the monomer are on the outer surfaces of the oligomer and helices B and D are on the inner surface. The oligomer consists of a shell of 24 subunits and is viewed down a 4-fold axis illustrating its 423 (cubic) symmetry. The molecule is illustrated further in Fig. 16-3. From Trikha *et al.*[74] Courtesy of Elizabeth Theil.

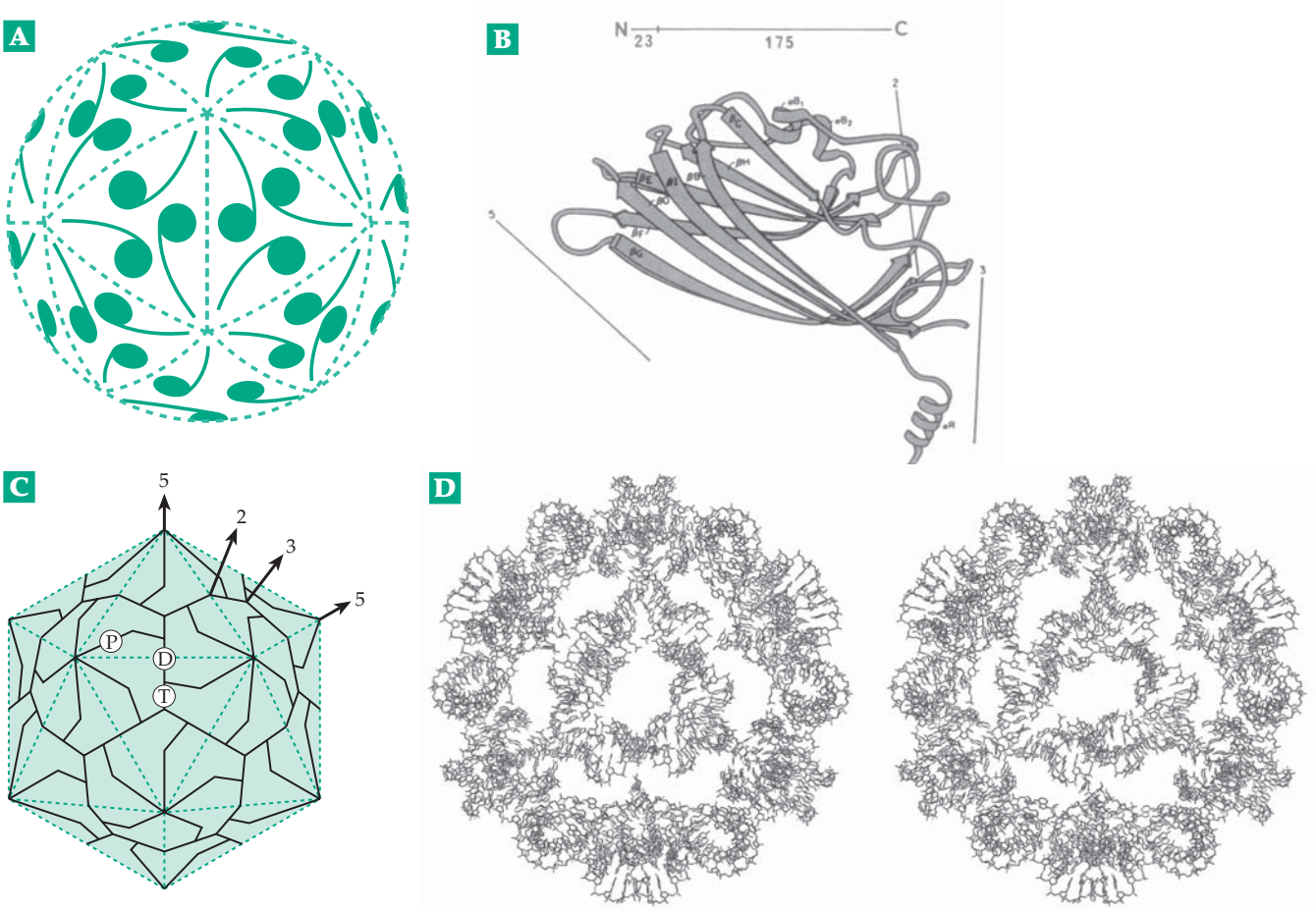



**Figure 7-14** (A) Schematic drawing illustrating an icosahedrally symmetric structure with sixty identical asymmetric subunits all in equivalent positions. The 5-fold axes are located at the vertices of the icosohedron and the 2-fold and 3-fold axes can readily be seen. (B) Ribbon drawing of the 195-residue polypeptide chain of the coat subunit of satellite tobacco necrosis virus. The protein folds into an inwardly projecting N-terminal segment and a "β-jellyroll" domain. The packing of this subunit in the virus particle is shown schematically in (C). The symmetry axes drawn next to the subunit diagram (B) can be used to position it in the structure. Contacts between subunits are labeled D, T, and P ('dimer', 'trimer', 'pentamer'). Diagrams courtesy of Drs. Strandberg, Liljas, and Harrison.[68] (D) The distribution of RNA helical segments in a hemisphere of a virion of a similar small virus, the satellite tobacco mosaic virus. The virion viewed down a 3-fold axis from the virus exterior. The helical axes of the RNA segments are along icosahedral edges. From Larson *et al.*[75]

be made is the **icosahedron**, a regular solid with 20 triangular faces. Sixty subunits, or some multiple of 60, are required and at each vertex they form a heterologous pentamer. As with the tetrahedron, each face contains a heterologous trimer, while isologous bonds across the edges form dimers (Fig. 7-14C). Many viruses consist of roughly spherical protein shells (coats) containing DNA or RNA inside.[68–70] As with the filamentous viruses, the protein coats consist of many identical subunits, a fact that can be rationalized in terms of economy from the genetic viewpoint. *Only one gene is needed to specify the structure of a large number of subunits.*[70,71] Under the electron microscope the viruses often have an icosahedral appearance (Figs. 5-41A, 7-14), and chemical studies show that the number of the most abundant subunits is usually a multiple of 60. An example is the tiny **satellite tobacco necrosis virus**,[72] diameter ~18 nm, whose coat contains just 60 subunits of a 195-residue protein. Its genome is a 1239-nucleotide molecule of RNA. The structure of the coat has been determined to 0.25-nm resolution.[73]

Many virus coats have 180 subunits or a number that is some other multiple of 60. However, in these coats the subunits cannot all be in identical environments. Two cases may be distinguished. If all of the subunits have identical amino acid sequences they probably exist in more than one distinct conformation that permit them to pack efficiently. (Next section) Alternatively, two or more subunits of differing sequence and structure may associate to form 60 larger subunits that do pack with icosahedral symmetry. For example, the polioviruses (diameter 25 nm) contain three major coat proteins (α, β, and γ or VP1, VP2, and VP3). These are formed by cleavage of a large precursor protein into at least four pieces.[76,77] The three largest pieces of ~33-, 30-, and 25-kDa mass (306, 272, and 238 residues, respectively) aggregate as $(\alpha\beta\gamma)_{60}$. Sixty copies of a fourth subunit of 60 residues are found within the shell. Related picorna viruses such as human rhinoviruses (Fig. 7-15),[69,78,79] foot-and-mouth disease virus, parvovirus,[80] and Mengo virus[81] have similar architectures. The small (diameter 25 nm) single-stranded DNA bacteriophages such as ϕX174 also have three different coat proteins, one of which forms small hollow spikes at the vertices of the icosahedral shell (Fig. 5-41A).[82]

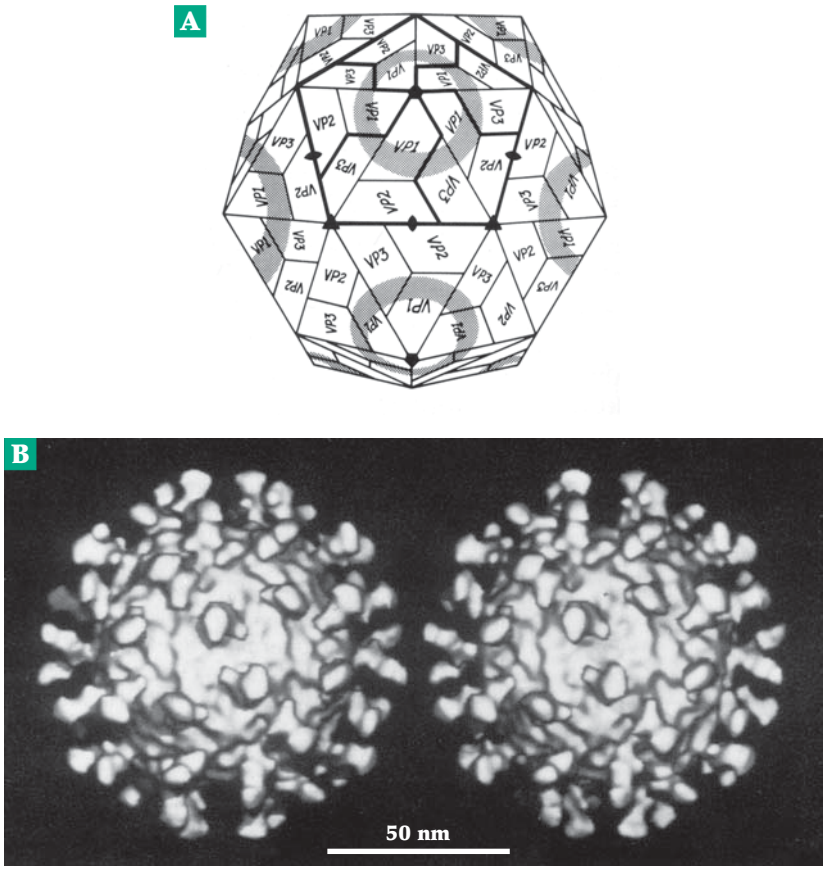



**Figure 7-15** (A) Schematic diagram of the icosahedral shell of a human rhinovirus showing the arrangement of the three subunits VP1, VP2, and VP3, each present as 60 copies. (B) Stereoscopic view of an image of the virus "decorated" by the binding of two immunoglobulinlike domains of the intercellular adhesion molecule ICAM-1, a natural receptor for the virus. Part of this receptor binds into a groove or "canyon," which in marked in (A) by the dark bands. From Olson *et al.*[78] Courtesy of Michael Rossmann.

***Asymmetry and quasi-equivalence in oligomers.*** It is natural to think about association of subunits in symmetric ways. Consequently the observation of square, pentagonal, and hexagonal arrangements of subunits directly with the electron microscope led to a ready acceptance of the idea that protomers tend to associate symmetrically. However, consider the predicament of the two molecules shown in Fig. 7-16. They might get together to form an isologous dimer if it were not for the fact that their "noses" are in the way. Despite the obvious steric hindrance, an isologous dimer can be formed in this case if one subunit is able to undergo a small change in conformation (Fig. 7-16). In the resulting dimer the two subunits are only **quasi-equivalent**.

Unsymmetrical dimerization of proteins appears to be a common phenomenon that is often observed in protein crystals. For example, the

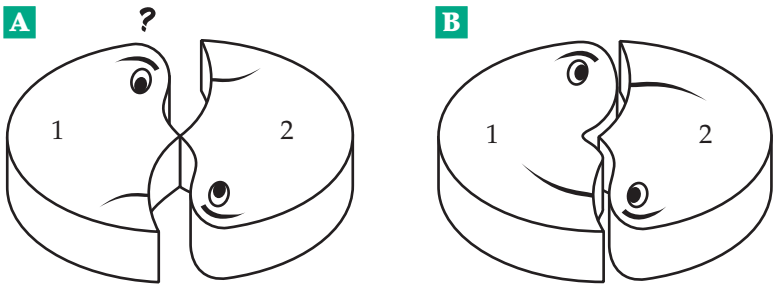



**Figure 7-16** Nonsymmetric bonding in a dimer. (A) Two molecules which cannot dimerize because of a bad fit at the center. (B) A solution: Molecule 1 has refolded its peptide chain a little, changing shape enough to fit to molecule 2.

enzymes malic dehydrogenase and glyceraldehyde phosphate dehydrogenase (Chapter 15) are both tetramers of approximate dihedral symmetry but X-ray crystallography revealed distinct asymmetries[83,84] which include a weaker binding of the coenzyme $NAD^+$ in one subunit. This may simply reflect differences in environment within the crystal lattice. However, negative cooperativity in coenzyme binds has also been revealed by kinetic experiments.[85]

The polypeptide hormone **insulin** is a small protein made up of two chains (designated A and B) which are held together by disulfide bridges (Fig. 7-17A). Figure 7-17B is a sketch of the structure as revealed by X-ray crystallography,[86,87] with only the backbone of the peptide chains and a few side chains shown. In the drawing, the B chain lies behind the A chain. Beginning with the N-terminal Phe 1 of the B chain the peptide backbone makes a broad curve, and then falls into an helix of three turns lying more or less in the center of the molecule. After a sharp turn, it continues upward on the left side of the drawing in a nearly completely extended β structure. The A chain has an overall U shape with two roughly helical portions. The U shape is partly maintained by a disulfide bridge running between two parts of the A chain. Two disulfide bridges hold the A and B chains together, and hydrophobic bonding of internal side chain groups helps to stabilize the molecule.

Insulin in solution dimerizes readily, the subunits occupying quasi-equivalent positions. Figure 7-17C shows some details of the bonding between the subunits in the insulin dimer with a view from the outside of the molecule down the 2-fold axis (marked by the X in the center of the Phe 25 ring in the right-hand chain) through the dimer. The C-terminal ends of the B chains are seen in an extended conformation. The two antiparallel chains form a β structure with two pairs of hydrogen bonds. If there were perfect isologous bonding, the two pairs would be entirely equivalent and symmetrically related one to the other. A straight line drawn from a position in one chain and passing through the twofold axis (X) would also pass approximately through the corresponding position in the other chain. However, there are many deviations from perfect symmetry, the most striking of which is at the center where the Phe 25 from the right-hand chain projects upward and to the left. If the symmetry were perfect the corresponding side chain from the left-hand chain would project upward and to the right and the two phenylalanines would collide, exactly as do the "noses" in Fig. 7-16. In insulin one phenylalanine side chain has been flipped back out of the way.

Under proper conditions, three insulin dimers associate to form a hexamer of approximate dihedral ($D_3$) symmetry that is stabilized by the presence of two zinc ions. Figure 7-17D is a crude sketch of the hexamer showing the three dimers, the 3-fold axis of symmetry, and the two pseudo 2-fold axes, one passing between the two subunits of the dimer and the other between two adjacent dimers. Figure 7-18 is a stereoscopic ribbon diagram of the atomic structure, with the A chains omitted, as obtained by X-ray diffraction.[87] The structure has also been obtained by NMR spectroscopy.[88] Note that each of the two zinc atoms lies on the threefold axis and is bound by three imidazole rings from histidines B-10. The significance of the zinc binding is uncertain but these hexamers readily form rhombohedral crystals, even within the pancreatic cells that synthesize insulin. The structure illustrates a feature that is common to many oligomers of circular or dihedral symmetry. A central "channel" is often quite open and protruding side chain groups, such as the imidazole groups in insulin, form handy nests into which ions or molecules regulating activity of proteins can fit. Conformational differences in insulin are induced by the binding of phenol. In Fig. 7-18A the C-terminal ends of the chains are extended but in the phenol complex (B) they have coiled to extend the α helices.

***Quasi-equivalence in virus coats.*** A large number of icosahedral viruses have coats consisting of 180 identical subunits. For example, the small RNA-containing bacteriophage MS2 consists of an eicosahedral shell of 180 copies of a 129-residue protein that encloses one molecule of a 3569-residue RNA.[89] The virus also contains a single molecule of a 44-kDa protein, the A protein, which binds the virus to a bacterial pilus to initiate infection. Related bacteriophages GA, fr, f2, and Qβ[90,91] have a similar

architecture. Many RNA-containing viruses of plants also have 180 subunits in their coats.[68] Much studied are the **tomato bushy stunt virus** (diameter ~33 nm, 40-kDa subunits),[68] and the related **southern bean mosaic virus**.[92] The human **wart virus** (diameter ~ 56 nm) contains 420 subunits, seven times the number in a regular icosahedron. **Adenoviruses** (diameter ~100 nm) have 1500 subunits, 25 times more than the 60 in a regular icosahedron.[93,94] Caspar and Klug[95] proposed a theory of quasi-equivalence of subunits according to which the distances between the centers of subunits are preserved in a family of **icosadeltahedra** containing subunits in multiples of 20. However, the angles must vary somewhat from those in a regular icosahedron (compare with geodesic shells in which the angles are constant but the distances are not all the same). The resulting polyhedra contain hexamers as well as pentamers at vertices; for example, the shells of the 180-subunit viruses contain clusters of subunits forming 12 pentamers and 20 hexamers. There are

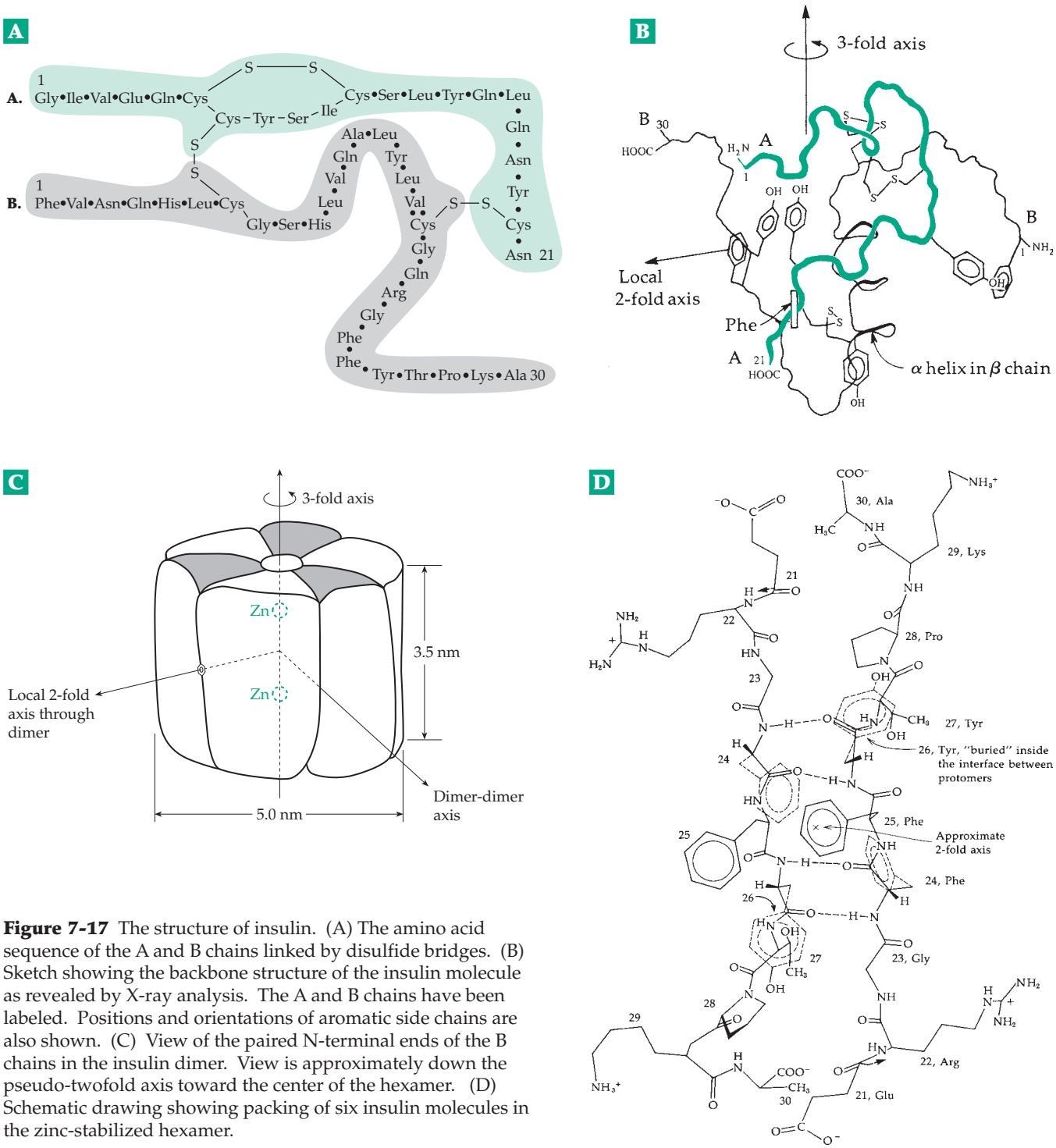



**Figure 7-17** The structure of insulin. (A) The amino acid sequence of the A and B chains linked by disulfide bridges. (B) Sketch showing the backbone structure of the insulin molecule as revealed by X-ray analysis. The A and B chains have been labeled. Positions and orientations of aromatic side chains are also shown. (C) View of the paired N-terminal ends of the B chains in the insulin dimer. View is approximately down the pseudo-twofold axis toward the center of the hexamer. (D) Schematic drawing showing packing of six insulin molecules in the zinc-stabilized hexamer.

also 60 trimers (on the faces) and 90 dimers (across the edges) (Fig. 7-19). Such structures can be formed only for certain values of *T* where the number of subunits is 60*T* and there are 12 pentamers (pentons) and 10 (*T*-1) hexamers hexons. *T* can assume values of $h^2 + hk + k^2$ where h and k may be positive integers or zero. Some allowed *T* values are 1, 3, 4, 7, 9, 13, 25.[68,70,96,97]

The subunits in virus coats with *T* greater than one are not all in equivalent positions. For example, the three subunits labeled A, B, and C in Fig. 7-19 are each slightly differently positioned with respect to neighboring subunits. Since virus coats are usually tightly packed the subunits must assume more than one conformation. One kind of conformational change that allows quasi-equivalence of subunits is observed in the tomato bushy stunt virus. Two structural domains are connected by a hinge which allows an outer protruding domain to move slightly to preserve good isologous interactions with a corresponding domain in another subunit.[68]

The southern bean mosaic virus has an eight-stranded antiparallel β-barrel structure closely similar to that of the major domain of the bushy stunt viruses but lacking the second hinged domain. The problem of quasi-equivalence is resolved by the presence of an N-terminal extension that binds onto a subunit across the quasi-six-fold axis to give a set of three subunits (labeled C in Fig. 7-19) that associate with true three-fold symmetry and another set (B) with a slightly different conformation fitting between them.[68,92] The subunits A, which have a third conformation, fit together around the five-fold axis in true cyclic symmetry.

A surprising finding is that the polyoma virus coat, which was expected to contain 420 (7 x 60) subunits, apparently contains only 360. The result is that the hexavalent morphological unit is a pentamer and that quasi-eqivalence appears

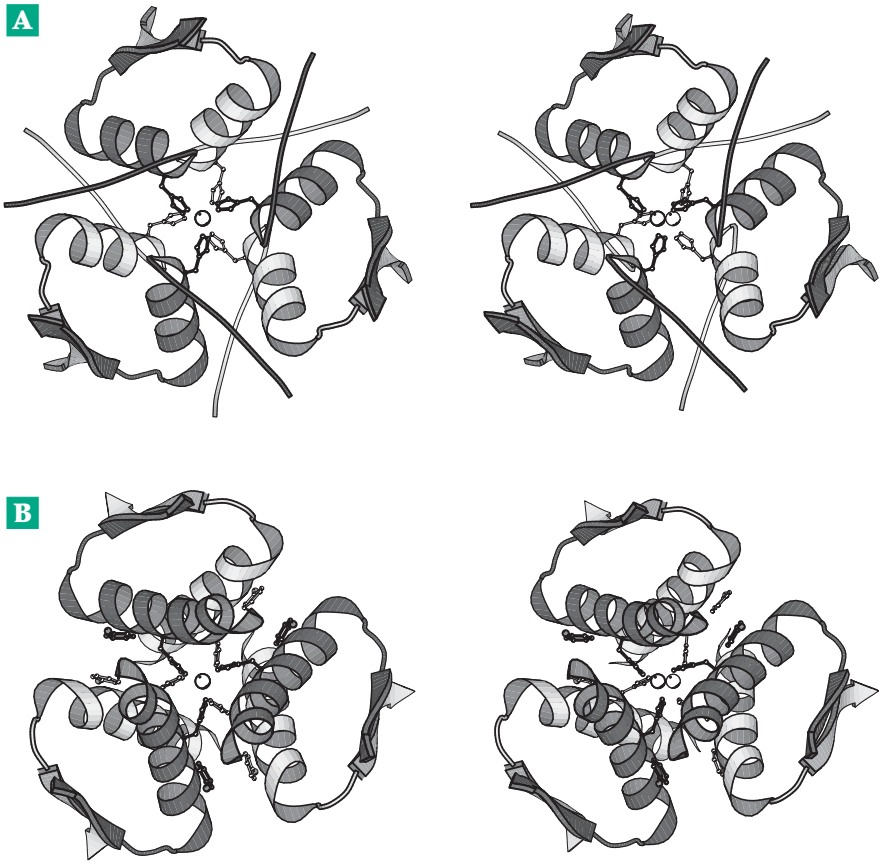



**Figure 7-18** Stereoscopic MolScript ribbon drawings of the B chains (A chains omitted) of (A) hexameric 2-zinc pig insulin. (B) A phenol complex of the same protein. Within each dimer the B chains are shaded differently. The $Zn^{2+}$ ions are represented by white spheres and the coordinating histidine side chains are shown. Six noncovalently bound phenol molecules can be seen, as can several conformational differences. From Whittingham *et al*.[87] Courtesy of Peter C. E. Moody.

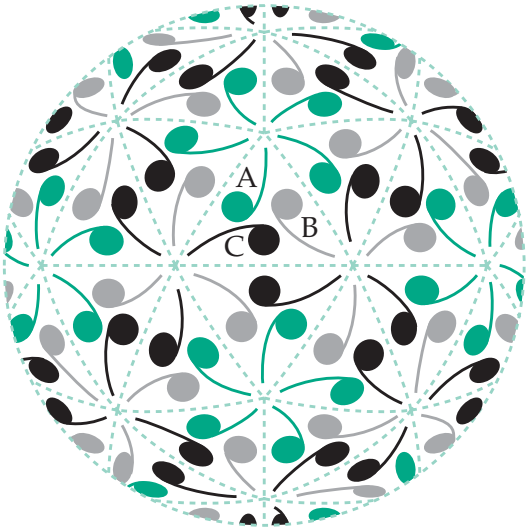



**Figure 7-19** Schematic icosahedrally symmetric structure with 180 subunits. The quasi-equivalent units A, B, and C are necessarily somewhat differently positioned with respect to their neighbors and must therefore assume different conformations in order to fit together tightly. From Harrison.[68]

to be violated.[98,99] Flexible arms tie the pentamers together.

Quasi-equivalence of subunits also provides the supercoil in bacterial flagella (Chapter 19) and accounts for some interesting aspects of the structure of tobacco mosaic virus. The protein subunits of the virus can exist either as a helix with 16.3 subunits per turn (Fig. 7-8) or as a flat ring of 17 subunits.[100] A very small conformational difference is involved. These rings dimerize but do not form larger aggregates. What is surprising is that the dimeric rings do not have dihedral symmetry, all of the subunits in the dimeric disk being oriented in one direction but with two different conformations. The disk may serve as an intermediate in virus assembly. The inner portions of the quasi-equivalent disk subunits have a jawlike appearance as if awaiting the incorporation of RNA. As the RNA becomes bound, the disks could dislocate to a "lock-washer" conformation to initiate and to propagate growth of the helical virus particle.[100,44a] However, there is uncertainty about this interpretation.[45,101]

Some enzymes, such as yeast hexokinase and creatine kinase (Chapter 12), associate in extremely asymmetric ways.[102] A dimer is formed by means of heterologous interactions but steric hindrance prevents the unsatisfied sets of interacting groups from joining with additional monomers to form higher polymers. As Galloway pointed out, many biological structures are not completely ordered but nevertheless possess well-defined and functionally important local relationships.[103]

***Regulatory subunits and multienzyme complexes.*** Proteins are often organized into large complexes, sometimes for the purpose of regulating metabolism. An example is **aspartate carbamoyltransferase** which catalyzes the first step in the synthesis of the pyrimidine rings of DNA and RNA (Chapter 25). The 310-kDa enzyme from *E. coli* can be dissociated into two 100-kDa trimers, referred to as **catalytic subunits**, and three 34-kDa dimers, the **regulatory subunits** which alter their conformations in response to changes in the ATP, UTP, and CTP concentrations.[104–107] The molecule is roughly triangular in shape[47,108] with a thickness of 9.2 nm and a length of the triangular side of 10.5 nm (Fig. 7-20). The symmetry is 3:2, i.e., it is dihedral with one 3-fold axis of rotation and three 2-fold axes. The two trimers of catalytic subunits lie face-to-face with the dimeric regulatory subunits fitting between them into the grooves around the edges of the trimers (Fig. 7-20). The dimers are not aligned exactly parallel with the 3-fold axis, but to avoid eclipsing, the upper half of the array is rotated around the 3-fold axis with respect to the lower half. In the center is an aqueous cavity of dimensions ~2.5 x 5.0 x 5.0 nm. The active sites of the enzyme are inside this cavity which is reached through six ~1.5-nm opening around the sides.

Many other oligomeric enzymes and other complex assemblies of more than one kind of protein subunit are known. For example, the **2-oxoacid dehydrogenases** are huge 2000- to 4000-kDa complexes containing three different proteins with different enzymatic activities in a cubic array (Fig. 15-14). The filaments of striated muscle (Chapter 19), antibodies and complement of blood (Chapter 31), and the tailed bacteriophages (Box 7-C ) all have complex molecular architectures.

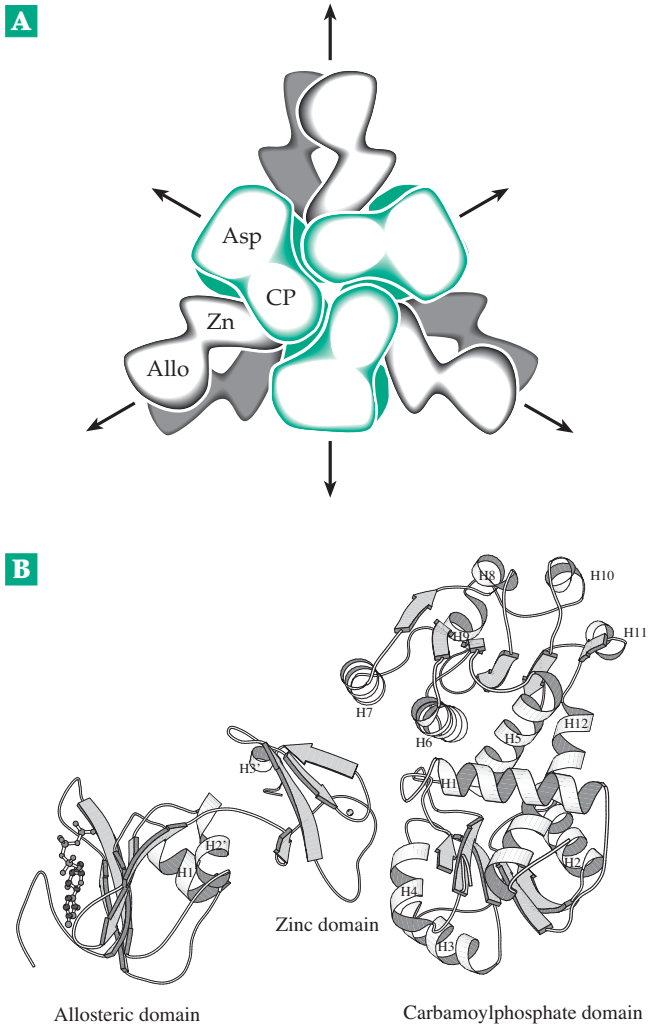


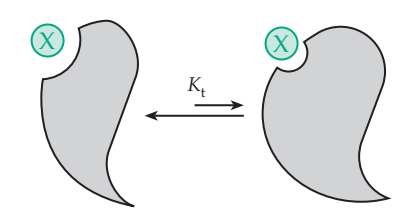

**Figure 7-20** (A) Subunit assembly of two $C_3$ catalytic trimers (green) and three $R_2$ regulatory dimers around the periphery in aspartate carbamoyltransferase. After Krause *et al.*[109] Courtesy of William N. Lipscomb. The aspartate- and carbamoylphosphate-binding domains of the catalytic subunits are labeled Asp and CP, respectivley, while the zinc and allosteric domains of the regulatory subunits are labeled Allo and Zn, respectively. (B) Ribbon drawing of a single pair of regulatory (left) and catalytic (right) subunits with the structural domains labeled. MolScript drawing from Thomas *et al.*[110]

## C. Cooperative Changes in Conformation

A substrate will bind better to some conformations of a protein than it will to others. This simple fact, together with the tendency for protein monomers to associate into clusters, allows for cooperative changes in conformation within oligomeric proteins. These changes provide the basis for important aspects of the regulation of enzymes and of metabolism. They impart cooperativity to the binding of small molecules such as that of oxygen to hemoglobin and of substrates and regulating molecules to enzymes. Many of the most fundamental and seemingly mysterious properties of living things are linked directly to cooperative changes within the fibrils, membranes, and other structures of the cell.

In 1965 a simple, appealing mathematical description of cooperative phenomena was suggested by Monod, Wyman, and Changeux[30,110a,110b] and focused new attention on the phenomenon. They suggested that conformational changes in protein subunits, which could be associated with altered binding characteristics, occur cooperatively within an oligomer. For example, binding of phenol to hexameric 2-zinc insulin (Fig. 7-18) could induce all six individual subunits to change their conformation together, preserving the $D_6$ symmetry of the complex. (In fact, it is more complex than this.[111]) The four subunits of hemoglobin could likewise change their conformation and affinity for $O_2$ synchronously. This is very nearly true and is of major physiological significance.

Consider an equilibrium (Eqs. 7-27 and 7-28) between protein molecules in two different conformations A and B (T and R in the MWC terminology) and containing a single binding site for molecule X. In the Monod–Wyman–Changeux (MWC) model the conformations are designated T (tense) and R (relaxed) but in the interest of providing a more general treatment the terminology used in this book is that of Koshland *et al*.[13,112–115]




(7-27)

$$K_t = [B] / [A] \tag{7-28}$$

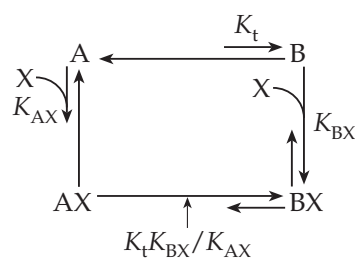
If the equilibrium constant $K_t$ is approximately 1, the two **conformers** have equal energies, but if $K_t < 1$, A is more stable than B.

### 1. Unequal Binding of Substrate and "Induced Fit"

Assume that conformer B binds X more strongly than does conformer A (as is suggested by the shapes of the binding sites in Eq. 7-27). The intrinsic binding constants to the A and B conformers $K_{AX}$ and $K_{BX}$ (or $K_T$ and $K_R$) are defined by Eq. 7-29:

$$K_{AX} = [AX] / [A][X]$$
$$K_{BX} = [BX] / [B][X]) \tag{7-29}$$

The entire set of equilibria for this system are shown in Eq. 7-30. Note that the constant relating



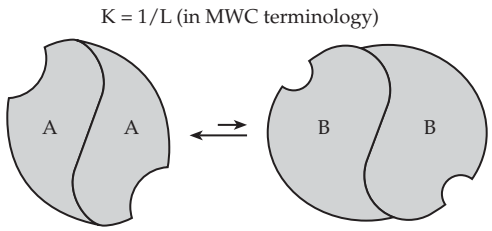

(7-30)

BX to AX is not independent of the other three constants but is given by the expression $K_tK_{BX} / K_{AX}$. Now consider the following situation. Suppose that A predominates in the absence of X but that X binds more tightly to B than to A. There will be largely either free A or BX in the equilibrium mixture with smaller amounts of AX and B. An interesting kinetic question arises. By which of the two possible pathways from A to BX (Eq. 7-30) will the reaction take place? The first possibility, assumed in the MWC model, is that X binds only to preformed B, which is present in a small amount in equilibrium with A. The second possibility is that X can bind to A but that AX is then rapidly converted to BX. We could say that X *induces a conformational change* that leads to a better fit. This is the basis for the **induced fit** theory of Koshland. Bear in mind that the equilibrium constants can give us the equilibrium concentrations of all four forms in Eq. 7-30. However, rates of reaction are often important in metabolism and we cannot say *a priori* which of the two pathways will be followed.

If $K_{BX} / K_{AX}$ is very large, an insignificant amount of AX will be present at equilibrium. In such a case there is no way experimentally to determine $K_{AX}$. The two constants $K_t$ and $K_{BX}$ are sufficient to describe the *equilibria* but an induced fit mechanism may still hold.

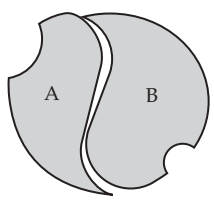
Now consider the association of A and B to form oligomers in which the intrinsic binding constants $K_{AX}$ and $K_{BX}$ have the same values as in the monomers. Since more enzymes apparently exist as isologous dimers than as any other oligomeric form,[116] it is appropriate to consider the behavior of such dimers in some detail. Monod *et al*. emphasized that both

conformers A and B (T and R) can associate to form isologous dimers in which symmetry is preserved (Eq. 7-31).




(7-31)

On the other hand, association of B and A would lead to an unsymmetric dimer in which bonding between subunits might be poor:




Mixed AB dimer
which associates weakly

In the MWC treatment, the assumption is made that the mixed dimer AB can be neglected entirely. However, a more general treatment requires that we consider all dimeric forms. The formation constants $K_{AA}$, $K_{BB}$, and $K_{AB}$ are defined as follows[13,113] (Eqs. 7-32 to 7-34; note the statistical factor of 2 relating $K_{AB}$ to the association constant $K_f$):

$$2A \rightleftarrows A_2 \qquad K_{AA} = \frac{[A_2]}{[A]^2} \tag{7-32}$$

$$2B \rightleftarrows B_2 \qquad K_{BB} = \frac{[B_2]}{[B]^2} = \frac{[B_2]}{K_t^2[A]^2} \tag{7-33}$$

$$A + B \rightleftarrows AB \qquad K_f = 2K_{AB} = \frac{[AB]}{[A][B]} = \frac{[AB]}{K_t[A]^2} \tag{7-34}$$

## 2. Binding Equilibria for a Dimerizing Protein

All of the equilibria of Eqs. 7-28 through 7-34 involved in formation of dimers $A_2$, AB, and $B_2$ and in the binding of one or two molecules of X per dimer are depicted in Fig. 7-21. Above each arrow the microscopic constant associated with that step is shown multiplied by an appropriate statistical factor. The fractional saturation Y is given by Eq. 7-35. Each of the nine terms in the numerator gives the concentration of

$$2Y \text{ (based on dimer)} = \frac{[AX] + [BX] + [A_2X] + 2[A_2X_2] + [ABX] + [AXB] + 2[ABX_2] + [B_2X] + 2[B_2X_2]}{\frac{1}{2}([A] + [AX] + [B] + [BX]) + [A_2] + [A_2X] + [A_2X_2] + [AB] + [ABX] + [AXB] + [ABX_2] + [B_2] + [B_2X] + [B_2X_2]} \tag{7-35}$$

bound X represented by one of the nine forms containing X in Fig. 7-21. The 14 terms in the denominator represent the concentration of protein in each form including those containing no bound X. Protein concentrations are given in terms of the molecular mass of the dimer; hence, some of the terms in the denominator are multiplied by 1/2.

All of the terms in both the numerator and the denominator of Eq. 7-35 can be related back to [X], using the microscopic constants from Fig. 7-21 to give an equation (comparable to Eq. 7-8) which presents Y in terms of [X], $K_{AX}$ and $K_{BX}$, $K_t$, and the interaction constants $K_{AA}$, $K_{AB}$, and $K_{BB}$. Since the equation is too complex to grasp immediately, let us consider several specific cases in which it can be simplified.

***The Monod-Wyman-Changeux (MWC) model.*** If both $K_{AA}$ and $K_{BB}$ are large enough, there will be no dissociation into monomers. The transition between conformation A and conformation B can occur cooperatively within the dimer or higher oligomer, and the mathematical relationships shown in Fig. 7-21 are still appropriate. One further restriction is needed to describe the MWC model. Only symmetric dimers are allowed. That is, $K_{AA}$ and $K_{BB} >> K_{AB}$ (see Eq. 7-31), and only those equilibria indicated with green arrows in Fig. 7-21 need be considered.[30] In the absence of ligand X, the ratio $[B_2]/[A_2]$ is a constant, 1/L in the MWC terminology (Eq. 7-36; see also Eq. 7-31).

$$\frac{[B_2]}{[A_2]} = \frac{1}{L} = \frac{K_{BB}}{K_{AA}} K_t^2 \tag{7-36}$$

Both of the association constants $K_{AA}$ and $K_{BB}$ and the transformation constant $K_t$ affect the position of the equilibrium. Thus, a low ratio of $[B_2]$ to $[A_2]$ could result if $K_{BB}$ and $K_{AA}$ were similar but $K_t$ was small. If $K_t$ were ~1 a low ratio could still arise because $K_{AA} >$

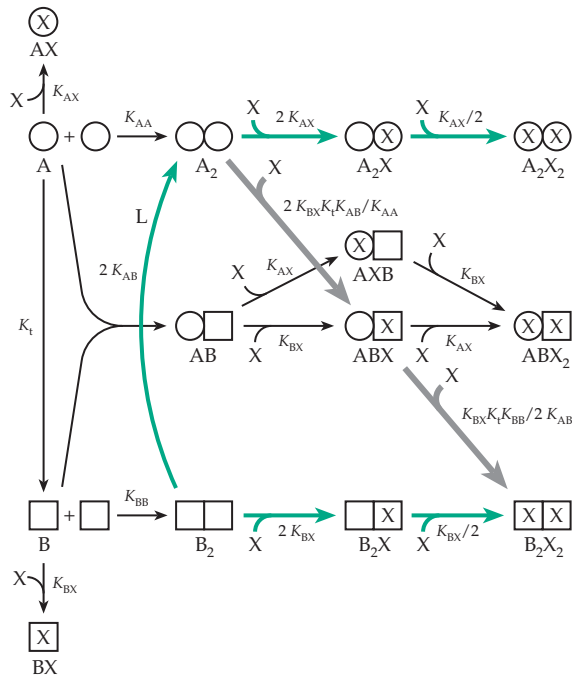



**Figure 7-21** Possible forms of a dimerizing protein existing in two conformations with a single binding site per protomer for X. Green arrows indicate equilibria considered by MWC. Solid arrows indicate equilibria considered by Koshland *et al.*[13,114] Heavy gray arrows are for the simplest induced fit model with no dissociation of the dimer. Note that all equilibria are regarded as reversible (despite the unidirectional arrows). $K_{AX}$ and $K_{BX}$ are assumed the same for subunits in monomeric and dimeric forms.

$K_{BB}$, i.e., because the subunits are associated more tightly in $A_2$ than in $B_2$. For this case Eq. 7-35 simplifies to Eq. 7-37.

$$2Y = \frac{[A_2X] + 2\,[A_2X_2] + [B_2X] + 2\,[B_2X_2]}{[A_2] + [A_2X] + [A_2X_2] + [B_2] + [B_2X] + [B_2X_2]}$$

$$= \frac{2K_{AA}K_{AX}[X] + 2K_{AA}K_{AX}^2[X]^2 + 2K_{BB}K_{BX}K_t^2[X] + 2K_{BB}K_{BX}K_t^2[X]^2}{K_{AA} + 2K_{AA}K_{AX}[X] + K_{AA}K_{AX}^2[X]^2 + K_{BB}K_t^2 + 2K_{BB}K_{BX}K_t^2[X] + K_{BB}K_{BX}^2K_t^2[X]^2} \tag{7-37}$$

Substituting from Eq. 7-36 into Eq. 7-37 we obtain (Eq. 7-38):

$$Y \text{ (for dimer)} = \frac{L \cdot K_{AX}[X](1 + K_{AX}[X]) + K_{BX}[X](1 + K_{BX}[X])}{L(1 + K_{AX}[X])^2 + (1 + K_{BX}[X])^2} \tag{7-38}$$

For an oligomer with $n$ subunits Monod *et al.* assumed that all sites in either conformer are independent and equivalent. The equation for Y (based on Eq. 7-17) is

$$Y = \frac{L \cdot K_{AX}[X](1 + K_{AX}[X])^{n-1} + K_{BX}[X](1 + K_{BX}[X])^{n-1}}{L(1 + K_{AX}[X])^n + (1 + K_{BX}[X])^n} \tag{7-39}$$

We assume initially that $B_2$ binds X more strongly than does $A_2$. Hence, if the equilibrium in Eq. 7-36 favors $B_2$ strongly ($L$ is small), the addition of X to the system will not shift the equilibrium between the two conformations and binding will be noncooperative (Eq. 7-39 will reduce to Eq. 7-18). However, if the equilibrium favors $A_2$ ($L$ is large), addition of X will shift the equilibrium in favor of $B_2$ (which binds X more tightly). Furthermore, since the expression for Y (Eq. 7-39) contains a term in $K^2_{BX}[X]^2$ in the numerator, binding will tend to be cooperative. In the extreme case that $L$ is large and $K_{AX} \sim 0$, most of the terms in Eq. 7-39 drop out and it approaches the equation previously given for completely cooperative binding (Eq. 7-21) with $K = K_{BX}^2L$. With other values of $K_{AX}$, $K_{BX}$, and $L$ incomplete cooperativity is observed.[112]

Further development of the MWC theory as it applies to enzyme kinetics is given in Chapter 9, Section B.

***The induced fit model.*** In this model, only $A_2$, ABX, and $B_2X_2$ are considered (heavy arrows in Fig. 7-15).[13,114] The expression for 2Y is:

$$2\,Y = \frac{[ABX] + 2\,[B_2X_2]}{[A_2] + [ABX] + [B_2X_2]} = \frac{2\,K_{BX}K_t\,\dfrac{K_{AB}}{K_{AA}}[X] + 2\,(K_{BX}K_t)^2\,\dfrac{K_{BB}}{K_{AA}}[X]^2}{1 + 2\,K_{BX}K_t\,\dfrac{K_{AB}}{K_{AA}}[X] + (K_{BX}K_t)^2\,\dfrac{K_{BB}}{K_{AA}}[X]^2} \tag{7-40}$$

The constants used here are defined by Eqs. 7-8 through 7-10 and differ from those of Koshland, who sometimes arbitrarily set $K_{AA} = 1$ and redefined $K_{BB}$ as an *interaction constant* equal to $K_{BB} / K_{AA}$. Although this simplifies the algebra it is appropriate only for completely associated systems and might prove confusing.

When $K_{AB}$ is small (no "mixed" dimer) Eq. 7-16 also simplifies to Eq. 7-45 for completely cooperative binding with the value $K$ given by Eq. 7-17. On the other hand, if $K_{AB}$ is large compared to $K_{AA}$ and $K_{BB}$, anticooperativity (negative cooperativity) will be observed. The saturation curve will contain two separate steps just as in the binding of protons by succinate dianion (Fig. 7-5).

$$\overline{K} = K_{BX}{}^2 K_t{}^2 \frac{K_{BB}}{K_{AA}} = \frac{K_{BX}{}^2}{L} \tag{7-41}$$

***One conformational state dissociated.*** It may happen that $A_2$ is a dimer but that $B_2$ dissociates into monomers because $K_{BB}$ is very small. In such a case binding of X leads to dissociation of the protein. A well-known example is provided by hemoglobin of the lamprey which is a dimer and which dissociates to a monomer upon binding of oxygen.[117] Equation 7-11 simplifies to Eq. 7-42. The reader may wish to consider whether the weakly cooperative binding of oxygen by lamprey hemoglobin is predicted by this equation.

$$2Y = \frac{[BX]}{[A_2] + \frac{1}{2}[BX]} \tag{7-42}$$

Look again at the expression for $L$, the constant determining the relative amounts of a protein in conformations A and B in the absence of ligand. From Eq. 7-36 we see that a large value of $L$ (conformer A favored) can result either because $K_t$ is very small or because $K_{BB} << K_{AA}$. Thus, if $K_t \sim 1$ and $L$ is large, the subunits must associate much more weakly in $B_2$ than in $A_2$ and the chances are that binding of X will dissociate the molecule as in the case of lamprey hemoglobin. On the other hand, if $K_t$ is very small, implying that the molecule is held in conformation A because of some intrinsically more stable folding pattern in that conformation, $K_{BB}$ might exceed $K_{AA}$ very much; if $K_{AA}$ were low enough $A_2$ could be completely dissociated. Binding of ligand would lead to association and to cooperative binding. This can be verified by writing down the appropriate terms from Eq. 7-35.

## 3. Higher Oligomers

Mathematical treatment of binding curves for oligomers containing more than two subunits is complex, but the algebra is straightforward. A computer can be programmed to do necessary calculations. Avoid picking an equation from the literature and assuming that it will be satisfactory. Consider the two tetrameric structures in Fig. 7-22. In the isologous square (also shown in Fig. 7-11) separate contributions to the free energy of binding can be assigned to the individual pairs of interactions *aj* and *bk*.

Thus, following Cornish-Bowden and Koshland[114] for assembly of the tetramer (Eq. 7-43):

$$\Delta G_f = 2\,\Delta G_{ajAA} + 2\,\Delta G_{bkAA}$$
$$K_f = K_{ajAA}^2 K_{bkAA}^2 \tag{7-43}$$

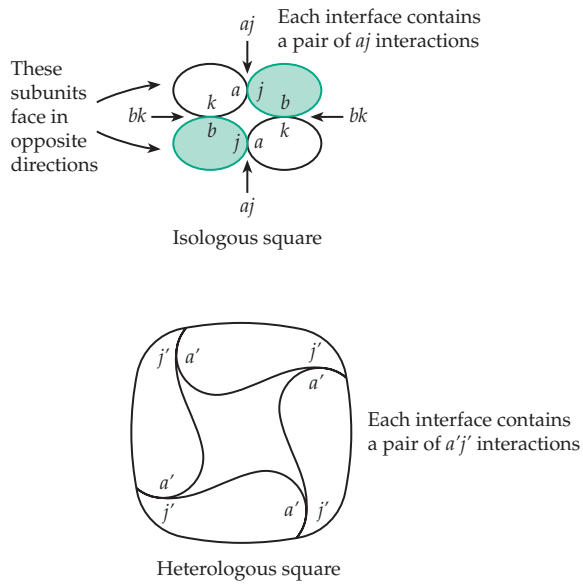
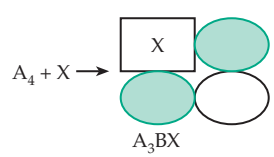



**Figure 7-22** Comparison of the interactions in isologous (dihedral) and heterologous (cyclic) square configurations of subunits.

Since Gibbs energies are additive, the formation constant will be the product of formation constants representing the individual interactions; thus, $K_{ajAA}$ represents the formation constant of a dimer in which only the *aj* pair of bonds is formed.

In the isologous tetrahedron (Fig. 7-5) the third set of paired interactions *cl* must be taken into account. (However, the third interaction constant will not be an association constant of the type represented by $K_{ajAA}$ and $K_{bkAA}$ but a dimensionless number.) On the other hand, the heterologous square has only a single interaction constant.

Now consider the binding of one molecule of X to the isologous tetramer with a conformational change in one subunit (Eq. 7-44). We see that one pair of *aj* interactions and one pair of *bk* interactions have been

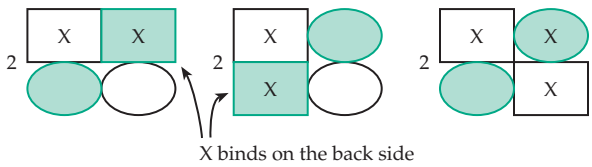



(7-44)

altered. The equilibrium constant for the binding of X to the tetramer will be (Eq. 7-45) in which the 4 is a

$$K = 4\,\frac{K_{ajAB}K_{bkAB}}{K_{ajAA}K_{bkAA}}\,K_{BX}K_t \tag{7-45}$$

statistical factor arising from the fact that there are four different ways in which to form $A_3BX$. When a second molecule of X is added three geometrical arrangements are possible:




Each one can be formed in two ways. It is a simple matter to write down the microscopic constants for addition of the second molecule of X as the sum of three terms. Because values of the constants for *aj* and *bk* differ, it will be clear that the three ways of adding the second molecule of X are not equally probable. Thus, the oligomer will show preferred orders of "loading" with ligand X.

Two different geometries for the heterologous tetramer are possible in form $A_2B_2X_2$. Again, the different arrangements need not be equally probable and the relative distribution of each will be determined by the specific values of the interaction constants. In the heterologous tetramer $A_4$ only one type of an interaction is present between subunits. However, as soon as a single molecule of X is bound and one subunit of conformation B is present, two kinds of *aj* interactions exist. (One in which group *a* is present in conformation A and the other in which it is present in conformation B.) Since these interactions always occur in equal numbers they can be lumped together.

While the foregoing may seem like an unnecessarily long exercise, it should provide a basic approach which can be applied to specific problems. However, remember that mathematical models require simplification. Real proteins often have more than two stable conformations.[117] The entire outside surface of a protein is made up of potential binding sites for a number of different molecules, both small and large. Filling of almost any of these sites can affect the functioning of a protein.

## D. The Oxygen-Carrying Proteins

### 1. Myoglobin and Hemoglobin

The most studied example of a conformational change in a multisubunit protein induced by binding of a small molecule is provided by the cooperative binding of oxygen to hemoglobin.[118–120] Mammalian hemoglobin is an $\alpha_2\beta_2$ tetramer of ~16-kDa subunits, each containing 140 – 150 residues. Within each subunit the peptide chain folds in a characteristic largely α-helical pattern around a single large flat iron-containing ring structure called **heme** (Fig. 7-23). The folding is essentially the same in all hemoglobins, both in the α and β subunits and in the monomeric muscle oxygen storage protein, myoglobin. Amino acid residues are customerily designated by their position in one of the eight helices A–H. The imidazole group of histidine F-8 is coordinated with the iron in the center of the heme on the "proximal" side. The other side of the iron atom (the "distal" side) is the site of binding of a single molecule of $O_2$.

Although the folding of the peptide chain is almost the same in both subunits, and almost identical to that of myoglobin,[118,121,122] there are numerous differences in the amino acid sequence. If it were not for these differences, hemoglobin would be a highly symmetric molecule with the bonding pattern indicated in Fig. 7-5 with three 2-fold axes of rotation. In fact, hemoglobin has one true axis of rotation and two pseudo-twofold axes. There are two sets of true isologous interactions (those between the two α subunits and between the two β subunits) and two pairs of unsymmetrical interactions (between α and β subunits). The nearly symmetric orientation of different portions of the peptide backbone is clearly seen in the beautiful drawings of Geis.[119]

The contact region involved in one pair of interactions in hemoglobin ($\alpha_1\beta_1$) is more extensive than the other. There is close contact between 34 different amino acid side chains and 110 atoms lie within 0.4 nm of each other.[118] Hydrophobic bonding is the principal force holding the two subunits together, and only a few reciprocal contacts of the type found in a true isologous bond remain. The second contact designated $\alpha_1\beta_2$ involves only 19 residues and a total of 80 atoms. Because this interaction is weaker, hemoglobin dissociates relatively easily into αβ dimers held together by the $\alpha_1\beta_1$ contacts and motion occurs along the $\alpha_1\beta_2$ contacts during oxygenation. The truly isologous interactions (i.e., αα and ββ ) are weak because the identical protomers hardly touch each other.

***The binding of oxygen.*** Curves of percentage oxygenation (Y) vs the partial pressure of $O_2$ are given in Fig. 7-24 and illustrate the high degree of cooperativity. Depending upon conditions, values of $n_{Hill}$ (Eq. 7-24) may be as high as 3. As a result of this cooperativity the hemoglobin, in the capillaries of the lungs at a partial pressure of oxygen of ~100 mm of mercury, is nearly saturated with oxygen. However, when the red cells pass through the capillaries of tissues in which oxygen is utilized the partial pressure of oxygen falls to about 5 mm of mercury. The cooperativity means that the oxygen is more completely "unloaded" than it would be if all four heme groups acted independently.

Deoxyhemoglobin has a low affinity for $O_2$, but the observed cooperativity in binding implies that in the fully oxygenated state the $O_2$ is held with a high

affinity. The monomeric myoglobin also has a high affinity for oxygen, as does the abnormal **hemoglobin H**, which is made up of four β subunits. The latter also completely lacks cooperativity in binding.[123] These results can be interpreted according to the MWC model to indicate that deoxyhemoglobin exists in the T (A) conformation, whereas oxyhemoglobin is in the R (B) conformation. Myoglobin stays in the R conformation in *both* states of oxygenation as do the separated α and β chains of hemoglobin. The subunits of hemoglobin H also appear to be frozen in the R conformation, even though the quaternary structure is similar to that of deoxyhemoglobin.[123,124]

Oxygenation curves of hemoglobin are often fitted with the Adair equation (Eq. 7-12). Thus, at pH 7.4 under the conditions given in Table 7-2, Imai[125] found for the successive formation constants $K_1 = 0.004$, $K_2 = 0.009$, $K_3 = 0.002$ and $K_4 = 0.95$ in units of mm $Hg^{-1}$. From the definition of a formation constant the oxygen pressure $P_{O_2}$ required for 50% oxygenation in the first step will be at $P_{O_2} = 1/K_f$ or $\log P_{O_2} = \log(1/.004) = 2.4$. This is a high oxygen pressure, far to the right side of the oxygenation curve in Fig. 7-24A. However, $\log K_4$ is about 0.02, well to the left on the oxygenation curve. From these formation constants we can say that after three of the subunits have become oxygenated the affinity of the remaining subunit has increased about 300-fold when the concentration of the effector 2,3-bisphosphoglycerate is present at the normal physiological concentraton (see Section 4).[125–127]

However, we must ask what uncertainties are present in the data used to obtain these constants. To extract four successive binding constants from a curve like that in Fig. 7-24A is extremely difficult.[129,130] This fact has encouraged the widespread use of the simpler MWC model.[30a,127a] When the same data were treated by Imai[131] using the MWC model it was found that $L = 2.8 \times 10^6$ and $c = K_{f(T)} / K_{f(R)} = 0.0038$. Changes in

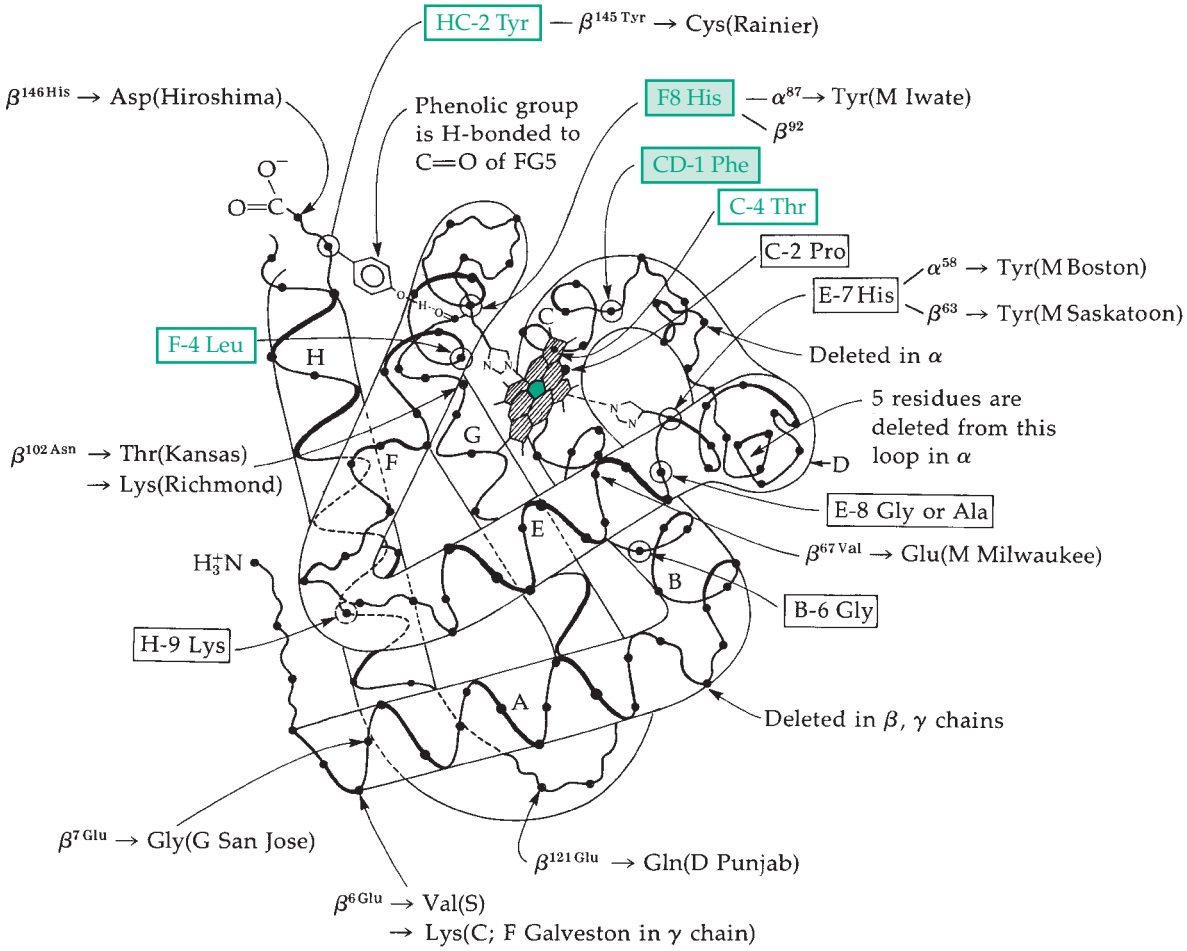



**Figure 7-23** Folding pattern of the hemoglobin monomers. The pattern shown is for the β chain of human hemoglobin. Some of the differences between this and the α chain and myoglobin are indicated. Evolutionarily conserved residues are indicated by boxes, ▭ highly conserved, ▭ invariant. Other markings show substitutions observed in some abnormal human hemoglobins. Conserved residues are numbered according to their location in one of the helices A–H, while mutant hemoglobins are indicated by the position of the substitution in the entire α and β chain.

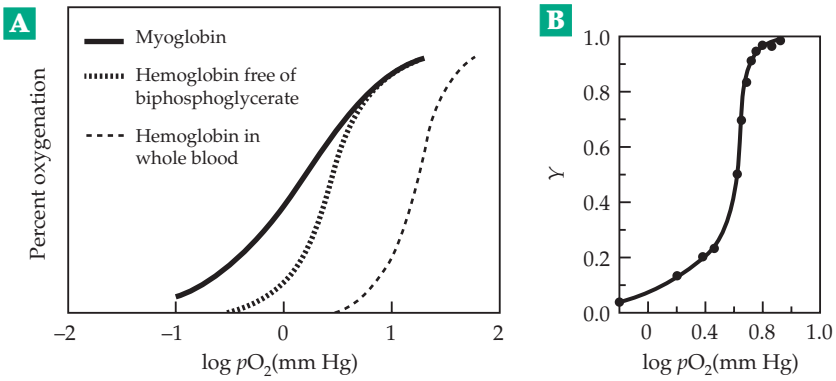



**Figure 7-24** Cooperative binding of oxygen by hemoglobins. (A) Binding curve for myoglobin (noncooperative) and for hemoglobin in the absence and presence (in whole blood) of 2,3-bisphosphoglycerate. Oxygen affinity is decreased by bisphosphoglycerate. (B) Saturation curve for hemoglobin (erythrocruorin) of *Arenicola*, a spiny annelid worm. The molecule contains 192 subunits and 96 hemes. It shows very strong cooperativity with $n_{Hill} \sim 6$. From (A) Benesch and Benesch,[123] and (B) Waxman.[128]

enthalpy, entropy, and Gibbs energy are given in Table 7-2. Kinetic data[131a] as well as $O_2$-binding measurements with single crystals[132] are partially consistent with the MWC model.[133] However, the discovery of a third quaternary structure of hemoglobin, similar to the R state but distinct from it,[134–136a] emphasizes the complexity of this allosteric molecule.

Hemoglobin tetramers tend to stay tightly associated but some dissociation of oxyhemoglobin into dimers does occur ($K_f = 7 \times 10^5$ M).[137,138] Deoxyhemoglobin is about 40,000 times more tightly associated. All of the equilibria involved are strongly affected by pH and by the presence of salts such as NaCl.[126,139] This is in part a reflection of the strong role of ionic interactions in holding together the subunits in the T state as is discussed in the following sections.[127]

**TABLE 7-2**
**Thermodynamic Functions for Oxygenation of Hemoglobin[a]**

| Reaction | $\Delta H$ (kJ mol$^{-1}$) | $\Delta S$ (J°K$^{-1}$ mol$^{-1}$) | $\Delta G$ (kJ mol$^{-1}$) | $K_f$ |
|---|---|---|---|---|
| $T \rightarrow T(O_2)_4$ | $-51 \pm 1$ | $-154 \pm 4$ | $-5.0 \pm 1.7$ | 7.5 |
| $R \rightarrow R(O_2)_4$ | $-62 \pm 2$ | $-146 \pm 6$ | $-19 \pm 2$ | $2.0 \times 10^3$ |
| $T \rightarrow R$ (unoxygenated) | $-70 \pm 7$ | $-111 \pm 25$ | $-37 \pm 10$ | $3.6 \times 10^{-7}$ |
| $T(O_2)_4 \rightarrow R(O_2)_4$ (oxygenated) | | | $-19$ | $2.1 \times 10^3$ |

Parameters for MWC model

$L = (3.6 \times 10^{-7})^{-1} = 2.8 \times 10^6$

$c = K_{f(T)}/K_{f(R)} = 0.0038$

[a] From Imai, K. (1979) *J. Mol. Biol.* **133**, 233–247. The measurements were made at pH 7.4 in the presence of 0.1 M chloride ion and 2 mM 2,3-bisphosphoglycerate to mimic physiological conditions. The values of $\Delta H$, $\Delta S$, and $\Delta G$ given are per mole of heme, i.e., per monomer unit. They must be multiplied by 4 to correspond to the reactions as shown for the tetramer.

***Structural changes accompanying oxygen binding.*** Perutz and associates, using X-ray crystallography, found *small* but real differences in the conformation of the subunits of deoxy- and oxy- hemoglobin.[140,141] More striking is the fact that upon oxygenation, both α and β subunits undergo substantial amounts of *rotation*, the net result being that the hemes of the two β subunits move about 0.07 nm closer together in the oxy form than in the deoxy form. Within the $\alpha_1\beta_1$ contacts (Fig. 7-25) little change is seen. On the other hand, contact $\alpha_1\beta_2$, the "allosteric interface," is altered drastically. As Perutz expressed it, there is a "jump in the dovetailing" of the CD region of the subunit relative to the FG region of the β subunit. The hydrogen-bonding pattern is also changed. A major difference is seen in the hydrogen-bonded salt bridges present at the ends of the molecules of deoxyhemoglobin. The $-NH_3^+$ group of Lys H-10 in each α subunit is hydrogen bonded to the carboxyl group of the C-terminal arginine of the opposite α chain. The guanidinium group of each C-terminal arginine is hydrogen bonded to the carboxyl group of Asp H-9 in the opposite α chain. It is also hydrogen bonded to an inorganic anion (phosphate or $Cl^-$), which in turn is hydrogen-bonded to the α amino group of Val 1 of the opposite α chain[142] forming a pair of isologous interactions.

At the other end of the molecule, the C-terminal group of His 146 of each β chain binds to the amino group of Lys C-6 of the α chain, while the imidazole side chain binds to Asp FG-1 of the same β chain (Fig. 7-25). These salt bridges appear to provide extra stability to

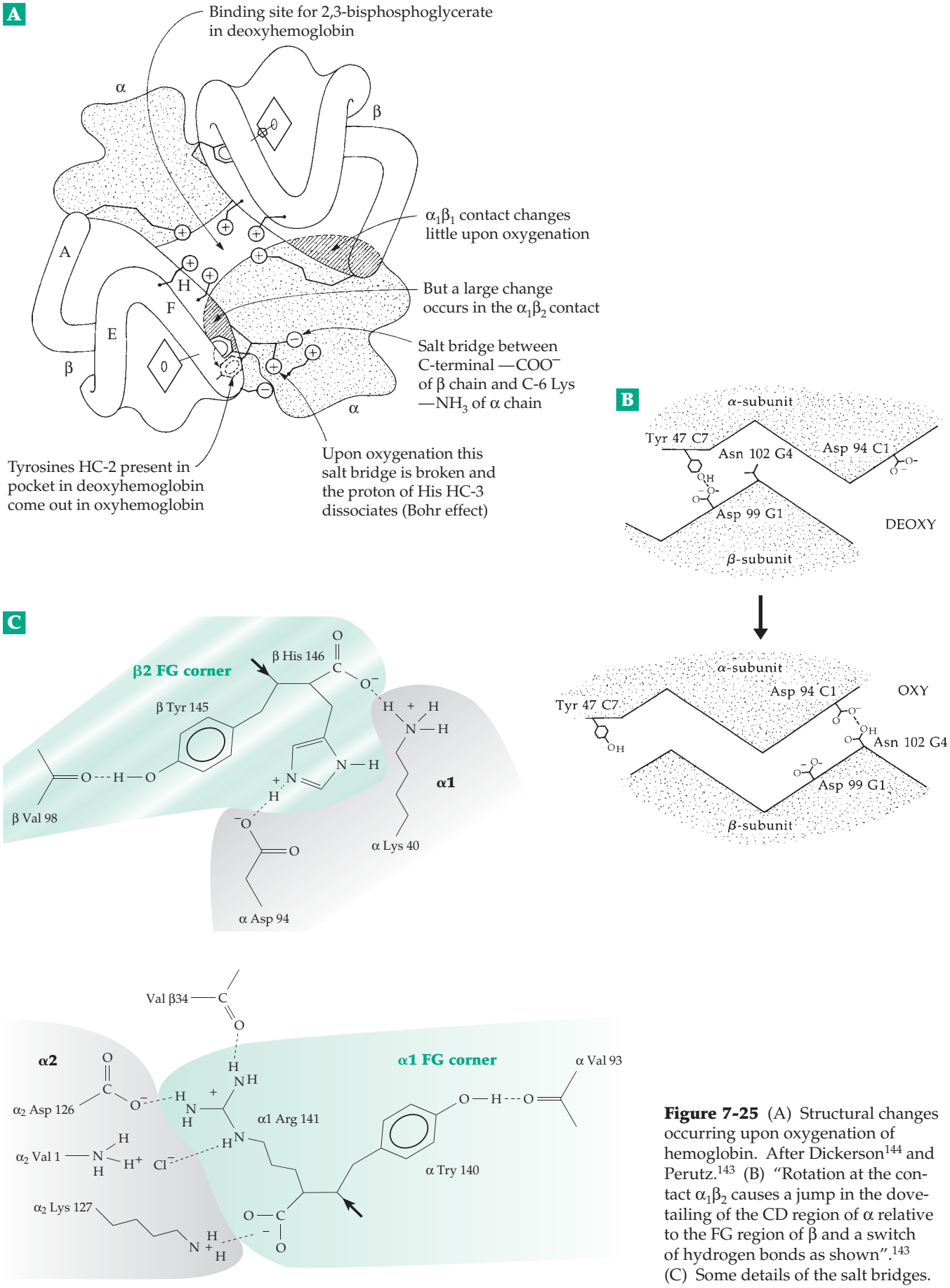



**Figure 7-25** (A) Structural changes occurring upon oxygenation of hemoglobin. After Dickerson[144] and Perutz.[143] (B) "Rotation at the contact $\alpha_1\beta_2$ causes a jump in the dovetailing of the CD region of α relative to the FG region of β and a switch of hydrogen bonds as shown".[143] (C) Some details of the salt bridges.

deoxyhemoglobin and account for the high value of the constant L. In deoxyhemoglobin the side chain of the highly conserved Tyr HC-2 lies tucked into a pocket between the H and F helices and is hydrogen bonded to the main chain carbonyl of residue FG-5 (Figs. 7-23 and 7-25). Upon oxygenation this tyrosine in each subunit is released from its pocket; the salt bridges at the ends of the molecules are broken and the subunit shifts into the new bonding pattern characteristic of oxyhemoglobin.[143,144] Cooperativity in $O_2$ binding is absent or greatly decreased in mutant hemoglobins with substitutions in the residues involved in these salt bridges[145–147] or in residues lying in the $\alpha_1\beta_2$ interface.[148,149]

How does the binding of $O_2$ to the iron of heme trigger the conformational change in hemoglobin? An enormous amount of effort by many people has been expended in trying to answer this question. As is pointed out in Chapter 16, the iron atom in deoxyhemoglobin lies a little outside the plane of the heme rings. When oxygenation occurs the iron atom moves toward the oxygen and into the plane of the heme.[150,151] This movement probably amounts to only about 0.05 nm. Nevertheless, this small displacement evidently induces the other structural changes that are observed. The iron pulls the side chain of histidine F-8 with it and moves helix F which is also hydrogen bonded to this imidazole ring. Because of the tight packing of the various groups this motion cannot occur freely but is accompanied by a movement of the F helix by 0.1 nm across the heme plane. These movements may induce additional structural changes in the irregularly folded FG corners that allow the subunits to shift to the new stable position of the R state. All four subunits appear to change conformation together. This conformational change must also cause the affinity for oxygen of any unoxygenated subunits to rise dramatically, presumably by shifting the iron atoms into the planes of the heme rings. This ensures the cooperative loading of the protein by $O_2$.

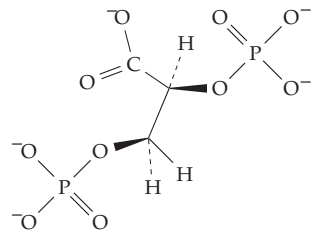
While there is no doubt that the iron atom moves upon oxygenation, it is not obvious that this will lead to the observed cooperativity. Oxygenated heme has some of the characteristics of an Fe(III)–peroxide anion complex.[152] The iron atom acquires an increased positive charge upon oxygenation by donating an electron for bond formation.

$$Fe^{2+}\ (\text{deoxy}) + O_2 \longrightarrow Fe^{3+}\text{–}\ O_2^{-} \qquad (7\text{-}46)$$

This may transmit an electronic effect through either His F-8 or the heme ring to the nearby $\alpha_1\beta_2$ interface and also affect the subunit interactions. The reaction of various heme proteins with oxygen is discussed further in Chapter 16.

***The Bohr effect and allosteric regulators.*** The breaking of the salt bridges at the ends of the hemoglobin molecule upon oxygenation has another important result. The p$K_a$ values of the N-terminal valines of the α subunits and of His HC-3 of the β subunits are abnormally high in the deoxy form because they are tied up in the salt bridges. In the oxy form in which the groups are free, the p$K_a$ values are lower. If hemoglobin is held at a constant pH of 7, these protons dissociate upon oxygenation. This **Bohr effect**, described in 1904,[153–156] is important because acidification of hemoglobin stabilizes the deoxy form. In capillaries in which oxygen pressure is low and in which carbon dioxide and lactic acid may have accumulated, the lowering of the pH causes oxyhemoglobin to release oxygen more efficiently. These effects are also strongly dependent on the presence of chloride ions.[139,142,143,157,158]

Just as the conformational equilibria in hemoglobin can be shifted by attachment of oxygen to the heme groups, so the binding of certain other molecules at different sites can also affect the conformation. Such compounds are called **allosteric effectors** or **regulators** because they bind at a site other than the "active site." They are considered in more detail in Chapter 9. An important allosteric effector for human hemoglobin is **2,3-bisphosphoglycerate**, a compound found in human red blood cells in a high concentration approximately equimolar with that of hemoglobin. One molecule of bisphosphoglycerate binds to a hemoglobin tetramer in the deoxy form with $K_f = 1.4 \times 10^5$ but has only half this affinity for oxyhemoglobin.[159] X-ray crystallography shows that bisphosphoglycerate binds between the two β chains of deoxyhemoglobin directly on the twofold axis (Fig. 7-26).[159] Because of the presence



2,3-Bisphosphoglycerate

of 2,3-bisphosphoglycerate in erythrocytes the affinity of oxygen for hemoglobin in whole blood is less than that for isolated hemoglobin[160,161] (Fig. 7-24). This is important because it allows a larger fraction of the oxygen carried to be unloaded from red corpuscles in body tissues. The bisphosphoglycerate level of red cells varies with physiological conditions, e.g., people living at higher elevations have a higher concentration.[161] It has been suggested that artificial manipulation of the level of this regulatory substance in erythrocytes may be of clinical usefulness for disorders in oxygen transport.

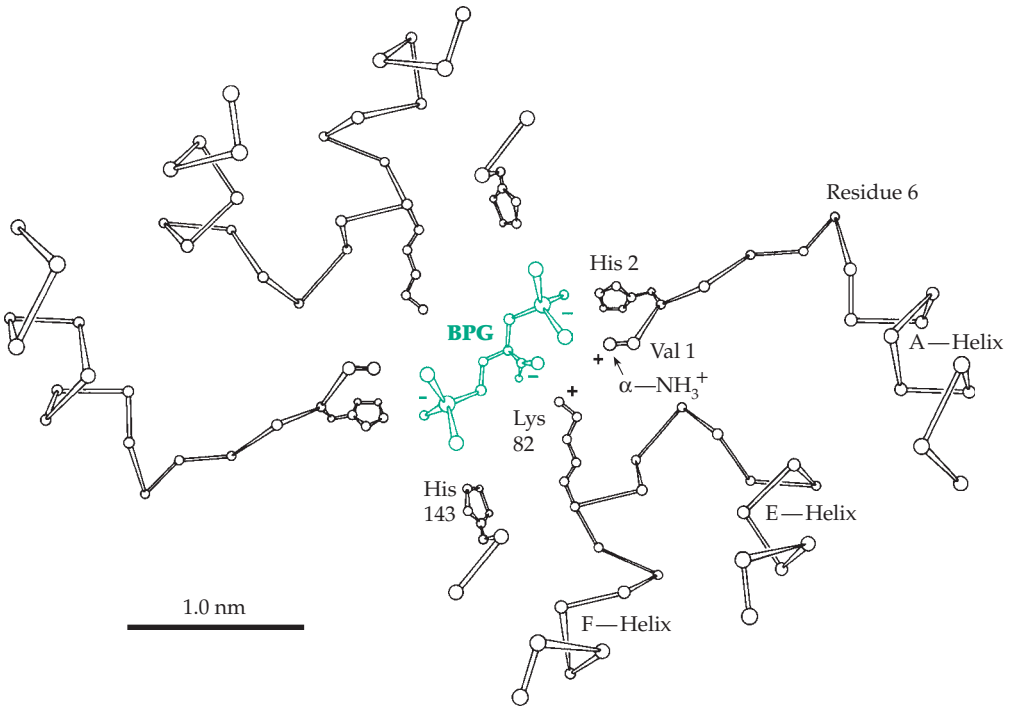
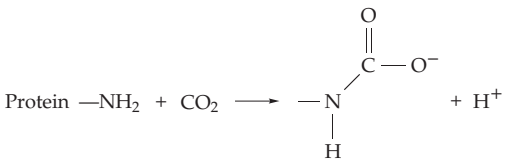



**Figure 7-26** The allosteric effects of 2,3-bisphosphoglycerate (BPG) bound to the β chains of human deoxyhemoglobin. The phosphate groups of the BPG form salt bridges with valines 1 and histidines 2 and 143 of both β chains and with lysine 82 of one chain. This binding pulls the A helix and residue 6 toward the E helix. From Arnone.[159]

Not all species contain 2,3-bisphosphoglycerate in their erythrocytes. In birds and turtles its function appears to be served by inositol pentaphosphate. In crocodiles the site between the two β chains that binds organic phosphates in other species has been modified so that it binds bicarbonate ion, $HCO_3^-$, specifically. This ion, which accumulates in tissues as crocodiles lie under water, acts as an allosteric regulator in these animals.[162,163] It allows the animals to more completely utilize the $O_2$ from the hemoglobin and to remain under water longer. The Bohr effect, which was considered in the preceding section, can be viewed as resulting from the action of **protons** as allosteric effectors that bind to the amino and imidazole groups of the salt linkages. **Carbon dioxide** also acts as a physiological effector in mammalian blood by combining reversibly with $NH_2$-terminal groups of the α and β subunits to form **carbamino** (– NH – COO ⁻) groups (Eq. 7-47).[119,164] It is the N-terminal amino groups rather than lysyl side chain groups that undergo this reaction. Because of their relatively low p$K_a$ values there is a significant




(7-47)

fraction of unprotonated $-NH_2$ groups at the pH of blood. The affinity for $CO_2$ is highest in deoxygenated hemoglobin. Consequently, unloading of $O_2$ is facilitated in the $CO_2$-rich respiring tissues. Hemoglobin carries a significant fraction of $CO_2$ to the lungs, and there the oxygenation of hemoglobin facilitates the dissociation of $CO_2$ from the carbamino groups. Hemoglobin is also one of the major pH buffers of blood.

***Carbon monoxide, cyanide, and nitric oxide.*** A danger to hemoglobin and other heme proteins is posed by competing ligands such as CO, $CN^-$, and NO. All of these are present within organisms and both CO and NO act as hormones. Hemoglobin and myoglobin are partially protected from carbon monoxide by the design of the binding site for $O_2$. The distal imidazole of histidine E7 hydrogen bonds to $O_2$ but not to the nonpolar CO. The site also accommodates the geometry of the bound $O_2$ better than that of CO.[165] Bound CO can be released from hemes by the action of light. Using X-ray diffraction[166,166a] and X-ray absorption measurements[167] at cryogenic temperatures, it has been possible to observe the motions of both the released CO and the heme in myoglobin, motions which may shed light on the normal oxygen transport cycle. Cooperativity in the binding of CO to hemoglobins has been studied extensively,[158,168,169] as has binding to model heme compounds.[170] Cyanide ions bind most tightly and also cooperatively[171–173] to the oxidized $Fe^{3+}$ form, which is called **methemoglobin**.

Nitric oxide is a reactive, paramagnetic gaseous free radical which is formed in the human body and in other organisms by an enzymatic oxidation of L-arginine (Eq. 18-65). Since about 1980, NO has been recognized as a hormone with a broad range of effects

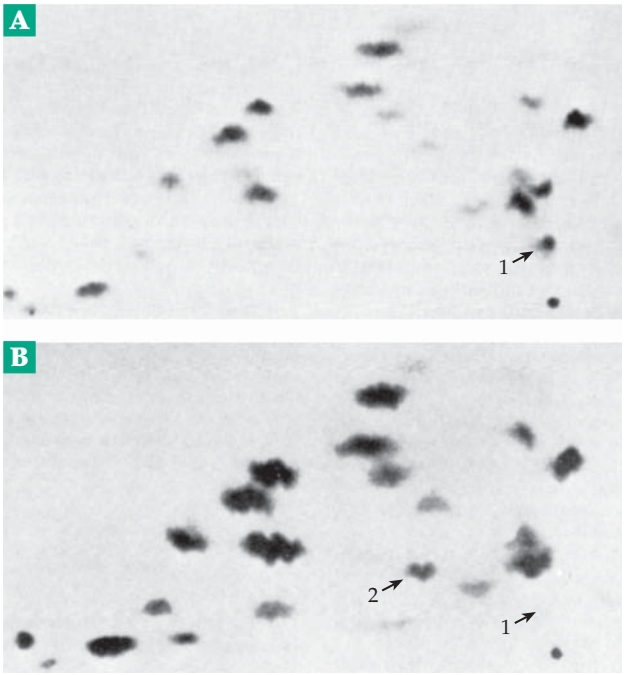



**Figure 7-27** "Fingerprints" of human hemoglobins. The denatured hemoglobin was digested with trypsin and the 28 resulting peptides were separated on a sheet of paper by electrophoresis in one direction (horizontal in the figures; anode to the left) and by chromatography in the other direction (vertical in the figure). The peptides were visualized by spraying with ninhydrin or with specific reagents for histidine or tyrosine residues. Since trypsin cuts only next to lysine, which occurs infrequently, the petide pattern provides a fingerprint, characteristic for any pure protein. (A) The fingerprint of normal adult hemoglobin A. (B) Fingerprint of hemoglobin S (sickle cell hemoglobin). One histidine-containing peptide (1) is missing and a new one (2) is present. This altered peptide contains the first eight residues of the N-terminal chain of the subunit of the protein. From H. Lehmann and R. G. Huntsman, *Man's Haemoglobin*.[187]

(Chapters 11, 18). It binds to the iron of heme groups in either the $Fe^{2+}$ or $Fe^{3+}$ form and also reacts with thiol groups of proteins and small molecules to form *S*-nitrosothiols (R–S–N=O).[174–176] It reacts with the heme iron of myoglobin and hemoglobin and, by transfer of one electron, can oxidize the iron of hemoglobin to the $Fe^{3+}$ methemoglobin with formation of the nitroxyl ion $NO^-$.[177,178] This reaction may be a major cause of methemoglobin formation.

One of the major effects of NO is to induce the relaxation of smooth muscle of blood vessels, an important factor in the regulation of blood pressure. Hemoglobin can carry NO both on its heme and on the thiol group of cysteine β93.[174,175] The affinity for NO is high in the T state and low in the R state. This allows hemoglobin to carry NO from the lungs to tissues, where it can be released and participate in the regulation of blood pressure.[174,179]

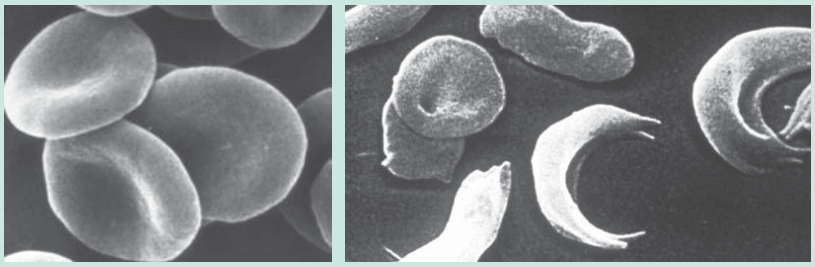
A cytoplasmic hemoglobin of the clam *Lucina pectinata* has evolved to carry oxygen to symbiotic chemoautotrophic bacteria located within cells of the host's gills. It is also readily oxidized to the $Fe^{3+}$ methemoglobin form which binds **sulfide ions** extremely tightly[180–182] and is thought to transport sulfide to the bacteria.

## 2. Abnormal Human Hemoglobins

Many alterations in the structure of hemoglobin have arisen by mutations in the human population. It is estimated that one person in 20 carries a mutation that will cause a hemoglobin disorder in a homozygote.[183] There are also many unrecognized and harmless substitutions of one amino acid for another. However, substitutions near the heme group often adversely affect the binding of oxygen and substitutions in the interfaces between subunits may decrease the cooperative interaction between subunits.[184] One of the most common and serious abnormal hemoglobins is **hemoglobin S**, which is present in individuals suffering from **sickle cell disease** (see Box 7-B). In Hb S, glutamic acid 6 of the β chain is replaced by valine. Replacement of the same amino acid by lysine leads to **Hb C**[185] and is associated with a mild disease condition . A few of the many other substitutions that have been studied are indicated in Fig. 7-23. The locations of the defects in the hemoglobin structure have been established with the aid of protein "fingerprinting" (Fig. 7-27).

A group of serious defects are represented by the **hemoglobins M**. Only heterozygotic individuals survive. Their blood is dark because in Hb M the iron in half of the subunits is oxidized irreversibly to the ferric state. The resulting methemoglobin is present in normal blood to the extent of about 1%. While normal methemoglobin is reduced by a **methemoglobin reductase** system (Box 15-H), methemoglobins M cannot be reduced. All of the five hemoglobins M result from substitutions near the heme group. In four of them, one of the heme-linked histidines (either F-8 or E-7) of either the α or the β subunits is substituted by tyrosine. In the fifth, valine 67 of the β chains is substituted by glutamate. The two hemoglobins M that carry substitutions in the α subunits ($M_{Boston}$ and $M_{Iwate}$) are frozen in the T (deoxy) conformation and therefore have low oxygen affinities and lack cooperativity.

## BOX 7-B SICKLE CELL DISEASE, MALARIA, AND BLOOD SUBSTITUTES



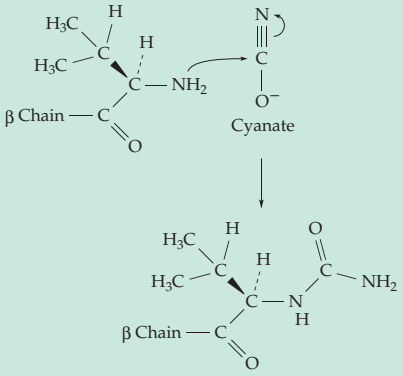
Left: Normal erythrocytes, © Biophoto Associates, Photo Researchers.
Right: Sickled erythrocytes, © Nigel Calder.

Many persons, especially if they are of west African descent, suffer from the crippling and often lethal sickle cell disease.[a,b] In 1949, Pauling, Itano, and associates discovered that hemoglobin from such individuals migrated unusually rapidly upon electrophoresis.[c] Later, Ingram devised the method of protein fingerprinting illustrated in Fig. 7-27 and applied it to hemoglobin.[d] He split the hemoglobin molecule into 15 tryptic peptides which he separated by electrophoresis and chromatography. From these experiments the abnormality in sickle cell hemoglobin (hemoglobin S; Hb S) was located at position 6 in the β chain (see Fig. 7-23). The glutamic acid present in this position in hemoglobin A was replaced by valine in Hb S. This was the first instance in which a genetic disease was traced directly to the presence of a single amino acid substitution in a specific protein. The DNA of the normal gene for the β globin chain has since been sequenced and found to have the glutamic acid codon GAG at position 6. A single base change to GTG (see Table 5-5) causes the sickle cell mutation. Persons homozygous for this altered gene have sickle cell disease, while the much more numerous heterozygotes have, at most, minor problems.

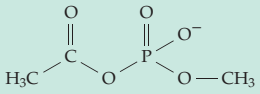
When HbS is deoxygenated it tends to "crystallize" in red blood cells, which contain 33% by weight hemoglobin. The crystallization (actually gel formation) distorts the cells into a sickle shape and these distorted corpuscles are easily destroyed, leading to anemia. The introduction of the hydrophobic valine residue in Hb S at position 6 near the end of the molecule helps form a new bonding domain by which the hemoglobin tetramers associate to form long semicrystalline microfilamentous arrays.[b,e–g]

Why is there such a high incidence of the sickle cell gene, estimated to be present in three million Americans? The occurrence and spread of the gene in Africa was apparently the result of a balance between its harmful effects and a beneficial effect under circumstances existing there. The malaria parasite, the greatest killer of all time, lives in red blood cells during part of its life cycle (see Fig. 1-9). Red cells that contain Hb S as well as Hb A are apparently less suitable than cells containing only Hb A for growth of the malaria organism. Thus, heterozygotic carriers of the sickle cell gene survived epidemics of malaria but at the price of seeing one-fourth of their offspring die of sickle cell disease.

What is the outlook for the many (50,000 in the United States alone) sufferers of sickle cell disease today? Careful medical care, including blood transfusion, can prolong life greatly[h] and intense efforts are under way to find drugs that will prevent Hb S from crystallizing.[i] The problem arises from a hydrophobic interaction of valine B6 with phenylalanine B85 and leucine B88 of another molecule in the filaments of Hb S. The latter two residues are on the outside surface of helix F (see Fig. 7-23). It is difficult to modify one of these residues chemically but various alterations at the nearby N-termini of the β chains do inhibit sickling. Cyanate does so by specifically carbamoylating these amino groups. However, although it was tested in humans, cyanate is too toxic for use.[j] Another approach employs an aldehyde that will



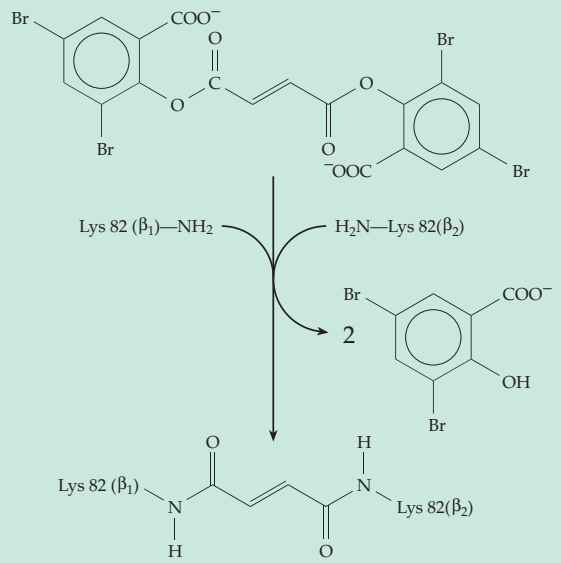

form Schiff bases (Eq. 13-4)[k] with the same amino groups.[k,l] A third approach is to use an acylating reagent. For example, methylacetyl phosphate[m] acetylates the same β Lys 82 amino groups that react with bisphosphoglycerate and with cyanate.




Aspirin (2-acetoxybenzoic acid) is also a mild acetylating reagent and "two-headed" aspirins such

**BOX 7-B (continued)**

as the following react specifically to crosslink the hemoglobin β chains.




These compounds bind into the bisphosphoglycerate binding site (Fig. 7-26) and prevent the chains from spreading apart as far as they normally do in the deoxy (T) state. Since it is only the latter that crystallizes, the compounds have a powerful antisickling action.[n,o] Various other crosslinking reagents have been developed and more than one could be used together.[p–r] New drugs that serve as allosteric modifiers in the same fashion as bisphosphoglycerate may also be useful.[s]

A fouth approach to treatment of sickle cell disease is gene therapy. This might allow patients to produce, in addition to Hb S, an engineered hemoglobin with compensating mutations that would mix with the Hb S and prevent gelling.[t,u,v] This is impractical at present but there is another approach. Persons with sickle cell disease sometimes also have the disorder of hereditary persistence of fetal hemoglobin. They continue to make Hb F into adulthood. Great amelioration of sickle cell disease is observed in patients with 20–25% Hb F.[t] Hydroxyurea stimulates a greater production of Hb F and in patients with hereditary persistence of Hb F hydroxyurea may raise its level in erythocytes to ~50% of the total hemoglobin.[w]

Crosslinking of the *alpha* chains of normal deoxyhemoglobin through lysines 99 yields a hemoglobin with normal oxygen-binding behavior and an increased stability.[o] It makes a practical emergency blood substitute, whereas unmodified hemoglobin is unsatisfactory.[v] Unless encapsulated in an erythrocyte the hemoglobin tetramers tend to dissociate to dimers, losing cooperativity and escaping through kidneys. Suitable crosslinking helps to solve this problem.[x,y,z] Both α and β chains can be produced from cloned genes and reassembled to form hemoglobin.[aa,bb] This will probably allow genetic engineering to form more stable but suitably cooperative hemoglobins that can be used to avoid hazards of transmission of viruses by transfusion.


[a] Weatherall, D. J., Clegg, J. B., Higgs, D. R., and Wood, W. G. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 3417–3484, McGraw-Hill, New York

[b] Harrington, D. J., Adachi, K., and Royer, W. E., Jr. (1998) *J. Biol. Chem.* **273**, 32690–32696

[c] Strasser, B. J. (1999) *Science* **286**, 1488–1490

[d] Ingram, V. M. (1957) *Nature (London)* **180**, 326–328

[e] Cretegny, I., and Edelstein, S. J. (1993) *J. Mol. Biol.* **230**, 733–738

[f] Padlan, E. A., and Love, W. E. (1985) *J. Biol. Chem.* **260**, 8272–8279

[g] Cao, Z., and Ferrone, F. A. (1996) *J. Mol. Biol.* **256**, 219–222

[h] Acquaye, C., Wilchek, M., and Gorecki, M. (1981) *Trends Biochem. Sci.* **6**, 146–149

[i] Klotz, I. M., Haney, D. N., and King, L. C. (1981) *Science* **213**, 724–731

[j] Harkness, D. R. (1976) *Trends Biochem. Sci.* **1**, 73–

[k] Acharya, A. S., Sussman, L. G., and Manning, J. M. (1983) *J. Biol. Chem.* **258**, 2296–2302

[l] San George, R. C., and Hoberman, H. D. (1986) *J. Biol. Chem.* **261**, 6811–6821

[m] Ueno, H., Pospischil, M. A., and Manning, J. M. (1989) *J. Biol. Chem.* **264**, 12344–12351

[n] Walder, J. A., Walder, R. Y., and Arnone, A. (1980) *J. Mol. Biol.* **141**, 195–216

[o] Chatterjee, R., Welty, E. V., Walder, R. Y., Pruitt, S. L., Rogers, P. H., Arnone, A., and Walder, J. A. (1986) *J. Biol. Chem.* **261**, 9929–9937

[p] Benesch, R. E., and Kwong, S. (1988) *Biochem. Biophys. Res. Commun.* **156**, 9–14

[q] Kluger, R., Wodzinska, J., Jones, R. T., Head, C., Fujita, T. S., and Shih, D. T. (1992) *Biochemistry* **31**, 7551–7559

[r] Jones, R. T., Shih, D. T., Fujita, T. S., Song, Y., Xiao, H., Head, C., and Kluger, R. (1996) *J. Biol. Chem.* **271**, 675–680

[s] Abraham, D. J., Wireko, F. C., Randad, R. S., Poyart, C., Kister, J., Bohn, B., Liard, J.-F., and Kunert, M. P. (1992) *Biochemistry* **31**, 9141–9149

[t] McCune, S. L., Reilly, M. P., Chomo, M. J., Asakura, T., and Townes, T. M. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9852–9856

[u] Cole-Strauss, A., Yoon, K., Xiang, Y., Byrne, B. C., Rice, M. C., Gryn, J., Holloman, W. K., and Kmiec, E. B. (1996) *Science* **273**, 1386–1388

[v] May, C., Rivella, S., Callegari, J., Heller, G., Gaensler, K. M. L., Luzzatto, L., and Sadelain, M. (2000) *Nature (London)* **406**, 82–86

[w] Eaton, W. A., and Hofrichter, J. (1995) *Science* **268**, 1142–1143

[x] Snyder, S. R., Welty, E. V., Walder, R. Y., Williams, L. A., and Walder, J. A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7280–7284

[y] Dick, L. A., Heibel, G., Moore, E. G., and Spiro, T. G. (1999) *Biochemistry* **38**, 6406–6410

[z] Manjula, B. N., Malavalli, A., Smith, P. K., Chan, N.-L., Arnone, A., Friedman, J. M., and Acharya, A. S. (2000) *J. Biol. Chem.* **275**, 5527–5534

[aa] Yamaguchi, T., Pang, J., Reddy, K. S., Witkowska, H. E., Surrey, S., and Adachi, K. (1996) *J. Biol. Chem.* **271**, 26677–26683

[bb] Jeong, S. T., Ho, N. T., Hendrich, M. P., and Ho, C. (1999) *Biochemistry* **38**, 13433–13442

In hemoglobins Rainier and Nancy the usually invariant C-terminal tyrosine 145 of the β chains is substituted by cysteine and by aspartate, respectively. Oxygen affinity is high and cooperativity is lacking.[186] Hemoglobin Kansas, in which the $\beta^{102}$ asparagine is substituted by threonine, also lacks cooperativity and has a very low oxygen affinity, while hemoglobin Richmond, in which the same amino acid is substituted by lysine, functions normally. In hemoglobin Creteil the β89 serine is replaced by asparagine with the result that the adjacent C-terminal peptide carrying tyrosine 145 becomes disordered. In hemoglobin Hiroshima the C-terminal histidine in the β chain is replaced by aspartic acid. This histidine is one that donates a Bohr proton and in the mutant hemoglobin the oxygen affinity is increased 3-fold and the Bohr effect is halved.[188] In hemoglobin Suresnes the C-terminal arginines 141 of the α chains are replaced by histidine with loss of one of the anion-binding sites mentioned in Section 3.[189]

### 3. Comparative Biochemistry of Hemoglobin

Even within human beings there are several hemoglobins. In addition to myoglobin, a brain protein neuroglobin,[189a] and adult hemoglobin A (**Hb A**, $\alpha_2\beta_2$), there is a minor hemoglobin $A_2$ ($\alpha_2\delta_2$). Prior to birth the blood contains **fetal hemoglobin**, also called hemoglobin F (**Hb F**, $\alpha_2\gamma_2$). In the presence of 2,3-bisphosphoglycerate Hb F has a 6-fold higher oxygen affinity than Hb A as befits its role in obtaining oxygen from the mother's blood.[190–192] Hemoglobin F disappears a few months after birth and is replaced by Hb A. Each of the hemoglobins differs from the others in amino acid sequence.

In other species the amino acid composition of hemoglobins varies more, as do the interactions between subunits. Hemoglobins and myoglobins are found throughout the animal kingdom and even in plants.[192a] The **leghemoglobins**[193–195] are formed in root nodules of legumes and are involved in nitrogen fixation by symbiotic bacteria. Other hemoglobins apparently function in the roots of plants.[196–197] Hemoglobins or myoglobins are found in some cyanobacteria[198] and in many other bacteria.[197,199] The globin fold of the polypeptide is recognizable in all of these.[200]

The quaternary structure of hemoglobin also varies. Myoglobin is a monomer, as is the leghemoglobin. Hemoglobin of the sea lamprey dissociates to monomers upon oxygenation.[201,201a] The clam *Scapharca inaequivalvis* has a dimeric hemoglobin that binds $O_2$ cooperatively even though the interactions between subunits are very different from those in mammalian hemoglobins.[202–204] Hemoglobin of the nematode *Ascaris* is an octamer.[205,206] It has puzzling properties, including a very high affinity for $O_2$ and a slow dissociation rate. The distal His E-7 is replaced by glutamine, which has a hydrogen-bonding ability closely similar to that of histidine.

Earthworms,[207] polychaete worms,[208,209] and leeches[210] have enormous hemoglobin molecules consisting of as many as 144 globin chains arranged into 12 dodecamers and held together by 36–42 linker chains. These hemoglobins are often called **erythrocruorins**. In a few families of polychaetes chloroheme (Fig. 16-5) substitutes for heme and the proteins are called **chlorocruorins**.[208]

What is common to all of the hemoglobins? The same folding pattern of the peptide chain is always present. The protein is always wrapped around the heme group in an identical or very similar manner. In spite of this striking conservation of overall structure, when animal hemoglobins are compared, *there are only ten residues that are highly conserved*. They are indicated in Fig. 7-23 by the boxes. The two glycines (or alanine) at B-6 and E-8 are conserved because the close contact between the B and E helices does not permit a larger side chain. Proline C-2 helps the molecule turn a corner. Four of the other conserved residues are directly associated with the heme group. Histidine E-7 and His F-8 are the "heme-linked" histidines. Tyrosine HC-2, as previously mentioned, plays a role in the cooperativity of oxygen binding. Only Lys H-9 is on the outside of the molecule. The reasons for its conservation are unclear.[211] When sequences from a broader range of organisms were determined five residues (see Fig. 7-23) were found to be highly conserved; *only two are completely conserved.* These are His F-8 and Phe CD-1, which binds the heme noncovalently.[212] Hemoglobins are not the only biological oxygen carriers. The two-iron **hemerythrins** (Fig. 16-20) are used by a few phyla of marine invertebrates, while the copper containing **hemocyanins** (Chapter 16, Section D,4) are used by many molluscs and arthropods.

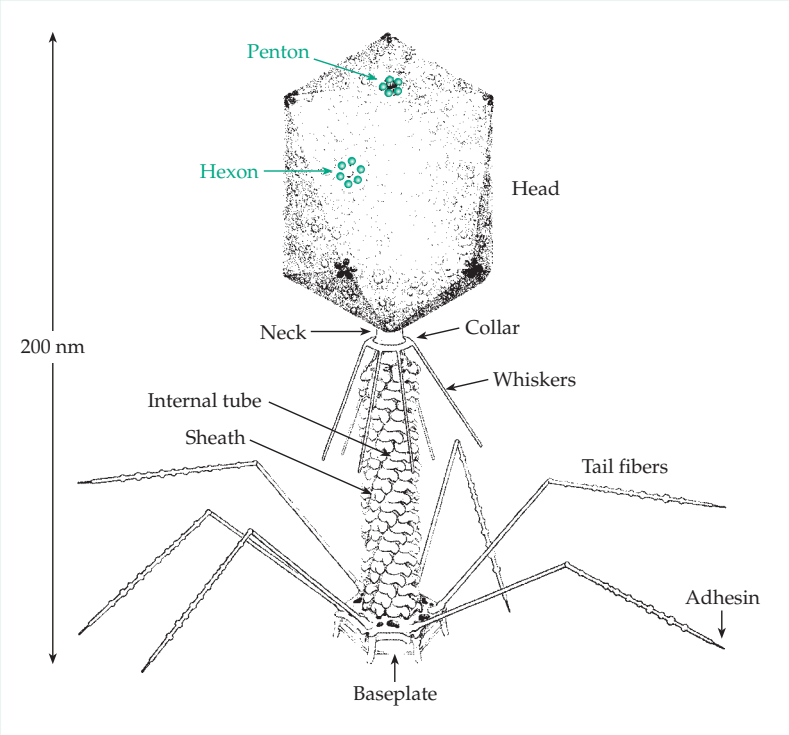
## E. Self-Assembly of Macromolecular Structures

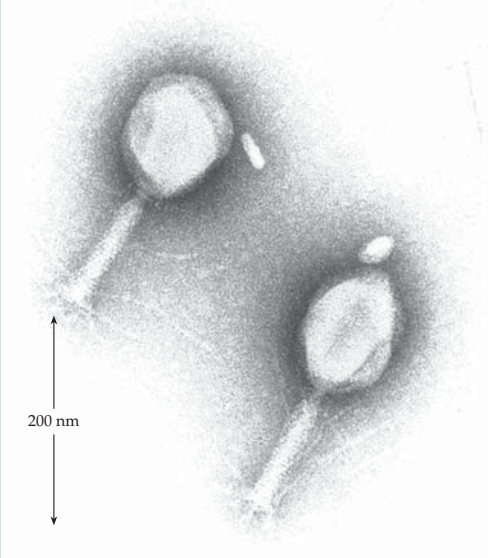
While it is easy to visualize the assembly of oligomeric proteins, it is not as easy to imagine how complex objects such as eukaryotic cilia (Fig. 1-8) or the sarcomeres of muscle (Fig. 19-6) are formed. However, study of the assembly of bacteriophage particles and other small biological objects has led to the concepts of **self-assembly** and **assembly pathways**, concepts that are now applied to every aspect of the architecture of cells.

### 1. Bacteriophages

A remarkable example of self-assembly is that of the T-even phage (Box 7-C).[213–215] From genetic analysis (Chapter 26) at least 22 genes are known to be required for formation of the heads, 21 genes for the tails, and 7 genes for the tail fibers. Many of these genes encode

## BOX 7-C THE T-EVEN BACTERIOPHAGES




Bacteriophage T4. × 240,000
Micrograph courtesy of Tom Moninger.

Drawing courtesy of F. Eiserling and the American Society for Microbiology.[a]

Among the most remarkable objects made visible by the electron microscope are the T-even bacteriophage (T2, T4, and T6) which attack *E. coli*.[a–e] While it is not often evident how a virus gains access to a cell, these "molecular syringes" literally inject their DNA through a hole dissolved in the cell wall of the host bacterium. The viruses, of length ~200 nm and mass ~225 x $10^6$ Da each, contain 130 x $10^6$ Da of DNA in a 100 x 70 nm head of elongated icosahedral shape. The head surface appears to be formed from ~840 copies of a 45-kDa protein known as gp23 (gene product 23; it is encoded by gene 23; see map in Fig. 26-2). These protein molecules are arranged as 140 hexamers (hexons) and together with ~55 copies of protein gp24, arranged as 11 pentamers (pentons), make up the bulk of the shell.[f] The head also contains at least nine other proteins, including three internal, basic proteins that enter the bacterium along with the DNA. Additional proteins form the **neck**, **collar**, and **whiskers**. The phage **tail**, which fastens to the collar via a **connector** protein,[g] contains an **internal tube** with a 2.5-nm hole, barely wide enough to accommodate the flow of the DNA molecule into the bacterium. The tube is made up of 144 subunits of gp19. The 8 x $10^6$-Da **sheath** that surrounds the tail tube is made up of 144 subunits of gp18, each of mass 55 kDa, arranged in the form of 24 rings of six subunits each.[h,i] The sheath has contractile properties. After the virus has become properly attached to the host it shortens from ~80 to ~30 nm, forcing the inner tube through a hole etched in the wall of the bacterium. At the end of the tail is a **baseplate**, a hexagonal structure bearing six short **pins**, each a trimer of a 55 kDa zinc metalloprotein.[j] One of the ten proteins known to be present in the baseplate is the enzyme T4 lysozyme (Chapter 12). The baseplate also contains six molecules of the coenzyme 7,8-dihydropteroyl-hexaglutamate (Chapter 15, Section D).

Six elongated, "jointed" **tail fibers** are attached to the baseplate. The proximal segment of each fiber is a trimer of the 1140-kDa protein gp34. A globular domain attaches it to the baseplate. The distal segment is composed of three subunits of the 109-kDa gp37, three subunits of the 23-kDa gp36, and a single copy of the 30-kDa gp35.[k] The C-terminal ~140 residues of the 1026-residue gp37 are the specific **adhesin** that binds to a lipopolysaccharide of *E. coli* type B cells or to the outer membrane protein OmpC (Chapter 8).[l] Among the smaller molecules present in the virus are the polyamines **putrescine** and **spermidine** (Chapter 24), which neutralize about 30% of the basic groups of the DNA.

How is infection by a T-even virus initiated? Binding of the tail fibers to specific receptor sites on the bacterial surface triggers a sequence of conformational changes in the fibers, baseplate, and

**BOX 7-C THE T-EVEN BACTERIOPHAGES (continued)**

sheath. The lysozyme is released from the baseplate and etches a hole in the bacterial cell wall. Contraction of the sheath is initiated at the baseplate and continues to the upper end of the sheath. The tail tube is forced into the bacterium and the DNA rapidly flows through the narrow hole into the host cell.

During contraction the subunits of the sheath undergo a remarkable rearrangement into a structure containing 12 larger rings of 12 subunits each.[h] Thus, a kind of mutual "intercalation" of subunits occurs. In its unidirectional and irreversible nature the shortening of the phage tail differs from the contraction of muscle. The protein subunits of the sheath seem to be in an unstable high energy state when the tail sheath of the phage is assembled. The stored energy remains available for later contraction.


[a] Mathews, C. K., Kutter, E. M., Mosig, G., and Berget, P. B., eds. (1983) *Bacteriophage $T_4$*, Am. Soc. Microbiology, Washington, D.C.
[b] Wood, W. B., and Edgar, R. S. (1967) *Sci. Am.* **217**(Jul), 60–74
[c] Mathews, C. K. (1971) *Bacteriophage Biochemistry*, Van Nostrand-Reinhold, Princeton, New Jersey
[d] Voyles, B. A. (1993) *The Biology of Viruses*, Mosby, St. Louis, Missouri
[e] Tikhonenko, A. S. (1970) *Ultrastructure of Bacterial Viruses*, Plenum, New York
[f] Branton, D., and Klug, A. (1975) *J. Mol. Biol.* **92**, 559–565
[g] Cerritelli, M. E., and Studier, F. W. (1996) *J. Mol. Biol.* **258**, 299–307
[h] Moody, M. F., and Makowski, L. (1981) *J. Mol. Biol.* **150**, 217–244
[i] Müller, D. J., Engel, A., Carrascosa, J. L., and Vélez, M. (1997) *EMBO J.* **16**, 2547–2553
[j] Zorzopulos, J., and Kozloff, L. M. (1978) *J. Biol. Chem.* **253**, 5543–5547
[k] Cerritelli, M. E., Wall, J. S., Simon, M. N., Conway, J. F., and Steven, A. C. (1996) *J. Mol. Biol.* **260**, 767–780
[l] Tétart, F., Repoila, F., Monod, C., and Krisch, H. M. (1996) *J. Mol. Biol.* **258**, 726–731


sequences of proteins that are incorporated into the mature virus, but several specify enzymes needed in the assembly process. Several mutant strains of the viruses are able to promote synthesis of all but one of the structural proteins of the virion. Proteins accumulating within these defective bacterial hosts have no tendency to aggregate spontaneously. However, when the missing protein (synthesized by bacteria infected with another strain of virus) is added complete virus particles are formed rapidly. Investigations resulting from this and other observations have led to the conclusion that during assembly *each different protein is added to the growing aggregate in a strictly specified sequence or assembly pathway.* The addition of each protein creates a binding site for the next protein. In some cases the protein that binds is an enzyme that cuts off a piece from the growing assembly of subunits and thereby creates a site for the next protein to bind.

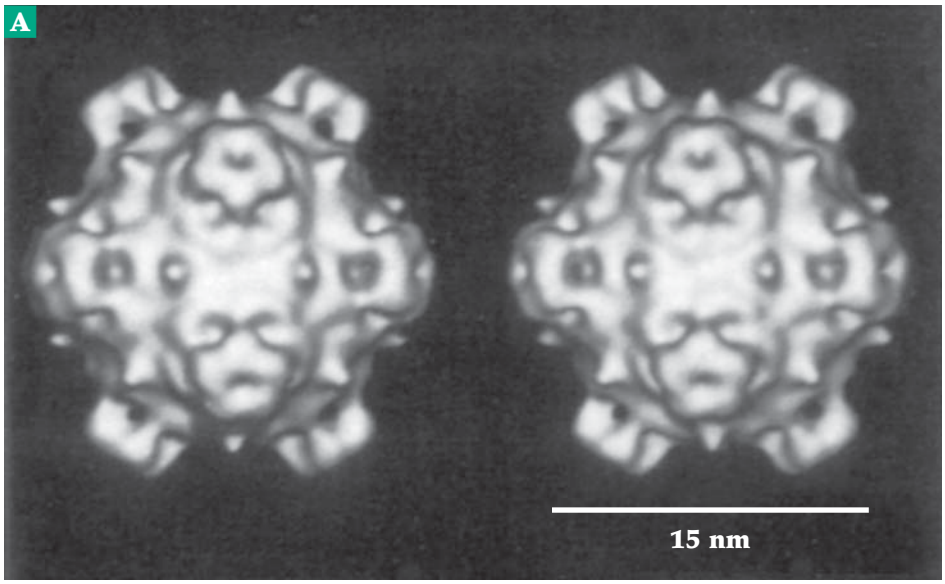
Before considering this complex process further, let's look at the assembly of simpler filamentous bacteriophages (Fig. 7-7) and bacterial pili (Fig. 7-9). The filamentous bacteriophages are put together from hydrophobic protein subunits and DNA. After their synthesis the protein subunits are stored within the cytoplasmic membrane of the infected bacteria.[216,217] These small, largely α-helical rods can easily fit within the membrane and remain there until a DNA molecule also enters the membrane.[218,219] Two additional proteins (gene I proteins) also enter the membrane. One has a 348-residue length, while the second is a 107-residue protein formed by translational initiation at a later point in the DNA sequence of the gene. These proteins help to create an assembly site at a place where the inner and outer membrane of the host bacterium are close together.[220] It isn't clear how the process is initiated, but it is likely that each subunit of the viral coat contains a nucleotide binding site that interacts with the DNA. Adjacent sides of the subunits are hydrophobic and interact with other subunits to spontaneously coat the DNA. As the rod is assembled, the hydrophobic groups become "buried." It is postulated that the remaining side chain groups on the outer surface of the virus are hydrophilic and that the formation of this hydrophilic rod provides a driving force for automatic extrusion of the phage from the membrane.[216]

Bacterial pili appear to be extruded in a similar manner. They arise rapidly and may possibly be retracted again into the bacterial membrane. The P pilus in Fig. 7-9A is made up of subunits PapA, G, F, E, and K which must be assembled in the correct sequence. A chaperonin PapD is also required as is an "**usher protein**," PapC,[50] and also the disulfide exchange protein DsbA (Chapter 10). DsbA helps PapD to form the correct disulfide bridges as it folds and PapD binds and protects the various pilus subunits as they accumulate in the periplasmic space of the host. The usher protein displaces the chaperonin PapD and "escorts" the subunits into the membrane where the extrusion occurs.[50,55]

Because eicosahedra are regular geometric solids and the faces can be made up of hexons and pentons of identical subunits, it might seem that self-assembly of eicosahedral viruses would occur easily. However, the subunits usually must be able to assume three or more different conformations and the shells can easily be assembled incorrectly. Several stategies are em-

ployed to avert this problem.[221,222] Some viruses assemble an empty shell into which the DNA flows, but many others first form an internal **scaffolding** or **assembly core** around which the shell is assembled. An external scaffold may also be needed.[223] The RNA virus MS2 forms its T=3 capsid by using the RNA molecule as the assembly core.[224] Other viruses may have one or more core proteins which dissociate from the completed shell or are removed by the action of proteases. This is a feature of the small ϕX phage (Fig. 7-28),[225,226,230] the tailed phages,[227,228] and double-stranded RNA viruses including human reoviruses.[229] Bacteriophage PRD1, another virus of *E. coli* and *Salmonella typhimurium*, has a membrane inside the capsid apparently playing a role in assembly.[232]

A general concept that seems to hold in all cases is one of "**local rule**." Several conformers of a virus subunit may equilibrate within a cell. However, *they*



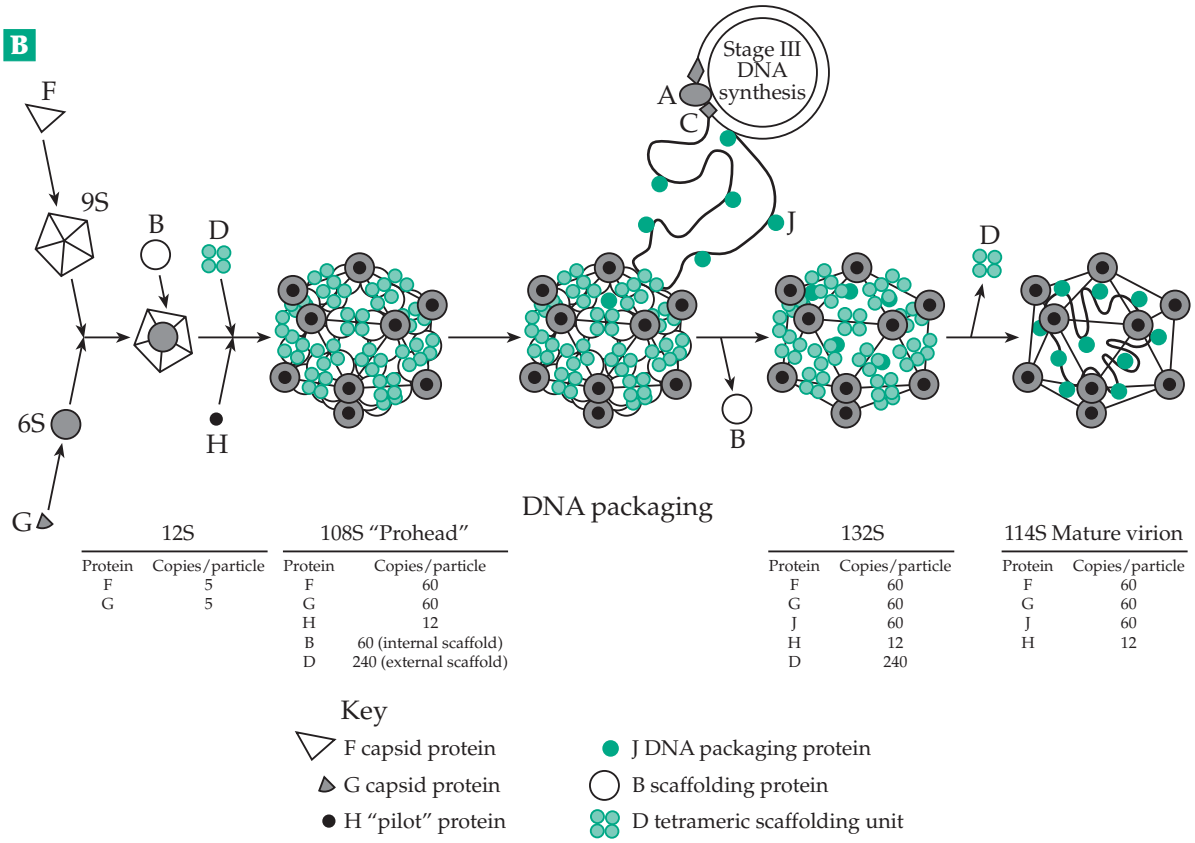




**Figure 7-28** (A) Stereoscopic view of the ϕX174 114 S mature virion viewed down a twofold axis after a cryoelectron microscopy reconstruction. From McKenna *et al*.[225] (B) Morphogenesis of ϕX174 (based on a report of Hayashi *et al*.[231]). Proteins A and C are required for DNA synthesis. Drawing from McKenna *et al*.[225] Courtesy of Michael G. Rossmann.

*can associate only through surfaces that are complementary.* A conformer that allows pentons to form cannot assemble into a hexon and only certain combinations of other conformers can give rise to hexons, etc.[233] If there is only one conformation and the shape is right a T=1 shell will be formed. If there are three conformers a T= 3 shell may arise. Another generalization is that in most cases the **procapsid** or **prohead** that is formed initially is fragile. Subunits may still be undergoing conformational changes. However, a final conformational alteration, which may include chain cleavage by a protease, usually occurs. This often expands the overall dimensions of the capsid and creates new intersubunit interactions which greatly strengthen the mature capsid.[234] Scaffolding proteins are then removed and DNA or RNA enters the capsid, again in a precise sequence. There are many variations and the detail that is known about virus assembly is far too great to describe here.

Figure 7-29 illustrates the assembly pathway for the very small ϕX174, a T=1 virus. The major capsid protein F is a 426-residue eight-stranded β-barrel. The 175-residue G protein forms pentameric spikes while 60 copies of the internal scaffolding protein and 240 copies of the external scaffolding protein D and 12 copies of the pilot protein are required to form the prohead. The single-stranded DNA enters along with 60 copies of a DNA packaging protein J.

Assembly of the tailed bacteriophages (Box 7-C) is even more complex. The genome of the viruses is large. The 166-kb circular dsDNA of phage T4 contains ~250 genes, many of which encode proteins of the virion or enzymes or chaperonins needed in assembly. The assembly pathway for the bacteriophage heads requires at least 22 gene products.[235] Seven of these form the assembly core which serves as a scaffolding around which the 840 copies of gp23 and 55 copies of gp24 (see Box 7-C) are added to give the elongated icosahedral prohead I. Most of the internal proteins are then dissolved by proteases, one of which is the phage-specified gp21. A protease also cuts a piece from each molecule of gp23 to form the smaller gp23*, the major protein of the mature **prohead II**. This cleavage also triggers the conformational change leading to head expansion. The empty proheads are now filled with DNA in a process which is assisted by another series of catalytic proteins.

The T4 phage tail is assembled in a separate sequence. Six copies of each of three different proteins form a "hub" with hexagonal symmetry (Fig. 7-29). In another assembly sequence, seven different proteins form wedge-shaped pieces, six of which are then joined to the hub to form the hexagonal baseplate. Two more proteins then add to the surface of the base plate and activate it for the growth of the internal tail tube. Only after assembly of the internal tube begins does the sheath

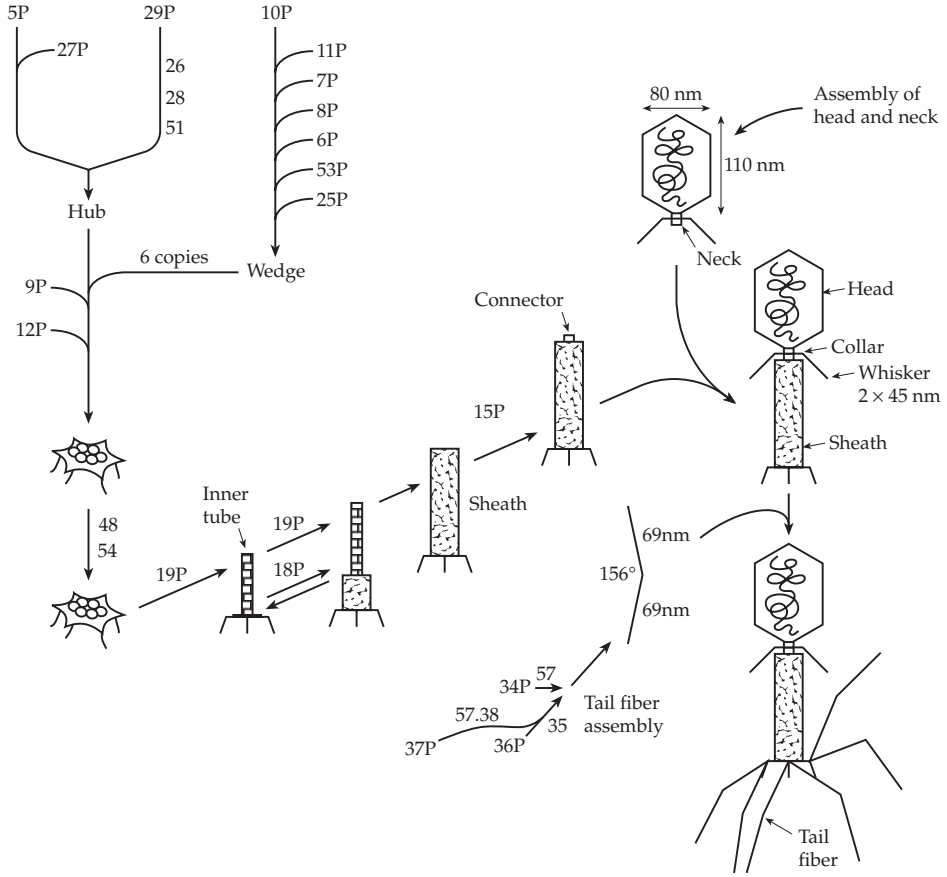



**Figure 7-29** Assembly sequence for bacteriophage T4 with details for the tail. The numbers refer to the genes in the T4 chromosome map (Fig. 26-2). A "P" after the number indicates that the protein gene product is incorporated into the phage tail. Other numbers indicate gene products that are thought to have essential catalytic functions in the assembly process. Adapted from King and Mykolajewycz[236] and Kikuchi and King.[214]

begin to grow, and only when both of these tubular structures have reached the correct length, is a cap protein placed on top. The DNA-filled head is then attached by a special connector and only then do the tail fibers, which have been assembled separately, join at the opposite end.

How can each step in this complex assembly process set the stage for the next step? Apparently the structure of each newly synthesized protein monomer is stable only until a specific interaction with another protein takes place. The binding energy of this interaction is sufficient to induce a conformational alteration that affects a distant part of the protein surface and generates complementarity toward a binding site on the next protein that is to be added. Every one of the baseplate proteins must have such self-activating properties! Sometimes proteolytic cleavage of a subunit is required. If it occurs at an appropriate point in the sequence it provides thermodynamic drive for the assembly process.

The induction of a change in one protein by interaction with another protein is a phenomenon that is met also in the construction of microtubules, ribosomes, cilia, and myofibrillar assemblies of muscle. It is basic to the assembly of the many labile but equally real cascade systems of protein–protein interactions such as that involved in the clotting of blood (Chapter 12) and signaling at membrane surfaces.

## 2. "Kringles" and Other Recognition Domains and Motifs

The assemble of either transient or long lasting complexes of proteins is often dependent upon the presence of conserved structural domains of 30–100 residues. A similar domain may occur in many different proteins and often two or more times within a single protein. The sequences within such a domain are homologous, allowing it to be recognized from protein or gene sequences alone.[237,238] Domains are often named after the protein in which they were first discovered. For example, **EGF-like domains** resemble the 53-residue epidermal growth factor. **SH2** and **SH3** domains are src-homology domains, named after **Src** (c-src), the protein encoded by the *src* protooncogene (Table 11-3).[239] The SH2 and SH3 domains are found near the N terminus of this 60-kDa protein. They are also found in many other proteins. An adapter protein called **Grb2**, important in cell signaling, consists of nothing but one SH2 domain and two SH3 domains (Figs. 11-13, 11-14).[240] The SH2 domains bind to phosphotyrosyl side chains of various proteins, while SH3 domains bind to a polyproline motif. Another phosphotyrosyl binding domain, the plekstrin homology or **PH** domain, is named for the protein in which it was discovered. **Kringle** and **apple** describe the appearances of the folded proteins in those domains. Structural domains often function to hold two proteins together or to help anchor them at a membrane surface by binding to specific protein groups, such as phosphotyrosyl, or calcium ions. Table 7-3 lists a few well-known folding domains and Fig. 7-30 shows three-dimensional structures of two of them.

Recognition domains often function transiently. For example, SH2 domains are often found in proteins that interact with phosphotyrosyl groups of "activated" cell surface receptors. The receptors become activated

**TABLE 7-3**
**A Few Well-Known Structural Domains**

| Name | Length in amino acid residues | Specific ligands |
|---|---|---|
| EGF-like[241–244] | ~45 | $Ca^{2+}$ |
| SH2[238,245–247] | ~100 | Phosphotyrosine |
| Structure[248–253] | | |
| SH3[239,246,254] | ~60 | Polyproline, PXXP |
| Structure[255–257] | | |
| PTB[238,258,259] | | Proline-rich sequence |
| PH (plekstrin homology)[260,261] | | Phosphotyrosine |
| Structure[262,263] | | |
| PDZ[264–266] | 80–100 | C-terminal XS / TXV–$COO^-$ |
| Immunoglobulin repeat (Fig. 2-16)[267] | ~100 | |
| Kringles, blood clotting proteins[268–270] | 80-85 | Calcium binding |
| Apple, Blood clotting Factor X[271] | 90 | Calcium binding |
| WW (Trp–Trp)[272] | ~38 | Proline |
| Serine protease[273] | | |
| P (Trefoil)[274] | ~50 | |
| TPR (Tetratrico peptide repeat)[247,275,276] | | |
| ZBD (Zinc-binding domains) | | |
| Zinc finger (Fig. 5-37)[277] | | |
| Others[277–280] | | |

by conformational alteration resulting from the binding. The src protein is a tyrosine kinase and, when activated, uses ATP to phosphorylate tyrosyl groups of other proteins, and using its SH2 domains it will bind to such groups forming and passing an intracellular message to them.[249]

## F. The Cytoskeleton

The cytoplasm of eukaryotic cells contains a complex network of slender rods and filaments that serve as a kind of internal skeleton. The properties of this **cytoskeleton** affect the shape and mechanical properties of cells. For example, the cytoskeleton is responsible

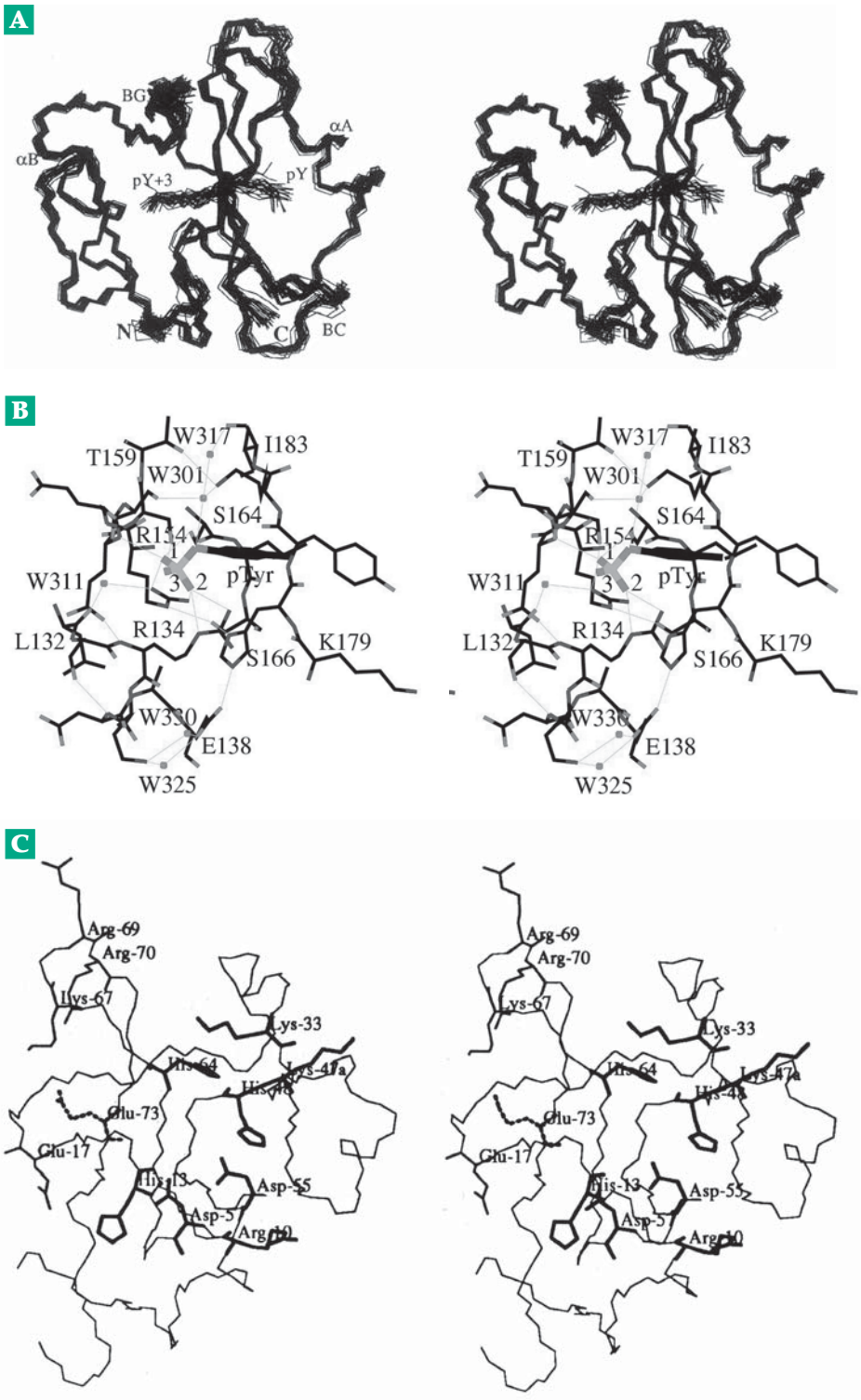



**Figure 7-30** (A) Stereoscopic MolScript view showing 30 superposed solution structures of the SH2 / phosphopeptide complex from protein Shc calculated from NMR data. The N and C termini of the protein as well as the phosphotyrosine (pY) and (pY + 3) residues of the phosphopeptide are indicated. From Zhou *et al*.[252] Courtesy of Stephen W. Fesik. (B) View of the phosphotyrosine side chain of the peptide pYEEI bound to a high-affinity SH2 domain from the human *src* tyrosine kinase called p56$^{lck}$. The phosphate group forms a series of hydrogen bonds with groups in the protein and with water molecules (small dots) and an ion pair with the guanidinium group of R134. From Tong *et al*.[248] Courtesy of Liang Tong. (C) Structure of kringle 2 from human tissue plasminogen activator (see Chapter 12). From de Vos *et al*.[269] Courtesy of Abraham M. de Vos.

for the biconcave disc shape of erythrocytes and for the ameba's ability to rapidly interconvert gel-like and fluid regions of the cytoplasm.[281–283]

Three principal components of the cytoskeleton are **microfilaments** of ~6 nm diameter, **microtubules** of 23–25 nm diameter, and **intermediate filaments** of ~10 nm diameter. A large number of associated proteins provide for interconnections, for assembly, and for disassembly of the cytoskeleton. Other proteins act as **motors** that provide motion. One of these motors is present in **myosin** of muscle. This protein is not only the motor for muscular work but also forms **thick filaments** of 12–16 nm diameter, which are a major structural component of muscle (see Fig. 19-6).

## 1. Intermediate Filaments

In most cells the intermediate filaments provide the scaffolding for the cytoskeleton[284–286] They may account for only 1% of the protein in a cell but provide up to 85% of the protein in the tough outer layers of skin. Intermediate filament proteins are encoded by over 50 human genes[286] which specify proteins of various sizes, structures, and properties. However, all of them have central 300- to 330-residue α-helical regions through which the molecules associate in parallel pairs to form coiled-coil rods with globular domains at the ends (Fig. 7-31). Some of these proteins, such as the **keratin** of skin, are insoluble. Others, including the nuclear **lamins** (Chapter 27)[287] and **vimentin**,[288–289a] dissociate and reform filaments reversibly.

Vimentin is found in most cells and predominates in fibroblasts and other cells of mesenchymal origin. **Desmin** (55-kDa monomer) is found in both smooth and skeletal muscle.[289b,290,290a] In the latter, it apparently ties the contractile myofibrils to the rest of the cytoskeletal network and the individual myofibrils to each other at Z disc (see Fig. 19-6). The **glial filaments** from the astroglial cells of the brain are composed mainly of a single type of 55-kDa subunits of the **glial fibrillar acidic protein** but the **neurofilaments** of mammalian neurons are composed of three distinct subunits of 68-, 150-, and 200-kDa mass.[291–293] The larger subunits have C-terminal tails that are not required for filament formation but which can be phosphorylated and form bridges to neighboring neurofilaments and other cytoskeletal components and organelles. Keratin filaments, which eventually nearly fill the highly differentiated epidermal cells, are also made up of several different subunits.[294] Extensions of the keratin chains are rich in cysteine side chains which form disulfide crosslinkages to adjacent molecules to provide a network that can be dehydrated to form hair and the tough outer layers of skin.[286] Elastin-associated microfibrils are important constituents of elastic tissues of blood vessels, lungs, and skin.[295]

A common architecture of intermediate filaments is a staggered head-to-tail and side-by-side association of pairs of the coiled-coil dimers into 2- to 3-nm protofilaments and further association of about eight protofilaments to form the 10-nm intermediate filaments.[286,290,296]

## 2. Microfilaments

The most abundant microfilaments are composed of fibrous actin (F-actin; Fig. 7-10). The **thin filaments** of F-actin are also one of the two major components of the contractile fibers of skeletal muscle. There is actually a group of closely related actins encoded by a multigene family. At least four vertebrate actins are specific to various types of muscle, while two (β- and γ-actins) are cytosolic.[298,299] Actins are present in all animal cells and also in fungi and plants as part of the cytoskeleton. The microfilaments can associate to

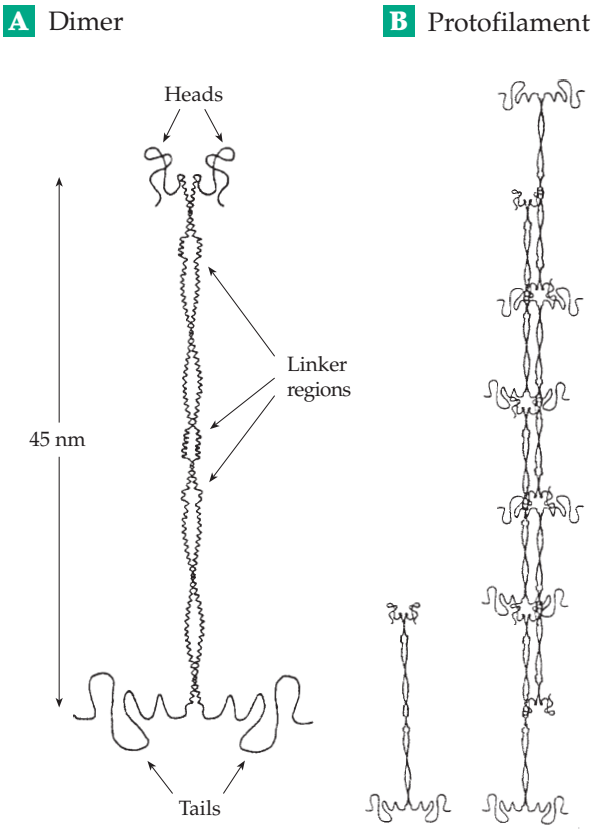



**Figure 7-31** A model for the structure of keratin microfibrils of intermediate filaments. (A) A coiled-coil dimer, 45-nm in length. The helical segments of the rod domains are interrupted by three linker regions. The conformations of the head and tail domains are unknown but are thought to be flexible. (B) Probable organization of a protofilament, involving staggered antiparallel rows of dimers. From Jeffrey A. Cohlberg[297]

form larger arrays and actin often exists as thicker "cables," some of which form the **stress fibers** seen in cultured cells adhering to a glass surface. In the red blood cells the **spectrin–actin** meshwork (Fig. 8-14), which lies directly beneath the plasma membrane, together with the proteins that anchor it to the membrane, form the cytoskeleton.[284,300,301] Its mechanical properties appear to be responsible for the biconcave disc shape of the cell.

The **acrosomal process** of some invertebrate sperm cells is an actin cable that sometimes forms almost instantaneously by polymerization of the actin monomers and shoots out to penetrate the outer layers of the egg during fertilization (Chapter 32). The **stereocilia**, the "hairs" of the hair cells in the inner ear, contain bundles of actin filaments.[302] Motion of the stereocilia caused by sound produces changes in the membrane potential of the cells initiating a nerve impulse. In certain lizards each hair cell contains about 75 stereocilia of lengths up to 30 μm and diameter 0.8 μm and containing more than 3000 actin filaments in a semicrystalline array. Microvilli (Fig. 1-6) contain longitudinal arrays of actin filaments.

In every instance, groups of microfilaments are held together by other proteins. Stress fibers of higher eukaryotes contain the "muscle proteins" **tropomyosin**, **α-actinin**, and myosin (Chapter 19), although the latter is usually not in fibrillar form. **Filamin** (250-kDa) and a 235-kDa protein are associated with actin in platelets.[303] The high-molecular-weight **synemin** crosslinks vimentin and desmin filaments,[304] while the smaller, highly polar **filaggrin** provides a matrix around the keratin filaments in the external layers of the skin.[305]

Postsynthetic modifications of cytoskeletal microfilaments can also occur. For example, epidermal keratin has been found to contain lanthionine, (γ-glutamyllysine) and lysinoalanine, both presumably arising from crosslinkages.[306]

## 3. Microtubules

A prominent component of cytoplasm consists of microtubules which appear under the electron microscope to have a diameter of 24 ± 2 nm and a 13 - to 15-nm hollow core.[307–310] However, the true diameter of a hydrated microtubule is about 30 nm and the microtubule may be further surrounded by a 5–20 nm low density layer of associated proteins. Microtubules are present in the most striking form in the flagella and cilia of eukaryotic cells (Fig. 1-8). The **stable microtubules** of cilia are integral components of the machinery causing their motion (Chapter 19). **Labile microtubules**, which form and then disappear, are often found in cytoplasm in which motion is taking place, for example, in the pseudopodia of the ameba. The mitotic spindle consists largely of microtubules which function in the movement of chromosomes in a dividing cell (Box 7-D and Chapter 26).

Microtubules in the long axons of nerve cells function as "rails" for the "fast transport" of proteins and other materials from the cell body down the axons. In fact, microtubules appear to be present throughout the cytoplasm of virtually all eukaryotic cells (Fig. 7-32) and also in spirochetes.[311] Motion in microtubular systems depends upon motor proteins such as **kinesin**, which moves bound materials toward what is known as the "negative" end of the microtubule,[312] **dyneins** which move toward the positive end.[310] These motor proteins are driven by the Gibbs energy of hydrolysis of ATP or GTP and in this respect, as well as in some structural details (Chapter 19), resemble the muscle protein myosin. Dynein is present in the arms of the microtubules of cilia (Fig. 1-8) whose motion results from the sliding of the microtubules driven by the action of this protein (Chapter 19).

Microtubules are assembled from ~ 55-kDa **tubulins**, which are mixed dimers of α subunits (450 residues) and β subunits (445 residues) with 40% sequence identity. The αβ dimers, whose structure is shown in

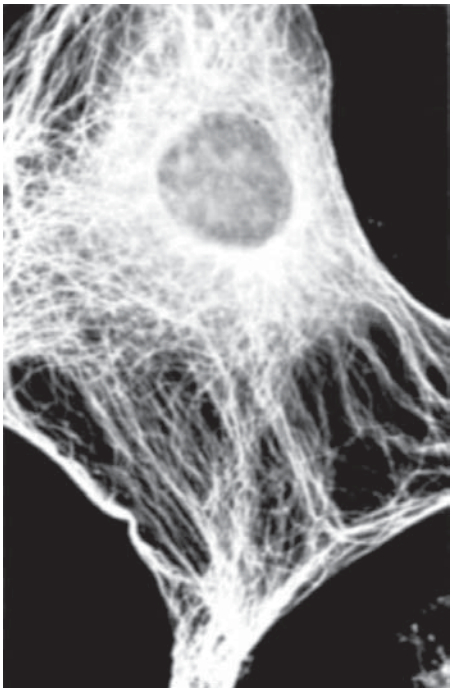
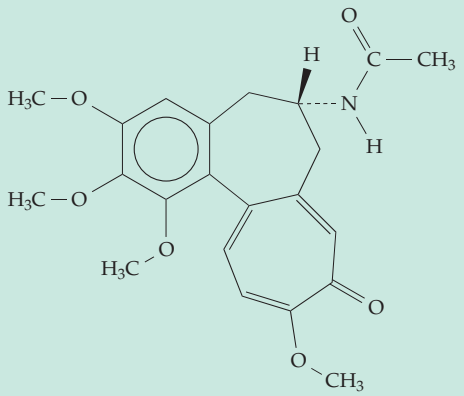


**Figure 7-32** Micrograph of a mouse embryo fibroblast was obtained using indirect immunofluorescence techniques.[313] The cells were fixed with formaldehyde, dehydrated, and treated with antibodies (formed in a rabbit) to microtubule protein. The cells were then treated with fluorescent goat antibodies to rabbit γ-globulins (see Chapter 31) and the photograph was taken by fluorescent light emission. Courtesy of Klaus Weber.

## BOX 7-D MITOSIS, TETRAPLOID PLANTS, AND ANTICANCER DRUGS

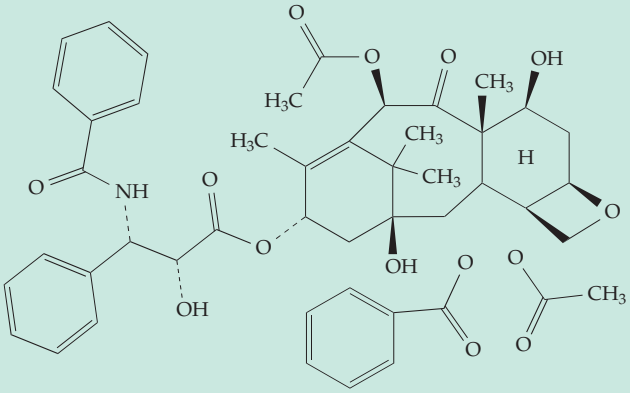
Microtubules in cells undergoing mitosis are the target of several important drugs. One of these is the alkaloid **colchicine** which is produced by various members of the lily family and has been used since ancient Egyptian times for the alleviation of the symptoms of gout.[a,b]



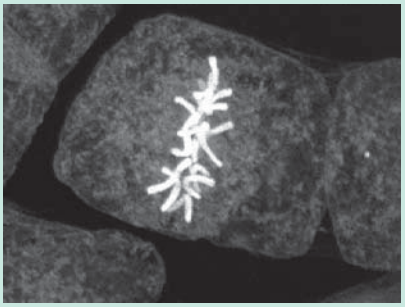

Colchicine

This compound, with its tropolone ring system, binds specifically and tightly and prevents assembly of microtubules, including those of the mitotic spindle.[b] Colchicine forms a complex with soluble tubulin,[c] perhaps a dimeric αβ complex of the two subunits.[d] Dividing cells treated with colchicine appear to be blocked at metaphase (Chapter 26) and daughter cells with a high degree of polyploidy are formed. This has led to the widespread use of colchicine in inducing formation of tetraploid varieties of flowering plants. Similar effects upon microtubules are produced by the antitumor agents **vincristine** and **vinblastine**, alkaloids formed by the common plant *Vinca* (periwinkle), and also by a variety of other drugs.[e]

The more recently discovered Taxol (paclitaxel) was extracted from the bark of the Pacific yew.[f] It *stabilizes* microtubules, inhibiting their disassembly.[g] Taxol also blocks mitosis and causes the cells which fail to complete mitosis to die. Taxol has been synthesized[f] and is a promising drug that is being used in treatment of breast, ovarian, and other cancers.[f] Binding sites for the compound have been located in β tubulin subunits (Fig. 7-33).[h,i] Attempts are being made to develop "taxoids" and other drugs more effective than taxol against cancer cells.[j,k]




Taxol (paclitaxel)



Laser scanning confocal micrograph of chromosomes at metaphase. Courtesy of Tom Moninger

Another group of drugs that bind to microtubules are **benzimidazole** and related compounds. These have been used widely to treat infection by parasitic nematodes in both humans and animals. Unfortunately resistance has developed rapidly. In a nematode that infects sheep a single tyrosine to phenylalanine mutation at position 200 in the β-tubulin subunit confers resistance.[l]

---

[a] Margulis, T. N. (1974) *J. Am. Chem. Soc.* **96**, 899–902
[b] Chakrabarti, G., Sengupta, S., and Bhattacharyya, B. (1996) *J. Biol. Chem.* **271**, 2897–2901
[c] Panda, D., Daijo, J. E., Jordan, M. A., and Wilson, L. (1995) *Biochemistry* **34**, 9921–9929
[d] Shearwin, K. E., and Timasheff, S. N. (1994) *Biochemistry* **33**, 894–901
[e] Hastie, S. B., Williams, R. C., Jr., Puett, D., and Macdonald, T. L. (1989) *J. Biol. Chem.* **264**, 6682–6688
[f] Nicolaou, K. C., Nantermet, P. G., Ueno, H., Guy, R. K., Couladouros, E. A., and Sorensen, E. J. (1995) *J. Am. Chem. Soc.* **117**, 624–633
[g] Derry, W. B., Wilson, L., and Jordan, M. A. (1995) *Biochemistry* **34**, 2203–2211
[h] Rao, S., He, L., Chakravarty, S., Ojima, I., Orr, G. A., and Horwitz, S. B. (1999) *J. Biol. Chem.* **274**, 37990–37994
[i] Makowski, L. (1995) *Nature (London)* **375**, 361–362
[j] Nicolaou, K. C., Guy, R. K., and Potier, P. (1996) *Sci. Am.* **274**(Jun), 94–98
[k] Kowalski, R. J., Giannakakou, P., and Hamel, E. (1997) *J. Biol. Chem.* **272**, 2534–2541
[l] Kwa, M. S. G., Veenstra, J. G., Van Dijk, M., and Roos, M. H. (1995) *J. Mol. Biol.* **246**, 500–510

Fig. 7-33[314–316] are thought to be packed into an imperfect helix as indicated in Fig. 7-34. The structure can also be regarded as an array of longitudinal protofilaments. Naturally formed microtubules usually have precisely 13 protofilaments and a discontinuity in the helical stacking of subunits as shown in Fig. 7-34. When grown in a laboratory the microtubules usually have 14 protofilaments[317] and rarely 10 or 16 protofilaments with regular helical packing.[318] Microtubules of some moths and also of male germ cells of *Drosophila* have 16-protofilament microtubules without a discontinuity, an architecture that is specified by the geometry of a specific β-tubulin isoform.[319]

Each tubulin dimer binds one molecule of GTP strongly in the α subunit and a second molecule of GTP or GDP more loosely in the β subunit. In this respect, tubulin resembles actin, whose subunits are about the same size. However, there is little sequence similarity. Labile microtubules of cytoplasm can be formed or disassembled very rapidly. GTP is essential for the fast growth of these microtubules and is hydrolyzed to GDP in the process.[320] However, nonhydrolyzable analogs of GTP, such as the one containing the linkage P–$CH_2$–P between the terminal and central phosphorus atoms of the GTP, also support polymerization.[321] Since microtubules have a distinct **polarity**, the two ends have different tubulin surfaces exposed, and polymerization and depolymerization can occur at different rates at the two ends. As a consequence, microtubules often grow at one end and disassemble at the other. Such **"treadmilling"** may be important in movement of chromosomes in neuronal migration[322] and in fast axonal transport of macromolecules (Chapter 30).[323] During mitosis the minus ends of the microtubules are believed to be tightly anchored at the centrosome while subunit exchanges can occur at the plus ends[323,324] where the β subunits are exposed. Using a phage display system (see Fig. 3-16) it could be shown that the N termini of the α subunits are exposed at the minus ends.[325] Kinesin can bind to the β subunits all along the microtubule.[326] Microtubules are formed by growth from microtubule nucleation sites in **microtubule organizing centers** found in centrosomes, spindle poles, and other locations.[327] Several proteins, including **γ-tubulin**, are required.[317,328,329] A proposed assembly pathway is illustrated in Fig. 7-34.

Isolated microtubules always contain small amounts of larger ~300-kDa **microtubule-associated proteins** (MAPS).[330] These elongated molecules may in part lie in the grooves between the tubulin subunits and in part be extended outward to form a low-density layer around the tubule.[283,309] Nerve cells that contain stable microtubules have associated stabilizing proteins.[331] A family of proteins formed by differential splicing of mRNA are known as **tau**. The tau proteins are prominent components of the cytoskeleton of neurons. They not only interact with microtubules but also undergo reversible phosphorylation. Hyperphosphorylated tau is the primary component of the paired helical filaments found in the brains of persons with Alzheimer disease.[330]

**Figure 7-33** Stereoscopic ribbon diagram of the tubulin dimer with α-tubulin with bound GTP at the top and β-tubulin with bound GDP at the bottom. The β-tubulin subunit also contains a bound molecule of taxotere (see Box 7-D) which is labeled TAX. This model is based upon electron crystallography of zinc-induced tubulin sheets at 0.37-nm resolution and is thought to approximate closely the packing of the tubulin monomers in microtubules.[315] The arrow at the left points toward the plus end of the microtubule. Courtesy of Kenneth H. Downing.

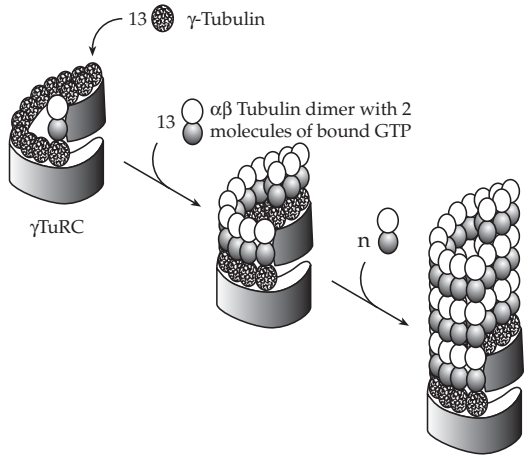



**Figure 7-34** Growth of a microtubule from a γ-tubulin ring complex (γTURC). The helical γ-tubulin rings are formed in the microtubule organizing centers which, in animal cells, are the centrosomes. Thirteen γ-tubulin subunits are shown in a hypothetical array formed together with a base of other molecules of unknown structure. The microtubule grows by addition of successive layers of α/β-tubulin dimers, each a split ring of 13 dimers with the β-tubulin subunits toward the base, the **negative end**, and the α-tubulin subunits toward the growing **positive end**. After Zheng *et al*.[317]

## References


1. Williams, D. H., Searle, M. S., Mackay, J. P., Gerhard, U., and Maplestone, R. A. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1172–1178
2. Weber, G. (1992) *Protein Interactions*, Chapman and Hall, London
3. Scatchard, G. (1949) *Ann. N.Y. Acad. Sci.* **51**, 660–672
4. Zierler, K. (1989) *Trends Biochem. Sci.* **14**, 314–317
5. Munson, P. J., Rodbard, D., and Klotz, I. M. (1983) *Science* **220**, 979–981
6. Deranleau, D. A. (1969) *J. Am. Chem. Soc.* **91**, 4044–4049
7. Dowd, J. E., and Riggs, D. S. (1965) *J. Biol. Chem.* **240**, 863–869
8. Feldman, H. A. (1983) *J. Biol. Chem.* **258**, 12865–12867

8a. Dunford, H. B. (1984) *J. Chem. Educ.* **61**, 129–132

8b. Klotz, I. M. (1997) *Ligand-Receptor Energetics*, Wiley, New York

9. Conners, K. A. (1987) *Binding Constants*, Wiley, New York
10. Motulsky, H. J., and Ransnas, L. A. (1987) *FASEB J.* **1**, 365–374
11. Schejter, A., and Margoliash, E. (1985) *Trends Biochem. Sci.* **10**, 490–492
12. Steinhardt, J., and Reynolds, J. A. (1969) *Multiple Equilibria in Proteins*, Academic Press, New York
13. Koshland, D. E., Jr. (1970) in *The Enzymes*, 3rd ed., Vol. 1 (Boyer, P. D., ed), pp. 341–396, Academic Press, New York
14. Tanford, C. (1961) *Physical Chemistry of Macromolecules*, Wiley, New York (pp. 533–534)
15. Edsall, J. T., and Wyman, J. (1958) *Biophysical Chemistry*, Academic Press, New York
16. Westheimer, F. H., and Shookhoff, M. W. (1939) *J. Am. Chem. Soc.* **61**, 555–560
17. Honig, B., and Nicholls, A. (1995) *Science* **268**, 1144–1149
18. Rajasekaran, E., Jayaram, B., and Honig, B. (1994) *J. Am. Chem. Soc.* **116**, 8238–8240
19. Tanford, C. (1961) *Physical Chemistry of Macromolecules*, Wiley, New York (Chapter 8)
20. Maier, G. D., and Metzler, D. E. (1957) *J. Am. Chem. Soc.* **79**, 4386–4391
21. Metzler, D. E. (1960) in *The Enzymes*, 2nd ed., Vol. 2 (Boyer, P. D., Lardy, H., and Myrback, K., eds), pp. 295–337, Academic Press, New York
22. Applequist, J. (1977) *J. Chem. Educ.* **54**, 417–419
23. Hill, T. (1962) *Introduction to Statistical Thermodynamics*, Addison–Wesley, Reading, Massachusetts (pp. 235–241)
24. Weiss, J. N. (1997) *FASEB J.* **11**, 835–841
25. Wang, X.-G., and Engel, P. C. (1995) *Biochemistry* **34**, 11417–11422
26. Damle, V. N. (1972) *Biopolymers* **11**, 1789–816
27. Watson, J. D. (1976) *Molecular Biology of the Gene*, 3rd ed., Benjamin, Menlo Park, California (p. 98)
28. Bernal, J. D. (1967) *J. Mol. Biol.* **24**, 379–390
29. Blundell, T. L., and Johnson, L. N. (1976) *Protein Crystallography*, Academic Press, New York
30. Monod, J., Wyman, J., and Changeux, J. D. (1965) *J. Mol. Biol.* **12**, 88–118

30a. Dill, K. A. (1997) *J. Biol. Chem.* **272**, 701–704

30b. Boresch, S., and Karplus, M. (1995) *J. Mol. Biol.* **254**, 801–807

30c. Tamura, A., and Privalov, P. L. (1997) *J. Mol. Biol.* **273**, 1048–1060

31. Zhang, R.-G., Westbrook, M. L., Westbrook, E. M., Scott, D. L., Otwinowski, Z., Maulik, P. R., Reed, R. A., and Shipley, G. G. (1995) *J. Mol. Biol.* **251**, 550–562

31a. Marvin, D. A. (1998) *Current Opinion in Structural Biology* **8**, 150–158

32. Marvin, D. A., Hale, R. D., Nave, C., and Citterich, M. H. (1994) *J. Mol. Biol.* **235**, 260–286
33. Glucksman, M. J., Bhattacharjee, S., and Makowski, L. (1992) *J. Mol. Biol.* **226**, 455–470
34. Day, L. A., Marzec, C. J., Reisberg, S. A., and Casadevall, A. (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 509–539
35. Williams, K. A., Glibowicka, M., Li, Z., Li, H., Khan, A. R., Chen, Y. M. Y., Wang, J., Marvin, D. A., and Deber, C. M. (1995) *J. Mol. Biol.* **252**, 6–14
36. Overman, S. A., Tsuboi, M., and Thomas, G. J., Jr. (1996) *J. Mol. Biol.* **259**, 331–336
37. Makowski, L. (1992) *J. Mol. Biol.* **228**, 885–892
38. Rakonjac, J., and Model, P. (1998) *J. Mol. Biol.* **282**, 25–41

38a. Welsh, L. C., Marvin, D. A., and Perham, R. N. (1998) *J. Mol. Biol.* **284**, 1265–1271

39. Liu, D. J., and Day, L. A. (1994) *Science* **265**, 671–674
40. Bhyravbhatla, B., Watowich, S. J., and Caspar, D. L. D. (1998) *Biophys. J.* **74**, 604–615
41. Fraenkel-Conrat, H. (1964) *Sci. Am.* **211**(Oct), 47–54
42. Butler, P. J. G., and Klug, A. (1978) *Sci. Am.* **239**(Nov), 62–69
43. Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R., and Klug, A. (1978) *Nature (London)* **276**, 362–368
44. Holmes, K. C. (1980) *Trends Biochem. Sci.* **5**, 4–7

44a. Eisenstein, M., Shariv, I., Koren, G., Friesem, A. A., and Katchalski-Katzir, E. (1997) *J. Mol. Biol.* **266**, 135–143

45. Raghavendra, K., Kelly, J. A., Khairallah, L., and Schuster, T. M. (1988) *Biochemistry* **27**, 7583–7588
46. Wang, H., and Stubbs, G. (1994) *J. Mol. Biol.* **239**, 371–384
47. Namba, K., Casper, D. L. D., and Stubbs, G. J. (1985) *Science* **227**, 773–776
48. Eisenstein, B. I. (1987) Escherichia coli *and* Salmonella typhimurium, Am. Soc. Microbiology, Washington, D.C., FC Niedhardt, ed. (pp. 84–90)
49. Bullitt, E., and Makowski, L. (1995) *Nature (London)* **373**, 164–167
50. Bullitt, E., Jones, C. H., Striker, R., Soto, G., Jacob-Dubuisson, F., Pinkner, J., Wick, M. J., Makowski, L., and Hultgren, S. J. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12890–12895
51. Bullitt, E., and Makowski, L. (1998) *Biophys. J.* **74**, 623–632
52. Gong, M., and Makowski, L. (1992) *J. Mol. Biol.* **228**, 735–742
53. Kuehn, M. J., Ogg, D. J., Kihlberg, J., Slonim, L. N., Flemmer, K., Bergfors, T., and Hultgren, S. J. (1993) *Science* **262**, 1234–1241

53a. Sauer, F. G., Fütterer, K., Pinkner, J. S., Dodson, K. W., Hultgren, S. J., and Waksman, G. (1999) *Science* **285**, 1058–1061

53b. Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S. J., and Knight, S. D. (1999) *Science* **285**, 1061–1065

53c. Skerker, J. M., and Shapiro, L. (2000) *EMBO J.* **19**, 3223–3234

53d. Hazes, B., Sastry, P. A., Hayakawa, K., Read, R. J., and Irvin, R. T. (2000) *J. Mol. Biol.* **299**, 1005–1017

53e. Collinson, S. K., Parker, J. M. R., Hodges, R. S., and Kay, W. W. (1999) *J. Mol. Biol.* **290**, 741–756

53f. Murphy, F. V., IV, Sweet, R. M., and Churchill, M. E. A. (1999) *EMBO J.* **18**, 6610–6618

54. Hanson, M. S., and Brinton, C. C., Jr. (1988) *Nature (London)* **332**, 265–268
55. St. Geme, J. W., III, Pinkner, J. S., Krasan, G. P., Heuser, J., and Bullitt, E. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11913–11918
56. Parge, H. E., Bernstein, S. L., Deal, C. D., McRee, D. E., Christensen, D., Capozza, M. A., Kays, B. W., Fieser, T. M., Draper, D., So, M., Getzoff, E. D., and Tainer, J. A. (1990) *J. Biol. Chem.* **265**, 2278–2285

## References


57. Parge, H. E., Forest, K. T., Hickey, M. J., Christensen, D. A., Getzoff, E. D., and Tainer, J. A. (1995) *Nature (London)* **378**, 32–38
58. Folkhard, W., Leonard, K. R., Malsey, S., Marvin, D. A., Dubochet, J., Engel, A., Achtman, M., and Helmuth, R. (1979) *J. Mol. Biol.* **130**, 145–160
59. Paiva, W. D., Grossman, T., and Silverman, P. M. (1992) *J. Biol. Chem.* **267**, 26191–26197
60. Rubenstein, P. A. (1990) *BioEssays* **12**, 309–315
61. Venkatesh, B., Tay, B. H., Elgar, G., and Brenner, S. (1996) *J. Mol. Biol.* **259**, 655–665
62. Holmes, K. C., Popp, D., Gebhard, W., and Kabsch, W. (1990) *Nature (London)* **347**, 44–49
63. Lorenz, M., Poole, K. J. V., Popp, D., Rosenbaum, G., and Holmes, K. C. (1995) *J. Mol. Biol.* **246**, 108–119
64. Valentine, R. C. (1969) in *Symmetry and Function of the Biological Systems at the Macromolecular Level* (Engstrom, A., and Strandberg, B., eds), p. 165, Wiley, New York (11th Nobel Symp.)
65. Klotz, I. M., Darnall, D. W., and Langerman, N. R. (1975) in *The Proteins*, 3rd ed., Vol. 1 (Neurath, H., and Hill, R. L., eds), p. 293, Academic Press, New York
66. Momany, C., Ernst, S., Ghosh, R., Chang, N.-L., and Hackert, M. L. (1995) *J. Mol. Biol.* **252**, 643–655
67. Almassy, R. J., Janson, C. A., Hamlin, R., Xuong, N.-H., and Eisenberg, D. (1986) *Nature (London)* **323**, 304–309
68. Harrison, S. C. (1984) *Trends Biochem. Sci.* **9**, 345–351
69. Rossmann, M. G., and Johnson, J. E. (1989) *Ann. Rev. Biochem.* **58**, 533–573
70. Johnson, J. E., and Speir, J. A. (1997) *J. Mol. Biol.* **269**, 665–675
71. Watson, J. D. (1976) *Molecular Biology of the Gene*, 3rd ed., Benjamin, Menlo Park, California (p. 107)
72. Unge, T., Liljas, L., Strandberg, B., Vaara, I., Kannan, K. K., Fridborg, K., Nordman, C. E., and Lentz, P. J., Jr. (1980) *Nature (London)* **285**, 373–377
73. Küpper, H., Keller, W., Kurz, C., Forss, S., Schaller, H., Franze, R., Stohmaier, K., Marquardt, O., Zaslavsky, V. G., and Hofschneider, P. H. (1981) *Nature (London)* **289**, 555–559
74. Trikha, J., Theil, E. C., and Allewell, N. M. (1995) *J. Mol. Biol.* **248**, 949–967
75. Larson, S. B., Day, J., Greenwood, A., and McPherson, A. (1998) *J. Mol. Biol.* **277**, 37–59
76. Hogle, J. M., Chow, M., and Filman, D. J. (1985) *Science* **229**, 1358–1365
77. Hogle, J. M., Chow, M., and Filman, D. J. (1987) *Sci. Am.* **256** (Mar), 42–49
78. Olson, N. H., Kolatkar, P. R., Oliveira, M. A., Cheng, R. H., Greve, J. M., McClelland, A., Baker, T. S., and Rossmann, M. G. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 507–511
79. Hadfield, A. T., Oliveira, M. A., Kim, K. H., Minor, I., Kremer, M. J., Heinz, B. A., Shepard, D., Pevear, D. C., Rueckert, R. R., and Rossmann, M. G. (1995) *J. Mol. Biol.* **253**, 61–73
80. Wu, H., and Rossmann, M. G. (1993) *J. Mol. Biol.* **233**, 231–244
81. Luo, M., Vriend, G., Kamer, G., Minor, I., Arnold, E., Rossmann, M. G., Boege, U., Scraba, D. G., Duke, G. M., and Palmenberg, A. C. (1987) *Science* **235**, 182–191
82. Tikhonenko, A. S. (1970) *Ultrastructure of Bacterial Viruses*, Plenum, New York
83. Birktoft, J. J., Rhodes, G., and Banaszak, L. J. (1989) *Biochemistry* **28**, 6065–6081
84. Skarzynski, T., Moody, P. C. E., and Wonacott, A. J. (1987) *J. Mol. Biol.* **193**, 171–187
85. Zimmerle, C. T., and Alter, G. M. (1993) *Biochemistry* **32**, 12743–12748
86. Hodgkin, D. C. (1975) *Nature (London)* **255**, 103
87. Whittingham, J. L., Chaudhuri, S., Dodson, E. J., Moody, P. C. E., and Dodson, G. G. (1995) *Biochemistry* **34**, 15553–15563
88. Chang, X., Jorgensen, A. M. M., Bardrum, P., and Led, J. J. (1997) *Biochemistry* **36**, 9409–9422
89. Golmohammadi, R., Valegård, K., Fridborg, K., and Liljas, L. (1993) *J. Mol. Biol.* **234**, 620–639
90. Liljas, L., Fridborg, K., Valegård, K., Bundule, M., and Pumpens, P. (1994) *J. Mol. Biol.* **244**, 279–290
91. Tars, K., Bundule, M., Fridborg, K., and Lijas, L. (1997) *J. Mol. Biol.* **271**, 759–773
92. Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D., and Tsukihara, T. (1980) *Nature (London)* **286**, 33–39
93. Athappilly, F. K., Murali, R., Rux, J. J., Cai, Z., and Burnett, R. M. (1994) *J. Mol. Biol.* **242**, 430–455
94. Stewart, P. L., Chiu, C. Y., Huang, S., Muir, T., Zhao, Y., Chait, B., Mathias, P., and Nemerow, G. R. (1997) *EMBO J.* **16**, 1189–1198
95. Caspar, D. L. D., and Klug, A. (1962) *Cold Spring Harbor Symposia on Quantitative Biology* **27**, 1–24
96. Johnson, J. E. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 27–33
97. Steven, A. C., Trus, B. L., Booy, F. P., Cheng, N., Zlotnick, A., Caston, J. R., and Conway, J. F. (1997) *FASEB J.* **11**, 733–742
98. Rayment, I., Baker, T. S., Caspar, D. L. D., and Murakami, W. T. (1982) *Nature (London)* **295**, 110–115
99. Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L., and Harrison, S. C. (1991) *Nature (London)* **354**, 278–284
100. Champness, J. N., Bloomer, A. C., Bricogne, G., Butler, P. J. G., and Klug, A. (1976) *Nature (London)* **259**, 20–24
101. Fukuda, M., and Okada, Y. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4035–4038
102. McDonald, R. C., Engelman, D. M., and Steita, T. A. (1979) *J. Biol. Chem.* **254**, 2942–2943
103. Galloway, J. (1984) *Trends Biochem. Sci.* **9**, 233–238
104. Kantrowitz, E. R., and Lipscomb, W. N. (1990) *Trends Biochem. Sci.* **15**, 53–59
105. Zhou, B.-B., and Schachman, H. K. (1993) *Protein Sci.* **2**, 103–112
106. Xi, X. G., De Staercke, C., Van Vliet, F., Triniolles, F., Jacobs, A., Stas, P. P., Ladjimi, M. M., Simon, V., Cunin, R., and Hervé, G. (1994) *J. Mol. Biol.* **242**, 139–149
107. Fetler, L., Vachette, P., Hervé, G., and Ladjimi, M. M. (1995) *Biochemistry* **34**, 15654–15660
108. Honzatko, R. B., Crawford, J. L., Monaco, H. L., Ladner, J. E., Edwards, B. F. D., Evans, D. R., Warren, S. G., Wiley, D. C., Ladner, R. C., and Lipscomb, W. N. (1982) *J. Mol. Biol.* **160**, 219–263
109. Krause, K. L., Volz, K. W., and Lipscomb, W. N. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1643–1647
110. Thomas, A., Field, M. J., and Perahia, D. (1996) *J. Mol. Biol.* **261**, 490–506

110a. Brunori, M. (1999) *Trends Biochem. Sci.* **24**, 158–161

110b. Di Cera, E., ed. (1998) *Advances in Protein Chemistry, Linkage Thermodynamics of Macromolecular Interactions*, Vol. 51, Academic Press, San Diego, California

111. Bloom, C. R., Choi, W. E., Brzovic, P. S., Ha, J. J., Huang, S.-T., Kaarsholm, N. C., and Dunn, M. F. (1995) *J. Mol. Biol.* **245**, 324–330
112. Hammes, G. G., and Wu, C.-W. (1971) *Science* **172**, 1205–1211
113. Koshland, D. E., Jr., Némethy, G., and Filmer, D. (1966) *Biochemistry* **5**, 365–385
114. Cornish-Bowden, A. J., and Koshland, D. E., Jr. (1970) *J. Biol. Chem.* **245**, 6241–6250
115. Cornish-Bowden, A. J., and Koshland, D. E., Jr. (1971) *J. Biol. Chem.* **246**, 3092–3102
116. Darnall, D. W., and Klotz, I. M. (1972) *Arch. Biochem. Biophys.* **149**, 1–14
117. Doci, Y., Sugita, Y., and Yoneyama, Y. (1973) *J. Biol. Chem.* **248**, 2354–2363
118. Perutz, M. F., Muirhead, H., Cox, J. M., and Goaman, L. C. G. (1968) *Nature (London)* **219**, 131–139
119. Dickerson, R. E., and Geis, I. (1983) *Hemoglobin: Structure, Function, Evolution and Pathology*, Benjamin / Cummings Publ., Redwood City, California
120. Antonini, E., and Brunori, M. (1971) *Hemoglobin and Myoglobin in Their Reactions with Ligands*, North-Hollard, Publ., Amsterdam
121. Kendrew, J. C. (1961) *Sci. Am.* **205**(Dec), 96–110
122. Fermi, G., Perutz, M. F., Shaanan, B., and Fourme, R. (1984) *J. Mol. Biol.* **175**, 159–174
123. Benesch, R., and Benesch, R. E. (1974) *Science* **185**, 905–908
124. Borgstahl, G. E. O., Rogers, P. H., and Arnone, A. (1994) *J. Mol. Biol.* **236**, 831–843
125. Imai, K. (1979) *J. Mol. Biol.* **133**, 233–247
126. Amiconi, G., Antonini, E., Brunori, M., Wyman, J., and Zolla, L. (1981) *J. Mol. Biol.* **152**, 111–129
127. Flanagan, M. A., Ackers, G. K., Matthew, J. B., Hanania, G. I. H., and Gurd, F. R. N. (1981) *Biochemistry* **20**, 7439–7449

127a. Shibayama, N. (1999) *J. Mol. Biol.* **285**, 1383–1388

128. Waxman, L. (1971) *J. Biol. Chem.* **246**, 7318–7327
129. Doyle, M. L., Di Cera, E., and Gill, S. J. (1988) *Biochemistry* **27**, 820–824
130. Edelstein, S. J. (1996) *J. Mol. Biol.* **257**, 737–744
131. Imai, K. (1983) *J. Mol. Biol.* **167**, 741–749

131a. Gibson, Q. H. (1999) *Biochemistry* **38**, 5191–5199

132. Paoli, M., Liddington, R., Tame, J., Wilkinson, A., and Dodson, G. (1996) *J. Mol. Biol.* **256**, 775–792
133. Henry, E. R., Jones, C. M., Hofrichter, J., and Eaton, W. A. (1997) *Biochemistry* **36**, 6511–6528
134. Silva, M. M., Rogers, P. H., and Arnone, A. (1992) *J. Biol. Chem.* **267**, 17248–17256
135. Jayaraman, V., and Spiro, T. G. (1995) *Biochemistry* **34**, 4511–4515
136. Schumacher, M. A., Zheleznova, E. E., Poundstone, K. S., Kluger, R., Jones, R. T., and Brennan, R. G. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7841–7844

136a. Tame, J. R. H. (1999) *Trends Biochem. Sci.* **24**, 372–377

137. Manning, L. R., Jenkins, W. T., Hess, J. R., Vandegriff, K., Winslow, R. M., and Manning, J. M. (1996) *Protein Sci.* **5**, 775–781
138. Holt, J. M., and Ackers, G. K. (1995) *FASEB J.* **9**, 210–218
139. Perutz, M. F., Shih, D. T.-b, and Williamson, D. (1994) *J. Mol. Biol.* **239**, 555–560
140. Perutz, M. F., Muirhead, H., Cox, J. M., Goaman, L. C. G., Mathews, F. S., McGandy, L. E., and Webb, L. E. (1968) *Nature (London)* **219**, 29–32
141. Perutz, M. (1990) *Mechanisms of Cooperativity and Allosteric Regulation in Proteins*, Cambridge Univ. Press, London
142. O'Donnell, S., Mandaro, R., Schuster, T. M., and Arnone, A. (1979) *J. Biol. Chem.* **254**, 12204–12208
143. Perutz, M. F. (1970) *Nature (London)* **228**, 726–739
144. Dickerson, R. E. (1972) *Ann. Rev. Biochem.* **41**, 815–842

# References


145. Ishimori, K., Hashimoto, M., Imai, K., Fushitani, K., Miyazaki, G., Morimoto, H., Wada, Y., and Morishima, I. (1994) *Biochemistry* **33**, 2546–2553
146. Chao, H., Sönnichsen, F. D., DeLuca, C. I., Sykes, B. D., and Davies, P. L. (1994) *Protein Sci.* **3**, 1760–1769
147. Kim, H.-W., Shen, T.-J., Ho, N. T., Zou, M., Tam, M. F., and Ho, C. (1996) *Biochemistry* **35**, 6620–6627
148. Kavanaugh, J. S., Weydert, J. A., Rogers, P. H., and Arnone, A. (1998) *Biochemistry* **37**, 4358–4373
149. Kiger, L., Klinger, A. L., Kwiatkowski, L. D., De Young, A., Doyle, M. L., Holt, J. M., Noble, R. W., and Ackers, G. K. (1998) *Biochemistry* **37**, 4336–4345
150. Hoard, J. L., and Scheidt, W. R. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3919–3922
151. Fermi, G., Perutz, M. F., and Shulman, R. G. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6167–6168
152. De Baere, I., Perutz, M. F., Kiger, L., Marden, M. C., and Poyart, C. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1594–1597
153. Bohr, C., Hasselbalch, K., and Krogh, A. (1904) *Skand. Arch. Physiol.* **16**, 402–
154. Wyman, J., Jr. (1948) *Adv. Prot. Chem.* **4**, 407–531
155. Benesch, R. E., and Benesch, R. (1974) *Adv. Prot. Chem.* **28**, 211–237
156. Fang, T.-Y., Zou, M., Simplaceanu, V., Ho, N. T., and Ho, C. (1999) *Biochemistry* **38**, 13423–13432
157. Fronticelli, C., Pechik, I., Brinigar, W. S., Kowalczyk, J., and Gilliland, G. L. (1994) *J. Biol. Chem.* **269**, 23965–23969
158. Perrella, M., Ripamonti, M., and Caccia, S. (1998) *Biochemistry* **37**, 2017–2028
159. Arnone, A. (1972) *Nature (London)* **237**, 146–149
160. Benesch, R., and Benesch, R. E. (1969) *Nature (London)* **221**, 618–622
161. Brewer, G. J., and Eaton, J. W. (1971) *Science* **171**, 1205–1211
162. Perutz, M. F., Bauer, C., Gros, G., Leclercq, F., Vandecasserie, C., Schnek, A. G., Braunitzer, G., Friday, A. E., and Joysey, K. A. (1981) *Nature (London)* **291**, 682–684
163. Komiyama, N. H., Miyazaki, G., Tame, J., and Nagai, K. (1995) *Nature (London)* **373**, 244–246
164. Fantl, W. J., Di Donato, A., Manning, J. M., Rogers, P. H., and Arnone, A. (1987) *J. Biol. Chem.* **87**, 12700–12713
165. Sage, J. T., and Jee, W. (1997) *J. Mol. Biol.* **274**, 21–26
166. Teng, T.-Y., Srajer, V., and Moffat, K. (1997) *Biochemistry* **36**, 12087–12100
166a. Ostermann, A., Waschipky, R., Parak, F. G., and Nienhaus, G. U. (2000) *Nature (London)* **404**, 205–208
167. Chance, M. R., Miller, L. M., Fischetti, R. F., Scheuring, E., Huang, W.-X., Sclavi, B., Hai, Y., and Sullivan, M. (1996) *Biochemistry* **35**, 9014–9023
168. Ösapay, K., Theriault, Y., Wright, P. E., and Case, D. A. (1994) *J. Mol. Biol.* **244**, 183–197
169. Martin, K. D., and Parkhurst, L. J. (1990) *Biochemistry* **29**, 5718–5726
170. Tetreau, C., Lavalette, D., Momenteau, M., Fischer, J., and Weiss, R. (1994) *J. Am. Chem. Soc.* **116**, 11840–11848
171. Huang, Y., and Ackers, G. K. (1995) *Biochemistry* **34**, 6316–6327
172. Paoli, M., Dodson, G., Liddington, R. C., and Wilkinson, A. J. (1997) *J. Mol. Biol.* **271**, 161–167
173. Daugherty, M. A., Shea, M. A., and Ackers, G. K. (1994) *Biochemistry* **33**, 10345–10357
174. Stamler, J. S., Jia, L., Eu, J. P., McMahon, T. J., Demchenko, I. T., Bonaventura, J., Gernert, K., and Piantadosi, C. A. (1997) *Science* **276**, 2034–2037
175. Upmacis, R. K., Hajjar, D. P., Chait, B. T., and Mirza, U. A. (1997) *J. Am. Chem. Soc.* **119**, 10424–10429
176. Miller, L. M., Pedraza, A. J., and Chance, M. R. (1997) *Biochemistry* **36**, 12199–12207
177. Gow, A. J., and Stamler, J. S. (1998) *Nature (London)* **391**, 169–173
178. Lutter, L. C., Halvorson, H. R., and Calladine, C. R. (1996) *J. Mol. Biol.* **261**, 620–633
179. Jia, L., Bonaventura, C., Bonaventura, J., and Stamler, J. S. (1996) *Nature (London)* **380**, 221–226
180. Rizzi, M., Wittenberg, J. B., Coda, A., Fasano, M., Ascenzi, P., and Bolognesi, M. (1994) *J. Mol. Biol.* **244**, 86–99
181. Rizzi, M., Wittenberg, J. B., Coda, A., Ascenzi, P., and Bolognesi, M. (1996) *J. Mol. Biol.* **258**, 1–5
182. Nguyen, B. D., Zhao, X., Vyas, K., La Mar, G. N., Lile, R. A., Brucker, E. A., Phillips, G. N., Jr., Olson, J. S., and Wittenberg, J. B. (1998) *J. Biol. Chem.* **273**, 9517–9526
183. Weatherall, D. J., Clegg, J. B., Higgs, D. R., and Wood, W. G. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 3417–3484, McGraw-Hill, New York
184. Bank, A., Mears, J. G., and Ramirez, F. (1980) *Science* **207**, 486–493
185. Hirsch, R. E., Juszczak, L. J., Fataliev, N. A., Friedman, J. M., and Nagel, R. L. (1999) *J. Biol. Chem.* **274**, 13777–13782
186. Arnone, A., Thillet, J., and Rosa, J. (1981) *J. Biol. Chem.* **256**, 8545–8552
187. Lehmann, H., and Huntsman, R. G. (1966) *Man's Haemoglobins*, North-Holland Publ. Co., Amsterdam (p. 246)
188. Olson, J. S., Gibson, Q. H., Nagel, R. L., and Hamilton, H. B. (1972) *J. Biol. Chem.* **247**, 7485–7493
189. Poyart, C., Bursaux, E., Arnone, A., Bonaventura, J., and Bonaventura, C. (1980) *J. Biol. Chem.* **255**, 9465–9473
189a. Burmester, T., Weich, B., Reinhardt, S., and Hankeln, T. (2000) *Nature* **407**, 520–523
190. Tomita, S. (1981) *J. Biol. Chem.* **256**, 9495–9500
191. Giardina, B., Scatena, R., Clementi, M. E., Cerroni, L., Nuutinen, M., Brix, O., Sletten, S. N., Castagnola, M., and Condò, S. G. (1993) *J. Mol. Biol.* **229**, 512–516
192. Clementi, M. E., Scatena, R., Mordente, A., Condò, S. G., Castagnola, M., and Giardina, B. (1996) *J. Mol. Biol.* **255**, 229–234
192a. Burr, A. H. J., Hunt, P., Wagar, D. R., Dewilde, S., Blaxter, M. L., Vanfleteren, J. R., and Moens, L. (2000) *J. Biol. Chem.* **275**, 4810–4815
193. Lee, H. C., Wittenberg, J. B., and Peisach, J. (1993) *Biochemistry* **32**, 11500–11506
194. Harutyunyan, E. H., Safonova, T. N., Kuranova, I. P., Popov, A. N., Teplyakov, A. V., Obmolova, G. V., Rusakov, A. A., Vainshtein, B. K., Dodson, G. G., Wilson, J. C., and Perutz, M. F. (1995) *J. Mol. Biol.* **251**, 104–115
195. Hargrove, M. S., Barry, J. K., Brucker, E. A., Berry, M. B., Phillips, G. N., Jr., Olson, J. S., Arredondo-Peter, R., Dean, J. M., Klucas, R. V., and Sarath, G. (1997) *J. Mol. Biol.* **266**, 1032–1042
196. Trevaskis, B., Watts, R. A., Andersson, C. R., Llewellyn, D. J., Hargrove, M. S., Olson, J. S., Dennis, E. S., and Peacock, W. J. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12230–12234
196a. Hargrove, M. S., Brucker, E. A., Stec, B., Sarath, G., Arredondo-Peter, R., Klucas, R. V., Olson, J. S., and Phillips, G. N., Jr. (2000) *Structure* **8**, 1005–1014
197. Hardison, R. C. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5675–5679
198. Potts, M., Angeloni, S. V., Ebel, R. E., and Bassam, D. (1992) *Science* **256**, 1690–1692
199. Yeh, S.-R., Couture, M., Ouellet, Y., Guertin, M., and Rousseau, D. L. (2000) *J. Biol. Chem.* **275**, 1679–1684
200. Aronson, H.-E. G., Royer, W. E., Jr., and Hendrickson, W. A. (1994) *Protein Sci.* **3**, 1706–1711
201. Honzatko, R. B., Hendrickson, W. A., and Love, W. E. (1985) *J. Mol. Biol.* **184**, 147–164
201a. Qiu, Y., Maillett, D. H., Knapp, J., Olson, J. S., and Riggs, A. F. (2000) *J. Biol. Chem.* **275**, 13517–13528
202. Condon, P. J., and Royer, W. E. J. (1994) *J. Biol. Chem.* **269**, 25259–25267
203. Royer, W. E., Jr., Heard, K. S., Harrington, D. J., and Chiancone, E. (1995) *J. Mol. Biol.* **253**, 168–186
204. Mozzarelli, A., Bettati, S., Rivetti, C., Rossi, G. L., Colotti, G., and Chiancone, E. (1996) *J. Biol. Chem.* **271**, 3627–3632
205. Yang, J., Kloek, A. P., Goldberg, D. E., and Mathews, F. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4224–4228
206. Huang, S., Huang, J., Kloek, A. P., Goldberg, D. E., and Friedman, J. M. (1996) *J. Biol. Chem.* **271**, 958–962
207. Lamy, J., Kuchumov, A., Taveau, J.-C., Vinogradov, S. N., and Lamy, J. N. (2000) *J. Mol. Biol.* **298**, 633–647
208. de Haas, F., Taveau, J.-C., Boisset, N., Lambert, O., Vinogradov, S. N., and Lamy, J. N. (1996) *J. Mol. Biol.* **255**, 140–153
209. de Haas, F., Zal, F., You, V., Lallier, F., Toulmond, A., and Lamy, J. N. (1996) *J. Mol. Biol.* **264**, 111–120
210. de Haas, F., Boisset, N., Taveau, J.-C., Lambert, O., Vinogradov, S. N., and Lamy, J. N. (1996) *Biophys. J.* **70**, 1973–1984
211. Zuckerkandl, E. (1965) *Sci. Am.* **212**(May), 110–118
212. Perutz, M. F. (1986) *Nature (London)* **322**, 405
213. Wood, W. B., and Edgar, R. S. (1967) *Sci. Am.* **217**(Jul), 60–74
214. Kikuchi, Y., and King, J. (1975) *J. Mol. Biol.* **99**, 645–672
215. Casjens, S., and King, J. (1975) *Ann. Rev. Biochem.* **44**, 555–611
216. Pollard, T. D., and Craig, S. W. (1982) *Trends Biochem. Sci.* **7**, 55–58
217. Ito, K., Date, T., and Wickner, W. (1980) *J. Biol. Chem.* **255**, 2123–2130
218. McDonnell, P. A., Shon, K., Kim, Y., and Opella, S. J. (1993) *J. Mol. Biol.* **233**, 447–463
219. Sanders, J. C., Haris, P. I., Chapman, D., Otto, C., and Hemminga, M. A. (1993) *Biochemistry* **32**, 12446–12454
220. Guy-Caffey, J. K., and Webster, R. E. (1993) *J. Biol. Chem.* **268**, 5496–5503
221. Duda, R. L., Martincic, K., and Hendrix, R. W. (1995) *J. Mol. Biol.* **247**, 636–647
222. Xie, Z., and Hendrix, R. W. (1995) *J. Mol. Biol.* **253**, 74–85
223. Marvik, C. J., Dokland, T., Nokling, R. H., Jacobsen, E., Larsen, T., and Lindqvist, B. H. (1995) *J. Mol. Biol.* **251**, 59–75
224. LeCuyer, K. A., Behlen, L. S., and Uhlenbeck, O. C. (1995) *Biochemistry* **34**, 10600–10606
225. McKenna, R., Ilag, L. L., and Rossmann, M. G. (1994) *J. Mol. Biol.* **237**, 517–543
226. Dokland, T., McKenna, R., Ilag, L. L., Bowman, B. R., Incardona, N. L., Fane, B. A., and Rossmann, M. G. (1997) *Nature (London)* **389**, 308–313
227. Thuman-Commike, P. A., Greene, B., Jakana, J., Prasad, B. V. V., King, J., Prevelige, P. E., Jr., and Chiu, W. (1996) *J. Mol. Biol.* **260**, 85–98
228. Parker, M. H., Stafford, W. F., III, and Prevelige, P. E., Jr. (1997) *J. Mol. Biol.* **268**, 655–665

## References


229. Butcher, S. J., Dokland, T., Ojala, P. M., Bamford, D. H., and Fuller, S. D. (1997) *EMBO J.* **16**, 4477–4487
230. Olson, N. H., Baker, T. S., Willingmann, P., and Incardona, N. L. (1992) *J. Structural Biol.* **108**, 168–175
231. Hayashi, M., Aoyama, A., Richardson, D. L., Jr., and Hayashi, M. N. (1988) in *The Bacteriophages* (Calendar, R., ed), pp. 1–71, Plenum, New York
232. Butcher, S. J., Bamford, D. H., and Fuller, S. D. (1995) *EMBO J.* **14**, 6078–6086
233. Berger, B., Shor, P. W., Tucker-Kellogg, L., and King, J. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7732–7736
234. Steven, A. C., Greenstone, H. L., Booy, F. P., Black, L. W., and Ross, P. D. (1992) *J. Biol. Chem.* **228**, 870–884
235. van Driel, R. (1980) *J. Mol. Biol.* **138**, 27–42
236. King, J., and Mykolajewycz, N. (1973) *J. Mol. Biol.* **75**, 339–358
237. Feller, S. M., Ren, R., Hanafusa, H., and Baltimore, D. (1994) *Trends Biochem. Sci.* **19**, 453–458
238. Pawson, T. (1995) *Nature (London)* **373**, 573–580
239. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C., and Pawson, T. (1991) *Science* **252**, 668–674
240. Maignan, S., Guilloteau, J.-P., Fromage, N., Arnoux, B., Becquart, J., and Ducruix, A. (1995) *Science* **268**, 291–293
241. Ullner, M., Selander, M., Persson, E., Stenflo, J., Drakenberg, T., and Teleman, O. (1992) *Biochemistry* **31**, 5974–5983
242. Meininger, D. P., Hunter, M. J., and Komives, E. A. (1995) *Protein Sci.* **4**, 1683–1695
243. Bersch, B., Hernandez, J.-F., Marion, D., and Arlaud, G. J. (1998) *Biochemistry* **37**, 1204–1214
244. Rand, M. D., Lindblom, A., Carlson, J., Villoutreix, B. O., and Stenflo, J. (1997) *Protein Sci.* **6**, 2059–2071
245. Müller, K., Gombert, F. O., Manning, U., Grossmüller, F., Graff, P., Zaegel, H., Zuber, J. F., Freuler, F., Tschopp, C., and Baumann, G. (1996) *J. Biol. Chem.* **271**, 16500–16505
246. Fry, M. J., Panayotou, G., Booker, G. W., and Waterfield, M. D. (1993) *Protein Sci.* **2**, 1785–1797
247. Malek, S. N., Yang, C. H., Earnshaw, W. C., Kozak, C. A., and Desiderio, S. (1996) *J. Biol. Chem.* **271**, 6952–6962
248. Tong, L., Warren, T. C., King, J., Betageri, R., Rose, J., and Jakes, S. (1996) *J. Mol. Biol.* **256**, 601–610
249. Xu, R. X., Word, J. M., Davis, D. G., Rink, M. J., Willard, D. H., Jr., and Gampe, R. T., Jr. (1995) *Biochemistry* **34**, 2107–2121
250. Mikol, V., Baumann, G., Zurini, M. G. M., and Hommel, U. (1995) *J. Mol. Biol.* **254**, 86–95
251. Pascal, S. M., Yamazaki, T., Singer, A. U., Kay, L. E., and Forman-Kay, J. D. (1995) *Biochemistry* **34**, 11353–11362
252. Zhou, M.-M., Meadows, R. P., Logan, T. M., Yoon, H. S., Wade, W. S., Ravichandran, K. S., Burakoff, S. J., and Fesik, S. W. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7784–7788
253. Metzler, W. J., Leiting, B., Pryor, K., Mueller, L., and Farmer, B. T., II. (1996) *Biochemistry* **35**, 6201–6211
254. Viguera, A. R., Arrondo, J. L. R., Musacchio, A., Saraste, M., and Serrano, L. (1994) *Biochemistry* **33**, 10925–10933
255. Liang, J., Chen, J. K., Schreiber, S. L., and Clardy, J. (1996) *J. Mol. Biol.* **257**, 632–643
256. Lim, W. A., Richards, F. M., and Fox, R. O. (1994) *Nature (London)* **372**, 375–379
257. Guruprasad, L., Dhanaraj, V., Timm, D., Blundell, T. L., Gout, I., and Waterfield, M. D. (1995) *J. Mol. Biol.* **248**, 856–866
258. van der Geer, P., and Pawson, T. (1995) *Trends Biochem. Sci.* **20**, 277–280
259. Zhou, M.-M., Ravichandran, K. S., Olejniczak, E. T., Petros, A. M., Meadows, R. P., Sattler, M., Harlan, J. E., Wade, W. S., Burakoff, S. J., and Fesik, S. W. (1995) *Nature (London)* **378**, 584–592
260. Rameh, L. E., Arvidsson, A.-k, Carraway, K. L., III, Couvillon, A. D., Rathbun, G., Crompton, A., VanRenterghem, B., Czech, M. P., Ravichandran, K. S., Burakoff, S. J., Wang, D.-S., Chen, C.-S., and Cantley, L. C. (1997) *J. Biol. Chem.* **272**, 22059–22066
261. Hemmings, B. A. (1997) *Science* **275**, 1899
262. Blomberg, N., Baraldi, E., Nilges, M., and Saraste, M. (1999) *Trends Biochem. Sci.* **24**, 441–445
263. Koshiba, S., Kigawa, T., Kim, J.-H., Shirouzu, M., Bowtell, D., and Yokoyama, S. (1997) *J. Mol. Biol.* **269**, 579–591
264. Saras, J., and Heldin, C.-H. (1996) *Trends Biochem. Sci.* **21**, 455–458
265. Cabral, J. H. M., Petosa, C., Sutcliffe, M. J., Raza, S., Byron, O., Poy, F., Marfatia, S. M., Chishti, A. H., and Liddington, R. C. (1996) *Nature (London)* **382**, 649–652
266. Ponting, C. P. (1997) *Protein Sci.* **6**, 464–468
267. Resnick, D., Pearson, A., and Krieger, M. (1994) *Trends Biochem. Sci.* **19**, 5–8
268. Chang, Y., Mochalkin, I., McCance, S. G., Cheng, B., Tulinsky, A., and Castellino, F. J. (1998) *Biochemistry* **37**, 3258–3271
269. de Vos, A. M., Ultsch, M. H., Kelley, R. F., Padmanabhan, K., Tulinsky, A., Westbrook, M. L., and Kossiakoff, A. A. (1992) *Biochemistry* **31**, 270–279
270. De Serrano, V. S., and Castellino, F. J. (1992) *Biochemistry* **31**, 11698–11706
271. Sun, M.-F., Zhao, M., and Gailani, D. (1999) *J. Biol. Chem.* **274**, 36373–36377
272. Macias, M. J., Hyvönen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. (1996) *Nature (London)* **382**, 646–649
273. Bogusky, M. J., Dobson, C. M., and Smith, R. A. G. (1989) *Biochemistry* **28**, 6728–6735
274. Hoffmann, W., and Hauser, F. (1993) *Trends Biochem. Sci.* **18**, 239–243
275. Lamb, J. R., Tugendreich, S., and Hieter, P. (1995) *Trends Biochem. Sci.* **20**, 257–259
276. Das, A. K., Cohen, P. T. W., and Barford, D. (1998) *EMBO J.* **17**, 1192–1199
277. Barlow, P. N., Luisi, B., Milner, A., Elliott, M., and Everett, R. (1994) *J. Mol. Biol.* **237**, 201–211
278. Lichtarge, O., Yamamoto, K. R., and Cohen, F. E. (1997) *J. Mol. Biol.* **274**, 325–337
279. Saurin, A. J., Borden, K. L. B., Boddy, M. N., and Freemont, P. S. (1996) *Trends Biochem. Sci.* **21**, 208–214
280. Pérez-Alvarado, G. C., Kosa, J. L., Louis, H. A., Beckerle, M. C., Winge, D. R., and Summers, M. F. (1996) *J. Mol. Biol.* **257**, 153–174
281. Porter, K. R., and Tucker, J. B. (1981) *Sci. Am.* **244**(Mar), 57–67
282. Weber, K., and Osborn, M. (1985) *Sci. Am.* **253**(Oct), 110–120
283. Amos, L. A., and Amos, B. W. (1991) *Molecules of the Cytoskeleton*, The Guilford Press, New York
284. Lazarides, E. (1980) *Nature (London)* **283**, 249–256
285. Fuchs, E., and Weber, K. (1994) *Ann. Rev. Biochem.* **63**, 345–382
286. Fuchs, E., and Cleveland, D. W. (1998) *Science* **279**, 514–519
287. Kaufmann, S. H. (1989) *J. Biol. Chem.* **264**, 13946–13955
288. Ip, W., Hartzer, M. K., Pang, Y.-Y. S., and Robson, R. M. (1985) *J. Mol. Biol.* **183**, 365–375
289. Herrmann, H., Häner, M., Brettel, M., Müller, S. A., Goldie, K. N., Fedtke, B., Lustig, A., Franke, W. W., and Aebi, U. (1996) *J. Mol. Biol.* **264**, 933–953
289a. Herrmann, H., Strelkov, S. V., Feja, B., Rogers, K. R., Brettel, M., Lustig, A., Häner, M., Parry, D. A. D., Steinert, P. M., Burkhard, P., and Aebi, U. (2000) *J. Mol. Biol.* **298**, 817–832
289b. Heimburg, T., Schuenemann, J., Weber, K., and Geisler, N. (1996) *Biochemistry* **35**, 1375–1382
290. Geisler, N., Schünemann, J., and Weber, K. (1992) *Eur. J. Biochem.* **206**, 841–852
290a. Bellin, R. M., Sernett, S. W., Becker, B., Ip, W., Huiatt, T. W., and Robson, R. M. (1999) *J. Biol. Chem.* **274**, 29493–29499
291. Leterrier, J. F., Käs, J., Hartwig, J., Vegners, R., and Janmey, P. A. (1996) *J. Biol. Chem.* **271**, 15687–15694
292. Betts, J. C., Blackstock, W. P., Ward, M. A., and Anderton, B. H. (1997) *J. Biol. Chem.* **272**, 12922–12927
293. Carter, J., Gragerov, A., Konvicka, K., Elder, G., Weinstein, H., and Lazzarini, R. A. (1998) *J. Biol. Chem.* **273**, 5101–5108
294. Osborn, M., and Weber, K. (1986) *Trends Biochem. Sci.* **11**, 469–472
295. Gibson, M. A., Kumaratilake, J. S., and Cleary, E. G. (1989) *J. Biol. Chem.* **264**, 4590–4598
296. Steinert, P. M. (1991) *J. Structural Biol.* **107**, 157–174
297. Cohlberg, J. A. (1993) *Trends Biochem. Sci.* **18**, 360–362
298. Nakajima-Iijima, S., Hamada, H., Reddy, P., and Kakunaga, T. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6133–6137
299. Orlova, A., Chen, X., Rubenstein, P. A., and Egelman, E. H. (1997) *J. Mol. Biol.* **271**, 235–243
300. Fulton, A. B. (1984) *The Cytoskeleton: Cellular Architecture and Choreography*, Chapman and Hall, New York
301. Bershadsky, A. D., and Vasiliev, J. M. (1988) *Cytoskeleton*, Plenum, New York
302. DeRosier, D. J., Tilney, L. G., and Egelman, E. (1980) *Nature (London)* **287**, 291–296
303. Collier, N. C., and Wang, K. (1982) *J. Biol. Chem.* **257**, 6937–943
304. Moon, R. T., and Lazarides, E. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5494–5499
305. Meek, R. L., Lonsdale-Eccles, J. D., and Dale, B. A. (1983) *Biochemistry* **22**, 4867–4877
306. Steinert, P. M., and Idler, W. W. (1979) *Biochemistry* **18**, 5664–5669
307. Dustin, P. (1980) *Sci. Am.* **243**(Aug.), 67–76
308. Mandelkow, E., and Mandelkow, E.-M. (1994) *Current Opinion in Structural Biology* **4**, 171–179
309. Hyams, J. S., and Lloyd, C. W., eds. (1994) *Microtubules*, Wiley-Liss, New York
310. Sosa, H., and Milligan, R. A. (1996) *J. Mol. Biol.* **260**, 743–755
311. Margulis, L., To, L., and Chase, D. (1978) *Science* **200**, 1118–1124
312. Thormählen, M., Marx, A., Müller, S. A., Song, Y.-H., Mandelkow, E.-M., Aebi, U., and Mandelkow, E. (1998) *J. Mol. Biol.* **275**, 795–809
313. Weber, K., Pollack, R., and Bibring, T. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 459–463
314. Wolf, S. G., Nogales, E., Kikkawa, M., Gratzinger, D., Hirokawa, N., and Downing, K. H. (1996) *J. Mol. Biol.* **262**, 485–501
315. Nogales, E., Wolf, S. G., and Downing, K. H. (1998) *Nature (London)* **391**, 199–203
316. Pennisi, E. (1998) *Science* **279**, 176–177
317. Zheng, Y., Wong, M. L., Alberts, B., and Mitchison, T. (1995) *Nature (London)* **378**, 578–583
318. Hackney, D. D. (1995) *Nature (London)* **376**, 215–216

## References


319. Raff, E. C., Fackenthal, J. D., Hutchens, J. A., Hoyle, H. D., and Turner, F. R. (1997) *Science* **275**, 70–73
320. Angelastro, J. M., and Purich, D. L. (1992) *J. Biol. Chem.* **267**, 25685–25689
321. Terry, B. J., and Purich, D. L. (1980) *J. Biol. Chem.* **255**, 10532–10536
322. Rakic, P., Knyihar-Csillik, E., and Csillik, B. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9218–9222
323. Rodionov, V. I., and Borisy, G. G. (1997) *Science* **275**, 215–218
324. Hoenger, A., and Milligan, R. A. (1996) *J. Mol. Biol.* **263**, 114–119
325. Fan, J., Griffiths, A. D., Lockhart, A., Cross, R. A., and Amos, L. A. (1996) *J. Mol. Biol.* **259**, 325–330
326. Hirose, K., Fan, J., and Amos, L. A. (1995) *J. Mol. Biol.* **251**, 329–333
327. Knop, M., and Schiebel, E. (1997) *EMBO J.* **16**, 6985–6995
328. Moritz, M., Braunfeld, M. B., Sedat, J. W., Alberts, B., and Agard, D. A. (1995) *Nature (London)* **378**, 638–640
329. Berridge, M. J. (1990) *J. Biol. Chem.* **265**, 9583–9586
330. Arnold, C. S., Johnson, G. V. W., Cole, R. N., Dong, D. L.-Y., Lee, M., and Hart, G. W. (1996) *J. Biol. Chem.* **271**, 28741–28744
331. Bosc, C., Cronk, J. D., Pirollet, F., Watterson, D. M., Haiech, J., Job, D., and Margolis, R. L. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2125–2130


## Study Questions

1. Rewrite Equations 6-75 through 6-77 in terms of dissociation constants. These may be labeled $K_1$, $K_2$, $K_i$, etc., as is conventional, but you may prefer to use $K_{1d}$, $K_{2d}$, $K_{id}$, etc., to avoid confusion.

2. A molecule has two identical binding sites for a ligand X. The Gibbs energy of interaction between ligands bound to the same molecule, ε, is defined as the change in Gibbs energy of binding of the ligand to the molecule that results from the prior binding of a ligand at the adjacent site. If the saturation fraction is Y, show from the equation for the binding isotherm that the following equation holds when Y = 1/2:

$$dY/d\ln[X] = \frac{1}{2(1 + e^{\varepsilon/RT})}$$

3. The hydrogen ion binding curve for succinate is shown is Fig. 7-4. From the curve estimate ε and the microscopic association constants.

4. A linear chain molecule has a very large number of identical binding sites for a ligand X. The Gibbs energy of interaction between ligands bound to adjacent sites is ε. Interactions between non-nearest neighbors are considered negligible. If the binding constant for a site adjacent to unoccupied sites is $K_r$, the binding isotherm is given by

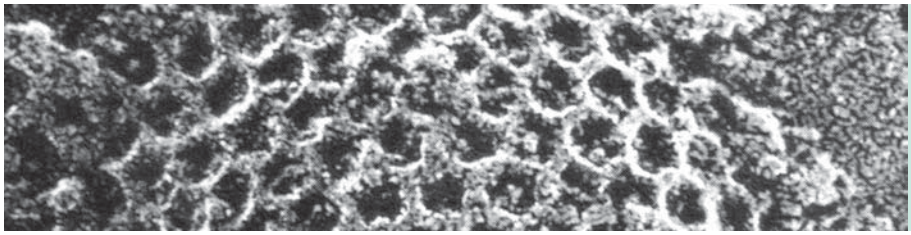
$$Y = \frac{1}{2} + \frac{K[X]e^{-\varepsilon/RT} - 1}{2\{K[X]e^{-\varepsilon/RT} - 1)^2 + 4K[X]\}^{1/2}}$$

[Applequist, J. (1977) *J. Chem. Ed.* **54**, 417]. Show from the equation for the binding isotherm that the following equation holds at Y = 1/2;

$$dY/d\ln[X] = 1/4e^{\varepsilon/2RT}$$

5. The binding of adenosine to polyribouridylic acid [poly(U)] has been studied by the method of equilibrium dialysis [Huang and Ts'o (1966) *J. Mol. Biol.* **16**, 523]. The table below gives the fraction of poly(U) sites occupied, Y at various molar concentrations of free adenosine [A] at 5°C. Assuming that the nearest-neighbor interaction model is correct, determine the intrinsic association constant for the binding of adenosine to poly U and the free energy of interaction of adjacent bound adenosines. Do the bound molecules attract or repel each other?

| [A] x $10^3$ | Y | [A] x $10^3$ | Y |
|---|---|---|---|
| 0.51 | 0 | 3.07 | 0.72 |
| 2.10 | 0 | 4.00 | 0.92 |
| 2.70 | 0.15 | 6.50 | 0.93 |
| 2.96 | 0.36 | 8.50 | 0.93 |
| 3.01 | 0.52 | 10.00 | 1.00 |



Portion of an endocytic vacuole forming in the plasma membrane of a cultured fibroblast. The view is from the inside of the cell and shows a large clathrin cage assembling to form a coated vesicle. The overall diameter of the vacuole is ~0.2 µm. Clathrin cages vary in size and in the number of faces but are typically ~0.1 µm in diameter (see Fig. 8-27 and associated references). Courtesy of Barbara Pearse.

# Contents


380 **A. Lipid Structures**
380 1. Fatty Acids, Fatty Alcohols, and Hydrocarbons
382 2. Acylglycerols, Ether Lipids, and Waxes
382 3. Phospholipids
387 4. Glycolipids
389 5. Sphingolipids
389 6. Sterols and Other Isoprenoid Lipids
390 **B. Membranes**
390 1. The Structure of Membranes
392 *Lipids of membranes*
392 *Liquid crystals, liposomes, and artificial membranes*
394 *Physical properties of membrane lipids*
396 *Functions of phospholipid head groups*
397 *Non-bilayer structures of phospholipids*
397 *Membrane fluidity and life*
400 *Electrical properties of membranes*
401 *The two sides of a membrane*
401 2. Membrane Proteins
401 *Integral membrane proteins*
402 *Anchors for proteins*
403 *Analyzing erythocyte membranes*
404 *Glycoproteins*
405 *Connections to the cytoskeleton*
405 *Integrins and focal adhesions*
406 *Other cell adhesion molecules*
407 *Peripheral proteins of the outer membrane surface*
409 3. Enzymes and Membrane Metabolism
410 **C. The Transport of Molecules through Membranes**
410 1. Facilitated Diffusion and Active Transport
411 2. Pores, Channels, and Carriers
411 *Porins*
411 *Aquaporins*
412 *Ion channels*
414 *Channel-forming toxins and antibiotics*
414 *Ionophores and other mobile carriers*
415 3. The 12-Helix Major Facilitator Superfamily
415 *Entrance of sugars into cells*
416 *Cotransport of sugars and other nutrients with $H^+$ or $Na^+$*
417 4. Active Transport Systems
417 *Periplasmic permeases*
419 *The bacterial phosphotransferase system*
420 5. Transport of Ions
420 *Anions*
421 *Cation channels*
422 *Active transport of cations*
422 *ATP-driven ion pumps*
425 6. Exocytosis, Endocytosis, and the Flow of Membrane Constituents
427 **D. Communication**
427 **E. The Extracellular Matrix and Cell Walls**
427 1. The Structure of Bacterial Cell Walls
428 *The outer membrane of gram-negative bacteria*
431 *Teichoic and teichuronic acids*
431 *Mycobacteria*
431 *Other bacterial coats*
431 2. The Surroundings of Animal Cells
432 *The collagens*
436 *Elastic fibers*
436 *Cartilage and basement membranes*
440 *Fibrillin and Marfan's syndrome*
440 *The cuticles of invertebrates*
440 *Bones, teeth, and shells*
443 3. Cell Walls of Fungi and Green Plants

444 **References**
452 **Study Questions**


**Boxes**

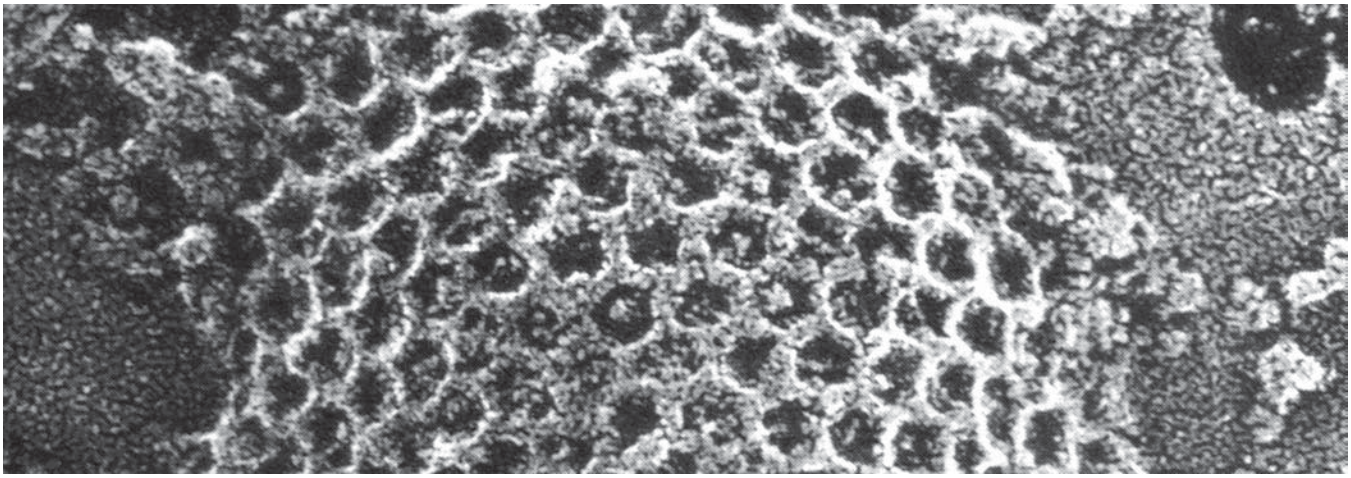

384 Box 8-A The Platelet-Activating Factor
386 Box 8-B Dipalmitoylphosphatidylcholine and the Surfactant System of the Lungs
398 Box 8-C Electron Paramagnetic Resonance (EPR) Spectra and "Spin Lables"
418 Box 8-D Colicins: Antibiotic Proteins
438 Box 8-E Genetic Defects of Collagen Structure
439 Box 8-F Skin
442 Box 8-G The Biochemistry of Teeth


**Tables**


380 Table 8-1 Some Important Fatty Acids
381 Table 8-2 Fatty Acid Composition (in %) of Some Typical Fats and Oils
393 Table 8-3 Estimated Chemical Compositions of Some Membranes
435 Table 8-4 Types of Vertebrate Collagen

# Lipids, Membranes, and Cell Coats

# 8



The boundary between a living cell and its surroundings is the incredibly thin (7–10 nm) plasma membrane. This vital partition, which controls the flow of materials into and out of a cell and which senses and controls the response of cells to hormones and other external signals, consists largely of **phospholipids** together with embedded proteins. The nonpolar chains of the phospholipids stick together to form a double molecular layer or **bilayer** which provides the basic structure of almost all biological membranes.

Phospholipids, together with other natural materials that have a high solubility in apolar solvents or are structurally related to compounds with such solubility properties, are classified as **lipids**. The most abundant lipids are the fats, compounds that are stored by animals and by many plants as an energy reserve (Fig. 8-1). Other lipids form the outer cuticle of plants and yet others serve as protective coatings on feathers and hair. Vitamins A, D, K, and E and ubiquinone are all lipids as are a variety of hormones and such light-absorbing plant pigments as the chlorophylls and carotenoids. Many of these compounds are dissolved in or partially embedded in the plasma membrane of bacteria or in the mitochondrial and chloroplast membranes of higher organisms. Membranes serve many purposes. The most obvious is to divide space into compartments. Thus, the plasma membrane forms cell boundaries and mitochondrial membranes separate the enzymes and metabolites of mitochondria from those of the cytosol. Membranes are semipermeable and regulate the penetration into cells and organelles of both ionic and nonionic substances. Many of these materials are brought into the cell against a concentration gradient. Hence, osmotic work must be done in a process known as **active transport**. Many enzymes,

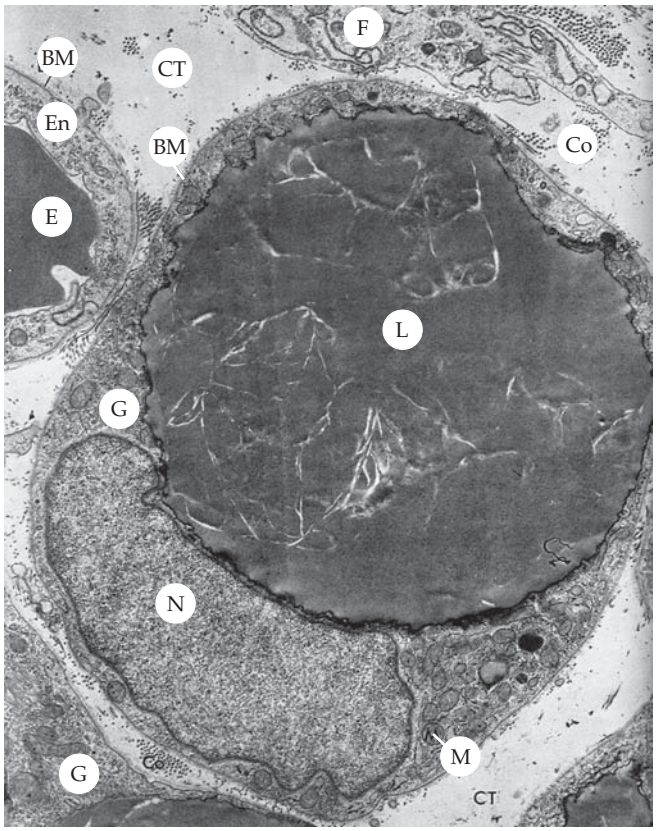


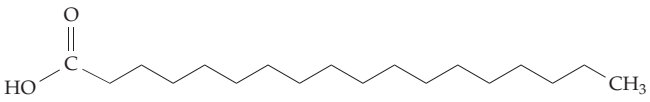

**Figure 8-1** Electron micrograph of a thin section of a fat storage cell or adipocyte. L, the single large fat droplet; N, nucleus; M, mitochondria; En, endothelium of a capillary containing an erythrocyte (E); CT, connective tissue ground substance which contains collagen fibers (Co) and fibroblasts (F). The basement membranes (BM) surrounding the endothelium and the fat cell are also marked. From Porter and Bonneville.[6] Courtesy of Mary Bonneville.

including those responsible for most of the oxidative metabolism of cells, are found in membranes of bacteria and of mitochondria. Within the chloroplasts of green leaves, highly folded membranes containing chlorophyll absorb energy from the sunlight. Thin membranes contain the photoreceptor proteins that function in vision. Electrical impulses are transmitted along the membranes of nerve cells.

The outer surfaces of membranes are designed to interact with the cell's external world. Special receptors sense the presence of hormones. Binding proteins await the arrival of needed nutrients and help to bring them into cells. Highly individual arrangements of protein and of the carbohydrate "fuzz" of glycoproteins and sphingolipids screen the outer surface, helping to prevent attack by foreign bacteria, viruses, and toxins.

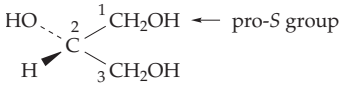
## A. Lipid Structures

Unlike proteins, polysaccharides, and nucleic acids, most lipids are not polymers. However, they are made by linking together smaller molecules.[1–5] Among the "building blocks" of lipids are **fatty acids, glycerol, phosphoric acid**, and **sugars**. Many lipids have both polar and nonpolar regions. This gives them an **amphipathic** character, i.e., a tendency toward both hydrophobic and hydrophilic behavior, and accounts for their tendency to aggregate into membranous structures. Notice that in the following structure the carbon atoms of glycerol have been numbered 1 to 3. Although glycerol is *not* chiral, the positioning of the two hydroxymethyl groups is not equivalent. If the *priority*, used in the *RS* system (page 42) for the C-1 group is raised (e.g., by ester formation) to be higher than that for the C-3 group, the molecule would have the *S* configuration. The C-1 group is said to be pro-*S* and the C-3 group pro-*R*. According to the stereochemical numbering (*sn*) system, which is discussed further in Chapter 9, the carbon in the pro-*S* position is numbered 1.




A fatty acid

Stearic acid, 18 carbon atoms, major component of animal triacylglycerols




Glycerol

Labeled by the stereochemical numbering (*sn*) system

### 1. Fatty Acids, Fatty Alcohols, and Hydrocarbons

The names and structures of some fatty acids are summarized in Table 8-1. Notice that these acids have straight carbon chains and may contain one or more double bonds. Except for the smallest members of the series, which are soluble in water, fatty acids are strongly hydrophobic. However, they are all acids with $pK_a$ values in water of ~4.8. To the extent that free fatty acids occur in nature, they are likely to be found in interfaces between lipid and water with the carboxyl groups dissociated and protruding into the water. However, most naturally occurring fatty acids

**TABLE 8-1**
**Some Important Fatty Acids**

| No. carbon atoms | Systematic name | Common name | Abbreviation[a] | Common name of acyl group |
|---|---|---|---|---|
| | | **Saturated fatty acids** | | |
| 1 | Methanoic | Formic | | Formyl |
| 2 | Ethanoic | Acetic | | Acetyl |
| 3 | Propanoic | Propionic | | Propionyl |
| 4 | Butanoic | Butyric | 4:0 | Butyryl |
| 12 | Dodecanoic | Lauric | 12:0 | Lauroyl[b] |
| 14 | Tetradecanoic | Myristic | 14:0 | Myristoyl |
| 16 | Hexadecanoic | Palmitic | 16:0 | Palmitoyl |
| 18 | Octadecanoic | Stearic | 18:0 | Stearoyl |
| 20 | Eicosanoic | Arachidic | 20:0 | |
| 22 | Docosanoic | Behenic | 22:0 | |
| 24 | Tetracosanoic | Lignoceric | 24:0 | |
| | | **Unsaturated fatty acids[c]** | | |
| 4 | | Crotonic | 4:1(2*t*) | Crotonoyl |
| 16 | | Palmitoleic | 16:1(9*c*) | |
| 18 | | Oleic | 18:1(9*c*) | Oleoyl |
| 18 | | Vaccenic | 18:1(11*c*) | |
| 18 | | Linoleic | 18:2(9c,12*c*) | |
| 18 | | Conjugated linoleic | 18:2(9c,11*t*) | |
| 18 | | Linolenic | 18:3(9*c*,12*c*,15*c*) | |
| 20 | | Arachidonic | 20:4(5*c*,8*c*,11*c*,14*c*) | |

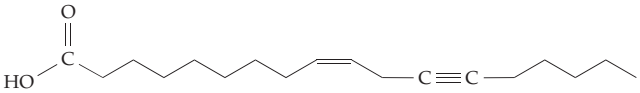
[a] The number of carbon atoms is given first, then the number of double bonds. The positions of the lowest numbered carbon of each double bond and whether the configuration is *cis* (*c*) or *trans* (*t*) are indicated in parentheses.

[b] Official IUPAC names of these and other acyl groups have been designated by the Commission of the Nomenclature of Organic Chemistry in *Pure and Applied Chemistry* **10**, 111–125 (1965). In a number of cases IUPAC inserted an *o* in the traditional name, e.g., palmityl became palmitoyl and crotonyl became crotonoyl. However, acetyl was not changed. In many cases the systematic names, e.g., hexadecanoyl (from hexadecanoic acid), are preferable and IUPAC–IUB recommends that alkyl radicals always be designated by systematic names, e.g., hexadecyl, *not* palmityl alcohol. The older use of palmityl for both acyl and alkyl radicals was one reason for IUPAC's adoption of new names for acyl radicals.

[c] Systematic names are not often used because of their complexity, e.g., linolenic acid is *cis,cis,cis*-9,12,15-octadecatrienoic acid.

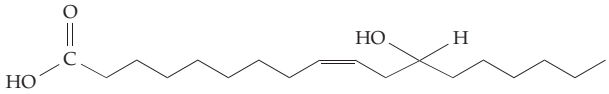
are esterified or combined via amide linkages in complex lipids. For example, ordinary fats are largely the fatty acid esters of glycerol called **triacylglycerols** (triglycerides).

There is a seemingly endless variety of fatty acids, but only a few of them predominate in any single organism. Most fatty acid chains contain an even number of carbon atoms. In higher plants the $C_{16}$ **palmitic acid** and the $C_{18}$ unsaturated **oleic** and **linoleic acids** predominate. The $C_{18}$ saturated **stearic acid** is almost absent from plants and $C_{20}$ to $C_{24}$ acids are rarely present except in the outer cuticle of leaves. Certain plants contain unusual fatty acids which may be characteristic of a taxonomic group. For example, the Compositae (daisy family) contain acetylenic fatty acids and the castor bean contains the hydroxy fatty acid **ricinoleic acid**.




Crepenynic acid 18:2 (9c, 12a)

Accounts for 60% of the fatty acids in seeds of *Crepis foetida*, a member of the compositae family




Ricinoleic acid (12-hydroxyoleic acid)

Accounts for up to 90% of the fatty acids of *Ricinus communis* (castor bean)

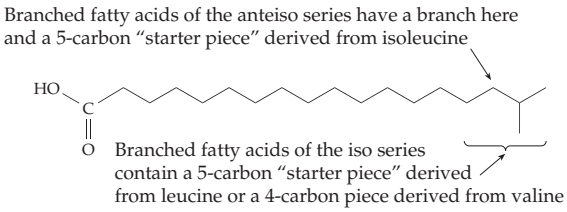
Like plants, animals contain palmitic and oleic acids. In addition, large amounts of stearic acid and small amounts of the $C_{20}$, $C_{22}$, and $C_{24}$ acids are also present. Phospholipids of photoreceptor membranes of the retina contain fatty acid chains as long as $C_{36}$.[7] The variety of fatty acids found in animals is greater than in a given plant species. A large fraction of the fatty acids present in most higher organisms are *unsaturated* and contain strictly *cis* double bonds. Table 8-2 shows the fatty acid composition of some typical triacylglycerol mixtures.

**TABLE 8-2**
**Fatty Acid Composition (in %) of Some Typical Fats and Oils**[a]

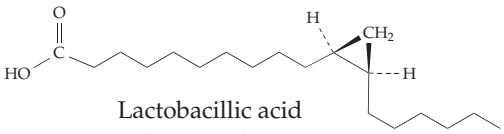
| | No. of carbon atoms and (following colon) the number of double bonds | | | | | |
|---|---|---|---|---|---|---|
| Fats and oils | 14 | 16 | 18 | 16:1 | 18:1 | 18:2 |
| Human depot fat | 3 | 23 | 4 | 8 | 45 | 10 |
| Beef tallow | 4 | 30 | 25 | 5 | 36 | 1 |
| Corn oil | | 13 | 2 | | 31 | 54 |
| Lard | 1 | 28 | 15 | 3 | 42 | 9 |

[a] From Gunstone.[8]

Bacteria usually lack polyunsaturated fatty acids but often contain branched fatty acids, cyclopropane-containing acids, hydroxy fatty acids, and unesterified fatty acids. Mycobacteria, including the human pathogen *Mycobacterium tuberculosis*, contain **mycolic acids**. In these compounds the complex grouping R contains a variety of functional groups including –OH, $-OCH_3$,

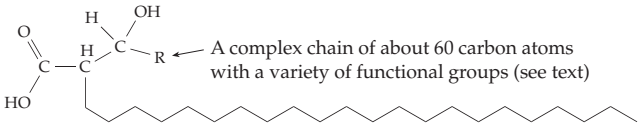



Branched fatty acids




Lactobacillic acid

A major fatty acid of lactobacilli

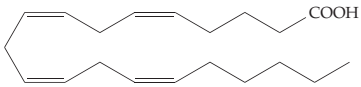



$C_{22}$ or $C_{24}$ alkyl group

A mycolic acid

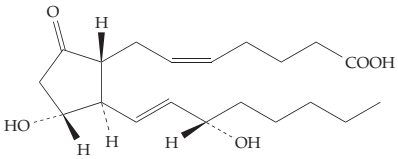
C=O, –COOH, cyclopropane rings, methyl branches, and C=C bonds. Each species of *Mycobacterium* contains about two dozen different mycolic acids[9,10] as well as other complex $C_{30}$–$C_{56}$ fatty acids (see Eq. 21-5).[11]

Certain polyunsaturated fatty acids are essential in the human diet (see Box 21-B). One of these, **arachidonic acid** (which may be formed from dietary linoleic acid), serves as a precursor for the formation of the hormones known as **prostaglandins** and a series of related **prostanoids**. Lipids of animal origin also




Arachidonic acid

A nutritionally essential fatty acid (Box 21-A) is shown here in a folded conformation

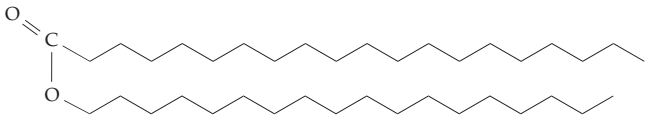



Prostaglandin $PGE_2$

One of a family of hormones, this one is synthesized in tissues from arachidonic acid

contain unusual unsaturated fatty acids. Among them, **conjugated linoleic acids** are receiving attention for their possible cancer-preventive action. The predominant form in meats, dairy products, and the human body is the $C_{18}$ 9-*cis*, 11-*trans* isomer whose two double bonds are conjugated.[11a]

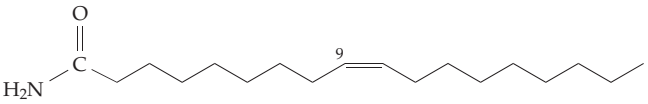
Other lipid components include the **fatty alcohols** which are formed by reduction of the acids. These are esterified with fatty acids to form **waxes**. Both fatty alcohols and free fatty acids occur in waxes together with the esterified forms. These mixtures are found on exterior surfaces of plants and animals. Plants and, to a limited extent, animals are able to decarboxylate fatty acids in a multistep process to **alkanes** and these too are important constituents of some waxes. Small amounts of fatty acid amides such as *cis*-9,10-octadecenoamide are present in low concentrations in



A wax

An ester of a fatty acid and a fatty alcohol

the cerebrospinal fluid of cats and rats as well as humans. This compound accumulates in cats that are deprived of sleep. When the synthetic compound was injected into rats they fell into apparently normal sleep.[12]

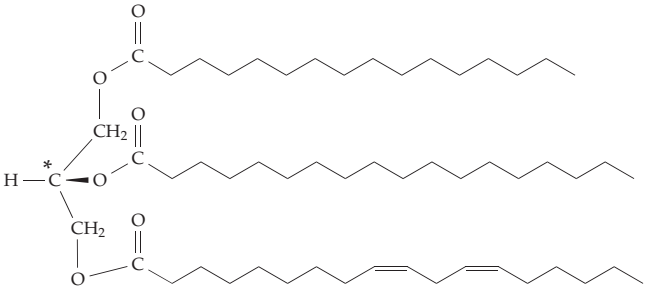


*cis*-9,10-Octadecenoamide (Oleic acid amide)

Insects make unsaturated as well as saturated hydrocarbons. The former as well as long-chain alcohols and their esters often form the volatile **pheromones** with which insects communicate. Thus, the female pink bollworm attracts a male with a sex pheromone consisting of a mixture of the *cis*,*cis* and *cis*,*trans* isomers of 7,11-hexadecadienyl acetate,[13] and European corn borer males are attracted across the cornfields of Iowa by *cis*-11-tetradecenyl acetate.[14] Addition of a little of the *trans* isomer makes the latter sex attractant much more powerful. Since more than one species uses the same attractant, it is possible that the males can distinguish between different ratios of isomers or of mixtures of closely related substances.

## 2. Acylglycerols, Ether Lipids, and Waxes

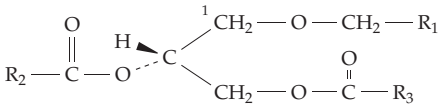
The components of complex lipids are linked in a variety of ways. Often, glycerol acts as the central unit, e.g., combining in ester linkage with three fatty acids to form **triacylglycerols** (triglycerides), the common fats of adipose tissues and plant oils. Diacyl- and monoacylglycerols (diglycerides, and monoglycerides) are



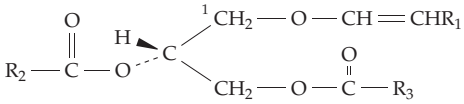
A triacylglycerol (fat)

Notice the chiral center, designated by the asterisk

present to a lesser extent. In addition, small amounts of **alkyl ethers** or **alkenyl ethers** are often present in isolated lipids. They are especially abundant in fish liver oils.



1-Alkyl-2,3-diacyl-*sn*-glycerol, an alkyl ether lipid



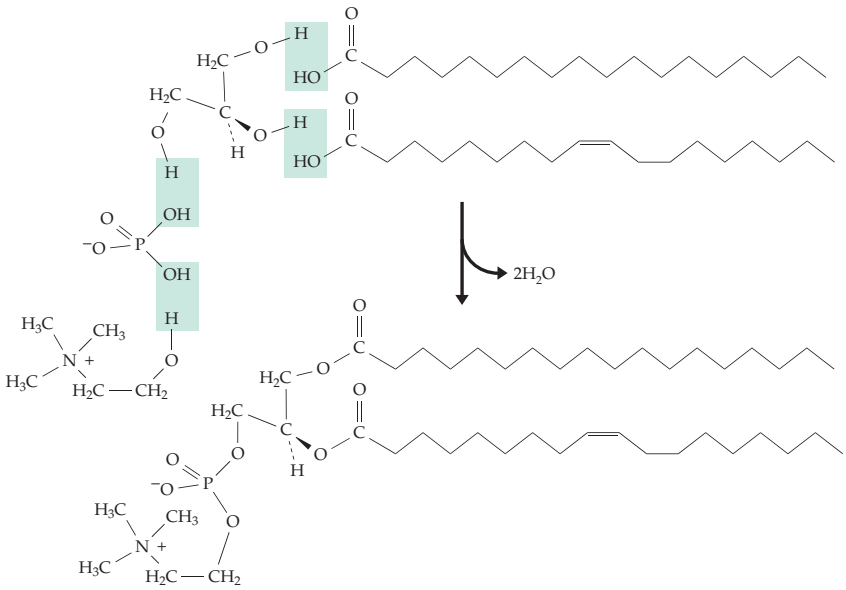
1-Alkenyl-2,3-diacyl-*sn*-glycerol, an alkenyl ether

These ether lipids are all chiral molecules with an *R* configuration but are derivatives of the nonchiral glycerol. The carbon atoms of glycerol are numbered using the stereochemical system which is described on p. 470. The ether linkage is to the *sn*-1 carbon atom. Most phospholipids are derivatives of the *sn*-3 phosphate ester of glycerol.

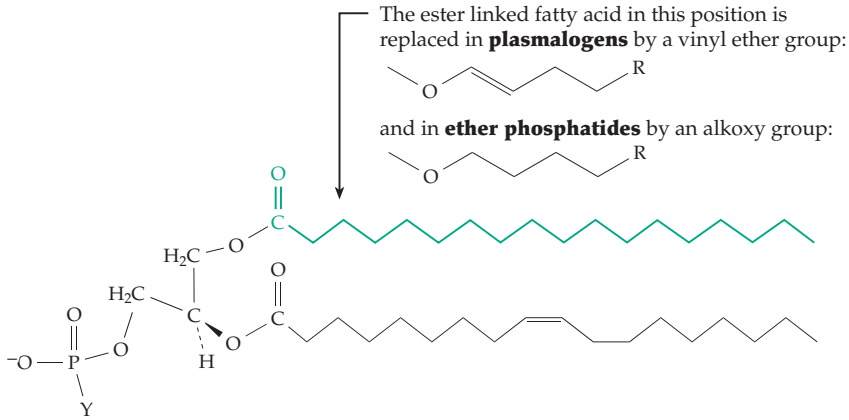
Triacylglycerols and the ether lipids described in the previous section are classified as **neutral lipids**. Other neutral lipids are alcohols, waxes, aldehydes, and hydrocarbons derived from fatty acids. These sometimes have specific biological functions. For example, fatty aldehydes are important in the bioluminescence of bacteria (Eq. 23-47).

## 3. Phospholipids

As major constituents of biological membranes, phospholipids play a key role in all living cells. The two principal groups of phospholipids are the




A phosphatidylcholine (lecithin). (8-1)



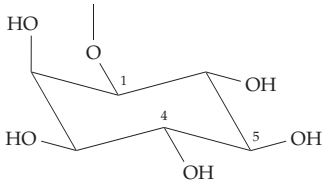
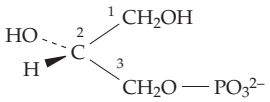

| | | Abbrev. |
|---|---|---|
| Y = —OH | Phosphatidic acid | PA |
| —O—$CH_2CH_2$—$N^+(CH_3)_3$ | Phosphatidylcholine (lecithin) | PC |
| —O—$CH_2$—$CH_2$(COO⁻)—$NH_3^+$ | Phosphatidylserine (serine cephalin) | PS |
| —O—$CH_2CH_2$—$NH_3^+$ | Phosphatidylethanolamine (ethanolamine cephalin) | PE |
|  | Phosphatidylinositol | PI |
| —O—$CH_2$—CH(OH)—$CH_2OH$ | Phosphatidylglycerol | |
| —O—$CH_2$—CH(OH)—$CH_2$—O— | Diphosphatidylglycerol (cardiolipin), a special component of bacterial and mitochondrial membranes | |

**Figure 8-2** Structures of some phosphatides (glycerophospholipids).

**glycerophospholipids** (glycerophosphatides; Fig. 8-2) which contain the alcohol glycerol and the **sphingophospholipids** which contain the alcohol **sphingosine** (Section 5). The glycerophospholipids can be thought of as arising from the building blocks glycerol, fatty acids, the dihydrogen phosphate ion $H_2PO_4^-$, and the appropriate alcohol by removal of four molecules of water (Eq. 8-1). They are derivatives of *sn*-glycerol-3-phosphate. Esterification of this alcohol with two fatty acids gives a **phosphatidic acid** (Fig. 8-2). Formation of a phosphate diester linkage to one of the alcohols



*sn*-Glycerol 3-phosphate
(L-α-glycerol phosphate)

choline, serine, or ethanolamine yields a glycerophospholipid. The resulting three groups of phospholipids are called **phosphatidylcholine** (lecithin), **phosphatidylserine**, and **phosphatidylethanolamine**, respectively (Fig. 8-2).

The phosphate and choline, ethanolamine, or serine portions of the phosphatide are electrically charged and provide a polar "head" for the molecule. In all three cases the positively charged group is able to fold back and form an ion pair with the negatively charged phosphate group. However, the methyl groups surrounding the nitrogen in phosphatidylcholine prevent a very close approach and with phosphatidylserine the adjacent carboxylate group weakens this electrostatic interaction. Unlike the triacylglycerols, most of which are liquid at body temperature, phospholipids are solid at this temperature. This property, like the ionic properties of the phosphatides, is doubtless related to their suitability for functioning in biological membranes.

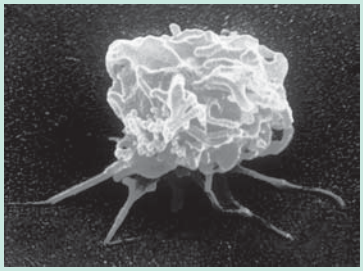
Lecithins and related phospholipids usually contain a saturated fatty acid in the C-1 position but an unsaturated acid, which may contain from one to four double bonds, at C-2. Arachidonic acid is often present here. Hydrolysis of the ester linkage at C-2 yields a 1-acyl-3-phosphoglycerol, better known as a **lysophosphatidylcholine**. The name comes from the powerful detergent action of these substances which leads to lysis of cells. Some snake venoms contain phospholipases that form lysophosphatidylcholine. Lysophosphatidic acid (1-acyl-glycerol-3-phosphate) is both an intermediate in phospholipid biosynthesis (Chapter 21) and also a signaling molecule released into the bloodstream by activated platelets.[15]

Another group of phosphatides contain the hexahydroxycyclohexane known as **inositol** (Fig. 8-2, see also Chapter 21).[16] Phosphatidylinositol, as well as smaller amounts of phosphatides derived from phosphate esters of inositol are present in membranes of all eukaryotes and have a specific role in regulating responses of cells to hormones and other external agents. See Chapter 11 for details. Phosphatidylinositol also forms part of "anchors" used to hold certain proteins onto membrane surfaces (see Fig. 8-13).

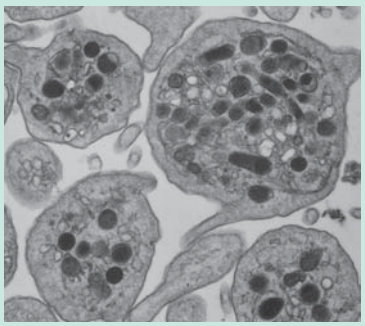
Bacteria and plants often make the anionic **phosphatidyglycerol** in which the second glycerol is esterified at its *sn*-1 position with the phosphate. Bacteria, as well as mitochondria, contain diphosphatidylglycerol (**cardiolipin**) in which phosphatidyl groups are attached at both the 1 and 3 positions of glycerol (Fig. 8-2). Ether phospholipids, analogous to the ether lipids described in Section 2, are also widely distributed. The alkenyl ether analogs of phosphatidylcholine (Fig. 8-2) are called **plasmalogens**.[17] In neutrophils the 1-*O*-alkyl ethers contain the major share of the cell's arachidonic acid, which is esterified in the 2 position.[18,19]

In halophilic (salt loving), thermophilic, and methanogenic bacteria, most of the lipids present are either

## BOX 8-A THE PLATELET-ACTIVATING FACTOR



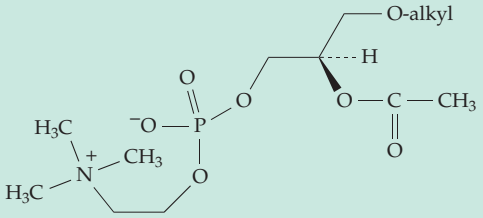
Scanning EM of activated blood platelets. © Quest, Photo Researchers



Transmission EM of thin section through activated blood platelets. © Photo Researchers

Part of the immune response consists of the release from stimulated neutrophils, macrophages, and other cells of a **platelet activating factor (PAF)**, a material that "activates" blood platelets. Activated platelets aggregate, a response that initiates clot formation. They may also be lysed and release stored substances that include platelet-derived growth factors (see Chapter 11) and fibrin stabilizing factor, a proenzyme that is converted to the protein-crosslinking enzyme transglutaminase (Eq. 2-17). See Fig. 12-17. However, the principal interest in the platelet-activating factor has arisen from its powerful effect in inducing **inflammation** in surrounding tissues.

PAF has been identified as the following simple ether phospholipid:[a–d]

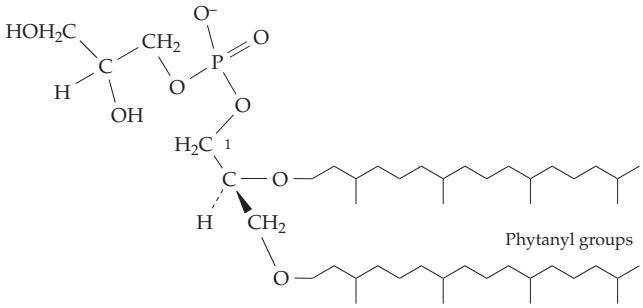



Platelet-activating factor (PAF)
1-*O*-alkyl-2-acetyl-*sn*-glycerophosphocholine

A remarkably potent compound, its effects on platelets are observed at concentrations of $10^{-11}$ to $10^{-10}$ M. Both lyso-PAF and the glycero-1- phosphocholine enantiomer are inactive. Specific receptors for this factor are evidently present on platelet surfaces.[e–g]

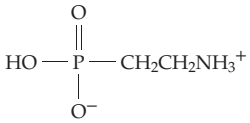
One effect of PAF on platelets is to induce a rapid (5–10s) cleavage of phosphatidylinositol 4,5-bisphosphate by phospholipase C to give **diacylglycerol** and **inositol 1,4,5-trisphosphate**. The subsequent effects of these two substances in causing a rapid influx of $Ca^{2+}$ and in inducing a series of secondary responses are outlined in Fig. 11-9. Among these responses are the release of the materials stored in the platelet's granules. PAF also appears to inhibit adenylate cyclase[b] and causes vasodilation, a property not expected for a compound that stimulates clotting. Receptors for PAF are also present in the brain, where this phosholipid may function in regulation of development.[h]

phospholipids and glycolipids containing the $C_{20}$ isoprenoid **phytanyl** group or the $C_{40}$ **diphytanyl** group[20–25] (see also Section 4), related isoprenoid alcohols, or long-chain 1,2-diols.[25] An example of a diphytanylglycerophospholipid is the following:




A phospholipid derived from 2,3-di-*O*-phytanyl-*sn*-glycerol

Many other phospholipids are present in small amounts or in a limited number of species. These include **phosphonolipids**, which contain a C–P bond and are abundant in ciliate protozoa such as *Tetrahymena* and in some other invertebrates.[26] Phosphonoethylamine replaces phosphoethanolamine in these lipids. A consequence of this structural alteration is a high



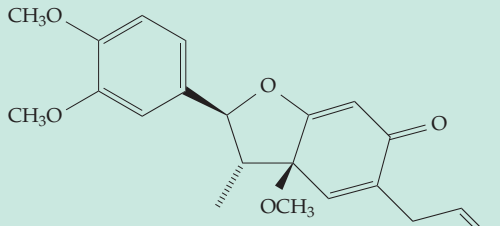

Phosphonoethylamine

degree of resistance to the action of the enzyme phospholipase C. The phosphonolipids of the external membrane of *Tetrahymena* are also ether lipids with an alkoxy group in the *sn*-1 position. This makes them

## BOX 8-A (continued)

Stimulated platelets release arachidonic acid rapidly from their phospholipids, apparently as a result of activation of phospholipase $A_2$. The released arachidonate can in turn be metabolized to endoperoxides and thromboxane $A_2$ (Chapter 21). These compounds are also potent activators of platelets and cause a self-activating or **autocrine** effect.[i,j] While PAF has a beneficial function, it can under some conditions contribute in a dangerous way to inflammation and to allergic responses including **anaphylaxis**,[j] **asthma**[g] and **cold-induced urticaria**.[k] Although the effect of PAF is separate from those of histamine and of leukotrienes, these agents may act cooperatively to induce inflammation.[l]

The biosynthesis of PAF is discussed in Chapter 21, Section C. One pathway is deacylation at the glycerol C-2 of a longer chain alkyl phospholipid followed by acetylation at the same position.[m,n] Platelets can inactivate PAF by the inverse sequence: deacetylation followed by acylation with arachidonic acid. PAF can also be hydrolyzed by a PAF acetylhydrolase, a phospholipase.[o] Absence of this enzyme in the brain may be related to a human brain malformation.[h] Some tissues may contain a lipid that inhibits binding of PAF.[p] The following compound from a Chinese herb binds to PAF receptors and may be the forerunner of useful drugs.[q]




Kadsurenone


[a] Cusack, N. J. (1980) *Nature (London)* **285**, 193
[b] Hanahan, D. J. (1986) *Ann. Rev. Biochem.* **55**, 483–509
[c] Prescott, S. M., Zimmerman, G. A., and McIntyre, T. M. (1990) *J. Biol. Chem.* **265**, 17381–17384
[d] Winslow, C. M., and Lee, M. L., eds. (1987) *New Horizons in Platelet Activating Factor Research*, Wiley, New York
[e] Hwang, S.-B., Lam, M.-H., and Pong, S.-S. (1986) *J. Biol. Chem.* **261**, 532–537
[f] Ishii, I., Izumi, T., Tsukamoto, H., Umeyama, H., Ui, M., and Shimizu, T. (1997) *J. Biol. Chem.* **272**, 7846–7854
[g] Bazan, N. G. (1995) *Nature (London)* **374**, 501–502
[h] Hattori, M., Adachi, H., Aoki, J., Tsujimoto, M., Arai, H., and Inoue, K. (1995) *J. Biol. Chem.* **270**, 31345–31352
[i] Bussolino, F., Sironi, M., Bocchietto, E., and Mantovani, A. (1992) *J. Biol. Chem.* **267**, 14598–14603
[j] Darius, H., Lefer, D. J., Smith, J. B., and Leefer, A. M. (1986) *Science* **232**, 58–60
[k] Grandel, K. E., Farr, R. S., Wanderer, A. A., Eisenstadt, T. C., and Wasserman, S. I. (1985) *N. Engl. J. Med.* **313**, 405–409
[l] Tomeo, A. C., Egan, R. W., and Durán, W. N. (1991) *FASEB J.* **5**, 2850–2855
[m] Billah, M. M., Eckel, S., Myers, R. F., and Siegel, M. I. (1986) *J. Biol. Chem.* **261**, 5824–5831
[n] Lee, T.-c, Ou, M.-c, Shinozaki, K., Malone, B., and Snyder, F. (1996) *J. Biol. Chem.* **271**, 209–217
[o] Stafforini, D. M., McIntyre, T. M., Zimmerman, G. A., and Prescott, S. M. (1997) *J. Biol. Chem.* **272**, 17895–17898
[p] Miwa, M., Hill, C., Kumar, R., Sugatan, J., Olson, M. S., and Hanahan, D. J. (1987) *J. Biol. Chem.* **262**, 527–530
[q] Hwang, S.-B., Lam, M.-H., Biftu, T., Beattie, T. R., and Shen, T.-Y. (1985) *J. Biol. Chem.* **260**, 15639–15645

## BOX 8-B DIPALMITOYLPHOSPHATIDYLCHOLINE AND THE SURFACTANT SYSTEM OF THE LUNGS

When air is exhaled the small alveoli of the lungs could collapse if it were not for the surface active material (surfactant) present in the fluid that coats the lungs.[a–e] In fact, the lack of adequate surfactant is the cause of **respiratory distress syndrome**, a major cause of death among premature infants and a disease that may develop in acute form in adults. The surfactant material forms a thin film of high fluidity at the air–liquid interface and lowers the surface tension from the 72 mN/m of pure water to <10 mN/m.[f,g] (Pay attention to the definition of surface tension.[h]) About 65% by weight of the surfactant is lecithin, mostly dipalmitoylphosphatidylcholine (see Fig. 8-4), a phospholipid resistant to attack by oxygen. Phosphatidylglycerol, in an unusually high concentration, accounts for ~ 12% of the human surfactant. Other phospholipids, plasmalogen,[i] cholesterol, proteins, and calcium ions are also present.

The surfactant contains four unique proteins, designated SP-A, SP-B, SP-C, and SP-D.[c,e] The major protein (SP-A) is a sialic acid-rich glycoprotein derived from a 26-kDa peptide, which contains a short collagen-like domain.[e,j,k] Like collagen, this domain contains glycosylated hydroxyproline. The C-terminal domain is a $Ca^{2+}$-dependent C-type lectin (Chapter 4), while the N-terminal domain is involved in oligomer formation through disulfide bridges. The overall structure is similar to that of the complement protein C1q (Chapter 31).[e,j] Protein D is also collagen-like[l] but evidently plays a very different functional role than SP-A. The latter associates with the major surfactant lipids but SP-D does not. It does bind phosphatidylinositol[m] and glucosylceramide, lipids that are present in small amounts. Perhaps SP-D helps to remove these polar lipids which might interfere with surfactant action.[e] Both proteins A and D may also have functions in the immune system.[l]

Proteins SP-B and SP-C are small extremely hydrophobic polypeptides consisting of 79 and 35 amino acid residues, respectively.[n,o] Aliphatic branched amino acids constitute 23 of the 35 residues of the C-terminal part of protein C, which is also palmitoylated on two cysteine residues. SP-B is formed from a large 381-residue precursor. The mature protein contains seven cysteines and disulfide bridges. Both proteins have major effects on the properties of the surfactant mixture. They promote rapid reorganization of lipid layers, an important consideration for the functioning of the surfactant. Infants lacking SP-B suffer severe respiratory failure with high mortality.[e]

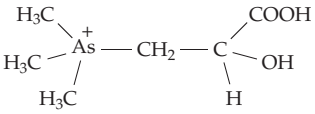
The properties of the surfactant allow for rapid formation of a large area of lipid monolayer–air interface. The low surface tension and the ability to rapidly spread the mixture of lipids and proteins are essential.[o] At the end of the expiration stage of breathing the surfactant is present in the interface as a strong, tightly packed monolayer whose properties reflect the rigidity of the dipalmitoylphosphatidylcholine. The excess surfactant in the alveolar fluid forms liposome-like bilayer structures and also associates with proteins and calcium ions to form a lattice-like material called **tubular myelin**. Lipid in this form must be transferred rapidly into the air–liquid interface during inspiration.[p] The ability of phospholipids to pass through a hexagonal phase (Fig. 8-12) may also be important for this transition.[q,r]

One enzyme present in the surfactant fluid is an acid phosphatase able to hydrolyze phosphatidylglycerol phosphate, perhaps functioning in the final step of biosynthesis of the phosphatidylglycerol present in the surfactant.[c] Study of the action of the natural lung surfactant has led to development of artificial surfactant mixtures that are being used effectively to save many lives.[d]

---


[a] Rooney, S. A. (1979) *Trends Biochem. Sci.* **4**, 189–191
[b] Persson, A., Chang, D., Rust, K., Moxley, M., Longmore, W., and Crouch, E. (1989) *Biochemistry* **28**, 6361–6367
[c] Rooney, S. A., Young, S. L., and Mendelson, C. R. (1994) *FASEB J.* **8**, 957–967
[d] Jobe, A. H. (1993) *N. Engl. J. Med.* **328**, 861–868
[e] Kuroki, Y., and Voelkers, D. R. (1994) *J. Biol. Chem.* **269**, 25943–25946
[f] Shiffer, K., Hawgood, S., Haagsman, H. P., Benson, B., Clements, J. A., and Goerke, J. (1993) *Biochemistry* **32**, 590–597
[g] Pastrana-Rios, B., Flach, C. R., Brauner, J. W., Mautone, A. J., and Mendelsohn, R. (1994) *Biochemistry* **33**, 5121–5127
[h] Bangham, A. D. (1992) *Nature (London)* **359**, 110
[i] Rana, F. R., Harwood, J. S., Mautone, A. J., and Dluhy, R. A. (1993) *Biochemistry* **32**, 27–31
[j] McCormack, F. X., Calvert, H. M., Watson, P. A., Smith, D. L., Mason, R. J., and Voelker, D. R. (1994) *J. Biol. Chem.* **269**, 5833–5841
[k] Taneva, S., McEachren, T., Stewart, J., and Keough, K. M. W. (1995) *Biochemistry* **34**, 10279–10289
[l] Crouch, E., Persson, A., Chang, D., and Heuser, J. (1994) *J. Biol. Chem.* **269**, 17311–17319
[m] Ogasawara, Y., Kuroki, Y., and Akino, T. (1992) *J. Biol. Chem.* **267**, 21244–21249
[n] Korimilli, A., Gonzales, L. W., and Guttentag, S. H. (2000) *J. Biol. Chem.* **275**, 8672–8679
[o] Pastrana, B., Mautone, A. J., and Mendelsohn, R. (1991) *Biochemistry* **30**, 10058–10064
[p] Lipp, M. M., Lee, K. Y. C., Zasadzinski, J. A., and Waring, A. J. (1996) *Science* **273**, 1196–1199
[q] Perkins, W. R., Dause, R. B., Parente, R. A., Minchey, S. R., Neuman, K. C., Gruner, S. M., Taraschi, T. F., and Janoff, A. S. (1996) *Science* **273**, 330–332
[r] Discher, B. M., Maloney, K. M., Grainger, D. W., Sousa, C. A., and Hall, S. B. (1999) *Biochemistry* **38**, 374–383

resistant to phospholipase $A_1$ as well. These two properties appear to protect the naked cell membranes of the protozoa from their own phospholipases which may be secreted into the environment.[26] Marine algae form an arsenic-containing phospholipid *O*-phosphatidyltrimethylarsonium lactic acid.[27]

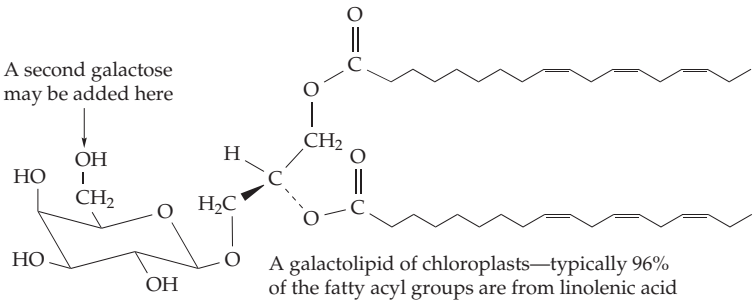


Trimethylarsonium lactic acid

They apparently do this as part of a scheme for detoxifying arsenate taken up with phosphate from the phosphate-poor ocean water.

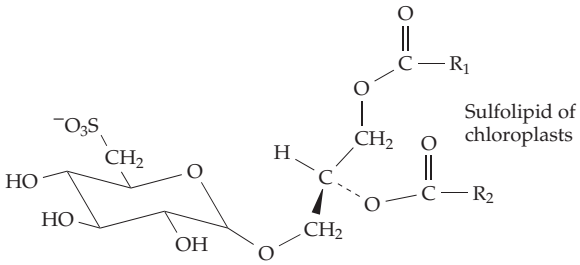
## 4. Glycolipids

The polar heads of the **glycoglycerolipids** lack phospho groups but contain *sugars* in glycosidic linkage.[28] Large amounts of the galactolipids shown in the following structure are found in chloroplasts.[29] The monogalactosyl diacylglycerol is said to be the most abundant polar lipid in nature.



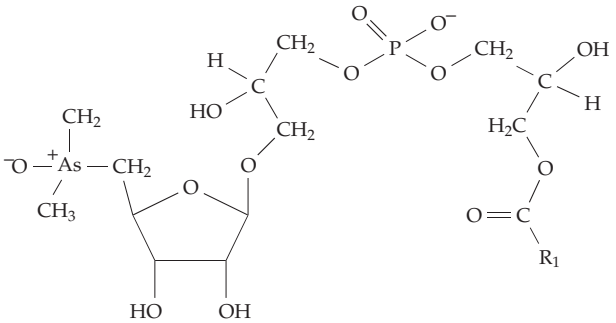

A galactolipid of chloroplasts—typically 96% of the fatty acyl groups are from linolenic acid

Chloroplasts also contain the following **sulfolipid**, an anionic sulfonate.



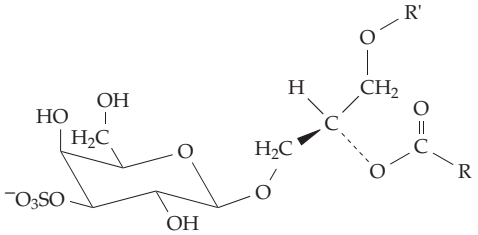

6-Sulfo-6-deoxy-α-D glucopyranosyl diglyceride

Marine algae as well as aquatic higher plants accumulate **arsenophospholipids**.[30]



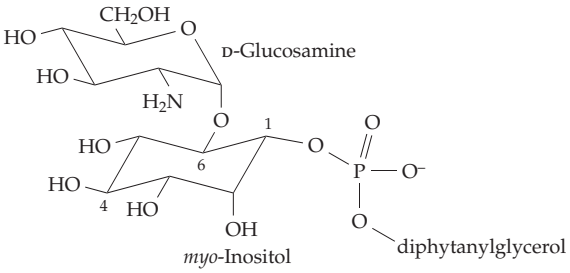
Arsenophospholipid of aquatic plants

The plasma membrane of mammalian male germ cells contains the following **sulfogalactosylglycerolipid**. It is found only in spermatozoa and testes, in which it accounts for 5–8% of total lipid, and in the brain, in which it accounts for only 0.2% of total lipid.[31]

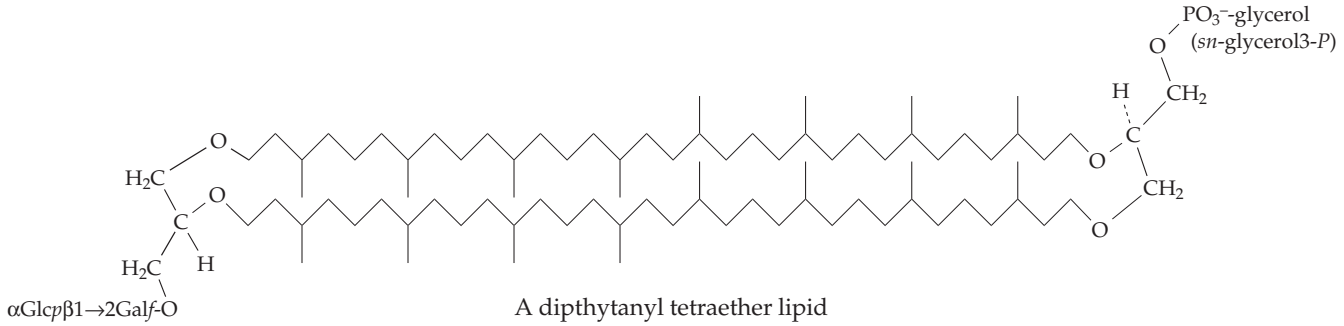


Sulfogalactosylglycerolipid, a sulfate ester

A variety of acylated glucolipids and phosphoglucolipids, including monoglucosyl and diglucosyl diacylglycerols, have been identified in membranes of the cell-wall-less bacterium *Acholeplasma laidlawii*.[32,32a] The following glycolipid from the methanogen *Methanosarcina*[33] is identical to the core structure of eukaryotic glycosylphosphatidylinositol membrane protein anchors (Fig. 8-13).




In addition to the previously mentioned phytanyl ether phospholipids, methanogens contain **diphytanyl tetraether lipids** that are both glycerophospholipids and glycolipids.

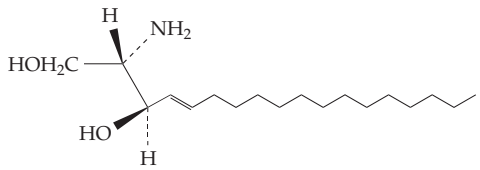



A dipthytanyl tetraether lipid

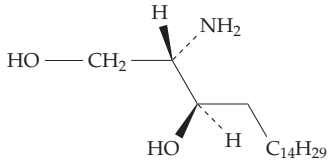
From a methanogenic bacterium.[21]

## 5. Sphingolipids

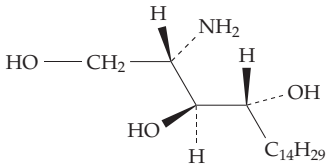
The backbone of the sphingolipids is the basic alcohol **sphingosine** (sphingenine) or a related long chain base. At least 60 such bases have been identified.[34,35] They vary in chain length from $C_{14}$ to $C_{22}$ and include members of the branched iso and anteiso series. Up to two double bonds may be present. Sphingosine contains 18 carbon atoms and is formed from palmitic acid and serine (Fig. 21-6). An intermediate in the formation of sphingosine is the saturated **sphinganine** (dihydrosphingosine), which is also a common component of animal sphingolipids. Hydroxylation of sphinganine to **phytosphingosine** occurs in both plants and animals, especially within glycolipids. The name comes from the fact that phytosphingosine was first discovered in plants.



Sphingosine (sphingenine)



Sphinganine (dihydrosphingosine)

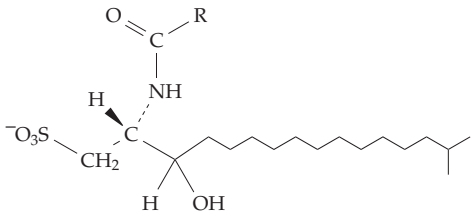


Phytosphingosine

Sphingosine-containing lipids are classified as **sphingophospholipids** (sphingomyelins) and **sphingoglycolipids**. In both cases the sphingosine is combined in amide linkage with a fatty acid to form a **ceramide** (Fig. 8-3) which still contains a free hydroxyl group able to combine with another component. In the sphingomyelins, which were first isolated from human brain by Thudicum in 1884, the additional component is usually phosphocholine (Fig. 8-3). Ceramide aminoethylphosphonates and related glycolipids occur in some invertebrates.[36,37]

The **cerebrosides** are glycosides of ceramide containing galactose or glucose. They are found in relatively large amounts in the brain where monogalactosylceramide predominates. Cerebrosides also occur in other animal tissues and to a lesser extent in plants. Many glycosphingolipids contain oligosaccharides of various sizes. When the oligosaccharide contains one or more residues of sialic acid the compound is known as a **ganglioside**.[38] Sialic acids are never found in plants. However, plants and fungi contain **phytoglycolipids**, which resemble gangliosides. Some contain inositol phosphates as well as sugars.[39–41] The structures of gangliosides may be very complex. Like glycoproteins, these substances are often located at the outer surface of cells where they may act as receptors for toxins, viruses, and hormones (see Section D).[42] Some of them carry attached blood group antigens (Box 4-C).[43,44] The sulfate esters of cerebrosides, known as **sulfatides** (Fig. 8-3), are also important components of membranes.[45]

Gliding bacteria such as *Capnocytophaga* contain sulfonolipids:[46]



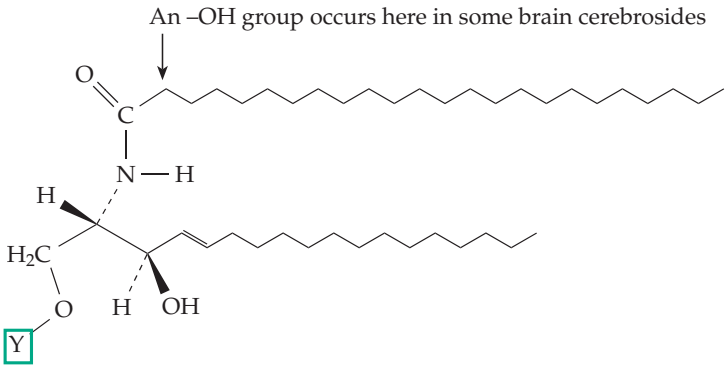
*N*-acylcapnine

A 1-deoxy-1-sulfonate analog of ceramides from gliding bacteria. R is a long-chain alkyl group.

These are analogs of the ceramides. The sulfonolipids seem to be necessary for the gliding movement of these bacteria across solid surfaces.[47]

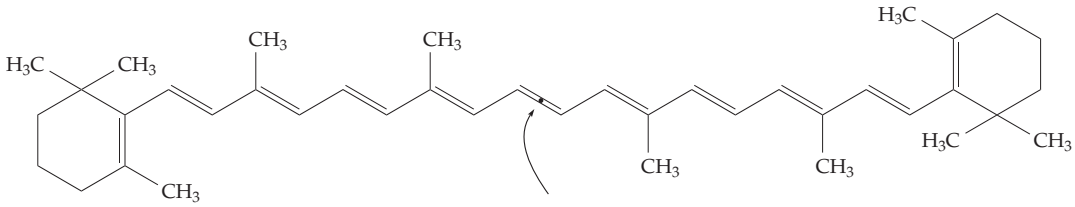
## 6. Sterols and Other Isoprenoid Lipids

A large group of isoprenoid lipids, including **sterols**, **terpenes**, and **carotenoid compounds**, are often present in membranes or in extracted lipids. Among these are the fat-soluble vitamins A, D, E, and K and the high polymers **rubber** and **gutta-percha**. The phytyl group of chlorophyll (Fig. 23-30), the phytanyl and diphytanyl groups of the lipids of methanogens, and the side chain of the pigment heme *a* (Fig. 16-5) are all related and are all derived from the precursor **prenyl pyrophosphate** (isopentenyl diphosphate). The whole group of compounds are often referred to as **polyprenyl** or by the older designation **isoprenoid**. The major discussion of polyprenyl compounds is found in Chapter 22 but the role of cholesterol and related compounds in membranes and the function of polyprenyl groups as membrane anchors for some proteins will be considered in this chapter.



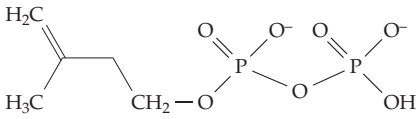

| | |
|---|---|
| Y = —H | The compounds in which Y = H are **ceramides**. Sphingolipids are often named as ceramide derivatives. |
| Y = —O—P(=O)(O⁻)—O—$CH_2CH_2$—$N^+(CH_3)_3$ (Phosphocholine) | **Sphingomyelins** |
| Y = —P(=O)(O⁻)—$CH_2CH_2NH_3^+$ | Ceramide aminoethyl phosphonates |
| Y = D-Galβ or D-Galβ1→4Glc– (The galactose bears a 3-sulfate group in cerebroside sulfatides, e.g., in lactosyl ceramide sulfate) | **Cerebrosides** or ceramide mono- and oligosaccharides |
| Y = GalNAc1→3Gal1→4Gal1→4Glc | Present in sphingolipid of red blood cell membranes |
| Y = —O—P(=O)(O⁻)—O—Inositol—Mannose | Present in **phytoglycolipid** of yeast |

**Figure 8-3** Structures of some sphingolipids.

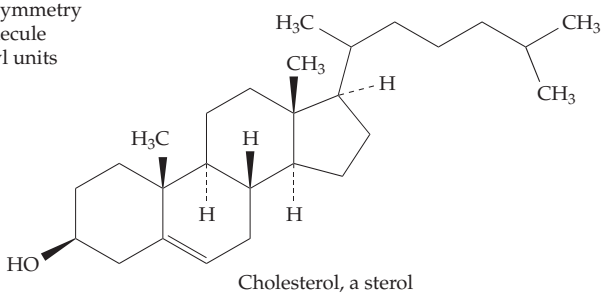



β-Carotene, a yellow pigment of carrots and other plants; converted in the human body to vitamin A.

Notice the center of symmetry in this 40-carbon molecule derived from 8 prenyl units




Prenyl diphosphate (pyrophosphate) (isopentenyl pyrophosphate), precursor of steroid,carotenoid, and other isoprenoid substances




Cholesterol, a sterol

## B. Membranes

In Chapter 7 we examined ways in which protein subunits can be stacked to form helices and closed oligomers. Another important arrangement of cell constituents is that of flat sheets or membranes.[48–54] Chemists, physicists, and biologists have mounted a sustained effort to understand these thin but remarkably tough outer surfaces of cells. However, consider the fact that a 7- to10-nm-thick plasma membrane of a cell of 10-µm diameter has less than 1/1000 the thickness of the cell and occupies only 0.5% of the total volume. The technical difficulties in studying such a membrane are great and are compounded by the fact that a cell contains more than one kind of membrane.

Membranes from many sources have been studied. One of these is **myelin**, the multilayered insulation that surrounds the axons of many nerve cells.[55–57] Myelin is derived from the plasma membrane of **Schwann cells** which lie adjacent to many neurons. Schwann cells literally wrap themselves around neuronal axons. Their cytoplasm is squeezed out leaving little but tightly packed membrane layers. Myelin membranes are the most stable known and also have the highest lipid content (80%). Another readily available experimental material is the plasma membrane of the human red bood cell, which can be prepared by osmotic rupture of the cells. The remaining **erythrocyte ghosts** contain ~1% of the dry matter of the cell and may have been studied more than any other membrane. A much investigated specialized membrane is the outer portion of the visual receptor cells known as **rods** (Chapter 23), which contains a closely packed and regular array of flat discs, each one consisting of a pair of membranes. Both membranes and cell walls of many kinds of bacteria have also been investigated.

### 1. The Structure of Membranes

Membranes consist largely of protein and lipid. The ratio (by weight) of protein to lipid varies from 0.25 in myelin to ~3.0 in bacterial membranes. In membranes of erythrocytes it is about 1.2 and a ratio of about 1.0 may be regarded as typical for animal cells. Small amounts of carbohydrates (<5%) are present, as are traces of RNA (<0.1%).

In 1926, Gorter and Grendel calculated that the erythrocyte ghost contained just enough lipid to form a 3.0- to 4.0-nm-thick layer around the cell. Apparently they reached this correct conclusion only because their measurements of pressures of surface films contained compensating errors.[58] Nevertheless, this information, together with the known propensity of lipids to aggregate in **micelles** in which the hydrocarbon "tails" clustered together and the polar "heads" protruded into the surrounding water,[59] led Danielli and Davsen in 1935 to propose the **lipid bilayer** structure for membranes.[60] Its essential features are indicated in Fig. 8-4. Hydrophobic bonding holds the extended hydrocarbon chains together, while the polar groups of the phospholipid molecules may interact with proteins on the sides of the bilayer. The original proposal assumed an extended β structure for the proteins, which would allow them to coat the bilayer uniformly on both sides. However, this is not correct. Proteins are sometimes embedded in the bilayer, sometimes protrude through it, and sometimes are attached on one surface, most often the cytoplasmic surface of the plasma membrane. These concepts were brought together by Singer and Nicolson in 1972 in the **fluid mosaic model** of membrane structure[61] (Fig. 8-5). The lipid bilayer still provides the basic structure upon which the complex membranes of living organisms are assembled.[52,54,62–65] The term "fluid" refers to the fact

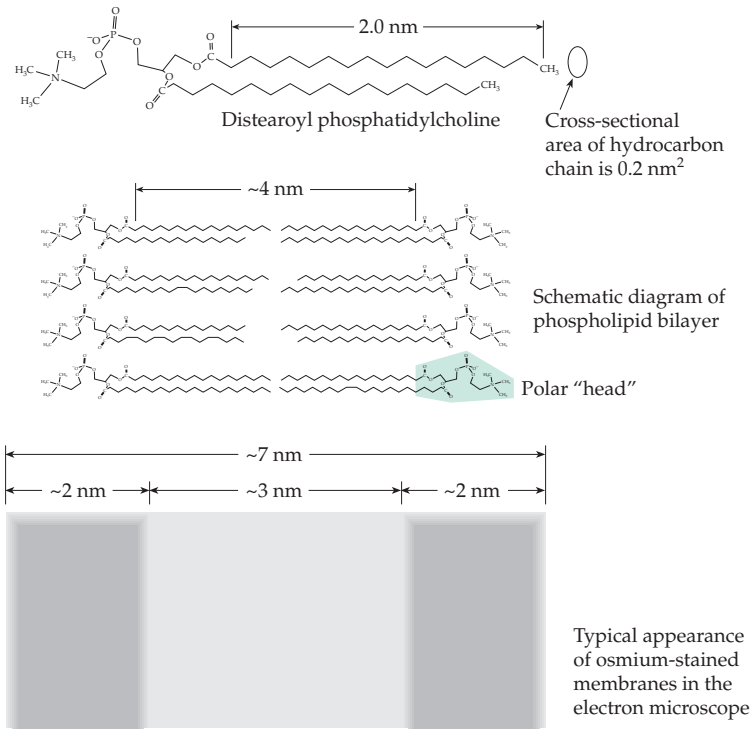



**Figure 8-4** Bimolecular lipid layers and membranes. (Top) A molecule of phosphatidylcholine. (Center) Lipid bilayer structure. (Bottom) Bilayer structure as seen by the electron microscope with osmium tetroxide staining.

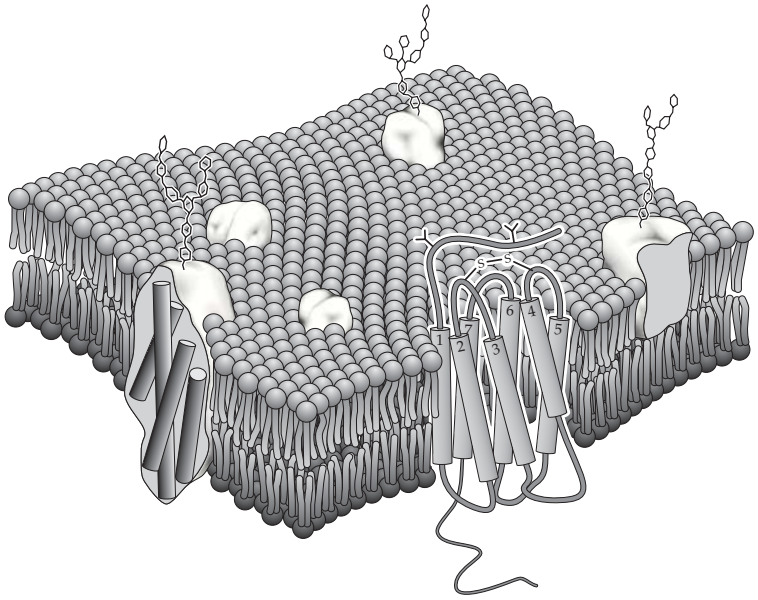


**Figure 8-5** The fluid mosaic model of Singer and Nicolson.[61] Some integral membrane proteins, which are shown as irregular solids, are dissolved in the bilayer. Transmembrane proteins protrude from both sides. One of these is pictured as a seven-helix protein, a common type of receptor for hormones and for light absorption by visual pigments. Other proteins adhere to either the outer or the inner surface. Many membrane proteins carry complex oligosaccharide groups which protrude from the outer surface (Chapter 4). A few of these are indicated here as chains of sugar rings.

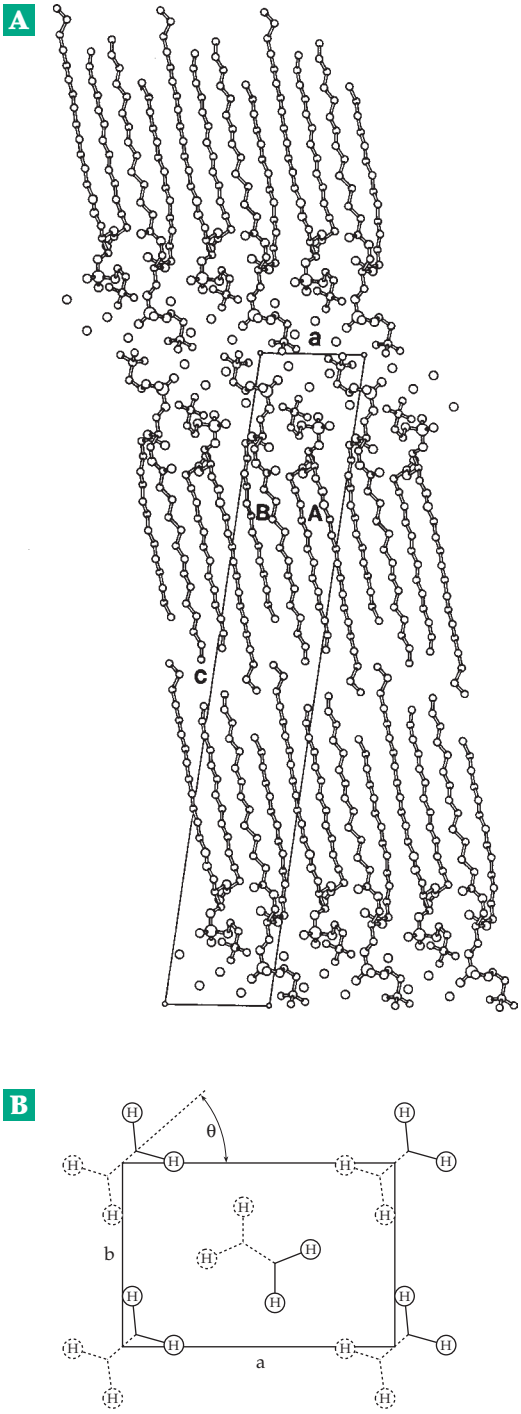



**Figure 8-6** (A) Molecular packing of 2,3-dimyristoyl-D-glycero-1-phosphocholine dihydrate. The two molecules in the asymmetric unit are labeled 1 and 2. The position of the water molecules is indicated either by W1–W4 or by small open circles. Hydrogen bonds are represented by dotted lines. From Pascher *et al.*[66] (B) Two-dimensional "orthorhombic" arrangement of hydrocarbon chains in a crystalline alkane. The *a*–*b* plane corresponds to the plane of the bilayer surface; the long axes of the acyl chains project from the page. From Cameron *et al.*[67]

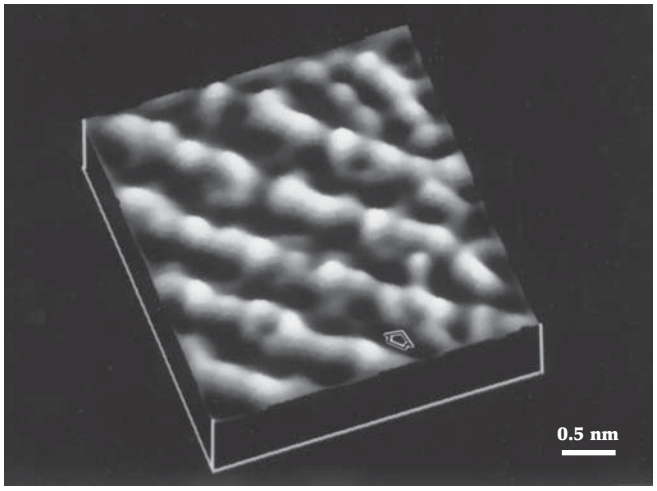



**Figure 8-7** Atomic force microscope image of a dimyristoylphosphatidylethanolamine bilayer deposited by the Langmuir–Blodgett technique (see Fig. 8-8) at a specific molecular area of 0.41 $nm^2$ and a surface pressure of 40 mN/m on a freshly cleaved mica substrate. The images were taken under water. The long, uniformly spaced rows are roughly 0.7–0.9 nm in spacing. The modulation along the rows, with rounded bright spots roughly every 0.5 nm, corresponds to the individual headgroups of the phosphatidylethanolamine molecules. The lattice spacing is identical to that measured by X-ray diffraction at the air–water interface. The area per molecule in the AFM image is ~0.4 $nm^2$. From Zasadzinski *et al.*[68]

that at temperatures suitable for growth and metabolism the hydrocarbon chains are not rigidly packed in the center of the bilayer but are "molten" (see Section 2). However, at a low enough temperature they become rigid and pack together in a manner similar to that of the chains in crystals of the phosphatidylcholine shown in Fig. 8-6A. These crystals consist of stacked bilayers of thickness 5.5 nm.[66,69] The scanning tunneling and atomic force microscopes have provided direct views of a similar arrangement of side chains in a monomolecular fatty acid layer[68,70,71] (Fig. 8-7). Measurements on multilamellar vesicles of dipalmitoylphosphatidylcholine give bilayer thicknessess from 5.4 nm for dehydrated vesicles to 6.7 nm for the biologically relevant fully hydrated bilayers.[72]

***Lipids of membranes.*** Approximately 1500 different lipids have been identified in the myelin of the central nervous system of humans. About 30 of these are present in substantial amounts.[73] The distribution of the different lipids varies markedly between membranes from different sources (Table 8-3) making generalization difficult. However, phospholipids are apparently always present and, except in chloroplasts, make up from 40% to over 90% of the total lipid (Table 8-3).

Five kinds of phospholipid predominate: phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, phosphatidylglycerols, and sphingomyelin. Usually there are also small amounts of phosphatidylinositol. The major phospholipid in animal cells is phosphatidylcholine, but in bacteria it is phosphatidylethanolamine. The phospholipids of *E. coli* consist of 80% phosphatidylethanolamine, 15% phosphatidylglycerol, and 5% diphosphatidylglycerol (cardiolipin). Significant amounts of cardiolipin are found only in bacteria and in the inner membrane of mitochondria. Sphingomyelin is almost absent from mitochondria, endoplasmic reticulum, or nuclear membranes.

Glycolipids are important constituents of the plasma membranes, of the endoplasmic reticulum, and of chloroplasts. The cerebrosides and their sulfate esters, the sulfatides, are especially abundant in myelin. In plant membranes, the predominant lipids are the galactosyl diglycerides.[29,74] The previously described ether phospholipids (archaebacteria), ceramide aminoethylphosphonate (invertebrates), and sulfolipid (chloroplasts) are also important membrane components.

Cholesterol makes up 17% of myelin and is present in plasma membranes. However, it usually does not occur in bacteria and is present only in trace amounts in mitochondria. Related sterols are present in plant membranes. Esters of sterols occur as transport forms but are not found in membranes. Membrane bilayers, likewise, contain little or no triacylglycerols, the latter being found largely as droplets in the cytoplasm.

Quantitatively minor membrane components with important biological functions include **ubiquinone**, which is present in the inner mitochondrial membrane, and the **tocopherols**. Plant chloroplast membranes contain chlorophyll, carotenes, and other lipid-soluble pigments.

***Liquid crystals, liposomes, and artificial membranes.*** Phospholipids dissolve in water to form true solutions only at very low concentrations (~$10^{-10}$ M for distearoyl phosphatidylcholine). At higher concentrations they exist in **liquid crystalline phases** in which the molecules are partially oriented. Phosphatidylcholines (lecithins) exist almost exclusively in a **lamellar** (smectic) phase in which the molecules form bilayers. In a warm phosphatidylcholine–water mixture containing at least 30% water by weight the phospholipid forms **multilamellar vesicles**, one lipid bilayer surrounding another in an "onion skin" structure. When such vesicles are subjected to ultrasonic vibration they break up, forming some very small vesicles of diameter down to 25 nm which are surrounded by a single bilayer. These unilamellar vesicles are often used for study of the properties of bilayers. Vesicles of both types are often called **liposomes**.[75–77]

When liposomes are stained with osmium tetroxide or potassium permanganate, embedded, and sectioned