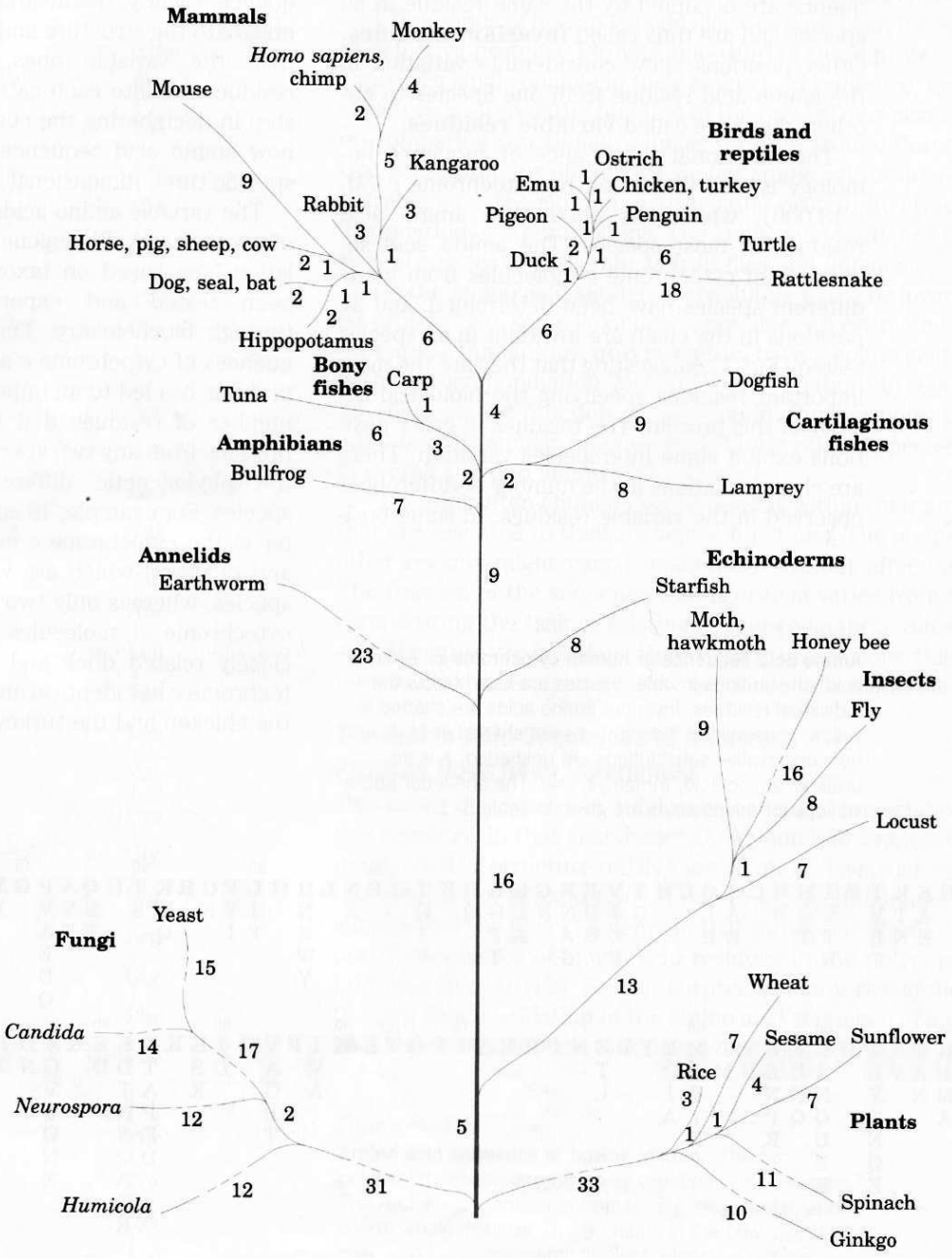
and sheep. Information on the number of residue differences between homologous proteins of different species allows the construction of evolutionary trees that show the origin and sequence of appearance of different species during the course of evolution (Fig. 2). The relationships established by anatomic and biochemical taxonomy are in close agreement.

figure 2
Main branches of the eukaryotic evolutionary tree constructed from the number of amino acid differences between cytochrome *c* molecules of various species. The numbers represent the number of residues by which the cytochrome *c* of a given line of organism differs from its ancestor. Branch points reflect a common ancestor.



EMPOWER - 1001 - Pt. 2

sequence in DNA and the amino acid sequence in proteins were somehow related. Barely a decade after these discoveries, the role of DNA nucleotide sequence in determining the amino acid sequence of protein molecules had been revealed (Chapter 27).

The amino acid sequences of thousands of different proteins from many species have been determined using principles first developed by Sanger. These methods are still in use, although with many variations and improvements in detail. Chemical protein sequencing now complements a growing list of newer methods, providing multiple avenues to obtain amino acid sequence data. Such data are now critical to every area of biochemical investigation.

Short Polypeptides Are Sequenced Using Automated Procedures

Various procedures are used to analyze protein primary structure. One is to hydrolyze the protein and determine its amino acid composition (Fig. 5–25a). This information is often valuable in interpreting the results of other procedures. Because amino acid composition differs from one protein to the next, it can serve as a kind of fingerprint. It can be used, for example, to help determine whether proteins isolated by different laboratories are the same or different. Hydrolysis alone, however, cannot be used to determine the *sequence* of amino acids in a protein.

A procedure that is often used in conjunction with hydrolysis is to label and identify the amino-terminal amino acid residue (Fig. 5–25b). For this purpose Sanger developed the reagent 1-fluoro-2,4-dinitrobenzene (FDNB).

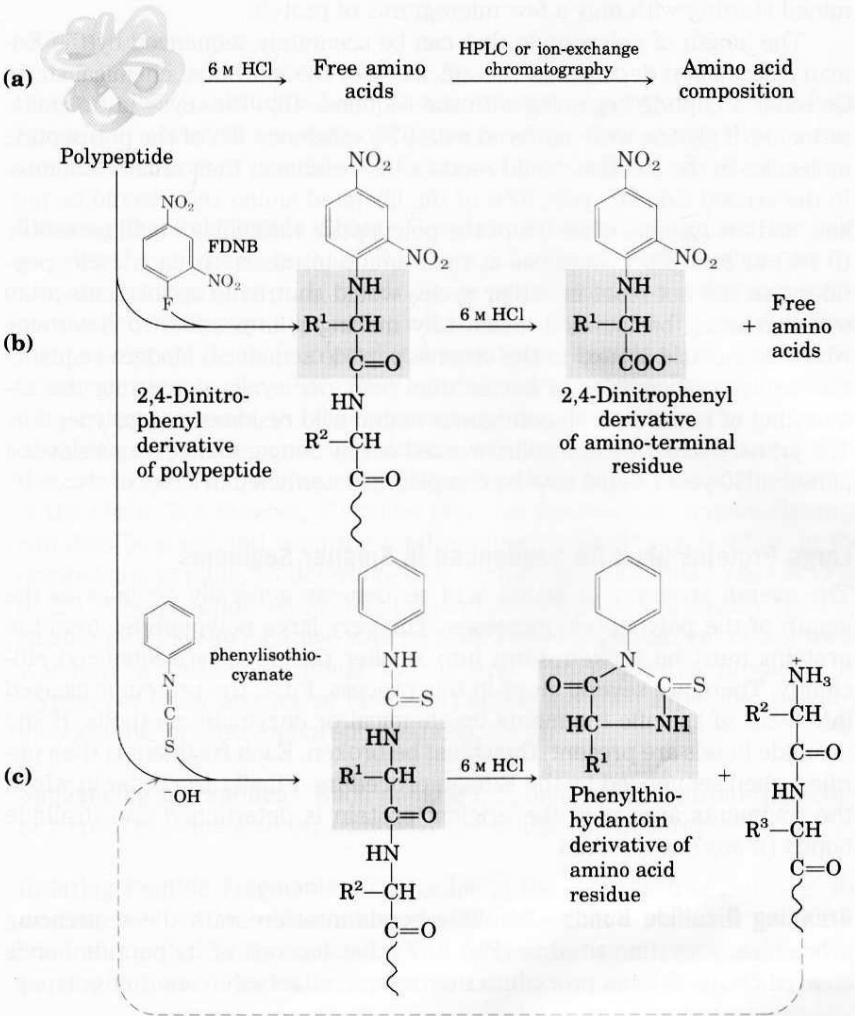
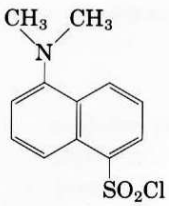
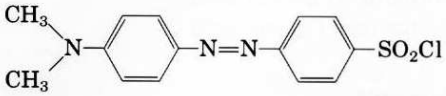


figure 5–25

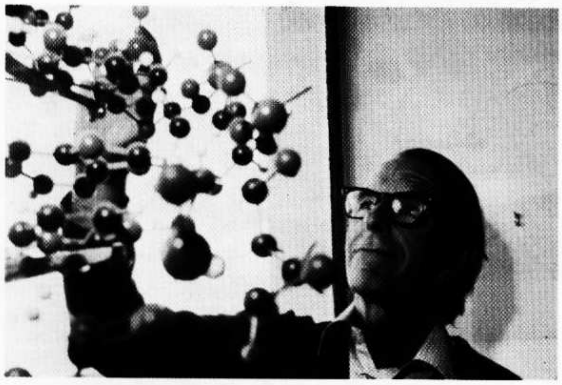
Steps in sequencing a polypeptide. (a) Determination of amino acid composition and **(b)** identification of the amino-terminal residue are often the first steps in sequencing a polypeptide. Sanger's method for identifying the amino-terminal residue is shown here. The Edman degradation procedure **(c)** reveals the entire sequence of a peptide. For shorter peptides, this method alone readily yields the entire sequence, and steps **(a)** and **(b)** are often omitted. The latter procedures are useful in the case of larger polypeptides, which are often fragmented into smaller peptides for sequencing (see Fig. 5–27).



Dansyl chloride



Dabsyl chloride



Frederick Sanger

Other reagents used to label the amino-terminal residue, dansyl chloride and dabsyl chloride, yield derivatives that are more easily detectable than the dinitrophenyl derivatives. After the amino-terminal residue is labeled with one of these reagents, the polypeptide is hydrolyzed to its constituent amino acids and the labeled amino acid is identified. Because the hydrolysis stage destroys the polypeptide, this procedure cannot be used to sequence a polypeptide beyond its amino-terminal residue. However, it can help determine the number of chemically distinct polypeptides in a protein, provided each has a different amino-terminal residue. For example, two residues—Phe and Gly—would be labeled if insulin (Fig. 5–24) were subjected to this procedure.

To sequence an entire polypeptide, a chemical method devised by Pehr Edman is usually employed. The **Edman degradation** procedure labels and removes only the amino-terminal residue from a peptide, leaving all other peptide bonds intact (Fig. 5–25c). The peptide is reacted with phenylisothiocyanate, and the amino-terminal residue is ultimately removed as a phenylthiohydantoin derivative. After removal and identification of the amino-terminal residue, the *new* amino-terminal residue so exposed can be labeled, removed, and identified through the same series of reactions. This procedure is repeated until the entire sequence is determined. The Edman degradation is carried out on a machine, called a **sequenator,** which mixes reagents in the proper proportions, separates the products, identifies them, and records the results. These methods are extremely sensitive. Often, the complete amino acid sequence can be determined starting with only a few micrograms of protein.

The length of polypeptide that can be accurately sequenced by the Edman degradation depends on the efficiency of the individual chemical steps. Consider a peptide beginning with the sequence Gly–Pro–Lys– at its amino terminus. If glycine were removed with 97% efficiency, 3% of the polypeptide molecules in the solution would retain a Gly residue at their amino terminus. In the second Edman cycle, 97% of the liberated amino acids would be proline, and 3% glycine, while 3% of the polypeptide molecules would retain Gly (0.1%) or Pro (2.9%) residues at their amino terminus. At each cycle, peptides that did not react in earlier cycles would contribute amino acids to an ever-increasing background, eventually making it impossible to determine which amino acid is next in the original peptide sequence. Modern sequenators achieve efficiencies of better than 99% per cycle, permitting the sequencing of more than 50 contiguous amino acid residues in a polypeptide. The primary structure of insulin, worked out by Sanger and colleagues over a period of 10 years, could now be completely determined in a day or two.

Large Proteins Must Be Sequenced in Smaller Segments

The overall accuracy of amino acid sequencing generally declines as the length of the polypeptide increases. The very large polypeptides found in proteins must be broken down into smaller pieces to be sequenced efficiently. There are several steps in this process. First, the protein is cleaved into a set of specific fragments by chemical or enzymatic methods. If any disulfide bonds are present, they must be broken. Each fragment is then purified, then sequenced by the Edman procedure. Finally, the order in which the fragments appear in the original protein is determined and disulfide bonds (if any) are located.

Breaking Disulfide Bonds Disulfide bonds interfere with the sequencing procedure. A cystine residue (Fig. 5–7) that has one of its peptide bonds cleaved by the Edman procedure may remain attached to another polypep-

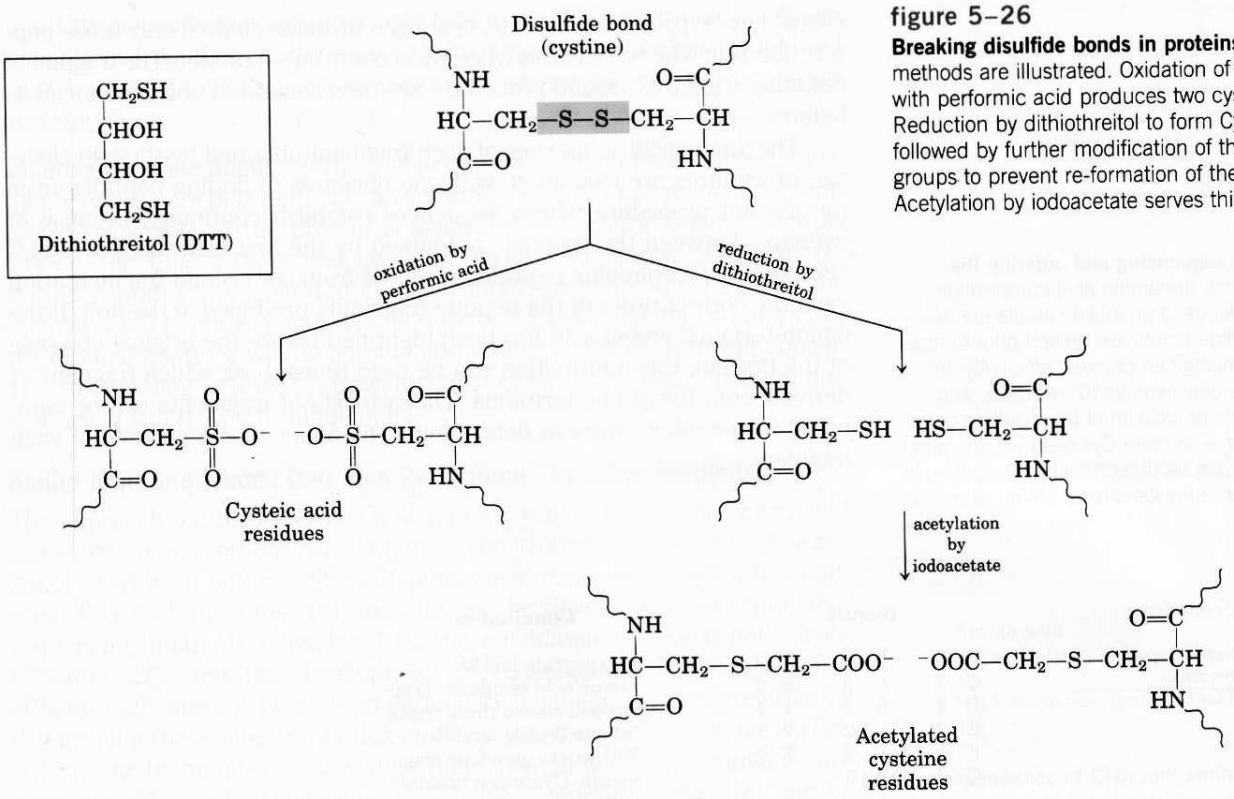


figure 5–26
Breaking disulfide bonds in proteins. Two common methods are illustrated. Oxidation of a cystine residue with performic acid produces two cysteic acid residues. Reduction by dithiothreitol to form Cys residues must be followed by further modification of the reactive —SH groups to prevent re-formation of the disulfide bond. Acetylation by iodoacetate serves this purpose.

tide strand via its disulfide bond. Disulfide bonds also interfere with the enzymatic or chemical cleavage of the polypeptide. Two approaches to irreversible breakage of disulfide bonds are outlined in Figure 5–26.

Cleaving the Polypeptide Chain Several methods can be used for fragmenting the polypeptide chain. Enzymes called **proteases** catalyze the hydrolytic cleavage of peptide bonds. Some proteases cleave only the peptide bond adjacent to particular amino acid residues (Table 5–7) and thus fragment a polypeptide chain in predictable and reproducible ways. A number of chemical reagents also cleave the peptide bond adjacent to specific residues.

Among proteases, the digestive enzyme trypsin catalyzes the hydrolysis of only those peptide bonds in which the carbonyl group is contributed by either a Lys or an Arg residue, regardless of the length or amino acid sequence of the chain. The number of smaller peptides produced by trypsin cleavage can thus be predicted from the total number of Lys or Arg residues in the original polypeptide, as determined by hydrolysis of an intact sample (Fig. 5–27). A polypeptide with five Lys and/or Arg residues will usually yield six smaller peptides on cleavage with trypsin. Moreover, all except one of these will have a carboxyl-terminal Lys or Arg. The fragments produced by trypsin (or other enzyme or chemical) action are then separated by chromatographic or electrophoretic methods.

Sequencing of Peptides Each peptide fragment resulting from the action of trypsin is sequenced separately by the Edman procedure.

Ordering Peptide Fragments The order of the "trypsin fragments" in the original polypeptide chain must now be determined. Another sample of the intact polypeptide is cleaved into fragments using a different enzyme or reagent, one that cleaves peptide bonds at points other than those

table 5–7

The Specificity of Some Common Methods for Fragmenting Polypeptide Chains

Treatment*	Cleavage points†
Trypsin	Lys, Arg (C)
Submaxillarus protease	Arg (C)
Chymotrypsin	Phe, Trp, Tyr (C)
Staphylococcus aureus V8 protease	Asp, Glu (C)
Asp-*N*-protease	Asp, Glu (N)
Pepsin	Phe, Trp, Tyr (N)
Endoproteinase Lys C	Lys (C)
Cyanogen bromide	Met (C)

*All except cyanogen bromide are proteases. All are available from commercial sources.

†Residues furnishing the primary recognition point for the protease or reagent; peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.

cleaved by trypsin. For example, cyanogen bromide cleaves only those peptide bonds in which the carbonyl group is contributed by Met. The fragments resulting from this second procedure are then separated and sequenced as before.

The amino acid sequences of each fragment obtained by the two cleavage procedures are examined, with the objective of finding peptides from the second procedure whose sequences establish continuity, because of overlaps, between the fragments obtained by the first cleavage procedure (Fig. 5–27). Overlapping peptides obtained from the second fragmentation yield the correct order of the peptide fragments produced in the first. If the amino-terminal amino acid has been identified before the original cleavage of the protein, this information can be used to establish which fragment is derived from the amino terminus. The two sets of fragments can be compared for possible errors in determining the amino acid sequence of each fragment.

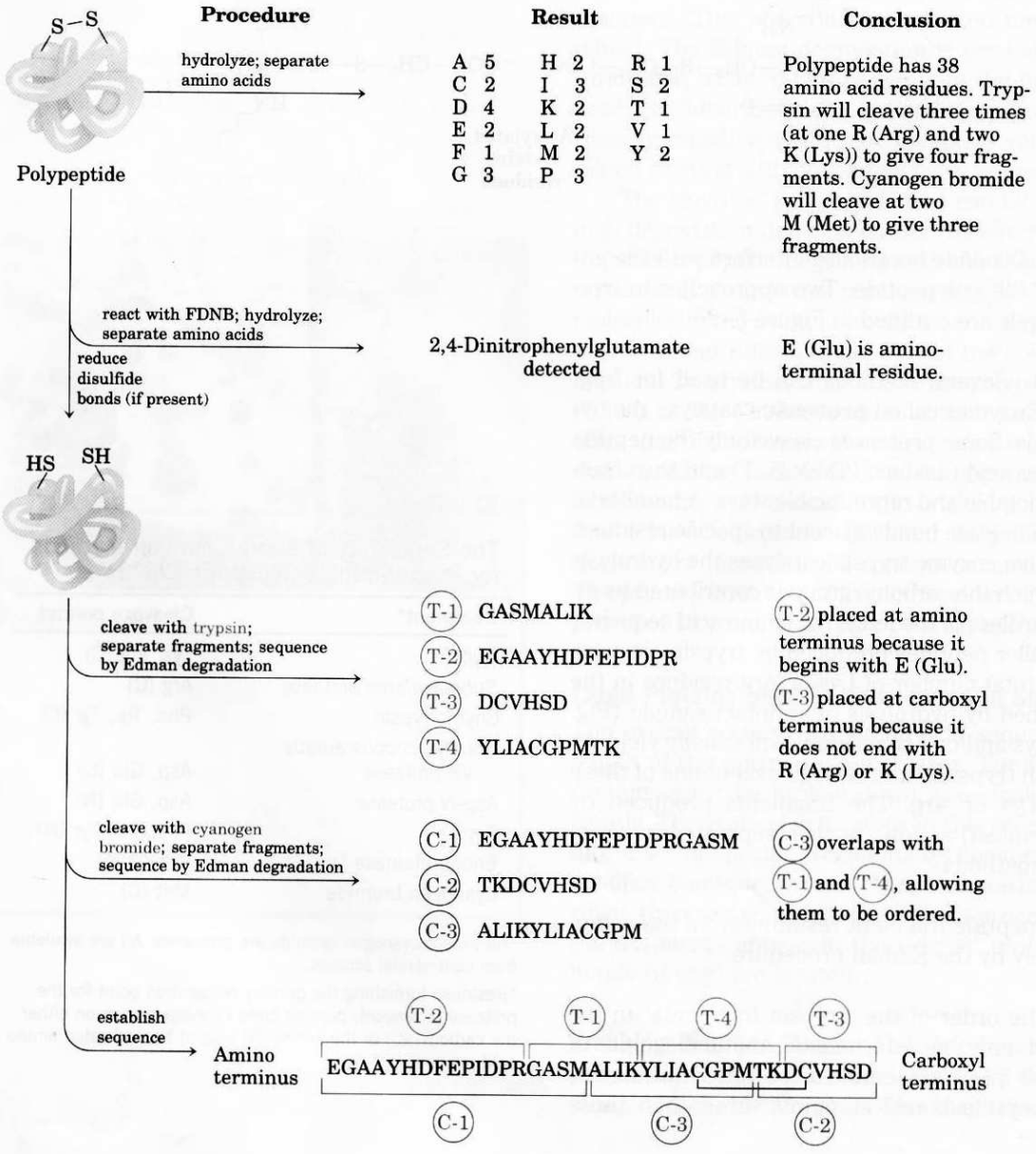


figure 5–27
Cleaving proteins and sequencing and ordering the peptide fragments. First, the amino acid composition and amino-terminal residue of an intact sample are determined. Then any disulfide bonds are broken prior to fragmenting so that sequencing can proceed efficiently. In this example, there are only two Cys (C) residues, and thus only one possibility for location of the disulfide bond. In polypeptides with three or more Cys residues, the position of disulfide bonds can be determined as described in the text. (The one-letter abbreviations for amino acids are given in Table 5–1.)

If the second cleavage procedure fails to establish continuity between all peptides from the first cleavage, a third or even a fourth cleavage method must be used to obtain a set of peptides that can provide the necessary overlap(s).

Locating Disulfide Bonds If the primary structure includes disulfide bonds, their locations are determined in an additional step after sequencing is completed. A sample of the protein is again cleaved with a reagent such as trypsin, this time without first breaking the disulfide bonds. When the resulting peptides are separated by electrophoresis and compared with the original set of peptides generated by trypsin, for each disulfide bond, two of the original peptides will be missing and a new, larger peptide will appear. The two missing peptides represent the regions of the intact polypeptide that are linked by the disulfide bond.

Amino Acid Sequences Can Also Be Deduced by Other Methods

The approach outlined above is not the only way to determine amino acid sequences. New methods based on mass spectrometry permit the sequencing of short polypeptides (20 to 30 amino acid residues) in just a few minutes (Box 5–3, pp. 146–149). In addition, the development of rapid DNA sequencing methods (Chapter 10), the elucidation of the genetic code (Chapter 27), and the development of techniques for isolating genes (Chapter 29) make it possible to deduce the sequence of a polypeptide by determining the sequence of nucleotides in the gene that codes for it (Fig. 5–28). The techniques used to determine protein and DNA sequences are complementary. When the gene is available, sequencing the DNA can be faster and more accurate than sequencing the protein. Most proteins are now sequenced in this indirect way. If the gene has not been isolated, direct sequencing of peptides is necessary, and this can provide information (e.g., the location of disulfide bonds) not available in a DNA sequence. In addition, a knowledge of the amino acid sequence of even a part of a polypeptide can greatly facilitate the isolation of the corresponding gene (Chapter 29).

The array of methods now available to analyze both proteins and nucleic acids is ushering in a new discipline of whole cell biochemistry. The complete sequence of an organism's DNA, its genome, is now available for organisms ranging from viruses to bacteria to multicellular eukaryotes (see Table 1–1). Genes are being discovered by the thousands, including many that encode proteins with no known function. To describe the entire protein complement encoded by an organism's DNA, researchers have coined the term **proteome.** Analysis of a cell's proteome is an increasingly important and informative adjunct to the completion of its genomic sequence. Proteins from a cell are separated and displayed by two-dimensional gel electrophoresis (Fig. 5–22). Individual protein spots can be extracted from such a gel. Small peptides derived from the proteins are sequenced by mass spectrometry (Box 5–3), and these sequences are compared with the genomic sequence to identify the protein. Often, knowledge of the sequence of a segment of six to eight amino acid residues is enough to pinpoint the gene encoding the entire protein. Inevitably, some of these proteins are already known and have well-studied functions; others are more mysterious. Once most of the proteins are matched to a gene, changes in a cell's protein complement brought on by the environment, nutritional changes, stress, or disease can be examined. This work can provide clues to the role of proteins whose functions are as yet unknown. Eventually, such studies will complement work carried out on cellular intermediary metabolism and nucleic acid metabolism to provide a new and increasingly complete picture of biochemistry at the level of cells and even organisms.

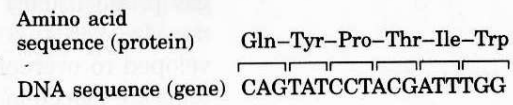


figure 5–28
Correspondence of DNA and amino acid sequences. Each amino acid is encoded by a specific sequence of three nucleotides in DNA. The genetic code is described in detail in Chapter 27.

box 5–3 Investigating Proteins with Mass Spectrometry

The mass spectrometer has long been an indispensable tool in chemistry. Molecules to be analyzed, referred to as **analytes,** are first ionized in a vacuum. When the newly charged molecules are introduced into an electric and/or magnetic field, their paths through the field are a function of their mass-to-charge ratio, m/z. This measured property of the ionized species can be used to deduce the mass (M) of the analyte with very high precision.

Although mass spectrometry has been in use for many years, it could not be applied to macromolecules such as proteins and nucleic acids. The m/z measurements are made on molecules in the gas phase, and the heating or other treatment needed to bring a macromolecule into the gas phase usually caused its rapid decomposition. In 1988, two different techniques were developed to overcome this problem. In one, proteins are placed in a light-absorbing matrix. With a short pulse of laser light, the proteins are ionized and then desorbed from the matrix into the vacuum system. This process, known as **matrix-assisted laser desorption/ionization mass spectrometry,** or **MALDI MS,** has been successfully used to measure the mass of a wide range of macromolecules. In a second and equally successful method, macromolecules in solution are forced directly from the liquid to gas phase. A solution of analytes is passed through a charged needle that is kept at a high electrical potential, dispersing the solution into a fine mist of charged microdroplets. The solvent surrounding the macromolecules rapidly evaporates and the resulting multiply charged macromolecular ions are thus introduced nondestructively into the gas phase. This technique is called **electrospray ionization mass spectrometry,** or **ESI MS.** Protons added during passage through the needle give additional charge to the macromolecule. The m/z of the molecule can be analyzed in the vacuum chamber.

Mass spectrometry provides a wealth of information for proteome research, enzymology, and protein chemistry in general. The techniques require only miniscule amounts of sample, so they can be readily applied to the small amounts of protein that can be extracted from a two-dimensional electrophoretic gel. The accurately measured molecular mass of a protein is one of the critical parameters in its identification. Once the mass of a protein is accurately known, mass spectrometry is a convenient and accurate method for detecting changes in mass due to the presence of bound cofactors, bound metal ions, covalent modifications, and so on.

The process for determining the molecular mass of a protein with ESI MS is illustrated in Figure 1. As it is injected into the gas phase, a protein acquires a variable number of protons,

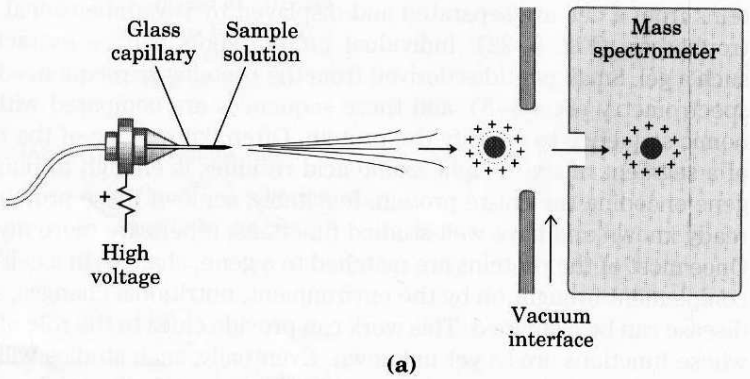


figure 1
Electrospray mass spectrometry of a protein. (a) A protein solution is dispersed into highly charged droplets by passage through a needle under the influence of a high-voltage electric field. The droplets evaporate, and the ions (with added protons in this case) enter the mass spectrometer for m/z measurement. The spectrum generated **(b)** is a family of peaks, with each successive peak (from right to left) corresponding to a charged species increased by 1 in both mass and charge. A computer-generated transformation of this spectrum is shown in the inset.

and thus positive charges, from the solvent. This creates a spectrum of species with different mass-to-charge ratios. Each successive peak corresponds to a species that differs from that of its neighboring peak by a charge difference of 1 and a mass difference of 1 (1 proton). The mass of the protein can be determined from any two neighboring peaks. The measured m/z of one peak is

$$(m/z)_2 = \frac{M + n_2X}{n_2}$$

where M is the mass of the protein, n_2 is the number of charges, and X is the mass of the added groups (protons in this case). Similarly for the neighboring peak,

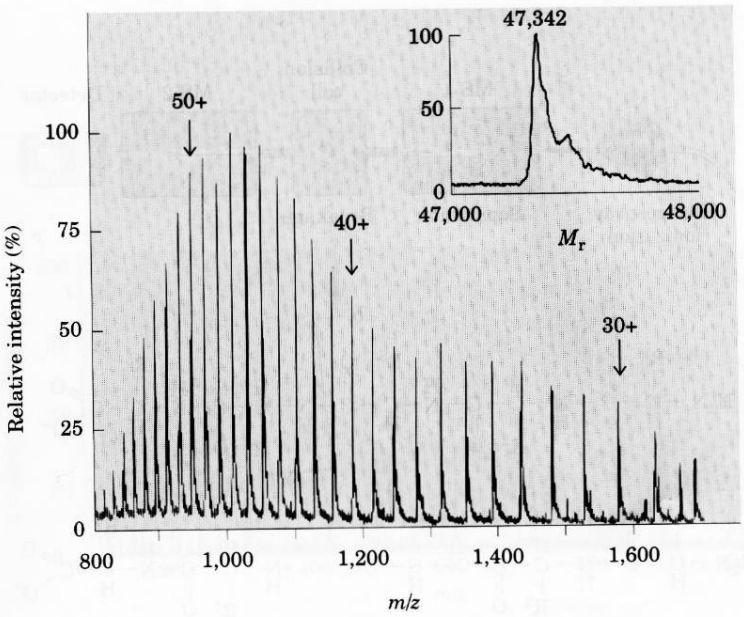
$$(m/z)_1 = \frac{M + (n_2 + 1)X}{n_2 + 1}$$

We now have two unknowns (M and n_2) and two equations. We can solve first for n_2 and then for M:

$$n_2 = \frac{(m/z)_2 - X}{(m/z)_2 - (m/z)_1}$$

$$M = n_2[(m/z)_2 - X]$$

This calculation using the m/z values for any two peaks in a spectrum such as that shown in Figure 1b will usually provide the mass of the protein (in this case, aerolysin k; 47,342 Da) with an error of only ±0.01%. Generating several sets of peaks, repeating the calculation, and averaging the results generally provides an even more accurate value for M. Computer algorithms can transform the m/z spectrum into a single peak that also provides a very accurate mass measurement (Fig. 1b, inset).



(b)

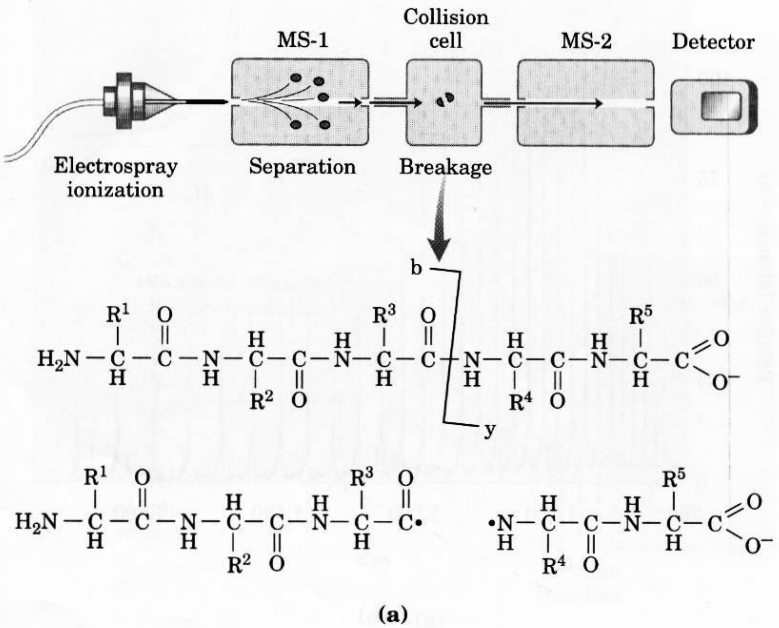
Continued on next page

Mass spectrometry can also be used to sequence short stretches of polypeptide, an application that has emerged as an invaluable tool for quickly identifying unknown proteins. Sequence information is extracted using a technique called **tandem MS,** or **MS/MS.** A solution containing the protein under investigation is first treated with a protease or chemical reagent to reduce it by hydrolytic cleavage to a mixture of shorter peptides. The mixture is then injected into a device that is essentially two mass spectrometers in tandem (Fig. 2a, top). In the first, the peptide mixture is sorted and the ionized fragments are manipulated so that only one of the several types of peptides produced by cleavage emerges at the other end. The sample of the selected peptide, each molecule of which has a charge somewhere along its length, then travels through a vacuum chamber between the two mass spectrometers. In this collision cell, the peptide is further fragmented by high-energy impact with a "collision gas," a small amount of a noble gas such as helium or argon that is bled into the vacuum chamber. This procedure is designed to fragment many of the peptide molecules in the sample, with each individual peptide broken in only one place on average. Most breaks occur at peptide bonds. This fragmentation does not involve the addition of water (it is done in a near-vacuum), so the products may include molecular ion radicals such as carbonyl radicals (Fig. 2a, bottom). The charge on the original peptide is retained on one of the fragments generated from it.

The second mass spectrometer then measures the *m/z* ratios of all the charged fragments (uncharged fragments are not detected). This generates one or more sets of peaks. A given set of peaks (Fig. 2b) consists of all the charged fragments that were generated by breaking the same type of bond (but at different points in the peptide), and derived from the same side of the bond breakage, either the carboxyl-terminal or amino-terminal side. Each successive peak in a given set has one less amino acid than the peak before. The difference in mass from peak to peak identifies the amino acid that was lost in each

figure 2

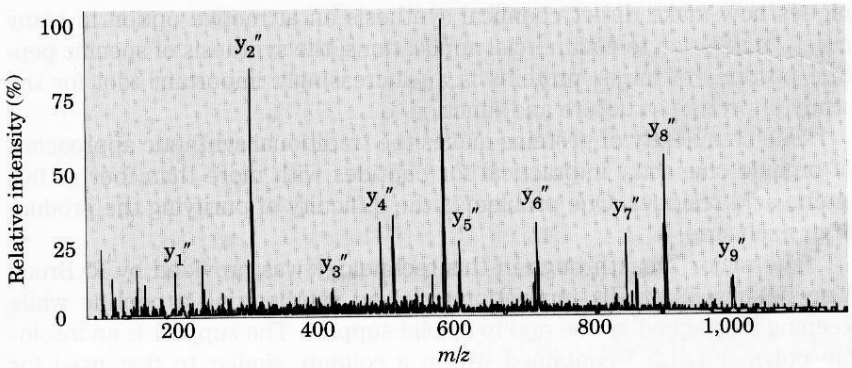
Obtaining protein sequence information with tandem mass spectrometry. (a) After proteolytic hydrolysis, a protein solution is injected into a mass spectrometer (MS-1). The different peptides are sorted so that only one type is selected for further analysis. The selected peptide is further fragmented in a chamber between the two mass spectrometers, and *m/z* for each fragment is measured in the second mass spectrometer (MS-2). Many of the ions generated during this second fragmentation result from breakage of the peptide bond, as shown. These are called b-type or y-type ions, depending on whether the charge is retained on the amino- or carboxyl-terminal side, respectively. **(b)** A typical spectrum with peaks representing the peptide fragments generated from a sample of one small peptide (10 residues). The labeled peaks are y-type ions. The large peak next to y_5'' is a doubly charged ion and is not part of the y set. The successive peaks differ by the mass of a particular amino acid in the original peptide. In this case, the deduced sequence was Phe–Pro–Gly–Gln–(Ile/Leu)–Asn–Ala–Asp–(Ile/Leu)–Arg. Note the ambiguity about Ile and Leu residues because they have the same molecular mass. In this example, the set of peaks derived from y-type ions predominates, and the spectrum is greatly simplified as a result. This is because an Arg residue occurs at the carboxyl terminus of the peptide, and most of the positive charges are retained on this residue.



case, thus revealing the sequence of the peptide. The only ambiguities involve leucine and isoleucine, which have the same mass.

The charge on the peptide can be retained on either the carboxyl- or amino-terminal fragment, and bonds other than the peptide bond can be broken in the fragmentation process, with the result that multiple sets of peaks are usually generated. The two most prominent sets generally consist of charged fragments derived from breakage of the peptide bonds. The set consisting of the carboxyl-terminal fragments can be unambiguously distinguished from that consisting of amino-terminal fragments. Because the bond breaks generated between the spectrometers (in the collision cell) do not yield full carboxyl and amino groups at the sites of the breaks, the only intact α-amino and α-carboxyl groups on the peptide fragments are those at the very ends (Fig. 2a). The two sets of fragments can thereby be assigned by the resulting slight differences in mass. The amino acid sequence derived from one set can be confirmed by the other, improving the confidence in the sequence information obtained.

Even a short sequence is often enough to permit unambiguous association of a protein with its gene, if the gene sequence is known. Sequencing by mass spectrometry cannot replace the Edman degradation procedure for the sequencing of long polypeptides, but it is ideal for proteome research aimed at cataloging the hundreds of cellular proteins that might be separated on a two-dimensional gel. In the coming decades, detailed genomic sequence data will be available from hundreds, eventually thousands, of organisms. The ability to rapidly associate proteins with genes using mass spectrometry will greatly facilitate the exploitation of this extraordinary information resource.



(b)

Amino Acid Sequences Provide Important Biochemical Information

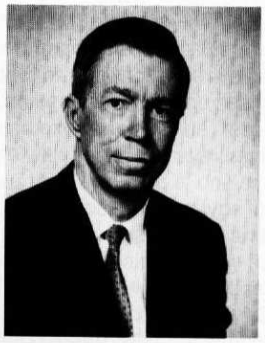
Knowledge of the sequence of amino acids in a protein can offer insights into its three-dimensional structure and its function, cellular location, and evolution. Most of these insights are derived by searching for similarities with other known sequences. Thousands of sequences are known and available in databases accessible through the Internet. The comparison of a newly obtained sequence with this large bank of stored sequences often reveals relationships both surprising and enlightening.

Exactly how the amino acid sequence determines three-dimensional structure is not understood in detail, nor can we always predict function from sequence. However, protein families that have some shared structural or functional features can be readily identified on the basis of amino acid sequence similarities. Individual proteins are assigned to families based on the degree of similarity in amino acid sequence. Members of a family are usually identical across 25% or more of their sequences, and proteins in these families generally share at least some structural and functional characteristics. Some families are defined, however, by identities involving only a few amino acid residues that are critical to a certain function. A number of similar substructures (to be defined in Chapter 6 as "domains") occur in many functionally unrelated proteins. These domains often fold up into structural configurations that have an unusual degree of stability or that are specialized for a certain environment. Evolutionary relationships can also be inferred from the structural and functional similarities within protein families.

Certain amino acid sequences often serve as signals that determine the cellular location, chemical modification, and half-life of a protein. Special signal sequences, usually at the amino terminus, are used to target certain proteins for export from the cell, while other proteins are targeted for distribution to the nucleus, the cell surface, the cytosol, and other cellular locations. Other sequences act as attachment sites for prosthetic groups, such as sugar groups in glycoproteins and lipids in lipoproteins. Some of these signals are well characterized and are easily recognized if they occur in the sequence of a newly characterized protein.

Small Peptides and Proteins Can Be Chemically Synthesized

Many peptides are potentially useful as pharmacologic agents, and their production is of considerable commercial importance. There are three ways to obtain a peptide: (1) purification from tissue, a task often made difficult by the vanishingly low concentrations of some peptides; (2) genetic engineering (Chapter 29); or (3) direct chemical synthesis. Powerful techniques now make direct chemical synthesis an attractive option in many cases. In addition to commercial applications, the synthesis of specific peptide portions of larger proteins is an increasingly important tool for the study of protein structure and function.



R. Bruce Merrifield

The complexity of proteins makes the traditional synthetic approaches of organic chemistry impractical for peptides with more than four or five amino acid residues. One problem is the difficulty of purifying the product after each step.

The major breakthrough in this technology was provided by R. Bruce Merrifield in 1962. His innovation involved synthesizing a peptide while keeping it attached at one end to a solid support. The support is an insoluble polymer (resin) contained within a column, similar to that used for chromatographic procedures. The peptide is built up on this support one amino acid at a time using a standard set of reactions in a repeating cycle (Fig. 5–29). At each successive step in the cycle, protective chemical groups block unwanted reactions.

figure 5–29

Chemical synthesis of a peptide on an insoluble polymer support. Reactions ① through ④ are necessary for the formation of each peptide bond. The 9-fluorenylmethoxycarbonyl (Fmoc) group (shaded blue) prevents unwanted reactions at the α-amino group of the residue (shaded red). Chemical synthesis proceeds from the carboxyl terminus to the amino terminus, the reverse of the direction of protein synthesis in vivo (Chapter 27).

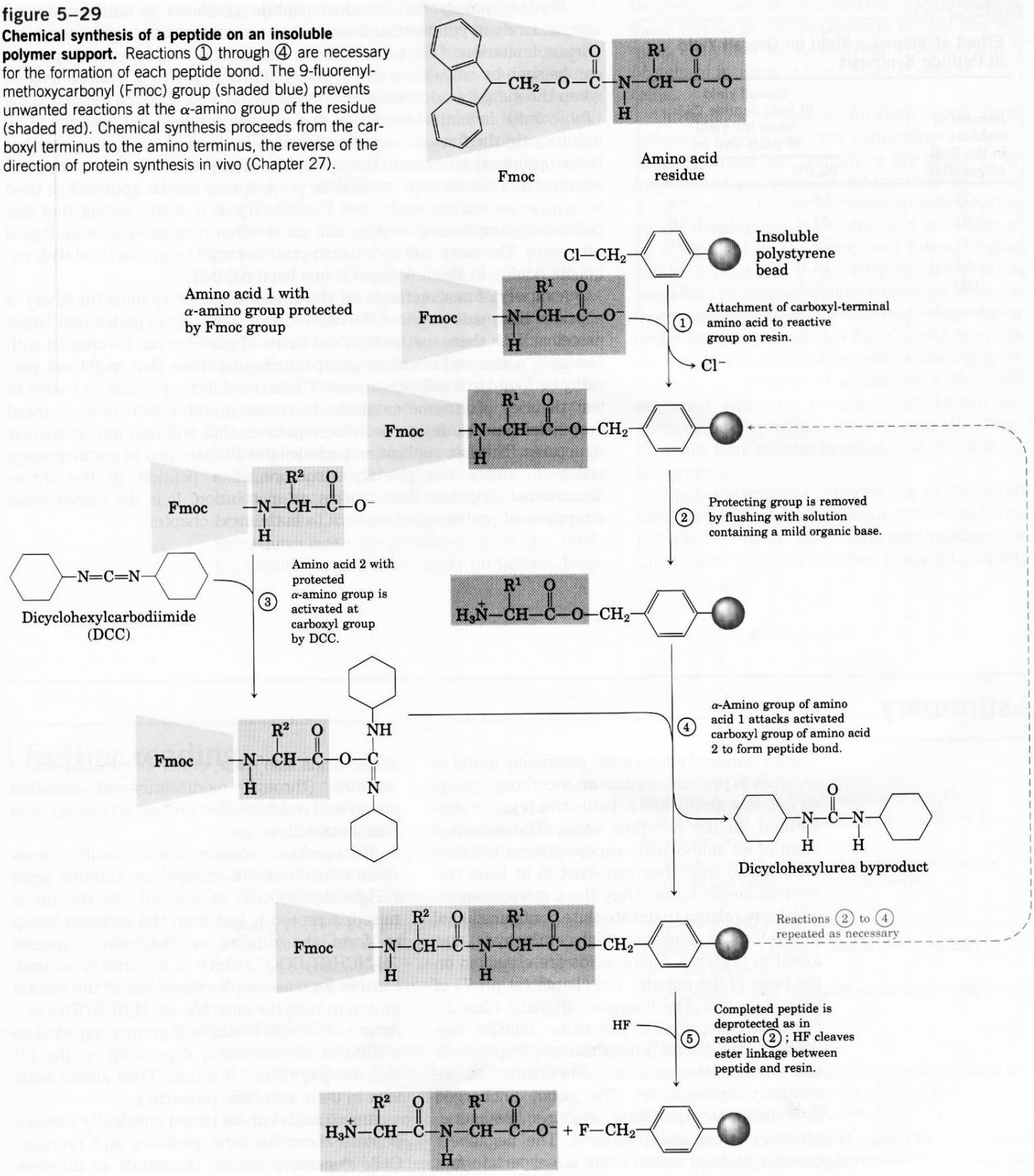


table 5–8

Effect of Stepwise Yield on Overall Yield in Peptide Synthesis

Number of residues in the final polypeptide	Overall yield of final peptide (%) when the yield of each step is:	
	96.0%	99.8%
11	66	98
21	44	96
31	29	94
51	13	90
100	1.7	82

The technology for chemical peptide synthesis is automated, and several kinds of commercial instruments are now available. The most important limitation of the process is the efficiency of each chemical cycle, as can be seen by calculating the overall yields of peptides of various lengths when the yield for addition of each new amino acid is 96.0% versus 99.8% (Table 5–8). Incomplete reaction at one stage can lead to formation of an impurity (in the form of a shorter peptide) in the next. The chemistry has been optimized to permit the synthesis of proteins of 100 amino acid residues in a few days in reasonable yield. A very similar approach is used to synthesize nucleic acids (see Fig. 10–37). It is worth noting that this technology, impressive as it is, still pales when compared with biological processes. The same 100 amino acid protein would be synthesized with exquisite fidelity in about 5 seconds in a bacterial cell.

A variety of new methods for the efficient ligation (joining together) of peptides has made possible the assembly of synthetic peptides into larger proteins. With these methods, novel forms of proteins can be created with precisely positioned chemical groups, including those that might not normally be found in a cellular protein. These novel forms provide new ways to test theories of enzyme catalysis, to create proteins with new chemical properties, and to design protein sequences that will fold into particular structures. This last application provides the ultimate test of our increasing ability to relate the primary structure of a peptide to the three-dimensional structure that it takes up in solution. It is the higher-order structures of proteins that we turn to in the next chapter.

summary

The 20 standard amino acids commonly found as residues in proteins contain an α-carboxyl group, an α-amino group, and a distinctive R group substituted on the α-carbon atom. The α-carbon atom of all amino acids except glycine is asymmetric, and thus they can exist in at least two stereoisomeric forms. Only the L stereoisomers, which are related to the absolute configuration of the reference molecule L-glyceraldehyde, are found in proteins. Amino acids are classified on the basis of the polarity and charge (at pH 7) of their R groups. The nonpolar, aliphatic class includes alanine, glycine, isoleucine, leucine, methionine, and valine. Phenylalanine, tryptophan, and tyrosine have aromatic side chains and are relatively hydrophobic. The polar, uncharged class includes asparagine, cysteine, glutamine, proline, serine, and threonine. The negatively charged (acidic) amino acids are aspartate and glutamate; the positively charged (basic) ones are arginine, histidine, and lysine. Nonstandard amino acids also exist, either as constituents of proteins (through modification of standard amino acid residues after protein synthesis) or as free metabolites.

Monoamino monocarboxylic amino acids (with nonionizable R groups) are diprotic acids ($^+H_3NCH(R)COOH$) at low pH. As the pH is raised, a proton is lost from the carboxyl group to form the dipolar or zwitterionic species $^+H_3NCH(R)COO^-$, which is electrically neutral. Further increase in pH causes loss of the second proton to yield the ionic species $H_2NCH(R)COO^-$. Amino acids with ionizable R groups may exist as additional ionic species, depending on the pH and the pK_a of the R group. Thus amino acids vary in their acid-base properties.

Amino acids can be joined covalently through peptide bonds to form peptides and proteins. Cells generally contain thousands of different proteins, each with a different function or biological activity. Proteins can be very long

polypeptide chains of 100 to several thousand amino acid residues. However, some naturally occurring peptides have only a few amino acid residues. Some proteins are composed of several noncovalently associated polypeptide chains, which are referred to as subunits. Simple proteins yield only amino acids on hydrolysis; conjugated proteins contain in addition some other component, such as a metal ion or organic prosthetic group.

There are four generally recognized levels of protein structure. Primary structure refers to the amino acid sequence and the location of disulfide bonds. Secondary structure is the spatial relationship of adjacent amino acids in localized stretches. Tertiary structure is the three-dimensional conformation of an entire polypeptide chain. Quaternary structure involves the spatial relationship of multiple polypeptide chains (subunits) that are stably associated.

Proteins are purified by taking advantage of various properties in which they differ. Proteins can be selectively precipitated by the addition of certain salts. A wide range of chromatographic procedures make use of differences in size, binding affinities, charge, and other properties. Electrophoresis can separate proteins on the basis of mass or charge. All purification procedures require a method for quantifying or assaying the protein of interest in the presence of other proteins.

Differences in protein function result from differences in amino acid composition and sequence. Amino acid sequences are deduced by fragmenting polypeptides into smaller peptides using reagents known to cleave specific peptide bonds, determining the amino acid sequence of each fragment by the automated Edman degradation procedure, then ordering the peptide fragments by finding sequence overlaps between fragments generated by different reagents. A protein sequence can also be deduced from the nucleotide sequence of its corresponding gene in DNA. Comparison of a protein's amino acid sequence with the thousands of known sequences often provides insights into the structure, function, cellular location, and evolution of the protein.

Short proteins and peptides (up to about 100 residues long) can be chemically synthesized. The peptide is built up, one amino acid residue at a time, while remaining tethered to a solid support.

further reading

General

Creighton, T.E. (1993) *Proteins: Structures and Molecular Properties,* 2nd edn, W.H. Freeman and Company, New York.

Very useful general source.

Sanger, F. (1988) Sequences, sequences, sequences. *Annu. Rev. Biochem.* **57,** 1–28.

A nice historical account of the development of sequencing methods.

Amino Acids

Greenstein, J.P. & Winitz, M. (1961) *Chemistry of the Amino Acids,* 3 Vols, John Wiley & Sons, New York.

Kreil, G. (1997) D-Amino acids in animal peptides. *Annu. Rev. Biochem.* **66,** 337–345.

An update on the occurrence of these unusual stereoisomers of amino acids.

Meister, A. (1965) *Biochemistry of the Amino Acids,* 2nd edn, Vols 1 and 2, Academic Press, Inc., New York.

Encyclopedic treatment of the properties, occurrence, and metabolism of amino acids.

Peptides and Proteins

Doolittle, R.F. (1985) Proteins. *Sci. Am.* **253** (October), 88–99.

An overview that highlights evolutionary relationships.

Working with Proteins

Dunn, M.J. (1997) Quantitative two-dimensional gel electrophoresis: from proteins to proteomes. *Biochem. Soc. Trans.* **25,** 248–254.

Dunn, M.J. & Corbett, J.M. (1996) Two-dimensional polyacrylamide gel electrophoresis. *Methods Enzymol.* **271,** 177–203.

A detailed description of the technology.

Kornberg, A. (1990) Why purify enzymes? *Methods Enzymol.* **182,** 1–5.

The critical role of classical biochemical methods in a new age.

Scopes, R.K. (1994) *Protein Purification: Principles and Practice,* 3rd edn, Springer-Verlag, New York.

A good source for more complete descriptions of the principles underlying chromatography and other methods.

Covalent Structure of Proteins

Andersen, J.S., Svensson, B., & Roepstorff, P. (1996) Electrospray ionization and matrix assisted laser desorption/ionization mass spectrometry: powerful analytical tools in recombinant protein chemistry. *Nat. Biotechnol.* **14,** 449–457.

A summary emphasizing applications.

Bork, P. & Koonin, E.V. (1998) Predicting functions from protein sequences—where are the bottlenecks? *Nat. Genet.* **18,** 313–318.

A good description of the technology and the roadblocks still limiting its use.

Dongre, A.R., Eng, J.K., & Yates, J.R. III. (1997) Emerging tandem-mass-spectrometry techniques for the rapid identification of proteins. *Trends Biotechnol.* **15,** 418–425.

A detailed description of methods.

Gibney, B.R., Rabanal, F., & Dutton, P.L. (1997) Synthesis of novel proteins. *Curr. Opin. Chem. Biol.* **1,** 537–542.

Koonin, E.V., Tatusov, R.L., & Galperin, M.Y. (1998) Beyond complete genomes: from sequence to structure and function. *Curr. Opin. Struct. Biol.* **8,** 355–363.

A good discussion of what we will do with the tremendous amount of protein sequence information becoming available.

Mann, M. & Wilm, M. (1995) Electrospray mass spectrometry for protein characterization. *Trends Biochem. Sci.* **20,** 219–224.

An approachable summary for beginners.

Wallace, C.J. (1995) Peptide ligation and semisynthesis. *Curr. Opin. Biotechnol.* **6,** 403–410.

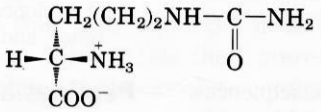
Good summary of methods available for peptide ligation. Includes some case studies

Wilken, J. & Kent, S.B. (1998) Chemical protein synthesis. *Curr. Opin. Biotechnol.* **9,** 412–426.

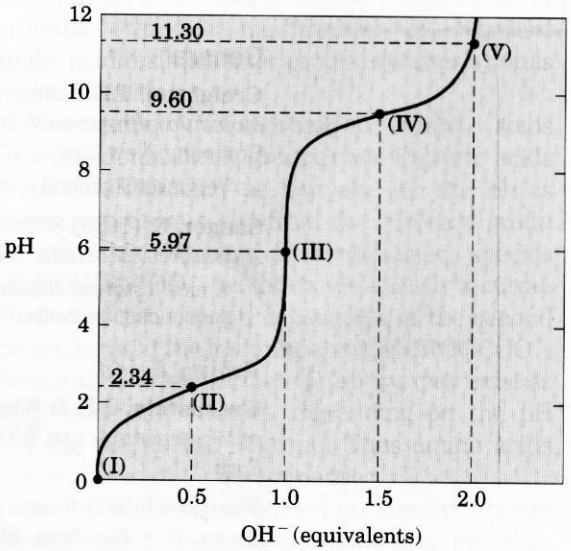
A good overview of chemical synthesis, focusing on peptide ligation methods and applications.

problems

1. Absolute Configuration of Citrulline The citrulline isolated from watermelons has the structure shown below. Is it a D- or L-amino acid? Explain.



2. Relationship between the Titration Curve and the Acid-Base Properties of Glycine A 100 mL solution of 0.1 M glycine at pH 1.72 was titrated with 2 M NaOH solution. The pH was monitored and the results were plotted on a graph, as shown at right. The key points in the titration are designated I to V. For each of the statements (a) to (o), *identify* the appropriate key point in the titration and *justify* your choice.



(a) Glycine is present predominantly as the species $^{+}H_3N{-}CH_2{-}COOH$.

(b) The *average* net charge of glycine is $+\frac{1}{2}$.

(c) Half of the amino groups are ionized.

(d) The pH is equal to the pK_a of the carboxyl group.

(e) The pH is equal to the pK_a of the protonated amino group.

(f) Glycine has its maximum buffering capacity.

(g) The *average* net charge of glycine is zero.

(h) The carboxyl group has been completely titrated (first equivalence point).

(i) Glycine is completely titrated (second equivalence point).

(j) The predominant species is $^{+}H_3N{-}CH_2{-}COO^{-}$.

(k) The *average* net charge of glycine is -1.

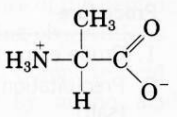
(l) Glycine is present predominantly as a 50:50 mixture of $^{+}H_3N{-}CH_2{-}COOH$ and $^{+}H_3N{-}CH_2{-}COO^{-}$.

(m) This is the isoelectric point.

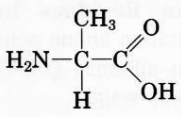
(n) This is the end of the titration.

(o) These are the *worst* pH regions for buffering power.

3. How Much Alanine Is Present as the Completely Uncharged Species? At a pH equal to the isoelectric point of alanine, the *net* charge on alanine is zero. Two structures can be drawn that have a net charge of zero, but the predominant form of alanine at its pI is zwitterionic.



Zwitterionic



Uncharged

(a) Why is alanine predominantly zwitterionic rather than completely uncharged at its pI?

(b) What fraction of alanine is in the completely uncharged form at its pI? Justify your assumptions.

4. Ionization State of Amino Acids Each ionizable group of an amino acid can exist in one of two states, charged or neutral. The electric charge on the functional group is determined by the relationship between its pK_a and the pH of the solution. This relationship is described by the Henderson-Hasselbalch equation.

(a) Histidine has three ionizable functional groups. Write the equilibrium equations for its three ionizations and assign the proper pK_a for each ionization. Draw the structure of histidine in each ionization state. What is the net charge on the histidine molecule in each ionization state?

(b) Draw the structures of the predominant ionization state of histidine at pH 1, 4, 8, and 12. Note that the ionization state can be approximated by treating each ionizable group independently.

(c) What is the net charge of histidine at pH 1, 4, 8, and 12? For each pH, will histidine migrate toward the anode (+) or cathode (−) when placed in an electric field?

5. Separation of Amino Acids by Ion-Exchange Chromatography Mixtures of amino acids are analyzed by first separating the mixture into its components through ion-exchange chromatography. Amino acids placed on a cation-exchange resin containing sulfonate groups (see Fig. 5–18) flow down the column at different rates because of two factors that influence their movement: (1) ionic attraction between the $-SO_3^-$ residues on the column and positively charged functional groups on the amino acids, and (2) hydrophobic interactions between amino acid side chains and the strongly hydrophobic backbone of the polystyrene resin. For each pair of amino acids listed, determine which will be eluted first from an ion-exchange column using a pH 7.0 buffer.

(a) Asp and Lys

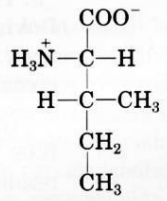
(b) Arg and Met

(c) Glu and Val

(d) Gly and Leu

(e) Ser and Ala

6. Naming the Stereoisomers of Isoleucine The structure of the amino acid isoleucine is



(a) How many chiral centers does it have?

(b) How many optical isomers?

(c) Draw perspective formulas for all the optical isomers of isoleucine.

7. Comparing the pK_a Values of Alanine and Polyalanine The titration curve of alanine shows the ionization of two functional groups with pK_a values of 2.34 and 9.69, corresponding to the ionization of the carboxyl and the protonated amino groups, respectively. The titration of di-, tri-, and larger oligopeptides of alanine also shows the ionization of only two functional groups, although the experimental pK_a values are different. The trend in pK_a values is summarized in the table.

Amino acid or peptide	pK_1	pK_2
Ala	2.34	9.69
Ala–Ala	3.12	8.30
Ala–Ala–Ala	3.39	8.03
Ala–(Ala)$_n$–Ala, $n \geq 4$	3.42	7.94

(a) Draw the structure of Ala–Ala–Ala. Identify the functional groups associated with pK_1 and pK_2.

(b) Why does the value of pK_1 *increase* with each addition of an Ala residue to the Ala oligopeptide?

(c) Why does the value of pK_2 *decrease* with each addition of an Ala residue to the Ala oligopeptide?

8. The Size of Proteins What is the approximate molecular weight of a protein with 682 amino acid residues in a single polypeptide chain?

9. The Number of Tryptophan Residues in Bovine Serum Albumin A quantitative amino acid analysis reveals that bovine serum albumin (BSA) contains 0.58% tryptophan (M_r 204) by weight.

(a) Calculate the *minimum* molecular weight of BSA (i.e., assuming there is only one tryptophan residue per protein molecule).

(b) Gel filtration of BSA gives a molecular weight estimate of 70,000. How many tryptophan residues are present in a molecule of serum albumin?

10. Net Electric Charge of Peptides A peptide has the sequence

Glu – His – Trp – Ser – Gly – Leu – Arg – Pro – Gly

(a) What is the net charge of the molecule at pH 3, 8, and 11? (Use pK_a values for side chains and terminal amino and carboxyl groups as given in Table 5–1.)

(b) Estimate the pI for this peptide.

11. Isoelectric Point of Pepsin Pepsin is the name given to several digestive enzymes that are secreted (as larger precursor proteins) by glands that line the stomach. These glands also secrete hydrochloric acid, which dissolves the particulate matter in food, allowing pepsin to enzymatically cleave individual protein molecules. The resulting mixture of food, HCl, and digestive enzymes is known as chyme and has a pH near 1.5. What pI would you predict for the pepsin proteins? What functional groups must be present to confer this pI on pepsin? Which amino acids in the proteins would contribute such groups?

12. The Isoelectric Point of Histones Histones are proteins found in eukaryotic cell nuclei, tightly bound to DNA, which has many phosphate groups. The pI of histones is very high, about 10.8. What amino acid residues must be present in relatively large numbers in histones? In what way do these residues contribute to the strong binding of histones to DNA?

13. Solubility of Polypeptides One method for separating polypeptides makes use of their differential solubilities. The solubility of large polypeptides in water depends upon the relative polarity of their R groups, particularly on the number of ionized groups: the more ionized groups there are, the more soluble the polypeptide. Which of each pair of polypeptides below is more soluble at the indicated pH?

(a) $(Gly)_{20}$ or $(Glu)_{20}$ at pH 7.0

(b) $(Lys\text{-}Ala)_3$ or $(Phe\text{-}Met)_3$ at pH 7.0

(c) $(Ala\text{-}Ser\text{-}Gly)_5$ or $(Asn\text{-}Ser\text{-}His)_5$ at pH 6.0

(d) $(Ala\text{-}Asp\text{-}Gly)_5$ or $(Asn\text{-}Ser\text{-}His)_5$ at pH 3.0

14. Purification of an Enzyme A biochemist discovers and purifies a new enzyme, generating the purification table below.

Procedure	Total protein (mg)	Activity (units)
1. Crude extract	20,000	4,000,000
2. Precipitation (salt)	5,000	3,000,000
3. Precipitation (pH)	4,000	1,000,000
4. Ion-exchange chromatography	200	800,000
5. Affinity chromatography	50	750,000
6. Size-exclusion chromatography	45	675,000

(a) From the information given in the table, calculate the specific activity of the enzyme solution after each purification procedure.

(b) Which of the purification procedures used for this enzyme is most effective (i.e., gives the greatest relative increase in purity)?

(c) Which of the purification procedures is least effective?

(d) Is there any indication based on the results shown in the table that the enzyme after step 6 is now pure? What else could be done to estimate the purity of the enzyme preparation?

15. Sequence Determination of the Brain Peptide Leucine Enkephalin A group of peptides that influence nerve transmission in certain parts of the brain has been isolated from normal brain tissue. These peptides are known as opioids, because they bind to specific receptors that also bind opiate drugs, such as morphine and naloxone. Opioids thus mimic some of the properties of opiates. Some researchers consider these peptides to be the brain's own pain killers. Using the information below, determine the amino acid sequence of the opioid leucine enkephalin. Explain how your structure is consistent with each piece of information.

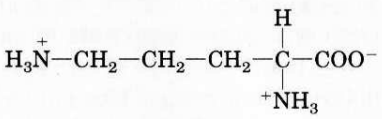
(a) Complete hydrolysis by 6 M HCl at 110 °C followed by amino acid analysis indicated the presence of Gly, Leu, Phe, and Tyr, in a 2:1:1:1 molar ratio.

(b) Treatment of the peptide with 1-fluoro-2,4-dinitrobenzene followed by complete hydrolysis and chromatography indicated the presence of the 2,4-dinitrophenyl derivative of tyrosine. No free tyrosine could be found.

(c) Complete digestion of the peptide with pepsin followed by chromatography yielded a dipeptide containing Phe and Leu, plus a tripeptide containing Tyr and Gly in a 1:2 ratio.

16. Structure of a Peptide Antibiotic from *Bacillus brevis* Extracts from the bacterium *Bacillus brevis* contain a peptide with antibiotic properties. This peptide forms complexes with metal ions and apparently disrupts ion transport across the cell membranes of other bacterial species, killing them. The structure of the peptide has been determined from the following observations.

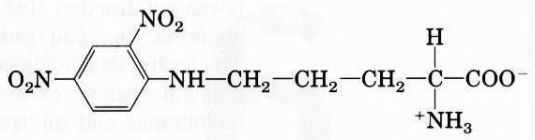
(a) Complete acid hydrolysis of the peptide followed by amino acid analysis yielded equimolar amounts of Leu, Orn, Phe, Pro, and Val. Orn is ornithine, an amino acid not present in proteins but present in some peptides. It has the structure



(b) The molecular weight of the peptide was estimated as about 1,200.

(c) The peptide failed to undergo hydrolysis when treated with the enzyme carboxypeptidase. This enzyme catalyzes the hydrolysis of the carboxyl-terminal residue of a polypeptide unless the residue is Pro or does not contain a free carboxyl group for some reason.

(d) Treatment of the intact peptide with 1-fluoro-2,4-dinitrobenzene, followed by complete hydrolysis and chromatography, yielded only free amino acids and the following derivative:



(Hint: Note that the 2,4-dinitrophenyl derivative involves the amino group of a side chain rather than the α-amino group.)

(e) Partial hydrolysis of the peptide followed by chromatographic separation and sequence analysis yielded the following di- and tripeptides (the amino-terminal amino acid is always at the left):

Leu–Phe Phe–Pro Orn–Leu Val–Orn
Val–Orn–Leu Phe–Pro–Val Pro–Val–Orn

Given the above information, deduce the amino acid sequence of the peptide antibiotic. Show your reasoning. When you have arrived at a structure, demonstrate that it is consistent with *each* experimental observation.

17. Efficiency in Peptide Sequencing A peptide with the primary structure Lys–Arg–Pro–Leu–Ile–Asp–Gly–Ala is sequenced by the Edman procedure. If each Edman cycle were 96% efficient, what percentage of the amino acids liberated in the fourth cycle would be leucine? Do the calculation a second time, but assume a 99% efficiency for each cycle.

18. Biochemistry Protocols: Your First Protein Purification As the newest and least experienced student in a biochemistry research lab, your first few weeks are spent washing glassware and labeling test tubes. You then graduate to making buffers and stock solutions for use in various laboratory procedures. Finally, you are given responsibility for purifying a protein. It is a citric acid cycle enzyme, citrate synthase, located in the mitochondrial matrix. Following a protocol for the purification, you proceed through the steps below. As you work, a more experienced student questions you about the rationale for each procedure. Supply the answers. (Hint: See Chapter 2 for information on separation of organelles from cells, and Chapter 4 for information about osmolarity).

(a) You pick up 20 kg of beef hearts from a nearby slaughterhouse. You transport the hearts on ice, and perform each step of the purification in a walk-in cold room or on ice. You homogenize the beef heart tissue in a high-speed blender in a medium containing ~0.2 M sucrose, buffered to a pH of 7.2. *Why do you use beef heart tissue, and in such large quantity? What is the purpose of keeping the tissue cold and suspending it in 0.2 M sucrose, at pH 7.2? What happens to the tissue when it is homogenized?*

chapter

7

Protein Function

Knowing the three-dimensional structure of a protein is an important part of understanding how the protein functions. However, the structure shown in two dimensions on a page is deceptively static. Proteins are dynamic molecules whose functions almost invariably depend on interactions with other molecules, and these interactions are affected in physiologically important ways by sometimes subtle, sometimes striking changes in protein conformation.

In this chapter, we explore how proteins interact with other molecules and how their interactions are related to dynamic protein structure. The importance of molecular interactions to a protein's function can hardly be overemphasized. In Chapter 6, we saw that the function of fibrous proteins as structural elements of cells and tissues depends on stable, long-term quaternary interactions between identical polypeptide chains. As we will see in this chapter, the functions of many other proteins involve interactions with a variety of different molecules. Most of these interactions are fleeting, though they may be the basis of complex physiological processes such as oxygen transport, immune function, and muscle contraction, the topics we examine in detail in this chapter. The proteins that carry out these processes illustrate the following key principles of protein function, some of which will be familiar from the previous chapter:

The functions of many proteins involve the reversible binding of other molecules. A molecule bound reversibly by a protein is called a **ligand.** A ligand may be any kind of molecule, including another protein. The transient nature of protein-ligand interactions is critical to life, allowing an organism to respond rapidly and reversibly to changing environmental and metabolic circumstances.

A ligand binds at a site on the protein called the **binding site,** which is complementary to the ligand in size, shape, charge, and hydrophobic or hydrophilic character. Furthermore, the interaction is specific: the protein can discriminate among the thousands of different molecules in its environment and selectively bind only one or a few. A given protein may have separate binding sites for several different ligands. These specific molecular interactions are crucial in maintaining the high degree of order in a living system. (This discussion excludes the binding of water, which may interact weakly and nonspecifically with many parts of a protein. In Chapter 8, we consider water as a specific ligand for many enzymes.)

Proteins are flexible. Changes in conformation may be subtle, reflecting molecular vibrations and small movements of amino acid residues

throughout the protein. A protein flexing in this way is sometimes said to "breathe." Changes in conformation may also be quite dramatic, with major segments of the protein structure moving as much as several nanometers. Specific conformational changes are frequently essential to a protein's function.

The binding of a protein and ligand is often coupled to a conformational change in the protein that makes the binding site more complementary to the ligand, permitting tighter binding. The structural adaptation that occurs between protein and ligand is called **induced fit.**

In a multisubunit protein, a conformational change in one subunit often affects the conformation of other subunits.

Interactions between ligands and proteins may be regulated, usually through specific interactions with one or more additional ligands. These other ligands may cause conformational changes in the protein that affect the binding of the first ligand.

Enzymes represent a special case of protein function. Enzymes bind and chemically transform other molecules—they catalyze reactions. The molecules acted upon by enzymes are called reaction **substrates** rather than ligands, and the ligand-binding site is called the **catalytic site** or **active site.** In this chapter we emphasize the noncatalytic functions of proteins. In Chapter 8 we consider catalysis by enzymes, a central topic in biochemistry. You will see that the themes of this chapter—binding, specificity, and conformational change—are continued in the next chapter, with the added element of proteins acting as reactants in chemical transformations.

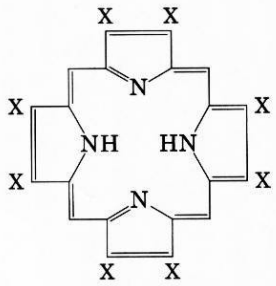
Reversible Binding of a Protein to a Ligand: Oxygen-Binding Proteins

Myoglobin and hemoglobin may be the most-studied and best-understood proteins. They were the first proteins for which three-dimensional structures were determined, and our current understanding of myoglobin and hemoglobin is garnered from the work of thousands of biochemists over several decades. Most important, they illustrate almost every aspect of that most central of biochemical processes: the reversible binding of a ligand to a protein. This classic model of protein function will tell us a great deal about how proteins work.

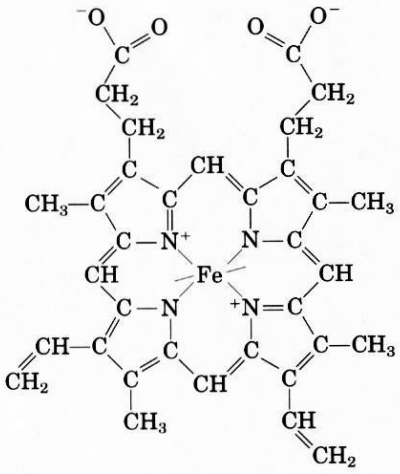
Oxygen Can Be Bound to a Heme Prosthetic Group

Oxygen is poorly soluble in aqueous solutions (see Table 4–3) and cannot be carried to tissues in sufficient quantity if it is simply dissolved in blood serum. Diffusion of oxygen through tissues is also ineffective over distances greater than a few millimeters. The evolution of larger, multicellular animals depended on the evolution of proteins that could transport and store oxygen. However, none of the amino acid side chains in proteins is suited for the reversible binding of oxygen molecules. This role is filled by certain transition metals, among them iron and copper, that have a strong tendency to bind oxygen. Multicellular organisms exploit the properties of metals, most commonly iron, for oxygen transport. However, free iron promotes the formation of highly reactive oxygen species such as hydroxyl radicals that can damage DNA and other macromolecules. Iron used in cells is therefore bound in forms that sequester it and/or make it less reactive. In multicellular organisms—especially those in which iron, in its oxygen-carrying capacity, must be transported over large distances—iron is often incorporated into a protein-bound prosthetic group called **heme.** (A prosthetic group is a compound permanently associated with a protein that contributes to the protein's function.)

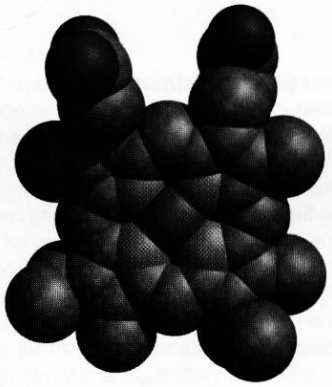
figure 7–1
Heme. The heme group is present in myoglobin, hemoglobin, and many other proteins, designated heme proteins. Heme consists of a complex organic ring structure, protoporphyrin IX, to which is bound an iron atom in its ferrous (Fe^{2+}) state. Porphyrins, of which protoporphyrin IX is only one example, consist of four pyrrole rings linked by methene bridges **(a),** with substitutions at one or more of the positions denoted X. Two representations of heme are shown in **(b)** and **(c).** The iron atom of heme has six coordination bonds: four in the plane of, and bonded to, the flat porphyrin ring system, and two perpendicular to it **(d).**

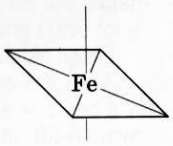


(a)



(b)


(c)



(d)

Heme (or haem) consists of a complex organic ring structure, **protoporphyrin,** to which is bound a single iron atom in its ferrous (Fe^{2+}) state (Fig. 7–1). The iron atom has six coordination bonds, four to nitrogen atoms that are part of the flat **porphyrin ring** system and two perpendicular to the porphyrin. The coordinated nitrogen atoms (which have an electron-donating character) help prevent conversion of the heme iron to the ferric (Fe^{3+}) state. Iron in the Fe^{2+} state binds oxygen reversibly; in the Fe^{3+} state it does not bind oxygen. Heme is found in a number of oxygen-transporting proteins, as well as in some proteins, such as the cytochromes, that participate in oxidation-reduction (electron transfer) reactions (Chapter 19).

In free heme molecules, reaction of oxygen at one of the two "open" coordination bonds of iron (perpendicular to the plane of the porphyrin molecule, above and below) can result in irreversible conversion of Fe^{2+} to Fe^{3+}. In heme-containing proteins, this reaction is prevented by sequestering the heme deep within a protein structure where access to the two open coordination bonds is restricted. One of these two coordination bonds is occupied by a side-chain nitrogen of a His residue. The other is the binding site for molecular oxygen (O_2) (Fig. 7–2). When oxygen binds, the electronic properties of heme iron change; this accounts for the change in color from the dark purple of oxygen-depleted venous blood to the bright red of oxygen-rich arterial blood. Some small molecules, such as carbon monoxide (CO) and nitric oxide (NO), coordinate to heme iron with greater affinity than does O_2. When a molecule of CO is bound to heme, O_2 is excluded, which is why CO is highly toxic to aerobic organisms. By surrounding and sequestering heme, oxygen-binding proteins regulate the access of CO and other small molecules to heme iron.

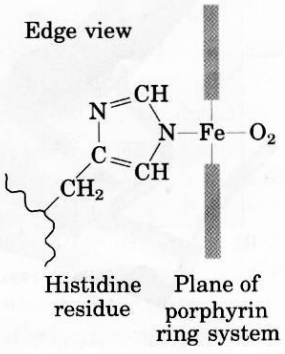


figure 7–2
The heme group viewed from the side. This view shows the two coordination bonds to Fe^{2+} perpendicular to the porphyrin ring system. One of these two bonds is occupied by a His residue, sometimes called the proximal His. The other is the binding site for oxygen. The remaining four coordination bonds are in the plane of, and bonded to, the flat porphyrin ring system.

Myoglobin Has a Single Binding Site for Oxygen

Myoglobin (M_r 16,700; abbreviated Mb) is a relatively simple oxygen-binding protein found in almost all mammals, primarily in muscle tissue. It is particularly abundant in the muscles of diving mammals such as seals and whales that must store enough oxygen for prolonged excursions undersea. Proteins very similar to myoglobin are widely distributed, occurring even in some single-celled organisms. Myoglobin stores oxygen for periods when energy demands are high and facilitates its distribution to oxygen-starved tissues.

Myoglobin is a single polypeptide of 153 amino acid residues with one molecule of heme. It is typical of the family of proteins called **globins,** which have similar primary and tertiary structures. The polypeptide is made up of eight α-helical segments connected by bends (Fig. 7–3). About 78% of the amino acid residues in the protein are found in these α helices.

Any detailed discussion of protein function inevitably involves protein structure. Our treatment of myoglobin will be facilitated by introducing some structural conventions peculiar to globins. As seen in Figure 7–3, the helical segments are labeled A through H. An individual amino acid residue may be designated either by its position in the amino acid sequence or by its location within the sequence of a particular α-helical segment. For example, the His residue coordinated to the heme in myoglobin, His^{93} (the 93rd amino acid residue from the amino-terminal end of the myoglobin polypeptide sequence), is also called His F8 (the 8th residue in α helix F). The bends in the structure are labeled AB, CD, EF, and so forth, reflecting the α-helical segments they connect.

Protein-Ligand Interactions Can Be Described Quantitatively

The function of myoglobin depends on the protein's ability not only to bind oxygen, but also to release it when and where it is needed. Function in biochemistry often revolves around a reversible protein-ligand interaction of this type. A quantitative description of this interaction is therefore a central part of many biochemical investigations.

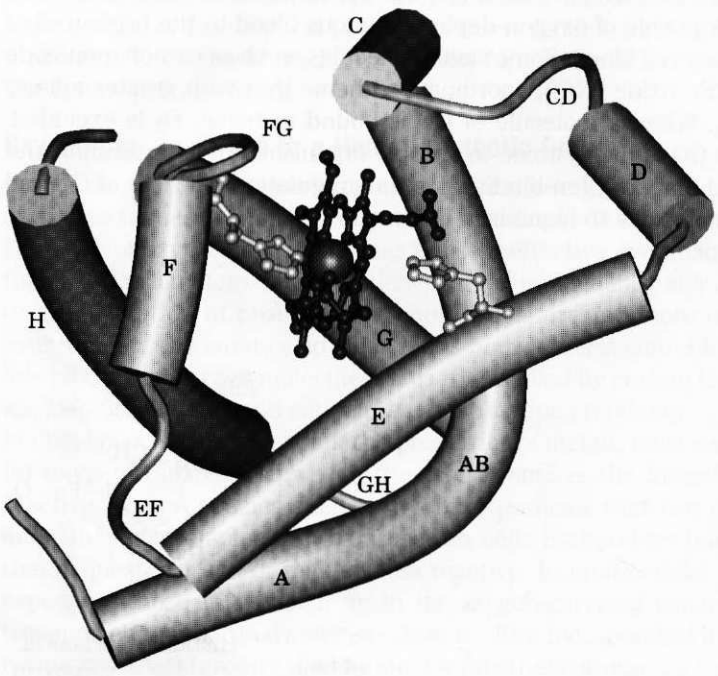


figure 7–3
The structure of myoglobin. The eight α-helical segments (shown here as cylinders) are labeled A through H. Nonhelical residues in the bends that connect them are labeled AB, CD, EF, and so forth, indicating the segments they interconnect. A few bends, including BC and DE, are abrupt and do not contain any residues; these are not normally labeled. (The short segment visible between D and E is an artifact of the computer representation.) The heme is bound in a pocket made up largely of the E and F helices, although amino acid residues from other segments of the protein also participate.

In general, the reversible binding of a protein (P) to a ligand (L) can be described by a simple **equilibrium expression:**

$$P + L \rightleftharpoons PL \tag{7–1}$$

The reaction is characterized by an equilibrium constant, K_a, such that

$$K_a = \frac{[PL]}{[P][L]} \tag{7–2}$$

The term $\boldsymbol{K_a}$ is an **association constant** (not to be confused with the K_a that denotes an acid dissociation constant; see p. 98). The association constant provides a measure of the affinity of the ligand L for the protein. K_a has units of M^{-1}; a higher value of K_a corresponds to a higher affinity of the ligand for the protein. A rearrangement of Equation 7–2 shows that the ratio of bound to free protein is directly proportional to the concentration of free ligand:

$$K_a[L] = \frac{[PL]}{[P]} \tag{7–3}$$

When the concentration of the ligand is much greater than the concentration of ligand-binding sites, the binding of the ligand by the protein does not appreciably change the concentration of free (unbound) ligand—that is, [L] remains constant. This condition is broadly applicable to most ligands that bind to proteins in cells and simplifies our description of the binding equilibrium.

Thus we can consider the binding equilibrium from the standpoint of the fraction, θ (theta), of ligand-binding sites on the protein that are occupied by ligand:

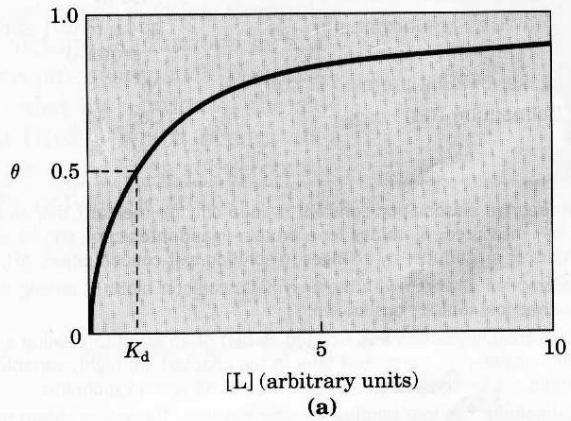
$$\theta = \frac{\text{binding sites occupied}}{\text{total binding sites}} = \frac{[PL]}{[PL] + [P]} \tag{7–4}$$

Substituting $K_a[L][P]$ for [PL] (see Eqn 7–3) and rearranging terms gives

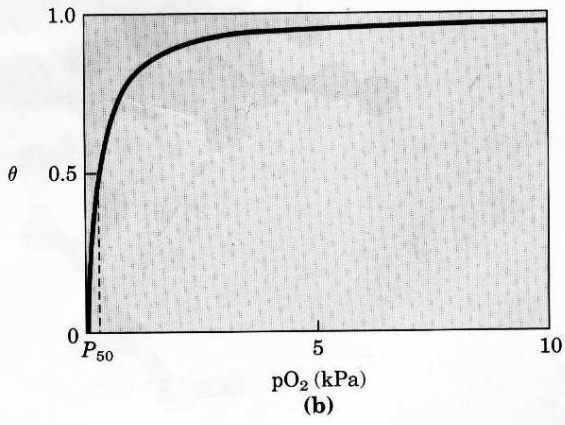
$$\theta = \frac{K_a[L][P]}{K_a[L][P] + [P]} = \frac{K_a[L]}{K_a[L] + 1} = \frac{[L]}{[L] + \frac{1}{K_a}} \tag{7–5}$$

The term K_a can be determined from a plot of θ versus the concentration of free ligand, [L] (Fig. 7–4a). Any equation of the form $x = y/(y + z)$ describes a hyperbola, and θ is thus found to be a hyperbolic function of [L]. The fraction of ligand-binding sites occupied approaches saturation asymptotically as [L] increases. The [L] at which half of the available ligand-binding sites are occupied (at $\theta = 0.5$) corresponds to $1/K_a$.

figure 7–4
Graphical representations of ligand binding. The fraction of ligand-binding sites occupied, θ, is plotted against the concentration of free ligand. Both curves are rectangular hyperbolas. **(a)** A hypothetical binding curve for a ligand L. The [L] at which half of the available ligand-binding sites are occupied is equivalent to $1/K_a$, or K_d. The curve has a horizontal asymptote at $\theta = 1$ and a vertical asymptote (not shown) at $[L] = -1/K_a$. **(b)** A curve describing the binding of oxygen to myoglobin. The partial pressure of O_2 in the air above the solution is expressed in terms of kilopascals (kPa). Oxygen binds tightly to myoglobin with a P_{50} of only 0.26 kPa.



(a)



(b)

It is sometimes intuitively simpler to consider the **dissociation constant, K_d,** which is the reciprocal of K_a ($K_d = 1/K_a$) and is given in units of molar concentration (M). K_d is the equilibrium constant for the release of ligand. The relevant expressions change to

$$K_d = \frac{[P][L]}{[PL]} \tag{7–6}$$

$$[PL] = \frac{[P][L]}{K_d} \tag{7–7}$$

$$\theta = \frac{[L]}{[L] + K_d} \tag{7–8}$$

When [L] is equal to K_d, half of the ligand-binding sites are occupied. When [L] is lower than K_d, little ligand binds to the protein. In order for 90% of the available ligand-binding sites to be occupied, [L] must be nine times greater than K_d. In practice, K_d is used much more often than K_a to express the affinity of a protein for a ligand. Note that a lower value of K_d corresponds to a higher affinity of ligand for the protein. The mathematics can be reduced to simple statements: K_d is the molar concentration of ligand at which half of the available ligand-binding sites are occupied. At this point, the protein is said to have reached half saturation with respect to ligand binding. The more tightly a protein binds a ligand, the lower the concentration of ligand required for half the binding sites to be occupied, and thus the lower the value of K_d. Some representative dissociation constants are given in Table 7–1.

The binding of oxygen to myoglobin follows the patterns discussed above, but because oxygen is a gas, we must make some minor adjustments to the equations. We can simply substitute the concentration of dissolved oxygen for [L] in Equation 7–8 to give

$$\theta = \frac{[O_2]}{[O_2] + K_d} \tag{7–9}$$

As for any ligand, K_d is equal to the $[O_2]$ at which half of the available ligand-binding sites are occupied, or $[O_2]_{0.5}$. Equation 7–9 becomes

$$\theta = \frac{[O_2]}{[O_2] + [O_2]_{0.5}} \tag{7–10}$$

table 7–1

Some Protein Dissociation Constants

Protein	Ligand	K_d (M)*
Avidin (egg white)†	Biotin	1×10^{-15}
Insulin receptor (human)	Insulin	1×10^{-10}
Anti-HIV immunoglobulin (human)‡	gp41 (HIV-1 surface protein)	4×10^{-10}
Nickel-binding protein (*E. coli*)	Ni^{2+}	1×10^{-7}
Calmodulin (rat)§	Ca^{2+}	3×10^{-6}
		2×10^{-5}

*A reported dissociation constant is valid only for the particular solution conditions under which it was measured. K_d values for a protein-ligand interaction can be altered, sometimes by several orders of magnitude, by changes in solution salt concentration, pH, or other variables.

†Interaction of avidin with the enzymatic cofactor biotin is among the strongest noncovalent biochemical interactions known.

‡This immunoglobulin was isolated as part of an effort to develop a vaccine against HIV. Immunoglobulins (described later in the chapter) are highly variable, and the K_d reported here should not be considered characteristic of all immunoglobulins.

§Calmodulin has four binding sites for calcium. The values shown reflect the highest- and lowest-affinity binding sites observed in one set of measurements.

The concentration of a volatile substance in solution, however, is always proportional to its partial pressure in the gas phase above the solution. In experiments using oxygen as a ligand, it is the partial pressure of oxygen, pO_2, that is varied because this is easier to measure than the concentration of dissolved oxygen. If we define the partial pressure of oxygen at $[O_2]_{0.5}$ as P_{50}, substitution in Equation 7–10 gives

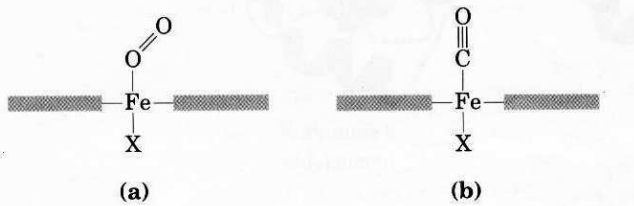
$$\theta = \frac{pO_2}{pO_2 + P_{50}} \qquad (7\text{–}11)$$

A binding curve for myoglobin that relates θ to pO_2 is shown in Figure 7–4b.

Protein Structure Affects How Ligands Bind

The binding of a ligand to a protein is rarely as simple as the above equations would suggest. The interaction is greatly affected by protein structure and is often accompanied by conformational changes. For example, the specificity with which heme binds its various ligands is altered when the heme is a component of myoglobin. CO binds to free heme molecules over 20,000 times better than does O_2 (the K_d or P_{50} for CO binding is more than 20,000 times lower than that for O_2) but binds only about 200 times better when the heme is bound in myoglobin. The difference is partly explained by steric hindrance. When O_2 binds to free heme, the axis of the oxygen molecule is positioned at an angle to the Fe—O bond (Fig. 7–5a). In contrast, when CO binds to free heme, the Fe, C, and O atoms lie in a straight line (Fig. 7–5b). In both cases, the binding reflects the geometry of hybrid orbitals in each ligand. In myoglobin, His^{64} (His E7), on the O_2-binding side of the heme, is too far away to coordinate with the heme iron, but it does interact with a ligand bound to heme. This residue, called the *distal His,* does not affect the binding of O_2 (Fig. 7–5c) but may preclude the linear binding of CO, providing one explanation for the diminished binding of CO to heme in myoglobin (and hemoglobin). This effect on CO binding is physiologically important, because CO is a low-level byproduct of cellular metabolism. Other factors, not yet well-defined, also seem to modulate the interaction of heme with CO in these proteins.

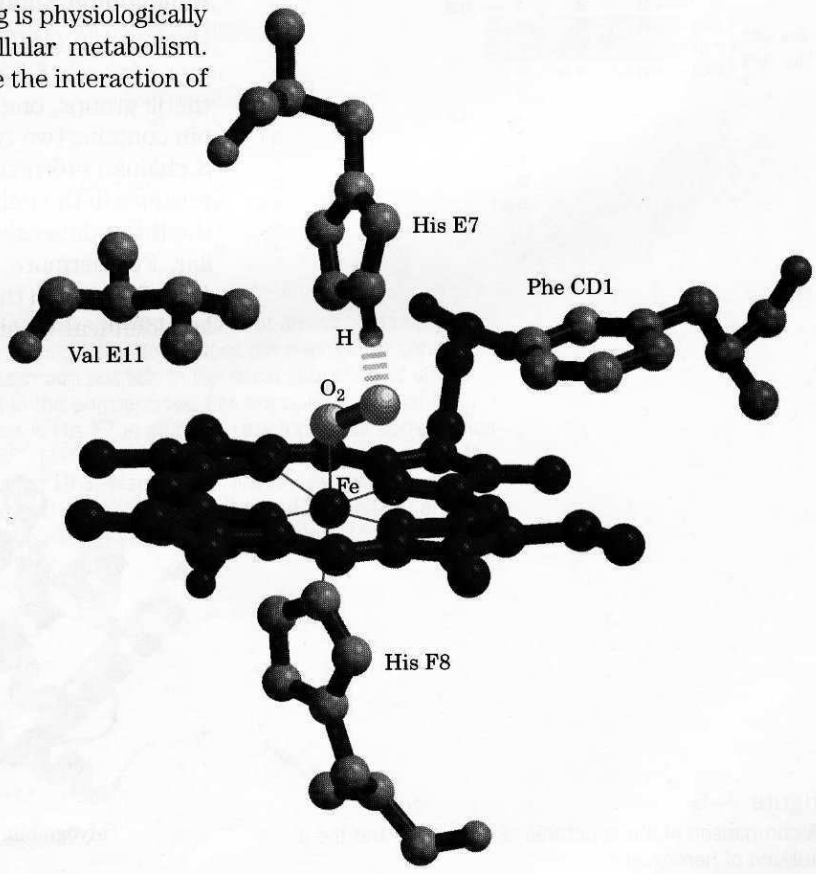
The binding of O_2 to the heme in myoglobin also depends on molecular motions, or "breathing," in the protein structure. The heme molecule is deeply buried in the folded polypeptide, with no direct path for oxygen to go from the surrounding solution to the ligand-binding site. If the protein were rigid, O_2 could not enter or leave the heme pocket at a measurable rate. However, rapid molecular flexing of the amino acid side chains produces transient cavities in the protein structure, and O_2 evidently makes its way in and out by moving through these cavities. Computer simulations of rapid structural fluctuations in myoglobin suggest that there are many such pathways. One major route is provided by rotation of the side chain of the distal His (His^{64}), which occurs on a nanosecond (10^{-9} s) time scale. Even subtle conformational changes can be critical for protein activity.



(a) (b)

figure 7–5

Steric effects on the binding of ligands to the heme of myoglobin. (a) Oxygen binds to heme with the O_2 axis at an angle, a binding conformation readily accommodated by myoglobin. **(b)** Carbon monoxide binds to free heme with the CO axis perpendicular to the plane of the porphyrin ring. CO binding to the heme in myoglobin is forced to adopt a slight angle because the perpendicular arrangement is sterically blocked by His E7, the distal His. This effect weakens the binding of CO to myoglobin. **(c)** Another view showing the arrangement of key amino acid residues around the heme of myoglobin. The bound O_2 is hydrogen-bonded to the distal His, His E7 (His^{64}), further facilitating the binding of O_2.



(c)

Oxygen Is Transported in Blood by Hemoglobin

Nearly all the oxygen carried by whole blood in animals is bound and transported by hemoglobin in erythrocytes (red blood cells). Normal human erythrocytes are small (6 to 9 μm in diameter), biconcave disks. They are formed from precursor stem cells called **hemocytoblasts.** In the maturation process, the stem cell produces daughter cells that form large amounts of hemoglobin and then lose their intracellular organelles—nucleus, mitochondria, and endoplasmic reticulum. Erythrocytes are thus incomplete, vestigial cells, unable to reproduce and, in humans, destined to survive for only about 120 days. Their main function is to carry hemoglobin, which is dissolved in the cytosol at a very high concentration (~34% by weight).

In arterial blood passing from the lungs through the heart to the peripheral tissues, hemoglobin is about 96% saturated with oxygen. In the venous blood returning to the heart, hemoglobin is only about 64% saturated. Thus, each 100 mL of blood passing through a tissue releases about one-third of the oxygen it carries, or 6.5 mL of O_2 gas at atmospheric pressure and body temperature.

Myoglobin, with its hyperbolic binding curve for oxygen (Fig. 7–4b), is relatively insensitive to small changes in the concentration of dissolved oxygen and so functions well as an oxygen-storage protein. Hemoglobin, with its multiple subunits and O_2-binding sites, is better suited to oxygen transport. As we will see, interactions between the subunits of a multimeric protein can permit a highly sensitive response to small changes in ligand concentration. Interactions among the subunits in hemoglobin cause conformational changes that alter the affinity of the protein for oxygen. The modulation of oxygen binding allows the O_2-transport protein to respond to changes in oxygen demand by tissues.

Hemoglobin Subunits Are Structurally Similar to Myoglobin

Hemoglobin (M_r 64,500; abbreviated Hb) is roughly spherical, with a diameter of nearly 5.5 nm. It is a tetrameric protein containing four heme prosthetic groups, one associated with each polypeptide chain. Adult hemoglobin contains two types of globin, two α chains (141 residues each) and two β chains (146 residues each). Although fewer than half of the amino acid residues in the polypeptide sequences of the α and β subunits are identical, the three-dimensional structures of the two types of subunits are very similar. Furthermore, their structures are very similar to that of myoglobin (Fig. 7–6), even though the amino acid sequences of the three polypeptides are identical at only 27 positions (Fig. 7–7). All three polypeptides are

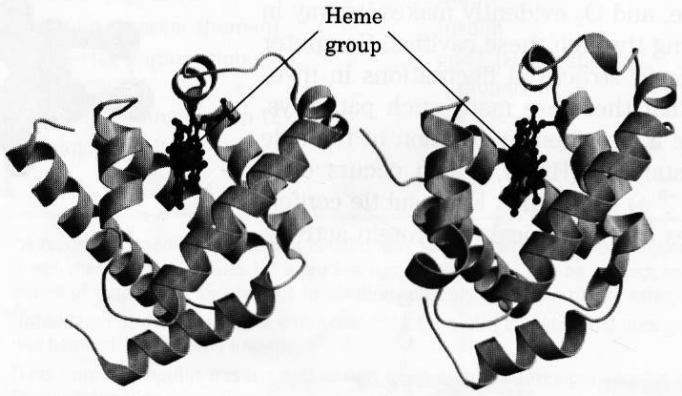


figure 7–6
A comparison of the structures of myoglobin and the β subunit of hemoglobin.

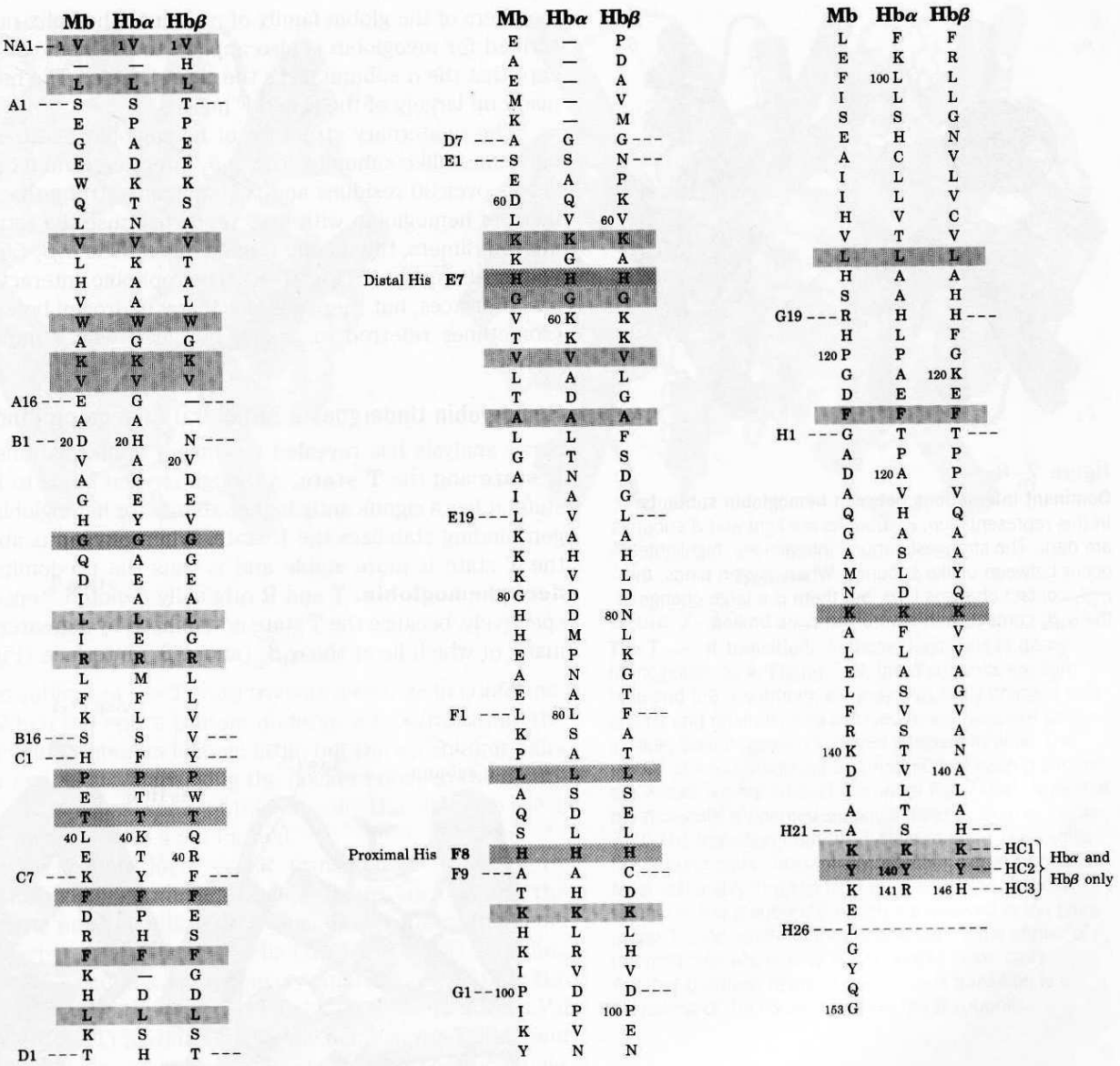


figure 7–7

The amino acid sequences of whale myoglobin and the α and β chains of human hemoglobin. Dashed lines mark helix boundaries. To align the sequences optimally, short breaks must be incorporated into both Hb sequences where a few amino acids are present in the other sequences. With the exception of the missing D helix in Hbα, this alignment permits the use of the helix lettering convention that emphasizes the common positioning of amino acid residues that are identical in all three structures (shaded). Residues shaded in red are conserved in all known globins. Note that a common letter-and-number designation for amino acids in two or three different structures does not necessarily correspond to a common position in the linear sequence of amino acids in the polypeptides. For example, the distal His residue is His E7 in all three structures, but corresponds to His^{64}, His^{58}, and His^{63} in the linear sequences of Mb, Hbα, and Hbβ, respectively. Nonhelical residues at the amino and carboxyl termini, beyond the first (A) and last (H) α-helical segments, are labeled NA and HC, respectively.

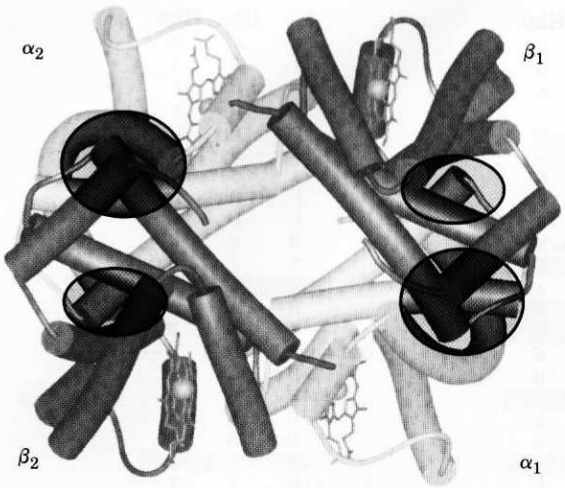


figure 7–8
Dominant interactions between hemoglobin subunits. In this representation, α subunits are light and β subunits are dark. The strongest subunit interactions, highlighted, occur between unlike subunits. When oxygen binds, the $\alpha_1\beta_1$ contact changes little, but there is a large change at the $\alpha_1\beta_2$ contact, with several ion pairs broken.

members of the globin family of proteins. The helix-naming convention described for myoglobin is also applied to the hemoglobin polypeptides, except that the α subunit lacks the short D helix. The heme-binding pocket is made up largely of the E and F helices.

The quaternary structure of hemoglobin features strong interactions between unlike subunits. The $\alpha_1\beta_1$ interface (and its $\alpha_2\beta_2$ counterpart) involves over 30 residues and is sufficiently strong that although mild treatment of hemoglobin with urea tends to cause the tetramer to disassemble into $\alpha\beta$ dimers, the dimers remain intact. The $\alpha_1\beta_2$ (and $\alpha_2\beta_1$) interface involves 19 residues (Fig. 7–8). Hydrophobic interactions predominate at the interfaces, but there are also many hydrogen bonds and a few ion pairs (sometimes referred to as salt bridges), whose importance is discussed below.

Hemoglobin Undergoes a Structural Change on Binding Oxygen

X-ray analysis has revealed two major conformations of hemoglobin: the **R state** and the **T state.** Although oxygen binds to hemoglobin in either state, it has a significantly higher affinity for hemoglobin in the R state. Oxygen binding stabilizes the R state. When oxygen is absent experimentally, the T state is more stable and is thus the predominant conformation of **deoxyhemoglobin.** T and R originally denoted "tense" and "relaxed," respectively, because the T state is stabilized by a greater number of ion pairs, many of which lie at the $\alpha_1\beta_2$ (and $\alpha_2\beta_1$) interface (Fig. 7–9). The binding

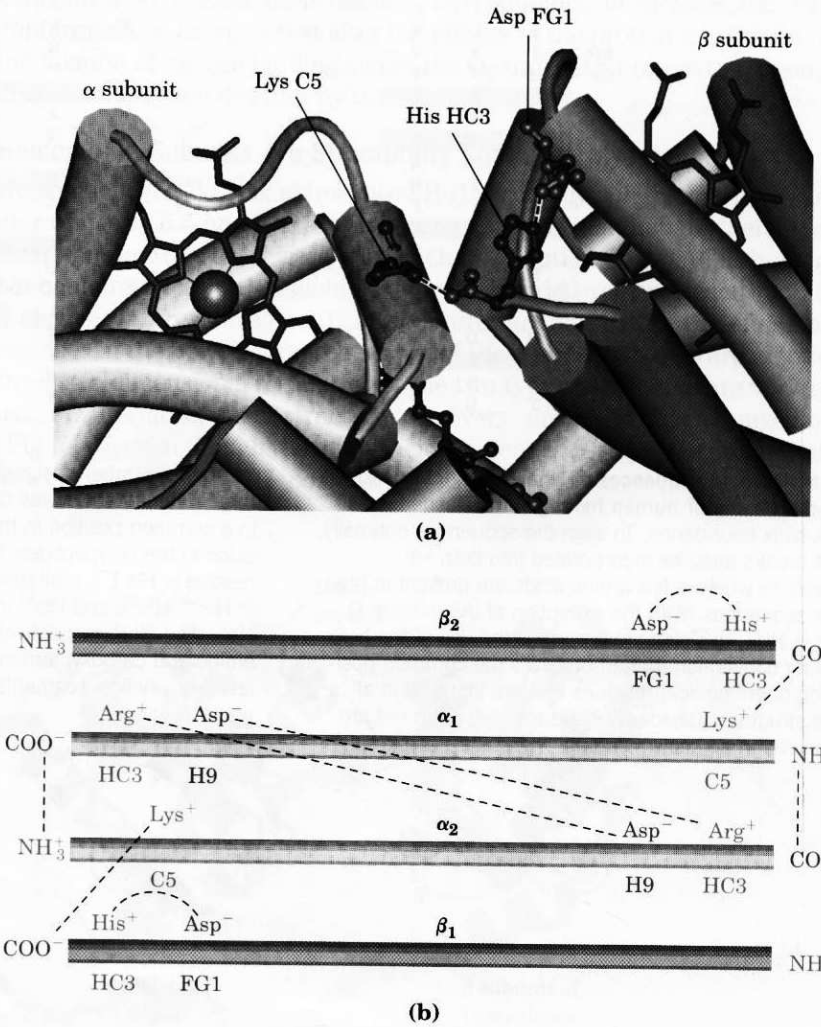


figure 7–9
Some ion pairs that stabilize the T state of deoxyhemoglobin. (a) A close-up view of a portion of a deoxyhemoglobin molecule in the T state. Interactions between the ion pairs His HC3 and Asp FG1 of the β subunit (blue) and between Lys C5 of the α subunit (gray) and the α-carboxyl group of His HC3 of the β subunit are shown with dashed lines. (Recall that HC3 is the carboxyl-terminal residue of the β subunit.) **(b)** The interactions between these ion pairs and others not shown in **(a)** are schematized in this representation of the extended polypeptide chains of hemoglobin.

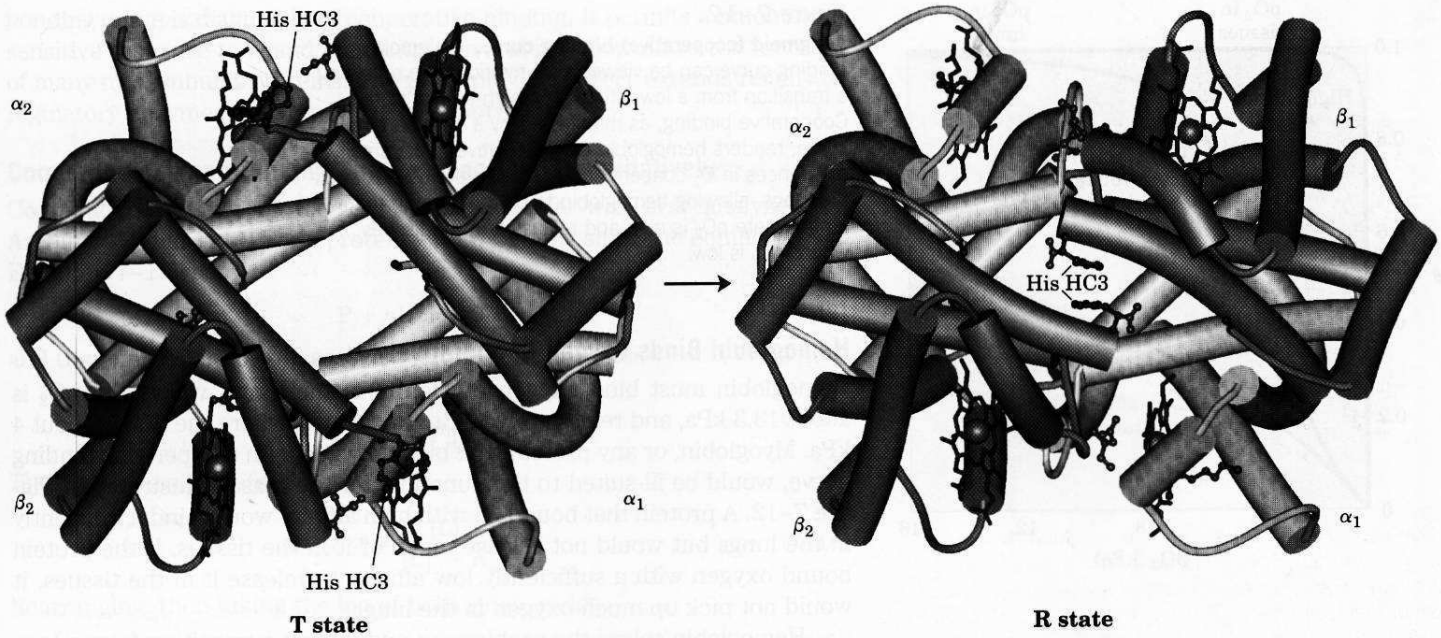


figure 7–10
The T ⟶ R transition. In these depictions of deoxyhemoglobin, as in Figure 7–9, the β subunits are light blue and the α subunits are gray. Positively charged side chains and chain termini involved in ion pairs are shown in blue, their negatively charged partners in pink. The Lys C5 of each α subunit and Asp FG1 of each β subunit are visible but not labeled (compare Fig. 7–9a). Note that the molecule is oriented slightly differently than in Figure 7–9. The transition from the T state to the R state shifts the subunit pairs substantially, affecting certain ion pairs. Most noticeably, the His HC3 residues at the carboxyl termini of the β subunits, which are involved in ion pairs in the T state, rotate in the R state toward the center of the molecule where they are no longer in ion pairs. Another dramatic result of the T ⟶ R transition is a narrowing of the pocket between the β subunits.

of O_2 to a hemoglobin subunit in the T state triggers a change in conformation to the R state. When the entire protein undergoes this transition, the structures of the individual subunits change little, but the $\alpha\beta$ subunit pairs slide past each other and rotate, narrowing the pocket between the β subunits (Fig. 7–10). In this process, some of the ion pairs that stabilize the T state are broken and some new ones are formed.

Max Perutz proposed that the T ⟶ R transition is triggered by changes in the positions of key amino acid side chains surrounding the heme. In the T state, the porphyrin is slightly puckered, causing the heme iron to protrude somewhat on the proximal His (His F8) side. The binding of O_2 causes the heme to assume a more planar conformation, shifting the position of the proximal His and the attached F helix (Fig. 7–11). Also, a Val residue in the E helix (Val E11) partially blocks the heme in the T state and must swing out of the way for oxygen to bind (Fig. 7–10). These changes lead to adjustments in the ion pairs at the $\alpha_1\beta_2$ interface.

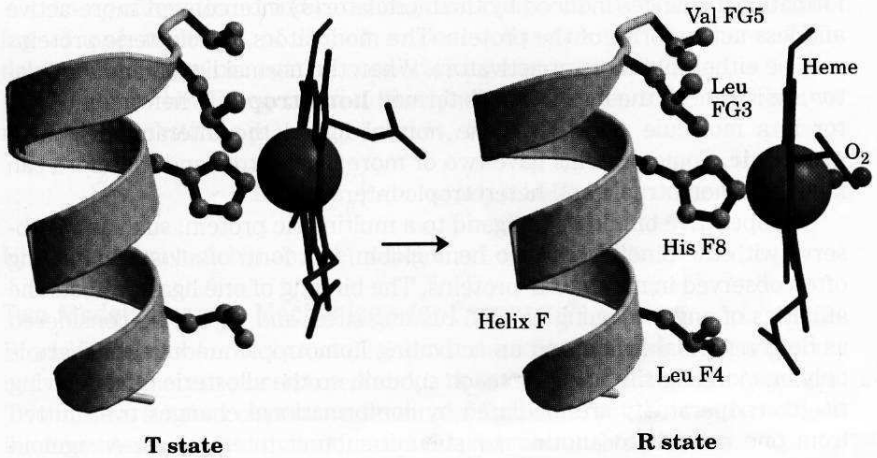


figure 7–11
Changes in conformation near heme on O_2 binding. The shift in the position of the F helix when heme binds O_2 is one of the adjustments that is believed to trigger the T ⟶ R transition.

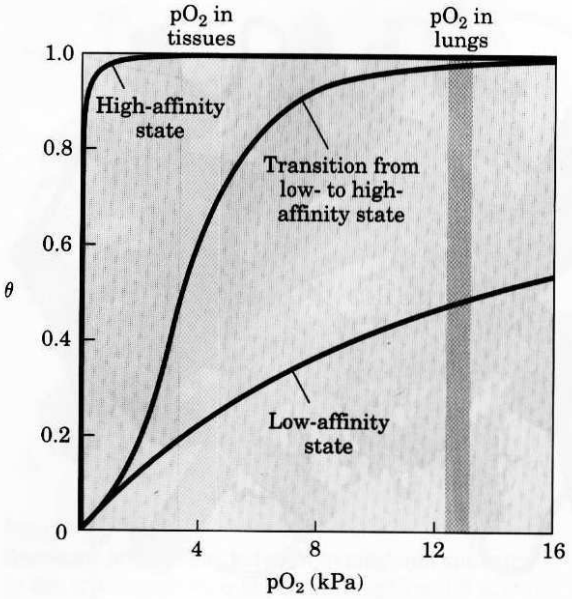


figure 7–12
A sigmoid (cooperative) binding curve. A sigmoid binding curve can be viewed as a hybrid curve reflecting a transition from a low-affinity to a high-affinity state. Cooperative binding, as manifested by a sigmoid binding curve, renders hemoglobin more sensitive to the small differences in O_2 concentration between the tissues and the lungs, allowing hemoglobin to bind oxygen in the lungs where pO_2 is high and release it in the tissues where pO_2 is low.

Hemoglobin Binds Oxygen Cooperatively

Hemoglobin must bind oxygen efficiently in the lungs, where the pO_2 is about 13.3 kPa, and release oxygen in the tissues, where the pO_2 is about 4 kPa. Myoglobin, or any protein that binds oxygen with a hyperbolic binding curve, would be ill-suited to this function, for the reason illustrated in Figure 7–12. A protein that bound O_2 with high affinity would bind it efficiently in the lungs but would not release much of it in the tissues. If the protein bound oxygen with a sufficiently low affinity to release it in the tissues, it would not pick up much oxygen in the lungs.

Hemoglobin solves the problem by undergoing a transition from a low-affinity state (the T state) to a high-affinity state (the R state) as more O_2 molecules are bound. As a result, hemoglobin has a hybrid S-shaped, or sigmoid, binding curve for oxygen (Fig. 7–12). A single-subunit protein with a single ligand-binding site cannot produce a sigmoid binding curve—even if binding elicits a conformational change—because each molecule of ligand binds independently and cannot affect the binding of another molecule. In contrast, O_2 binding to individual subunits of hemoglobin can alter the affinity for O_2 in adjacent subunits. The first molecule of O_2 that interacts with deoxyhemoglobin binds weakly, because it binds to a subunit in the T state. Its binding, however, leads to conformational changes that are communicated to adjacent subunits, making it easier for additional molecules of O_2 to bind. In effect, the T $\longrightarrow$ R transition occurs more readily in the second subunit once O_2 is bound to the first subunit. The last (fourth) O_2 molecule binds to a heme in a subunit that is already in the R state, and hence it binds with much higher affinity than the first molecule.

An **allosteric protein** is one in which the binding of a ligand to one site affects the binding properties of another site on the same protein. The term allosteric derives from the Greek *allos,* "other," and *stereos,* "solid" or "shape." Allosteric proteins are those having "other shapes" or conformations induced by the binding of ligands referred to as modulators. The conformational changes induced by the modulator(s) interconvert more-active and less-active forms of the protein. The modulators for allosteric proteins may be either inhibitors or activators. When the normal ligand and modulator are identical, the interaction is termed **homotropic.** When the modulator is a molecule other than the normal ligand the interaction is **heterotropic.** Some proteins have two or more modulators and therefore can have both homotropic and heterotropic interactions.

Cooperative binding of a ligand to a multimeric protein, such as we observe with the binding of O_2 to hemoglobin, is a form of allosteric binding often observed in multimeric proteins. The binding of one ligand affects the affinities of any remaining unfilled binding sites, and O_2 can be considered as both a normal ligand and an activating homotropic modulator. There is only one binding site for O_2 on each subunit, so the allosteric effects giving rise to cooperativity are mediated by conformational changes transmitted from one subunit to another by subunit-subunit interactions. A sigmoid

bonding curve is diagnostic of cooperative binding. It permits a much more sensitive response to ligand concentration and is important to the function of many multisubunit proteins. The principle of allostery extends readily to regulatory enzymes, as we will see in Chapter 8.

Cooperative Ligand Binding Can Be Described Quantitatively

Cooperative binding of oxygen by hemoglobin was first analyzed by Archibald Hill in 1910. For a protein with n binding sites, the equilibrium of Equation 7–1 becomes

$$P + nL \rightleftharpoons PL_n \tag{7–12}$$

and the expression for the association constant becomes

$$K_a = \frac{[PL_n]}{[P][L]^n} \tag{7–13}$$

The expression for θ (see Eqn 7–8) is

$$\theta = \frac{[L]^n}{[L]^n + K_d} \tag{7–14}$$

Rearranging, then taking the log of both sides, yields

$$\frac{\theta}{1-\theta} = \frac{[L]^n}{K_d} \tag{7–15}$$

$$\log\left(\frac{\theta}{1-\theta}\right) = n \log [L] - \log K_d \tag{7–16}$$

Equation 7–16 is the **Hill equation,** and a plot of log $[\theta/(1-\theta)]$ versus log [L] is called a **Hill plot.** Based on the equation, the Hill plot should have a slope of n. However, the experimentally determined slope actually reflects not the number of binding sites, but the degree of interaction between them. The slope of a Hill plot is therefore denoted n_H, the **Hill coefficient,** which is a measure of the degree of cooperativity. If n_H equals 1, ligand binding is not cooperative, a situation that can arise even in a multisubunit protein if the subunits do not communicate. An n_H of greater than 1 indicates positive cooperativity in ligand binding. This is the situation observed in hemoglobin, in which the binding of one molecule of ligand facilitates the binding of others. The theoretical upper limit for n_H is reached when $n_H = n$. In this case the binding would be completely cooperative: all binding sites on the protein would bind ligand simultaneously, and no protein molecules partially saturated with ligand would be present under any conditions. This limit is never reached in practice, and the measured value of n_H is always less than the actual number of ligand-binding sites in the protein.

An n_H of less than 1 indicates negative cooperativity, in which the binding of one molecule of ligand *impedes* the binding of others. Well-documented cases of negative cooperativity are rare.

To adapt the Hill equation to the binding of oxygen to hemoglobin we must again substitute pO_2 for [L] and P_{50} for K_d:

$$\log\left(\frac{\theta}{1-\theta}\right) = n \log pO_2 - \log P_{50} \tag{7–17}$$

Hill plots for myoglobin and hemoglobin are given in Figure 7–13.

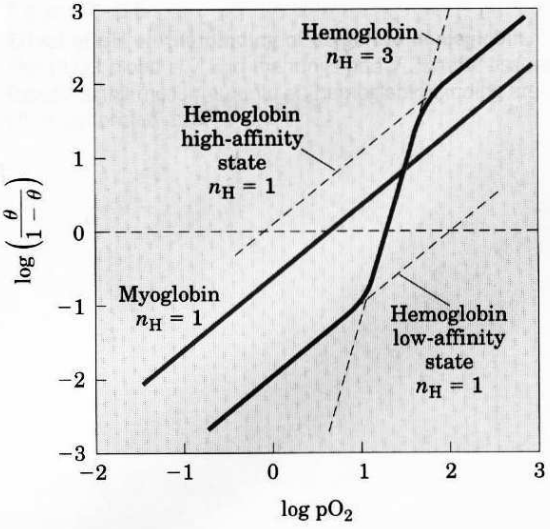


figure 7–13

Hill plots for the binding of oxygen to myoglobin and hemoglobin. When $n_H = 1$, there is no evident cooperativity. The maximum degree of cooperativity observed for hemoglobin corresponds approximately to $n_H = 3$. Note that while this indicates a high level of cooperativity, n_H is less than n, the number of O_2-binding sites in hemoglobin. This is normal for a protein that exhibits allosteric binding behavior.

Two Models Suggest Mechanisms for Cooperative Binding

Biochemists now know a great deal about the T and R states of hemoglobin, but much remains to be learned about how the T ⟶ R transition occurs. Two models for the cooperative binding of ligands to proteins with multiple binding sites have greatly influenced thinking about this problem.

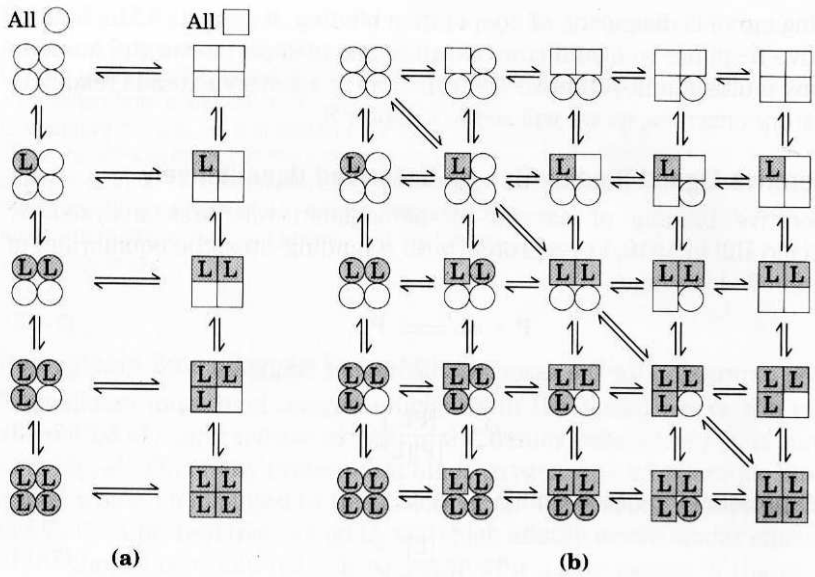


figure 7–14

Two general models for the interconversion of inactive and active forms of cooperative ligand-binding proteins. Although the models may be applied to any protein—including any enzyme (Chapter 8)—that exhibits cooperative binding, four subunits are shown because the model was originally proposed for hemoglobin. In the concerted, or all-or-none, model **(a)** all the subunits are postulated to be in the same conformation, either all ○ (low affinity or inactive) or all □ (high affinity or active). Depending on the equilibrium, K_1, between ○ and □ forms, the binding of one or more ligand molecules (L) will pull the equilibrium toward the □ form. Subunits with bound L are shaded. In the sequential model **(b)** each individual subunit can be in either the ○ or □ form. A very large number of conformations is thus possible.

The first model was proposed by Jacques Monod, Jeffries Wyman, and Jean-Pierre Changeux in 1965, and is called the **MWC model** or the **concerted model** (Fig. 7–14a). The concerted model assumes that the subunits of a cooperatively binding protein are functionally identical, that each subunit can exist in (at least) two conformations, and that all subunits undergo the transition from one conformation to the other simultaneously. In this model, no protein has individual subunits in different conformations. The two conformations are in equilibrium. The ligand can bind to either conformation, but binds each with different affinity. Successive binding of ligand molecules to the low-affinity conformation (which is more stable in the absence of ligand) makes a transition to the high-affinity conformation more likely.

In the second model, the **sequential model** (Fig. 7–14b), proposed in 1966 by Daniel Koshland and colleagues, ligand binding can induce a change of conformation in an individual subunit. A conformational change in one subunit makes a similar change in an adjacent subunit, as well as the binding of a second ligand molecule, more likely. There are more potential intermediate states in this model than in the concerted model. The two models are not mutually exclusive; the concerted model may be viewed as the "all-or-none" limiting case of the sequential model. In Chapter 8 we will use these models when we investigate allosteric enzymes.

Hemoglobin Also Transports H^+ and CO_2

In addition to carrying nearly all the oxygen required by cells from the lungs to the tissues, hemoglobin carries two end products of cellular respiration—H^+ and CO_2—from the tissues to the lungs and the kidneys, where they are excreted. The CO_2, produced by oxidation of organic fuels in mitochondria, is hydrated to form bicarbonate:

$$CO_2 + H_2O \rightleftharpoons H^+ + HCO_3^-$$

This reaction is catalyzed by **carbonic anhydrase,** an enzyme particularly abundant in erythrocytes. Carbon dioxide is not very soluble in aqueous solution, and bubbles of CO_2 would form in the tissues and blood if it were not converted to bicarbonate. As you can see from the equation, the hydration of CO_2 results in an increase in the H^+ concentration (a decrease in pH) in

the tissues. The binding of oxygen by hemoglobin is profoundly influenced by pH and CO_2 concentration, so the interconversion of CO_2 and bicarbonate is of great importance to the regulation of oxygen binding and release in the blood.

Hemoglobin transports about 20% of the total H^+ and CO_2 formed in the tissues to the lungs and the kidneys. The binding of H^+ and CO_2 is inversely related to the binding of oxygen. At the relatively low pH and high CO_2 concentration of peripheral tissues, the affinity of hemoglobin for oxygen decreases as H^+ and CO_2 are bound, and O_2 is released to the tissues. Conversely, in the capillaries of the lung, as CO_2 is excreted and the blood pH consequently rises, the affinity of hemoglobin for oxygen increases and the protein binds more O_2 for transport to the peripheral tissues. This effect of pH and CO_2 concentration on the binding and release of oxygen by hemoglobin is called the **Bohr effect,** after Christian Bohr, the Danish physiologist (and father of physicist Niels Bohr) who discovered it in 1904.

The binding equilibrium for hemoglobin and one molecule of oxygen can be designated by the reaction

$$Hb + O_2 \rightleftharpoons HbO_2$$

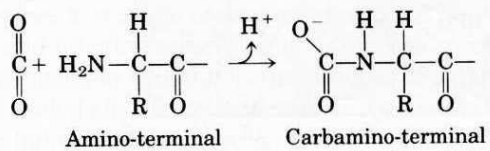
but this is not a complete statement. To account for the effect of H^+ concentration on this binding equilibrium, we rewrite the reaction as

$$HHb^+ + O_2 \rightleftharpoons HbO_2 + H^+$$

where HHb^+ denotes a protonated form of hemoglobin. This equation tells us that the O_2-saturation curve of hemoglobin is influenced by the H^+ concentration (Fig. 7–15). Both O_2 and H^+ are bound by hemoglobin, but with inverse affinity. When the oxygen concentration is high, as in the lungs, hemoglobin binds O_2 and releases protons. When the oxygen concentration is low, as in the peripheral tissues, H^+ is bound and O_2 is released.

Oxygen and H^+ are not bound at the same sites in hemoglobin. Oxygen binds to the iron atoms of the hemes, whereas H^+ binds to any of several amino acid residues in the protein. A major contribution to the Bohr effect is made by His^{146} (His HC3) of the β subunits. When protonated, this residue forms one of the ion pairs—to Asp^{94} (Asp FG1)—that helps stabilize deoxyhemoglobin in the T state (Fig. 7–9). The ion pair stabilizes the protonated form of His HC3, giving this residue an abnormally high pK_a in the T state. The pK_a falls to its normal value of 6.0 in the R state because the ion pair cannot form, and this residue is largely unprotonated in oxyhemoglobin at pH 7.6, the blood pH in the lungs. As the concentration of H^+ rises, protonation of His HC3 promotes release of oxygen by favoring a transition to the T state. Protonation of the amino-terminal residues of the α subunits, certain other His residues, and perhaps other groups has a similar effect.

Thus we see that the four polypeptide chains of hemoglobin communicate with each other not only about O_2 binding to their heme groups, but also about H^+ binding to specific amino acid residues. And there is still more to the story. Hemoglobin also binds CO_2, again in a manner inversely related to the binding of oxygen. Carbon dioxide binds as a carbamate group to the α-amino group at the amino-terminal end of each globin chain, forming carbaminohemoglobin:



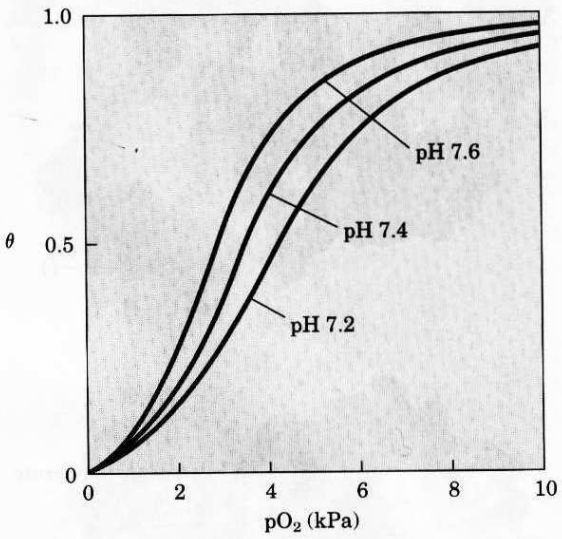


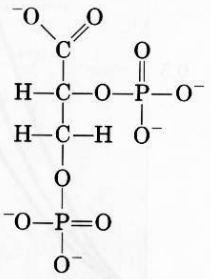
figure 7–15
Effect of pH on the binding of oxygen to hemoglobin. The pH of blood is 7.6 in the lungs and 7.2 in the tissues. Experimental measurements on hemoglobin binding are often performed at pH 7.4.

This reaction produces H^+, contributing to the Bohr effect. The bound carbamates also form additional salt bridges (not shown in Fig. 7–9) that help to stabilize the T state and promote the release of oxygen.

When the concentration of carbon dioxide is high, as in peripheral tissues, some CO_2 binds to hemoglobin and the affinity for O_2 decreases, causing its release. Conversely, when hemoglobin reaches the lungs, the high oxygen concentration promotes binding of O_2 and release of CO_2. It is the capacity to communicate ligand-binding information from one polypeptide subunit to the others that makes the hemoglobin molecule so beautifully adapted to integrating the transport of O_2, CO_2, and H^+ by erythrocytes.

Oxygen Binding to Hemoglobin Is Regulated by 2,3-Bisphosphoglycerate

The interaction of **2,3-bisphosphoglycerate** (BPG) with hemoglobin provides an example of heterotropic allosteric modulation. BPG is present in relatively high concentrations in erythrocytes. When hemoglobin is isolated, it contains substantial amounts of bound BPG, which can be difficult to remove completely. In fact, the O_2-binding curves for hemoglobin that we have examined to this point were obtained in the presence of bound BPG. 2,3-Bisphosphoglycerate is known to greatly reduce the affinity of hemoglobin for oxygen—there is an inverse relationship between the binding of O_2 and the binding of BPG. We can therefore describe another binding process for hemoglobin:



2,3-Bisphosphoglycerate

$$HbBPG + O_2 \rightleftharpoons HbO_2 + BPG$$

BPG binds at a site distant from the oxygen-binding site and regulates the O_2-binding affinity of hemoglobin in relation to the pO_2 in the lungs. BPG plays an important role in the physiological adaptation to the lower pO_2 available at high altitudes. For a healthy human strolling by the ocean, the binding of O_2 to hemoglobin is regulated such that the amount of O_2 delivered to the tissues is equivalent to nearly 40% of the maximum that could be carried by the blood (Fig. 7–16). If the same person is quickly transported to a mountainside at an altitude of 4,500 meters, where the pO_2 is considerably lower, the delivery of O_2 to the tissues is reduced. However,

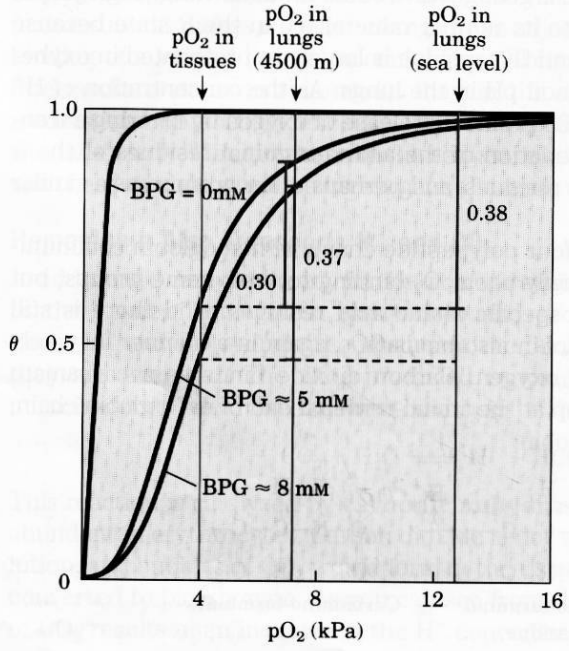


figure 7–16

Effect of BPG on the binding of oxygen to hemoglobin. The BPG concentration in normal human blood is about 5 mM at sea level and about 8 mM at high altitudes. Note that hemoglobin binds to oxygen quite tightly when BPG is entirely absent, and the binding curve appears to be hyperbolic. In reality, the measured Hill coefficient for O_2-binding cooperativity decreases only slightly (from 3 to about 2.5) when BPG is removed from hemoglobin, but the rising part of the sigmoid curve is confined to a very small region close to the origin. At sea level, hemoglobin is nearly saturated with O_2 in the lungs, but only 60% saturated in the tissues, so that the amount of oxygen released in the tissues is close to 40% of the maximum that can be carried in the blood. At high altitudes, O_2 delivery declines by about one-fourth, to 30% of maximum. An increase in BPG concentration, however, decreases the affinity of hemoglobin for O_2 so that nearly 40% of what can be carried is again delivered to the tissues.

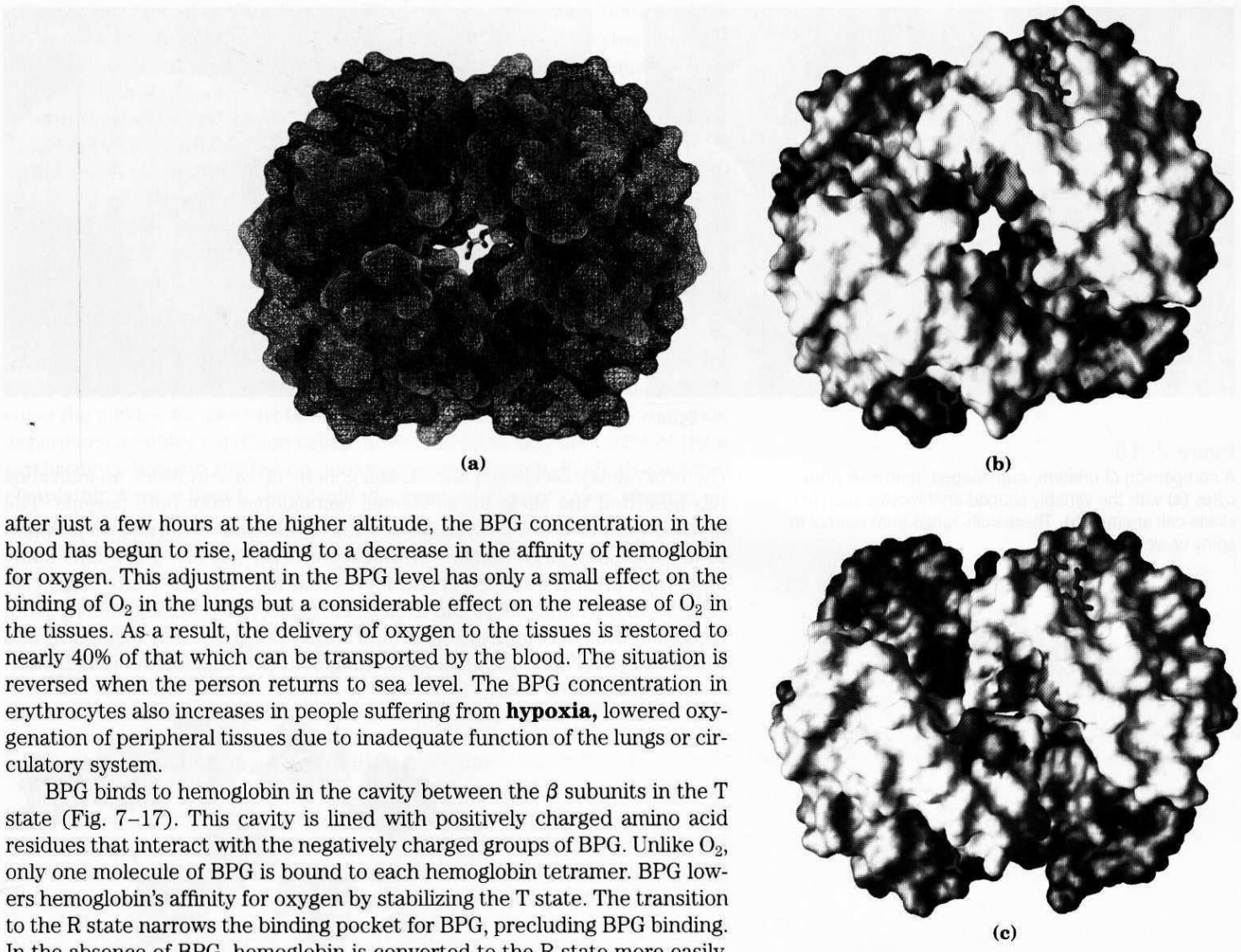


figure 7–17
Binding of BPG to deoxyhemoglobin. (a) BPG binding stabilizes the T state of deoxyhemoglobin, shown here as a mesh surface image. **(b)** The negative charges of BPG interact with several positively charged groups (shown in blue in this GRASP surface image) that surround the pocket between the β subunits in the T state. **(c)** The binding pocket for BPG disappears on oxygenation, following transition to the R state. (Compare **(b)** and **(c)** with Fig. 7–10.)

after just a few hours at the higher altitude, the BPG concentration in the blood has begun to rise, leading to a decrease in the affinity of hemoglobin for oxygen. This adjustment in the BPG level has only a small effect on the binding of O_2 in the lungs but a considerable effect on the release of O_2 in the tissues. As a result, the delivery of oxygen to the tissues is restored to nearly 40% of that which can be transported by the blood. The situation is reversed when the person returns to sea level. The BPG concentration in erythrocytes also increases in people suffering from **hypoxia,** lowered oxygenation of peripheral tissues due to inadequate function of the lungs or circulatory system.

BPG binds to hemoglobin in the cavity between the β subunits in the T state (Fig. 7–17). This cavity is lined with positively charged amino acid residues that interact with the negatively charged groups of BPG. Unlike O_2, only one molecule of BPG is bound to each hemoglobin tetramer. BPG lowers hemoglobin's affinity for oxygen by stabilizing the T state. The transition to the R state narrows the binding pocket for BPG, precluding BPG binding. In the absence of BPG, hemoglobin is converted to the R state more easily.

Regulation of oxygen binding to hemoglobin by BPG has an important role in fetal development. Because a fetus must extract oxygen from its mother's blood, fetal hemoglobin must have greater affinity than the maternal hemoglobin for O_2. In fetuses, γ subunits are synthesized rather than β subunits, and $\alpha_2\gamma_2$ hemoglobin is formed. This tetramer has a much lower affinity for BPG than normal adult hemoglobin, and a correspondingly higher affinity for O_2.

Sickle-Cell Anemia Is a Molecular Disease of Hemoglobin

The great importance of the amino acid sequence in determining the secondary, tertiary, and quaternary structures of globular proteins, and thus their biological functions, is strikingly demonstrated by the hereditary human disease sickle-cell anemia. More than 300 genetic variants of hemoglobin are known to occur in the human population. Most of these variations consist of differences in a single amino acid residue. The effects on hemoglobin structure and function are often minor but can sometimes be extraordinary. Each hemoglobin variation is the product of an altered gene. The variant genes are called alleles. Because humans generally have two copies of each gene, an individual may have two copies of one allele (thus being homozygous for that gene) or one copy of each of two different alleles

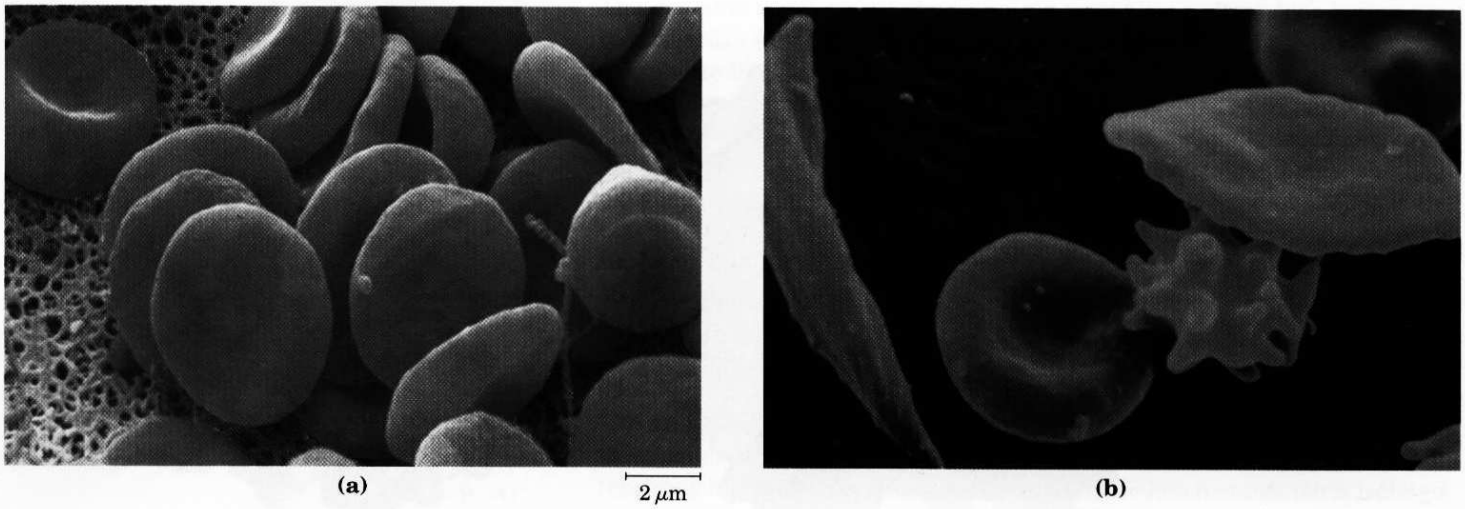


figure 7–18
A comparison of uniform, cup-shaped, normal erythrocytes **(a)** with the variably shaped erythrocytes seen in sickle-cell anemia **(b).** These cells range from normal to spiny or sickle-shaped.

(heterozygous). Sickle-cell anemia is a genetic disease in which an individual has inherited the allele for sickle-cell hemoglobin from both parents. The erythrocytes of these individuals are fewer and also abnormal. In addition to an unusually large number of immature cells, the blood contains many long, thin, crescent-shaped erythrocytes that look like the blade of a sickle (Fig. 7–18). When hemoglobin from sickle cells (called hemoglobin S) is deoxygenated, it becomes insoluble and forms polymers that aggregate into tubular fibers (Fig. 7–19). Normal hemoglobin (hemoglobin A) remains sol-

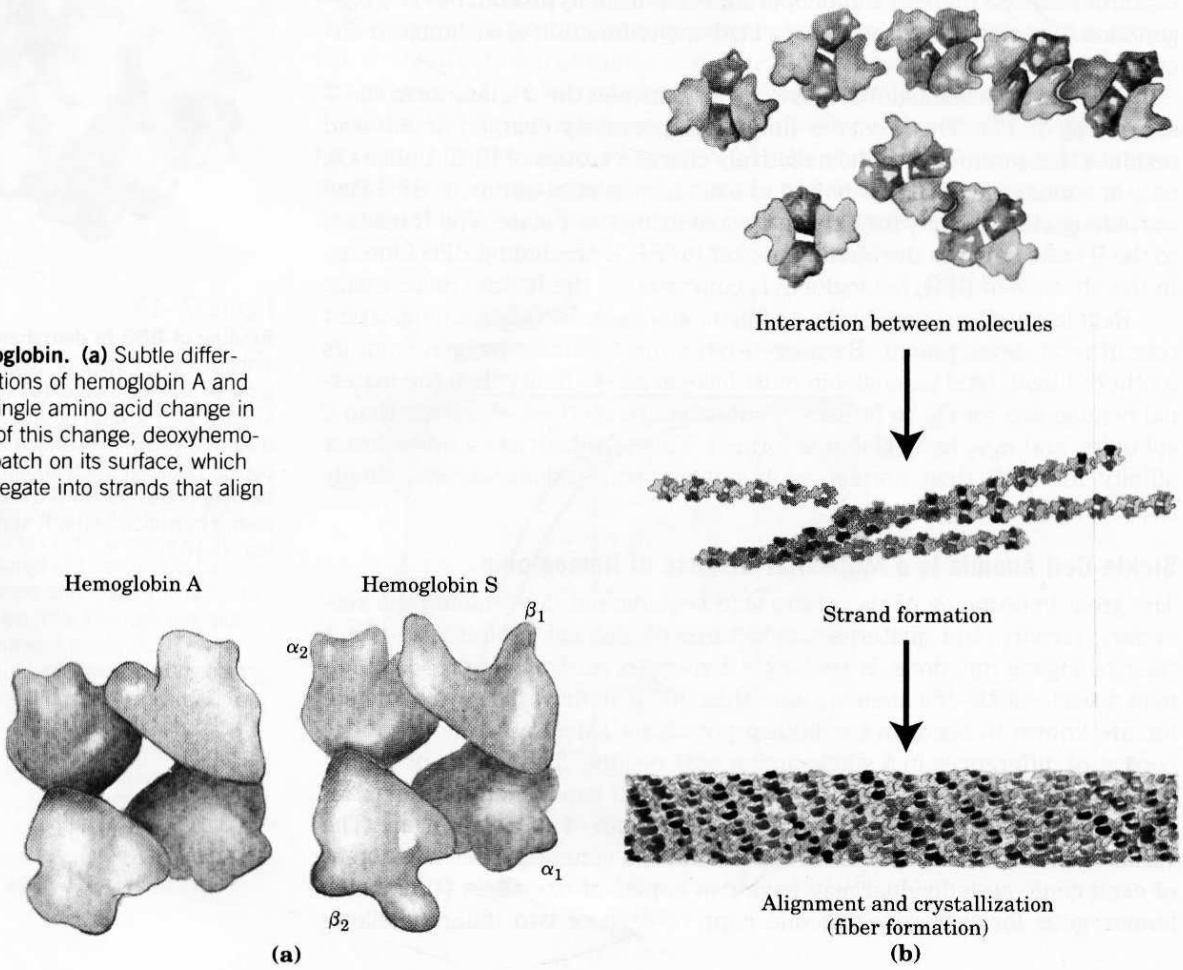


figure 7–19
Normal and sickle-cell hemoglobin. (a) Subtle differences between the conformations of hemoglobin A and hemoglobin S result from a single amino acid change in the β chains. **(b)** As a result of this change, deoxyhemoglobin S has a hydrophobic patch on its surface, which causes the molecules to aggregate into strands that align into insoluble fibers.

uble on deoxygenation. The insoluble fibers of deoxygenated hemoglobin S are responsible for the deformed sickle shape of the erythrocytes, and the proportion of sickled cells increases greatly as blood is deoxygenated.

The altered properties of hemoglobin S result from a single amino acid substitution, a Val instead of a Glu residue at position 6 in the two β chains. The R group of valine has no electric charge, whereas glutamate has a negative charge at pH 7.4. Hemoglobin S therefore has two fewer negative charges than hemoglobin A, one for each of the two β chains. Replacement of the Glu residue by Val creates a "sticky" hydrophobic contact point at position 6 of the β chain, which is on the outer surface of the molecule. These sticky spots cause deoxyhemoglobin S molecules to associate abnormally with each other, forming the long, fibrous aggregates characteristic of this disorder.

Sickle-cell anemia occurs in individuals homozygous for the sickle-cell allele of the gene encoding the β subunit of hemoglobin. Individuals who receive the sickle-cell allele from only one parent and are thus heterozygous experience a milder condition called sickle-cell trait; only about 1% of their erythrocytes become sickled on deoxygenation. These individuals may live completely normal lives if they avoid vigorous exercise or other stresses on the circulatory system.

People with sickle-cell anemia suffer from repeated crises brought on by physical exertion. They become weak, dizzy, and short of breath, and they also experience heart murmurs and an increased pulse rate. The hemoglobin content of their blood is only about half the normal value of 15 to 16 g/100 mL because sickled cells are very fragile and rupture easily; this results in anemia ("lack of blood"). An even more serious consequence is that capillaries become blocked by the long, abnormally shaped cells, causing severe pain and interfering with normal organ function—a major factor in the early death of many people with the disease.

Without medical treatment, people with sickle-cell anemia usually die in childhood. Nevertheless, the sickle-cell allele is surprisingly common in certain parts of Africa. Investigation into the persistence of an allele that is so obviously deleterious in homozygous individuals led to the finding that the allele confers a small but significant resistance to lethal forms of malaria in heterozygous individuals. Natural selection has resulted in an allele population that balances the deleterious effects of the homozygous condition against the resistance to malaria afforded by the heterozygous condition.

Complementary Interactions between Proteins and Ligands: The Immune System and Immunoglobulins

Our discussion of oxygen-binding proteins showed how the conformations of these proteins affect and are affected by the binding of small ligands (O_2 or CO) to the heme group. However, most protein-ligand interactions do not involve a prosthetic group. Instead, the binding site for a ligand is more often like the hemoglobin binding site for BPG—a cleft in the protein lined with amino acid residues, arranged to render the binding interaction highly specific. Effective discrimination between ligands is the norm at binding sites, even when the ligands have only minor structural differences.

All vertebrates have an immune system capable of distinguishing molecular "self" from "nonself" and then destroying those entities identified as nonself. In this way, the immune system eliminates viruses, bacteria, and other pathogens and molecules that may pose a threat to the organism. On a physiological level, the response of the immune system to an invader is an intricate and coordinated set of interactions among many classes of proteins,

molecules, and cell types. However, at the level of individual proteins, the immune response demonstrates how an acutely sensitive and specific biochemical system is built upon the reversible binding of ligands to proteins.

The Immune Response Features a Specialized Array of Cells and Proteins

Immunity is brought about by a variety of **leukocytes** (white blood cells), including **macrophages** and **lymphocytes,** all arising from undifferentiated stem cells in the bone marrow. Leukocytes can leave the bloodstream and patrol the tissues, each cell producing one or more proteins capable of recognizing and binding to molecules that might signal an infection.

The immune response consists of two complementary systems, the humoral and cellular immune systems. The **humoral immune system** (Latin *humor,* "fluid") is directed at bacterial infections and extracellular viruses (those found in the body fluids), but can also respond to individual proteins introduced into the organism. The **cellular immune system** destroys host cells infected by viruses and also destroys some parasites and foreign tissues.

The proteins at the heart of the humoral immune response are soluble proteins called **antibodies** or **immunoglobulins,** often abbreviated Ig. Immunoglobulins bind bacteria, viruses, or large molecules identified as foreign and target them for destruction. Making up 20% of blood protein, the immunoglobulins are produced by **B lymphocytes** or **B cells,** so named because they complete their development in the *b*one marrow.

The agents at the heart of the cellular immune response are a class of **T lymphocytes** or **T cells** (so called because the latter stages of their development occur in the *t*hymus) known as **cytotoxic T cells (T_C cells,** also called killer T cells). Recognition of infected cells or parasites involves proteins called **T-cell receptors** on the surface of T_C cells. Recall from Chapter 2 (p. 30) that receptors are proteins, usually found on the outer surface of cells and extending through the plasma membrane; they recognize and bind extracellular ligands, triggering changes inside the cell.

In addition to cytotoxic T cells, there are **helper T cells (T_H cells),** whose function it is to produce soluble signaling proteins called cytokines, which include the interleukins. T_H cells interact with macrophages. Table 7–2 summarizes the functions of the various leukocytes of the immune system.

Each recognition protein of the immune system, either an antibody produced by a B cell or a receptor on the surface of a T cell, specifically binds some particular chemical structure, distinguishing it from virtually all others. Humans are capable of producing over 10^8 different antibodies with distinct binding specificities. This extraordinary diversity makes it likely that any chemical structure on the surface of a virus or invading cell will be recognized and bound by one or more antibodies. Antibody diversity is derived from random reassembly of a set of immunoglobulin gene segments via genetic recombination mechanisms that are discussed in Chapter 25.

Some properties of the interactions between antibodies or T-cell receptors and the molecules they bind are unique to the immune system, and a specialized lexicon is used to describe them. Any molecule or pathogen capable of eliciting an immune response is called an **antigen.** An antigen may be a virus, a bacterial cell wall, or an individual protein or other macromolecule. A complex antigen may be bound by a number of different antibodies. An individual antibody or T-cell receptor binds only a particular molecular structure within the antigen, called its **antigenic determinant** or **epitope.**

It would be unproductive for the immune system to respond to small molecules that are common intermediates and products of cellular metabolism. Molecules of M_r <5,000 are generally not antigenic. However, small molecules can be covalently attached to large proteins in the laboratory, and in this

table 7–2

Some Types of Leukocytes Associated with the Immune System

Cell type	Function
Macrophages	Ingest large particles and cells by phagocytosis
B lymphocytes (B cells)	Produce and secrete antibodies
T lymphocytes (T cells)	
Cytotoxic (killer) T cells (T_C)	Interact with infected host cells through receptors on T-cell surface
Helper T cells (T_H)	Interact with macrophages and secrete cytokines (interleukins) that stimulate T_C, T_H, and B cells to proliferate.

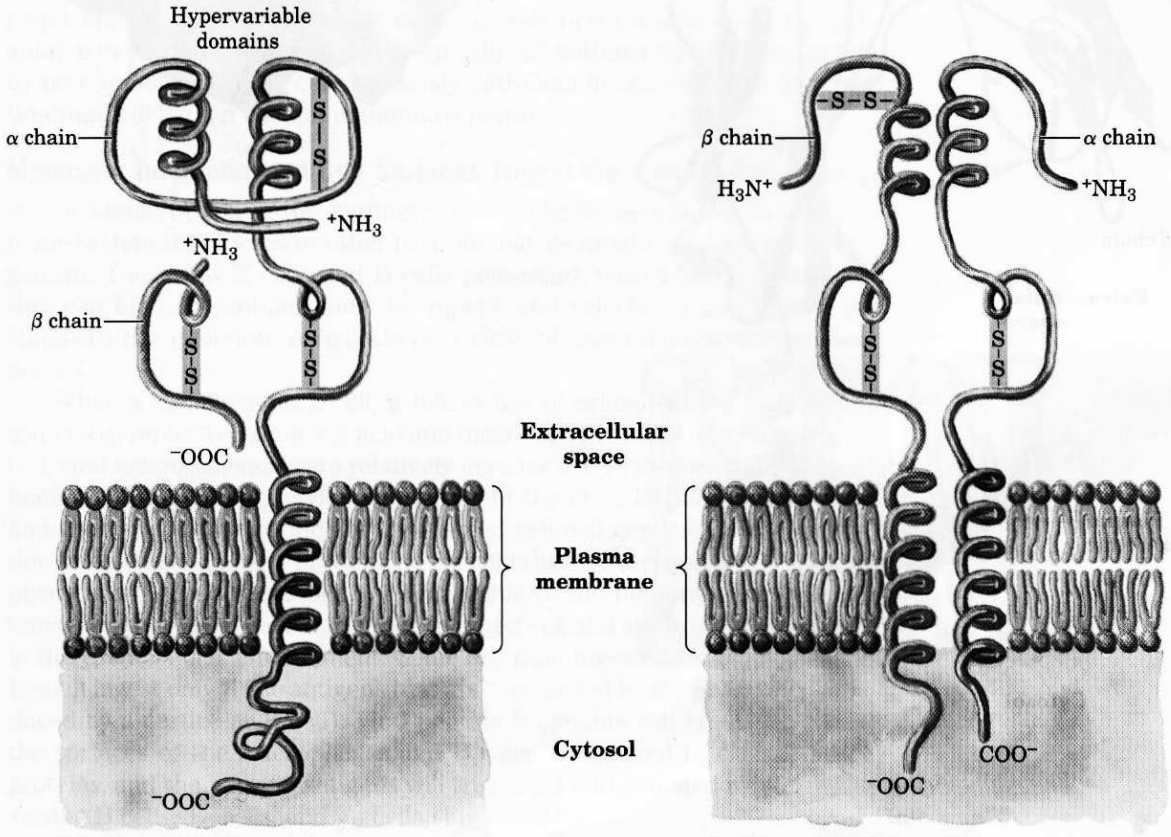
form they may elicit an immune response. These small molecules are called **haptens.** The antibodies produced in response to protein-linked haptens will then bind to the same small molecules when they are free. Such antibodies are sometimes used in the development of analytical tests described later in this chapter or as catalytic antibodies (described in Box 8–3).

The interactions of antibody and antigen are much better understood than are the binding properties of T-cell receptors. However, before focusing on antibodies, we need to look at the humoral and cellular immune systems in more detail to put the fundamental biochemical interactions into their proper context.

Self Is Distinguished from Nonself by the Display of Peptides on Cell Surfaces

The immune system must identify and destroy pathogens, but it must also recognize and *not* destroy the normal proteins and cells of the host organism—the "self." Detection of protein antigens in the host is mediated by **MHC (major histocompatibility complex) proteins.** MHC proteins bind peptide fragments of proteins digested in the cell and present them on the outside surface of the cell. These peptides normally come from the digestion of typical cellular proteins, but during a viral infection viral proteins are also digested and presented by MHC proteins. Peptide fragments from foreign proteins that are displayed by MHC proteins are the antigens the immune system recognizes as nonself. T-cell receptors bind these fragments and launch the subsequent steps of the immune response. There are two classes of MHC proteins (Fig. 7–20), which differ in their distribution among cell types and in the source of digested proteins whose peptides they display.

figure 7–20
MHC proteins These proteins consist of α and β chains. In class I MHC proteins **(a),** the small β chain is invariant but the amino acid sequence of the α chain exhibits a high degree of variability, localized in specific domains of the protein that appear on the outside of the cell. Each human produces up to six different α chains for class 1 MHC proteins. In class II MHC proteins **(b),** both the α and β chains have regions of relatively high variability near their amino-terminal ends.



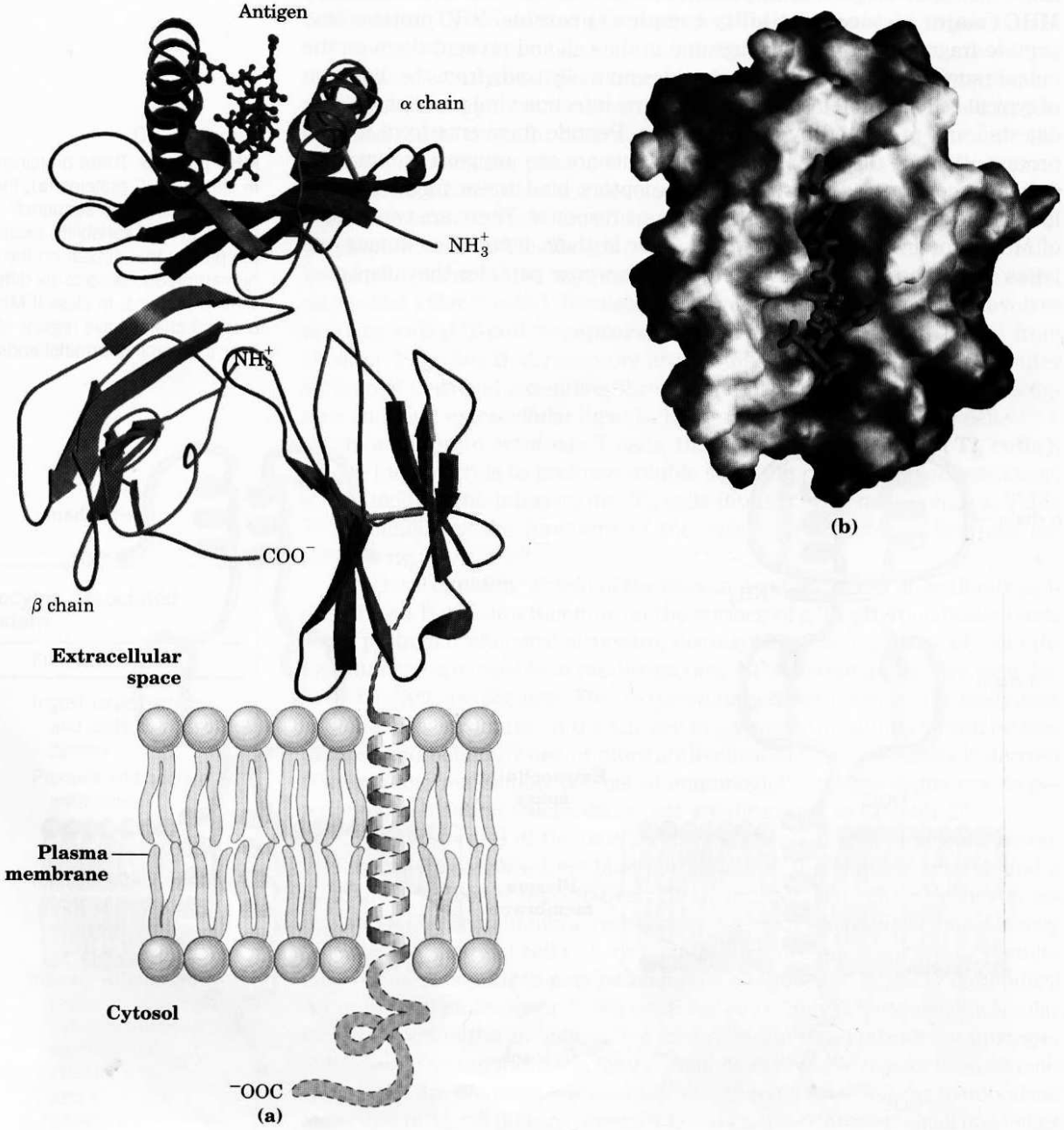
(a) Class I MHC protein **(b) Class II MHC protein**

Class I MHC proteins (Fig. 7–21) are found on the surface of virtually all vertebrate cells. There are countless variants in the human population, placing them among the most polymorphic of proteins. Because individuals produce up to six class I MHC protein variants, any two individuals are unlikely to have the same set. Class I MHC proteins bind and display peptides derived from the proteolytic degradation and turnover of proteins that occurs randomly within the cell. These complexes of peptides and class I MHC proteins are the recognition targets of the T-cell receptors of the T_C cells in the cellular immune system. The general pattern of immune system recognition was first described by Rolf Zinkernagel and Peter Doherty in 1974.

Each T_C cell has many copies of only one T-cell receptor that is specific for a particular class I MHC protein–peptide complex. To avoid creating a

figure 7–21

Structure of a human class I MHC protein. (a) This image is derived in part from the determined structure of the extracellular portion of the protein. The α chain of MHC is shown in gray; the small β chain is blue; the disulfide bonds are yellow. A bound ligand, a peptide derived from HIV, is shown in red. **(b)** Top view showing a surface contour image of the site where peptides are bound and displayed. The HIV peptide (red) occupies the site. This part of the class I MHC protein interacts with T-cell receptors.



legion of T_C cells that would set upon and destroy normal cells, the maturation of T_C cells in the thymus includes a stringent selection process that eliminates more than 95% of the developing T_C cells, including those that might recognize and bind class I MHC proteins displaying peptides from cellular proteins of the organism itself. The T_C cells that survive and mature are those with T-cell receptors that do not bind to the organism's own proteins. The result is a population of cells that bind foreign peptides bound to class I MHC proteins of the host cell. These binding interactions lead to the destruction of parasites and virus-infected cells. When an organ is transplanted, its foreign class I MHC proteins are also bound by T_C cells, leading to tissue rejection.

Class II MHC proteins occur on the surfaces of a few types of specialized cells that take up foreign antigens, including macrophages and B lymphocytes. Like class I MHC proteins, the class II proteins are highly polymorphic, with many variants in the human population. Each human is capable of producing up to 12 variants, and thus it is unlikely that any two individuals have an identical set of variants. The class II MHC proteins bind and display peptides derived not from cellular proteins but from external proteins ingested by the cells. The resulting class II MHC protein–peptide complexes are the binding targets of the T-cell receptors of the various helper T cells. T_H cells, like T_C cells, undergo a stringent selection process in the thymus, eliminating those that recognize the individual's own cellular proteins.

Despite the elimination of most T_C and T_H cells during the selection process in the thymus, a very large number survive, and these provide the immune response. Each survivor has a single type of T-cell receptor that can bind to one particular chemical structure. The T cells patrolling the bloodstream and the tissues carry millions of different binding specificities in the T-cell receptors. Within the highly varied T-cell population there is almost always a contingent of cells that can specifically bind any antigen that might appear. The vast majority of these cells never encounter a foreign antigen to which they can bind and typically die within a few days, replaced by new generations of T cells endlessly patrolling in search of the interaction that will launch the full immune response.

Molecular Interactions at Cell Surfaces Trigger the Immune Response

A new antigen is often the harbinger of an infection—a signal to the immune system that a virus or other parasite may be rapidly growing in the organism. Those few T cells and B cells possessing receptors or antibodies that can bind the antigen must be rapidly and selectively propagated to eliminate the infection. A hypothetical viral infection illustrates how this occurs.

When a virus invades a cell, it makes use of cellular functions and resources to replicate its nucleic acid and make viral proteins. Once inside the cell, viral macromolecules are relatively inaccessible to the antibodies of the humoral immune system. However, some of the class I MHC proteins that find their way to the surface of an infected cell will generally display peptide fragments from viral proteins, which can then be recognized by T_C lymphocytes. Mature viruses become vulnerable to the humoral immune system when they are released from the infected cell and are present for a time in the extracellular environment. Some are then ingested by macrophages (which ingest only those antigens that are recognized by the antibodies produced by a particular B cell). Viral peptide fragments will be displayed on the surfaces of the macrophages and B cells, complexed to class II MHC proteins, and the peptide antigens will trigger a multi-pronged response involving B cells, T_C cells, and T_H cells (Fig. 7–22).

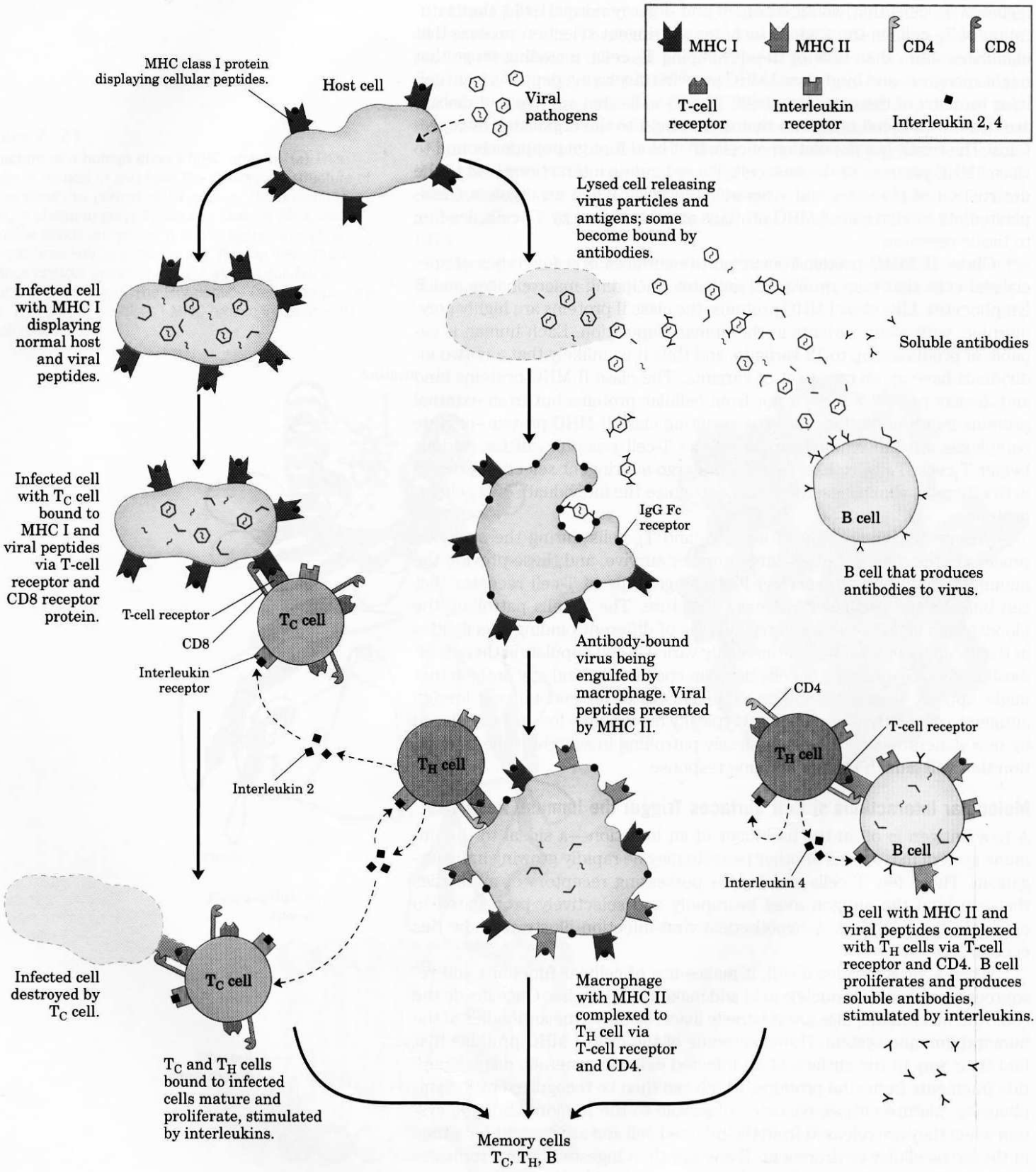


figure 7–22

Overview of the immune response to a viral infection. The individual steps are described in the text.

The class I MHC protein–peptide complexes on infected cells are recognized as foreign and bound by those T_C cells with T-cell receptors having the appropriate binding specificity. The T-cell receptors respond only to peptide antigens that are complexed to class I MHC proteins. The T_C cells have an additional receptor, **CD8,** also called a coreceptor, that enhances the binding interactions of T-cell receptors and MHC proteins (Fig. 7–22, middle left). The T_C cells live up to the name killer T cells by destroying the virally infected cell to which they are complexed through their T-cell receptors. Cell death is brought about by a number of mechanisms, not all well understood. One mechanism involves the release of a protein called **perforin,** which binds to and aggregates in the plasma membrane of the target cell, forming molecular pores that destroy the capacity of that cell to regulate its interior environment. T_C cells also induce a process called **programmed cell death,** or **apoptosis** (most commonly pronounced app′-a-toe′-sis), in which the cells complexed to T_C cells undergo metabolic changes that rapidly lead to the demise of the cell.

T_C cells with the proper specificity must proliferate selectively if large numbers of virus-infected cells are to be destroyed. To this end, T_C cells complexed to an infected cell generate cell-surface receptors for signaling proteins called **interleukins.** Interleukins, secreted by a variety of cells, stimulate the proliferation of only those T and B cells bearing the required interleukin receptors. Because T and B cells produce interleukin receptors only when they are complexed with an antigen, the only immune system cells that proliferate are those few that can respond to the antigen. The process of producing a population of cells by stimulated reproduction of a particular ancestor cell is called **clonal selection.**

The peptides complexed to class II MHC proteins and displayed on the surface of macrophages and B lymphocytes are similarly bound by the appropriate T-cell receptors of T_H cells. The T_H cells also have a coreceptor, called **CD4,** that enhances the binding interactions of the T-cell receptors. This overall binding interaction, in concert with secondary molecular signals that are currently being identified, activates the T_H cells. A subpopulation of activated T_H cells secrete a small signal protein called interleukin-2 (IL-2; M_r 15,000), which stimulates proliferation of nearby T_C cells and T_H cells having the appropriate interleukin receptors. This greatly increases the number of available immune system cells capable of recognizing and responding to the antigen. Another subpopulation of activated T_H cells complexed to macrophages or B lymphocytes secrete interleukin-4 (IL-4; M_r 20,000), which stimulates the proliferation of B cells that recognize the antigen (Fig. 7–22, bottom right). Proliferation of the responding B, T_C, and T_H cells continues as long as the appropriate antigen is present.

The proliferating B cells promote the destruction of any extracellular viruses or bacterial cells. They first secrete large amounts of soluble antibody that binds to the antigen. This bound antibody recruits a cellular system of about 20 proteins collectively called **complement** because they complement and enhance the action of the antibodies. The complement proteins disrupt the coats of many viruses or, in bacterial infections, produce holes in the cell walls of bacteria, causing them to swell and burst by osmotic shock.

Unlike T cells, B cells do not undergo selection in the thymus to eliminate those producing antibodies that recognize host (self) proteins. However, B cells do not contribute significantly to an immune response unless they are stimulated to proliferate by T_H cells. The T_H cells *do* undergo selection in the thymus, leaving no T_H cells capable of stimulating B cells that produce antibodies potentially dangerous to the host.

The T_H cells themselves participate only indirectly in the destruction of infected cells and pathogens, but their role is critical to the entire immune

response. This is dramatically illustrated by the epidemic produced by HIV (human immunodeficiency virus), the virus that causes AIDS (acquired immune deficiency syndrome). The primary targets of HIV infection are T_H cells. Elimination of these cells progressively incapacitates the entire immune system.

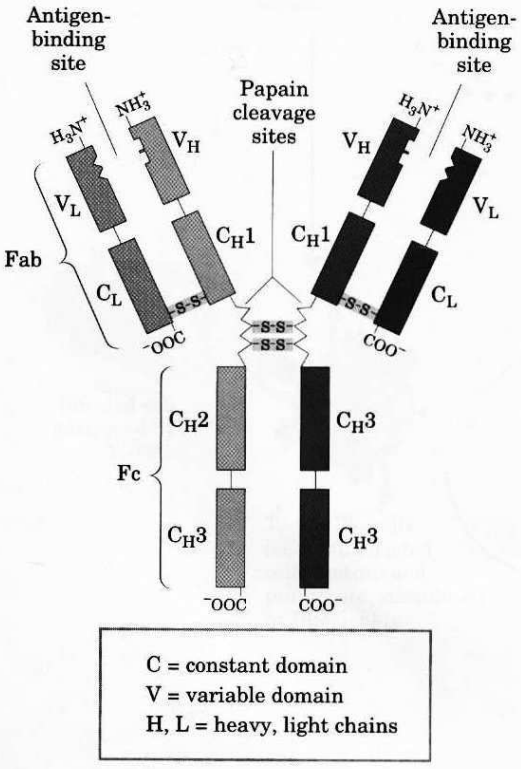
Once antigen is depleted, activated immune cells generally die in a matter of days by programmed cell death. However, a few of the stimulated B and T cells mature into **memory cells.** These are long-lived cells that do not participate directly in the primary immune response when the antigen is first encountered. Instead they become permanent residents of the blood, ready to respond to a reappearance of the same antigen. Memory cells, when subsequently challenged by the antigen, can mount a secondary immune response that is generally much more rapid and vigorous than the primary response because of prior clonal expansion. By this mechanism, vertebrates once exposed to a virus or other pathogen can respond quickly to the pathogen when exposed again. This is the basis of the long-term immunity conferred by vaccines and the natural immunity to repeated infections by the same strain of a virus.

Antibodies Have Two Identical Antigen-Binding Sites

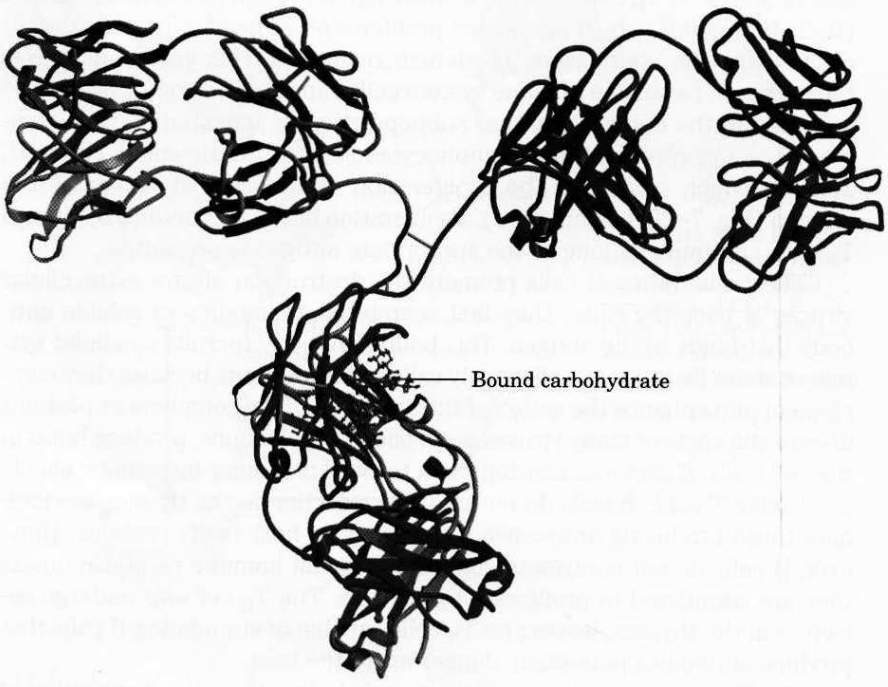
Immunoglobulin G (IgG) is the major class of antibody molecule and one of the most abundant proteins in the blood serum. IgG has four polypeptide chains: two large ones, called heavy chains, and two light chains, linked by noncovalent and disulfide bonds into a complex of M_r 150,000. The heavy chains of an IgG molecule interact at one end, then branch to interact separately with the light chains, forming a Y-shaped molecule (Fig. 7–23). At the "hinges" separating the base of an IgG molecule from its branches, the immunoglobulin can be cleaved with proteases. Cleavage with the protease papain liberates the basal fragment, called **Fc** because it usually *c*rystallizes readily, and the two branches, which are called **Fab,** the *a*ntigen-*b*inding fragments. Each branch has a single antigen-binding site.

figure 7–23

The structure of immunoglobulin G. (a) Pairs of heavy and light chains combine to form a Y-shaped molecule. Two antigen-binding sites are formed by the combination of variable domains from one light (V_L) and one heavy (V_H) chain. Cleavage with papain separates the Fab and Fc portions of the protein in the hinge region. The Fc portion of the molecule also contains bound carbohydrate. **(b)** A ribbon model of the first complete IgG molecule to be crystallized and structurally analyzed. Although the molecule contains two identical heavy chains (two shades of blue) and two identical light chains (two shades of red), it crystallized in the asymmetric conformation shown. Conformational flexibility may be important to the function of immunoglobulins.



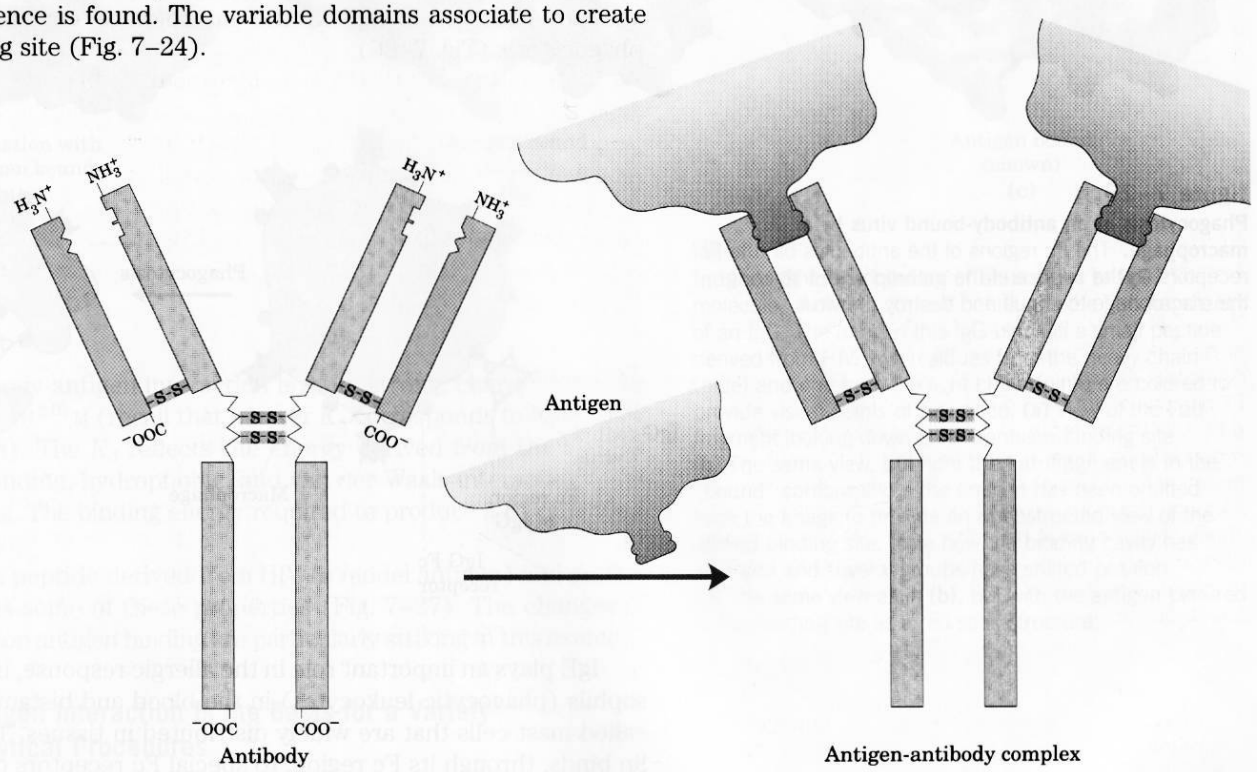
(a)



(b)

The fundamental structure of immunoglobulins was first established by Gerald Edelman and Rodney Porter. Each chain is made up of identifiable domains; some are constant in sequence and structure from one IgG to the next, others are variable. The constant domains have a characteristic structure known as the **immunoglobulin fold,** a well-conserved structural motif in the all-β class. There are three of these constant domains in each heavy chain and one in each light chain. The heavy and light chains also have one variable domain each, in which most of the variability in amino acid residue sequence is found. The variable domains associate to create the antigen-binding site (Fig. 7–24).

figure 7–24
Binding of IgG to an antigen. To generate an optimal fit for the antigen, the binding sites of IgG often undergo slight conformational changes. Such induced fit is common to many protein-ligand interactions.



In many vertebrates, IgG is only one of five classes of immunoglobulins. Each class has a characteristic type of heavy chain, denoted α, δ, ϵ, γ, and μ for IgA, IgD, IgE, IgG, and IgM, respectively. Two types of light chain, κ and λ, occur in all classes of immunoglobulins. The overall structures of **IgD** and **IgE** are similar to that of IgG. **IgM** occurs in either a monomeric, membrane-bound form or a secreted form that is a cross-linked pentamer of this basic structure (Fig. 7–25). **IgA,** found principally in secretions such as saliva, tears, and milk, can be a monomer, dimer, or trimer. IgM is the first antibody to be made by B lymphocytes and is the major antibody in the early stages of a primary immune response. Some B cells soon begin to produce IgD (with the same antigen-binding site as the IgM produced by the same cell), but the unique function of IgD is less clear.

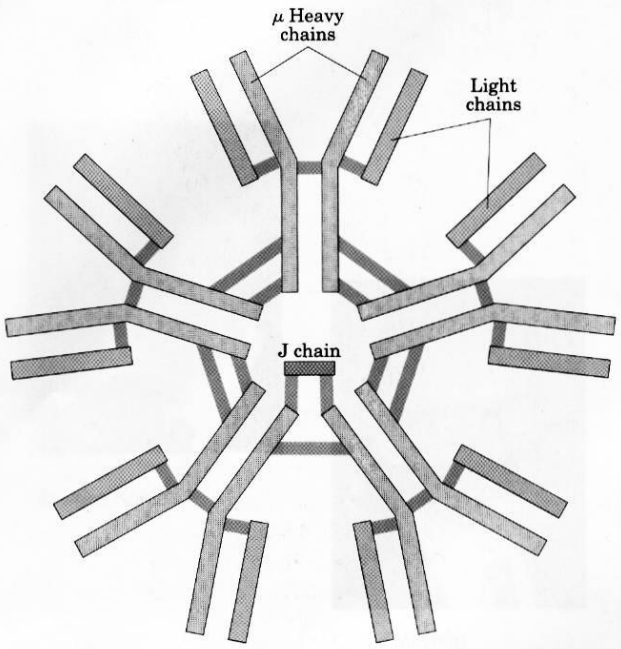


figure 7–25
IgM pentamer of immunoglobulin units. The pentamer is cross-linked with disulfide bonds. The J chain is a polypeptide of M_r 20,000 found in both IgA and IgM.

The IgG described above is the major antibody in secondary immune responses, which are initiated by memory B cells. As part of the organism's ongoing immunity to antigens already encountered and dealt with, IgG is the most abundant immunoglobulin in the blood. When IgG binds to an invading bacterium or virus, it not only activates the complement system, but also activates certain leukocytes such as macrophages to engulf and destroy the invader. Yet another class of receptors on the cell surface of macrophages recognizes and binds the Fc region of IgG. When these Fc receptors bind an antibody-pathogen complex, the macrophage engulfs the complex by phagocytosis (Fig. 7–26).

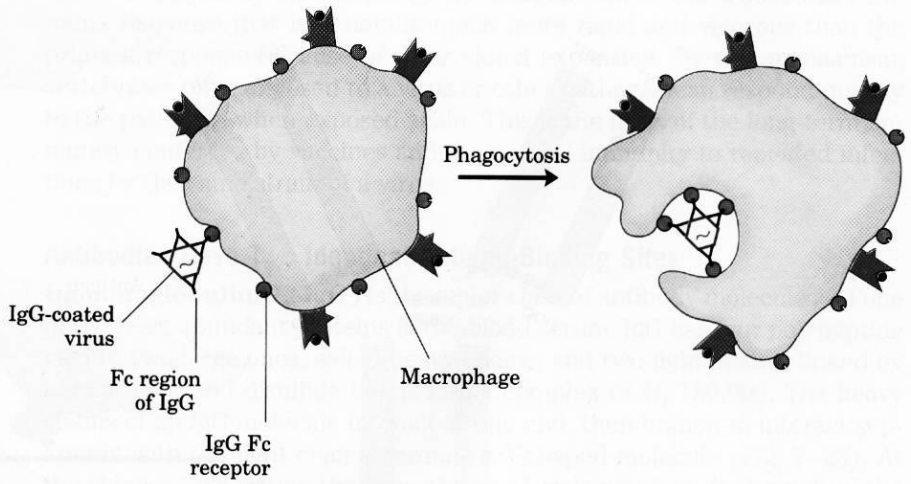


figure 7–26
Phagocytosis of an antibody-bound virus by a macrophage. The Fc regions of the antibodies bind to Fc receptors on the surface of the macrophage, triggering the macrophage to engulf and destroy the virus.

IgE plays an important role in the allergic response, interacting with basophils (phagocytic leukocytes) in the blood and histamine-secreting cells called mast cells that are widely distributed in tissues. This immunoglobulin binds, through its Fc region, to special Fc receptors on the basophils or mast cells. In this form, IgE serves as a kind of receptor for antigen. If antigen is bound, the cells are induced to secrete histamine and other biologically active amines that cause dilation and increased permeability of blood vessels. These effects on the blood vessels are thought to facilitate the movement of immune system cells and proteins to sites of inflammation. They also produce the symptoms normally associated with allergies. Pollen or other allergens are recognized as foreign, triggering an immune response normally reserved for pathogens.

Antibodies Bind Tightly and Specifically to Antigen

The binding specificity of an antibody is determined by the amino acid residues in the variable domains of its heavy and light chains. Many residues in these domains are variable, but not equally so. Some, particularly those lining the antigen-binding site, are hypervariable—especially likely to differ. Specificity is conferred by chemical complementarity between the antigen and its specific binding site, in terms of shape and the location of charged, nonpolar, and hydrogen-bonding groups. For example, a binding site with a negatively charged group may bind an antigen with a positive charge in the complementary position. In many instances, complementarity is achieved interactively as the structures of antigen and binding site are influenced by each other during the approach of the ligand. Conformational changes in the antibody and/or the antigen then occur that allow the complementary groups to interact fully. This is an example of induced fit (Fig. 7–27).

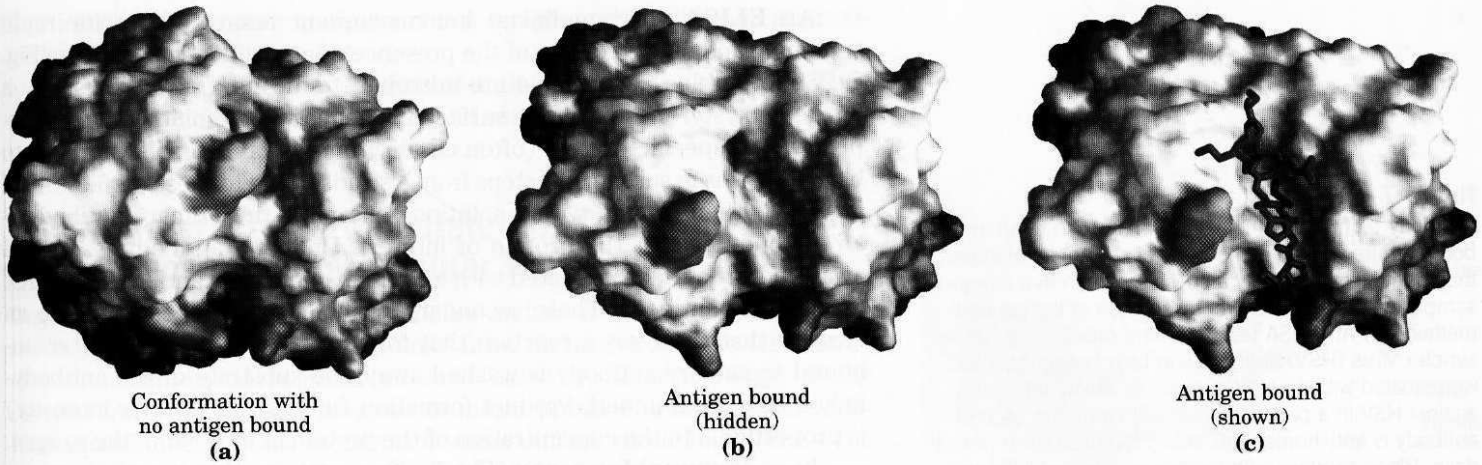


figure 7–27
Induced fit in the binding of an antigen to IgG. The molecule, shown in surface contour, is the Fab fragment of an IgG. The antigen this IgG binds is a small peptide derived from HIV. Two residues from the heavy chain (blue) and one from the light chain (pink) are colored to provide visual points of reference. **(a)** View of the Fab fragment looking down on the antigen-binding site. **(b)** The same view, but here the Fab fragment is in the "bound" conformation; the antigen has been omitted from the image to provide an unobstructed view of the altered binding site. Note how the binding cavity has enlarged and several groups have shifted position. **(c)** The same view as in **(b),** but with the antigen pictured in the binding site as a red stick structure.

A typical antibody-antigen interaction is quite strong, characterized by K_d values as low as 10^{-10} M (recall that a lower K_d corresponds to a stronger binding interaction). The K_d reflects the energy derived from the various ionic, hydrogen-bonding, hydrophobic, and van der Waals interactions that stabilize the binding. The binding energy required to produce a K_d of 10^{-10} M is about 65 kJ/mol.

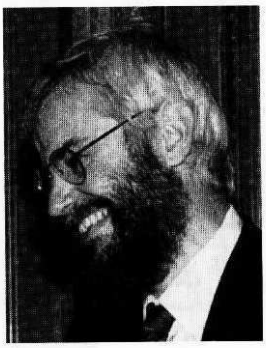
A complex of a peptide derived from HIV (a model antigen) and an Fab molecule illustrates some of these properties (Fig. 7–27). The changes in structure observed on antigen binding are particularly striking in this example.

The Antibody-Antigen Interaction Is the Basis for a Variety of Important Analytical Procedures

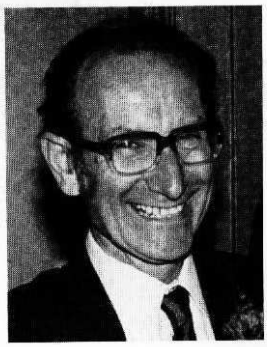
The extraordinary binding affinity and specificity of antibodies makes them valuable analytical reagents. Two types of antibody preparations are in use: polyclonal and monoclonal. **Polyclonal antibodies** are those produced by many different B lymphocytes responding to one antigen, such as a protein injected into an animal. Cells in the population of B lymphocytes produce antibodies that bind specific, different epitopes within the antigen. Thus, polyclonal preparations contain a mixture of antibodies that recognize different parts of the protein. **Monoclonal antibodies,** in contrast, are synthesized by a population of identical B cells (a **clone**) grown in cell culture. These antibodies are homogeneous, all recognizing the same epitope. The techniques for producing monoclonal antibodies were developed by Georges Köhler and Cesar Milstein.

The specificity of antibodies has practical uses. A selected antibody can be covalently attached to a resin and used in a chromatography column of the type shown in Figure 5–18c. When a mixture of proteins is added to the column, the antibody will specifically bind its target protein and retain it on the column while other proteins are washed through. The target protein can then be eluted from the resin by a salt solution or some other agent. This is a powerful tool for protein purification.

In another versatile analytical technique, an antibody is attached to a radioactive label or some other reagent that makes it easy to detect. When the antibody binds the target protein, the label reveals the presence of the protein in a solution or its location in a gel or even a living cell. Several variations of this procedure are illustrated in Figure 7–28.



Georges Köhler



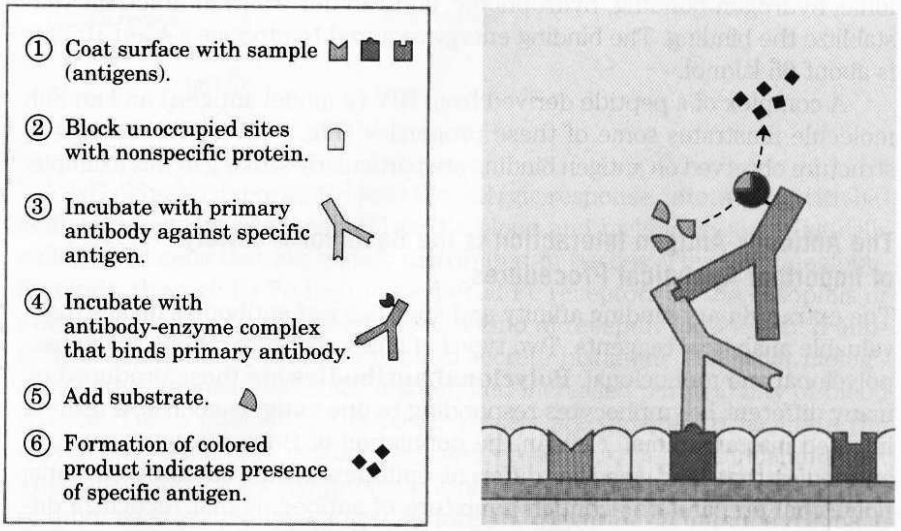
Cesar Milstein

An **ELISA** (enzyme-linked immunosorbent assay) allows for rapid screening and quantification of the presence of an antigen in a sample (Fig. 7–28b). Proteins in a sample are adsorbed to an inert surface, usually a 96-well polystyrene plate. The surface is washed with a solution of an inexpensive nonspecific protein (often casein from nonfat dry milk powder) to block proteins in subsequent steps from also adsorbing to these surfaces. The surface is then treated with a solution containing the primary antibody—an antibody against the protein of interest. Unbound antibody is washed away and the surface is treated with a solution containing antibodies against the primary antibody. These secondary antibodies have been linked to an enzyme that catalyzes a reaction that forms a colored product. After unbound secondary antibody is washed away, the substrate of the antibody-linked enzyme is added. Product formation (monitored as color intensity) is proportional to the concentration of the protein of interest in the sample.

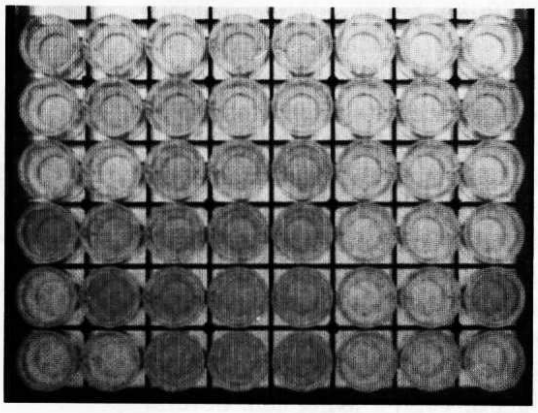
In an **immunoblot assay** (Fig. 7–28c), proteins that have been separated by gel electrophoresis are transferred electrophoretically to a nitrocellulose membrane. The membrane is blocked (as described above for ELISA), then treated successively with primary antibody, secondary antibody linked to enzyme, and substrate. A colored precipitate forms only along the band containing the protein of interest. The immunoblot allows

figure 7–28

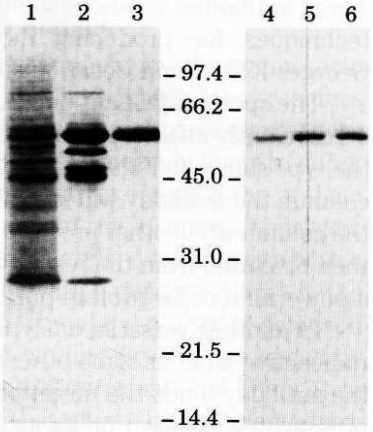
Antibody techniques. The specific reaction of an antibody with its antigen is the basis of several techniques that identify and quantify a specific protein in a complex sample. **(a)** A schematic representation of the general method. **(b)** An ELISA testing for the presence of herpes simplex virus (HSV) antibodies in blood samples. Wells were coated with an HSV antigen, to which antibodies against HSV in a patient's blood will bind. The second antibody is anti–human IgG linked to horseradish peroxidase. Blood samples with greater amounts of HSV antibody turn brighter yellow. **(c)** An immunoblot. Lanes 1 to 3 are from an SDS gel; samples from successive stages in the purification of a protein kinase have been separated and stained with Coomassie blue. Lanes 4 to 6 show the same samples, but these were electrophoretically transferred to a nitrocellulose membrane after separation on an SDS gel. The membrane was then "probed" with antibody against the protein kinase. The numbers between the gel and the immunoblot indicate M_r ($\times 10^{-3}$).



(a)



ELISA assay
(b)



SDS gel Immunoblot
(c)

the detection of a minor component in a sample and provides an approximation of its molecular weight.

We will encounter other aspects of antibodies in later chapters. They are extremely important in medicine and can tell us much about the structure of proteins and the action of genes.

Protein Interactions Modulated by Chemical Energy: Actin, Myosin, and Molecular Motors

Organisms move. Cells move. Organelles and macromolecules within cells move. Most of these movements arise from the activity of the fascinating class of protein-based molecular motors. Fueled by chemical energy, usually derived from ATP, large aggregates of motor proteins undergo cyclic conformational changes that accumulate into a unified, directional force—the tiny force that pulls apart chromosomes in a dividing cell, and the immense force that levers a pouncing, quarter-ton jungle cat into the air.

The interactions among motor proteins, as you might predict, feature complementary arrangements of ionic, hydrogen-bonding, hydrophobic, and van der Waals interactions at protein binding sites. In motor proteins, however, these interactions achieve exceptionally high levels of spatial and temporal organization.

Motor proteins underlie the contraction of muscles, the migration of organelles along microtubules, the rotation of bacterial flagella, and the movement of some proteins along DNA. As we noted in Chapter 2, proteins called kinesins and dyneins move along microtubules in cells, pulling along organelles or reorganizing chromosomes during cell division (see Fig. 2–19). An interaction of dynein with microtubules brings about the motion of eukaryotic flagella and cilia. Flagellar motion in bacteria involves a complex rotational motor at the base of the flagellum (see Fig. 19–32). Helicases, polymerases, and other proteins move along DNA as they carry out their functions in DNA metabolism (Chapter 25). Here, we focus on the well-studied example of the contractile proteins of vertebrate skeletal muscle as a paradigm for how proteins translate chemical energy into motion.

The Major Proteins of Muscle Are Myosin and Actin

The contractile force of muscle is generated by the interaction of two proteins, myosin and actin. These proteins are arranged in filaments that undergo transient interactions and slide past each other to bring about contraction. Together, actin and myosin make up over 80% of the protein mass of muscle.

Myosin (M_r 540,000) has six subunits: two heavy chains (M_r 220,000) and four light chains (M_r 20,000). The heavy chains account for much of the overall structure. At their carboxyl termini, they are arranged as extended α helices, wrapped around each other in a fibrous, left-handed coiled coil similar to that of α-keratin (Fig. 7–29a). At its amino termini, each heavy chain has a large globular domain containing a site where ATP is hydrolyzed. The light chains are associated with the globular domains.

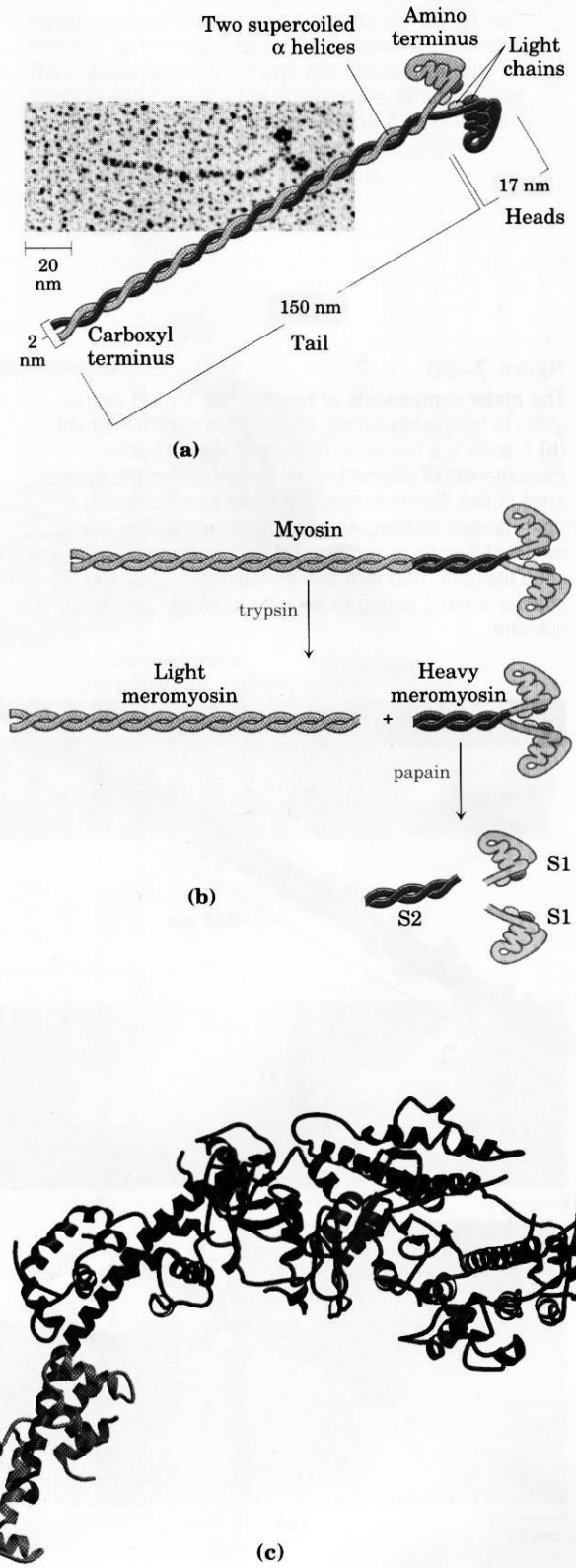


figure 7–29
Myosin. (a) Myosin has two heavy chains (in two shades of pink), the carboxyl termini forming an extended coiled coil (tail) and the amino termini having globular domains (heads). Two light chains (blue) are associated with each myosin head. **(b)** Cleavage with trypsin and papain separates the myosin heads (S1 fragments) from the tails. **(c)** Ribbon representation of the myosin S1 fragment. The heavy chain is in gray, the two light chains in two shades of blue.

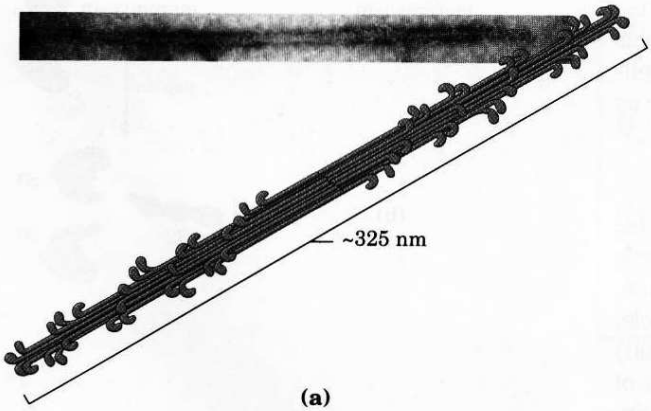
When myosin is treated briefly with the protease trypsin, much of the fibrous tail is cleaved off, dividing the protein into components called light and heavy meromyosin (Fig. 7–29b). The globular domain, called myosin subfragment 1, or S1, or simply the myosin head group, is liberated from heavy meromyosin by cleavage with papain. The S1 fragment produced by this procedure is the motor domain that makes muscle contraction possible. S1 fragments can be crystallized, and their structure has been determined. The overall structure of the S1 fragment as determined by Ivan Rayment and Hazel Holden is shown in Figure 7–29c.

In muscle cells, molecules of myosin aggregate to form structures called **thick filaments** (Fig. 7–30a). These rodlike structures serve as the core of the contractile unit. Within a thick filament, several hundred myosin molecules are arranged with their fibrous "tails" associated to form a long bipolar structure. The globular domains project from either end of this structure, in regular stacked arrays.

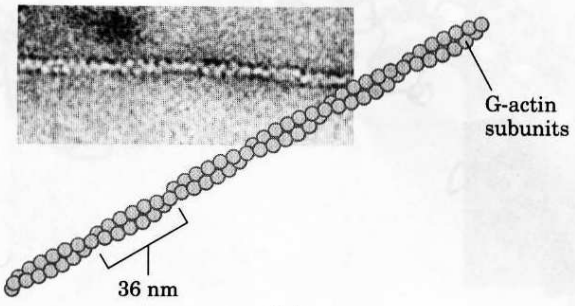
The second major muscle protein, **actin,** is abundant in almost all eukaryotic cells. In muscle, molecules of monomeric actin, called G-actin (*g*lobular actin; M_r 42,000), associate to form a long polymer called F-actin (*f*ilamentous actin). The **thin filament** (Fig. 7–30b) consists of F-actin, along with the proteins troponin and tropomyosin. The filamentous parts of thin filaments assemble as successive monomeric actin molecules add to one end. On addition, each monomer binds ATP, then hydrolyzes it to ADP, so all actin molecules in the filament are complexed to ADP. However, this ATP hydrolysis by actin functions only in the assembly of the filaments; it does not contribute directly to the energy expended in muscle contraction. Each actin monomer in the thin filament can bind tightly and specifically to one myosin head group (Fig. 7–30c).

figure 7–30

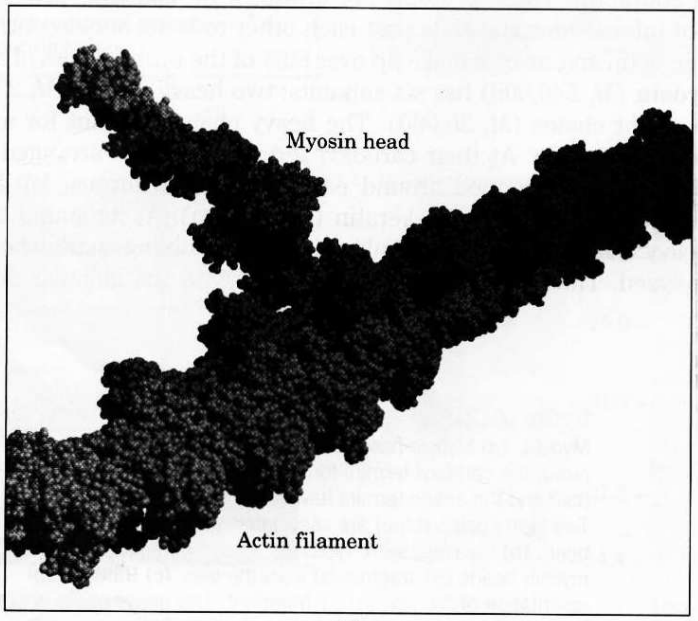
The major components of muscle. (a) Myosin aggregates to form a bipolar structure called a thick filament. **(b)** F-actin is a filamentous assemblage of G-actin monomers that polymerize two by two, giving the appearance of two filaments spiraling about one another in a right-handed fashion. An electron micrograph and a model of F-actin are shown. **(c)** Space-filling model of an actin filament (red) with one myosin head (gray and two shades of blue) bound to an actin monomer within the filament.



(a)



(b)



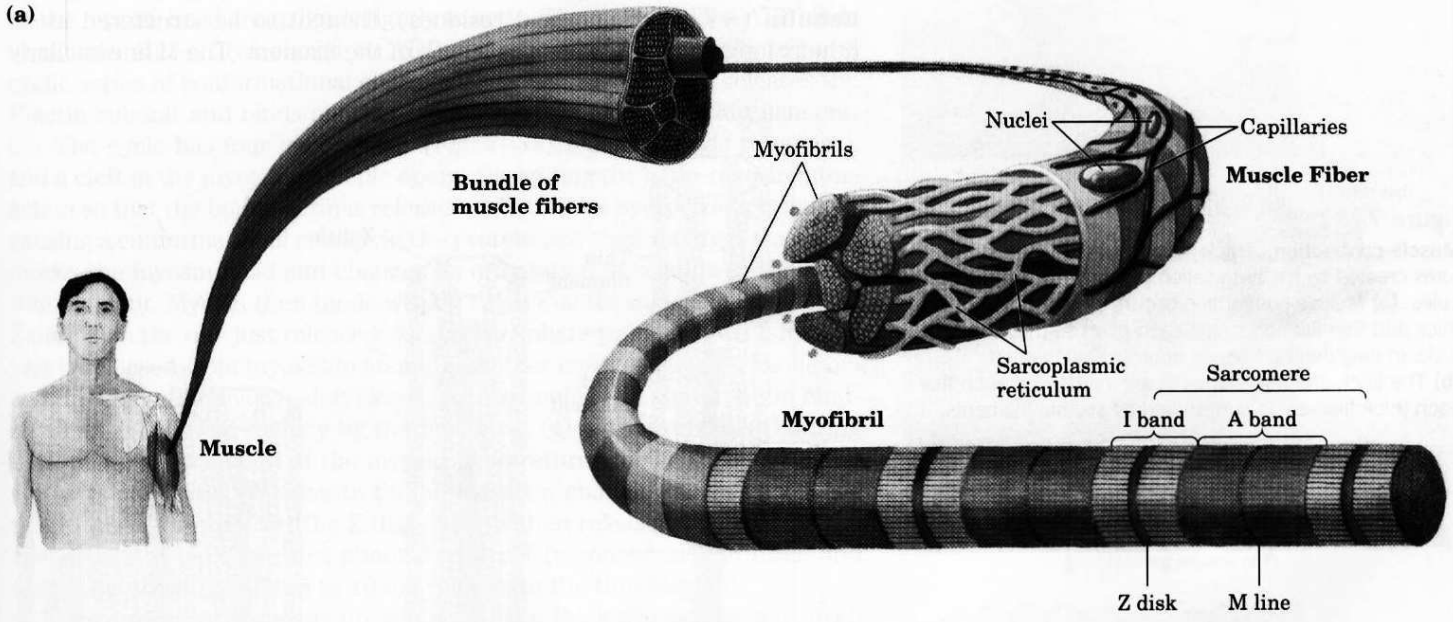
(c)

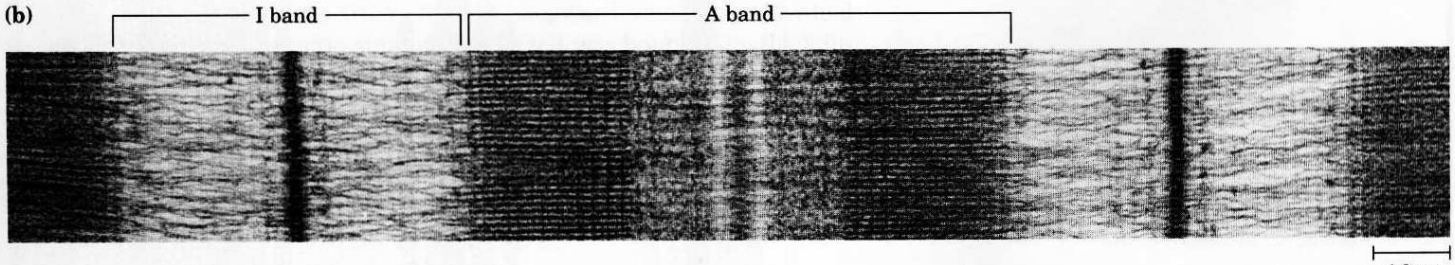
Additional Proteins Organize the Thin and Thick Filaments into Ordered Structures

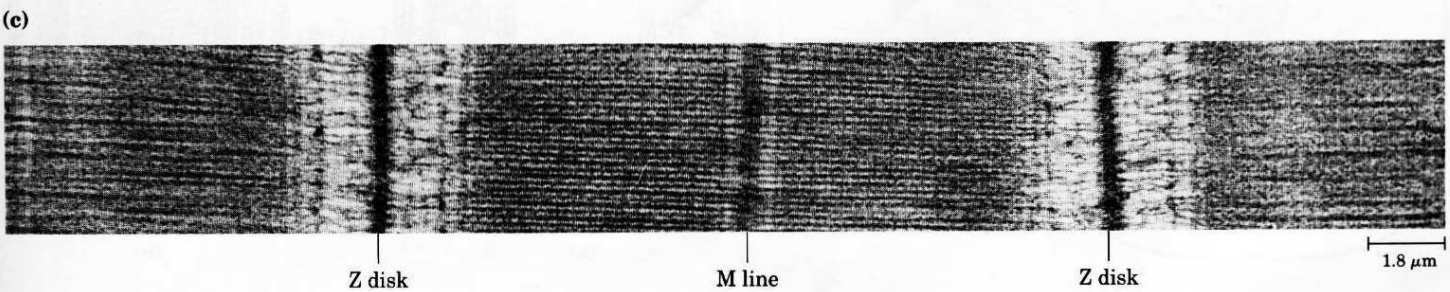
Skeletal muscle consists of parallel bundles of **muscle fibers,** each fiber a single, very large, multinucleated cell, 20 to 100 μm in diameter, formed from many cells fused together and often spanning the length of the muscle. Each fiber, in turn, contains about 1,000 **myofibrils,** 2 μm in diameter, each consisting of a vast number of regularly arrayed thick and thin filaments complexed to other proteins (Fig. 7–31). A system of flat membranous vesicles called the **sarcoplasmic reticulum** surrounds each myofibril. Examined under the electron microscope, muscle fibers reveal alternating regions of high and low electron density, called the **A** and **I bands** (Fig. 7–31b,c). The A and I bands arise from the arrangement of

figure 7–31

Structure of skeletal muscle. (a) Muscle fibers consist of single, elongated, multinucleated cells that arise from the fusion of many precursor cells. Within the fibers are many myofibrils (only six are shown here for simplicity) surrounded by the membranous sarcoplasmic reticulum. The organization of thick and thin filaments in the myofibril gives it a striated appearance. When muscle contracts, the I bands narrow and the Z disks come closer together, as seen in electron micrographs of relaxed **(b)** and contracted **(c)** muscle.







thick and thin filaments, which are aligned and partially overlapping. The I band is the region of the bundle that in cross section would contain only thin filaments. The darker A band stretches the length of the thick filament and includes the region where parallel thick and thin filaments overlap. Bisecting the I band is a thin structure called the **Z disk,** perpendicular to the thin filaments and serving as an anchor to which the thin filaments are attached. The A band too is bisected by a thin line, the **M line** or M disk, a region of high electron density in the middle of the thick filaments. The entire contractile unit, consisting of bundles of thick filaments interleaved at either end with bundles of thin filaments, is called the **sarcomere.** The arrangement of interleaved bundles allows the thick and thin filaments to slide past each other (by a mechanism discussed below), causing a progressive shortening of each sarcomere (Fig. 7–32).

The thin actin filaments are attached at one end to the Z disk in a regular pattern. The assembly includes the minor muscle proteins **α-actinin, desmin,** and **vimentin.** Thin filaments also contain a large protein called **nebulin** (~7,000 amino acid residues), thought to be structured as an α helix long enough to span the length of the filament. The M line similarly

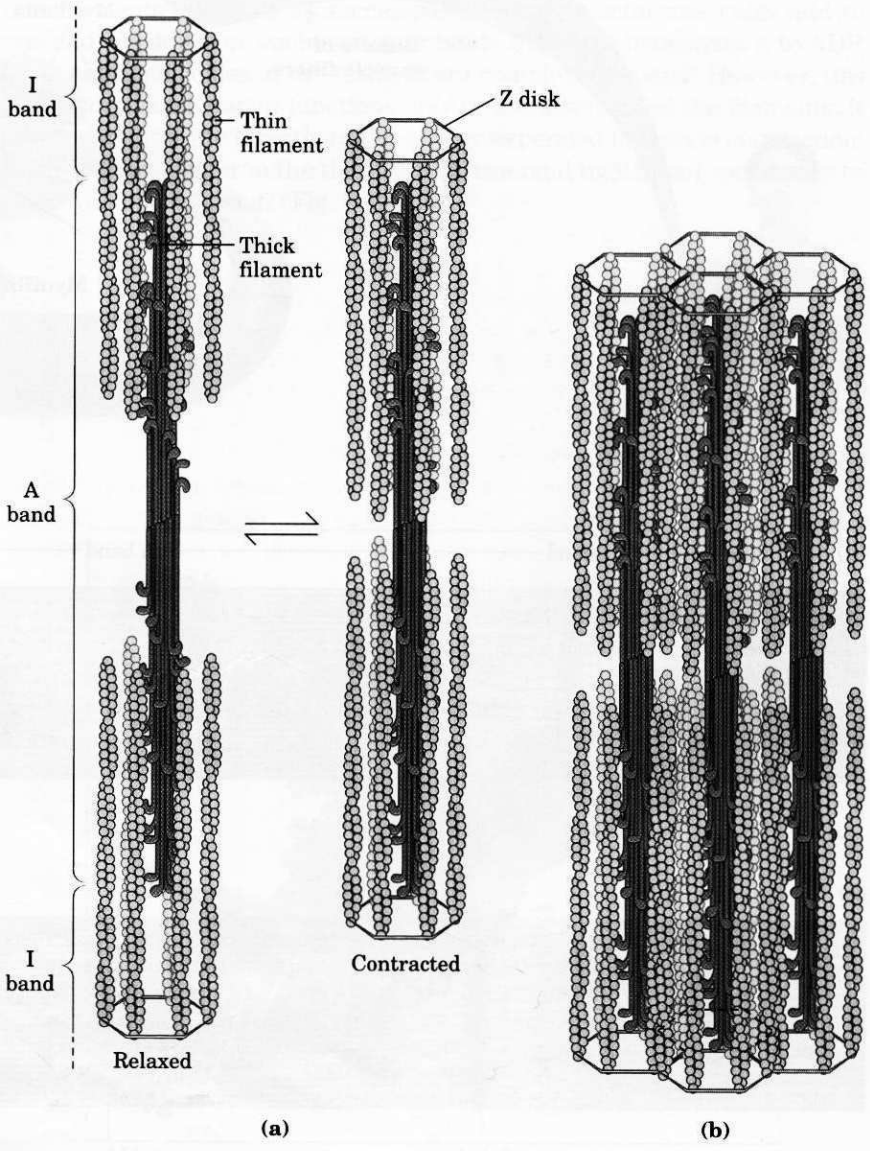


figure 7–32
Muscle contraction. Thick filaments are bipolar structures created by the association of many myosin molecules. **(a)** Muscle contraction occurs by the sliding of the thick and thin filaments past each other so that the Z disks in neighboring I bands approach each other. **(b)** The thick and thin filaments are interleaved such that each thick filament is surrounded by six thin filaments.

organizes the thick filaments. It contains the proteins **paramyosin, C-protein,** and **M-protein.** Another class of proteins called **titins,** the largest known single polypeptide chains (the titin of human cardiac muscle has 26,926 amino acid residues), link the thick filaments to the Z disk, providing additional organization to the overall structure. Among their structural functions, the proteins nebulin and titin are believed to act as "molecular rulers," regulating the length of the thin and thick filaments, respectively. Titin extends from the Z disk to the M line, regulating the length of the sarcomere itself and preventing overextension of the muscle. The characteristic sarcomere length varies from one muscle tissue to the next in a vertebrate organism, attributed in large part to the expression of different titin variants.

Myosin Thick Filaments Slide along Actin Thin Filaments

The interaction between actin and myosin, like that between all proteins and ligands, involves weak bonds. When ATP is not bound to myosin, a face on the myosin head group binds tightly to actin (Fig. 7–33). When ATP binds to myosin and is hydrolyzed to ADP and phosphate, a coordinated and cyclic series of conformational changes occur in which myosin releases the F-actin subunit and binds another subunit farther along the thin filament.

The cycle has four major steps (Fig. 7–33). ① ATP binds to myosin, and a cleft in the myosin molecule opens, disrupting the actin-myosin interaction so that the bound actin is released. ATP is then hydrolyzed (step ②), causing a conformational change in the protein to a "high-energy" state that moves the myosin head and changes its orientation in relation to the actin thin filament. Myosin then binds weakly to an F-actin subunit closer to the Z disk than the one just released. As the phosphate product of ATP hydrolysis is released from myosin in step ③, another conformational change occurs in which the myosin cleft closes, strengthening the myosin-actin binding. This is followed quickly by the final step, ④, a "power stroke" during which the conformation of the myosin head returns to the original resting state, its orientation relative to the bound actin changing so as to pull the tail of the myosin toward the Z disk. ADP is then released to complete the cycle. Each cycle generates about 3 to 4 pN (piconewtons) of force and moves the thick filament 5 to 10 nm relative to the thin filament.

Because there are many myosin heads in a thick filament, at any given moment some (probably 1% to 3%) are bound to the thin filaments. This prevents the thick filaments from slipping backward when an individual myosin head releases the actin subunit to which it was bound. The thick filament thus actively slides forward past the adjacent thin filaments. This process, coordinated among the many sarcomeres in a muscle fiber, brings about muscle contraction.

The interaction between actin and myosin must be regulated so that contraction occurs only in response to appropriate signals from the nervous

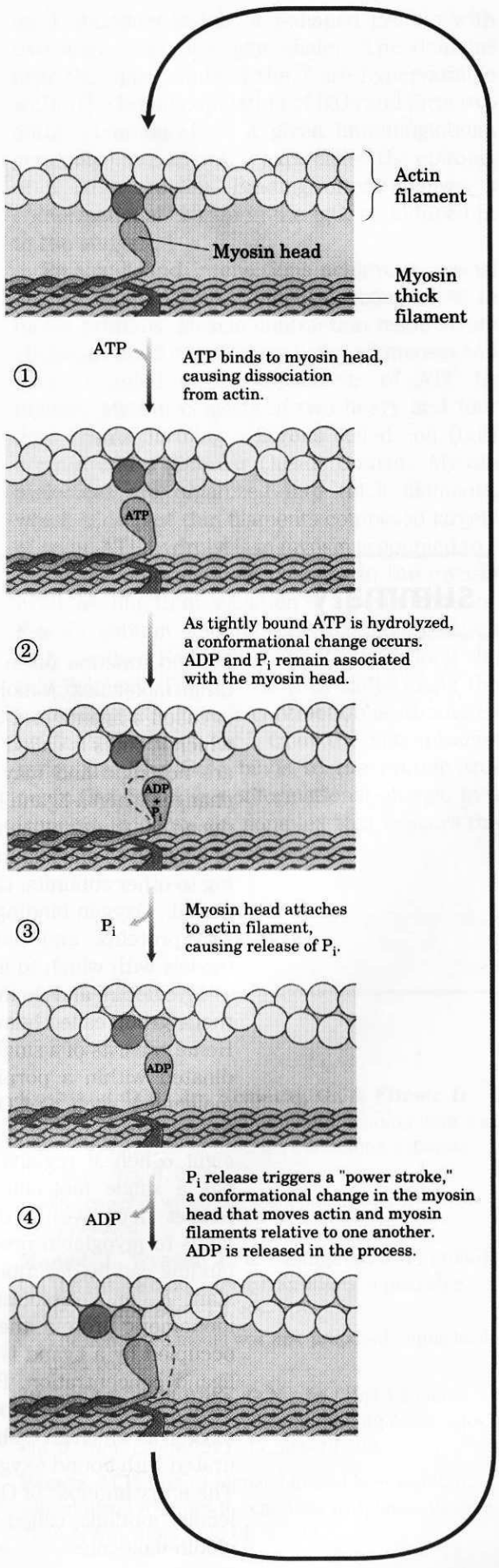


figure 7–33
Molecular mechanism of muscle contraction. Conformational changes in the myosin head that are coupled to stages in the ATP hydrolytic cycle cause myosin to successively dissociate from one actin subunit, then associate with another farther along the actin filament. In this way the myosin heads slide along the thin filaments, drawing the thick filament array into the thin filament array (see Fig. 7–32).

system. The regulation is mediated by a complex of two proteins, **tropomyosin** and **troponin.** Tropomyosin binds to the thin filament, blocking the attachment sites for the myosin head groups. Troponin is a Ca^{2+}-binding protein. A nerve impulse causes release of Ca^{2+} from the sarcoplasmic reticulum. The released Ca^{2+} binds to troponin (another protein-ligand interaction) and causes a conformational change in the tropomyosin-troponin complexes, exposing the myosin-binding sites on the thin filaments. Contraction follows.

Working skeletal muscle requires two types of molecular functions that are common in proteins—binding and catalysis. The actin-myosin interaction, a protein-ligand interaction like that of immunoglobulins with antigens, is reversible and leaves the participants unchanged. When ATP binds myosin, however, it is hydrolyzed to ADP and P_i. Myosin is not only an actin-binding protein, it is also an ATPase—an enzyme. The function of enzymes in catalyzing chemical transformations is the topic of the next chapter.

summary

Protein function often entails interactions with other molecules. A molecule bound by a protein is called a ligand, and the site on the protein to which it binds is called the binding site. Proteins are not rigid and may undergo conformational changes when a ligand binds, a process called induced fit. In a multisubunit protein, the binding of a ligand to one subunit may affect ligand binding to other subunits. Ligand binding can be regulated. Oxygen-binding proteins, immune system proteins, and motor proteins are useful models with which to illustrate these principles.

Myoglobin and hemoglobin contain a prosthetic group called heme to which oxygen binds. Heme consists of a single atom of Fe^{2+} iron coordinated within a porphyrin. Some other small molecules, such as CO and NO, can also bind heme. Myoglobin is a single polypeptide with eight α-helical regions connected by bends. It has a single molecule of heme, located in a pocket deep within the polypeptide. Oxygen binds to myoglobin reversibly. Simple reversible binding can be described by an association constant K_a or a dissociation constant K_d. For a monomeric protein, the fraction of binding sites occupied by a ligand is a hyperbolic function of ligand concentration. Because O_2 is a gas, the term P_{50}, which describes the partial pressure of oxygen at which an O_2-binding protein is half saturated with bound oxygen, is used in place of K_d. The entry and exit of O_2 depend upon small molecular motions, called "breathing," of the myoglobin molecule.

Normal adult hemoglobin has four heme-containing subunits, two α and two β. They are similar in structure to each other and to myoglobin. Strong interactions occur between unlike (α and β) subunits. Hemoglobin exists in two interchangeable states, called T and R. The T state is stabilized by several salt bridges and is most stable when oxygen is not bound. Oxygen binding promotes a transition to the R state.

Oxygen binding to hemoglobin is both allosteric and cooperative. Binding of O_2 to one binding site of hemoglobin affects binding of O_2 to other such sites, an example of allosteric binding behavior. Conformational changes between the T and R states, mediated by subunit-subunit interactions, give rise to a form of allostery called cooperative binding. Cooperative binding results in a sigmoid binding curve and can be analyzed by a Hill plot. Two major models have been proposed to explain the cooperative binding of ligands to multisubunit proteins. In the concerted model, all subunits are in the same conformation at any given time, and the entire protein undergoes a reversible transition between two possible conformations. Successive binding of ligand molecules to the low-affinity conformation facilitates transition to the high-affinity conformation. In the sequential model, individual subunits can undergo conformational changes. Binding of a ligand to one subunit alters that subunit's conformation, facilitating similar changes in, and binding of additional ligands to, the other subunits.

Hemoglobin also binds H^+ and CO_2. In both cases, binding results in the formation of ion pairs that stabilize the T state and O_2 binding is weakened, a phenomenon called the Bohr effect. The binding of H^+ and CO_2 to hemoglobin in the tissues promotes the release of O_2, and the binding of O_2 to hemoglobin in the lungs promotes the release of H^+ and CO_2. Oxygen binding to hemoglobin is also modulated by 2,3-bisphosphoglycerate, which binds to and stabilizes the T state.

Sickle-cell anemia is a genetic disease caused by a single amino acid substitution (Glu to Val) at position 6 in the β chains of hemoglobin. The change produces a hydrophobic patch on the surface of the protein that causes the hemoglobin molecules to aggregate into bundles of fibers. These bundles give the erythrocytes a sickle shape. This homozygous condition results in serious medical complications.

The immune response is mediated by interactions among an array of specialized leukocytes and their associated proteins. T lymphocytes produce T-cell receptors. B lymphocytes produce immunoglobulins. All cells produce MHC proteins, which display host (self) or antigenic (nonself) peptides on the cell surface. Helper T cells induce the proliferation of those B cells and cytotoxic T cells producing immunoglobulins or T-cell receptors that bind to a specific antigen, a process called clonal selection.

Humans have five classes of immunoglobulins, each with different biological functions. The most abundant is IgG, a Y-shaped protein with two heavy and two light chains. The domains near the upper ends of the Y are hypervariable within the broad population of IgGs and form two antigen-binding sites. A given immunoglobulin generally binds to only a part, called the epitope, of a large antigen. Binding often involves a conformational change in the IgG, an induced fit to the antigen.

Protein-ligand interactions achieve a special degree of spatial and temporal organization in motor proteins. Muscle contraction results from choreographed interactions between myosin and actin, coupled to the hydrolysis of ATP by myosin. Myosin consists of two heavy and four light chains, forming a fibrous coiled coil (tail) domain and a globular (head) domain. Myosin molecules are organized into thick filaments, which slide past thin filaments composed largely of actin. ATP hydrolysis in myosin is coupled to a series of conformational changes in the myosin head, leading to dissociation of myosin from one F-actin subunit and its eventual reassociation with another F-actin subunit farther along the thin filament. The myosin thus slides along the actin filaments. Muscle contraction is stimulated by the release of Ca^{2+} from the sarcoplasmic reticulum. The Ca^{2+} binds to the protein troponin, leading to a conformational change in a troponin-tropomyosin complex that triggers the cycle of actin-myosin interactions.

further reading

Oxygen-Binding Proteins

Ackers, G.K. & Hazzard, J.H. (1993) Transduction of binding energy into hemoglobin cooperativity. *Trends Biochem. Sci.* **18,** 385–390.

Changeux, J.-P. (1993) Allosteric proteins: from regulatory enzymes to receptors—personal recollections. *Bioessays* **15,** 625–634.

An interesting perspective from a leader in the field.

Dickerson, R.E. & Geis, I. (1982) *Hemoglobin: Structure, Function, Evolution, and Pathology,* The Benjamin/Cummings Publishing Company, Redwood City, CA.

di Prisco, G., Condò, S.G., Tamburrini, M., & Giardina, B. (1991) Oxygen transport in extreme environments. *Trends Biochem. Sci.* **16,** 471–474.

A revealing comparison of the oxygen-binding properties of hemoglobins from polar species.

Koshland, D.E., Jr., Nemethy, G., & Filmer, D. (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **6,** 365–385.

The paper in which the sequential model is introduced.

Monod, J., Wyman, J., & Changeux, J.-P. (1965) On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12,** 88–118.

The concerted model was first proposed in this landmark paper.

Olson, J.S. & Phillips, G.N., Jr. (1996) Kinetic pathways and barriers for ligand binding to myoglobin. *J. Biol. Chem.* **271,** 17,593–17,596.

Perutz, M.F. (1989) Myoglobin and haemoglobin: role of distal residues in reactions with haem ligands. *Trends Biochem. Sci.* **14,** 42–44.

Perutz, M.F., Wilkinson, A.J., Paoli, M., & Dodson, G.G. (1998) The stereochemical mechanism of the cooperative effects in hemoglobin revisited. *Annu. Rev. Biophys. Biomol. Struct.* **27,** 1–34.

Immune System Proteins

Blom, B., Res, P.C., & Spits, H. (1998) T cell precursors in man and mice. *Crit. Rev. Immunol.* **18,** 371–388.

Cohen, I.R. (1988) The self, the world and autoimmunity. *Sci. Am.* **258** (April), 52–60.

Davies, D.R. & Chacko, S. (1993) Antibody structure. *Acc. Chem. Res.* **26,** 421–427.

Davies, D.R., Padlan, E.A., & Sheriff, S. (1990) Antibody-antigen complexes. *Annu. Rev. Biochem.* **59,** 439–473.

Davis, M.M. (1990) T cell receptor gene diversity and selection. *Annu. Rev. Biochem.* **59,** 475–496.

Dutton, R.W., Bradley, L.M., & Swain, S.L. (1998) T cell memory. *Annu. Rev. Immunol.* **16,** 201–223.

Life, Death and the Immune System. (1993) *Sci. Am.* **269** (September).

A special issue on the immune system.

Marrack, P. & Kappler, J. (1987) The T cell receptor. *Science* **238,** 1073–1079.

Müller-Eberhard, H.J. (1988) Molecular organization and function of the complement system. *Annu. Rev. Biochem.* **57,** 321–337.

Parham, P. & Ohta, T. (1996) Population biology of antigen presentation by MHC class I molecules. *Science* **272,** 67–74.

Ploegh, H.L. (1998) Viral strategies of immune evasion. *Science* **280,** 248–253.

Thomsen, A.R., Nansen, A., & Christensen, J.P. (1998) Virus-induced T cell activation and the inflammatory response. *Curr. Top. Microbiol. Immunol.* **231,** 99–123.

Van Parjis, L. & Abbas, A.K. (1998) Homeostasis and self-tolerance in the immune system: turning lymphocytes off. *Science* **280,** 243–248.

York, I.A. & Rock, K.L. (1996) Antigen processing and presentation by the class-I major histocompatibility complex. *Annu. Rev. Immunol.* **14,** 369–396.

Molecular Motors

Finer, J.T., Simmons, R.M., & Spudich, J.A. (1994) Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature* **368,** 113–119.

Modern techniques reveal the forces affecting individual motor proteins.

Geeves, M.A. & Holmes, K.C. (1999) Structural mechanism of muscle contraction. *Annu. Rev. Biochem.* **68,** 687–728.

Goldman, Y.E. (1998) Wag the tail: structural dynamics of actomyosin. *Cell* **93,** 1–4.

Huxley, H.E. (1998) Getting to grips with contraction: the interplay of structure and biochemistry. *Trends Biochem. Sci.* **23,** 84–87.

An interesting historical perspective on deciphering the mechanism of muscle contraction.

Labeit, S. & Kolmerer, B. (1995) Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science* **270,** 293–296.

A structural and functional description of some of the largest known proteins.

Rayment, I. (1996) The structural basis of the myosin ATPase activity. *J. Biol. Chem.* **271,** 15,850–15,853.

Examining mechanism from a structural perspective.

Rayment, I. & Holden, H.M. (1994) The three-dimensional structure of a molecular motor. *Trends Biochem. Sci.* **19,** 129–134.

Spudich, J.A. (1994) How molecular motors work. *Nature* **372,** 515–518.

problems

1. Relationship between Affinity and Dissociation Constant Protein A has a binding site for ligand X with a K_d of 10^{-6} M. Protein B has a binding site for ligand X with a K_d of 10^{-9} M. Which protein has a higher affinity for ligand X? Explain your reasoning. Convert the K_d to K_a for both proteins.

2. Negative Cooperativity Which of the following situations would produce a Hill plot with $n_H < 1.0$? Explain your reasoning in each case.

(a) The protein has multiple subunits, each with a single ligand-binding site. Binding of ligand to one site decreases the binding affinity of other sites for the ligand.

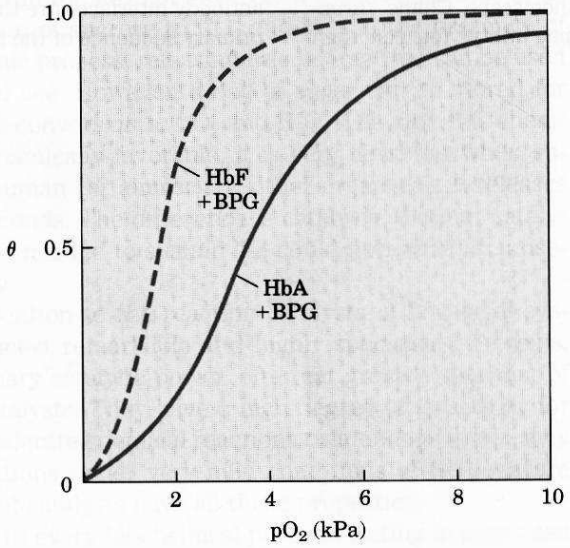
(b) The protein is a single polypeptide with two ligand-binding sites, each having a different affinity for the ligand.

(c) The protein is a single polypeptide with a single ligand-binding site. As purified, the protein preparation is heterogeneous, containing some protein molecules that are partially denatured and thus have a lower binding affinity for the ligand.

3. Affinity for Oxygen in Myoglobin and Hemoglobin What is the effect of the following changes on the O_2 affinity of myoglobin and hemoglobin? (a) A drop in the pH of blood plasma from 7.4 to 7.2. (b) A decrease in the partial pressure of CO_2 in the lungs from 6 kPa (holding one's breath) to 2 kPa (normal). (c) An increase in the BPG level from 5 mM (normal altitudes) to 8 mM (high altitudes).

4. Cooperativity in Hemoglobin Under appropriate conditions, hemoglobin dissociates into its four subunits. The isolated α subunit binds oxygen, but the O_2-saturation curve is hyperbolic rather than sigmoid. In addition, the binding of oxygen to the isolated α subunit is not affected by the presence of H^+, CO_2, or BPG. What do these observations indicate about the source of the cooperativity in hemoglobin?

5. Comparison of Fetal and Maternal Hemoglobins Studies of oxygen transport in pregnant mammals have shown that the O_2-saturation curves of fetal and maternal blood are markedly different when measured under the same conditions. Fetal erythrocytes contain a structural variant of hemoglobin, HbF, consisting of two α and two γ subunits ($\alpha_2\gamma_2$), whereas maternal erythrocytes contain HbA ($\alpha_2\beta_2$).



(a) Which hemoglobin has a higher affinity for oxygen under physiological conditions, HbA or HbF? Explain.

(b) What is the physiological significance of the different O_2 affinities?

(c) When all the BPG is carefully removed from samples of HbA and HbF, the measured O_2-saturation curves (and consequently the O_2 affinities) are displaced to the left. However, HbA now has a greater affinity for oxygen than does HbF. When BPG is reintroduced, the O_2-saturation curves return to normal, as shown in the graph. What is the effect of BPG on the O_2 affinity of hemoglobin? How can the above information be used to explain the different O_2 affinities of fetal and maternal hemoglobin?

6. Hemoglobin Variants There are almost 500 naturally occurring variants of hemoglobin. Most are the result of a single amino acid substitution in a globin polypeptide chain. Some variants produce clinical illness, though not all variants have deleterious effects. A brief sample is presented below:

HbS (sickle-cell Hb): substitutes a Val for a Glu on the surface

Hb Cowtown: eliminates an ion pair involved in T-state stabilization

Hb Memphis: substitutes one uncharged polar residue for another of similar size on the surface

Hb Bibba: substitutes a Pro for a Leu involved in an α helix

Hb Milwaukee: substitutes a Glu for a Val

Hb Providence: substitutes an Asn for a Lys that normally projects into the central cavity of the tetramer

Hb Philly: substitutes a Phe for a Tyr, disrupting hydrogen bonding at the $\alpha_1\beta_1$ interface

Explain your choices for each of the following:

(a) The Hb variant *least* likely to cause pathological symptoms.

(b) The variant(s) most likely to show pI values different from that of HbA when run on an isoelectric focusing gel.

(c) The variant(s) most likely to show a decrease in BPG binding and an increase in the overall affinity of the hemoglobin for oxygen.

7. Reversible (but Tight) Binding to an Antibody An antibody binds to an antigen with a K_d of 5×10^{-8} M. At what concentration of antigen will θ be (a) 0.2, (b) 0.5, (c) 0.6, (d) 0.8?

8. Using Antibodies to Probe Structure-Function Relationships in Proteins A monoclonal antibody binds to G-actin but not to F-actin. What does this tell you about the epitope recognized by the antibody?

9. The Immune System and Vaccines A host organism needs time, often days, to mount an immune response against a new antigen, but memory cells permit a rapid response to pathogens previously encountered. A vaccine to protect against a particular viral infection often consists of weakened or killed virus or isolated proteins from a viral protein coat. When injected into a human patient, the vaccine generally does not cause an infection and illness, but it effectively "teaches" the immune system what the viral particles look like, stimulating the production of memory cells. On subsequent infection, these cells can bind to the virus and trigger a rapid immune response. Some pathogens, including HIV, have developed mechanisms to evade the immune system, making it difficult or impossible to develop effective vaccines against them. What strategy could a pathogen use to evade the immune system? Assume that antibodies and/or T-cell receptors are available to bind to any structure that might appear on the surface of a pathogen and that, once bound, the pathogen is destroyed.

10. How We Become a "Stiff" When a higher vertebrate dies, its muscles stiffen as they are deprived of ATP, a state called rigor mortis. Explain the molecular basis of the rigor state.

11. Sarcomeres from Another Point of View The symmetry of thick and thin filaments in a sarcomere is such that six thin filaments ordinarily surround each thick filament in a hexagonal array. Draw a cross section (transverse cut) of a myofibril at the following points: (a) at the M line; (b) through the I band; (c) through the dense region of the A band; (d) through the less dense region of the A band, adjacent to the M line (see Fig. 7–31b).

Biochemistry on the Internet

12. Lysozyme and Antibodies To fully appreciate how proteins function in a cell, it is helpful to have a three-dimensional view of how proteins interact with other cellular components. Fortunately, this is possible using the Internet and on-line protein databases. Go to the biochemistry site at

http://www.worthpublishers.com/lehninger

to learn how to use the Chemscape Chime three-dimensional molecular viewing utility. You can then use the Protein Data Bank and Chemscape Chime to investigate the interactions between antibodies and antigens in more detail.

To examine the interactions between the enzyme lysozyme (Chapter 6) and the Fab portion of the anti-lysozyme antibody, go to the Protein Data Bank Website. Use the PDB identifier 1FDL to retrieve the data page for the IgG1 Fab Fragment-Lysozyme Complex (antibody-antigen complex). Open the structure using Chemscape Chime, and use the different viewing options to answer the following questions:

(a) Which chains in the three-dimensional model correspond to the antibody fragment and which correspond to the antigen, lysozyme?

(b) What secondary structure predominates in this Fab fragment?

(c) How many amino acid residues are in the heavy and light chains of the Fab fragment and in lysozyme? Estimate the percentage of the lysozyme that interacts with the antigen-binding site of the antibody fragment.

(d) Identify the specific amino acid residues in lysozyme and in the variable regions of the heavy and light chains that appear to be situated at the antigen-antibody interface. Are the residues contiguous in the primary sequence of the polypeptide chain?