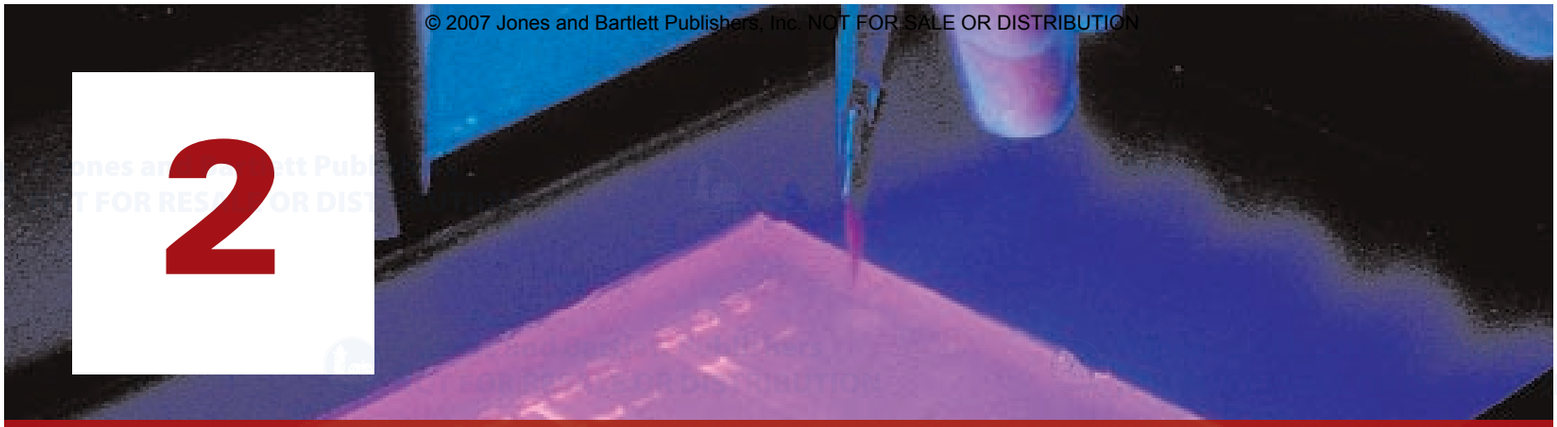
SECTION 1

Protein Structure and Function



2

Protein Structure

OUTLINE OF TOPICS

Photo courtesy of James Gathany / CDC

As described in Chapter 1, the Watson-Crick Model helped to bridge a major gap between genetics and biochemistry, and in so doing helped to create the discipline of molecular biology. The double helix structure showed the importance of elucidating a biological molecule's structure when attempting to understand its function. This chapter and Chapter 3 extend the study of structure-function relationships to polypeptides, which catalyze specific reactions, transport materials within a cell or across a membrane, protect cells from foreign invaders, regulate specific biological processes, and support various structures.

The basic building blocks for polypeptides are small organic molecules called **amino acids**. Amino acids can combine to form long linear chains known as **polypeptides**. Each type of polypeptide chain has a unique amino acid sequence. Although a polypeptide must have the correct amino acid sequence to perform its specific biological function, the amino acid sequence alone does not guarantee that the polypeptide will be biologically active. The polypeptide must fold into a specific three-dimensional structure before it can perform its biological function(s). Once folded into its biologically active form, the polypeptide is termed a **protein**. Polypeptides are unique among biological molecules in their flexibility, which allows them to fold into characteristic three-dimensional structures with specific binding properties. Mutations that alter a protein's ability to interact with its normal molecular partners often result in a loss of protein activity. One of the most common partners of a folded polypeptide is another folded polypeptide, which may be identical to it or different. A complex that contains two, three, or more identical polypeptides is called a **homodimer**, **homotrimer**, and so forth, whereas one that contains different polypeptides is called a **heterodimer**, **heterotrimer**, and so forth.

2.1 The α-Amino Acids

α-Amino acids have an amino group and a carboxyl group attached to a central carbon atom.

The typical amino acid building block for polypeptide synthesis has a central carbon atom that is attached to an amino ($-NH_2$) group, a carboxyl (-COOH) group, a hydrogen atom, and a side chain (-R). At pH 7, the amino group is protonated (i.e., the addition of a proton) to form $-NH_3^+$ and the carboxyl group is deprotonated to form $-COO^-$ so that the amino acid has the structure shown in **FIGURE 2.1**. These amino acids are termed **α-amino acids** in accordance with a pre-IUPAC nomenclature system, in which the atoms in a hydrocarbon chain attached to a carboxyl (-COOH) group are designated by Greek letters. The carbon atom closest to the carboxyl group is designated α, the next β, and so forth.

Each amino acid has characteristic physical and chemical properties that derive from its unique side chain. Amino acids with similar side chains usually have similar properties. This relationship is an important consideration when comparing amino acid sequences of two different polypeptides or when considering the effect that an amino

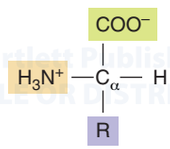


FIGURE 2.1 Structure of an α-amino acid. A typical α-amino acid in which the central carbon atom is attached to an amino ($-NH_3^+$) group, a carboxylate ($-COO^-$) group, a hydrogen atom, and a side chain (-R).

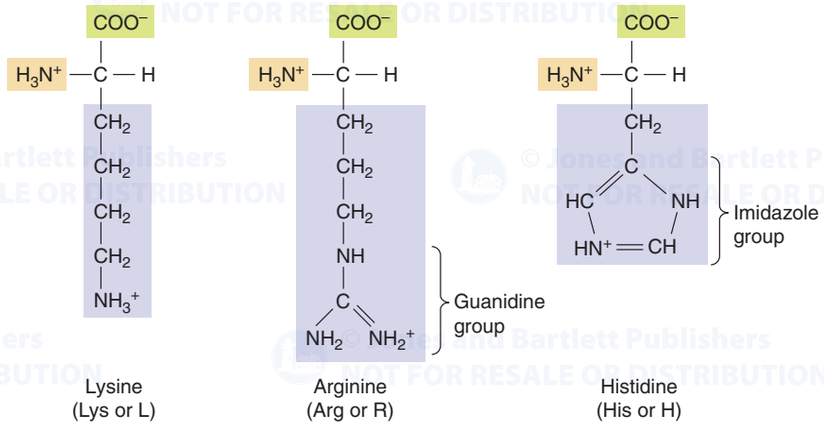


FIGURE 2.2 Amino acids with basic side chains.

acid substitution will have on protein function. Based on side chain structure, amino acids can be divided into four groups.

Side Chains with Basic Groups

Arginine, **lysine**, and **histidine** are called basic amino acids because their side chains are proton acceptors (**FIGURE 2.2**). The guanidino group in arginine's side chain is a relatively strong base. The amine group in lysine's side chain is a somewhat weaker base, and the imidazole group in histidine's side chain is the weakest of the three bases.

Side Chains with Acidic Groups

Aspartic acid and **glutamic acid** each has a carboxyl group as part of its side chain (**FIGURE 2.3**). Both the α-carboxyl and the side chain carboxyl groups are deprotonated and have negative charges at pH 7. However, the α-carboxyl group is a slightly stronger acid because the α-carbon is also attached to a positively charged amino group. When the side chain is deprotonated, aspartic and glutamic acids are more appropriately called aspartate and glutamate, respectively. Because aspartic acid and aspartate refer to the same amino acid at different pH values, the names are used interchangeably. The same is true for glutamic acid and glutamate.

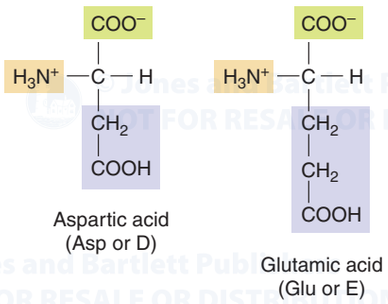


FIGURE 2.3 Amino acids with acidic side chains.

Side Chains with Polar but Uncharged Groups

Six amino acids have side chains with polar groups (**FIGURE 2.4**). **Asparagine** and **glutamine** are amide derivatives of aspartate and gluta-

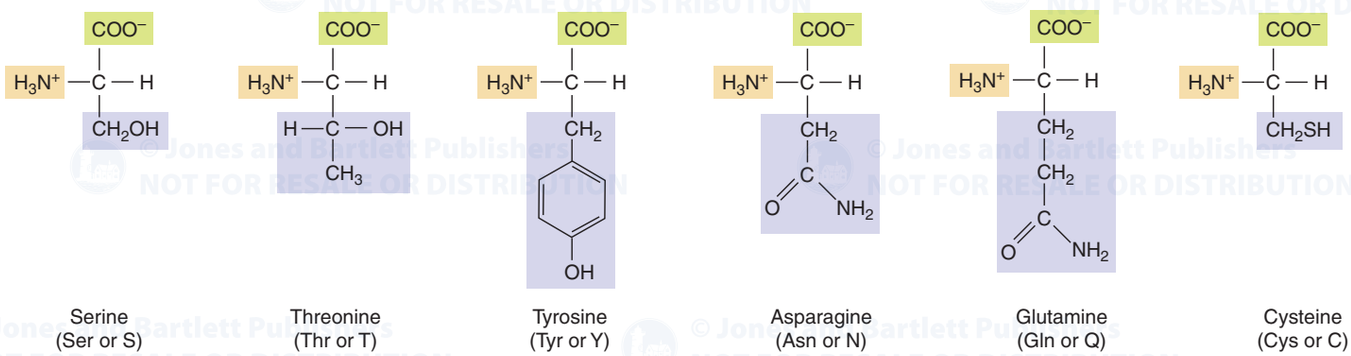


FIGURE 2.4 Amino acids with polar but uncharged side chains at pH 7.0.

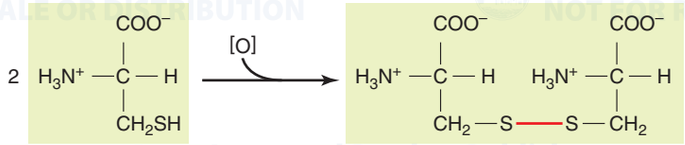


FIGURE 2.5 Oxidation of cysteine to form cystine.

mate, respectively. **Serine**, **threonine**, and **tyrosine** have side chains with hydroxyl (-OH) groups. The tyrosine side chain also has another interesting feature; it is aromatic. **Cysteine** is similar to serine but a sulfhydryl (-SH) group replaces the hydroxyl group. When exposed to oxygen or other oxidizing agents, sulfhydryl groups on two cysteine molecules react to form a disulfide (—S—S—) bond, resulting in the formation of cystine (**FIGURE 2.5**). Cystine, which is not a building block for polypeptide synthesis, is formed by the oxidation of cysteine side chains after the polypeptide has been formed.

Side Chains with Nonpolar Groups

Nine amino acids have side chains with nonpolar groups (**FIGURE 2.6**). **Glycine**, with a side chain consisting of a single hydrogen atom, is the smallest amino acid and the only one that lacks a stereogenic carbon atom. Because it is so small, glycine can fit into tight places and tends to behave like amino acids with polar but uncharged side chains when present in a polypeptide.

Alanine, **isoleucine**, **leucine**, and **valine** have hydrocarbon side chains. **Phenylalanine** and **tryptophan** have aromatic side chains. **Methionine** and **proline** have side chains with unique features. The

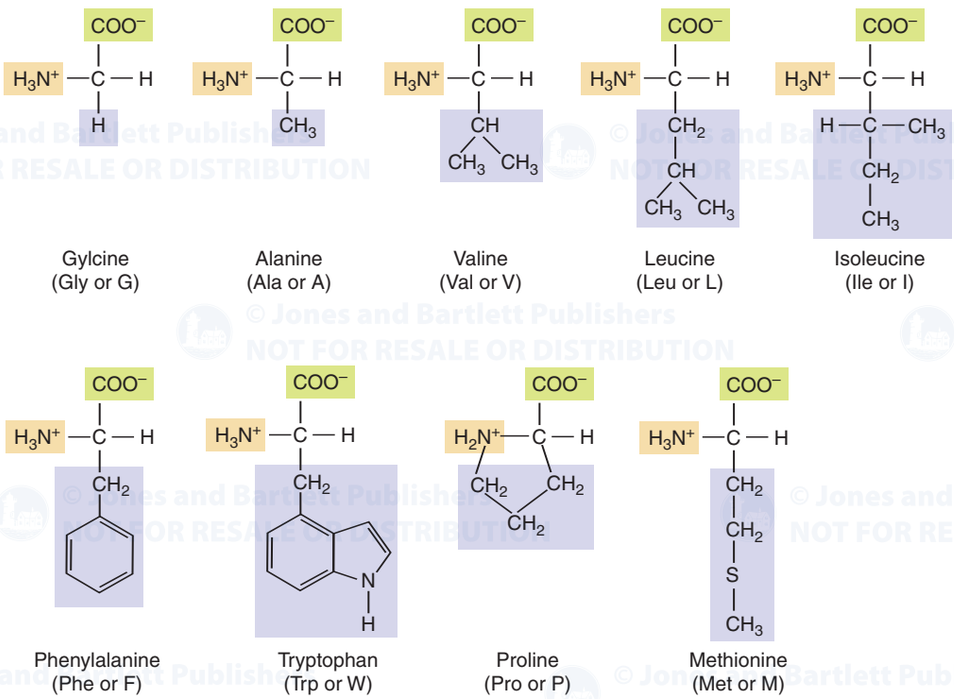


FIGURE 2.6 Amino acids with nonpolar side chains.

TABLE 2.1 **Amino Acid Abbreviations**

Amino Acid	Three Letter Abbreviation	One Letter Abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid (Aspartate)	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid (Glutamate)	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

methionine side chain contains a thioether ($-CH_2—S—CH_3$) group. Proline's side chain is part of a five-member ring that includes the α-amino group, making the α-amino group a secondary rather than a primary amine group. The rigid ring structure can influence the way a polypeptide chain folds by introducing a kink into the structure.

Amino acids are represented by three-letter and one-letter abbreviations.

Writing the full names of the amino acids is inconvenient, especially for polypeptide chains with many amino acids. Two systems of abbreviations listed in Table 2.1 offer more convenient methods for representing amino acids.

In the first system, each amino acid is represented by a three-letter abbreviation. For most amino acids, the first three-letters of the amino acid's name are used. For example, Arg is used for arginine, Phe for phenylalanine, and Lys for lysine. But four amino acids have unusual three-letter abbreviations; aspartate and asparagine have identical first three letters and the same is true for glutamate and glutamine. So the abbreviations for aspartate and glutamate are the expected Asp and Glu, respectively and those for asparagine and glutamine are Asn and Gln, respectively.

The other two amino acids with unusual three letter abbreviations are Trp for tryptophan and Ile for isoleucine. Even the three-letter abbreviation system requires too much space for many

applications. So investigators devised the one-letter abbreviation system shown in Table 2.1.

2.2 The Peptide Bond

α-Amino acids are linked by peptide bonds.

The α-carboxyl group of one amino acid can react with the α-amino group of a second amino acid to form an amide bond and release water (**FIGURE 2.7**). Amide bonds that link amino acids are designated **peptide bonds** and the resulting molecules are called **peptides**. Peptides with two amino acids are dipeptides, those with three are tripeptides, and so forth.

Systematic physical studies by Linus Pauling and Robert R. Corey in the late 1930s provide important information about bond distances and angles in dipeptides. The results of their studies, summarized in **FIGURE 2.8**, show that the carbon-nitrogen peptide bond is 1.33 Å long, placing it between the length of a carbon-nitrogen single bond (1.49 Å) and a carbon-nitrogen double bond (1.27 Å). Therefore, the peptide bond has some double bond character, producing an energy barrier to free rotation and imposing a planarity on the peptide bond. Two configurations are possible, one in which adjacent C_α atoms are on opposite sides (the *trans* isomer; Figure 2.8) and another in which they are on the same side of the peptide bond (the *cis* isomer; **FIGURE 2.9**). The *trans* isomer is the more stable of the two because there is less physical contact between the side chains of the two amino acids forming the peptide bond. Although rare, *cis* isomers do occur in polypeptides, especially when a proline is on the carboxyl side of a peptide bond.

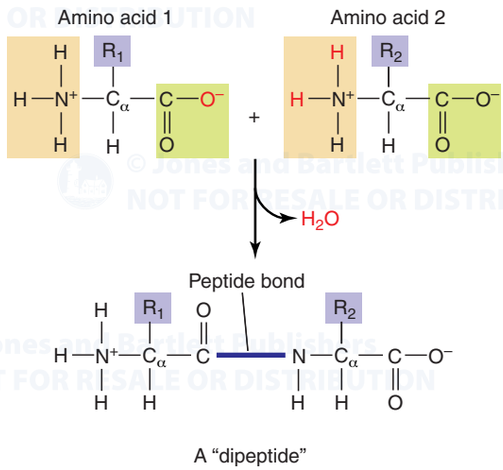


FIGURE 2.7 Peptide bond formation in the laboratory. Two amino acids combine with the loss of water to form a dipeptide. The peptide bond is shown in blue.

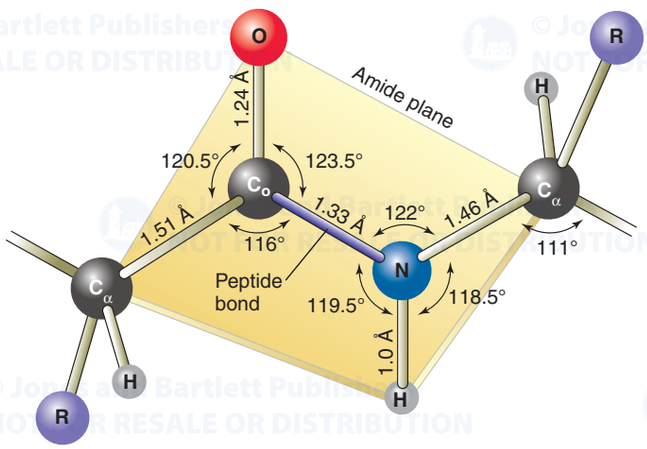


FIGURE 2.8 The *trans* peptide bond. The standard dimensions of bond lengths in Angstroms (Å) and bond angles in degrees (°) of the planar dipeptide were determined by averaging the corresponding quantities in x-ray crystal structures of peptides. Adapted from Voet, D., and Voet, J. G. *Biochemistry, 3/e*. John Wiley & Sons, Ltd. 2005.

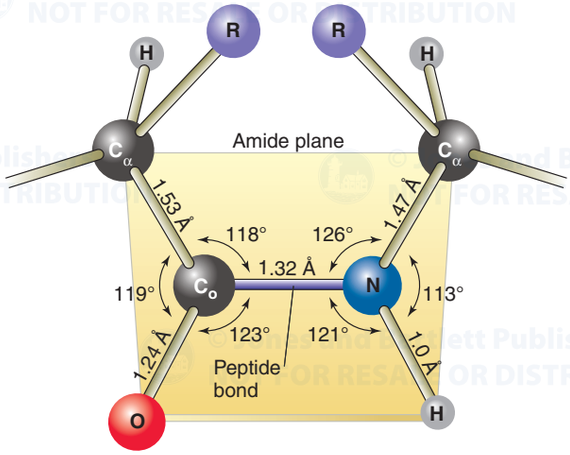


FIGURE 2.9 The *cis* peptide bond. Adapted from Voet, D., and Voet, J. G. *Biochemistry, 3/e*. John Wiley & Sons, Ltd. 2005.

Like polynucleotide chains, peptide chains have directionality. A free amino group is present at one end of the peptide and a free carboxyl group at the other end. By convention, the free amino group is drawn on the left. Once linked in a peptide chain, amino acids are called **amino acid residues** (or **residues** for short). Some important peptide nomenclature conventions are summarized in **FIGURE 2.10**.

Rather arbitrarily, peptides are divided by size into two major groups. Those with less than fifty amino acids are **oligopeptides**, while those with fifty or more residues are **polypeptides**. As stated above, the term *protein* is reserved for a polypeptide chain (or set of associated polypeptide chains) with a specific three-dimensional structure that is essential for biological function(s). Because peptides can vary in chain length, amino acid sequence, or both, one can imagine an almost limitless variety of peptides. For example, there are 20^{50} or slightly more than 1.12×10^{65} possible sequences for polypeptides with just 50 amino acid residues.

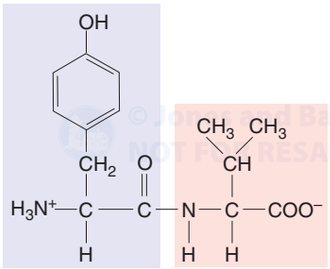
2.3 Protein Purification

Protein mixtures can be fractionated by chromatography.

A complex process such as DNA replication or RNA synthesis requires many different proteins that must work together. Each protein makes a specific contribution to the overall process. However, it's difficult to examine the structure and function(s) of an individual protein when it is present in a mixture of other proteins. Fortunately, most proteins are reasonably hardy and so retain their biological activity during purification. Nevertheless, it is usually desirable to fractionate proteins at 4°C and at or about pH 7 to prevent the loss of biological activity.

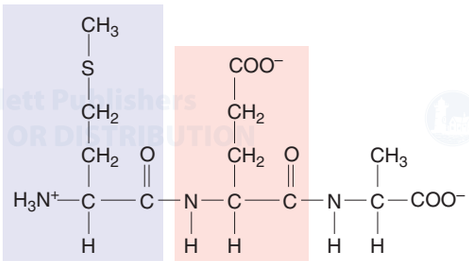
One general method for protein purification, called **column chromatography**, separates proteins in a mixture by repeated partitioning

Dipeptide



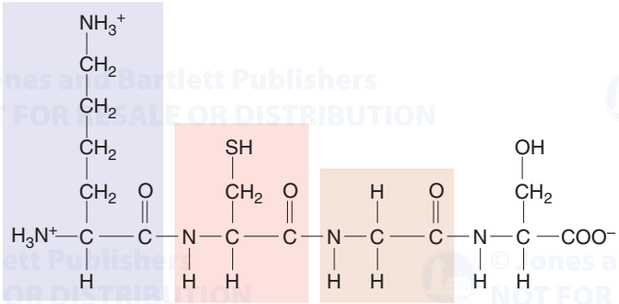
Tyrosylvaline
(Tyr-Val or YV)

Tripeptide



Methionylglutamylalanine
(Met-Glu-Ala or MEA)

Tetrapeptide



Lysylcysteinylglycylserine
(Lys-Cys-Gly-Ser or KCGS)

FIGURE 2.10 Conventions for drawing peptides. By convention, the amino acid terminus (N-terminus) is on the left and the carboxyl terminus (C-terminus) is on the right. Peptides are named as derivatives of the carboxyl terminal amino acid.

between a mobile aqueous solution and an immobile solid matrix. The solution containing the protein mixture is percolated through a column containing the immobile solid matrix consisting of thousands of tiny beads (**FIGURE 2.11**). As the solution passes through the column, proteins interact with the immobile matrix and are retarded. If the column is long enough, it can separate proteins that have different migration rates. Proteins released from the column can be de-

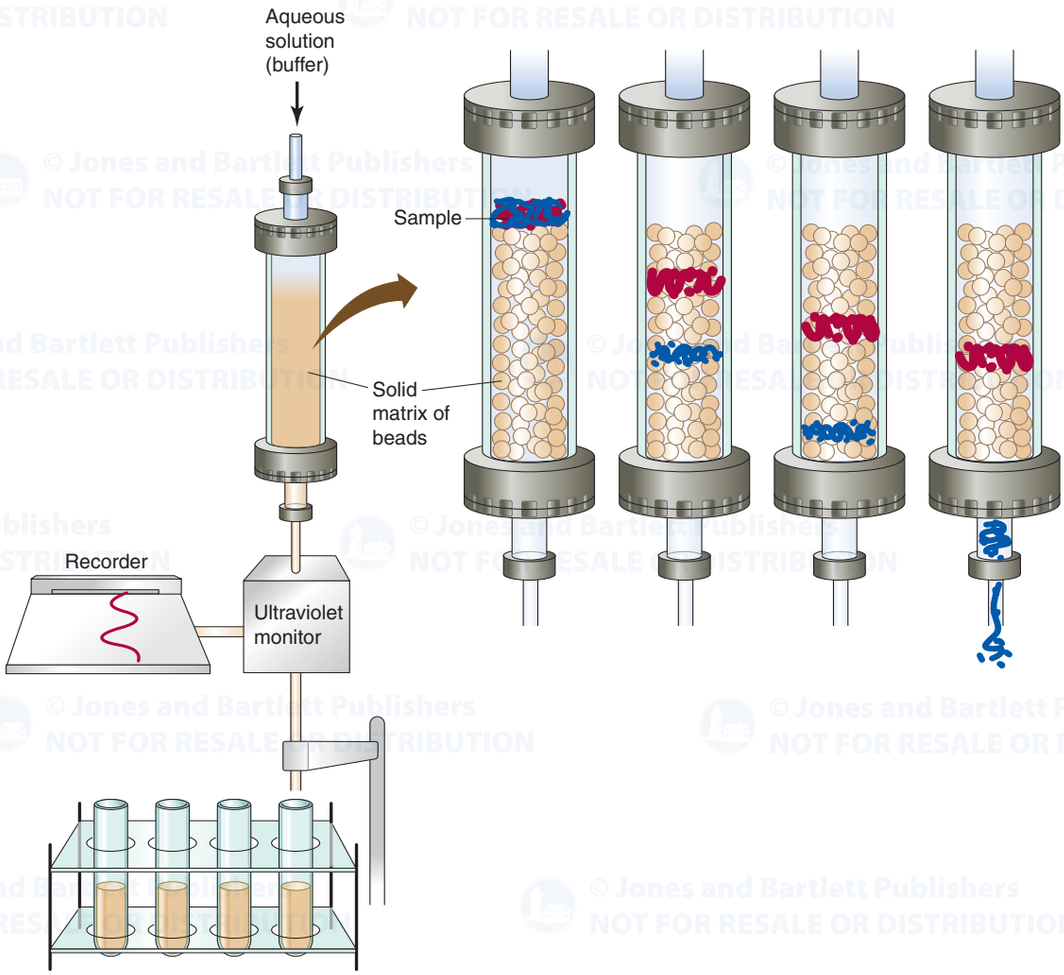


FIGURE 2.11 Schematic for column chromatography. Adapted from an illustration by Wilbur H. Campbell, Michigan Technological University (http://www.bio.mtu.edu/campbell/bl4820/lectures/lec6/482w62.htm).

tected by an ultraviolet monitor and then collected in tubes by a fraction collector. Protein separation can be improved by increasing the solid matrix's surface area through the use of a longer column or finer beads.

However, both methods for increasing the surface area reduce solvent flow through the column. Initial attempts to increase the flow rate by forcing the aqueous solution through the column under pressure were unsuccessful because the pressure compressed the beads, impeding solvent flow.

The flow rate problem was solved by developing hard chromatographic beads that permit the aqueous solution to be forced through the column under high pressure, shortening the time required to achieve a separation and increasing resolution. This modified procedure known as **high-performance liquid chromatography** (**HPLC**) is now widely used in protein purification. Several different kinds of at-

tractive interactions can be used to retard a protein migration through the solid matrix.

Ion-exchange chromatography uses electrostatic interactions between the protein and the solid matrix to fractionate proteins (**FIGURE 2.12**). A sample containing a mixture of proteins is allowed to percolate through a column packed with an immobile matrix, such as polysaccharide beads that are coated with positively (or negatively) charged groups. The beads' charged groups interact with the charged amino acid side-chains on the protein.

At pH 7, aspartate, glutamate, and carboxyl terminal residues will have negative charges and interact with positively charged resins (anion-exchange chromatography). Lysine, arginine, and amino terminal residues will have positive charges and interact with negatively charged resins (cation exchange resins).

Proteins are released by passing aqueous solutions with progressively higher salt concentrations through the column. The salt ions displace the charged side chains from the ion exchange beads. Proteins that interact with the column most weakly will migrate through the column fastest. Because proteins have both positively-charged and negatively-charged side chains on their surface, a specific protein may be fractionated by both anion and cation exchange chromatography.

Reverse phase chromatography (also called hydrophobic chromatography) uses the weak attractive interactions between nonpolar amino acid side chains and nonpolar groups such as phenyl or octyl groups attached to polysaccharide beads to retard protein migration (**FIGURE 2.13**). Proteins, dissolved in an aqueous solution with a high salt concentration, are applied to the column filled with the reverse-phase chromatography beads. Proteins stick to the beads more tightly when the salt concentration is high. The interaction of a given protein with the reverse phase resin depends on the number and placement of its nonpolar amino acid residues. Proteins are released

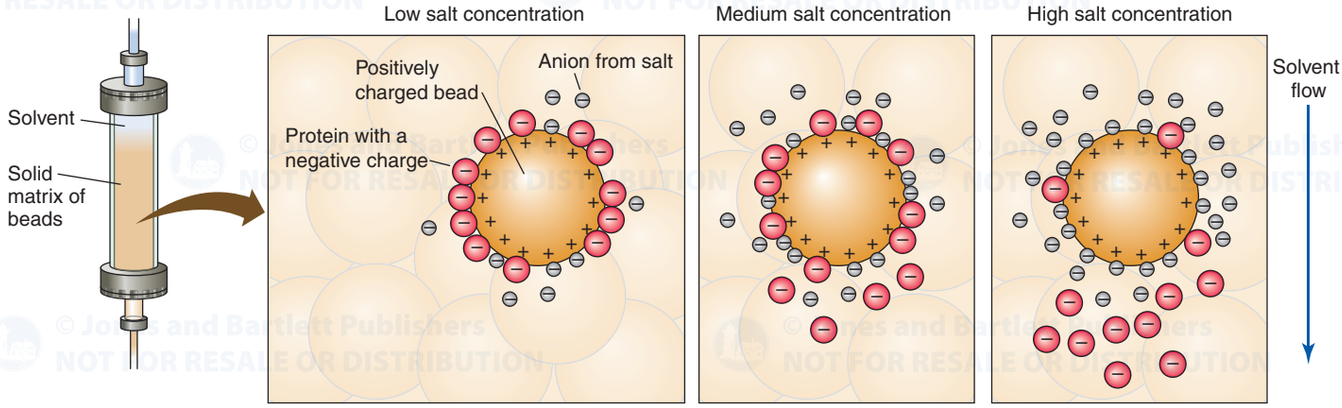


FIGURE 2.12 Anion exchange chromatography. Negatively charged groups on the proteins bind to positively charged groups on the anion exchange resin. Increasing salt concentrations produce anions that displace the proteins. Cation exchange resins work in a similar way except in this case, positively charged groups on the proteins bind to the negatively charged groups on the resin and cations displace the proteins.

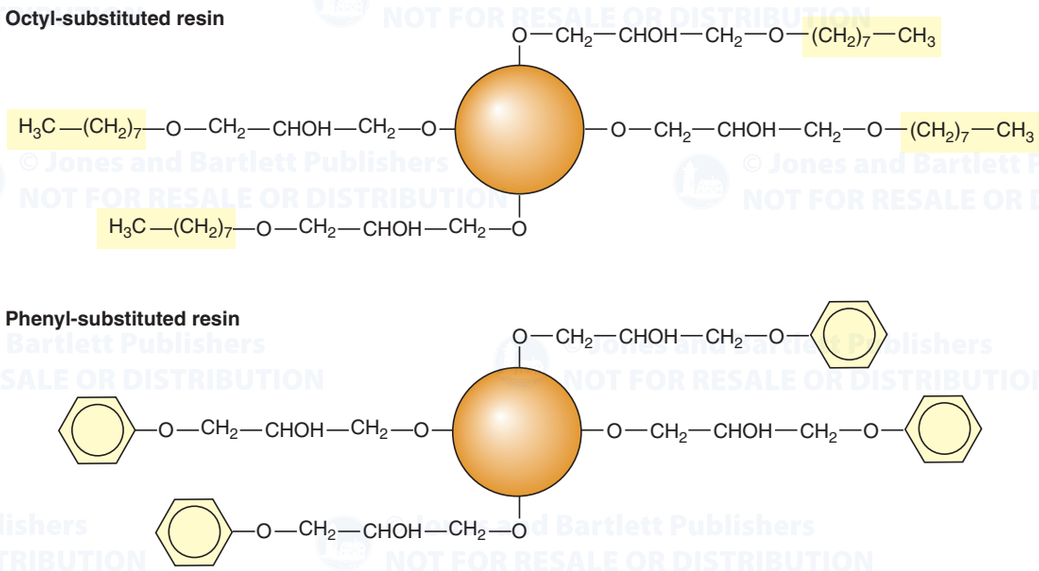


FIGURE 2.13 Resins used for reverse phase chromatography.

by passing aqueous solutions with progressively lower salt concentrations through the column. Proteins with the lowest affinity for the beads will be released first. Thus, the conditions used for binding and eluting in reverse phase chromatography are the opposite of those used in ion exchange chromatography.

Gel filtration or **molecular exclusion chromatography** separates protein molecules by size. This method depends upon special beads that permit small proteins to penetrate into their interior while excluding large proteins from this region (**FIGURE 2.14a**). A gel filtration column has two different water compartments: the internal compartment consists of the aqueous solution inside the beads and the external compartment consists of the aqueous solution outside the beads. Small protein molecules have access to both compartments, whereas large protein molecules only have access to the external compartment. Therefore, the large proteins appear in earlier fractions than do the small proteins (**FIGURE 2.14b** and **c**).

Affinity chromatography takes advantage of the fact that many proteins can bind specific small molecules termed **ligands**. Affinity chromatography exploits this specificity to purify the protein. Ligands are attached to tiny beads to form affinity beads that are suspended in an aqueous buffer and poured into a column. In one of the earliest experiments, the deoxythymidylic acid derivative shown in **FIGURE 2.15** was attached to an insoluble polysaccharide to form affinity beads that bind nucleases (enzymes that catalyze nucleic acid hydrolysis). An aqueous solution, containing a mixture of bacterial proteins, was passed through a column packed with these affinity beads. Most of the proteins did not bind to the ligand and passed through the column. However, the nuclease did bind to the ligand and was retained by the column. Active nuclease was recovered by washing the beads

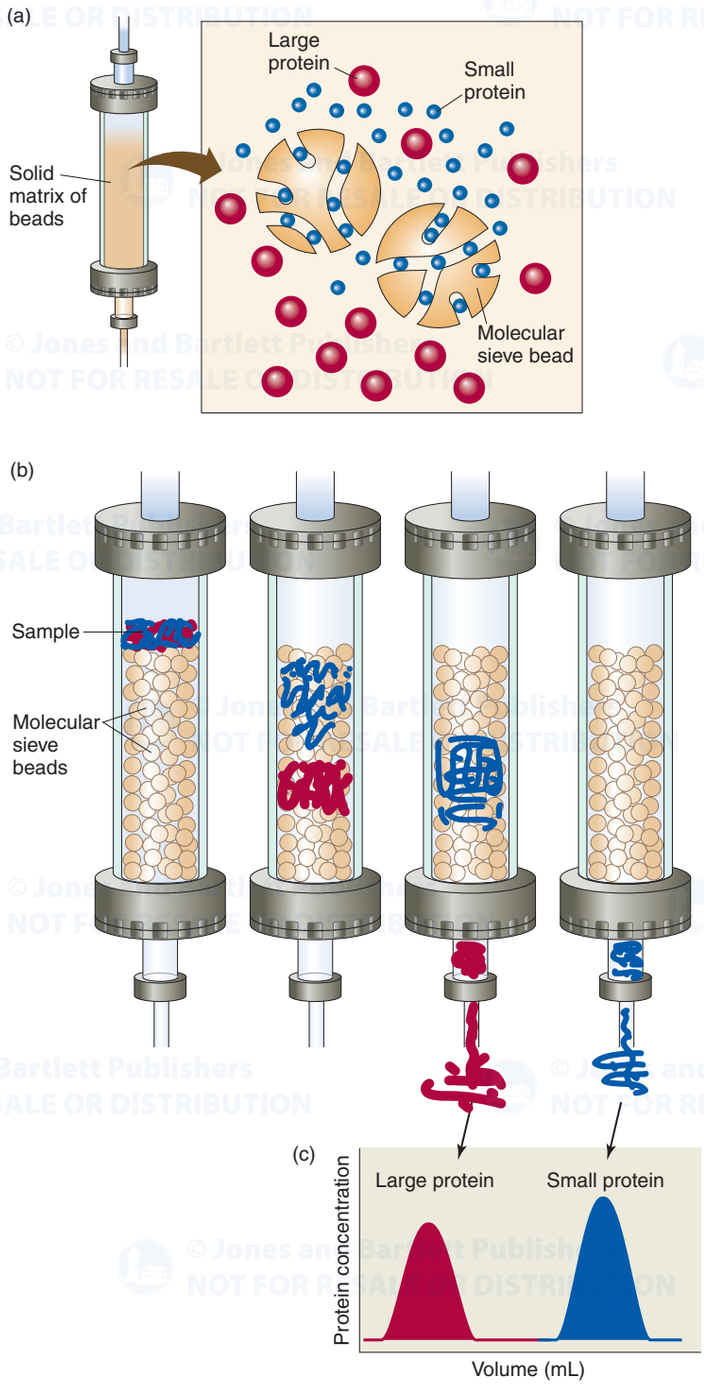


FIGURE 2.14 Gel filtration chromatography. (a) The gel filtration column has an internal compartment consisting of the aqueous solution inside specially designed beads and an external compartment consisting of the aqueous solution outside the beads. The beads permit small proteins to penetrate into their matrix while excluding large proteins from the region; the large proteins can only get into the external compartment. (b) Proteins are separated by size with the larger proteins appearing in earlier fractions and smaller proteins in later fractions. (c) Protein elution profile. Adapted from Tropp, B. E. *Biochemistry: Concepts and Applications.* Brooks/Cole Publishing Company, 1997.

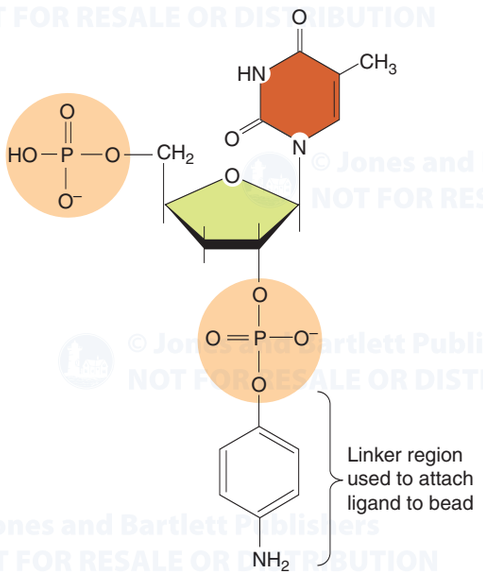


FIGURE 2.15 Affinity chromatography. Affinity chromatography exploits a protein's ability to bind to ligands. Nucleases bind to deoxyribonucleotide derivatives such as the one shown. Affinity resins are prepared by attaching the ligand to tiny water-insoluble beads. Adapted from Tropp, B. E. *Biochemistry: Concepts and Applications.* Brooks/Cole Publishing Company, 1997.

with a buffer solution at a low pH. Enzymes can also be eluted by washing the column with a solution that contains free ligand.

Affinity chromatography can provide a high degree of protein purification due to the specificity of the binding step. Thanks to recombinant DNA technology, it is possible to use a variant of affinity chromatography to purify almost any protein if the DNA that codes for it has been isolated. The basic approach is to modify the DNA so that the gene of interest now codes for the protein with an additional six histidine residues at its amino or carboxyl end. Specific resins have been devised that have a high affinity for proteins with a $(His)_6$ sequence. Because only the protein encoded by the modified DNA will have a $(His)_6$ sequence, the recombinant protein can be separated from all the other proteins in the mixture. Although addition of a $(His)_6$ tag usually does not alter a protein's biological properties, one must test to be certain that it does not.

Proteins and other charged biological polymers migrate in an electric field.

Multiply charged macromolecules such as proteins and nucleic acids migrate through a medium in response to an electric field (**FIGURE 2.16**). In **protein electrophoresis**, the medium is a porous matrix such as a polyacrylamide gel saturated with buffer solution. The protein sample is applied to one end of the gel and the electric field is generated by connecting a power source to electrodes attached at either end of the gel. Proteins that migrate through the gel at the fastest rate tend to have the greatest net charge, the most compact shape, and the smallest size. The net charge can be altered by changing the pH of the medium. The pH at which a protein has no net charge, and therefore will not migrate in an electric field, is called its **isoelectric pH**. Protein

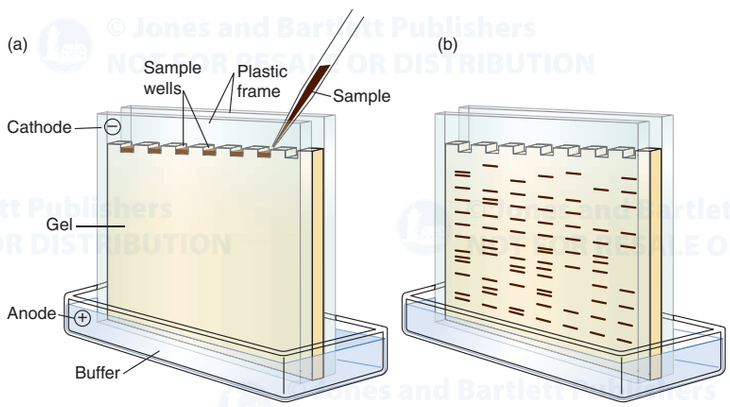


FIGURE 2.16 Gel electrophoresis. Samples are applied to slots in a porous matrix such as polyacrylamide gel, and an electric field is generated. Proteins that migrate through the gel at the fastest rate tend to have the greatest net charge, the most compact shape, and the smallest size. Adapted from Tropp, B. E. *Biochemistry: Concepts and Applications.* Brooks/Cole Publishing Company, 1997.

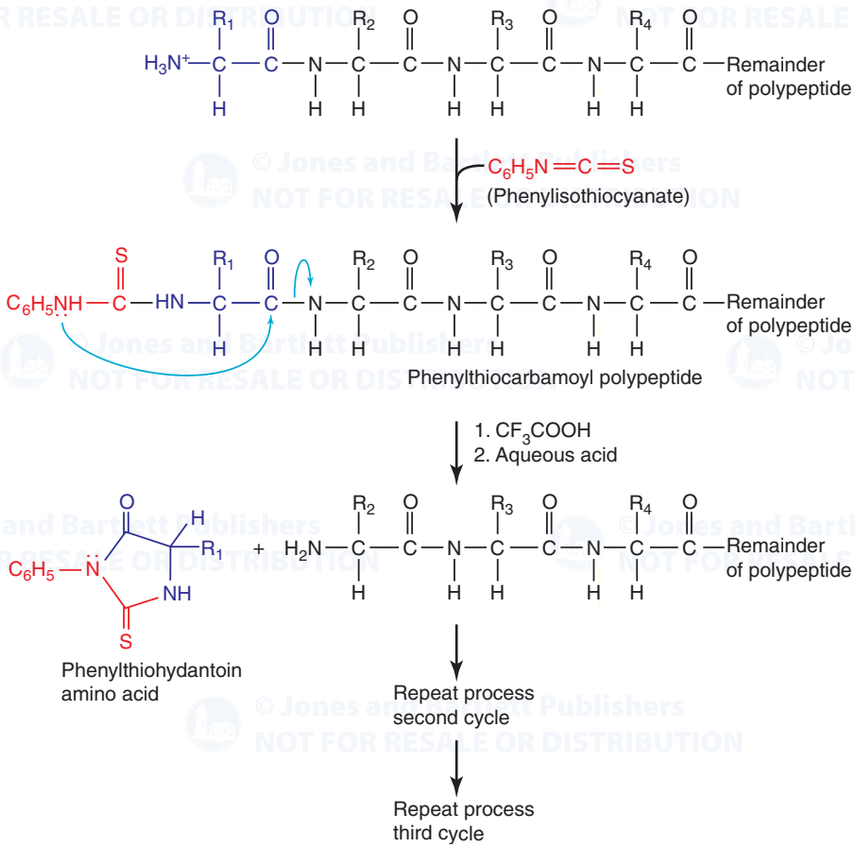


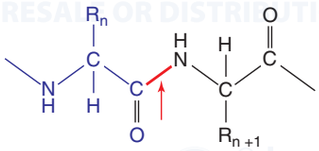
FIGURE 2.17 The Edman degradation. The Edman degradation, which selectively removes the N-terminal amino acid from a polypeptide chain, can be used to determine a polypeptide's amino acid sequence. Adapted from Tropp, B. E. *Biochemistry: Concepts and Applications.* Brooks/Cole Publishing Company, 1997.

bands are visualized by staining with dyes. A protein free of all contaminants will appear as a single band. Hence, electrophoresis is a very useful method for monitoring protein purity.

2.4 Primary Structure of Proteins

Amino acid sequences can be determined by using the Edman degradation procedure and the overlap method.

The amino acid sequence or primary structure of a purified protein can be determined. Once a protein has been purified, it must be characterized to learn more about its chemical and biological properties. A degradation technique devised by Pehr Edman in 1950, until recently, has been the most widely used method for amino acid sequence determination. Edman degradation involves a series of chemical steps that remove the amino acid from the amino terminal end of a polypeptide (**FIGURE 2.17**). The released amino acid deriva-



Enzyme or chemical agent	R_n
Trypsin	Lysine or arginine
Chymotrypsin	Phenylalanine, tyrosine, or tryptophan
Cyanogen bromide (CNBr)	Methionine

FIGURE 2.18 Specific cleavage. Trypsin, chymotrypsin, and cyanogen bromide (CNBr) cleave after specific amino acid residues. The arrow indicates the bond cleavage.

tive is then identified and the process repeated through several rounds of amino acid removal and identification.

Because cleavage efficiency is less than 100%, each successive cleavage cycle produces an increasingly heterogeneous peptide population. After about 50 cycles, the peptide population is so heterogeneous that it becomes nearly impossible to interpret the data. Cutting long polypeptides into well-defined fragments with a digestive enzyme such as trypsin or a chemical reagent such as cyanogen bromide (CNBr) solves this problem (**FIGURE 2.18**). Fragments produced by cleavage with a specific protease or CNBr are separated by chromatography and then sequenced by Edman degradation. Although the fragments are usually short enough to be sequenced completely, one must still determine the fragment order in the original polypeptide.

Fragment order is determined by the overlap method (**FIGURE 2.19**). This method depends upon having the sequences of two fragment collections, each generated by using an enzyme or chemical reagent that cuts after specific residues. For example, trypsin generates a set of fragments by cutting after arginine and lysine residues and cyanogen bromide (CNBr) generates a different set of fragments by cutting after methionine residues. The order of the fragments

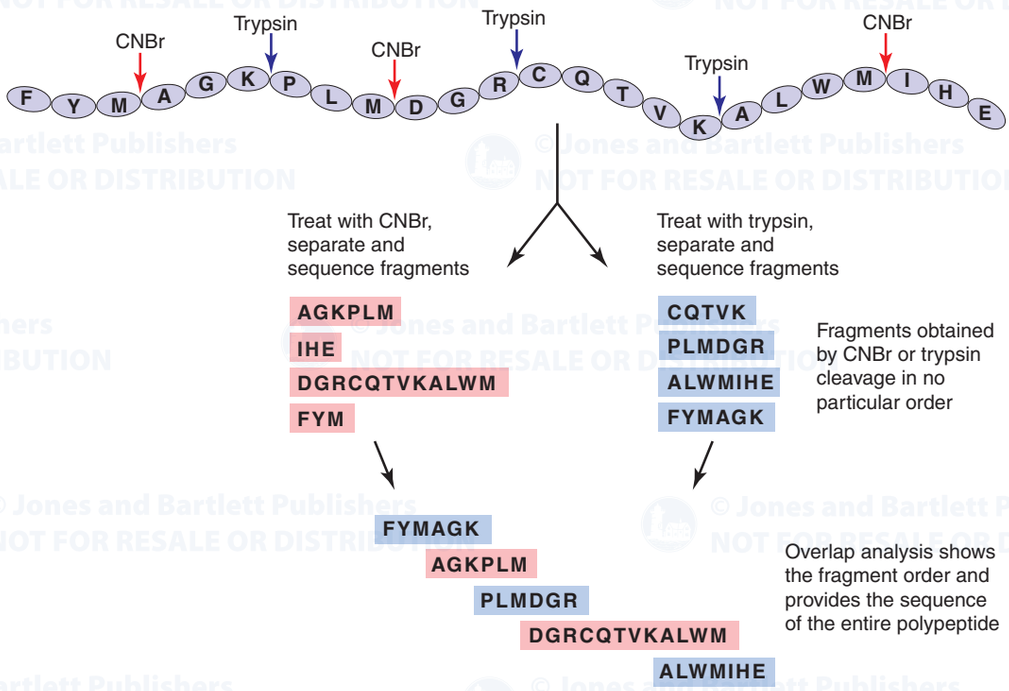


FIGURE 2.19 The overlap method for amino acid sequence determination. A polypeptide that has phenylalanine (F) as its N-terminal residue is divided into two samples. One sample is treated with CNBr, which cleaves after methionine (M) and the other is treated with trypsin, which cleaves after arginine (R) or lysine (K). The CNBr cleavage sites are indicated by the red arrows and the trypsin cleavage sites by the purple arrows. Then the fragments produced from each sample are resolved by chromatography and each purified fragment is sequenced by the Edman degradation procedure. Finally, the fragments are sequenced by searching for overlaps.

within the original polypeptide chain is determined by searching the two sets of fragments for overlapping sequences.

Despite its great value, Edman degradation also has limitations. It is time consuming, does not work when a peptide is blocked at its amino terminus, and does not provide information about amino acid residues that have been modified. A new approach that overcomes these difficulties takes advantage of **mass spectrometry**, a technique in which molecules are ionized and their masses are determined by following the specific trajectories of the ionized fragments in a vacuum system. Because mass spectrometry is very sensitive, it requires very little protein. Moreover, peptide fragmentation takes place in seconds rather than hours, and sequencing is possible even if the protein is not completely pure.

FIGURE 2.20 summarizes the main steps in the mass spectrometer sequence technique. A protein population is prepared from a biological source and the individual polypeptides separated by electrophoresis. After separation, the gel lane is cut into several slices. A specific protease or chemical agent is added to the gel slice of interest to digest the trapped protein, converting the protein into peptides. The peptides in the mixture that is generated by this digestion are separated by HPLC and then analyzed by the mass spectrometer.

Polypeptide sequences can be obtained from nucleic acid sequences.

The development of rapid DNA sequencing methods, which will be described in Chapter 5, greatly accelerated the pace at which polypeptide sequences were determined. Instead of sequencing purified polypeptides, investigators determined the entire DNA sequences of organisms and then translated this information to obtain the polypeptide sequences.

Although the DNA sequencing approach is very fast and quite accurate, extrapolating to polypeptide sequences does have serious limitations. A DNA sequence does not necessarily predict the chemical nature of the biologically active protein for the following reasons: (1) In eukaryotes large precursors of messenger RNA (mRNA) are converted to mRNA by a precise splicing mechanism in which intervening RNA sequences, called introns, are removed with concomitant joining of flanking sequences, called exons (**FIGURE 2.21**; see Chapter 17).

The resulting mRNA molecules that are missing sequences present in DNA will program ribosomes to form polypeptides that are shorter than those predicted from the DNA sequences. In many cases, we cannot predict the sequences that will be lost during splicing and therefore cannot predict the sequence of the biologically active polypeptide. (2) Many polypeptides are converted into their biologically active form by cleavage at specific sites. For instance, the polypeptide precursor to insulin, preproinsulin, is converted into the active hormone by specific peptide bond cleavage. (3) Many proteins are subject to covalent modifications such as disulfide bond formation or the addition of phosphate, sugar, acetyl, methyl, lipid, or other groups. These modifications, which

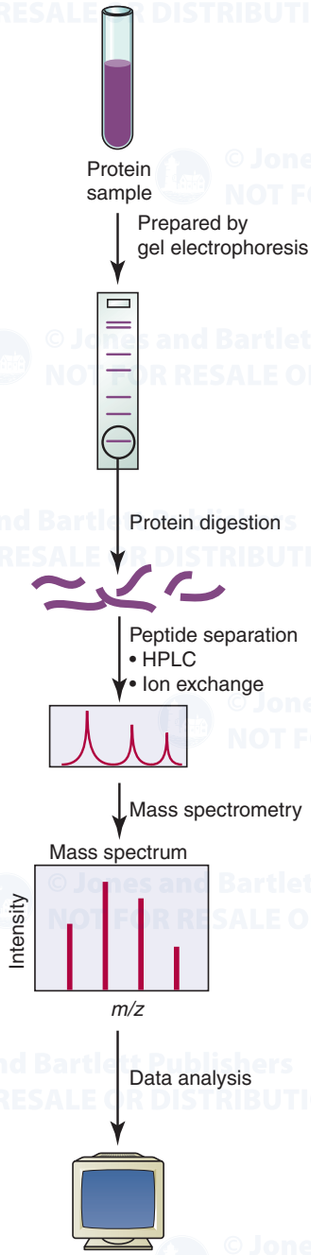


FIGURE 2.20 Mass spectrometric determination of peptide sequence. A protein population is prepared from a biological source such as a bacterial or cell culture. The protein of interest is purified; the last purification step is usually electrophoresis. The gel lane is cut to obtain the band that contains the desired protein. The protein is digested while still in the gel with specific enzymes, chemicals, or both. Then the peptide mixture that is generated is separated by chromatography and individual peptides are analyzed by mass spectrometry. The results are analyzed to determine the amino acid sequence within each peptide. The polypeptide sequence can be determined by digesting different samples of the same protein with different cleavage agents so that the overlap method can be used to order the proteins. Adapted from Steen, H., and Mann, M., *Nature Rev. Mol. Cell Biol.* 5 (2004): 699–711.

often influence the protein's biological activity and stability, can only be revealed by studying the purified protein. (4) Many polypeptides do not act alone but instead function as a part of a complex that contains other polypeptides of the same or different types. The true nature of these protein complexes can only be revealed by studying the purified complex.

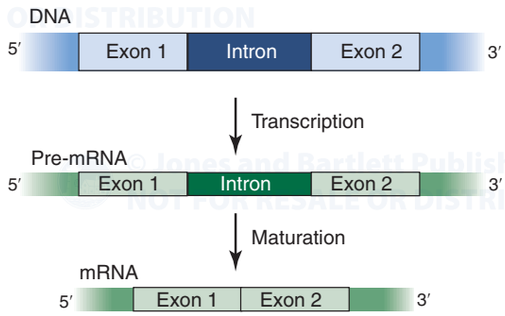


FIGURE 2.21 Split genes. Many eukaryotic genes have intervening sequences (introns) that are included in the precursor to messenger RNA (mRNA) molecules that are formed when the genes are transcribed but are removed during a maturation process in which precursor mRNA molecules are converted to mature mRNA molecules. The coding (or expressed) sequences, which are included in both the precursor mRNA and mature mRNA molecules, are called exons.

The BLAST program compares a new polypeptide sequence with all sequences stored in a data bank.

The genomes of more than 165 different organisms have now been sequenced. The world's most comprehensive catalog of protein sequence information is available from Uniprot. Comparison of a new sequence with sequences stored in the data bank is possible by using the BLAST (*B*asic *L*ocal *A*lignment *S*earch *T*ool) program available at the Uniprot site.

Five different BLAST programs offer fast, sensitive, and relatively easy ways to compare specific nucleic acid or polypeptide sequences (the query sequences) with all sequences (the subject sequences) in the data bank. Each program permits a different type of search. Here we consider just one of the five BLAST programs, the *blastp program*, because it is the one that compares the amino acid query sequence with all polypeptide sequences in the data bank. The blastp program searches the data bank by first looking for every tripeptide in the data bank that is similar to tripeptides in the query polypeptide, and then extends initial regions of similarity into larger alignments without gaps. Once alignments have been created, the blastp program determines and reports the probability of their arising by chance, lists the data bank sequences that are most similar to the query sequence, and shows a local alignment of the query sequence with matched data bank sequences. A newly sequenced polypeptide may be so similar to a polypeptide with a known function in another organism that we can safely conclude that the two polypeptides have the same function.

However, a search of the data banks may also reveal no similar polypeptides or it may reveal polypeptides with similar sequences but with no known function. Now we are in the rather unsettling position of knowing that a polypeptide exists and knowing its sequence but not having any idea about what the polypeptide actually does. Solving this "function problem" is one of the major challenges for molecular biologists in the coming decades.

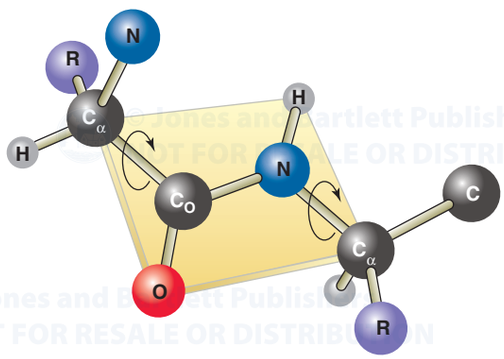


FIGURE 2.22 Rotation about C_α-C_o and N-C_α. Rotation does not take place about N-C_o but is free to take place about C_α-C_o and N-C_α. Adapted from Mathews, C. K., et al. *Biochemistry, 3/e.* Prentice Hall, 2000.

Proteins with just one polypeptide chain have primary, secondary, and tertiary structures while those with two or more chains also have quaternary structures.

The primary structure is just the first of four possible levels of polypeptide structures. The other three levels are determined by the way that the polypeptide is arranged in space. The polypeptide backbone has three types of bonds, C_α—C_o, C_o—N (the peptide bond), and N—C_α (**FIGURE 2.22**). Although rotation about peptide bonds is severely limited, rotation does occur about N—C_α and C_α—C_o, the

two single bonds in the polypeptide backbone. Rotation about the single bonds permits the polypeptide to fold into biologically active proteins. Because of this folding, proteins have three levels of structure in addition to their primary structure (**FIGURE 2.23a**). The **secondary structure** describes the folding pattern within a segment of a polypeptide chain containing neighboring residues. Among the many different possible secondary structures, the three most common are the α-helix, the β-conformation, and the loop or turn (**FIGURE 2.23b**). The **tertiary structure** provides a view of a protein's entire three-dimensional structure (**FIGURE 2.23c**), including spatial arrangements among different segments and among residues within the different segments. The **quaternary structure**, which will be examined in the next chapter, applies only to proteins with two or more polypeptide chains and indicates the way that the chains are arranged in space with respect to one another (**FIGURE 2.23d**).

2.5 Weak Non-Covalent Bonds

The polypeptide folding pattern is determined by weak non-covalent interactions.

The secondary, tertiary, and quaternary structures are each stabilized by weak non-covalent interactions, which are easily formed and easily broken. These interactions are in order of decreasing strength: the **ionic bond**, the **hydrogen bond**, **van der Waals interactions**, and **hydrophobic interactions**. Because each kind of weak interaction has a range of binding energies, there is considerable energy overlap among the four kinds of weak interactions.

Ionic Bond

The **ionic bond** results from the attraction between positively- and negatively-charged ionic groups. For example, the negatively-charged side chain in aspartate or glutamate can form an ionic bond with the positively charged side chain in arginine, lysine, or histidine. Ionic bonds can form between residue pairs that are near to one another in the primary structure or pairs that are far apart. The strength of the ionic bond varies depending on its surroundings. In an aqueous solution, the usual environment for most proteins, ionic bond strength ranges from 10 to 30 kJ • mol^{-1}, which is considerably lower than covalent bond strength (200-1000 kJ • mol^{-1}). Ionic bonds that stabilize protein structure can be easily disrupted by pH changes, which alter the charges on interacting side chains or by high concentrations of small ions that compete with the interacting side chains.

Hydrogen Bond

The hydrogen bond also results from an electrostatic attraction; however, this time the attractive interaction is between a partially negative electronegative atom and a partially positive hydrogen atom linked to a second electronegative atom. Only two electronegative atoms—oxygen and nitrogen—participate in hydrogen bond formation in biological molecules. (Fluorine, which can also participate in

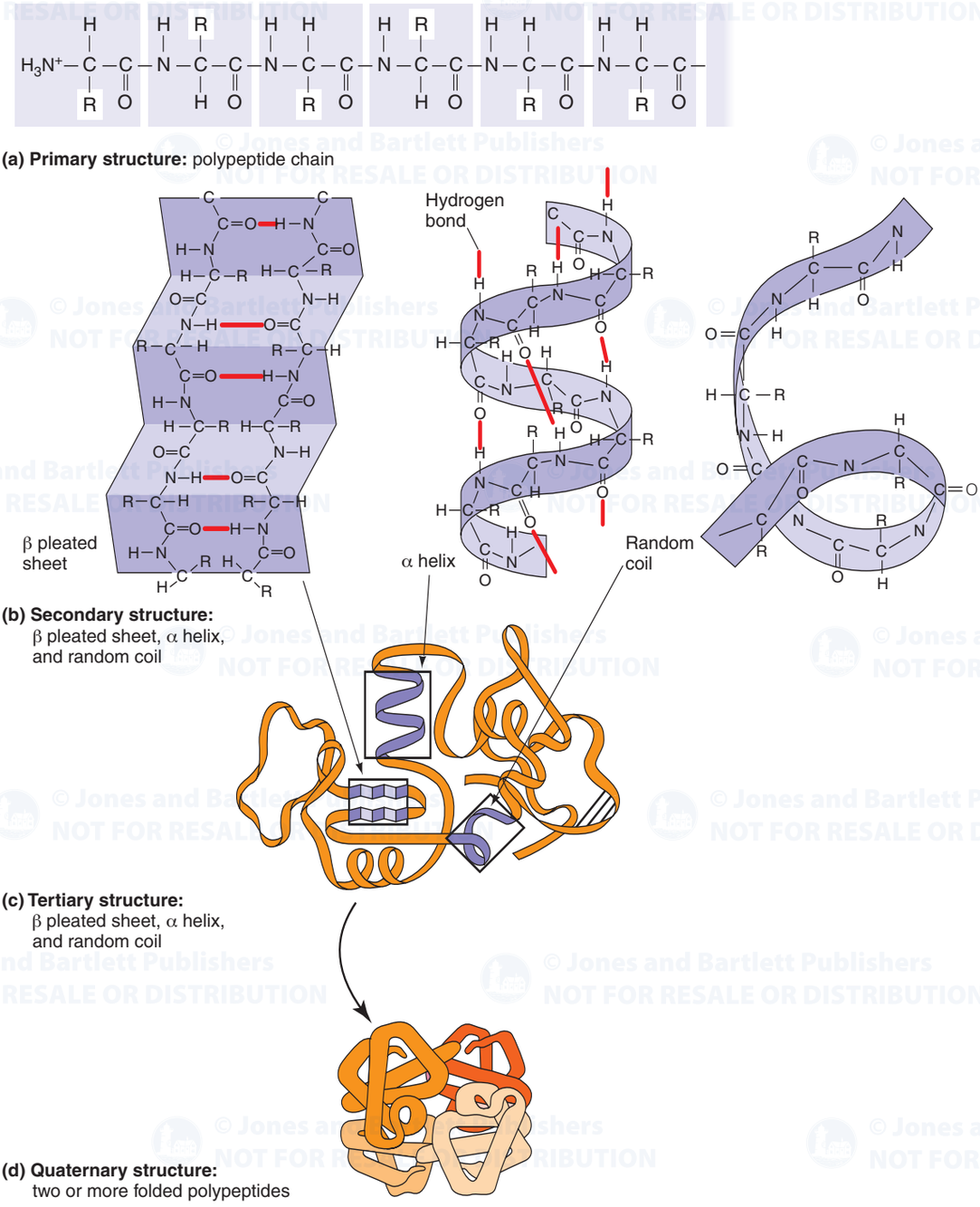


FIGURE 2.23 **Protein structure.** Polypeptides can be viewed at four different structural angles. The ribbon diagram (b and c) is a simple and effective way to represent secondary and tertiary structures. Adapted from Tropp, B. E. *Biochemistry: Concepts and Applications.* Brooks/Cole Publishing Company, 1997.

hydrogen bond formation, is rarely present in biological molecules.) Some typical examples of hydrogen bonds that occur in biological molecules are shown in **FIGURE 2.24**. Hydrogen bond strengths range from 5 to 20 kJ · mol^{-1}. In the strongest hydrogen bonds, the two electronegative atoms and the hydrogen atom lie on a straight line.

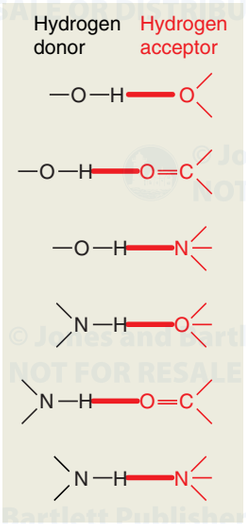


FIGURE 2.24 Some typical hydrogen bonds in proteins. The amino acid residue that supplies the hydrogen atom is designated the donor and the residue that binds the hydrogen atom is designated the acceptor.

The aggregate effect of many hydrogen bonds makes an important contribution to the way that nucleic acids and proteins fold. The Watson-Crick Model (Chapter 1) for DNA recognizes that hydrogen bonds between adenine-thymine base pairs and guanine-cytosine base pairs contribute to the stability of the double helix.

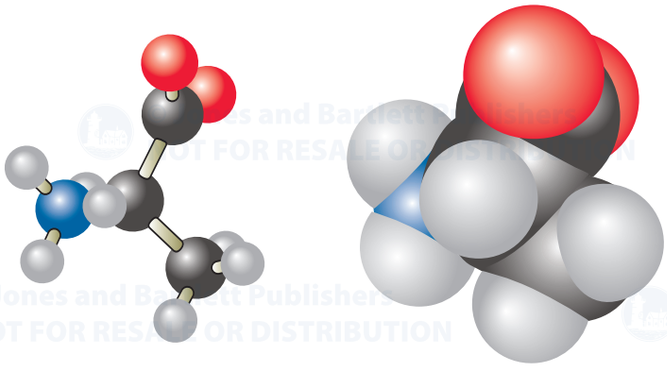
Van Der Waals Interactions

Van der Waals interactions result from weak electrostatic interactions between two polar groups, a polar group and a nonpolar group, or two nonpolar groups. Electrostatic attraction between two polar groups results from the attraction between a partially positive atom on one polar group and a partially negative atom of another polar group. Electrostatic attractions between a polar and a nonpolar group result from the polar group's ability to induce a short-lived polarity in the nonpolar group, which in turn leads to a weak attractive interaction between the oppositely charged regions on the two groups. Very weak electrostatic attractions between two nonpolar groups arise from fluctuating charge densities in the nonpolar groups. At any given time, there is a small probability that a nonpolar group will have an asymmetric electron distribution. A nonpolar group that experiences such a transitory perturbation of charge distribution can induce polarity in neighboring nonpolar groups. The combination of fluctuating and induced polarity accounts for the very weak forces of attraction that hold nonpolar molecules together. The strength of a van der Waals interaction ranges from 1 to 10 kJ • mol^{-1}. Because the attractive force between two atoms is proportional to $1/r^6$ (r is the distance between their nuclei), van der Waals interactions become significant only when two atoms are very near one another (0.1–0.2 nm apart). A powerful repulsive force also comes into play when the outer electron shells of the two atoms overlap. The **van der Waals radius** is defined as the distance at which the attractive and repulsive forces between the atoms balance precisely. Van der Waals radii differ from one kind of atom pair to another; some representative values are shown in **Table 2.2**. The shape of a molecule is in essence the surface formed by the van der Waals spheres of each atom. **FIGURE 2.25** shows the shapes of alanine and proline when defined in this way. The average energy of thermal motion at room temperature is about 2.5 kJ • mol^{-1}. Therefore, van der Waals interaction between two atoms is not sufficient to maintain these atoms in proximity. However, if the interactions of *several* pairs of atoms are combined, the cumulative attractive force can be great enough to withstand being disrupted by thermal motion. Thus, two molecules can attract one another if several of their component atoms can mutually interact. However, because of the $1/r^6$-dependence, the intermolecular fit must be nearly perfect. Therefore, two molecules will hold together if their shapes are complementary. Likewise, two separate regions of a polymer will hold together if their shapes match. Sometimes, the van der Waals attraction between two regions is not large enough to cause binding; however, it can significantly strengthen other weak interactions such as the hydrophobic interaction, if the fit is good.

TABLE 2.2 Van der Waals Radii

Atom	Van der Waals radii (nm)
Hydrogen	0.120
Oxygen	0.152
Nitrogen	0.155
Carbon	0.170
Sulfur	0.180
Phosphorus	0.180

(a) Alanine shown in ball and stick (left) and van der Waals displays



(b) Proline shown in ball and stick (left) and van der Waals displays

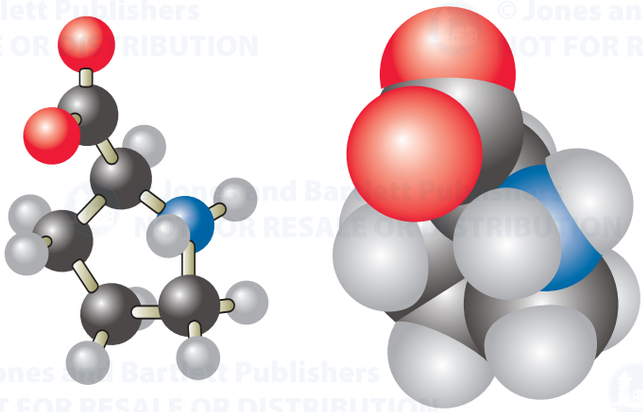
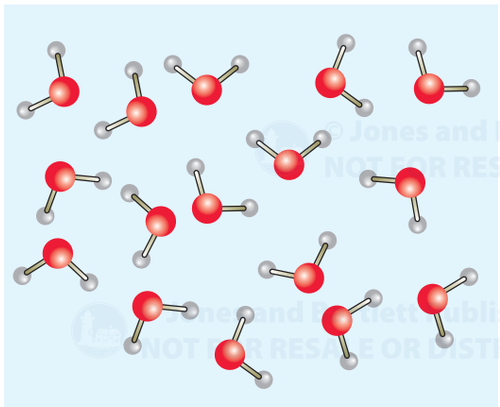


FIGURE 2.25 Alanine and proline as ball and stick and van der Waals displays. (a) Alanine shown in ball and stick (left) and van der Waals (right) displays. (b) Proline shown in ball and stick (left) and van der Waals (right) displays.

Hydrophobic Interaction

The driving force for hydrophobic interactions can be explained by considering the structural arrangement of surrounding water molecules. When a nonpolar molecule or group is placed in water, water molecules interact through hydrogen bonds to form highly ordered cages around the nonpolar molecule or group (**FIGURE 2.26**). The second law of thermodynamics tells us that disorder is favored over order. Stated another way, the second law indicates that **entropy**, the measure of disorder, increases for spontaneous processes. Thus, the ordering of water molecules to form a cage around a nonpolar molecule or group is an unfavorable process. Placing two nonpolar molecules or groups into water would require water molecules to form two highly ordered cages. If the nonpolar molecules or groups were close together, a single water cage would suffice. For strictly geometric reasons, the number of water molecules required to form a single cage around the pair of nonpolar molecules or groups is less than half

(a) Water molecules in bulk phase of water



(b) Water molecules in cage around hydrocarbon

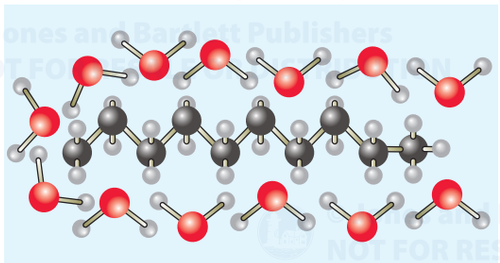


FIGURE 2.26 The hydrophobic effect. (a) Water molecules move randomly in the absence of nonpolar (hydrophobic) molecules. (b) Water molecules form ordered cages around hydrophobic molecules or groups. Because disorder is favored over order, the most favorable situation is one that uses the fewest possible water molecules to form cages. For strictly geometric reasons, the number of water molecules required to form a cage around a group of hydrophobic molecules is less than the number required to form separate cages around each molecule. Therefore, hydophobic molecules (or groups) tend to aggregate in water. Adapted from Tropp, B. E. *Biochemistry: Concepts and Applications.* Brooks/Cole Publishing Company, 1997.

the total number needed to form a cage around each molecule or group separately. In general, the number of ordered water molecules per nonpolar molecule or group is always smaller if the nonpolar molecules or groups are clustered or stacked. Hence, clustering or stacking of nonpolar molecules or groups is thermodynamically favored because a cluster or stack requires fewer water molecules to be arranged in highly ordered cages. Note that no bonds are formed; it is only that a cluster is the more probable arrangement. One can obtain an indication of the tendency of molecules or groups to aggregate in water, known as **hydrophobicity**, from their tendency to transfer from water to a nonpolar solvent. Many amino acid side chains are hydrophobic. For example, the hydrocarbon chains of alanine, leucine, isoleucine, and valine tend to form clusters and the aromatic side chains in tyrosine and tryptophan tend to form stacks in water. When interacting hydrophobic side chains are present on residues that are far apart, they bring distant parts of the polypeptide chain together. Thus, folding patterns of polypeptides are strongly influenced by hydrophobic interactions among the side chains of amino acid residues. Dozens of hydrophobicity scales have been proposed for amino acid side chains, each based on slightly different experimental data and theoretical principles. Unfortunately, values provided by each of the scales are only approximations. One widely used scale devised by J. Kyte and R. F. Doolittle is shown in **FIGURE 2.27**.

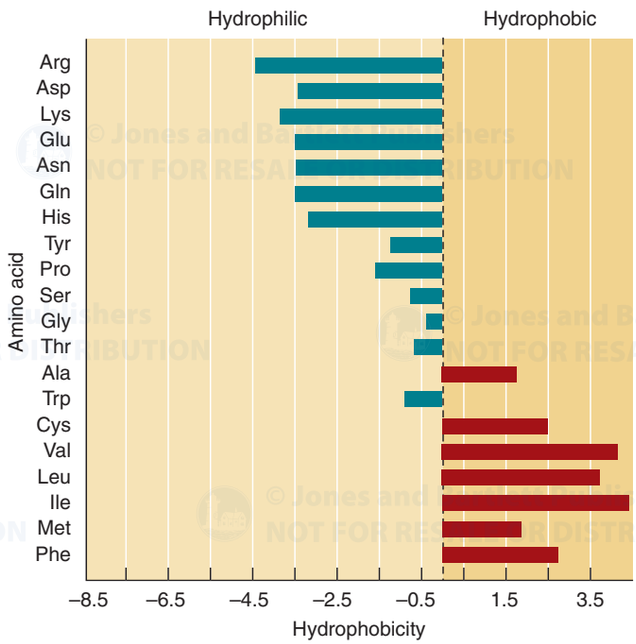


FIGURE 2.27 Hydrophobicity scale. Many hydrophobicity scales have been devised. The one shown here, which was devised by J. Kyte and R.F. Doolittle, is based on experimental data.

2.6 Secondary Structures

The α-helix is a compact structure that is stabilized by hydrogen bonds.

Adjacent residues on a polypeptide chain can fold into regular secondary structures. Linus Pauling predicted the existence of the first secondary structure, the α-**helix**, in 1951 while recovering from an illness at home. To occupy his time, Pauling drew a short peptide with correct bond angles and bond lengths on a piece of paper. Upon creasing the paper, he noticed that the peptide backbone folded into a helix. When he returned to the laboratory, Pauling and his colleague Robert Corey constructed a more accurate three-dimensional model that they named the α-helix. The model (**FIGURE 2.28**) has the

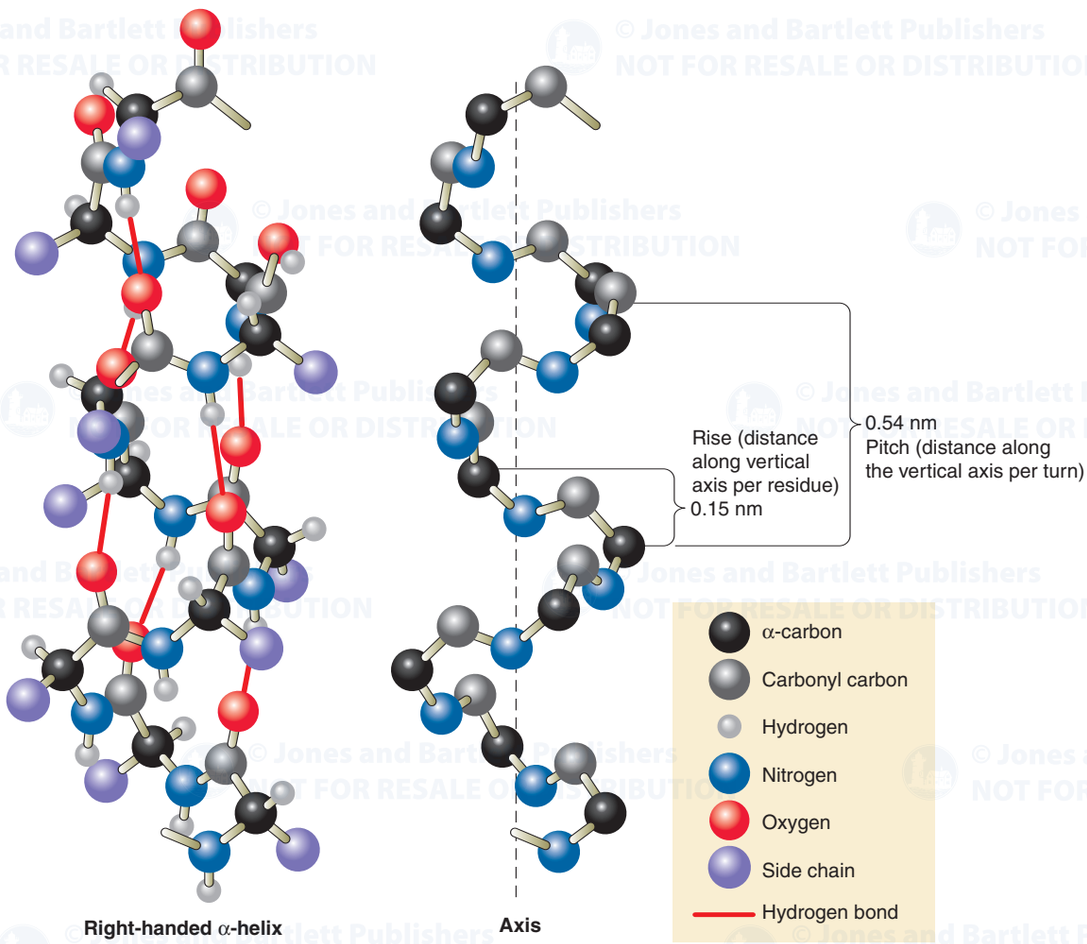


FIGURE 2.28 The α-helix. The N-terminal residue is at the bottom and the C-terminal residue at the top of the figure. The oxygen in each carbonyl group forms a hydrogen bond with the amide hydrogen that is four residues more toward the C-terminus, with the hydrogen bond approximately parallel to the long axis of the helix. All carbonyl groups point toward the C-terminus. In an ideal α-helix, equivalent positions reappear every 0.54 nm (the pitch of the helix), each amino acid residue advances the helix by 0.15 nm along the axis of the helix (the rise), and there are 3.6 amino acid residues per turn. In a right-handed helix the backbone turns in a clockwise direction when viewed along the axis from its N-terminus. Modified from Horton, R. H., et al. *Principles of Biochemistry,* 3/e. Prentice Hall, 2002.

following features: Each complete helical turn extends 0.54 nm along the vertical axis and requires 3.6 amino acid residues. The vertical rise per residue is 0.15 nm/residue (0.54 nm/3.6 residues).

Hydrogen bonds are located inside the helix, forming a regular repeating pattern. The oxygen atom on the carbonyl group of residue n forms a hydrogen bond with the hydrogen atom of the N-H group of residue n+4. The first and last four residues in the helix cannot form a full set of hydrogen bonds because hydrogen bond partners are not available. All C=O bonds point in one direction and all N-H bonds point in the opposite direction, producing a significant net dipole for the α-helix with a partial positive charge at the amino end and a partial negative charge at the carboxyl end (**FIGURE 2.29**). These electrostatic charges are often offset by acidic residues at the amino end of the helical segment and basic residues at the carboxyl end of the helical segment.

The α-helix is a right-handed helix. The right-handed nature of the helix is most easily visualized by picturing the threads of a screw. In a right-handed screw threads are shaped so that clockwise rotation produces tightening. However, the threads could also be in a left-handed form so that counterclockwise rotation produces tightening. Naturally occurring polypeptide chains, which are made of L-amino acids, fold into right-handed α-helices. When synthetic polypeptides are made of D-amino acids, the polypeptides fold into left-handed helices.

All amino acid residues can fit in the α-helix but they differ in their propensity to do so. Proline and glycine have the least tendency to fit into a helix. The nitrogen atom in a proline that is part of a polypeptide lacks the substituent hydrogen atom needed to form a

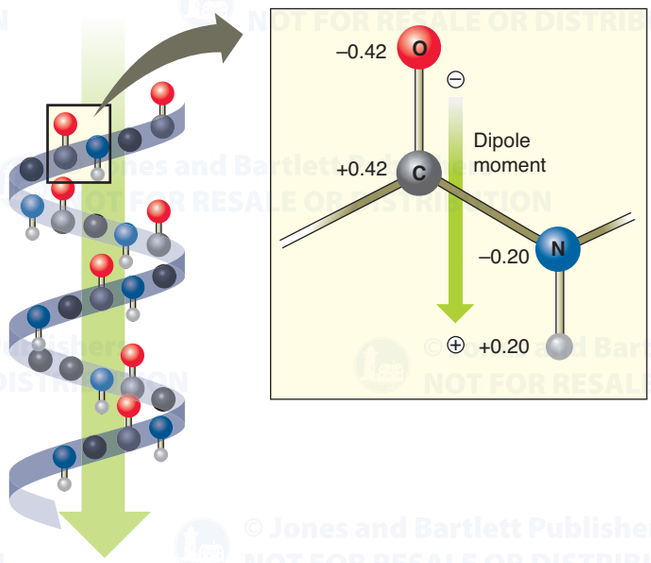


FIGURE 2.29 Macroscopic dipole in the α-helix. The individual dipole moments of the N-H and C=O groups along the helical axis generate a large net dipole for the helix. The numbers shown indicate fractional electric charges on the respective atoms. Adapted from Garrett, R. H., and Grisham, C. M. *Biochemistry (with BiochemistryNow and InfoTrac), 3/e.* Brooks/Cole Publishing Company, 2005.

hydrogen bond. Moreover, the proline side chain is rigid and so does not easily fit into the α-helix. Proline is occasionally present in the first helical turn, where its side chain geometry and inability to hydrogen bond do not create a problem. Glycine's tendency not to be part of an α-helix results from an entirely different problem. Glycine has great conformational freedom because its side chain, a single hydrogen atom, is so small. Furthermore, a single hydrogen atom is insufficient to protect the hydrogen bonds inside the helix from disruption by water. Certain combinations of adjacent residues also tend to disrupt or break the helical structure. When neighboring residues have bulky side chains, the side chains make physical contact that prevents them from fitting properly. For instance, neighboring isoleucine, tryptophan, and tyrosine residues would disrupt the helical structure. A run of positively-charged or negatively-charged side chains repel one another, destabilizing the helix. This phenomenon is best illustrated by comparing polyglutamate and polylysine structures at different pH values. Polyglutamate exists as an α-helix at pH 2 but not at pH 7 or pH 10. The explanation for this behavior is that the carboxylic acid side chains are uncharged at pH 2 but have negative charges at pH 7 and pH 10. Side chain repulsion at pH 7 and pH 10 prevents helix formation.

In contrast, polylysine exists as an α-helix at pH 10 but not at the two lower pH values. Once again the explanation is charge repulsion. The positively-charged lysine side chains repel one another at pH 2 and pH 7. Although the α-helix is the most common helix present in proteins, it is not the only one. For instance, the so-called 3,10 helix, which is much less common, has hydrogen bonds between residue n and n+3.

A few fibrous proteins, most notably α-keratin the major protein in hair, have no residues or combination of residues that disrupt the α-helix and so have a single regular structure throughout. However, most proteins do have helix breakers and so have helical segments surrounded by nonhelical segments. When present in a water-soluble protein, α-helical segments tend to be on the outside of the protein with one side facing the aqueous medium and the other the hydrophobic interior. The side facing the aqueous medium tends to have hydrophilic (acidic, basic, or polar) residues while the side facing the interior has mostly hydrophobic residues. Because the α-helix has 3.6 residues per turn, such an arrangement can be achieved by switching from hydrophobic to hydrophilic side chains with a 3 or 4 residue periodicity.

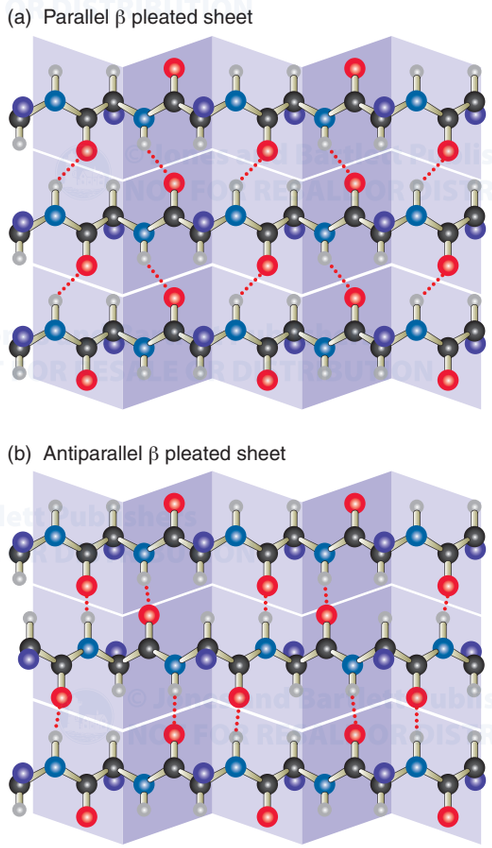


FIGURE 2.30 The β-conformation of polypeptide chains. Polypeptide chains with β-conformations can line up side by side so that C=O and N-H groups on adjacent chains interact through hydrogen bonds to produce an almost flat structure called a β-sheet. The polypeptide chains' zigzag structure gives the β-sheet the appearance of being pleated with consecutive C-α atoms slightly above or below the plane of the sheet. The sheets can be organized so that (a) all peptide chains have the same amino to carboxyl direction—parallel pleated sheets or (b) successive polypeptide chains are oriented in opposite directions—antiparallel pleated sheets. Adapted from Nelson, D. C., and Cox, M. M. *Lehninger Principles of Biochemistry.* W. H. Freeman & Company, 2004.

The β-conformation is also stabilized by hydrogen bonds

Pauling and Corey also predicted the existence of a second type of secondary structure, the β-conformation, in which the polypeptide chain is almost fully extended. Polypeptide chains with β-conformations can line up side by side so that C=O and N-H groups on adjacent chains interact through hydrogen bonds to produce an almost flat structure called a β-sheet (**FIGURE 2.30**). The polypeptide chains'

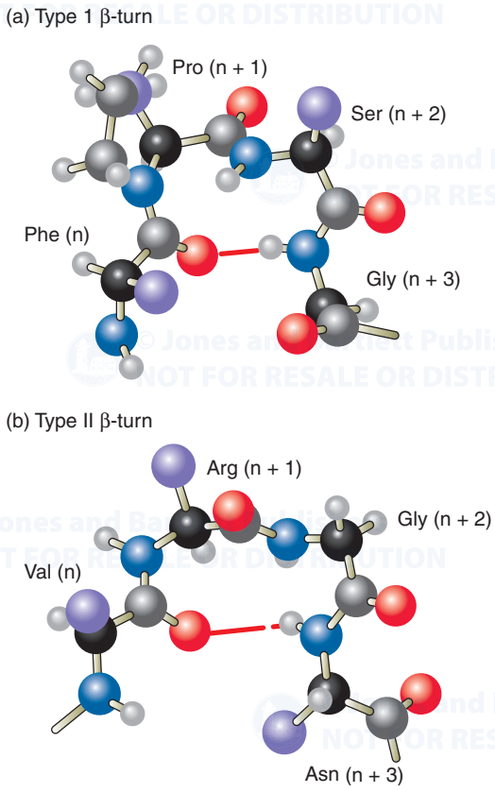


FIGURE 2.31 Two types of β-turns. The two most commonly occurring β-turns are called (a) type I and (b) type II. Type I β-turns occur with about twice the frequency of type II β-turns. In each case there is a hydrogen bond between residue n and residue n+3. Type II β-turns always have a glycine at position n+2.

zigzag structure gives the β-sheet the appearance of being **pleated** with consecutive C_α atoms slightly above or below the plane of the sheet. The sheets can be organized so that all polypeptide chains have the same amino to carboxyl direction, **parallel pleated sheets** (Figure 2.30a), or so that successive polypeptide chains are oriented in opposite directions, **antiparallel pleated sheets** (Figure 2.30b). In parallel pleated sheets, hydrogen bonds are evenly spaced and at angles to the long axes of the polypeptide chain. Hydrophobic side chains are present on both sides of the parallel pleated sheet. In antiparallel pleated sheets, pairs of narrowly-spaced hydrogen bonds alternate with pairs of more widely-spaced hydrogen bonds, with all hydrogen bonds perpendicular to the long axes of the polypeptide strands. The polypeptide chains that comprise antiparallel pleated sheets tend to have alternating hydrophilic and hydrophobic residues, so that hydrophobic side chains tend to be present on one side of the sheet and hydrophilic residues on the other.

Loops and turns connect different peptide segments, allowing polypeptide chains to fold back on themselves.

The average globular protein has a diameter of about 2.5 nm, corresponding to about 11 residues in an α-helix and only seven residues in the extended β-conformation. Secondary structures known as loops and turns connect α-helical and β-strand segments within a protein, allowing the polypeptide backbone to fold back upon itself and reverse direction. Loops and turns usually are on the protein surface, extending into the surrounding aqueous environment. They therefore tend to be made of hydrophilic residues but also exploit the special conformational properties of glycine and proline residues to reverse direction. Loops contain five or more residues, tend to be quite flexible, and lack a defined structure. Turns, which are smaller, usually have defined structures.

The most common kind of turn, the β-turn, consists of four residues and allows the polypeptide chain to reverse direction. The carbonyl oxygen of the first residue in a β-turn forms a hydrogen bond with the amino group of the fourth residue. The peptide bonds of the middle two residues do not interact through hydrogen bonds.

Glycine and proline are commonly present in β-turns. Glycine's conformational flexibility allows it to fit into the tight turn. Proline's steric constraints are also well suited to the β-turn. Two of the most common types of β-turns are shown in **FIGURE 2.31**. β-Turns are usually present near the surface of a protein, allowing the peptide groups of the central two amino acid residues to interact with the surrounding water molecules.

Certain combinations of secondary structures, called supersecondary structures or folding motifs, appear in many different proteins.

Folding patterns for a few thousand different proteins are now available. Although at first glance these patterns look unique, more care-

ful analysis reveals that some unifying principles do exist. One of the most important of these is that certain combinations of secondary structures, called **supersecondary structures** or **folding motifs**, are present in many different proteins. Supersecondary structures are represented by a schematic topology diagram in which β-strands are represented by arrows pointing from the amino to the carboxyl terminus, α-helices by cylinders or helical structures, and loops and turns by ribbons. Some common supersecondary structures are shown in **FIGURE 2.32**.

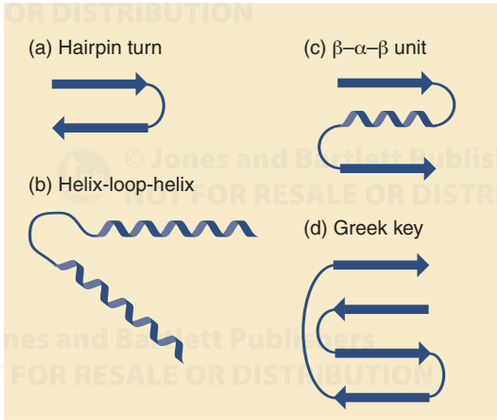


FIGURE 2.32 A sample of supersecondary protein structures.

We cannot yet predict secondary structures with absolute certainty.

Examination of amino acid side chain structures provides helpful insights into the way that a residue may fit into a secondary structure. For example, the methyl group in alanine fits well into an α-helix, whereas the larger isopropyl group in valine experiences some steric hindrance. Additional insights come from comparing folding patterns of many different polypeptides. A statistical analysis of data from such studies allows us to predict the likelihood that a particular residue will be in an α-helix or β-strand (**FIGURE 2.33**). However, residues do not always behave as expected. For example, studies by Daniel L. Minor, Jr. and Peter S. Kim in 1996 showed that an 11-amino acid sequence (Ala-Trp-Thr-Val-Glu-Lys-Ala-Phe-Lys-Thr-Phe) folds into an α-helix when in one position in the primary sequence of a polypeptide but as a β-sheet when in another position. Thus, peptide sequences can form different secondary structures when placed in different protein contexts.

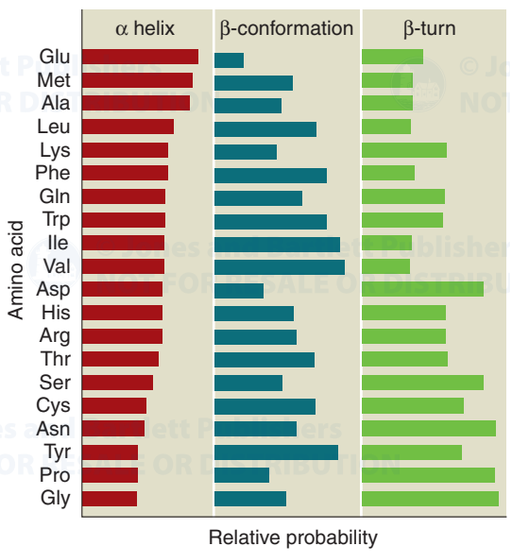


FIGURE 2.33 Relative probabilities that a given amino acid will occur in the three common types of secondary structure. Adapted from Nelson, D. C., and Cox, M. M. *Lehninger Principles of Biochemistry.* W. H. Freeman & Company, 2004.

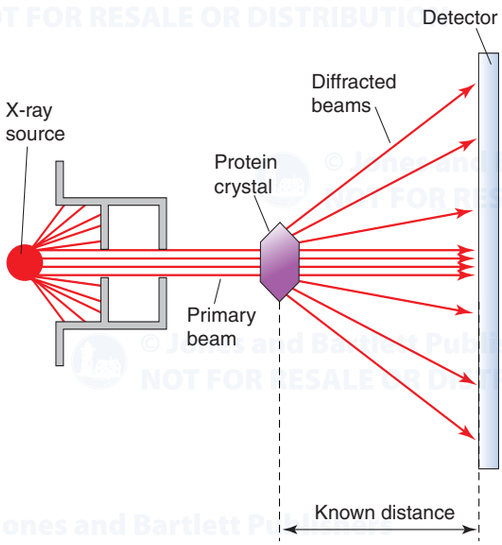


FIGURE 2.34 Schematic of x-ray crystallography experiment. When a beam of x-rays (red) hits a crystal, most of the electromagnetic radiation passes right through the crystal. The crystal scatters or diffracts the remaining light in many different directions. The diffracted x-rays produce a characteristic pattern of spots on a film or detector that is placed behind the crystal. Distances between spots and spot intensities provide the necessary information for determining protein structure. Adapted from Branden, C. I., and Tooze, J. *Introduction to Protein Structure, 2/e.* Routledge, 1999.

2.7 Tertiary Structure

X-ray crystallography and nuclear magnetic resonance studies have revealed the three-dimensional structures of many different proteins.

It is much more difficult to determine the three-dimensional structure of a globular protein that contains a combination of secondary structures than it is to determine the three-dimensional structure of a regular protein that is all α-helix or all β-conformation. Two physical methods, **x-ray crystallography** and **nuclear magnetic resonance (NMR) spectroscopy**, are used to elucidate the three-dimensional structure of proteins and nucleoproteins (protein-nucleic acid complexes) at atomic detail. A thorough examination of these two techniques is beyond the scope of this book, but a brief overview of how x-ray crystallography and NMR are used to study protein structure is provided.

X-Ray Crystallography

An ordinary optical instrument such as a light microscope does not permit us to see proteins in atomic detail because the distances between atoms are too small. In general, the wavelength required to resolve two objects (recognize the two objects as distinct entities) must be less than half the distance between the objects. Because distances between atoms linked by covalent bonds are about 0.15 nm, molecular resolution requires the very short wavelengths in x-rays. However, x-rays do not provide a direct image of protein molecules because currently available lenses and mirrors cannot focus such short wavelengths. X-ray crystallography allows interatomic distances in proteins to be measured by exploiting the fact that the electrons surrounding atoms in a crystalline protein scatter or diffract x-rays. The diffracted x-rays produce a characteristic pattern of spots on a film or detector that is placed behind the protein crystals (**FIGURE 2.34**). A heavy metal such as uranium is attached to a specific residue without altering protein structure to provide a reference point for data interpretation. Distances between spots and spot intensities provide the necessary information for determining protein structure.

The power of x-ray diffraction crystallography became apparent for all to see in 1957, when John Kendrew used the technique to determine the structure of myoglobin, an oxygen storage protein that is present in high concentrations in the muscles of diving mammals. A short time later, Max Perutz reported the three-dimensional structure of hemoglobin, the oxygen transport protein in red blood cells. Many improvements have been made in x-ray crystallography since the structures of myoglobin and hemoglobin were first reported. Three advances in particular are especially noteworthy: (1) faster computers facilitate data analysis; (2) high-intensity x-ray beams emanating from synchrotrons allow investigators to study protein crystals that are much smaller than those used in earlier studies of protein structure; and (3) synchrotron radiation at multiple wavelengths elimi-

nates the need to attach heavy metals to specific sites in the protein. These advances have greatly accelerated the rate at which protein crystal structures can be solved. Thanks to these technological advances, much less time is required to solve protein structure problems today. Recombinant DNA techniques have also played a critical role by allowing investigators to construct cells that produce large quantities of a desired protein.

Today, approximately 37,000 protein crystal structures are available at the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) and new structures are being added each day. The data are entered as a Protein Data Bank ("pdb") file. A program that serves as a viewer allows us to convert the pdb file into a three-dimensional image that can be manipulated on the computer. One excellent viewer, ViewerLite Software® (from Accelrys), permits the protein structure to be displayed in different forms and colors. **FIGURE 2.35** shows the crystal structure of pancreatic ribonuclease A (RNase A), an enzyme that digests ribonucleic acids, in a spacefill display, a stick display, and a ribbon display.

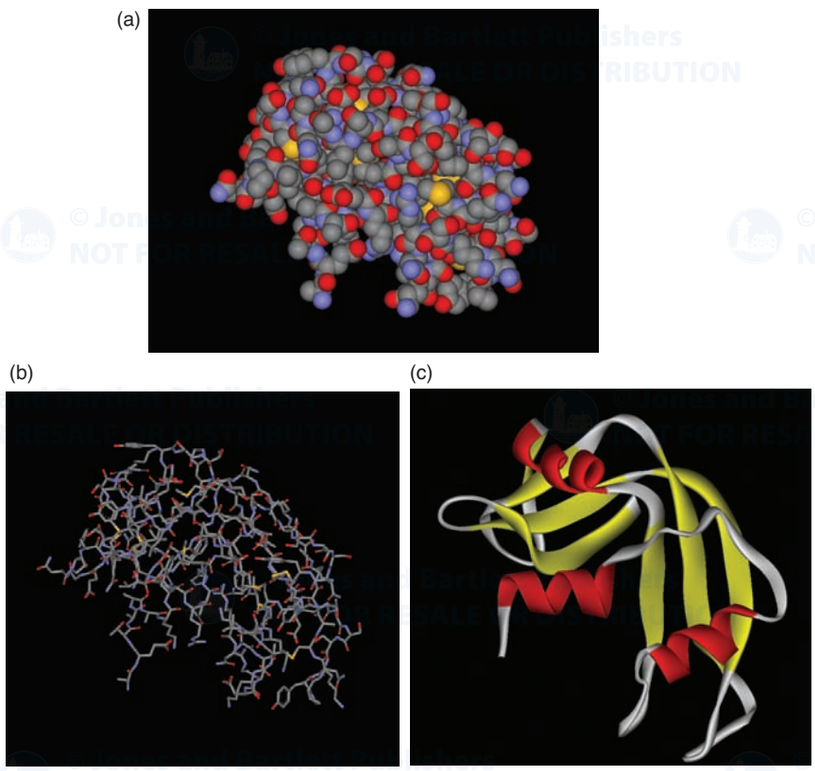


FIGURE 2.35 Crystal structure of ribonuclease A. (a) Crystal structure displayed in space filled form. The colors are in standard CPK (Corey, Pauling, Kultin) color scheme. Carbon is gray, hydrogen white, nitrogen blue, and sulfur yellow. (b) Crystal structure displayed in stick form. The standard CPK color scheme is used. The orientation is as in (a). (c) Crystal structure shown in ribbon form. α-helices are shown in red and β-conformations in yellow. The protein orientation is as in (a). Adapted from Vitagliano, L, *Proteins* 46 (2002): 97–104.

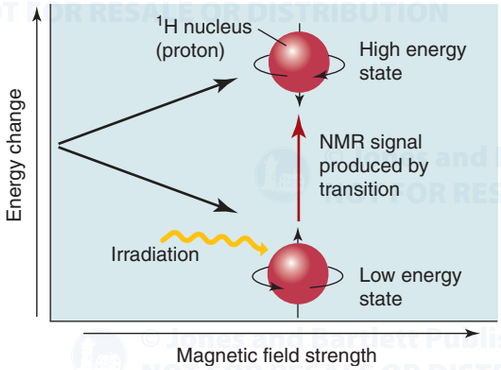


FIGURE 2.36 Basis of NMR spectroscopy. When a strong magnetic field is applied to proteins in solution, ^{1}H, ^{13}C, and ^{15}N nuclei line up parallel to the field (low energy) or antiparallel to the field (high energy). The energy difference between the two orientations is comparatively small so that the low energy state is populated by only slightly more protons than the high energy state. A pulse of electromagnetic radiation with an energy that exactly matches the energy difference between the two spin states can raise the lower energy state to the higher energy state. Adapted from Berg, J. M., et al. *Biochemistry, 5/e.* W. H. Freeman and Company, 2002.

Despite its enormous power, x-ray crystallography does have some shortcomings. First and foremost, proteins that do not form crystals cannot be examined by this technique. Second, protein molecules pack close together when they form a crystal. Sometimes this packing causes residues on the protein surface to assume positions that are slightly different from their position in solution. Third, some proteins have segments that are highly disordered, so the structures of these segments cannot be determined by x-ray crystallography. Initially there was concern that protein crystal structures might differ from protein structures in aqueous solution. However, it is now clear that protein crystals contain considerable water, which allows proteins to retain their biological activity. Furthermore, structures determined by NMR spectroscopy of proteins in solution are the same as the crystal structures.

Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy has been used to elucidate the structure of relatively small organic molecules since the 1950s. Thanks to the pioneering efforts of Kurt Wüthrich beginning in the 1980s, this physical technique can now also be used to determine the three-dimensional structure of proteins, provided that the protein has a molecular mass of less than 40 kDa. As the technology improves, it is likely that NMR spectroscopy will be able to determine structures of larger proteins. The protein under study must be in pure form and sufficiently soluble in water so that a concentration of at least 1 mM can be achieved. NMR spectroscopy offers two major advantages: (1) proteins can be studied in solution in an environment similar to that in the living cell, and (2) the proteins do not need to form crystals.

NMR takes advantage of the fact that certain atomic nuclei such as ^{1}H, ^{13}C, and ^{15}N are intrinsically magnetic and display a property called magnetic spin that results in the generation of a magnetic dipole. In essence, ^{1}H, ^{13}C, and ^{15}N nuclei act as tiny magnets. The hydrogen nucleus (^{1}H) is especially important in NMR spectroscopy structural studies of proteins because ^{1}H atoms are distributed throughout naturally occurring proteins. C-13 and N-15 isotopes have to be introduced by culturing cells in medium containing nutrients enriched for these isotopes. When a strong magnetic field is applied to proteins in solution, the H^1 nuclei line up parallel to the field (low energy) or antiparallel to the field (high energy) (**FIGURE 2.36**). The energy difference between these two orientations is comparatively small, so that the low energy state is populated by only slightly more protons than the high energy state. Because the solution contains a large population of identical protein molecules the effect is amplified. A spinning ^{1}H nucleus in the lower energy state can be raised to the higher energy state by applying a pulse of electromagnetic radiation in the radio-frequency range with an energy that exactly matches the energy difference between the two spin states. This radio-frequency is said to be in **resonance** with the proton when its energy exactly matches that required to convert the lower energy spin state to the higher energy spin state. The transition energy required to

induce a spin transition for a specific ^1H nucleus is influenced by the spin of protons that are covalently connected by only one or two other atoms. This type of interaction is called "interaction through bonds" (**FIGURE 2.37a**). The transition energy of a proton in a peptide is also influenced by protons that are located several hundred bonds away, provided that the peptide folds so that the interacting protons are within 5 Å of one another. This type of interaction is called *interaction through distance* (**FIGURE 2.37b**). Data obtained from NMR spectroscopy allow investigators to estimate the distances between specific pairs of distant atoms. The resulting set of distances, together with the amino acid sequence and known geometric constraints such as bond angles and distances, group planarity, stereoconfiguration, and van der Waals radii, are used to compute the proteins' three-dimensional structure. Because distances between proton pairs are not precise, one obtains an ensemble of very similar structures rather than a single structure. The structure obtained for ribonuclease A by NMR spectroscopy, shown in **FIGURE 2.38**, is remarkably similar to the crystalline structure for this protein (Figure 2.35a).

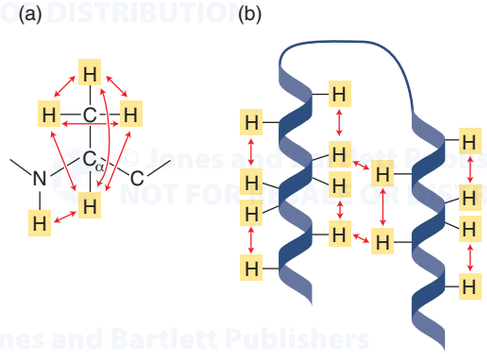


FIGURE 2.37 Two types of NMR information. The transition energy required to induce a spin transition for a specific ^1H nucleus is influenced by (a) the spin of protons that are covalently connected by only one or two other atoms (*interaction through bonds*) and (b) protons that are located several hundred bonds away provided that the polypeptide folds so that the interacting protons are within 5 Å of one another (*interaction through distance*). Adapted from Branden, C. I., and Tooze, J. *Introduction to Protein Structure, 2/e.* Routledge, 1999.

The primary structure of a polypeptide determines its tertiary structure.

Heat and certain chemical agents such as acids, detergents, and urea (NH_2CONH_2) cause proteins to lose their biological activity by disrupting the weak non-covalent bonds that stabilize secondary and tertiary structures. Before the weak non-covalent bonds are disrupted, the protein is said to be in its **native state**. After disruption, the protein is said to be in a **denatured state**, existing as a mixture of **random conformations.** Hydrophobic segments that are normally buried in the core of water-soluble proteins become exposed after denaturation and bind to hydrophobic segments of other denatured proteins to form water-insoluble aggregates. Protein denaturation is a rather common phenomenon. For example, milk spoils as a result of bacterial growth that produces acidic waste products, which cause the milk to have a sour taste and curdle. The curd is composed of denatured milk proteins that have become water-insoluble.

Some polypeptides can refold to their native state after denaturation. Studies of this process known as **renaturation** have provided a great deal of information about the folding process. In the mid-1950s Christian Anfinsen and colleagues selected bovine pancreatic ribonuclease A (RNase A) as a model protein for studying *in vitro* renaturation (**FIGURE 2.39**). Eight of the 124 residues in the polypeptide chain are cysteines that pair to form four disulfide bonds. These disulfide bonds help to lock the tertiary structure into place, making it very difficult to denature the enzyme. RNase A can be denatured by simultaneous treatment with urea and β-mercaptoethanol (CH_2OHCH_2SH). Neither chemical agent by itself is sufficient to denature RNase A. Urea interferes with hydrophobic interactions and disrupts hydrogen bonds, allowing the β-mercaptoethanol to gain access to the disulfide bonds. β-Mercaptoethanol disrupts the disulfide

FIGURE 2.38 High-resolution three-dimensional structure of ribonuclease A in solution by nuclear magnetic resonance spectroscopy.

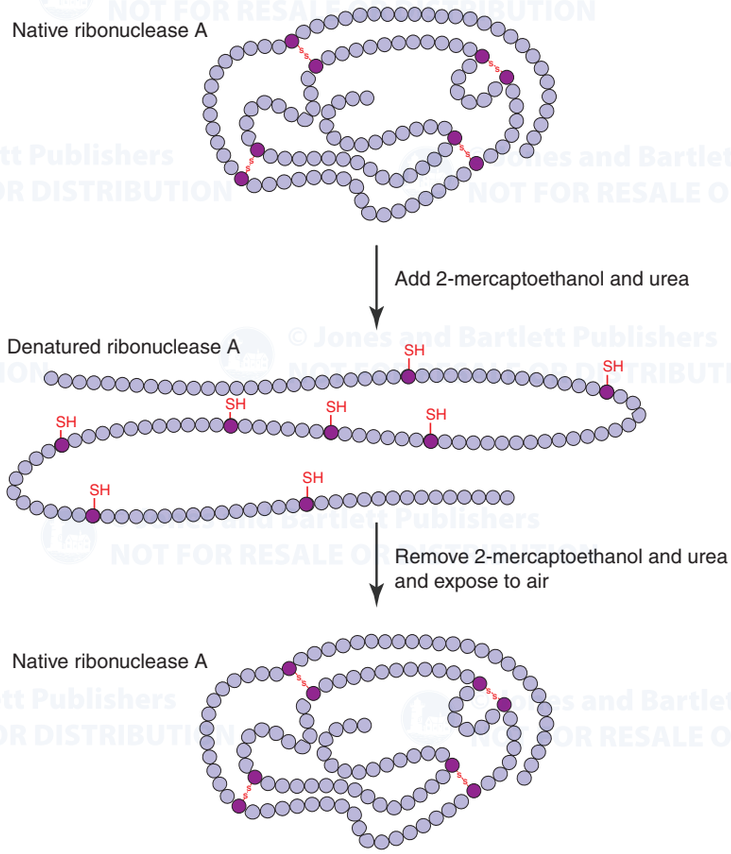


FIGURE 2.39 Denaturation and renaturation of RNase A. Denature pancreatic RNase A with urea in the presence of 2-mercaptoethanol. Renature enzyme by dialyzing denatured RNase to remove the urea and 2-mercaptoethanol, and then exposing the polypeptide chain to air to re-form disulfide bonds. Adapted from Moran, L. A. and Scrimgeour, K. G. *Biochemistry, 2/e.* Neil Patterson Publishers, 1994.

bonds, permitting further urea denaturation. If urea and the mercaptoethanol are slowly removed, perfect renaturation occurs, including formation of the four correct disulfide bonds (disulfide bonds can form spontaneously by oxidation in air). This latter finding is remarkable because there are 105 possible ways that eight cysteine residues can combine to form four disulfide bonds. Anfinsen's interpretation of this experiment was that the folding of RNase A is determined exclusively by its amino acid sequence and that the proper disulfide bonds are formed because, during folding, the cysteine residues are correctly placed for joining. Evidence that disulfide bond formation does not direct the folding comes from an experiment in which the β-mercaptoethanol was removed first and oxidation was allowed to occur prior to removal of the urea, that is, while the RNase A was a random coil. With this protocol, the native molecule was not formed.

The Anfinsen experiment shows that, at least for ribonuclease A,

the tertiary structure is determined by the amino acid sequence and is the one with the lowest energy. Similar observations have been made for many other proteins, but not for all. The fact that denatured RNase can fold to form its native structure is remarkable in view of the number of possible conformations. If we assume that each residue has ten possible conformations available to it, then the 124 residues would have to sample 10^{124} different conformations before achieving the correct one. If the polypeptide chain had to sample each possible conformation, the folding process could not possibly take place during the organism's lifetime. This problem, first recognized by Cyrus Levinthal, can only be solved if the polypeptide does not need to sample all possible conformations. One hypothesis is that local regions of secondary structure form rapidly, causing the polypeptide chain to fold into a compact structure, called a **molten globule**, that has many of the correct secondary structure features but that requires further folding to form the correct tertiary structure.

In addition to its great theoretical significance, the Anfinsen experiment also had practical applications. At the time of the Anfinsen experiment, no one had yet managed to synthesize a polypeptide chain as long as ribonuclease. Some investigators questioned whether it would be worth the effort, because a synthetic polypeptide might not be able to fold into a biologically active form outside of the cell. The Anfinsen experiment showed that polypeptides prepared in the laboratory would have a reasonable chance of folding into biologically active proteins, stimulating efforts to synthesize long polypeptides. Although it is now possible to synthesize polypeptides in the laboratory by standard organic chemical techniques, the more common practice is to isolate or synthesize DNA that codes for the desired protein and then introduce that DNA into a living cell so that it directs the synthesis of the protein.

Molecular chaperones help proteins to fold inside the cell.

Protein folding is a much more complex process inside the living cell, which has high concentrations of proteins and other macromolecules. Instead of interacting with one another to form the correct tertiary structure, hydrophobic segments in an unfolded polypeptide may interact with hydrophobic segments in other unfolded polypeptide chains to form biologically inactive aggregates. The tendency to form aggregates is increased by the fact that stable tertiary structure formation usually requires a complete folding unit (100–300 amino acid residues). However, the complete folding unit does not emerge from the protein synthetic factory, the ribosome, in a single discrete step. Instead, the folding unit slowly emerges from the ribosome as each succeeding amino acid is added to the growing end of the polypeptide chain. Correct folding takes place only after the entire folding unit has emerged from the ribosome. Exposure of hydrophobic segments on the growing polypeptide would make them susceptible to intermolecular aggregation were it not for the presence of specific proteins called **molecular chaperones** that help to stabilize

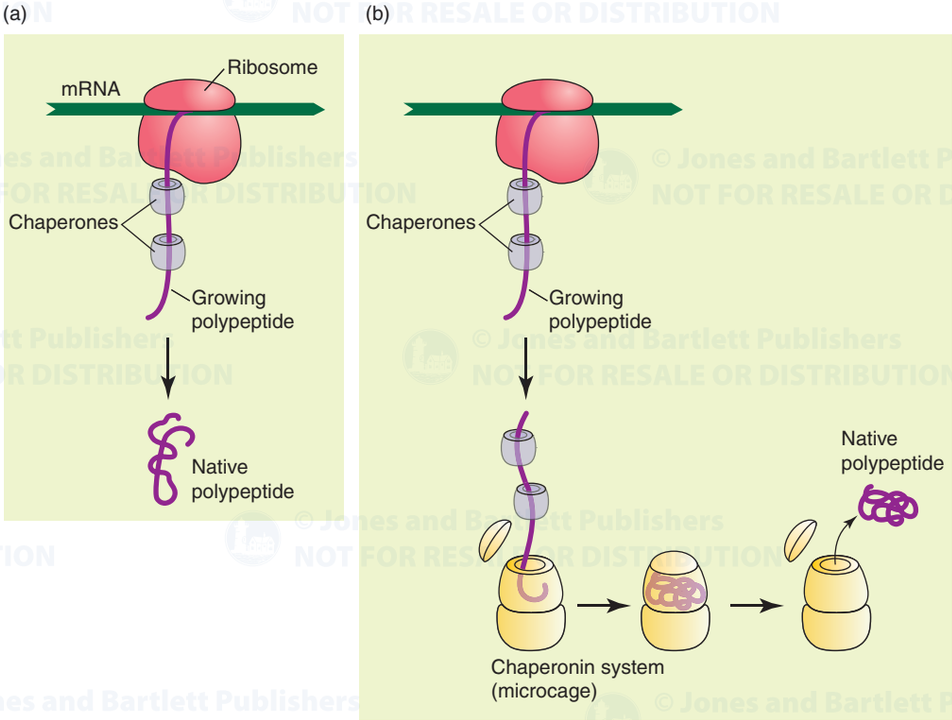


FIGURE 2.40 Protein folding in the cell. Chaperones assist polypeptides to fold as the polypeptides emerge from the ribosome. Chaperones bind to the hydrophobic segments as they emerge from the ribosome and prevent the hydrophobic segment from interacting with hydrophobic segments from other unfolded polypeptides. The chaperones are released after the entire folding unit has been extruded, freeing the polypeptide chain to (a) fold into its native state or (b) enter a microcage called a *chaperonin system* in which the polypeptide can fold without interference from other polypeptides. Adapted from Young, J. C., et al., *Nat. Rev. Mol. Cell Biol.* 5 (2004): 781–791.

the emerging polypeptide until the entire folding unit has been extruded by the ribosome. The chaperones assist the polypeptide to fold as it emerges from the ribosome (**FIGURE 2.40**). Some chaperones work by binding to hydrophobic segments as they emerge from the ribosome, which prevents the hydrophobic segment from interacting with hydrophobic segments from other unfolded polypeptides. Molecular chaperones are released after the entire folding unit has been extruded freeing the polypeptide chain to fold into its native state or to enter a microcage called a **chaperonin system** in which the polypeptide is free to fold without making contact with other unfolded polypeptides, thereby preventing aggregation.

Two enzymes, **protein disulfide isomerase** and **peptidylprolyl *cis-trans* isomerase**, speed the polypeptide folding process in cells (**FIGURE 2.41**). Protein disulfide isomerases catalyze thiol/disulfide interchange, facilitating the formation of the correct set of disulfide bonds by reshuffling disulfide bonds when incorrect pairings are formed. Peptidylprolyl *cis-trans* isomerase catalyzes the isomerization of amino acid-proline peptide bonds, accelerating the refolding of polypeptide chains that contain proline.

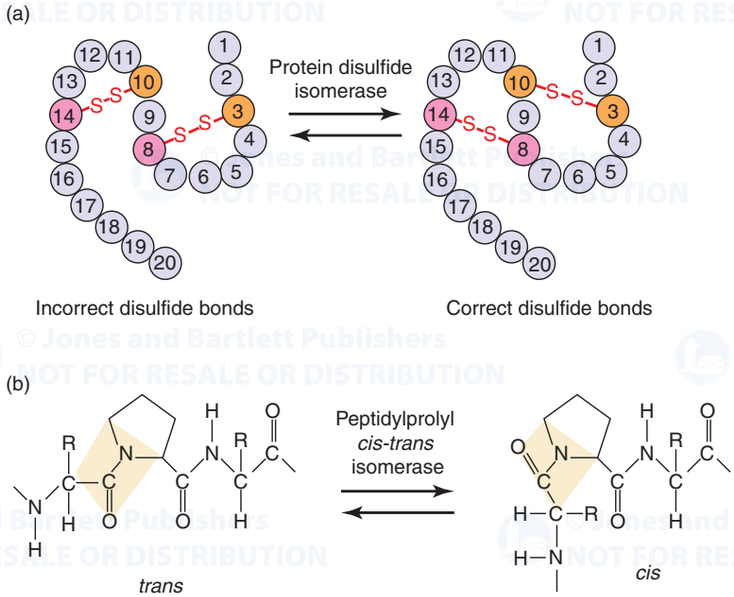


FIGURE 2.41 Enzymes that assist polypeptide folding. (a) Protein disulfide isomerase. (b) Peptidylprolyl *cis-trans* isomerase. Adapted from Tropp, B. E. *Biochemistry: Concepts and Applications.* Brooks/Cole Publishing Company, 1997.

2.8 Proteins and Biological Membranes

Proteins interact with lipids in biological membranes.

Until this point, we have been considering only those proteins that are water-soluble. However, many important proteins are present in biological membranes. The remainder of this chapter briefly examines the way that these proteins are organized in membranes, which contain lipids as their second major component. A lipid is defined as a biological molecule that is soluble in an organic solvent such as chloroform but not in water. The most notable feature of nearly all biological membranes is that they consist of two layers, called a **bilayer.** This structure is a consequence of the chemical nature of lipids, the water-insoluble molecules that form the bilayer. Three major lipid families, the glycerophospholipids, the sphingolipids, and the sterols are commonly found in biological membranes. We focus on the glycerophospholipids (**FIGURE 2.42**), which are glycerol-3-phosphate derivatives, because they are the only lipids present in all biological membranes. Members of this family are named as derivatives of phosphatidate, a lipid that is usually present in membranes in only trace amounts.

When placed in aqueous solution, glycerophospholipids aggregate to form bilayers. An examination of phosphatidylcholine structure shows why this is so (**FIGURE 2.43**). Each phosphatidylcholine molecule has a polar head group, phosphocholine, and two nonpolar hydrocarbon tail groups (Figure 2.43a). The two tails usually contain different numbers of carbon atoms (range, 14–24) and hence have different lengths. One tail typically has one or more *cis* double bonds, each of

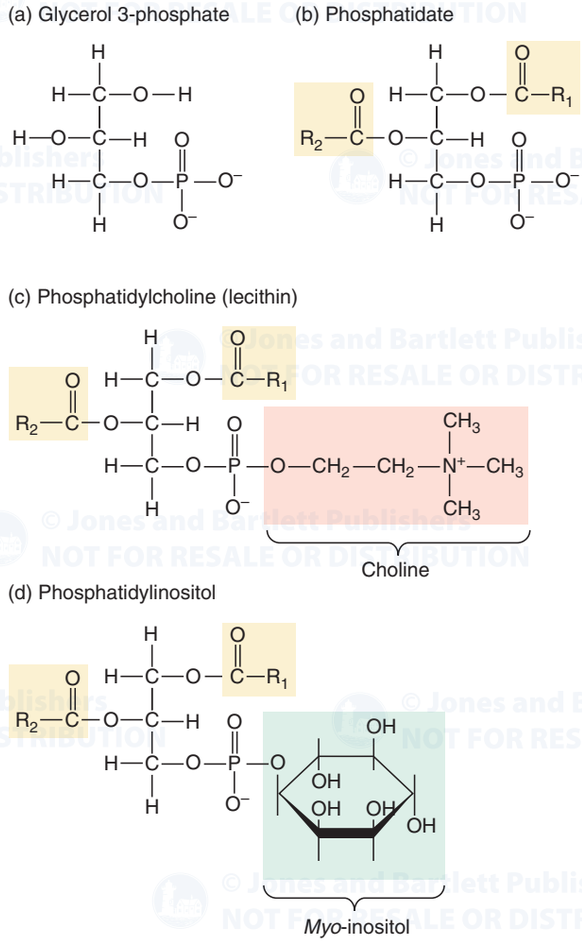


FIGURE 2.42 Some glycerophospholipids. (a) Glycerol-3-phosphate is the building block for glycerophospholipids. (b) Phosphatidate is usually present in membranes in only trace amounts but other glycerophospholipids are named as its derivatives. (c) Phosphatidylcholine (lecithin) is present in eukaryotic membranes but not always in bacterial membranes. (d) Phosphatidylinositol is a membrane lipid that is converted into bioactive molecules.

which causes a kink in the chain. At low concentrations, cylindrically shaped phosphatidylcholine molecules form a sheet that is one molecule thick called a **lipid monolayer** on the surface of an aqueous solution (Figure 2.43b). For convenience glycerophospholipids are represented as stick figures with hydrocarbon chains drawn as zigzag lines and polar head groups as circles. Hydrocarbon groups stick up in the air and polar head groups interact with water. This avoids an unfavorable entropy situation in which water molecules would be required to form highly ordered cages around the hydrocarbon chains. The lipid monolayer gains additional stability from van der Waals interactions among hydrocarbon chains. When there is no more room on the surface of the aqueous solution, glycerophospholipids form a **lipid bilayer** with the hydrocarbon chains sequestered inside and the polar head groups directed toward the surrounding aqueous solution (Figure 2.43c). The entropy of water also provides the driving force for bilayer formation.

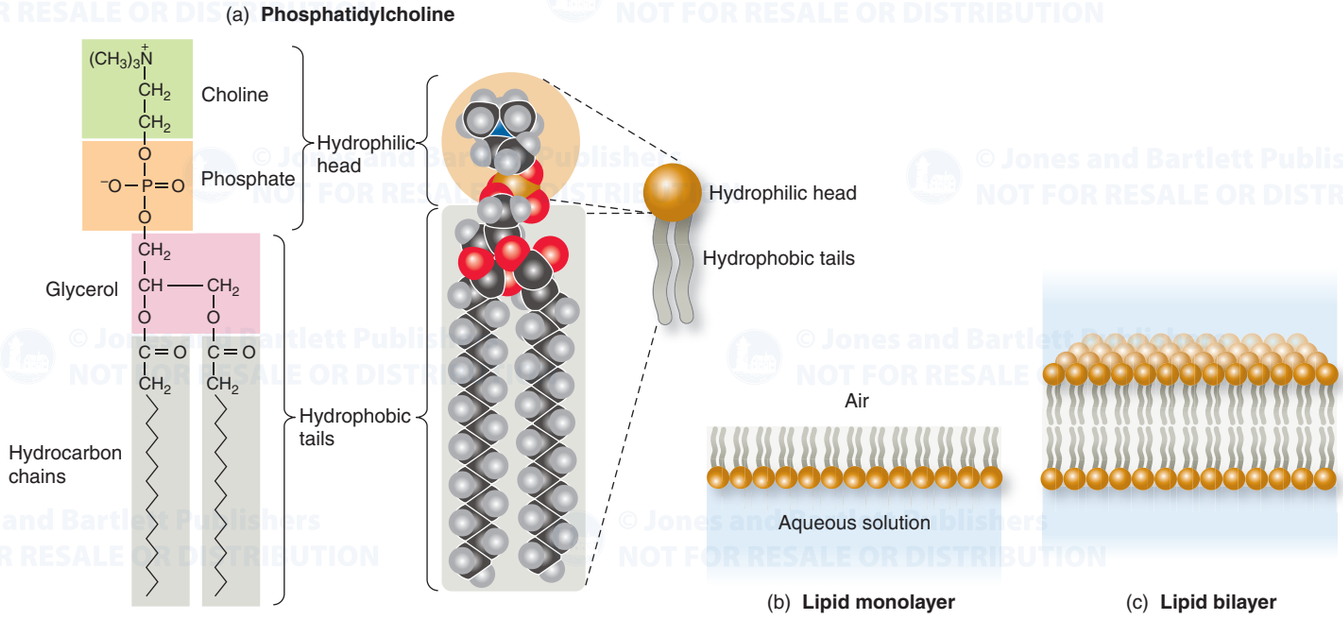


FIGURE 2.43 Lipid monolayer and lipid bilayer. (a) Glycerophospholipids are represented as stick figures with polar head groups shown as circles and hydrocarbon chains as zigzag lines. (b) When glycerophospholipids are placed in water, the polar head groups interact with water and the hydrocarbon tails stick up into the air to form a monolayer. (c) When there is no more room on the water's surface, glycerophospholipids tend to form a lipid bilayer with their polar head groups directed toward the water and their hydrophobic tails sequestered inside. Adapted from Alberts, B., et al. *Molecular Biology of the Cell, 4/e.* Garland Science, 2002.

A variety of physicochemical studies of artificial membranes show that the glycerophospholipids in a membrane drift laterally, indicating that the membrane is fluid. However, glycerophospholipids seldom, if ever, move from one leaflet of the bilayer to the other. Such transverse or flip-flop movement is energetically unfavorable because it requires the glycerophospholipid's polar head group to pass through the bilayer's hydrophobic core. Biological membranes have proteins called *flippases* that catalyze the movement of glycerophospholipid molecules from one leaflet to the other. The ends of a lipid bilayer are unstable because the hydrocarbon chains are exposed to water. Thus lipid bilayers tend to close upon themselves to form hollow bilayer spheres known as **vesicles** or **liposomes** (**FIGURE 2.44**).

The fluid mosaic model has been proposed to explain the structure of biological membranes.

Naturally occurring biological membranes contain many protein molecules (varying from about 30%–70% by weight), and membranes having different functions contain different proteins. There are three kinds of membrane proteins.

1. **Integral membrane proteins** can only be removed from the membrane by harsh treatments such as extraction with detergents, organic solvents, or protein denaturing agents that destroy the membrane. Even when separated from the membrane, integral membrane proteins often retain bound lipids, suggest-

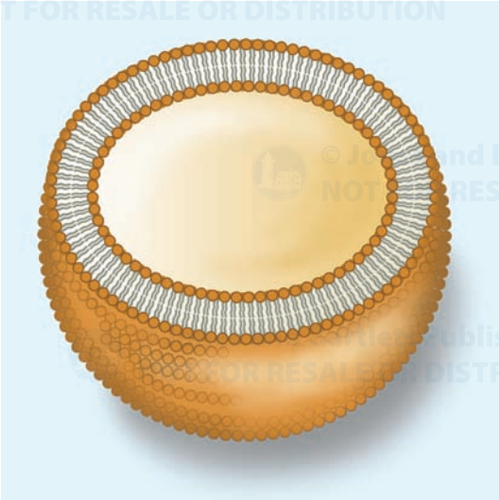


FIGURE 2.44 Lipid bilayer vesicle. Glycerophospholipids can form spherical lipid bilayer vesicles called liposomes. The hydrocarbon chains are sequestered within the bilayer and the polar head groups face the aqueous solution on either side of the membrane. The structure, which is actually spherical, is shown here in cross section.

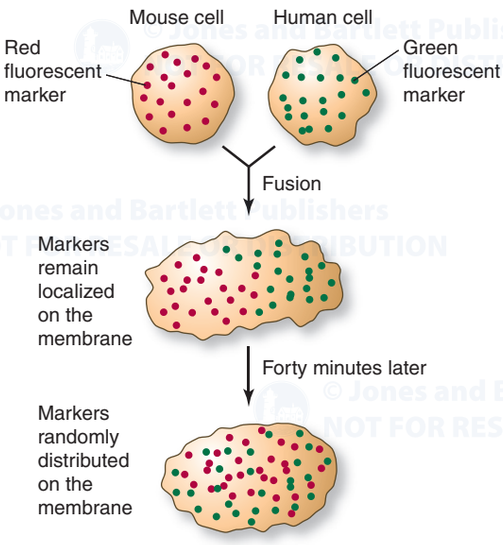


FIGURE 2.45 Diffusion of integral membrane proteins in fused cells. Mouse and human cells can be induced to fuse to form a hybrid cell. The experiment clearly indicates that many membrane proteins are able to move in the plane of the membrane. Adapted from Moran, L. A. and Scrimgeour, K. G. *Biochemistry, 1/e.* Neil Patterson Publishers, 1992.

ing that a similar association exists in the native membrane. When the lipids are removed, integral membrane proteins often form insoluble aggregates and lose their activity. Integral membrane proteins usually account for over 70% of the total membrane proteins. Some integral membrane proteins have a single region that is embedded in the membrane whereas others have two or more embedded regions. The most common type of secondary structure of embedded segments is the α-helix. However, some bacterial integral membrane proteins in the bacterial outer membrane have embedded segments with a β-sheet secondary structure.

2. **Peripheral membrane proteins** can be extracted from membranes by relatively mild treatments such as increasing the salt concentration or adding chelating agents. The former disrupts ionic bonds that link the proteins to the membrane, and the latter removes calcium and magnesium ions that serve as salt bridges between the proteins and the membrane. Once dissociated, peripheral membrane proteins are free of lipids and relatively soluble in aqueous solutions. These properties suggest that peripheral membrane proteins are held to the membrane by weak noncovalent bonds and are not strongly associated with membrane lipids.
3. **Lipoproteins** have a lipid group that is covalently attached to one of the amino acid residues. The hydrophobic lipid group is inserted into one of the monolayers in the membrane bilayer and anchors the lipoprotein to the membrane. The lipoproteins, like the integral membrane proteins, can only be extracted by detergents, organic solvents, or protein denaturing agents that destroy the membrane.

An experiment performed by L. D. Frye and Michael Edidin in 1970 demonstrated that many integral membrane proteins move randomly and freely. As shown in **FIGURE 2.45**, this experiment made use of a cell fusion technique in which human and mouse cells were stimulated to fuse by infection with Sendai virus inactivated by ultraviolet light. Antibodies were prepared against the integral membrane proteins of human cells and mouse cells. The antibody to mouse protein was covalently linked to rhodamine, a molecule that fluoresces red when activated with the appropriate wavelength of light. The human antibody was linked to fluorescein, which fluoresces green. Immediately after fusion, the two fluorescent antibodies were added and the joined cells (heterokaryons) were observed with a fluorescence microscope. Initially each heterokaryon had a red half and a green half. Observation at various times showed a gradual mixing of the colors. After 40 minutes at 37°C, both red and green fluorescence covered the cell surface, indicating that the human and mouse integral membrane proteins were completely and randomly mixed.

In 1972, Jonathan S. Singer and Garth Nicholson proposed the **fluid mosaic model** to explain how proteins are organized in biological membranes (**FIGURE 2.46**). According to this model, lipids and integral membrane proteins are free to move laterally within a mono-

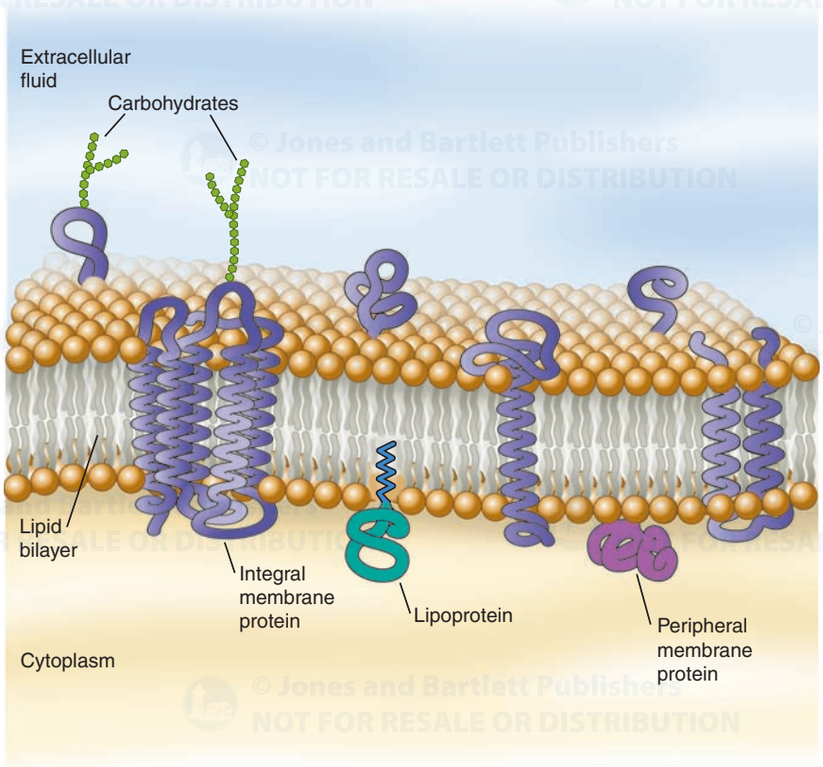


FIGURE 2.46 Fluid mosaic model for the cell membrane. Integral membrane proteins are embedded in the lipid bilayer. The embedded regions have amino acids with hydrophobic side chains, whereas regions that are exposed on either side of the bilayer tend to have amino acids with hydrophilic side chains. Lipoproteins have attached lipid groups such as fatty acids or glycolipids that are embedded in the bilayer. Peripheral membrane proteins associate with integral membrane proteins, lipoproteins, or lipids through weak, non-covalent interactions. Adapted from Cooper, G. M., and Hausman, R. E. *The Cell: A Molecular Approach, 2nd edition.* Sinauer Associates, Inc., 2000.

layer but not to move across the membrane in a transverse fashion (from one monolayer to another) known as *flip-flop*. Flip-flop motion does not occur because the polar regions of proteins and lipids cannot move through the hydrophobic core of the lipid bilayer without the assistance of special proteins. All copies of an integral membrane protein have the same amino-to-carboxyl orientation within the lipid bilayer. Therefore, the two sides of a biological membrane must be different. That is, a biological membrane is an asymmetric structure having (taking the cell as reference) an inside and an outside. Membrane asymmetry also extends to the lipid bilayer so that each monolayer has a unique lipid composition.

Although quite brief, this description of biological membranes allows us to examine molecular biology problems as they relate to biological membranes.

Suggested Reading

General Overview

Branden, C-I., and Tooze, J. 1999. *Introduction to Protein Structure* (2nd ed). New York: Garland Publishing.

Creighton, T. E. 1993. *Proteins: Structures and Molecular Properties*. New York: W. H. Freeman.

Pauling, L. 1993. How my interest in proteins developed. *Protein Sci* 2:1060–1063.

Petsko, G. A., and Ringe, D. 2003. *Protein Stucture and Function*. Sunderland, CT: Sinauer Associates.

Amino Acids and Peptide Bonds

Gilbert, H. F. 2001. Peptide bonds, disulfide bonds and properties of small peptides. *Encyclopedia of Life Sciences*. pp. 1–7. London, UK: Nature Publishing Co.

Deber, C. M., and Brodsky, B. 2001. Proline residues in proteins. *Encyclopedia of Life Sciences*. pp. 1–6. London, UK: Nature Publishing Co.

Cozzone, A. J. 2002. Proteins: Fundamental chemical properties. *Encyclopedia of Life Sciences*. pp. 1–10. London, UK: Nature Publishing Co.

Protein Purification

Deutscher, M. (ed). 1997. *Guide to Protein Purification*. San Diego, CA: Academic Press.

Fassina, G. 2001. Affinity chromatography. *Encyclopedia of Life Sciences*. pp. 1–3. London, UK: Nature Publishing Co.

Jennissen, H. P. 2001. Hydrophobic interaction chromatography. *Encyclopedia of Life Sciences*. pp. 1–8. London, UK: Nature Publishing Co.

Mansoor, M. A. 2002. Liquid chromatography. *Encyclopedia of Life Sciences*. pp. 1–3. London, UK: Nature Publishing Co.

Roe, S. (ed). 2001. *Protein Purification Techniques. A Practical Approach* (2nd ed). Oxford, UK: Oxford University Press.

Primary Structure

Aebersold, R., and Mann, M. 2003. Mass spectrometry-based proteomics. *Nature* 422:198–207.

Barker, W. C. 2001. Protein sequence databases. *Encyclopedia of Life Sciences*. pp. 1–3. London, UK: Nature Publishing Co.

Steen, H., and Mann, M. 2004. The abc's (and xyz's) of peptide sequencing. *Nat Rev Mol Cell Biol* 5:699–711.

Yates, J. R. III.1998. Mass spectrometry and the age of the proteome. *J Mass Spect* 33:1–19.

Weak Non-Covalent Bonds

Chan, H. S. 2002. Amino acid side-chain hydrophobicity. *Encyclopedia of Life Sciences*. pp. 1–7. London, UK: Nature Publishing Co.

Fersht, A. R. 1987. The hydrogen bond in molecular recognition. *Trend Biochem Sci* 12:301–304.

Herzfeld, J., and Olbris, D. J. 2002. Hydrophobic effect. *Encyclopedia of Life Sciences*. pp. 1–9. London, UK: Nature Publishing Co.

Hubbard, R. E. 2001. Hydrogen bonds in proteins: role and strength. *Encyclopedia of Life Sciences*. pp. 1–6. London, UK: Nature Publishing Co.

Sharp, K. A. 2001. Water: structure and properties. *Encyclopedia of Life Sciences*. pp. 1–7. London, UK: Nature Publishing Co.

Secondary Structure

Barton, G. J. 1995. Protein secondary structure prediction. *Curr Opin Struct Biol* 5:372–376.

Chandonia, J-M. Protein secondary structures: prediction. *Encyclopedia of Life Sciences*. pp. 1–6. London, UK: Nature Publishing Co.

Eisenhaber, F., Persson, B., and Argos, P. 1995. Protein structure prediction: recognition of primary, secondary, and tertiary structural features from amino acid sequence. *Crit Rev Biochem Mol Biol* 30:1–94.

Minor, D. L. Jr, and Kim, P. S. 1996. Context-dependent secondary structure formation of a designed protein sequence. *Nature* 380:730–734.

Rost, B. 2001. Protein secondary structure prediction continues to rise. *J Struct Biol* 134:204–218.

Sansom, M. S. P. 2001. Hydrophobicity plots. *Encyclopedia of Life Sciences*. pp. 1–4. London, UK: Nature Publishing Co.

Tertiary Structure

Anfinsen, C. B. 1973. The principles that govern the folding of protein chains. *Science* 181:223–230.

Carter, W. C. Jr, and Sweet, R. M. (eds). 1997. Macromolecular crystallography. *Methods in Enzymology*, vol 276 and 277. San Diego, CA: Academic Press.

Consalvi, V., and Chiaraluce, R. 2001. Chaperones, chaperonin and heat-shock proteins. *Encyclopedia of Life Sciences*. pp. 1–6. London, UK: Nature Publishing Co.

Daggett, V., and Fersht, A. R. 2003. Is there a unifying mechanism for protein folding? *Trend Biochem Sci* 28:18–25.

Ellis, R. J. 1993. The general concept of molecular chaperones. *Phil Trans Royal Soc Lond Series B* 339:257–261.

Feldman, D. E., and Frydman, J. 2000. Protein folding *in vitro*: the importance of molecular chaperones. *Curr Opin Struct Biol* 10:26–33.

Fink, A. L. 2001. Molten globule. *Encyclopedia of Life Sciences*. pp. 1–6. London, UK. Nature Publishing Co.

Fischer, G., and Schmid, F. X. 2001. Peptidylproline *cis-trans*-isomerases. *Encyclopedia of Life Sciences*. pp. 1–6. London, UK: Nature Publishing Co.

Gottesman, M. E., and Hendrickson, W. A. 2000. Protein folding and unfolding by *Escherichia coli* chaperones and chaperonins. *Curr Opin Microbiol* 3:197–202.

Guss, J. M., and King, G. F. 2002. Macromolecular structure determination: comparison of crystallography and NMR. *Encyclopedia of Life Sciences*. pp. 1–5. London, UK: Nature Publishing Co.

Hartl, F. U. 1996. Molecular chaperones in cellular protein folding. *Nature* 381:571–580.

Helliwell, J. R. 2001. X-ray diffraction at synchrotron light sources. *Encyclopedia of Life Sciences*. pp. 1–6. London, UK: Nature Publishing Co.

Kay, L. E., and Gardner, K. H. 1997. Solution NMR spectroscopy beyond 25 kDa. *Curr Opin Struct Biol* 7:722–731.

Kyte, J., and Doolittle, R. F. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157:105–132.

Pearl, F. M. G., Orengo, C. A., and Thornton, J. M. 2001. Protein structure classification. *Encyclopedia of Life Sciences*. pp. 1–7. London, UK: Nature Publishing Co.

Rhodes, R. 1993. *Crystallography made crystal clear*. San Diego, CA: Academic Press.

Schumann, W. 2001. Heat shock response. *Encyclopedia of Life Sciences*. pp. 1–7. London, UK: Nature Publishing Co.

Schwaller, M. D., and Gilbert, H. F. 2001. Protein disulfide isomerases. *Encyclopedia of Life Sciences*. pp. 1–7. London, UK: Nature Publishing Co.

Tugarinov, V., Hwang, P. M., and Kay, L. E. 2004. Nuclear magnetic resonance spectroscopy of high-molecular weight proteins. *Ann Rev Biochem* 73:107–146.

Young, J. C., Agashe, V. R., Siegers, K., and Hartl, F. U. 2004. Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev Mol Cell Biol* 5:781–791.

Wery, J. P., and Scehvitz, R. W. 1997. New trends in macromolecular x-ray crystallography. *Curr Opin Chem Biol* 1:365–369.

Wickner, S., Maurizi, M. R., and Gottesman, S. 1999. Posttranslational quality control: folding, refolding, and degrading proteins. *Science* 286:1888–1893.

Wüthrich, K. 2001. Nuclear magnetic resonance spectroscopy of proteins. *Encyclopedia of Life Sciences*. pp. 1–5. London: UK: Nature Publishing Co.

Wüthrich, K. 2003. NMR studies of structure and function of biological macromolecules. *J Biomol NMR* 27:13–39.

Yates, J. R. 2001. Mass spectrometry in biology. *Encyclopedia of Life Sciences*. pp. 1–5. London, UK: Nature Publishing Co.

Zhang, H. 2002. Protein tertiary structures: prediction from amino acid sequences. *Encyclopedia of Life Sciences*. pp. 1–7. London, UK: Nature Publishing Co.

Biological Membranes

Edidin, M. 2003. Lipids on the frontier: a century of cell-membrane bilayers. *Nat Rev Mol Cell Biol* 4:414–418.

Pietzsch, J. 2004. Mind the membrane. *Horizon Symposia.* pp.1–4. London, UK: Nature Publishing Group.

Raudino, A., and Sarpietro, M. G. 2001. Lipid bilayers. *Encyclopedia of Life Sciences*. pp. 1–6. London, UK: Nature Publishing Co.

Singer, S. J., and Nicolson, G. L. 1972. The fluid mosaic model of the structure of cell membranes. *Science* 175:720–731.

Tanford, C. 1980. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*. Hoboken, NJ: John Wiley & Sons.

Yeagle, P. (ed). 2004. *The Structure of Biological Membranes* (2nd ed). Boca Raton, FL: CRC Press.

Yeagle, P. L. 2001. Cell membrane features. *Encyclopedia of Life Sciences*. pp. 1–7. London, UK: Nature Publishing Co.

This book has a website, **http://biology.jbpub.com/book/molecular**, which contains a variety of resources, including links to in-depth information on topics covered in this chapter, and review material designed to assist in your study of molecular biology.