Proteins: Higher Orders of Structure

5

Victor W. Rodwell, PhD, & Peter J. Kennelly, PhD

BIOMEDICAL IMPORTANCE

Proteins catalyze metabolic reactions, power cellular motion, and form macromolecular rods and cables that provide structural integrity to hair, bones, tendons, and teeth. In nature, form follows function. The structural variety of human proteins therefore reflects the sophistication and diversity of their biologic roles. Maturation of a newly synthesized polypeptide into a biologically functional protein requires that it be folded into a specific three-dimensional arrangement, or **conformation.** During maturation, **posttranslational modifications** may add new chemical groups or remove transiently needed peptide segments. Genetic or nutritional deficiencies that impede protein maturation are deleterious to health. Examples of the former include Creutzfeldt-Jakob disease, scrapie, Alzheimer's disease, and bovine spongiform encephalopathy (mad cow disease). Scurvy represents a nutritional deficiency that impairs protein maturation.

CONFORMATION VERSUS CONFIGURATION

The terms configuration and conformation are often confused. **Configuration** refers to the geometric relationship between a given set of atoms, for example, those that distinguish L- from D-amino acids. Interconversion of *configurational* alternatives requires breaking covalent bonds. **Conformation** refers to the spatial relationship of every atom in a molecule. Interconversion between conformers occurs without covalent bond rupture, with retention of configuration, and typically via rotation about single bonds.

PROTEINS WERE INITIALLY CLASSIFIED BY THEIR GROSS CHARACTERISTICS

Scientists initially approached structure-function relationships in proteins by separating them into classes based upon properties such as solubility, shape, or the presence of nonprotein groups. For example, the proteins that can be extracted from cells using solutions at physiologic pH and ionic strength are classified as **soluble.** Extraction of **integral membrane proteins** requires dissolution of the membrane with detergents. **Globular proteins** are compact, are roughly spherical or ovoid in shape, and have **axial ratios** (the ratio of their shortest to longest dimensions) of not over 3. Most enzymes are globular proteins, whose large internal volume provides ample space in which to construct cavities of the specific shape, charge, and hydrophobicity or hydrophilicity required to bind substrates and promote catalysis. By contrast, many structural proteins adopt highly extended conformations. These **fibrous proteins** possess axial ratios of 10 or more.

Lipoproteins and **glycoproteins** contain covalently bound lipid and carbohydrate, respectively. Myoglobin, hemoglobin, cytochromes, and many other proteins contain tightly associated metal ions and are termed **metalloproteins.** With the development and application of techniques for determining the amino acid sequences of proteins (Chapter 4), more precise classification schemes have emerged based upon similarity, or **homology,** in amino acid sequence and structure. However, many early classification terms remain in common use.

PROTEINS ARE CONSTRUCTED USING MODULAR PRINCIPLES

Proteins perform complex physical and catalytic functions by positioning specific chemical groups in a precise three-dimensional arrangement. The polypeptide scaffold containing these groups must adopt a conformation that is both functionally efficient and physically strong. At first glance, the biosynthesis of polypeptides comprised of tens of thousands of individual atoms would appear to be extremely challenging. When one considers that a typical polypeptide can adopt $\geq 10^{50}$ distinct conformations, folding into the conformation appropriate to their biologic function would appear to be even more difficult. As described in Chapters 3 and 4, synthesis of the polypeptide backbones of proteins employs a small set of common building blocks or modules, the amino acids, joined by a common linkage, the peptide bond. A stepwise modular pathway simplifies the folding and processing of newly synthesized polypeptides into mature proteins.

THE FOUR ORDERS OF PROTEIN STRUCTURE

The modular nature of protein synthesis and folding are embodied in the concept of orders of protein structure: **primary structure,** the sequence of the amino acids in a polypeptide chain; **secondary structure,** the folding of short (3- to 30-residue), contiguous segments of polypeptide into geometrically ordered units; **tertiary structure,** the three-dimensional assembly of secondary structural units to form larger functional units such as the mature polypeptide and its component domains; and **quaternary structure,** the number and types of polypeptide units of oligomeric proteins and their spatial arrangement.

SECONDARY STRUCTURE

Peptide Bonds Restrict Possible Secondary Conformations

Free rotation is possible about only two of the three covalent bonds of the polypeptide backbone: the α-carbon (Cα) to the carbonyl carbon (Co) bond and the Cα to nitrogen bond (Figure 3–4). The partial double-bond character of the peptide bond that links Co to the α-nitrogen requires that the carbonyl carbon, carbonyl oxygen, and α-nitrogen remain coplanar, thus preventing rotation. The angle about the Cα—N bond is termed the phi (Φ) angle, and that about the Co—Cα bond the psi (Ψ) angle. For amino acids other than glycine, most combinations of phi and psi angles are disallowed because of steric hindrance (Figure 5–1). The conformations of proline are even more restricted due to the absence of free rotation of the N—Cα bond.

Regions of ordered secondary structure arise when a series of aminoacyl residues adopt similar phi and psi angles. Extended segments of polypeptide (eg, loops) can possess a variety of such angles. The angles that define the two most common types of secondary structure, the **α helix** and the **β sheet,** fall within the lower and upper left-hand quadrants of a Ramachandran plot, respectively (Figure 5–1).

The Alpha Helix

The polypeptide backbone of an α helix is twisted by an equal amount about each α-carbon with a phi angle of approximately −57 degrees and a psi angle of approximately −47 degrees. A complete turn of the helix contains an average of 3.6 aminoacyl residues, and the distance it rises per turn (its pitch) is 0.54 nm (Figure 5–2). The R groups of each aminoacyl residue in an α helix face outward (Figure 5–3). Proteins contain only L-amino acids, for which a right-handed α helix is by far the more stable, and only right-handed α helices occur in nature. Schematic diagrams of proteins represent α helices as cylinders.

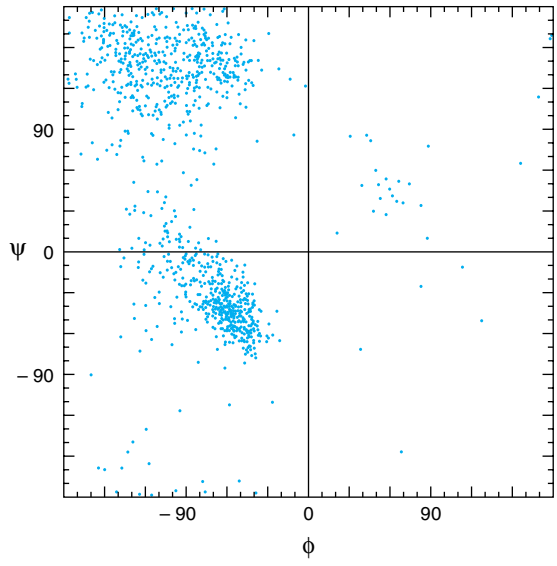

Figure 5–1. Ramachandran plot of the main chain phi (Φ) and psi (Ψ) angles for approximately 1000 nonglycine residues in eight proteins whose structures were solved at high resolution. The dots represent allowable combinations and the spaces prohibited combinations of phi and psi angles. (Reproduced, with permission, from Richardson JS: The anatomy and taxonomy of protein structures. Adv Protein Chem 1981;34:167.)

The stability of an α helix arises primarily from hydrogen bonds formed between the oxygen of the peptide bond carbonyl and the hydrogen atom of the peptide bond nitrogen of the fourth residue down the polypeptide chain (Figure 5–4). The ability to form the maximum number of hydrogen bonds, supplemented by van der Waals interactions in the core of this tightly packed structure, provides the thermodynamic driving force for the formation of an α helix. Since the peptide bond nitrogen of proline lacks a hydrogen atom to contribute to a hydrogen bond, proline can only be stably accommodated within the first turn of an α helix. When present elsewhere, proline disrupts the conformation of the helix, producing a bend. Because of its small size, glycine also often induces bends in α helices.

Many α helices have predominantly hydrophobic R groups on one side of the axis of the helix and predominantly hydrophilic ones on the other. These **amphipathic helices** are well adapted to the formation of interfaces between polar and nonpolar regions such as the hydrophobic interior of a protein and its aqueous envi-

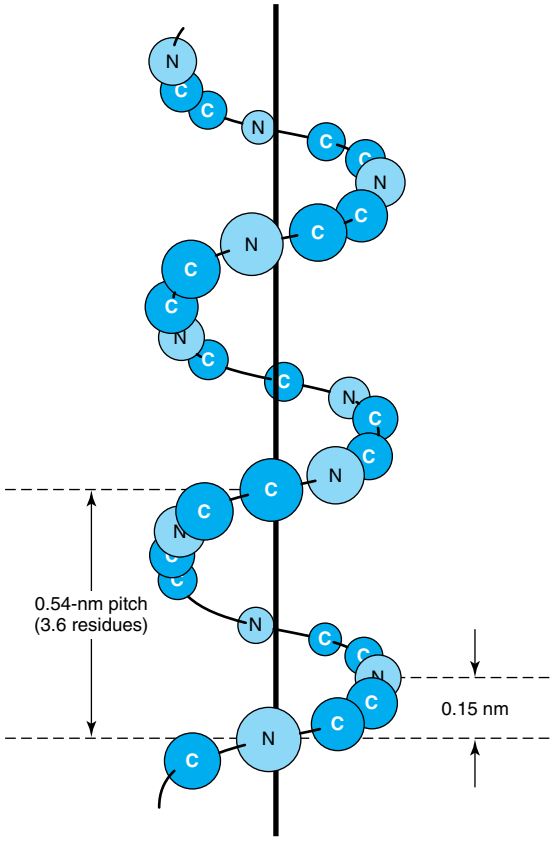

Figure 5–2. Orientation of the main chain atoms of a peptide about the axis of an α helix.

ronment. Clusters of amphipathic helices can create a channel, or pore, that permits specific polar molecules to pass through hydrophobic cell membranes.

The Beta Sheet

The second (hence "beta") recognizable regular secondary structure in proteins is the β sheet. The amino acid residues of a β sheet, when viewed edge-on, form a zigzag or pleated pattern in which the R groups of adjacent residues point in opposite directions. Unlike the compact backbone of the α helix, the peptide backbone of the β sheet is highly extended. But like the α helix, β sheets derive much of their stability from hydrogen bonds between the carbonyl oxygens and amide hydrogens of peptide bonds. However, in contrast to the α helix, these bonds are formed with adjacent segments of β sheet (Figure 5–5).

Interacting β sheets can be arranged either to form a **parallel** β sheet, in which the adjacent segments of the

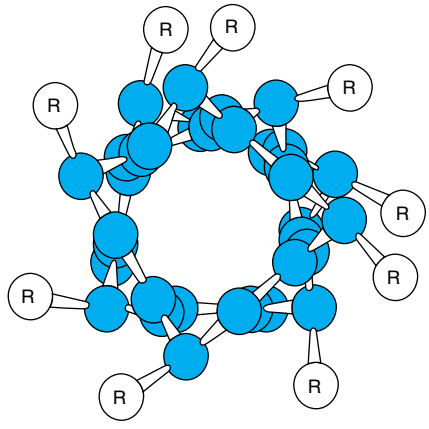

Figure 5–3. View down the axis of an α helix. The side chains (R) are on the outside of the helix. The van der Waals radii of the atoms are larger than shown here; hence, there is almost no free space inside the helix. (Slightly modified and reproduced, with permission, from Stryer L: *Biochemistry,* 3rd ed. Freeman, 1995. Copyright © 1995 by W.H. Freeman and Co.)

polypeptide chain proceed in the same direction amino to carboxyl, or an **antiparallel** sheet, in which they proceed in opposite directions (Figure 5–5). Either configuration permits the maximum number of hydrogen bonds between segments, or strands, of the sheet. Most β sheets are not perfectly flat but tend to have a right-handed twist. Clusters of twisted strands of β sheet form the core of many globular proteins (Figure 5–6). Schematic diagrams represent β sheets as arrows that point in the amino to carboxyl terminal direction.

Loops & Bends

Roughly half of the residues in a "typical" globular protein reside in α helices and β sheets and half in loops, turns, bends, and other extended conformational features. Turns and bends refer to short segments of amino acids that join two units of secondary structure, such as two adjacent strands of an antiparallel β sheet. A β turn involves four aminoacyl residues, in which the first residue is hydrogen-bonded to the fourth, resulting in a tight 180-degree turn (Figure 5–7). Proline and glycine often are present in β turns.

Loops are regions that contain residues beyond the minimum number necessary to connect adjacent regions of secondary structure. Irregular in conformation, loops nevertheless serve key biologic roles. For many enzymes, the loops that bridge domains responsible for binding substrates often contain aminoacyl residues

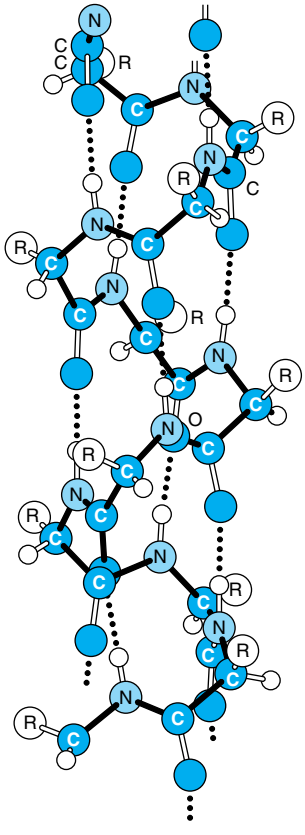

Figure 5–4. Hydrogen bonds (dotted lines) formed between H and O atoms stabilize a polypeptide in an α-helical conformation. (Reprinted, with permission, from Haggis GH et al: *Introduction to Molecular Biology.* Wiley, 1964.)

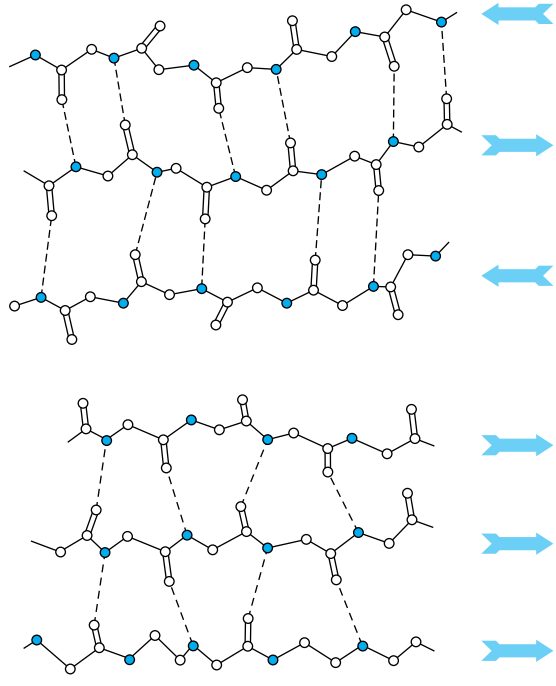

Figure 5–5. Spacing and bond angles of the hydrogen bonds of antiparallel and parallel pleated β sheets. Arrows indicate the direction of each strand. The hydrogen-donating α-nitrogen atoms are shown as blue circles. Hydrogen bonds are indicated by dotted lines. For clarity in presentation, R groups and hydrogens are omitted. **Top:** Antiparallel β sheet. Pairs of hydrogen bonds alternate between being close together and wide apart and are oriented approximately perpendicular to the polypeptide backbone. **Bottom:** Parallel β sheet. The hydrogen bonds are evenly spaced but slant in alternate directions.

that participate in catalysis. **Helix-loop-helix motifs** provide the oligonucleotide-binding portion of DNA-binding proteins such as repressors and transcription factors. Structural motifs such as the helix-loop-helix motif that are intermediate between secondary and tertiary structures are often termed **supersecondary structures.** Since many loops and bends reside on the surface of proteins and are thus exposed to solvent, they constitute readily accessible sites, or **epitopes,** for recognition and binding of antibodies.

While loops lack apparent structural regularity, they exist in a specific conformation stabilized through hydrogen bonding, salt bridges, and hydrophobic interactions with other portions of the protein. However, not all portions of proteins are necessarily ordered. Proteins may contain "disordered" regions, often at the extreme amino or carboxyl terminal, characterized by high conformational flexibility. In many instances, these disordered regions assume an ordered conformation upon binding of a ligand. This structural flexibility enables such regions to act as ligand-controlled switches that affect protein structure and function.

Tertiary & Quaternary Structure

The term "tertiary structure" refers to the entire three-dimensional conformation of a polypeptide. It indicates, in three-dimensional space, how secondary structural features—helices, sheets, bends, turns, and loops—assemble to form **domains** and how these domains relate spatially to one another. A domain is a section of protein structure sufficient to perform a particular chemical or physical task such as binding of a substrate

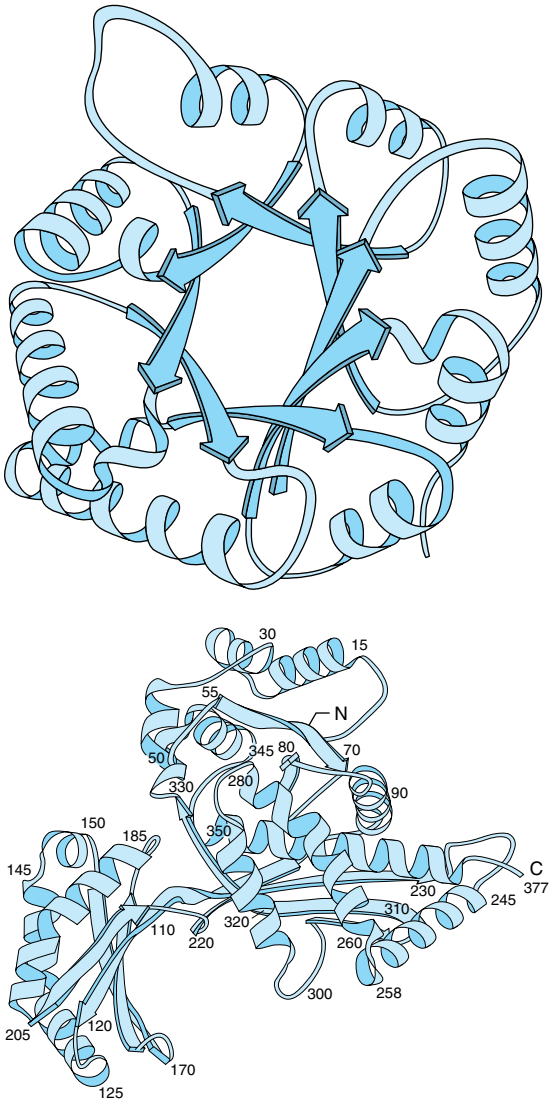

Figure 5–6. Examples of tertiary structure of proteins. **Top:** The enzyme triose phosphate isomerase. Note the elegant and symmetrical arrangement of alternating β sheets and α helices. (Courtesy of J Richardson.) **Bottom:** Two-domain structure of the subunit of a homodimeric enzyme, a bacterial class II HMG-CoA reductase. As indicated by the numbered residues, the single polypeptide begins in the large domain, enters the small domain, and ends in the large domain. (Courtesy of C Lawrence, V Rodwell, and C Stauffacher, Purdue University.)

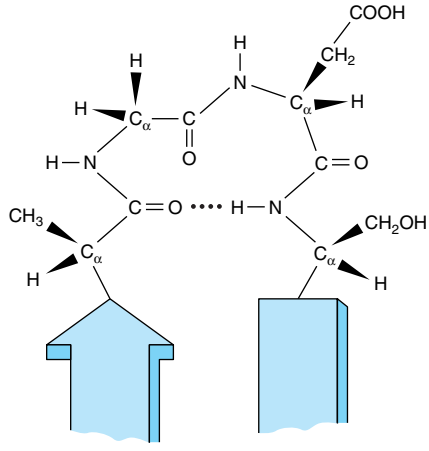

Figure 5–7. A β-turn that links two segments of antiparallel β sheet. The dotted line indicates the hydrogen bond between the first and fourth amino acids of the four-residue segment Ala-Gly-Asp-Ser.

or other ligand. Other domains may anchor a protein to a membrane or interact with a regulatory molecule that modulates its function. A small polypeptide such as triose phosphate isomerase (Figure 5–6) or myoglobin (Chapter 6) may consist of a single domain. By contrast, protein kinases contain two domains. Protein kinases catalyze the transfer of a phosphoryl group from ATP to a peptide or protein. The amino terminal portion of the polypeptide, which is rich in β sheet, binds ATP, while the carboxyl terminal domain, which is rich in α helix, binds the peptide or protein substrate (Figure 5–8). The groups that catalyze phosphoryl transfer reside in a loop positioned at the interface of the two domains.

In some cases, proteins are assembled from more than one polypeptide, or protomer. Quaternary structure defines the polypeptide composition of a protein and, for an oligomeric protein, the spatial relationships between its subunits or protomers. **Monomeric** proteins consist of a single polypeptide chain. **Dimeric** proteins contain two polypeptide chains. Homodimers contain two copies of the same polypeptide chain, while in a heterodimer the polypeptides differ. Greek letters (α, β, γ etc) are used to distinguish different subunits of a heterooligomeric protein, and subscripts indicate the number of each subunit type. For example, α_4 designates a homotetrameric protein, and $\alpha_2\beta_2\gamma$ a protein with five subunits of three different types.

Since even small proteins contain many thousands of atoms, depictions of protein structure that indicate the position of every atom are generally too complex to be readily interpreted. Simplified schematic diagrams thus are used to depict key features of a protein's ter-

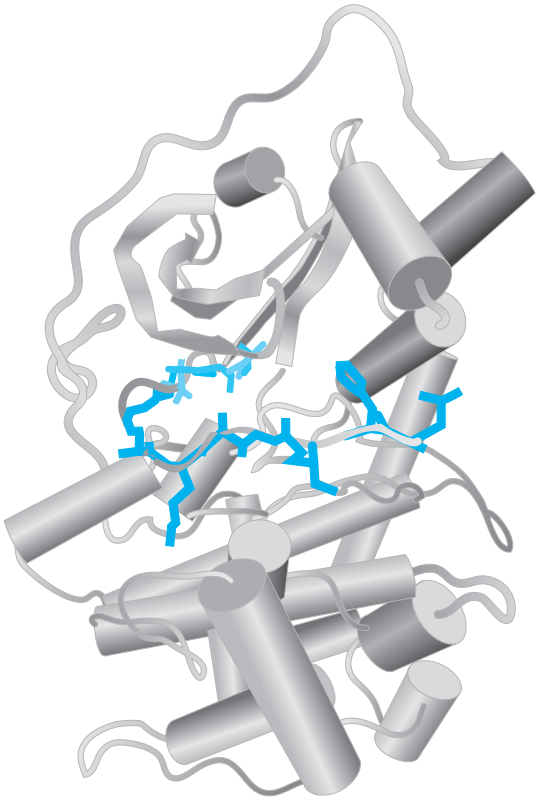

Figure 5–8. Domain structure. Protein kinases contain two domains. The upper, amino terminal domain binds the phosphoryl donor ATP (light blue). The lower, carboxyl terminal domain is shown binding a synthetic peptide substrate (dark blue).

tiary and quaternary structure. Ribbon diagrams (Figures 5–6 and 5–8) trace the conformation of the polypeptide backbone, with cylinders and arrows indicating regions of α helix and β sheet, respectively. In an even simpler representation, line segments that link the α carbons indicate the path of the polypeptide backbone. These schematic diagrams often include the side chains of selected amino acids that emphasize specific structure-function relationships.

MULTIPLE FACTORS STABILIZE TERTIARY & QUATERNARY STRUCTURE

Higher orders of protein structure are stabilized primarily—and often exclusively—by noncovalent interactions. Principal among these are hydrophobic interactions that drive most hydrophobic amino acid side chains into the interior of the protein, shielding them from water. Other significant contributors include hydrogen bonds and salt bridges between the carboxylates of aspartic and glutamic acid and the oppositely charged side chains of protonated lysyl, argininyl, and histidyl residues. While individually weak relative to a typical covalent bond of 80–120 kcal/mol, collectively these numerous interactions confer a high degree of stability to the biologically functional conformation of a protein, just as a Velcro fastener harnesses the cumulative strength of multiple plastic loops and hooks.

Some proteins contain covalent disulfide (S—S) bonds that link the sulfhydryl groups of cysteinyl residues. Formation of disulfide bonds involves oxidation of the cysteinyl sulfhydryl groups and requires oxygen. Intrapolypeptide disulfide bonds further enhance the stability of the folded conformation of a peptide, while interpolypeptide disulfide bonds stabilize the quaternary structure of certain oligomeric proteins.

THREE-DIMENSIONAL STRUCTURE IS DETERMINED BY X-RAY CRYSTALLOGRAPHY OR BY NMR SPECTROSCOPY

X-Ray Crystallography

Since the determination of the three-dimensional structure of myoglobin over 40 years ago, the three-dimensional structures of thousands of proteins have been determined by x-ray crystallography. The key to x-ray crystallography is the precipitation of a protein under conditions in which it forms ordered crystals that diffract x-rays. This is generally accomplished by exposing small drops of the protein solution to various combinations of pH and precipitating agents such as salts and organic solutes such as polyethylene glycol. A detailed three-dimensional structure of a protein can be constructed from its primary structure using the pattern by which it diffracts a beam of monochromatic x-rays. While the development of increasingly capable computer-based tools has rendered the analysis of complex x-ray diffraction patterns increasingly facile, a major stumbling block remains the requirement of inducing highly purified samples of the protein of interest to crystallize. Several lines of evidence, including the ability of some crystallized enzymes to catalyze chemical reactions, indicate that the vast majority of the structures determined by crystallography faithfully represent the structures of proteins in free solution.

Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy, a powerful complement to x-ray crystallography, mea-

sures the absorbance of radio frequency electromagnetic energy by certain atomic nuclei. "NMR-active" isotopes of biologically relevant atoms include ^{1}H, ^{13}C, ^{15}N, and ^{31}P. The frequency, or chemical shift, at which a particular nucleus absorbs energy is a function of both the functional group within which it resides and the proximity of other NMR-active nuclei. Two-dimensional NMR spectroscopy permits a three-dimensional representation of a protein to be constructed by determining the proximity of these nuclei to one another. NMR spectroscopy analyzes proteins in aqueous solution, obviating the need to form crystals. It thus is possible to observe changes in conformation that accompany ligand binding or catalysis using NMR spectroscopy. However, only the spectra of relatively small proteins, ≤ 20 kDa in size, can be analyzed with current technology.

Molecular Modeling

An increasingly useful adjunct to the empirical determination of the three-dimensional structure of proteins is the use of computer technology for molecular modeling. The types of models created take two forms. In the first, the known three-dimensional structure of a protein is used as a template to build a model of the probable structure of a homologous protein. In the second, computer software is used to manipulate the static model provided by crystallography to explore how a protein's conformation might change when ligands are bound or when temperature, pH, or ionic strength is altered. Scientists also are examining the library of available protein structures in an attempt to devise computer programs that can predict the three-dimensional conformation of a protein directly from its primary sequence.

PROTEIN FOLDING

The Native Conformation of a Protein Is Thermodynamically Favored

The number of distinct combinations of phi and psi angles specifying potential conformations of even a relatively small—15-kDa—polypeptide is unbelievably vast. Proteins are guided through this vast labyrinth of possibilities by thermodynamics. Since the biologically relevant—or native—conformation of a protein generally is that which is most energetically favored, knowledge of the native conformation is specified in the primary sequence. However, if one were to wait for a polypeptide to find its native conformation by random exploration of all possible conformations, the process would require billions of years to complete. Clearly, protein folding in cells takes place in a more orderly and guided fashion.

Folding Is Modular

Protein folding generally occurs via a stepwise process. In the first stage, the newly synthesized polypeptide emerges from ribosomes, and short segments fold into secondary structural units that provide local regions of organized structure. Folding is now reduced to the selection of an appropriate arrangement of this relatively small number of secondary structural elements. In the second stage, the forces that drive hydrophobic regions into the interior of the protein away from solvent drive the partially folded polypeptide into a "molten globule" in which the modules of secondary structure rearrange to arrive at the mature conformation of the protein. This process is orderly but not rigid. Considerable flexibility exists in the ways and in the order in which elements of secondary structure can be rearranged. In general, each element of secondary or supersecondary structure facilitates proper folding by directing the folding process toward the native conformation and away from unproductive alternatives. For oligomeric proteins, individual protomers tend to fold before they associate with other subunits.

Auxiliary Proteins Assist Folding

Under appropriate conditions, many proteins will spontaneously refold after being previously **denatured** (ie, unfolded) by treatment with acid or base, chaotropic agents, or detergents. However, unlike the folding process in vivo, refolding under laboratory conditions is a far slower process. Moreover, some proteins fail to spontaneously refold in vitro, often forming insoluble **aggregates,** disordered complexes of unfolded or partially folded polypeptides held together by hydrophobic interactions. Aggregates represent unproductive dead ends in the folding process. Cells employ auxiliary proteins to speed the process of folding and to guide it toward a productive conclusion.

Chaperones

Chaperone proteins participate in the folding of over half of mammalian proteins. The hsp70 (70-kDa heat shock protein) family of chaperones binds short sequences of hydrophobic amino acids in newly synthesized polypeptides, shielding them from solvent. Chaperones prevent aggregation, thus providing an opportunity for the formation of appropriate secondary structural elements and their subsequent coalescence into a molten globule. The hsp60 family of chaperones, sometimes called **chaperonins,** differ in sequence and structure from hsp70 and its homologs. Hsp60 acts later in the folding process, often together with an hsp70 chaperone. The central cavity of the donut-

shaped hsp60 chaperone provides a sheltered environment in which a polypeptide can fold until all hydrophobic regions are buried in its interior, eliminating aggregation. Chaperone proteins can also "rescue" proteins that have become thermodynamically trapped in a misfolded dead end by unfolding hydrophobic regions and providing a second chance to fold productively.

Protein Disulfide Isomerase

Disulfide bonds between and within polypeptides stabilize tertiary and quaternary structure. However, disulfide bond formation is nonspecific. Under oxidizing conditions, a given cysteine can form a disulfide bond with the —SH of any accessible cysteinyl residue. By catalyzing disulfide exchange, the rupture of an S—S bond and its reformation with a different partner cysteine, protein disulfide isomerase facilitates the formation of disulfide bonds that stabilize their native conformation.

Proline-*cis,trans*-Isomerase

All X-Pro peptide bonds—where X represents any residue—are synthesized in the trans configuration. However, of the X-Pro bonds of mature proteins, approximately 6% are cis. The cis configuration is particularly common in β-turns. Isomerization from trans to cis is catalyzed by the enzyme proline-*cis,trans*-isomerase (Figure 5–9).

SEVERAL NEUROLOGIC DISEASES RESULT FROM ALTERED PROTEIN CONFORMATION

Prions

The transmissible spongiform encephalopathies, or **prion diseases,** are fatal neurodegenerative diseases characterized by spongiform changes, astrocytic gliomas, and neuronal loss resulting from the deposition of insoluble protein aggregates in neural cells. They include Creutzfeldt-Jakob disease in humans, scrapie in sheep, and bovine spongiform encephalopathy (mad cow disease) in cattle. Prion diseases may manifest themselves as infectious, genetic, or sporadic disorders. Because no viral or bacterial gene encoding the pathologic prion protein could be identified, the source and mechanism of transmission of prion disease long remained elusive. Today it is believed that prion diseases are protein conformation diseases transmitted by altering the conformation, and hence the physical properties, of proteins endogenous to the host. Human prion-related protein, PrP, a glycoprotein encoded on the short arm of chromosome 20, normally is monomeric and rich in α helix. Pathologic prion proteins serve as the templates for the conformational transformation of normal PrP, known as PrPc, into PrPsc. PrPsc is rich in β sheet with many hydrophobic aminoacyl side chains exposed to solvent. PrPsc molecules therefore associate strongly with one other, forming insoluble protease-resistant aggregates. Since one pathologic prion or prion-related protein can serve as template for the conformational transformation of many times its number of PrPc molecules, prion diseases can be transmitted by the protein alone without involvement of DNA or RNA.

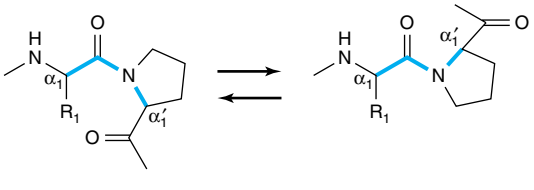

Figure 5–9. Isomerization of the N-α_1 prolyl peptide bond from a cis to a trans configuration relative to the backbone of the polypeptide.

Alzheimer's Disease

Refolding or misfolding of another protein endogenous to human brain tissue, β-amyloid, is also a prominent feature of Alzheimer's disease. While the root cause of Alzheimer's disease remains elusive, the characteristic senile plaques and neurofibrillary bundles contain aggregates of the protein β-amyloid, a 4.3-kDa polypeptide produced by proteolytic cleavage of a larger protein known as amyloid precursor protein. In Alzheimer's disease patients, levels of β-amyloid become elevated, and this protein undergoes a conformational transformation from a soluble α helix–rich state to a state rich in β sheet and prone to self-aggregation. Apolipoprotein E has been implicated as a potential mediator of this conformational transformation.

COLLAGEN ILLUSTRATES THE ROLE OF POSTTRANSLATIONAL PROCESSING IN PROTEIN MATURATION

Protein Maturation Often Involves Making & Breaking Covalent Bonds

The maturation of proteins into their final structural state often involves the cleavage or formation (or both) of covalent bonds, a process termed **posttranslational modification.** Many polypeptides are initially synthesized as larger precursors, called **proproteins.** The "extra" polypeptide segments in these proproteins often serve as leader sequences that target a polypeptide

to a particular organelle or facilitate its passage through a membrane. Others ensure that the potentially harmful activity of a protein such as the proteases trypsin and chymotrypsin remains inhibited until these proteins reach their final destination. However, once these transient requirements are fulfilled, the now superfluous peptide regions are removed by selective proteolysis. Other covalent modifications may take place that add new chemical functionalities to a protein. The maturation of collagen illustrates both of these processes.

Collagen Is a Fibrous Protein

Collagen is the most abundant of the fibrous proteins that constitute more than 25% of the protein mass in the human body. Other prominent fibrous proteins include keratin and myosin. These proteins represent a primary source of structural strength for cells (ie, the cytoskeleton) and tissues. Skin derives its strength and flexibility from a crisscrossed mesh of collagen and keratin fibers, while bones and teeth are buttressed by an underlying network of collagen fibers analogous to the steel strands in reinforced concrete. Collagen also is present in connective tissues such as ligaments and tendons. The high degree of tensile strength required to fulfill these structural roles requires elongated proteins characterized by repetitive amino acid sequences and a regular secondary structure.

Collagen Forms a Unique Triple Helix

Tropocollagen consists of three fibers, each containing about 1000 amino acids, bundled together in a unique conformation, the collagen triple helix (Figure 5–10). A mature collagen fiber forms an elongated rod with an axial ratio of about 200. Three intertwined polypeptide strands, which twist to the left, wrap around one another in a right-handed fashion to form the collagen triple helix. The opposing handedness of this superhelix and its component polypeptides makes the collagen triple helix highly resistant to unwinding—the same principle used in the steel cables of suspension bridges. A collagen triple helix has 3.3 residues per turn and a rise per residue nearly twice that of an α helix. The R groups of each polypeptide strand of the triple helix pack so closely that in order to fit, one must be glycine. Thus, every third amino acid residue in collagen is a glycine residue. Staggering of the three strands provides appropriate positioning of the requisite glycines throughout the helix. Collagen is also rich in proline and hydroxyproline, yielding a repetitive Gly-X-Y pattern (Figure 5–10) in which Y generally is proline or hydroxyproline.

Collagen triple helices are stabilized by hydrogen bonds between residues in *different* polypeptide chains. The hydroxyl groups of hydroxyprolyl residues also participate in interchain hydrogen bonding. Additional stability is provided by covalent cross-links formed between modified lysyl residues both within and between polypeptide chains.

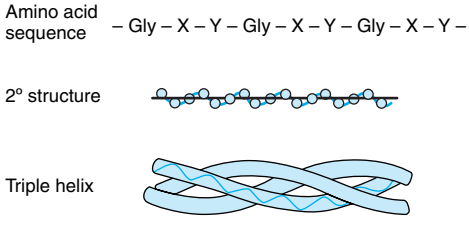

Figure 5–10. Primary, secondary, and tertiary structures of collagen.

Collagen Is Synthesized as a Larger Precursor

Collagen is initially synthesized as a larger precursor polypeptide, procollagen. Numerous prolyl and lysyl residues of procollagen are hydroxylated by prolyl hydroxylase and lysyl hydroxylase, enzymes that require ascorbic acid (vitamin C). Hydroxyprolyl and hydroxylysyl residues provide additional hydrogen bonding capability that stabilizes the mature protein. In addition, glucosyl and galactosyl transferases attach glucosyl or galactosyl residues to the hydroxyl groups of specific hydroxylysyl residues.

The central portion of the precursor polypeptide then associates with other molecules to form the characteristic triple helix. This process is accompanied by the removal of the globular amino terminal and carboxyl terminal extensions of the precursor polypeptide by selective proteolysis. Certain lysyl residues are modified by lysyl oxidase, a copper-containing protein that converts ε-amino groups to aldehydes. The aldehydes can either undergo an aldol condensation to form a C═C double bond or to form a Schiff base (eneimine) with the ε-amino group of an unmodified lysyl residue, which is subsequently reduced to form a C—N single bond. These covalent bonds cross-link the individual polypeptides and imbue the fiber with exceptional strength and rigidity.

Nutritional & Genetic Disorders Can Impair Collagen Maturation

The complex series of events in collagen maturation provide a model that illustrates the biologic consequences of incomplete polypeptide maturation. The best-known defect in collagen biosynthesis is scurvy, a result of a dietary deficiency of vitamin C required by

prolyl and lysyl hydroxylases. The resulting deficit in the number of hydroxyproline and hydroxylysine residues undermines the conformational stability of collagen fibers, leading to bleeding gums, swelling joints, poor wound healing, and ultimately to death. Menkes' syndrome, characterized by kinky hair and growth retardation, reflects a dietary deficiency of the copper required by lysyl oxidase, which catalyzes a key step in formation of the covalent cross-links that strengthen collagen fibers.

Genetic disorders of collagen biosynthesis include several forms of osteogenesis imperfecta, characterized by fragile bones. In Ehlers-Dahlos syndrome, a group of connective tissue disorders that involve impaired integrity of supporting structures, defects in the genes that encode α collagen-1, procollagen *N*-peptidase, or lysyl hydroxylase result in mobile joints and skin abnormalities.

SUMMARY

- Proteins may be classified on the basis of the solubility, shape, or function or of the presence of a prosthetic group such as heme. Proteins perform complex physical and catalytic functions by positioning specific chemical groups in a precise three-dimensional arrangement that is both functionally efficient and physically strong.
- The gene-encoded primary structure of a polypeptide is the sequence of its amino acids. Its secondary structure results from folding of polypeptides into hydrogen-bonded motifs such as the α helix, the β-pleated sheet, β bends, and loops. Combinations of these motifs can form supersecondary motifs.
- Tertiary structure concerns the relationships between secondary structural domains. Quaternary structure of proteins with two or more polypeptides (oligomeric proteins) is a feature based on the spatial relationships between various types of polypeptides.
- Primary structures are stabilized by covalent peptide bonds. Higher orders of structure are stabilized by weak forces—multiple hydrogen bonds, salt (electrostatic) bonds, and association of hydrophobic R groups.
- The phi (Φ) angle of a polypeptide is the angle about the C_α—N bond; the psi (Ψ) angle is that about the C_α-C_o bond. Most combinations of phi-psi angles are disallowed due to steric hindrance. The phi-psi angles that form the α helix and the β sheet fall within the lower and upper left-hand quadrants of a Ramachandran plot, respectively.
- Protein folding is a poorly understood process. Broadly speaking, short segments of newly synthesized polypeptide fold into secondary structural units. Forces that bury hydrophobic regions from solvent then drive the partially folded polypeptide into a "molten globule" in which the modules of secondary structure are rearranged to give the native conformation of the protein.
- Proteins that assist folding include protein disulfide isomerase, proline-*cis*,*trans*,-isomerase, and the chaperones that participate in the folding of over half of mammalian proteins. Chaperones shield newly synthesized polypeptides from solvent and provide an environment for elements of secondary structure to emerge and coalesce into molten globules.
- Techniques for study of higher orders of protein structure include x-ray crystallography, NMR spectroscopy, analytical ultracentrifugation, gel filtration, and gel electrophoresis.
- Silk fibroin and collagen illustrate the close linkage of protein structure and biologic function. Diseases of collagen maturation include Ehlers-Danlos syndrome and the vitamin C deficiency disease scurvy.
- Prions—protein particles that lack nucleic acid—cause fatal transmissible spongiform encephalopathies such as Creutzfeldt-Jakob disease, scrapie, and bovine spongiform encephalopathy. Prion diseases involve an altered secondary-tertiary structure of a naturally occurring protein, PrPc. When PrPc interacts with its pathologic isoform PrPSc, its conformation is transformed from a predominantly α-helical structure to the β-sheet structure characteristic of PrPSc.

REFERENCES

Branden C, Tooze J: *Introduction to Protein Structure.* Garland, 1991.

Burkhard P, Stetefeld J, Strelkov SV: Coiled coils: A highly versatile protein folding motif. Trends Cell Biol 2001;11:82.

Collinge J: Prion diseases of humans and animals: Their causes and molecular basis. Annu Rev Neurosci 2001;24:519.

Frydman J: Folding of newly translated proteins in vivo: The role of molecular chaperones. Annu Rev Biochem 2001;70:603.

Radord S: Protein folding: Progress made and promises ahead. Trends Biochem Sci 2000;25:611.

Schmid FX: Proly isomerase: Enzymatic catalysis of slow protein folding reactions. Ann Rev Biophys Biomol Struct 1993;22:123.

Segrest MP et al: The amphipathic alpha-helix: A multifunctional structural motif in plasma lipoproteins. Adv Protein Chem 1995;45:1.

Soto C: Alzheimer's and prion disease as disorders of protein conformation: Implications for the design of novel therapeutic approaches. J Mol Med 1999;77:412.