

Chap. 4. "Proteins: Three-Dimensional Structure and Function"

Reading Assignment: pp. 81-121.

Problem Assignment: 1-3, 8, & 11, 12 and 15

I. Introduction.

This chapter is concerned with the topic of protein structure. This is a key area of modern biochemistry as the functional properties of proteins are directly determined by their structural features. We will begin with a discussion of protein conformation--the spatial arrangements of atoms in a protein--and discuss basic structure elements such as the α helix from which proteins are built. Then, the folding and stability of proteins will be covered. Finally, the connection between structure and function will be illustrated through detailed discussions of the structure and properties of collagen, myoglobin, and hemoglobin.

II. The four levels of protein structure. (Fig. 4.1)

- Primary structure - Amino acid sequence of the polypeptide written from N- to C-terminus.
- Secondary structure - Regular local repetitive structures; e.g., α helix, β strands, β turns.
- Tertiary structure - Arrangement of amino acids in three dimensional space. Can be for the entire protein or for one of its domains.
- Quaternary structure - Spatial arrangement of subunits and the locations of contacts between them. Only for proteins with more than one subunit.

There are a number of ways to represent the three dimensional structure of a protein (**Fig. 4.3**). Space-filling molecules show the overall shape and surface exposed to water. Ribbon models are good for showing the segments of secondary structure elements within the protein. In ribbon models, α -helices are shown as spirals analogous to phone cords, and β strands are shown as arrows where the arrowhead is the C-terminal end of the strand.

III. Conformation of the peptide bond.

Protein secondary and tertiary structure are highly dependent on the conformations that are allowed over short stretches of amino acids. The range of variation that can be exhibited in local conformation is determined by the properties of peptide bonds and the sizes of the R-groups on the consecutive residues in the polypeptide chain.

X-ray crystallographic studies of Pauling and Corey showed that peptide bonds are rigid and planar due to the partial double bond character caused by resonance delocalization of electrons in the bond (**Fig. 4.5**). Double bond character prevents rotation, and the oxygen and hydrogen atoms almost always are located *trans* with respect to one another (**Fig. 4.7**). Proline is the only amino acid that is found in the *cis* conformation to a significant extent in proteins. In contrast, there is free rotation about the C_{α} -C and N- C_{α} bonds of the backbone (**Fig. 4.6**).

Secondary and tertiary structures are determined by the actual values of the angles of these bonds, which are referred to as the phi angle (for the N- C_{α} bond) and the psi angle (for the C_{α} -C bond) (**Fig. 4.8**). The allowed angles depend on the size of the R-group and are more

restricted when there is a bulky R-group such as in tyrosine, or the R-group has a branch at the β -carbon position as in isoleucine. By using models and analyzing the possibility of steric clash between R-groups, Ramachandran developed a plot showing the permissible values for phi and psi angles (**Fig. 4.9**). As would be expected, secondary structures such as the α helix and β sheet have angles that fall within the permissible areas of the plot.

IV. The α helix.

The α helix is one of the 2 main secondary structure elements found in proteins. In the α helix, backbone atoms of the polypeptide chain form a coil which is nearly always right-handed. There are 3.6 amino acids per turn (i.e., 100° rotation per residue) in the most common α helix found in proteins. The helix pitch (rise or length per turn) is 5.4 Å or 1 amino acid per 1.5 Å. H-bonds occur between the -C=O and H-N- groups of amino acids located in the n and $n + 4$ positions along the helix axis (**Fig. 4.10**). Due to this repetitive and regular H-bonding arrangement, the phi and psi angles for amino acids in α helices tend to fall within a narrow range, e.g., average phi = -57° and average psi = -47° (**Fig. 4.9**).

R-groups point outward from the surface of the α helix (**Fig. 4.11**). Within a protein, individual α helices contact one another via their R groups. In many cases, α helices are amphipathic, meaning the helix has a hydrophobic face and a polar face (**Fig. 4.12 & 4.13**). The polar face of an amphipathic helix may contact water or another polar helix, while the nonpolar face may contact a nonpolar helix in the interior of the protein.

α helical structure is highly abundant in insoluble fibrous structural proteins, such as α keratins (hair, wool, horns, finger nails) and in myosin molecules (muscle). α helices intertwine to form "coiled coils" in many proteins such as myosin. The so-called leucine zippers found in many DNA binding proteins is an example of a coiled-coil structure (**Fig. 4.14**).

V. β strands and β sheets.

β strands are the second of the 2 main types of secondary structure elements. In the β conformation, the polypeptide chain is extended and 1 amino acid occurs approximately every 3.5 Å along the chain. Backbone carbonyl group and amides point out on opposite sides of the strand. Likewise, the R-groups of alternating amino acids in the β strand face in opposite directions (**Fig. 4.15**).

In a β sheet assembly of several β strands, H-bonds occur between backbone amide and carbonyl groups in adjacent strands (inter-strand). This is unlike in the α helix, where H-bonds are "intra-strand." Adjacent strands can orient in the same direction (parallel) or in opposite directions (anti-parallel). Stacks of β strands form " β pleated sheets" in structural proteins such as spider silk. β strands also can be amphipathic, and in this case the strands consist of an alternating sequence of polar and nonpolar residues. This makes one side of the β polar and the other nonpolar. When the nonpolar faces of 2 β strands contact one another the structure is called a β sandwich (**Fig. 4.17**).

VI. Loops and turns.

Loops are surface-exposed sequences containing >5 amino acids that connect 2 segments of secondary structure. Turns are shorter versions (4 or 5 amino acids) of loops. Both function to

allow the polypeptide chain to reverse direction in space, producing the compact shape seen in the native conformation. β turns are prevalent between β strands that make up a β sheet. The C=O group of residue 1 in the turn is H-bonded to the H-N group of residue $n + 3$ across the turn (**Fig. 4.18**). Proline and glycine occur frequently in β turns, because proline naturally tends to cause the chain to kink (e.g., cis-proline), and because the α carbon of glycine can take up a wide range of phi and psi angles allowing the chain to twist. β turns have fairly well defined structures and fall into about 8 distinct groups such as Type I and Type II β turns (**Fig. 4.18**).

VII. Tertiary structure of proteins.

Tertiary structure refers to the three-dimensional arrangement of all atoms in a protein. Tertiary structure is formed by the folding in three dimensions of the secondary structure elements of a protein. While the α helical secondary structure is held together by interactions between the carbonyl and amide groups within the backbone, tertiary structure is held together by interactions between R-groups of residues brought together by folding. Disulfide bonds are also counted under the category of tertiary structure interactions. Proteins that are compact are known as globular proteins. Proteins that are extended are known as fibrous proteins.

A. Supersecondary structures.

These folding motifs are common elements composed of small collections of secondary structure elements. Several examples such as the "helix-loop-helix" supersecondary structure found in DNA binding proteins, and coiled-coils found in myosin are shown in **Fig. 4.19**. Supersecondary structures are also known as folding motifs.

B. Domains.

Domains are independently folded discrete structure units within a larger protein. Usually, a domain is composed of a contiguous stretch of amino acids rather than from stretches of amino acids occurring in separated segments of the primary structure. Common domain folds are shown in **Fig. 4.23**. As the structures of more and more proteins have been solved, it has become apparent that most proteins are modular, being built up from collections of 2 or more domains. It is speculated that early in evolution, these relatively small folding elements arose and then later were shuffled around and combined to create proteins with new or multiple functions. Many folding motifs such as the cytochrome *c* fold have been around since the origin of the primordial common ancestor organism.

VIII. Quaternary structure.

Quaternary structure refers to the arrangement of subunits and their contacts within a protein that containing 2 or more subunits. Each subunit is a separate polypeptide chain. Multisubunit proteins are referred to as oligomers. Each subunit within an oligomer is usually assigned a Greek letter to identify it. The chains within a multisubunit protein can be the same or different. Individual chains typically are held together by noncovalent interactions. Analysis of multisubunit proteins has revealed some generalities. First, the individual chains usually are more stable when combined in the multisubunit structure suggesting their folding may be

stabilized by contacts between subunits. Second, small molecule binding sites can be composed of residues from more than one subunit, i.e., within clefts between subunits. Third, regulatory enzymes often are multisubunit proteins that can undergo major conformation changes on binding of small molecules. Fourth, a given subunit may be found in more than one type of oligomeric protein. The shared subunit performs a common function in all of the proteins in which it is incorporated.

IX. Protein denaturation and renaturation.

Most proteins are poised on a precipice of folding instability. The disruption of only a few noncovalent interactions often is all that is required to cause them to unfold (denature). The most common cause of denaturation is heating. On heating, proteins unfold, come out of solution, and aggregate as occurs when cooking an egg white. For this reason, many microorganisms are very sensitive to temperature changes and have developed proteins known as the heat shock proteins which assist in re-folding proteins denatured by high temperatures. In the laboratory, proteins can be denatured by agents such as urea and guanidinium hydrochloride (**Fig. 4.27**). These compounds are known as chaotropic agents, and they alter the structure of water so that it can better contact hydrophobic amino acids that normally are buried inside a folded protein. Laboratory studies indicate that protein folding is a cooperative process--i.e., it occurs via the formation of multiple simultaneous interactions between non-covalently interacting residues. This explains why on heating, a protein denatures over a very narrow range of temperature (**Fig. 4.26**).

In a classic experiment, Christian Anfinsen demonstrated that the folding of a protein can be determined simply by its amino acid sequence (**Fig. 4.29**). Anfinsen treated pure ribonuclease A with urea and 2-mercaptoethanol to denature it. He then showed that the protein regained its native, active tertiary structure when these compounds were removed by dialysis. This experiment indicated that the final folded tertiary structure of a protein can be determined solely by its amino acid sequence. Note, that proteins called chaperones often assist in protein folding in vivo. Chaperones accelerate the rate of folding and help prevent some proteins from getting trapped in an incompletely folded state (see below).

X. Protein folding and stability.

The native conformation of a protein is its lowest free energy structure (**Fig. 4.30**) in which the maximum number of noncovalent stabilizing interactions have formed. Incompletely folded forms of a protein are higher energy structures in which fewer noncovalent interactions occur. The folding process is thought to be driven largely by a collapsing together of hydrophobic residues in the interior of the protein to get out of contact with water. At the same time, secondary structure elements form and begin to associate with one another in a loose arrangement known as a molten globule. Subsequently, more precisely aligned supersecondary structures form. Finally, the approximate tertiary structure is refined as more and more of the total possible noncovalent interactions are established. (Refer to **Fig. 4.31**).

Hydrophobic interactions play an important role in the initial stages of folding as does H-bonding which establishes local secondary structures. Later, H-bonding and van der Waals interactions play important roles in forming contacts within the lowest free energy structure. The types of H-bonds that occur in proteins are summarized in **Table 4.1**. Charge-charge interactions usually don't contribute much to overall structural stability, unless they occur in the nonpolar interior of the folded structure.

XI. Examples of medically important proteins.

A. Collagen.

Collagen is the most abundant protein in mammals making up 25 to 35% of the total protein in the body. Collagen is a fibrous protein meaning it takes on an extended rod-like conformation. It is a key structural component of skin, tendon, and bone and contributes great strength to these tissues.

Each collagen molecule adopts an extended left-handed helical shape. This is a unique structure that is distinctly different from both the α helix and the β conformation. Collagen polypeptides are organized as left-handed helices in which 3 amino acids occur per turn and the pitch is 0.94 nm. Thus, the collagen helix is much more extended than an α helix. Individual polypeptides associate in a triple-helical supercoil called a collagen rod (3,000 Å long) in which the collagen molecules are coiled about each other in a right-handed triple helix (**Fig. 4.36**). Because collagen molecules are extended in structure, collagen fibers do not have much ability to stretch.

The sequence of the collagen polypeptide consists largely of repeats of the form -Gly-X-Y- where X often is proline and Y often is 4-hydroxyproline (**Fig. 4.34**). Glycine with its small side chain allows the chains within the triple helix to come together closely. In addition, the -NH-group of glycine forms an inter-chain H-bond with the carbonyl group on a neighboring chain (**Fig. 4.35**). The hydroxyl group of hydroxyproline also is involved in H-bond stabilization of the triple helix. The synthesis of hydroxyproline requires vitamin C (ascorbic acid), and a deficiency of vitamin C causes the disease scurvy in which symptoms include skin lesions, bleeding gums, etc. Collagen molecules are also cross-linked by linkages involving modified lysine residues (**Fig. 4.37 & 4.38**). A number of hereditary diseases associated with deficiencies in collagen synthesis and cross-linking have been noted. Some of these allow the skin to be stretched to unusual lengths.

B. Myoglobin and hemoglobin.

1. Structure. Myoglobin and hemoglobin are the oxygen carrying proteins found in muscle and erythrocytes, respectively. Myoglobin and hemoglobin are some of the best characterized proteins at the structural level, and they are excellent models for demonstration of protein structure-function principles. Myoglobin is a globular, monomeric protein. It is somewhat unusual in being predominated by α helical structure. In fact, the fold of the peptide consists of 8 α helical segments with no β structure (**Fig. 4.40**). Hemoglobin is a tetrameric protein composed of 2 α and 2 β chains, i.e., it has an $\alpha_2\beta_2$ structure organization. The folds of each type of chain are nearly superimposable with that of myoglobin (**Fig. 4.43**). In the quaternary structure of hemoglobin, the 4 chains are located at the corners of a tetrahedron (**Fig.**

4.42). Each α chain tightly contacts a β chain, making the protein behave essentially as if it were an $(\alpha\beta)_2$ dimer.

Myoglobin and each type of hemoglobin chain contain a tightly bound molecule of heme (**Fig. 4.39**). Heme is called a prosthetic group (after prosthesis) and is the O_2 binding moiety of these proteins. In general, prosthetic groups are organic molecules that are required for the function of a protein. O_2 binds to an Fe^{2+} ion located in the center of the heme group. The heme group is located within an hydrophobic cleft in myoglobin and in each type of hemoglobin chain. Two histidines in the polypeptides interact with the heme iron. When O_2 is bound to the heme, both proteins become a bright red color.

2. Oxygen binding properties of myoglobin and hemoglobin. The oxygenated forms of these proteins are called oxymyoglobin and oxyhemoglobin. Their deoxy forms are called deoxymyoglobin and deoxyhemoglobin. The heme group must be in the ferrous (Fe^{2+}) state to bind O_2 and will not bind O_2 in the oxidized ferric state (Fe^{3+}). Oxidation of the heme group is prevented by sequestering it inside the tightly folded globin chains (**Fig. 4.45**).

The manner in which O_2 binds to the heme is illustrated in **Fig. 4.44**. The iron coordination positions are occupied by 4 nitrogens of the pyrrole groups of the heme and 1 nitrogen from a histidine residue known as the proximal histidine which is closely associated with the Fe^{2+} ion in the so-called 5th coordination position. A second histidine known as the distal histidine is near the Fe^{2+} ion but is not bonded to it. It forces the O_2 molecule to bind to the 6th coordination position of Fe^{2+} with a bent geometry, which O_2 can readily do due to the shape of its molecular orbitals. On the other hand, steric clash between the distal histidine prevents the linear carbon monoxide (CO) molecule from binding with high affinity to the iron, and this helps protect us from CO poisoning.

The myoglobin monomer contains a single O_2 binding site whereas there are 4 O_2 binding sites in hemoglobin. For structural reasons, the binding of O_2 to a given heme in hemoglobin affects the binding of O_2 s to the remaining hemes, as will be explained below. Modulation of the binding affinity of a ligand by the binding of a previous ligand is known as binding cooperativity. Cooperativity can be positive as it is for hemoglobin such that the binding of the first ligand increases the affinity of the protein for subsequent ligands. Alternatively, binding cooperativity can be negative. Cooperative binding is a hallmark of so-called allosteric proteins in which ligand binding affinity is affected by long-range structure interactions between the subunits of the protein. It turns out that O_2 , H^+ and CO_2 all are allosteric modulators of O_2 binding to hemoglobin.

The cooperative binding of O_2 to hemoglobin is evident from comparison of the O_2 binding curves of hemoglobin and myoglobin (**Fig. 4.46**). Myoglobin has a higher affinity for O_2 than hemoglobin. The O_2 binding curve is hyperbolic, indicating that only a single O_2 binding site is present in the molecule. Myoglobin is 50% saturated (i.e., 50% of the molecules have bound an O_2 molecule) at an O_2 partial pressure (pO_2) of 2.8 torr. In contrast, a higher O_2 pressure (26 torr) is required for 50% saturation of hemoglobin. Furthermore, the hemoglobin curve has a sigmoidal shape (s shape), with a small slope at low pO_2 and a larger slope as pO_2 increases. The changing slope indicates that the first O_2 molecule binds with relatively low affinity whereas subsequent molecules bind with higher affinity. As discussed below, changes in binding affinity are due to conformational changes within the hemoglobin tetramer that occur as O_2 molecules are bound.

The physiological consequences of positive binding cooperativity in hemoglobin are enormous. Due to positive O₂ binding cooperativity, hemoglobin can more efficiently deliver O₂ to peripheral tissues. At the O₂ pressure in the lung alveoli, hemoglobin is nearly 100% saturated with O₂, whereas in active muscle hemoglobin is only 32% saturated. Thus during each passage of an erythrocyte through the circulation, hemoglobin molecules give up 66% of their bound O₂ molecules. If binding were not cooperative, and the O₂ pressure required to give 50% saturation were the same, it can be calculated that hemoglobin would only be able to give up 36% of its O₂ molecules during a pass through the peripheral circulation. Thus positive cooperativity improves O₂ delivery by roughly a factor of 2.

3. Structural basis for cooperativity. Positive cooperativity effects can be traced to modification of the number of contacts between the α and β subunits of the hemoglobin tetramer as each O₂ binds. Salt-bridge contacts between side chains of amino acids located near the C-termini of interacting α and β chains help stabilize the structure of deoxyhemoglobin. When O₂ binds to the heme group of a subunit in deoxyhemoglobin, a conformational change occurs in that subunit. Namely, O₂ binding to the 6th coordination position pulls the iron directly into the plane of the heme, and pulls the helix containing the proximal histidine towards the heme as well (**Fig. 4.47**). Movement of the helix changes the shape of the subunit and alters how it contacts the other subunits; namely some of the salt-bridges are broken by O₂ binding. Due to disruption of salt-bridges, it becomes easier for the next O₂ to bind to the second subunit. In like fashion, binding of the second O₂ breaks additional salt-bridges between subunits and makes it easier for the third O₂ to bind, etc. Conversely, dissociation of the first O₂ from one subunit in oxyhemoglobin makes the dissociation of the next (then the next, etc.) easier. Because the structure of deoxyhemoglobin is constrained by salt-bridges its is referred to as the tense (T) form of hemoglobin. Oxyhemoglobin is much less constrained and is referred to as the relaxed (R) form. In summary, hemoglobin is an allosteric protein in which O₂ is an positive allosteric effector. Binding of successive O₂ molecules reduces the number of contacts between the 4 subunits increasing the O₂ binding affinity of the remaining subunits.

4. Allosteric effectors. 2,3-bisphosphoglycerate (2,3BPG) (**Fig. 4.48**) is a negative effector of O₂ binding to hemoglobin. 2,3BPG, which is present in red blood cells, lowers the affinity of hemoglobin for O₂ allowing it to release O₂ better at low pO₂s. BPG binds to hemoglobin in a cavity located between the amino-termini of the 2 β chains (**Fig. 4.49**). Binding is mediated by negatively charged phosphate groups in 2,3BPG and positively charged side chains in the β chains. The salt-bridges between 2,3BPG and the β chains provide additional interactions that must be broken on oxygenation. Thus 2,3BPG stabilizes deoxyhemoglobin and shifts the conformational equilibrium towards the T form, which is better able to release O₂ in the peripheral tissues. Indeed, without 2,3BPG the O₂ binding properties of hemoglobin become very much like that of myoglobin. At high altitude a higher concentration of 2,3BPG is synthesized to make O₂ delivery even more efficient under lower pO₂ conditions. Interestingly, the fetal form of hemoglobin lacks a high affinity binding site for 2,3BPG making it easier for O₂ to be transferred from maternal to fetal blood.

H⁺ and CO₂ also are allosteric effectors of O₂ binding to hemoglobin. When CO₂ is produced by metabolism and dissolves in the blood, an enzyme known as carbonic anhydrase catalyzes its combination with a water molecule forming carbonic acid. Carbonic acid then dissociates a proton forming bicarbonate, and blood pH will decrease. However, some protons

end up binding to basic functional groups in the hemoglobin molecule, and this helps in buffering the pH of blood. In addition, the protonation of basic groups leads to formation of salt-bridges that stabilize the deoxy form of hemoglobin. This results in a shift in the O₂ dissociation curve to the right, i.e., results in a decrease in O₂ binding affinity (**Fig. 4.50**). O₂ delivery to peripheral tissues where the pH is lower than in the lung is thereby improved. In the lungs the expulsion of CO₂ reverses the carbonic acid dissociation reaction and causes protons to be lost from hemoglobin which increased its O₂ binding affinity.

A large amount of the CO₂ produced in respiring tissue actually is carried covalently attached to hemoglobin. CO₂ groups combine with the free amino groups of the N-terminal residues of the β chains forming carbamate derivatives (**Fig. 4.51**). The carbamates form salt-bridges that stabilize the deoxy form and promote release of O₂ in the peripheral tissues.