Protein Structure

OBJECTIVES: After completing this exercise you should be able to:
* Describe the nature of an α amino acid.
* Explain the role of amino acids in proteins.
* Describe the roles of peptide bonds, disulfide bonds, ionic bonds, hydrogen bonds, and van der Waals attractions in protein structure.
* Define primary, secondary, tertiary and quaternary structure in proteins.
* Name two common protein secondary structures.

KEY WORDS

α helix
α amino acid
amino acid residue
amino acid side chain
ATP (adenosine triphosphate)
ß sheet
C terminus
covalent modification
disulfide bond
domain
homology
hydrogen bond (H bond)
hydrophobic effect

N terminus
native conformation
peptide bond
polypeptide
post-translational modification
proteolytic cleavage
primary structure quaternary
structure secondary structure
subunit
tertiary structure
van der Waals interactions

INTRODUCTION

It has been estimated that human cells can synthesize 100,000 different proteins, each of which has a different structure and accomplishes a different function. Proteins serve as catalysts, as structural components of the cell and the surrounding extracellular matrix, as molecular signals and signal receptors, as molecular motors, and as vehicles that transport and store other classes of molecules. This review is included in the curriculum at the beginning of the Prologue because many of the histology, cell biology and pharmacology lectures will build upon the concepts in it.

I.   PROTEINS ARE UNBRANCHED POLYMERS OF AMINO ACIDS.

Proteins are macromolecules. Typical proteins have molecular weights that range from 5000 to 350,000. Although they are large, proteins are relatively simple in

design. Each is a linear, unbranched polymer made up of L-α-amino acids. As shown in Figure 1, an α amino acid is a carboxylic acid that has an amino group, a hydrogen atom and a variable R group attached to its α carbon. Proteins are constructed using the twenty different amino acids shown in Table I.

Nineteen of the 20 amino acids differ only in the identity of the R group. But proline is different because it has a ring structure that includes its α carbon, α amino group, and the R group. Amino acid names are often abbreviated using the three-letter or one-letter abbreviations shown in Table I, e.g., arginine = arg = R.

Table I. The twenty amino acids used in protein synthesis.

*The structure shown is that which predominates at pH 7.4.

To form a protein, individual amino acids must be linked together in a chain by means of the α amino and α carboxyl groups. As each amino acid is added to the growing chain, its α amino group is joined to the α carboxyl group of the preceding amino acid. The covalent bond formed in the process is called a **peptide bond** (Figure 2). Consequently, a protein can also be called a **polypeptide**. After it has
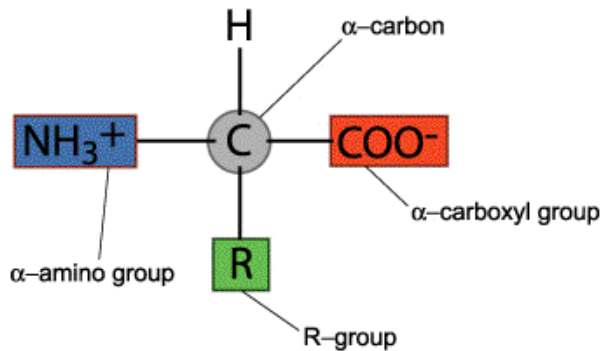
Figure 1. The general structure of an a amino acid. (Courtesy of Tony Le.)

been incorporated into a protein, each amino acid unit is referred to as an **amino acid residue** (Figure 3). Although most of the α amino and α carboxyl groups react to form peptide bonds, the first amino acid in the chain (referred to as the **N terminus**) still has an unreacted α amino group and the last amino acid incorporated (referred to as the **C terminus**) still has an unreacted α carboxyl group. By convention, the amino acids in a polypeptide chain are numbered from N terminus to C terminus. These numbers can be combined with one-letter or three-letter abbreviations to specify the amino acid present at a given position in the chain. For example, if a phenylalanine residue occupies position 508 of a polypeptide chain, that residue can be referred to as F508 or Phe508.

Figure 2. Amino acid residues in proteins are linked by peptide bonds.

Figure 3. The structure of a polypeptide chain containing five amino acid residues.

Polymerization of amino acids to form a polypeptide chain results in a structure that has a monotonous "backbone" with a repeating sequence formed by the a carbons and the atoms that participate in the peptide bond. The variety in protein structure comes from the R groups (also called **side chains**) attached to the a carbons. Each species of protein (such as hemoglobin or collagen) has a different sequence of side chains.

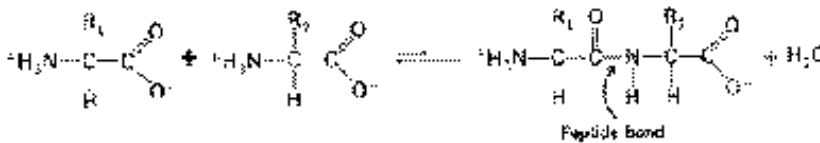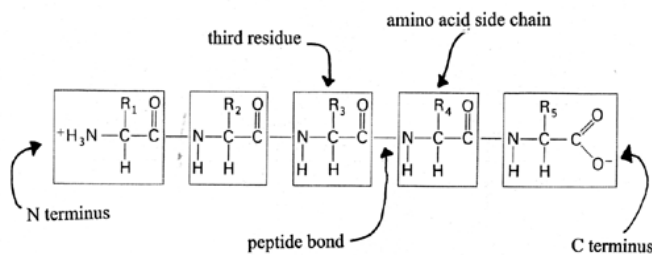## II.   MOST POLYPEPTIDE CHAINS FOLD INTO COMPACT, THREE-DIMENSIONAL SHAPES STABILIZED BY NONCOVALENT BONDS.

After they are synthesized, most proteins fold into compact, three-dimensional shapes. Folding is an essential step in the synthesis of a protein because proteins can perform their functions only if they can bind to other molecules. To do so, they must have specific binding sites within their structures. These binding sites are formed as the protein folds from its original extended form into a more compact, three-dimensional shape.

In its natural environment, each protein folds into a specific shape known as its **native conformation**. As a protein folds, different parts of the chain are brought close to each other and have the opportunity to form **noncovalent bonds** that stabilize the folded structure. Amino acids that are quite distant from each other in the polypeptide chain (e.g., Glu23 and His117) may be in contact with each other in the folded structure and therefore can form a noncovalent bond.

The folding pattern of a protein is ultimately dictated by the sequence of amino acids along the polypeptide chain because the sequence of side chains dictates the types of noncovalent bonds that can be formed. To understand why proteins adopt specific conformations, we need to understand what types of noncovalent bonds

Table II. Types of chemical groups and the nonvcovalent bonds they can form.

| GROUP TYPE	DEFINITION	EXAMPLES	BONDING OPTIONS
Charged polar	A group that has gained or lost electrons and therefore carries a full positive or negative charge	Amino group: full positive charge at pH 7. Carboxy group: full negative charge at pH 7.	Ionic Bond H bond van der Waals interactions
Neutral Polar	A group that does not carry a net charge but in which the electrons are unevenly distributed.	Hydroxyl group: no net charge, but electrons spend more time near the oxygen atom. Oxygen atom carries a partial negative charge and hydrogen atom a partial positive charge	H bond van der Waals interactions
Nonpolar	A group that does not carry either a net charge or a partial charge.	Methyl group Side chain of valine	van der Waals interactions
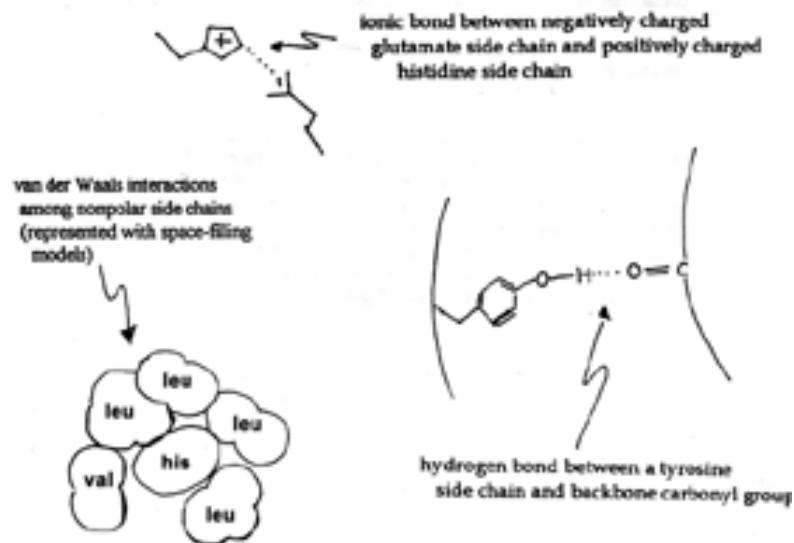
Figure 4. Examples of noncovalent interactions that are important in protein folding.

each part of the polypeptide chain can participate in (see Figure 4 and Table II). Only charged polar groups can form **ionic bonds** (the strongest type of noncovalent bond). Within the physiological range of pH (pH 5 – 8), the charged polar groups in proteins include the N terminus and C terminus as well as the side chains of certain amino acids (aspartate, glutamate, lysine, arginine and histidine). Both charged polar groups and neutral polar groups can form **hydrogen bonds** (H bond; the second strongest type of noncovalent bond). The neutral polar groups of proteins include the side chains of certain amino acids (serine, threonine, tyrosine, cysteine, asparagine, and glutamine) and the amino and carbonyl groups of the polypeptide backbone. All atoms can participate in **van der Waals interactions** (the weakest type of noncovalent attraction) but they are particularly important for nonpolar groups because nonpolar groups cannot form ionic or hydrogen bonds.

III. MOST PROTEINS FOLD IN AQUEOUS ENVIRONMENTS AND CONSEQUENTLY ARE AFFECTED BY THE POLARITY OF WATER.

Because both the interior of the cell and the extracellular space are aqueous, most proteins fold in an aqueous environment, and this influences how they fold. As shown in Figure 5, water is polar. The oxygen atom carries a partial negative charge and each of the hydrogen atoms a partial positive charge. Even in liquid water, most of the water molecules are H-bonded to other water molecules at any given moment. Compounds that have polar groups, such as hydroxyl, carboxyl, and amino groups, are termed **hydrophilic** (water-loving) because they form H bonds to water and therefore can easily dissolve in it. Nonpolar groups, such as methyl groups, are termed **hydrophobic** (water-hating) because they do not easily dissolve in water. They do not themselves form H bonds with the water molecules, and furthermore their presence in water prevents the water molecules from freely H bonding with each other. When a nonpolar substance is added to water, it tends to form a separate phase (think of mixing oil and water). This separation of polar and
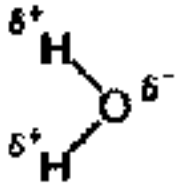
Figure 5: Water is polar.

nonpolar compounds into two phases is referred to as the **hydrophobic effect**.

The hydrophobic effect is central to protein folding. Each protein folds into the shape that maximizes the number of noncovalent bonds that both it **and** the water around it can form. A key factor in this process is the segregation of nonpolar side chains of the protein away from the polar groups of the protein and away from water. Proteins that fold in aqueous environments segregate as many of the nonpolar side chains as possible into a nonpolar region in the core of the protein, while most of the polar side chains end up in contact with the aqueous environment. However in most cases no folding pattern allows all of the hydrophobic side chains to move to the core of the folded structure. Thus a few nonpolar residues may end up on the surface of the protein. Furthermore some polar side chains and many of the H-bonding groups of the polypeptide backbone end up in the core of the protein. The most stable folded structure is one that allows H-bonding groups trapped inside the protein to bond with each other.

IV.   **PROTEINS ARE DESCRIBED IN TERMS OF FOUR LEVELS OF STRUCTURE: PRIMARY, SECONDARY, TERTIARY, AND QUATERNARY.**

A.   The **primary structure** of a protein consists of its linear sequence of amino acids. The primary structure of a protein determines its folding pattern and its function.

B.   The term **secondary structure** refers to characteristic _regular folding patterns_ that involve _only part of the polypeptide chain_. The regularity of secondary structures is due to the regular occurrence of H-bonding groups along the polypeptide chain – the carbonyl groups and amino group of the peptide bond. Carbonyl groups are H-bond acceptors and amino groups are H-bond donors. Relatively few of these backbone groups can be located at the surface of the protein where they could form H-bonds to water molecules and instead are trapped within the hydrophobic core of the protein. The most stable folding pattern of the protein is one that allows these trapped H-bonding groups to find bonding partners within the core. Secondary structures accomplish this by allowing the backbone carbonyl groups to bond with backbone amino groups. There are two major types of protein secondary structure: helices and ß sheets.

1.   **Helices**: The polypeptide chain can find H-bonding partners for all of the carbonyl and amino groups along its backbone if it folds into a helix. One type of
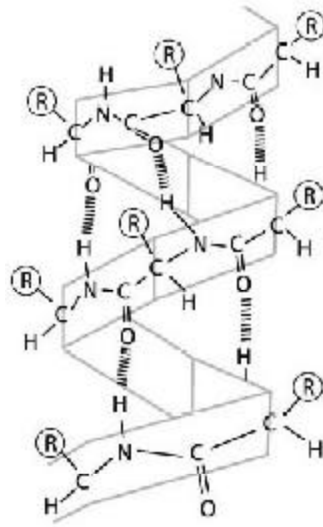
Figure 6. Structure of an α helix. (Reproduced with permission from D. Colby, Biochemistry: A Synopsis, Lange, 1985.)

helix commonly found in proteins is called the **a helix** (Figure 6). In an α helix, each peptide carbonyl group is H-bonded to a peptide amino group 4 amino acids later in the chain (amino acids n and n + 4 are linked). The polypeptide backbone lies in the core of the helix. The amino acid side chains project to the outside and coat its surface. Because the side chains project to the outside, the surface properties of a helix are dictated by the properties of the side chains. For example, a helix made entirely of amino acids with nonpolar side chains has a nonpolar surface. A helix of this type is likely to be found in the hydrophobic core of a protein or spanning the hydrophobic core of a biologic membrane (see Independent Learning Module: The Components and Properties of Cell Membranes). In contrast, an α helix located at the surface of a protein is likely to have one hydrophilic side and one hydrophobic side. Since it takes four amino acid residues to complete one turn of an α helix, the primary structure of this type of helix alternates two polar amino acids with two nonpolar ones.

2.   **The beta (ß) sheet**: A polypeptide may also assume a relatively extended conformation in which successive side chains project from the polypeptide backbone in opposite directions. When two or more such extended sections of chain are aligned on the same axis, they form another secondary structure called a **ß sheet**. In a ß sheet, the polypeptide backbone of one strand is hydrogen bonded to the polypeptide backbone of another strand (Figure 7). ß sheets can contain as few as two strands but they commonly contain as many as eight.

C.   The **tertiary structure** of a protein corresponds to the overall folding of the complete chain. What generalizations can we make about tertiary structure?

1.   Water-soluble proteins fold so as to form hydrophobic cores. Within the core of the protein, there is a minimum of unfilled space and groups are in very close contact. Because each amino acid side chain has a specific shape, the
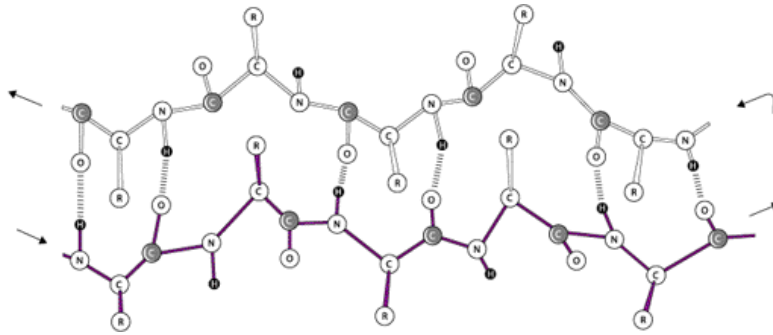
Figure 7. Two strands of a beta sheet. Arrows indicate the orientation of each strand in the N terminal to C terminal direction. In this example, the two strands are antiparallel to each other, i.e., the N terminus of one strand is H bonded to the C terminus of the other strand. (Courtesy of Jesse Gray and Tony Le.)

inside of a protein is like a three-dimensional jig-saw puzzle. The folding pattern that the protein adopts must provide enough room for each side chain without leaving big holes. This means that two segments of the proteins that fit together in the core of the protein must have surfaces that are complementary in shape. Although the van der Waals interactions in the core are individually weak, they make an important contribution to the stability of the folding pattern.

2.   Approximately 60% of the residues of a typical protein are involved in secondary structures. The parts of the protein that are not involved in secondary structure are folded irregularly but, in most cases, have a definite folding pattern.

3.   Some polypeptide chains have regions that fold independently. Each independent folding unit within a polypeptide chain is called a **domain**.

4.   Proteins that have related amino acid sequences are described as having **homology**.

5.   Proteins are not rigid structures. Because the noncovalent bonds that stabilize their tertiary structures are relatively weak, the folding pattern or conformation can rearrange if the protein interacts with another molecule.

Biochemists have several different ways of representing protein structures. Figure 8 shows three of the most commonly used models – the space filling model, the wire model, and the ribbon model.

D.   Many proteins consist of stable assemblies of multiple polypeptide chains or **subunits**. These proteins are said to have **quaternary structure**. A protein that has a quaternary structure may be made up of several copies of the same

WIRE MODEL

RIBBON MODEL

example of
β–strand

example of
α–helix
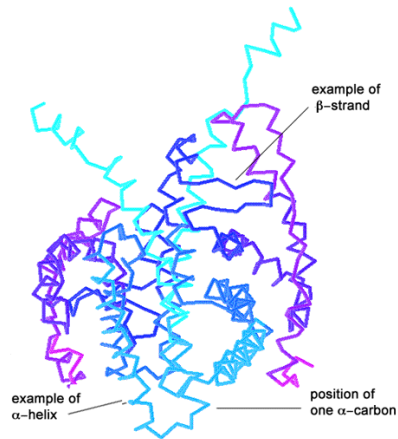
position of
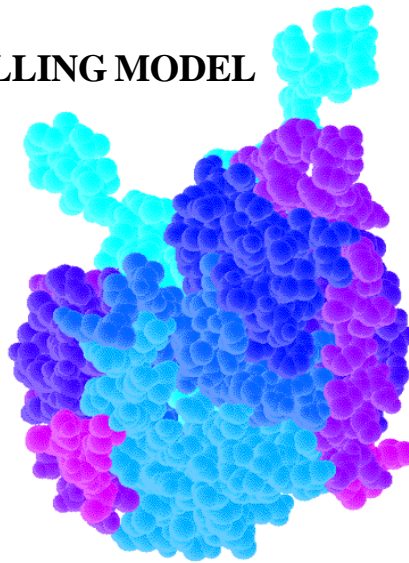one α–carbon
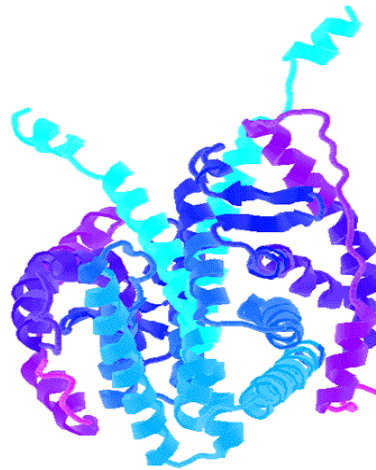
SPACE FILLING MODEL

Figure 8. Three ways to represent the three-dimensional structure of a protein. Shown is a part of the estrogen receptor. In the backbone wire model, each bend in the chain represents the position of an α carbon. Other atoms are not represented. In the ribbon model, the ribbon follows the path of the backbone. ß strands are represented as arrows, with the arrowhead representing the C terminus of that part of the chain. (Courtesty of Tony Le.)

polypeptide or may contain more than one type of polypeptide. Quaternary structures are brought together by the same forces that determine folding of tertiary structures (the hydrophobic effect). The individual polypeptide chains in a quaternary structure have hydrophobic patches on their surfaces, and assembly of the chains into a complex is favored because it removes these hydrophobic patches from the surrounding water. In order to assemble into a complex, two polypeptides
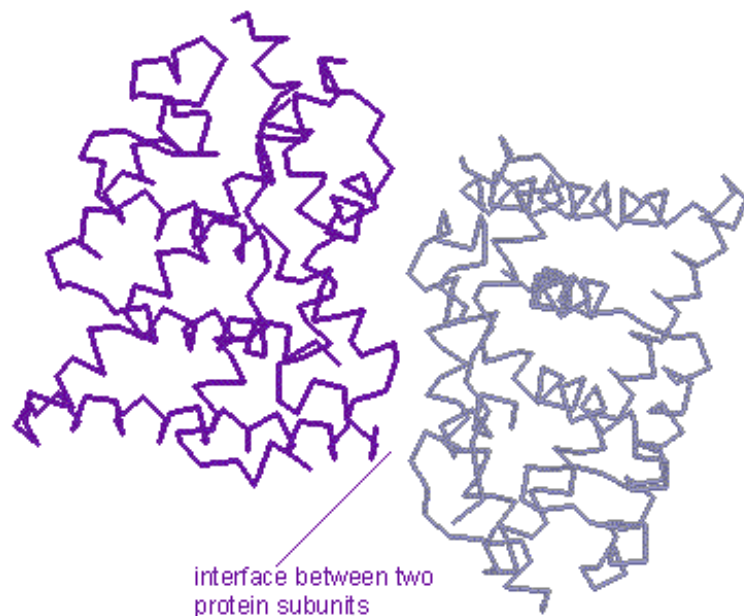

Figure 9. Another view of a portion of the estrogen receptor showing its quaternary structure. (Courtesy of Tony Le).

interface between two protein subunits

must have tertiary structures with complementary surfaces. Again, the structures fit together like jig-saw puzzles (Figure 9).

V.   SOME PROTEIN FOLDING PATTERNS ARE STABILIZED BY COVALENT CROSSLINKS.

Not all proteins produced by eukaryotic cells remain in the cytoplasm where they are synthesized. Many enter the endoplasmic reticulum and are sorted from there to other compartments of the cell or to the extracellular space. Although the conformation into which such a protein folds is determined by noncovalent interactions, its pattern may be further stabilized by covalent bonds between pairs of cysteine residues. These bonds, which are termed **disulfide bonds** or **S-S bridges**, are formed when two cysteine residues are brought into close proximity by the folding process. During the maturation of the protein in the endoplasmic reticulum, the adjacent cysteine residues are oxidized (Figure 10). Some disulfide
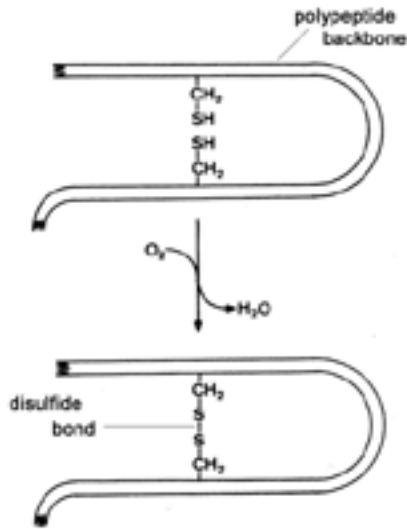
Figure 10. Formation of a disulfide bond between cysteine side chains in a protein. (Reproduced with permission from Colby, *Biochemistry, A Synopsis*, Lange, 1985.)

bonds link two regions of the same chain and stabilize its tertiary structure. Others link two different chains.

VI.   SOME PROTEINS ARE MODIFIED DURING OR AFTER FOLDING.

Proteins are synthesized using the catalog of 20 amino acids shown in Table I. But an examination of naturally occurring proteins shows that many contain unusual amino acids and a variety of covalently attached groups, such as carbohydrates, fatty acids and phosphate groups. Some of these modifications provide the protein with groups needed to perform specific functions, some help target the protein to membranes, and others are used to regulate the activity of the protein.   The process by which polypeptide chains are made in the cell is called *translation*. Attachment of a group to the polypeptide chain after it has been synthesized is termed **post-translational modification**.