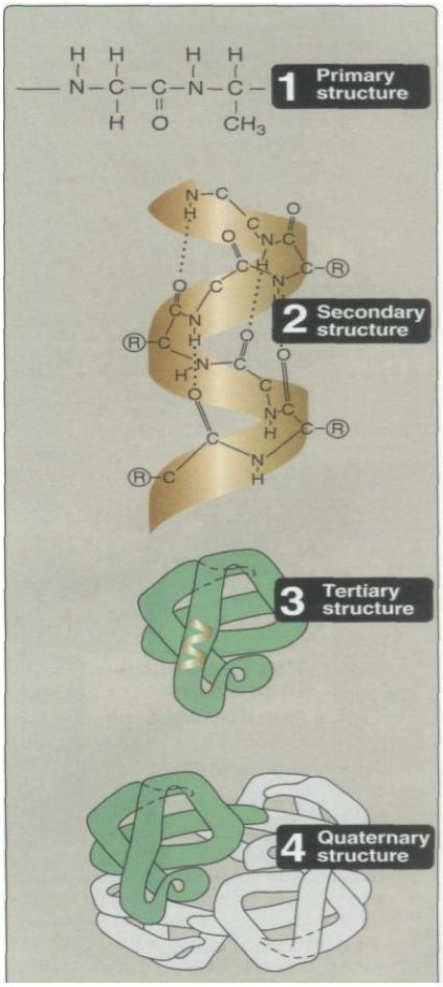
Protein structure

The twenty amino acids commonly found in proteins are joined together by peptide bonds. The linear sequence of the linked amino acids contains the information necessary to generate a protein molecule with a unique three-dimensional shape. The complexity of protein structure is best analyzed by considering the molecule in terms of four organizational levels, namely, primary, secondary, tertiary, and quaternary.

PRIMARY STRUCTURE OF PROTEINS



The sequence of amino acids in a protein is called the primary structure of the protein. Understanding the primary structure of proteins is important because many genetic diseases result in proteins with abnormal amino acid sequences, which cause improper folding and loss or impairment of normal function. If the primary structures of the normal and the mutated proteins are known, this information may be used to diagnose or study the disease.

A. Peptide bond

Proteins, amino acids are joined covalently by peptide bonds, which are amide linkages between the α-carboxyl group of one amino acid, and the α-amino group of another. For example, valine and alanine can form the dipeptide valylalanine through the formation of a peptide bond Peptide bonds are not broken by conditions that denature proteins, such as heating or high concentrations of urea. Prolonged exposure to a strong acid or base at elevated temperatures is required to hydrolyze these bonds nonenzymically.

1. Naming the peptide: By convention, the free amino end of the peptide chain (N-Terminal) is written to the left and the free carboxyl end (C-Terminal)to the right. Therefore, all amino sequences are read from the N- to the C-terminal end of the peptide

2. Characteristics of the peptide bond: The peptide bond has a partial double-bond character that is, it is shorter than a single bond, and is rigid and planar. This prevents free rotation around the bond between the carbonyl carbon and the nitrogen of the peptide bond. However, the bonds between the α-carbons and the α-amino or α-carboxyl groups can be freely rotated (although they are limited by the size and character of the R-groups). This allows the polypeptide chain to assume a variety of possible configurations. The peptide bond

is generally a trans bond in large part because of steric interference of the R-groups when in the cis position.

3.Polarity of the peptide bond: Like all amide linkages, the -C= 0 and -N H groups of the peptide bond are uncharged, and neither accept nor release protons over the P^H range of 2 to 12. Thus, the charged groups present in polypeptides consist solely of the N-terminal α-amino group, the C-terminal α-carboxyl group, and any ionized groups present in the side chains of the constituent amino acids.

B. Determination of the amino acid composition of a polypeptide

The first step in determining the primary structure of a polypeptide is to identify and quantitate its constituent amino acids. A purified sample of the polypeptide to be analyzed is first hydrolyzed by strong acid at 110^0C for 24 hours. This treatment cleaves the peptide bonds, and releases the individual amino acids, which can be separated by cation-exchange chromatography. In this technique, a mixture of amino acids is applied to a column that contains a resin to which a negatively charged group is tightly attached. The amino acids bind to the column with different affinities, depending on their charges, hydrophobicity, and other characteristics. Each amino acid is sequentially released from the chromatography column by eluting with solutions of increasing ionic strength and pH. The separated amino acids contained in the elute from the column are quantitated by heating them with Ninhydrin-a reagent that forms a purple compound with most.

C. Sequencing of the peptide from its end

Sequencing is a stepwise process of identifying the specific amino acids at each position in the peptide chain, beginning at the N-Terminal end. Phenylisothiocyanate, known as Edman's reagent, is used to label the amino terminal residue under mildly alkaline conditions. The resulting phenylthiohydantoin (PTH) derivative introduces an instability in the N-terminal peptide bond that can be selectively hydrolyzed without cleaving the other peptide bonds. The identity of the amino acid derivative can then be determined. Edman's reagent can be applied repeatedly to the shortened peptide obtained in each previous cycle. This process has been automated and, currently, the repetition of the method can be employed by a machine (sequenator) to determine the sequence of more than 100 amino acid residues, starting at the amino terminal end of a polypeptide.

D. Cleavage of the polypeptide into smaller fragments

Many polypeptides have a primary structure composed of more than100 amino acids. Such molecules cannot be sequenced directly from end to end by a sequenator. However, these large molecules can be cleaved at specific sites, and the resulting fragments sequenced. By using more than one cleaving agent (enzymes and/or chemicals) on separate samples of the purified polypeptide, overlapping fragments can be generated that permit the proper ordering of the sequenced fragments, thus providing a complete amino acid sequence of the large polypeptide.

E. Determination of a protein's primary structure by DNA sequencing

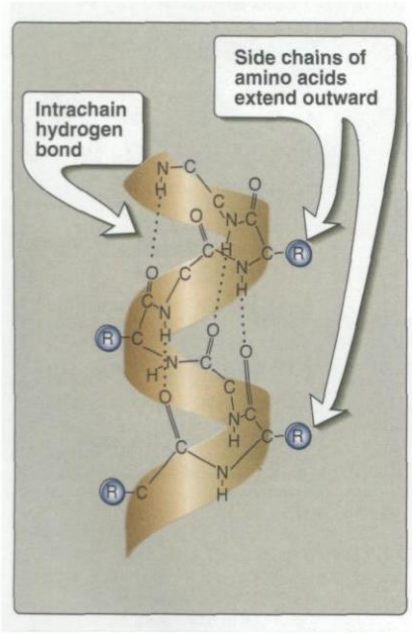
The sequence of nucleotides in a coding region of the DNA specifies the amino acid sequence of a polypeptide. Therefore, if the nucleotide sequence can be determined, it is possible, from knowledge of the genetic code, to translate the sequence of nucleotides into the corresponding amino acid sequence of that polypeptide. This process, although routinely used to obtain the amino acid sequences of proteins, has the limitations of not being able to predict the positions of disulfide bonds in the folded chain, and not identifying any amino acids that are modified after their incorporation into the polypeptide post translational modification,. Therefore, direct protein sequencing is an extremely important tool for determining the true character of the primary sequence of many polypeptides.

SECONDARY STRUCTURE OF PROTEINS

The polypeptide backbone does not assume a random three-dimensional structure, but instead generally forms regular arrangements of amino acids that are located near to each other in the linear sequence. These arrangements are termed the **secondary structure** of the polypeptide. The α-helix, β-sheet, and β-bend are examples of secondary structures frequently encountered in proteins.

A. α- Helix

There are several different polypeptide helices found in nature, but the α-helix is the most common. It is a spiral structure, consisting of a tightly packed, coiled polypeptide backbone core, with the side chains of the component amino acids extending outward from the central axis to avoid interfering sterically with each other. A very diverse group of proteins contains α-helices. For example, the keratins are a family of closely related, fibrous proteins whose structure is nearly entirely α-helical. They are a major component of tissues such as hair and skin, and their rigidity is determined by the number of disulfide bonds between the constituent polypeptide chains. In contrast to keratin, myoglobin, whose structure is approximately eighty percent α-helical, is a globular, flexible molecule.

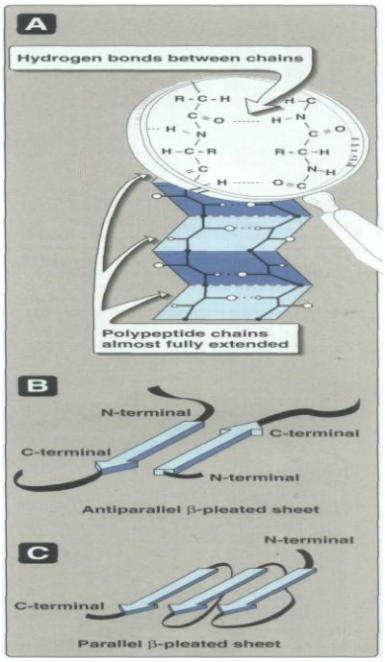


1. **Hydrogen bonds:** An α-helix is stabilized by extensive hydrogen bonding between the peptide-bond carbonyl oxygens and amide hydrogens that are part of the polypeptide backbone. The hydrogen bonds extend up the spiral from the carbonyl oxygen of one peptide bond to the -NH - group of a peptide linkage four residues ahead in the polypeptide. This ensures that all but the first and last peptide bond components are linked to each other through hydrogen bonds. Hydrogen bonds are individually weak, but they collectively serve to stabilize the helix.

2. Amino acids per turn: Each turn of an α-helix contains 3.6 amino acids. Thus, amino acid residues spaced three or

four apart in the primary sequence are spatially close together when folded in the α-helix.

3. Amino acids that disrupt an α-helix: Proline disrupts an α-helix because its imino group is not geometrically compatible with the right-handed spiral of the α-helix. Instead, it inserts a kink in the chain, which interferes with the smooth, helical structure. Large numbers of charged amino acids (for example, glutamate aspartate, histidine, lysine, or arginine) also disrupt the helix by forming ionic bonds, or by electrostatically repelling each other. Finally, amino acids with bulky side chains, such as tryptophan, or amino acids, such as valine or isoleucine, that branch at the β-carbon (the first carbon in the R-group, next to the α-carbon) can interfere with formation of the α-helix if they are present in large numbers.



B. β-sheet

The β-sheet is another form of secondary structure in which all of the peptide bond components are involved in hydrogen bonding. The surfaces of β-sheets appear "pleated," and these structures are, therefore, often called **"β-pleated sheets."** When illustrations are made of protein structure, β-strands are often visualized as broad arrows.

1. Comparison of a β-sheet and an α-helix: Unlike the α-helix, β-sheets are composed of two or more peptide chains (β-strands), or segments of polypeptide chains, which are almost fully extended. Note also that in β-sheets the hydrogen bonds are perpendicular to the polypeptide backbone.

2. **Parallel and antiparallel sheets:** A β-sheet can be formed from two or more separate polypeptide chains or segments of polypeptide chains that are arranged either antiparallel to each other (with the and ends of the β-strands alternating as shown in Figure B, or parallel. When the hydrogen bonds are formed between the

Poly peptide backbones of separate polypeptide chains, they are termed **inter-chain bonds. A** β-sheet can also be formed by a single polypeptide chain folding back on itself (see Figure C). In this case, the hydrogen bonds are **intrachain bonds. In** globular proteins, β-sheets always have a right-handed curl, or twist, when viewed along the polypeptide backbone.

C. β bends (reverse turns)

β - bends reverse the direction of a polypeptide chain, helping it forma compact, globular shape. They are usually found on the surface of protein molecules, and often include charged residues. β - bends are generally composed off our amino acids, one of which may be

Proline the imono acid that causes a "kink" in the polypeptide chain. Glycine, the amino acid with the smallest R-group, is also frequently found in β-bends. β -Bends are stabilized by the formation of hydrogen and ionic bonds.

D. Non repetitive secondary structure

Approximately one half of an average globular protein is organized into repetitive structures, such as the α-helix and/or β-sheet. The remainder of the polypeptide chain is described as having a loop or coil conformation. These nonrepetitive secondary structures are not "random," but rather simply have a less regular structure.

E. Supersecondary structures (motifs)

Globular proteins are constructed by combining secondary structural elements (α-helices, β-sheets, nonrepetitive sequences). These form primarily the core region-that is, the interior of the molecule. They are connected by loop regions (for example, β-bends) at the surface of the protein. Super secondary structures are usually produced by packing side chains from adjacent secondary structural elements close to each other. Thus, for example, α-helices and β-sheets that are adjacent in the amino acid sequence are also usually (but not always) adjacent in the final, folded protein. Some of the more common motifs are illustrated in Figure.

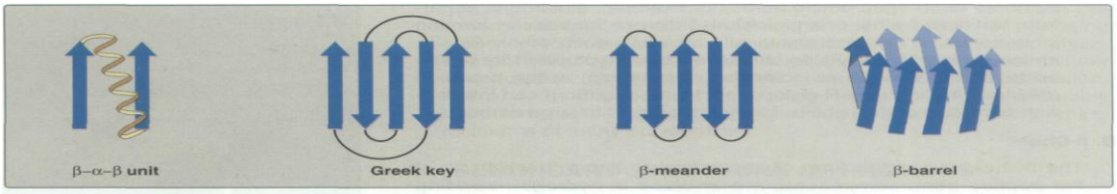


Figure common structural motifs

TERTIARY STRUCTURE OF GLOBULAR PROTEINS

The primary structure of a polypeptide chain determines its **tertiary structure.** [Note: "Tertiary" refers both to the folding of domains (the basic units of structure and function, see discussion below), and the final arrangement of domains in the polypeptide.] The structure of globular proteins in **aqueous solution** is compact, with a high-density (close packing) of the atoms in the core of the molecule. **Hydrophobic side chains** are buried in the **interior,** whereas **hydrophilic groups** are generally found on the **surface** of the molecule. All hydrophilic groups (including components of the peptide bond) located in the interior of the polypeptide are involved in hydrogen bonds or electrostatic interactions.

A. Domains

Domains are the fundamental functional and three-dimensional structural units of a polypeptide. Polypeptide chains that are greater than 200 amino acids in length generally consist of two or more domains. The core of a domain is built from combinations of **super-secondary structural elements (motifs).** Folding of the peptide chain within a domain usually occurs independently of folding in other

domains. Therefore, each domain has the characteristics of as mall, compact globular protein that is structurally independent of the other domains in the polypeptide chain.

B. Interactions stabilizing tertiary structure

The unique three-dimensional structure of each polypeptide is determined by its amino acid sequence. Interactions between the amino acid side chains guide the folding of the polypeptide to form a compact structure. Four types of interactions cooperate in stabilizing the tertiary structures of globular proteins.

1. **Disulfide bonds:** A disulfide bond is a covalent linkage formed from the sulfhydryl group (-SH) of each of **two cysteine residues,** to produce a **cysteine** residue (Figure 2.9). The two cysteine's may be separated from each other by many amino acids in the primary sequence of a polypeptide, or may even be located on two different polypeptide chains; the folding of the polypeptide chain(s)brings the cysteine residues into proximity, and permits covalent bonding of their side chains. A disulfide bond contributes to the stability of the three-dimensional shape of the protein molecule. For example, many disulfide bonds are found in proteins such as immunoglobulin's that are secreted by cells.

2. Hydrophobic interactions: Amino acids with nonpolar side chains tend to be located in the interior of the polypeptide molecule, where they associate with other hydrophobic amino acids. In contrast, amino acids with polar or charged side chains tend to be located on the surface of the molecule in contact with the polar solvent. Proteins located in nonpolar (lipid) environments, such as a membrane, exhibit the reverse arrangement – that is, hydrophilic amino acid side chains are located in the interior of the polypeptide, whereas hydrophobic amino acids are located on the surface of the molecule in contact with the nonpolar environment In each case, the segregation of R-groups occurs that is energetically most favorable.

3. Hydrogen bonds: Amino acid side chains containing oxygen- or nitrogen-bound hydrogen, such as in the alcohol groups of serine and threonine, can form hydrogen bonds with electron-rich atoms, such as the oxygen of a carboxyl group or carbonyl group of a peptide bond Formation of hydrogen bonds between polar groups on the surface of proteins and the aqueous solvent enhances the solubility of the protein.

4. Ionic interactions: Negatively charged groups, such as the carboxyl group ($-COO^-$ in the side chain of aspartate or glutamate, can interact with positively charged groups, such as the amino group ($-NH_3^+$ in the side chain of lysine.

C. Protein folding

Interactions between the side chains of amino acids determine how a long polypeptide chain folds into the intricate three-dimensional shape of the functional protein. Protein folding, which occurs within the cell in seconds to minutes, employs a shortcut through the maze of all folding possibilities. As a peptide folds, its amino acid side chains are attracted and repulsed according to their chemical properties. For example, positively and negatively charged side chains attract each other. Conversely, similarly charged side

chains repel each other. In addition, interactions involving hydrogen bonds, hydrophobic interactions, and disulfide bonds all seek to exert an influence on the folding process. This process of trial and error tests many, but not all, possible configurations, seeking a compromise in which attractions outweigh repulsions. This results in a correctly folded protein with a low energy state

D. Role of chaperones in protein folding

It is generally accepted that the information needed for correct protein folding is contained in the primary structure of the polypeptide. Given that premise, it is difficult to explain why most proteins when denatured do not resume their native conformations under favorable environmental conditions. One answer to this problem is that a protein begins to fold in stages during its synthesis, rather than waiting for synthesis of the entire chain to be totally completed. This limits competing folding configurations made available by longer stretches of nascent peptide. In addition, a specialized group of proteins, named **"chaperones,"** are required for the proper folding of many species of proteins. The chaperones- also known as **"heat shock" Protein-interact** with the polypeptide at various stages during the folding process. Some chaperones are important in keeping the protein unfolded until its synthesis is finished, or act as catalysts by increasing the rates of the final stages in the folding process. Others protect proteins as they fold so that their vulnerable, exposed regions do not become tangled in unproductive encounters.

QUATERNARY STRUCTURE OF PROTEINS

Many proteins consist of a single polypeptide chain, and are defined as monomeric **proteins.** However, others may consist of two or more polypeptide chains that may be structurally identical or totally unrelated. The arrangement of these polypeptide subunits is called the quaternary structure of the protein. [Note: If there are two subunits, the protein is called dimeric if three subunits trimeric and, if several subunits, multimeric Subunits are held together by non-covalent interactions (for example, hydrogen bonds, ionic bonds, and hydrophobic interactions). Subunits may either function independently of each other, or may work cooperatively, as in hemoglobin, in which the binding of oxygen to one subunit of the tetramer increases the affinity of the other subunit oxygen.

References

1. Wikipedia - https://en.wikipedia.org/wiki/Protein_structure
2. Introduction to Protein Structure 2nd Edition by Carl Branden, John Tooze