Three-dimensional structure of protein

Although proteins are large molecules, they are discrete chemical entities with unique structures.

- The three-dimensional structure of a protein is determined by its amino acid sequence.
- The function of a protein depends on its structure.
- An isolated protein has a unique, or nearly unique structure.
- The most important forces stabilizing the specific structure maintained by a given protein are noncovalent interactions.
- Amid the huge number of unique protein structures, some common structural patterns can help us organize the protein architecture.

In this chapter, we will discuss the four topics about protein, which include:

- Overview of protein Structure
- Protein Secondary Structure
- Protein Tertiary & Quaternary Structure
- Protein Denaturation & Folding

I. Overview of protein structure

Conformation of protein: The spatial arrangement of atoms in a protein. The possible conformations of a protein include any structural state that can be achieved without breaking covalent bonds.

Although protein contains so many possible conformations, there are at least one or more than one conformations predominate under biological conditions, and this kind of conformation normally is thermodynamic most stable with lowest Gibbs free energy (G). Proteins in any of their functional folded conformations are called native proteins.

A protein's conformation is stabilized largely by weak interactions

The term of protein stability is a protein's tendency to maintain its native conformation, in order to maintain its function.

The energy difference between the folded protein and unfolded protein is in the range of 20 – 60 kJ/mol, as protein keeps in unfolded conformation is entropy favorable state if

we think about protein alone; in addition, many polar functional groups exist in protein, and they can form weak interactions with solvent molecules, such as water. So, what kind of interaction keeps the native protein folded?

a) Disulfide bonds: As protein contains Cysteine, an amino acid with sulfhydryl group. Two sulfhydryl groups are easily oxidized to form disulfide covalent bond, and make a link between/within the peptide chains. This covalent bond is much stronger than the following weak interactions. Normally, proteins contain at least two Cysteines, which can form disulfide bond and force the protein to fold.

b) Hydrogen bonds: Although the polar functional groups can form hydrogen bond with water, but this kind of hydrogen bonds are normally weaker than the hydrogen bonds between water molecules per unit mass. When protein dissolves in water, it disturbs the hydrogen bonds between water molecules. In addition, protein has many nonpolar hydrocarbon chain, such as those found in valine, phenylalanine, leucine, and so on, these kind of non-polar groups will force water to form a cage around these groups, so the total process is entropy unfavorable. So, protein is likely to keep folded, in that state, almost all the non-polar groups will be kept interior of the folded conformation, and most of polar functional groups will face water molecules. Moreover, when protein folds, the interior polar groups will form hydrogen bond between each other also, and once one hydrogen bond forms, it will lead to the formation of many hydrogen bonds.

c) Ionic interaction: many of the proteins contain ionic groups, and positive and negative charged groups can form electrostatic interaction, and stabilize the folded proteins. In many of proteins, there is at least one metal cation, and the functional groups in Lysine, Histine, Serine residue will form coordination with the metal cations, and stablize the conformation of protein.

d) Hydrophobic and Lipophilic Interaction: As many of non-polar functional group exist in protein, these non-polar functional groups like to cluster together, and stay away from water or other polar solvents. So, when these non-polar functional groups stay in the interior of protein, they will contribute to the stablization of protein conformation.

e) Van der Waals interactions: Van der Waals interactions are very weak compare to other weak interactions, however, as these kind of interactions exist in any of functional groups, the numerous interactions add up will benefit the conformation of proteins.

The above five interactions or bond are the major factors to keep protein folded. However, as disulfide bond is easily to be broken down by reducing reagent, oxidizing reagent, radical, disulfide bond is not strong enough to keep the conformation of protein permanently. In addition, disulfide is weak, and can break when temperature increases, this is another reason that the energy difference between the folded and unfolded state is in the range of 20-60 KJ/mol. As we know the weak interactions are very weak, all these

weak interactions and the formation of disulfide bond contribute to the stabilization of folded protein.

Normally, proteins have their native functions when they are in folded states, as the energy difference between folded and unfolded conformation is about 20-60 kJ/mol, so this kind of energy difference can be easily overcome by increasing the temperature, this is one of the reason why enzymes or proteins can function at room temperature, not at higher temperature.

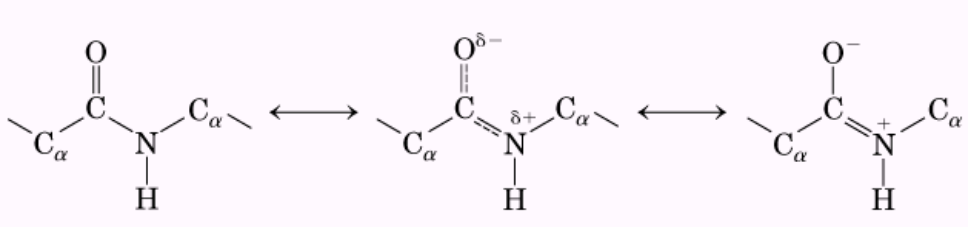
Based on the above description, we can outline the most of the structural patterns about proteins.

a) Hydrophobic residues are largely buried in the protein interior, away from water.

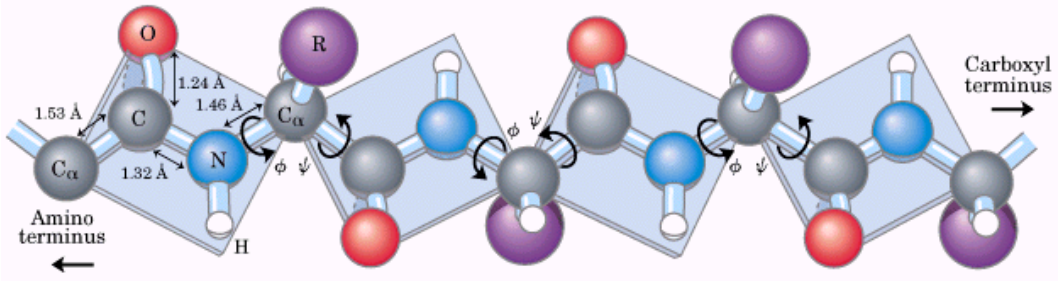
b) The number of hydrogen bonds within the protein is maximized.

The peptide bond is rigid and planar

According to the resonance theory, in peptide bond, the carbonyl oxygen has a partial negative charge and the amide nitrogen has a partial positive charge, setting up a small electric dipole. Virtually all peptide bonds in proteins occur in this trans configuration. Each peptide bond has some double-bond character due to resonance and cannot rotate. Because of this, the six atoms are on the same plane.



Three bonds separate sequential α carbons in a polypeptide chain. The N-C_α and C_α-C bonds can rotate, with bond angels designated ϕ and ψ respectively. The peptide C-N bond is not free to rotate. Other single bonds in the backbone may also be rotationally hindered, depending on the size and charge of the R groups.



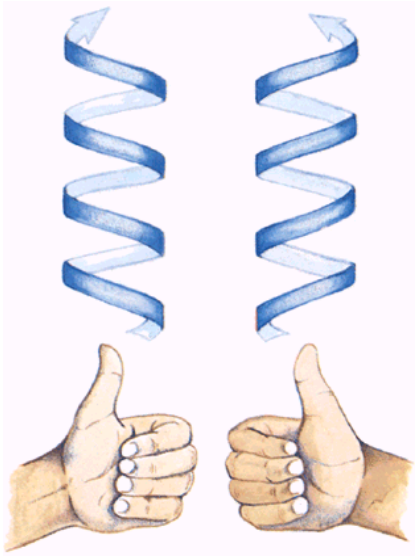
Remember, this kind of rotational prohibition also contributes to the conformation of protein.

II. Protein secondary structure

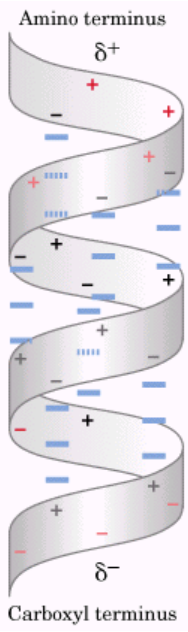
The protein secondary structure refers to particularly stable arrangements of amino acid residues giving rise to recurring structural patterns. It only refers to the local conformation of some part of the polypeptide. For the local arrangement of amino acid residues, there are many kind of conformation types, such as α-helices (helix), β-turns, γ-turns, β-sheets, β-bulges, and β-hairpins.

The α helix is a common protein secondary structure

In this structure, the polypeptide backbone is tightly wound around an imaginary axis drawn longitudinally through the middle of the helix, and the R groups of the amino acid residues protrude outward from the helical backbone. The repeating unit is a single turn of the helix, which extends about 0.54 nm along the long axis. The amino acid residues in an α helix have conformations with ϕ = -45° to –50° and ψ = -60° and each helical turn includes 3.6 amino acid residues. The helical twist of the α helix found in all proteins is right-handed. (See the following pictures)



The simple method for determining whether a helical structure is right-handed or left-handed. Make fists of your two hands with thumbs outstretched and pointing straight up.

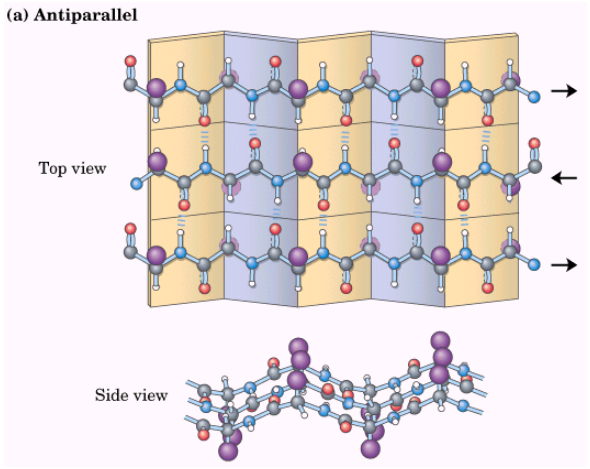


The electric dipole of a peptide bond is transmitted along an α-helical segment through the intra-chain hydrogen bonds, resulting in an overall helix dipole.

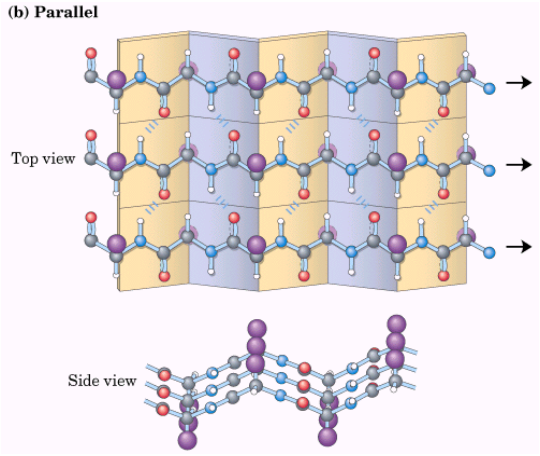
The five different kinds of constraints affect the stability of an α helix: a) the electrostatic repulsion (or attraction) between successive amino acid residues with charged R groups; b) the bulkiness of adjacent R groups; c) the interactions between amino acid side chains spaced three (or four) residues apart; d) the occurrence of Pro and Gly residues; and e) the interaction between amino acid residues at the ends of the helical segment and the electric dipole inherent to the α helix. Hence, the tendency of a given segment of a polypeptide chain to fold up as an α helix depends on the identity and sequence of amino acid residues within the segment.

The β conformation organizes polypeptide chains into sheets

β conformation is a more extended conformation about protein, in which the backbone of the polypeptide chain is extended into a zigzag rather than helical structure. The zigzag polypeptide chains can be arranged side by side to form a structure resembling a series of pleats. In this arrangement, called a β sheet, hydrogen bonds are formed between adjacent segments of polypeptide chain.



Antiparallel β sheet, in which the amino-terminal to carboxyl-terminal orientation of adjacent chains (arrows) is inverse.



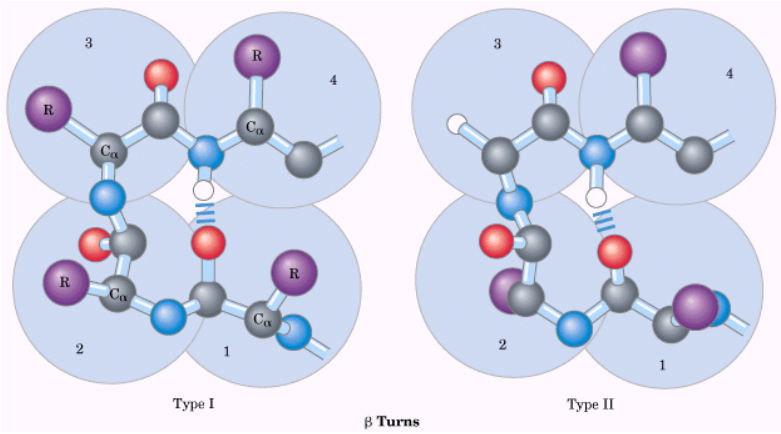
Parallel β sheet.

When two or more β sheets are layered closely together within a protein, the R groups of the amino acid residues on the touching surfaces must be relatively small. So, Gly and Ala are quite often in β sheets.

β turns are common in proteins

β turns are kind of conformations that connect the ends of two adjacent segments of an antiparallel β sheets. The structure is a 180 turn involving four amino acid residues,

with the carbonyl oxygen of the first amino acid residue forming a hydrogen bond with the amino-group hydrogen of the fourth. The peptide groups of the central two residues do not participate in any inter-residue hydrogen bonding, instead, they form hydrogen bond with water molecules. Gly and Pro residues often occur in β turns.



Two types of β turns, type I β turns occur more than twice as frequently as type II.

In γ turn segment, three amino acid residues turn with a hydrogen bond between the first and the third residues.

Protein tertiary and quaternary structures

The overall three-dimensional arrangement of all atoms in a protein is referred to as the protein's tertiary structure, which describes the three-dimensional folding of polypeptide chains. The location of bends (including β turns) in the polypeptide chain and the direction and angle of these bends are determined by the number and location of specific bend-producing residues, such as **Pro**, **Thr**, **Ser**, and **Gly**.

Quaternary structure describes the three-dimensional arrangement of more than two polypeptide chains, which might or might not be identical in terms of structure.

Name	Structure character	Sub-structure	Functions
Fibrous proteins	Long strands or sheet	Single type of secondary structure	Support, shape, external protection to vertebrates
Globular proteins	Spherical or globular shape	Several types of secondary structures	Enzymatic catalyst, regulation

Fibrous proteins are adapted for a structural function

Fibrous proteins include α-Keratin, collagen, and silk fibroin. Fibrous proteins share properties that give strength and/or flexibility to the structures. In each case, the

fundamental structural unit is a simple repeating element of secondary structure. All fibrous proteins are insoluble in water.

Secondary Structures and Properties of Fibrous Proteins

Structure	Characteristics	Examples of occurrence
α Helix, cross-linked by disulfide bonds	Tough, insoluble protective structures of varying hardness and flexibility	α-Keratin of hair, feathers, and nails
β Conformation	Soft, flexible filaments	Silk fibroin
Collagen triple helix	High tensile strength, without stretch	Collagen of tendons, bone matrix

α-Keratin

The α-keratins have evolved for strength. Found in hair, wool, nails, claws, quills, horns, hooves, and outer layer of skin. The strength of fibrous proteins is enhanced by covalent cross-links between polypeptide chains within the multi-helical "ropes" and between adjacent chains in a supramolecular assembly.

The α-keratin helix is a right-handed α helix. Two strands of α-keratin, oriented in parallel are wrapped about each other to form a supertwisted coiled coil, and this supertwisted coil is left-handed. α-Keratin is rich in hydrophobic amino acid residues, such as Ala, Val, Leu, Ile, Met, and Phe.

Collagen

Collagen also has evolved for strength. It is found in connective tissue such as tendons, cartilage, the organic matrix of bone and the cornea of the eye. Three separate polypeptides, called α chains are supertwisted about each other to form right-handed coiled coil, but the original α chain is left-handed.

Collagen is 35% Gly, 11% Ala, 21% Pro (or HyPro), and has the general repeating tripeptide unit, Gly-X-Pro (HyPro), where X can be any amino acid residue.

Silk Fibroin

Silk Fibroin produced by insects and spiders, is predominated by β conformation. Fibroin is rich in Ala and Gly residues, permitting the close packing of β sheets.

Structural diversity reflects functional diversity in globular proteins.

Globular proteins include enzymes, transport proteins, motor proteins, regulatory proteins, immunoglobulins, and proteins with many other functions.

Myoglobin provided early clues about the complexity of globular protein structure

The structure of myoglobin came from the X-ray diffraction study. Myoglobin is a small protein with molecular weight of 16,700 (153 amino acid residues), and is used to store oxygen and the facilitate oxygen diffusion in rapidly contracting muscle tissue. It can bind oxygen because it contains a iron protoporphyrin or a heme group. It contains 8 relatively straight segments of α helix interrupted by bends, some of which are β turns. Most of the hydrophobic R groups are in the interior of the myoglobin molecule, and all

but two of the polar R groups are located on the outer surface of the molecule, and all are hydrated. It is so compact that its interior has room for only four molecules of water.

The structure of protein can be determined by X-ray diffraction and NMR analysis.

Globular proteins have a variety of tertiary structures

Approximate Amounts of α Helix and β Conformation in Some Single-Chain Proteins*

	Residues (%)	
Protein (total residues)	**α Helix**	**β Conformation**
Chymotrypsin (247)	14	45
Ribonuclease (124)	26	35
Carboxypeptidase (307)	38	17
Cytochrome *c* (104)	39	0
Lysozyme (129)	40	12
Myoglobin (153)	78	0

Source: Data from Cantor, C.R. & Schimmel, P.R. (1980) *Biophysical Chemistry,* Part I: *The Conformation of Biological Macromolecules,* p. 100, W.H. Freeman and Company, New York.

*Portions of the polypeptide chains that are not accounted for by α helix or β conformation consist of bends and irregularly coiled or extended stretches. Segments of α helix and β conformation sometimes deviate slightly from their normal dimensions and geometry.

Different proteins, shown in Table 2, present the relationships between structures and functions. Small protein has low ration of volume to surface area, and less weak interactions for stabilizing the conformation of protein. In order to keep a stable conformation, these small proteins contain either disulfide bond or other functional group such as heme to provide the necessary conformation.

Analysis of many globular proteins reveals common structural patterns

Supersecondary structures, also called motifs or simply folds, are particularly stable arrangements of several elements of secondary structure and the connections between them.

Polypeptides with more than a few hundred amino acid residues often fold into two or more stable, globular units called domains. In many cases, a domain from a large protein will retain its correct three-dimensional structure even when it is separated from the remainder of the polypeptide chain. A protein with multiple domains may appear to have a distinct globular lobe for each domain. Different domains often have distinct functions.

Folding of polypeptides is subject to an array of physical and chemical constraints. A sampling of the prominent folding rules that have emerged provides an opportunity to introduce some simple motifs.

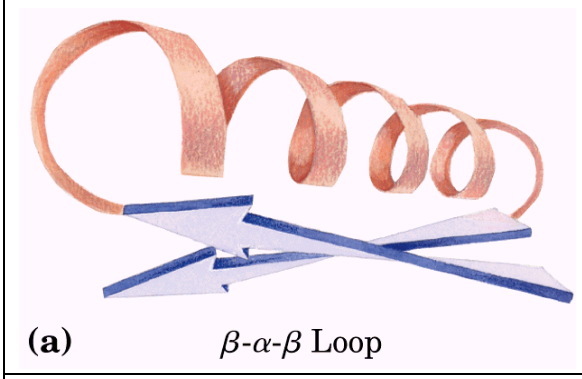
a) hydrophobic interactions make a large contribution to the stability of protein structures. Burial of hydrophobic amino acid R groups so as to exclude water requires at least two layers of secondary structure. Two simple motifs, the β-α-β loop and the α-α corner, create two layers.

b) Where they occur together in proteins, α helices and β sheets generally are found in different structural layers. This is because the backbone of a polypeptide segment in the β conformation cannot readily hydrogen bond to an α helix aligned with it.

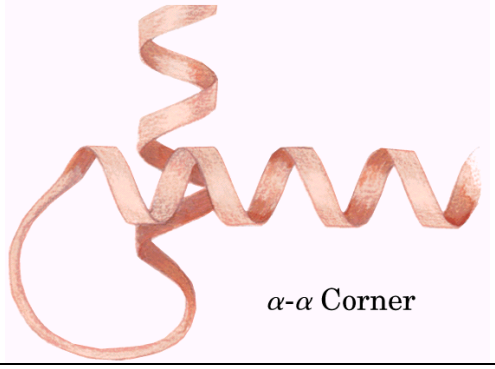
c) Polypeptide segments adjacent to each other in the primary sequence are usually stacked adjacent to each other in the folded structure. Although distant segments of a polypeptide may come together in the tertiary structure, this is not the norm.

d) Connections between elements of secondary structure cannot cross or form knots.

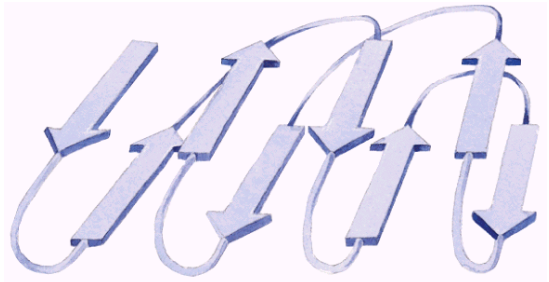
e) The β conformation is most stable when the individual segments are twisted slightly in a right-handed sense. This influences both the arrangement of β sheets relative to one another and the path of the polypeptide connection between them. The twisting of β sheets also leads to a characteristic twisting of the structure formed when many segment are put together.



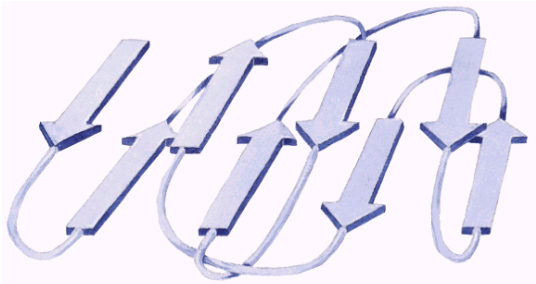
(a) β-α-β Loop



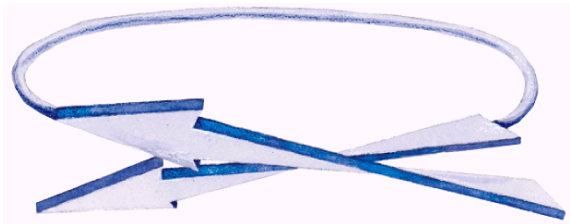
α-α Corner



(b) Typical connections in an all-β motif

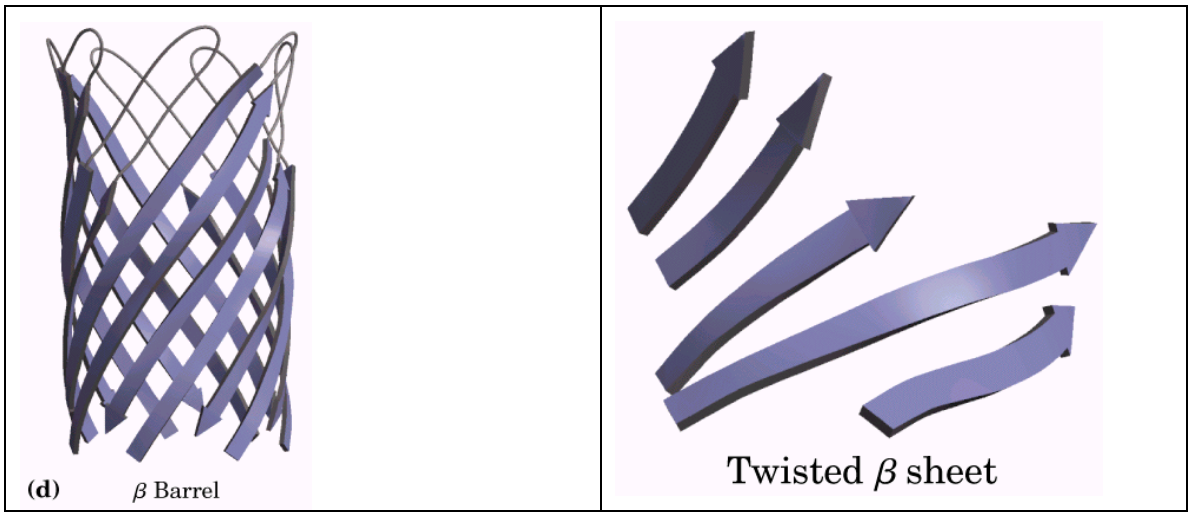


Crossover connection (not observed)



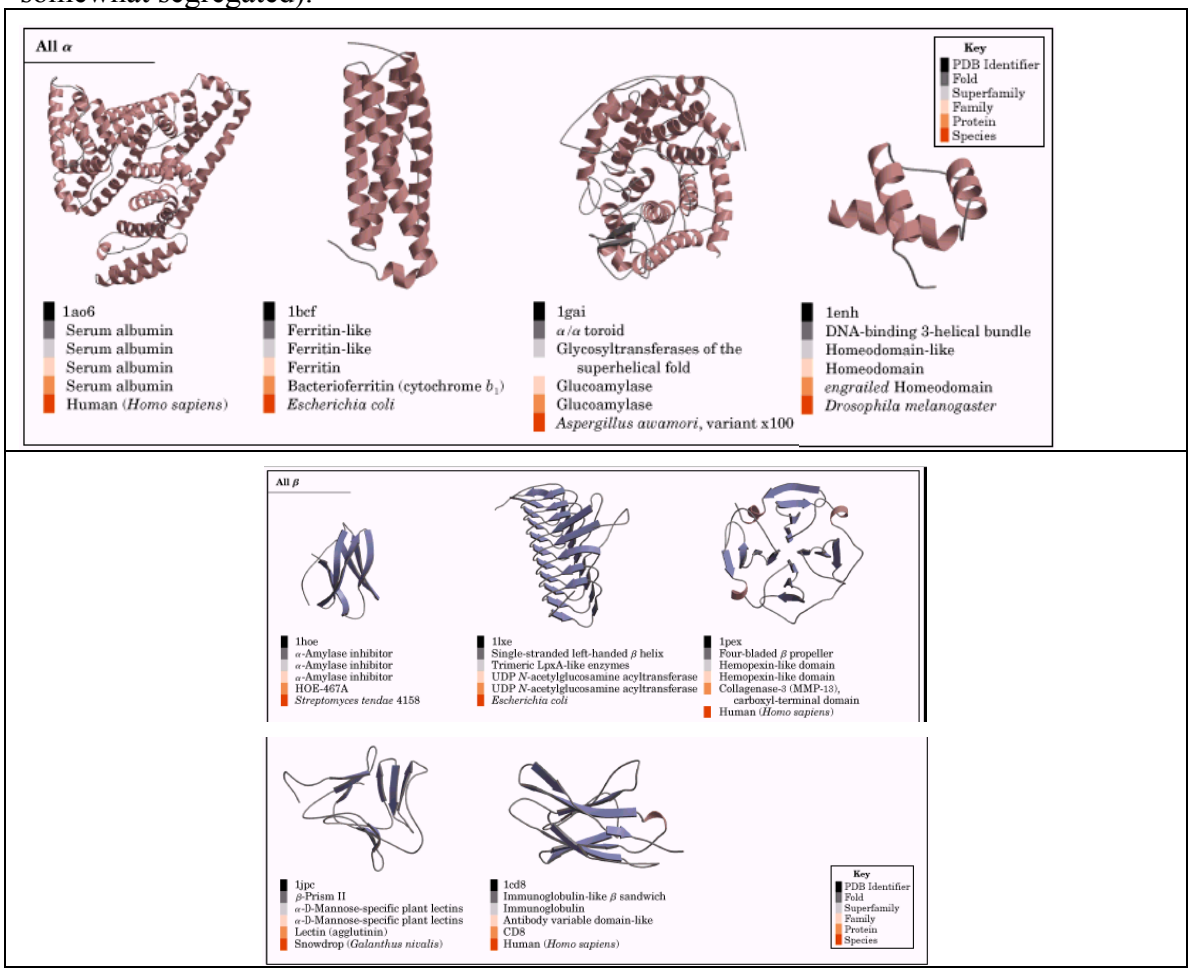
(c) Right-handed connection between β strands

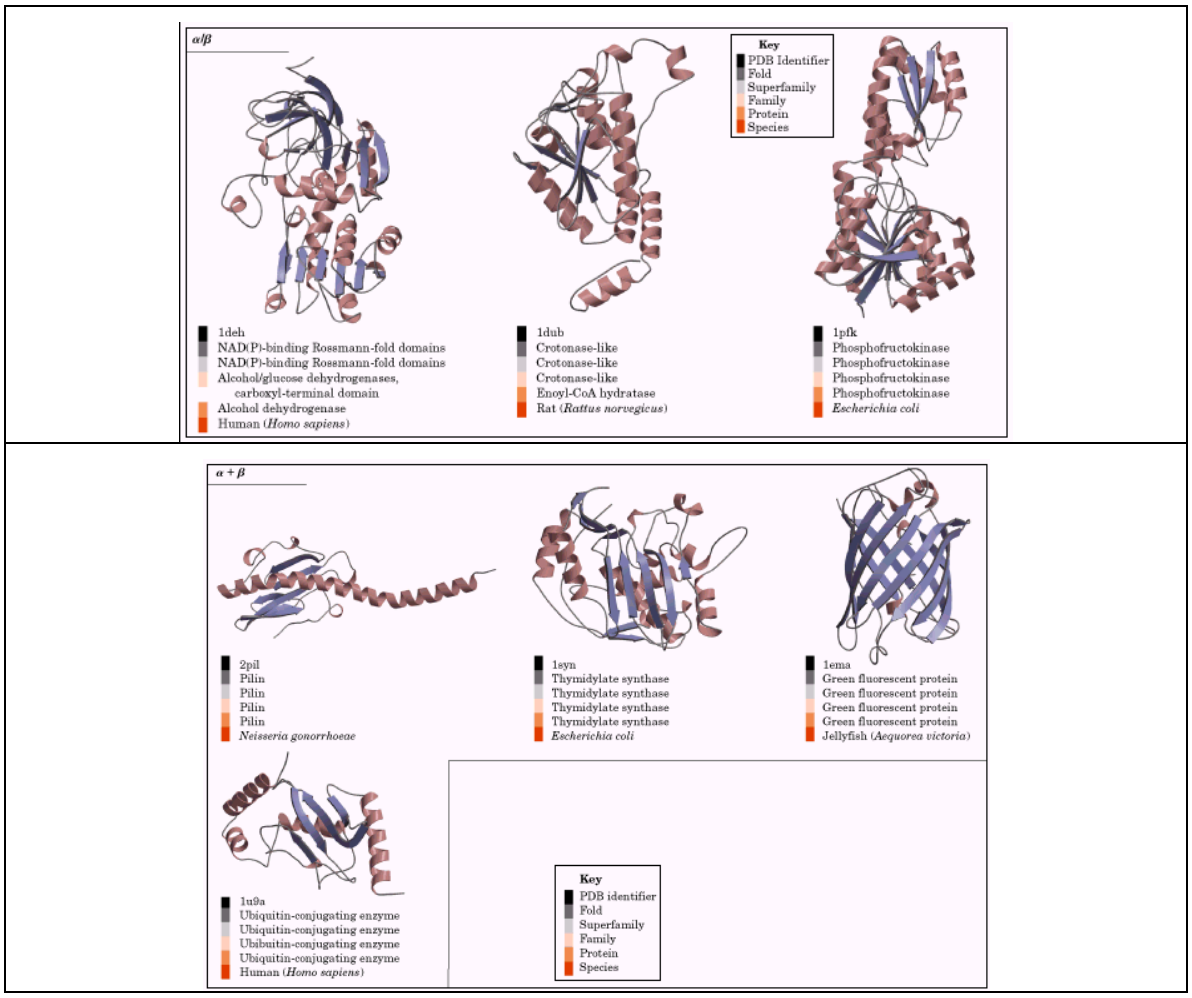
Left-handed connection between β strands (very rare)



Protein motifs are the basis for protein structural classification

Protein structures are divided into four classes: all α, all β, α/β (in which the α and β segments are interspersed or alternate), and α + β (in which the α and β regions are somewhat segregated).





Proteins with significant primary sequence similarity, and/or with demonstrably similar structure and function, are said to be in the same protein family. A strong evolutionary relationship is usually evident within a protein family.

Two or more families with little primary sequence similarity sometimes make use of the same major structural motif and have functional similarities, these families are grouped as superfamilies. An evolutionary relationship between the families in a superfamily is considered probable.

Protein Quaternary structures range from simple dimers to large complexes

A multisubunit protein is also referred to as a multimer. Multimeric proteins can have from two to hundreds of subunits. A multimer with just a few subunits is often called an oligomer. If a multimer is composed of a number of nonidentical subunits, the overall structure of the protein can be asymmetric and quite complicated. However, most multimers have identical subunits or repeating groups of nonidentical subunits, usually in symmetric arrangements. The repeating structural unit in such a multimeric protein, whether it is a single subunit or a group of subunits, is called a protomer.

Oligomers can have either rotational symmetry or helical symmetry, that means individual subunits can be superimposed on others by rotation about one or more rotational axes, or by a helical rotation. There are several forms of rotational symmetry.

There are limits to the size of proteins

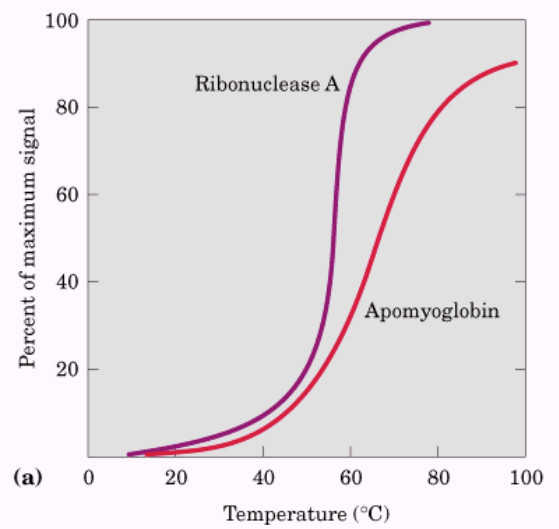
The proteins must be big enough to carry out their functions, however, proteins cannot be limitless. In fact, the size of protein is determined by two factors. 1) the available of genetic codes. As we know the amino acid sequence of protein is determined by genetic code, every three continuous nucleotides units is corresponding to a amino acid, as the limit of gene, therefore, the size of protein is limited too. So, the large proteins almost contain many copies of subunits, say more than 100,000 molecular weight of protein; and under this circumstance, the genetic source is efficiently used. 2) error frequency during protein biosynthesis. It is quite clear that the big the protein is, the high chance to involve the error during the biosynthesis of protein.

Protein denaturation and folding

Loss of protein structure results in loss of function

The function of protein is determined by its three-dimensional structure. A loss of three-dimensional structure sufficient to cause loss of function is called denaturation. The denatured state does not necessarily equate with complete unfolding of the protein and randomization of conformation.

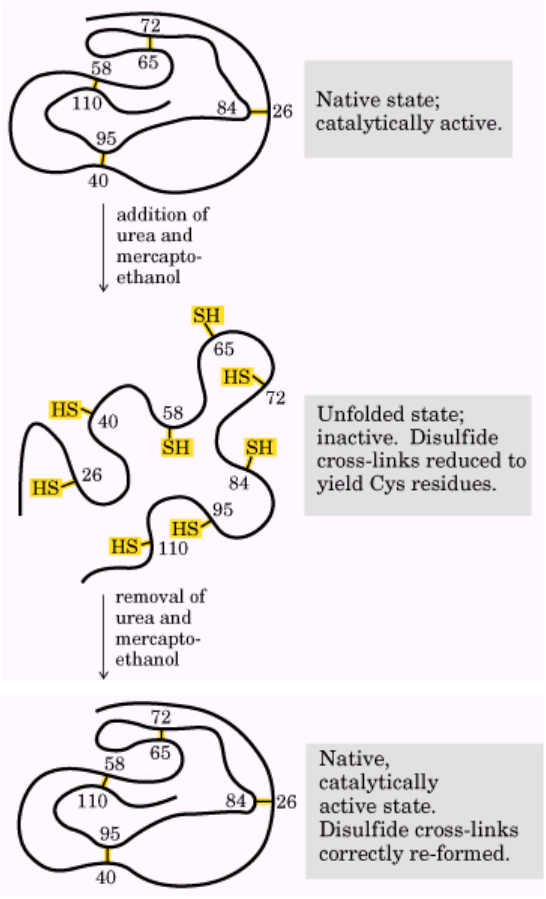
Most of proteins can be denatured by heat, which affects the weak interactions in a protein (primarily hydrogen bonds) in a complex manner. If the temperature is increased slowly, a protein's conformation generally remains intact until an abrupt loss of structure (and function also) occurs over a narrow temperature range. The abruptness of the change suggests that unfolding is a cooperative process: loss of structure in one part of the protein destabilizes other parts.



Proteins can be denatured not only by heat but also by 1) extremes of pH, 2) by certain miscible organic solvents such as alcohol or acetone, 3) by certain solutes such as urea and guanidine hydrochloride, or 4) by detergents.

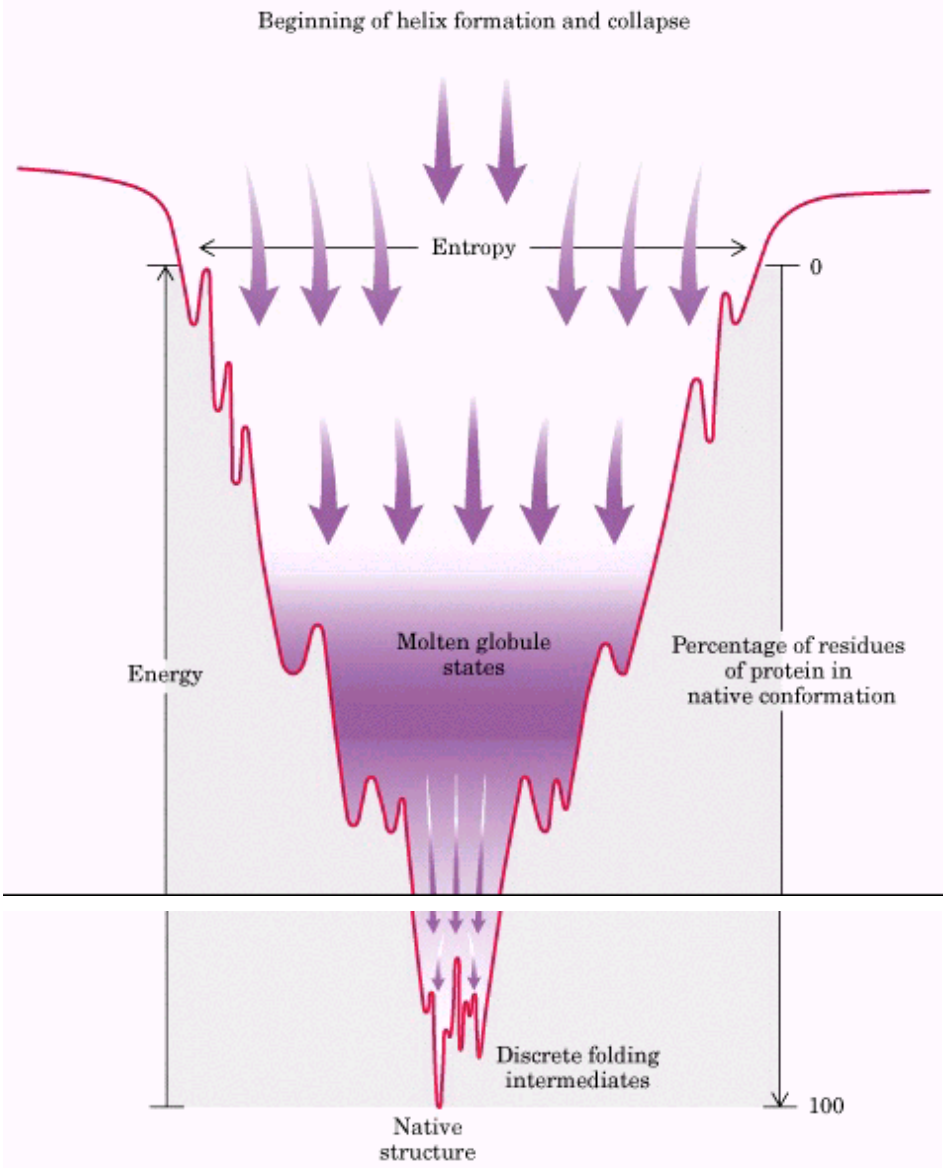
Amino acid sequence determines tertiary structure

Certain globular proteins denatured by heat, extremes of pH, or denaturing reagents will regain their native structure and their biological activity if returned to conditions in which the native conformation is stable. This process is called renaturation. See the following sample.



Polypeptides fold rapidly by a stepwise process

The thermodynamics of protein folding depicted as a free-energy funnel. At the top, the number of conformations, and hence the conformational entropy is large. Only a small fraction of the intramolecular interactions that will exist in the native conformation are present. As folding progresses, the thermodynamic path down the funnel reduces the number of states present (decreases entropy), increases the free energy. Depressions on the sides of the funnel represent semistable folding intermediates, which may, in some cases, slow the folding process.



Some proteins undergo assisted folding

Molecular chaperones are proteins that interact with partially folded or improperly folded polypeptides, facilitating correct folding pathways or providing microenvironments in which folding can occur. Two classes of molecular chaperones have been well-studied. The first class, a family of proteins called Hsp70 (heat shock proteins of Mr 70,000, or Hsp70). Hsp70 binds to regions of unfolded polypeptides that are rich in hydrophobic residues, preventing inappropriate aggregation.

The second class of chaperones are called chaperonins. These are elaborate protein complexes required for the folding of a number of cellular proteins that do not fold spontaneously.

Finally the folding pathways of a number of proteins require two enzymes that catalyze isomerization reactions. Protein disulfide isomerase (PDI) is a widely distributed enzyme that catalyzes the interchange or shuffling of disulfide bonds until the bonds of the native conformation are formed. Among its functions, PDI catalyzes the elimination of folding intermediates with inappropriate disulfide cross-links. Peptide prolyl cis-trans isomerase (PPI) catalyzes the interconversion of the cis and trans isomers of proline peptide bonds, which can be a slow step in the folding of proteins that contain some bonds in the cis conformation.