UNIT-I

THE THREE-DIMENSIONAL STRUCTURE OF PROTEINS

The covalent backbone of proteins is made up of hundreds of individual bonds. If free rotation were possible around even a fraction of these bonds, proteins could assume an almost infinite number of three dimensional structures. Each protein has a specific chemical or structural function; however, strongly suggesting that each protein has a unique three-dimensional structure The simple fact that proteins can be crystallized provides strong evidence that this is the case. The ordered arrays of molecules in a crystal can generally form only if the molecular units making up the crystal are identical. The enzyme urease (M_r 483,000) was among the first proteins crystallized, by James Sumner in 1926. This accomplishment demonstrated dramatically that even very large proteins are discrete chemical entities with unique structures, and it revolutionized thinking about proteins.

1. OVERVIEW OF PROTEIN STRUCTURE

The spatial arrangement of atoms in a protein is called a conformation. The term conformation refers to a structural state that can, without breaking any covalent bonds, interconvert with other structural states. A change in conformation could occur, for example, by rotation about single bonds. Of the innumerable conformations that are theoretically possible in a protein containing hundreds of single bonds, one generally predominates. This is usually the conformation that is thermodynamically the most stable, having the lowest Gibbs' free energy (G). Proteins in their functional conformation are called native proteins.

Four Levels of Architecture in Proteins

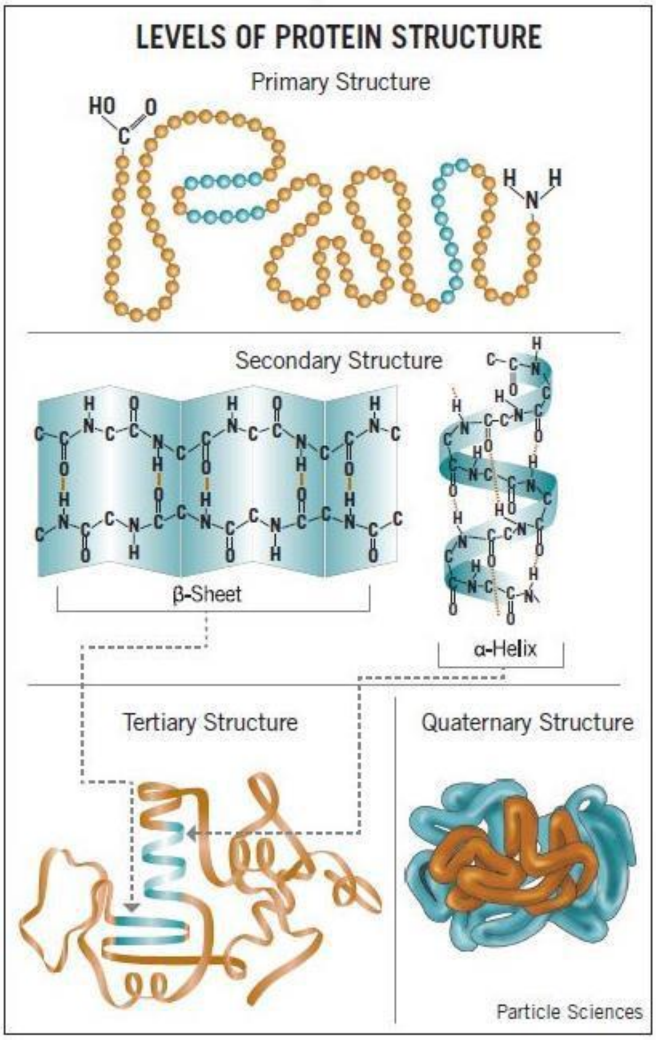


Figure 1 Levels of structure in proteins

Conceptually, protein structure can be considered at four levels (Fig. 1). **Primary structure** includes all the covalent bonds between amino acids and is normally defined by the sequence of peptide-bonded amino acids and locations of disulfide bonds. The relative spatial arrangement of the linked amino acids is unspecified. Polypeptide chains are not free to take up any three-dimensional structure at random. Steric constraints and many weak interactions stipulate that some arrangements will be more stable than others.

Secondary structure refers to regular, recurring arrangements in space of adjacent amino acid residues in a polypeptide chain. There are a few common types of secondary structure, the most prominent being the a helix and the β conformation.

Tertiary structure refers to the spatial relationship among all amino acids in a polypeptide; it is the complete three-dimensional structure of the polypeptide. The boundary between secondary and tertiary structure is not always clear. Several different types of secondary structure are often found within the three-dimensional structure of a large protein. Proteins with several polypeptide chains have one more level of structure: **quaternary structure**, which refers to the spatial relationship of the polypeptides, or subunits, within the protein.

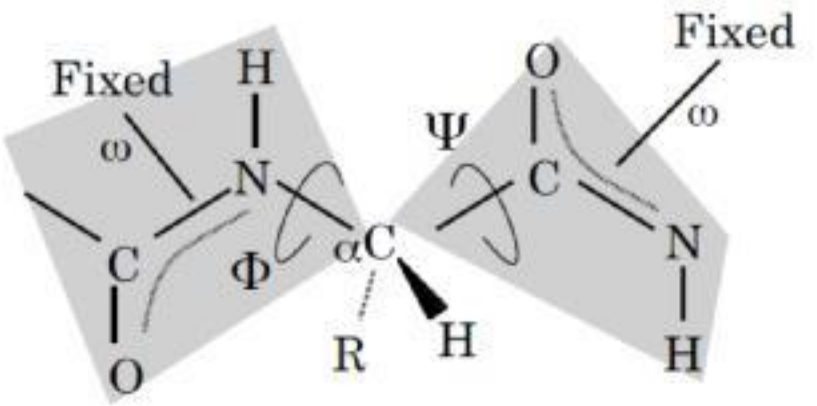
1.1.Protein Secondary Structure

Several types of secondary structure are particularly stable and occur widely in proteins. The most prominent are the α helix and β conformations. Using fundamental chemical principles and a few experimental observations, Linus Pauling and Robert Corey predicted the existence of these secondary structures in 1951, several years before the first complete protein structure was elucidated.

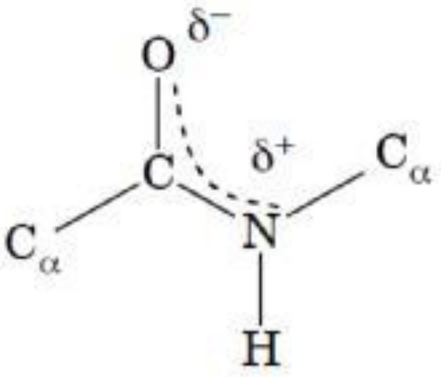
In considering secondary structure, it is useful to classify proteins into two major groups: fibrous proteins, having polypeptide chains arranged in long strands or sheets, and globular proteins, with polypeptide chains folded into a spherical or globular shape. Fibrous proteins play important structural roles in the anatomy and physiology of vertebrates, providing external protection, support, shape, and form. They may constitute one-half or more of the total body protein in larger animals. Most enzymes and peptide hormones are globular proteins. Globular proteins tend to be structurally complex, often containing several types of secondary structure; fibrous proteins usually consist largely of a single type of secondary structure. Because of this structural simplicity, certain fibrous proteins played a key role in the development of the modern understanding of protein structure and provide particularly clear examples of the relationship between structure and function; they are considered in some detail after the general discussion of secondary structure.

The Peptide Bond Is Rigid and Planar

In the peptide bond, the π-electrons from the carbonyl are delocalized between the oxygen and the nitrogen. This means that the peptide bond has ~40% double bond character. This partial double bond character is evident in the shortened bond length of the C–N bond. The length of a normal C–N single bond is 1.45 Å and a C=N double bond is 1.25 Å, while the peptide C–N bond length is 1.33 Å.



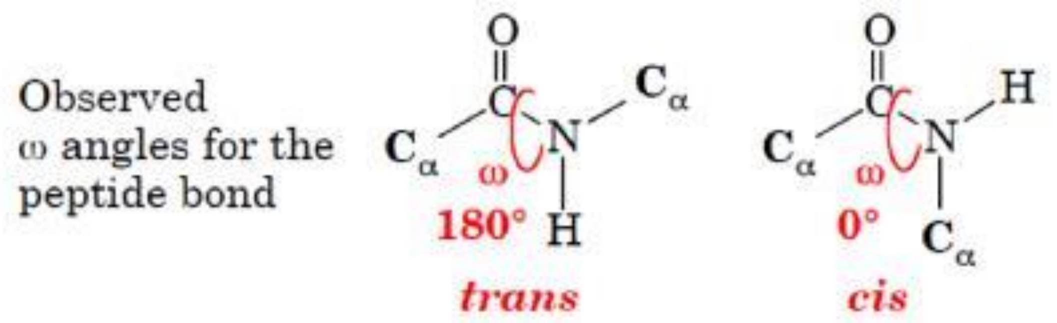
Because of its partial double bond character, rotation around the N–C bond is severely restricted. The peptide bond allows rotation about the bonds from the α- carbon, but not the amide C–N bond. Only the Φ and Ψ torsion angles (see below) can vary reasonably freely. In addition, the six atoms in the peptide bond (the two α-carbons, the amide O, and the amide N and H) are coplanar. Finally, the peptide bond has a dipole, with the O having a partial negative charge, and the Namide having a partial positive charge.



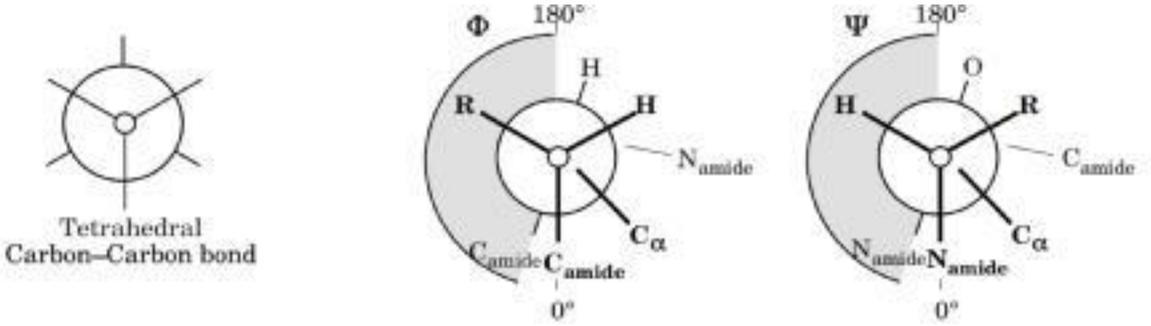
This allows the peptide bond to participate in electrostatic interactions, and contributes to the hydrogen bond strength between the backbone carbonyl and the Namide proton.

Peptide bond and protein structure

The peptide bond contains three sets of torsion angles (also known as dihedral angles). The least variable of these torsion angles is the ω angle, which is the dihedral angle around the amide bond. As discussed above, this angle is fixed by the requirement for orbital overlap between the carbonyl double bond and the Namide lone pair orbital. Steric considerations strongly favor the trans configuration (i.e. an ω angle of 180°), because of steric hindrance between the alpha carbons of adjacent amino acid residues. This means that nearly all peptide bonds in a protein will have an ω angle of 180°.



In considering peptide structures, it is usually much more important to look at the backbone angles that can vary more widely. These angles are the Φ (= phi, Cα–Namide) and Ψ (= psi, Cα–Camide) angles. By definition, the fully extended conformation corresponds to 180° for both Φ and Ψ. (Note that 180° = –180°). Numeric values of angles increase in the clockwise direction when looking away from the α-carbon



By definition, Φ = 0° when the Camide-Namide and Camide-Cα bonds are in the same plane, and Ψ = 0° when the Namide-Camide and Namide-Cα bonds are in the same plane. The (+) direction is clockwise while looking away from the Cα. The torsion angles that the atoms of the peptide bond can assume are limited by steric constraints. Some Φ / Ψ pairs will result in atoms being closer than allowed by the van der Waals radii of the atoms, and are therefore sterically forbidden (for example: 0°:0°, 180°:0°, and 0°:180° are forbidden because of backbone atom clashes).

For tetrahedral carbons, the substituents are typically found in staggered conformations (see figure, above). Peptide bonds are more complicated, because while the α-carbon is tetrahedral, the two other backbone atom types are not. However, the same principle applies: the preferred conformations for peptide bond atoms have the substituent atoms at maximal distances from one another.

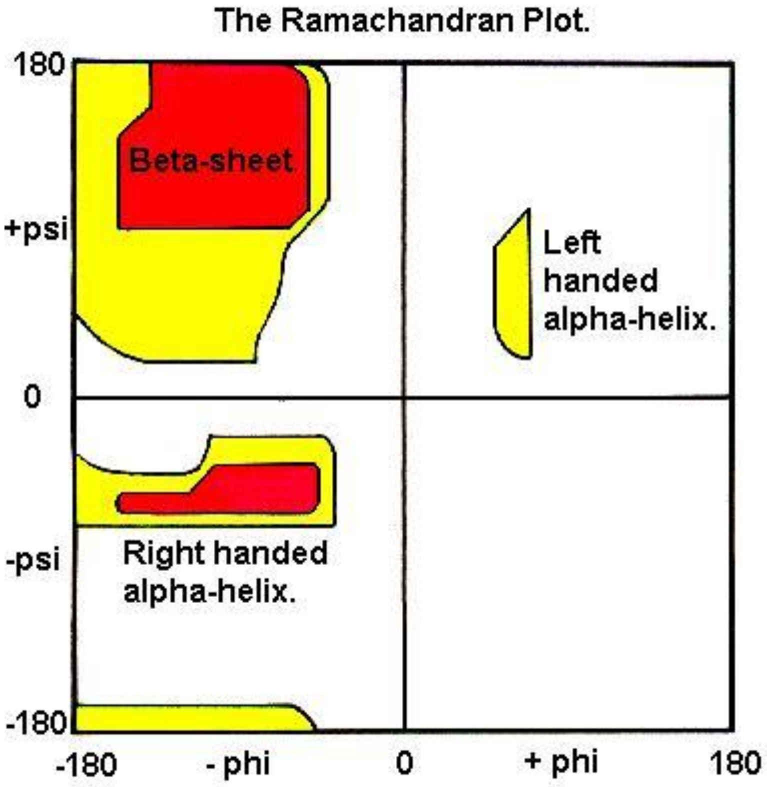
A Ψ angle of 180° results in an alignment of the Namide with the carbonyl oxygen from the same residue. This is allowed, although not especially favored. A Ψ angle of 0° places the

Namide from one residue very close to the Namide from the previous residue; this results in a steric clash (as well as an unfavorable electrostatic interaction, because both Namide have partial positive charges). The residue side-chains also impose steric constraints. Glycine, because of its small side chain, has a much large ranger of possible Φ / Ψ pairs than any other residue. Proline has a very limited range of Φ angles because its side-chain is covalently bonded to its Namide. Most other residues are limited to relatively few Φ / Ψ pairs (although more than proline). This is especially true for the β-branched residues threonine, valine, and isoleucine, which are the most restricted, because these residues have more steric bulk due to the presence of two groups attached their β- carbon. Allowed values for Φ and Ψ are graphically revealed when Ψ is plotted versus Φ in a **Ramachandran plot**, introduced by G. N. Ramachandran .

The Ramachandran Plot

In a polypeptide the main chain N-Calpha and Calpha-C bonds relatively are free to rotate. These rotations are represented by the torsion angles phi and psi, respectively.

G N Ramachandran used computer models of small polypeptides to systematically vary phi and psi with the objective of finding stable conformations. For each conformation, the structure was examined for close contacts between atoms. Atoms were treated as hard spheres with dimensions corresponding to their van der Waals radii. Therefore, phi and psi angles which cause spheres to collide correspond to sterically disallowed conformations of the polypeptide backbone.



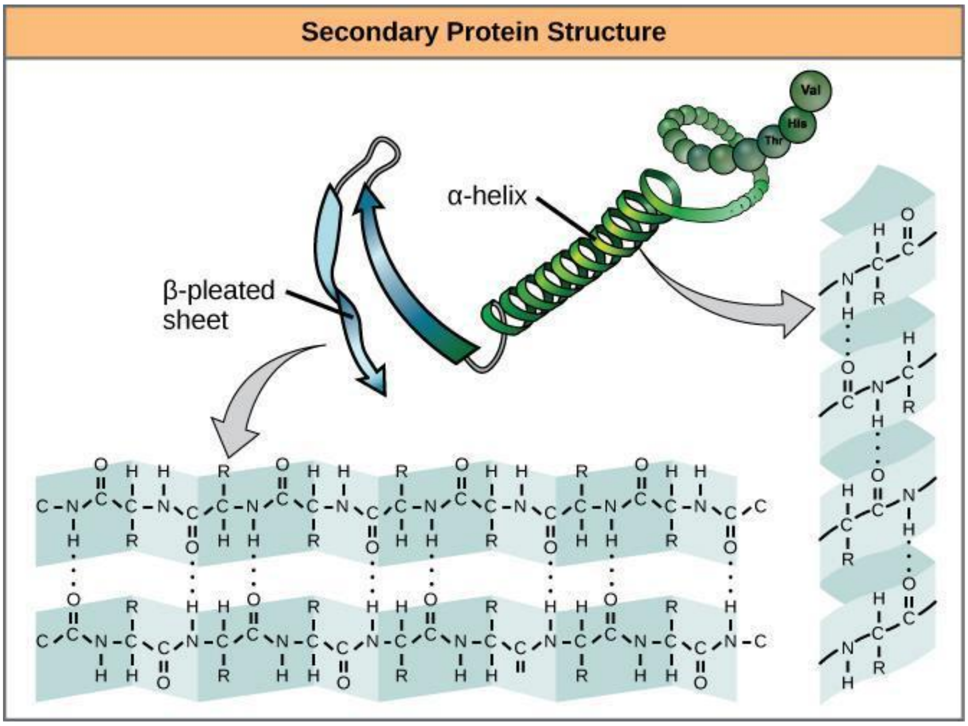
In the diagram above the white areas correspond to conformations where atoms in the polypeptide come closer than the sum of their van der Waals radi. These regions are sterically disallowed for all amino acids except glycine which is unique in that it lacks a side chain. The red regions correspond to conformations where there are no steric clashes, ie these are the allowed regions namely the alpha-helical and beta-sheet conformations. The yellow areas show the allowed regions if slightly shorter van der Waals radi are used in the calculation, ie the atoms are allowed to come a little closer together. This brings out an additional region which corresponds to the left-handed alpha-helix.

L-amino acids cannot form extended regions of left-handed helix but occassionally individual residues adopt this conformation. These residues are usually glycine but can also be asparagine or aspartate where the side chain forms a hydrogen bond with the main chain and therefore stabilises this otherwise unfavourable conformation. The 3(10) helix occurs close to the upper right of the alpha-helical region and is on the edge of allowed region indicating lower stability.

Disallowed regions generally involve steric hindrance between the side chain C-beta methylene group and main chain atoms. Glycine has no side chain and therefore can adopt phi and psi angles in all four quadrants of the Ramachandran plot. Hence it frequently occurs in turn regions of proteins where any other residue would be sterically hindered.

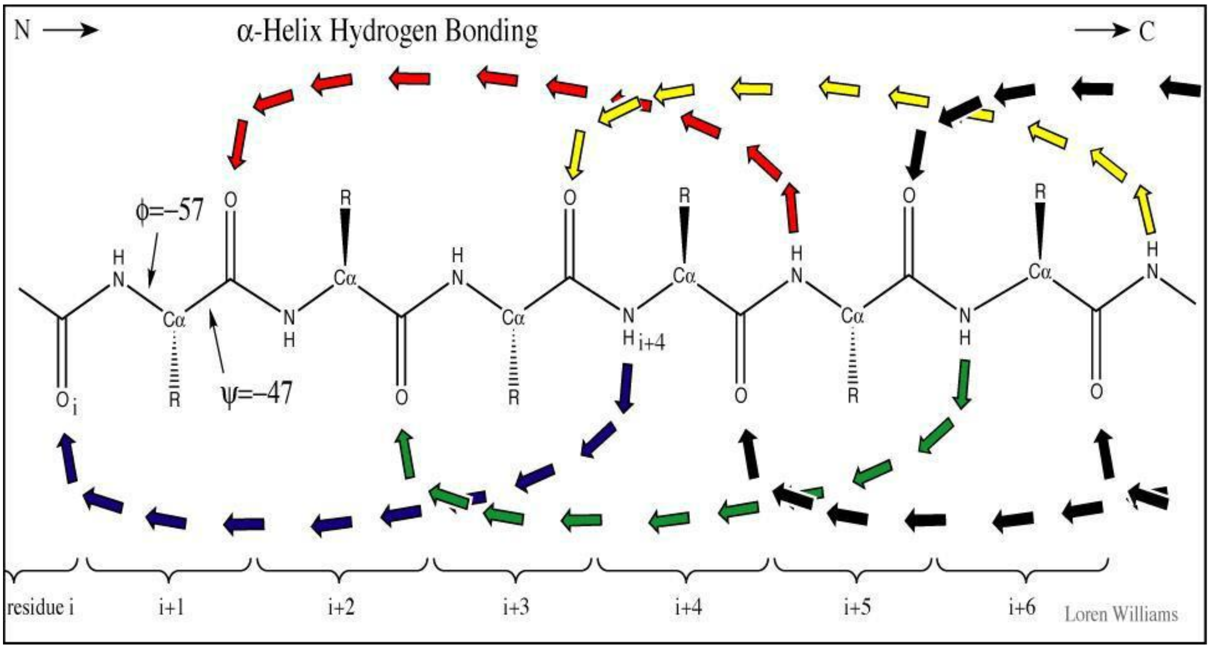
Secondary structure

The term secondary structure refers to the local conformation of some part of a polypeptide. The discussion of secondary structure most usefully focuses on common regular folding patterns of the polypeptide backbone. A few types of secondary structure are particularly stable and occur widely in proteins. The most prominent are the α-helix and β-sheet. Using fundamental chemical principles and a few experimental observations, Pauling and Corey predicted the existence of these secondary structures in 1951, several years before the first complete protein structure was elucidated.



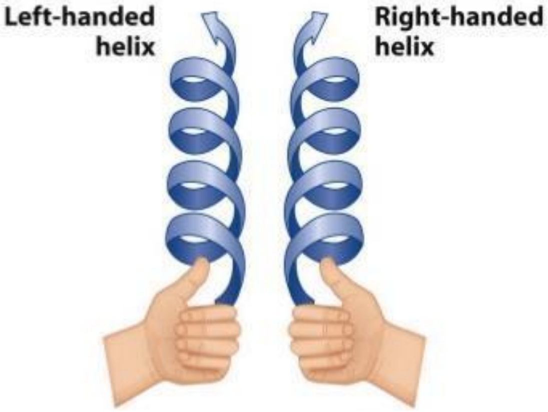
Alpha helix (α-helix)

The **alpha helix** (**α-helix**) is a common secondary structure of proteins and is a right hand-coiled or spiral conformation (helix) in which every backbone N-H group donates a hydrogen bond to the backbone C=O group of the amino acid four residues earlier ($i+4 \rightarrow i$ hydrogen bonding). This secondary structure is also sometimes called a classic **Pauling–Corey–Branson alpha helix** (see below). The name **3.6_{13}-helix** is also used for this type of helix, denoting the number of residues per helical turn, and 13 atoms being involved in the ring formed by the hydrogen bond. Among types of local structure in proteins, the α-helix is the most regular and the most predictable from sequence, as well as the most prevalent.

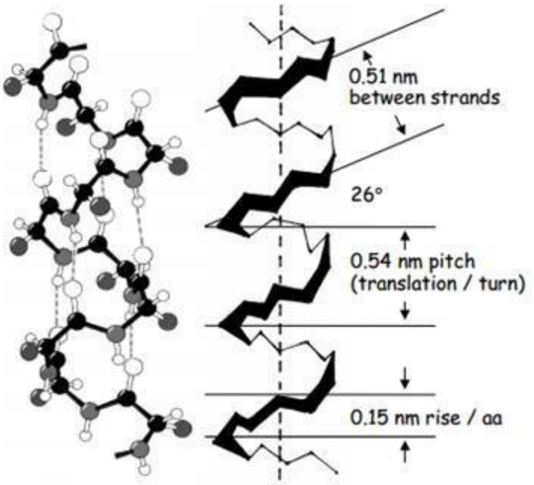


PROPERTIES

The amino acids in an α helix are arranged in a right-handed helical structure where each amino acid residue corresponds to a 100° turn in the helix (i.e., the helix has 3.6 residues per turn), and a translation of 1.5 Å (0.15 nm) along the helical axis.



Short pieces of left-handed helix sometimes occur with a large content of achiral glycine amino acids, but are unfavorable for the other normal, biological L-amino acids.



The pitch of the alpha-helix (the vertical distance between consecutive turns of the helix) is 5.4 Å (0.54 nm), which is the product of 1.5 and 3.6. What is most important is that the N-H group of an amino acid forms a hydrogen bond with the C=O group of the amino acid *four* residues earlier; this repeated $i + 4 \rightarrow i$ hydrogen bonding is the most prominent characteristic of an α-helix.

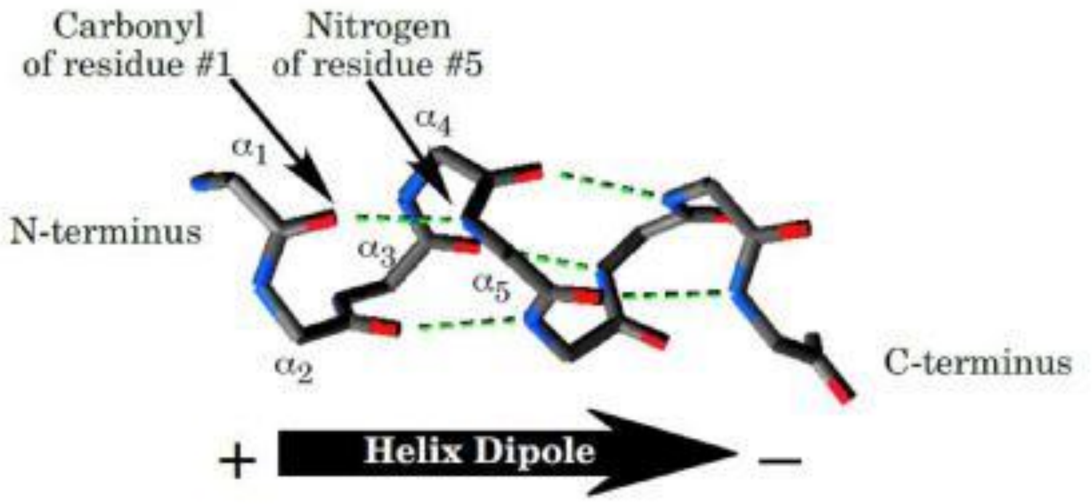
Similar structures include the 3_{10} helix ($i + 3 \rightarrow i$ hydrogen bonding) and the π-helix ($i + 5 \rightarrow i$ hydrogen bonding). The α helix can be described as a 3.6_{13} helix, since the i + 4 spacing adds 3 more atoms to the H-bonded loop compared to the tighter 3_{10} helix, and on average, 3.6 amino acids are involved in one ring of α helix. The subscripts refer to the number of atoms (including the hydrogen) in the closed loop formed by the hydrogen bond.

Residues in α-helices typically adopt backbone (φ, ψ) dihedral angles around (-60°, -45°), as shown in the image at right. In more general terms, they adopt dihedral angles such that the ψ dihedral angle of one residue and the φ dihedral angle of the *next* residue sum to roughly -105°. As a consequence, α-helical dihedral angles, in general, fall on a diagonal stripe on the Ramachandran diagram (of slope -1), ranging from (-90°, -15°) to (-35°, -70°). For comparison, the sum of the dihedral angles for a 3_{10} helix is roughly -75°, whereas that for the π-helix is roughly -130°.

Structural features of the three major forms of protein helices

Geometry attribute	**α-helix**	**3_{10} helix**	**π-helix**
Residues per turn	3.6	3.0	4.4
Translation per residue	1.5 Å (0.15 nm)	2.0 Å (0.20 nm)	1.1 Å (0.11 nm)
Radius of helix	2.3 Å (0.23 nm)	1.9 Å (0.19 nm)	2.8 Å (0.28 nm)
Pitch	5.4 Å (0.54 nm)	6.0 Å (0.60 nm)	4.8 Å (0.48 nm)

An α-helix has a dipole, with the partial positive charge toward N-terminus. This is true because all of the partial charges of the peptide bonds are in alignment.



The backbone of the helix is ~6 Å in diameter (ignoring side chains).

Two-dimensional representations of α-helices

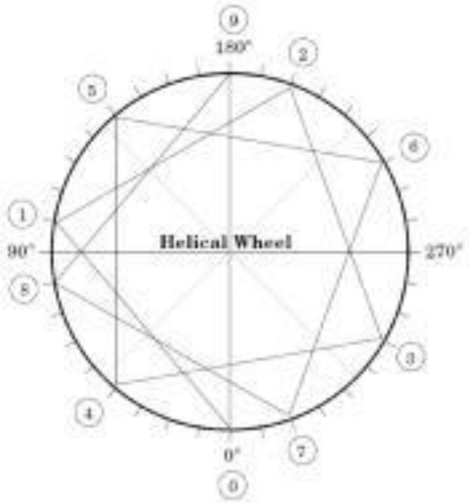
Drawing a three-dimensional helix on paper is difficult. Two types of two dimensional representations (helical wheel and helical net diagrams) are commonly used to simplify the analysis of helical segments of proteins. The two-dimensional representations are somewhat stylized, but show the major features more clearly than attempting to draw a three-dimensional structure accurately in two dimensions.

The first type of representation is a Helical Wheel diagram. In this diagram, the representation involves looking down the helix axis, and plotting the rotational angle around the helix for each residue. This representation is conceptually easily grasped, but tends to obscure the distance along the helix; residues 0 and 18 are exactly aligned on this diagram, but are actually separated in space by 27 Å.

Helical Wheel

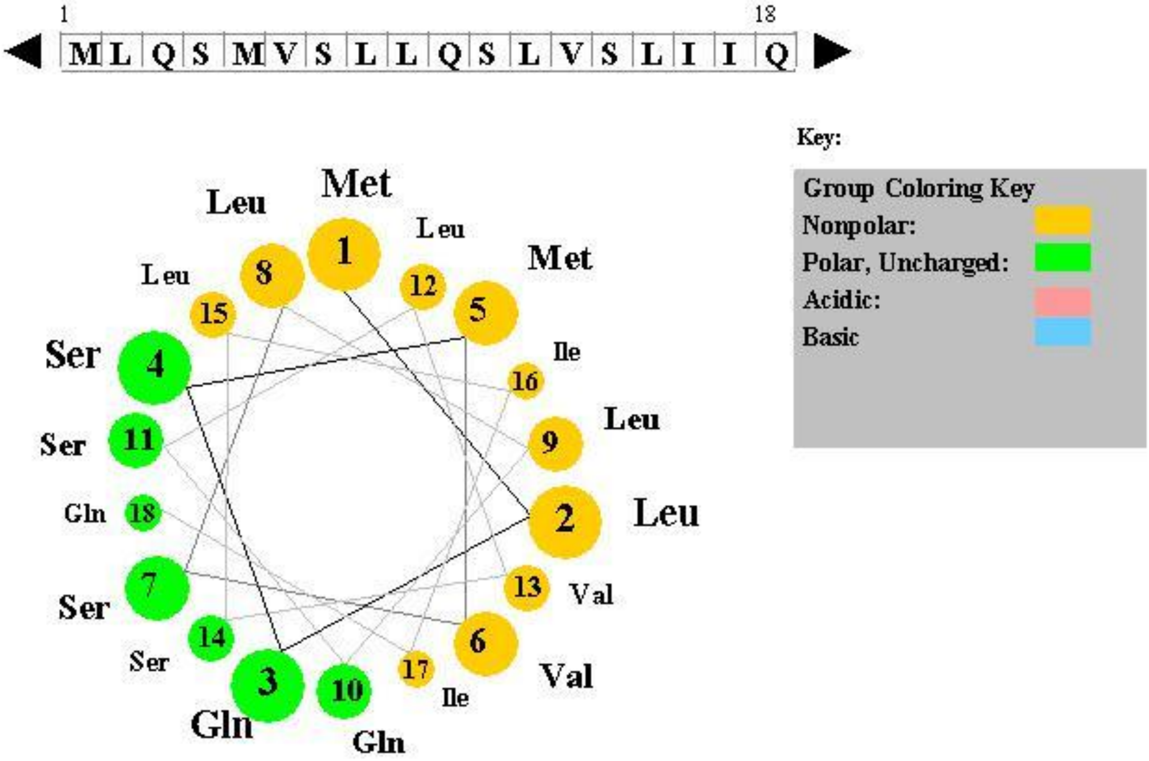
Residue #0 = 0° (by definition)
#1 = 100°
#2 = 200°
#3 = 300°
#4 = 400° = 40°
#5 = 140°
#6 = 240°
#7 = 340°
#8 = 440° = 80°
9# = 900° (from first) = 180°
These angles can be plotted on a circle.

Doing so results in a representation that corresponds to the view looking down the long axis of the helix. (**Note that the rotation is clockwise as the residue number increases**.)

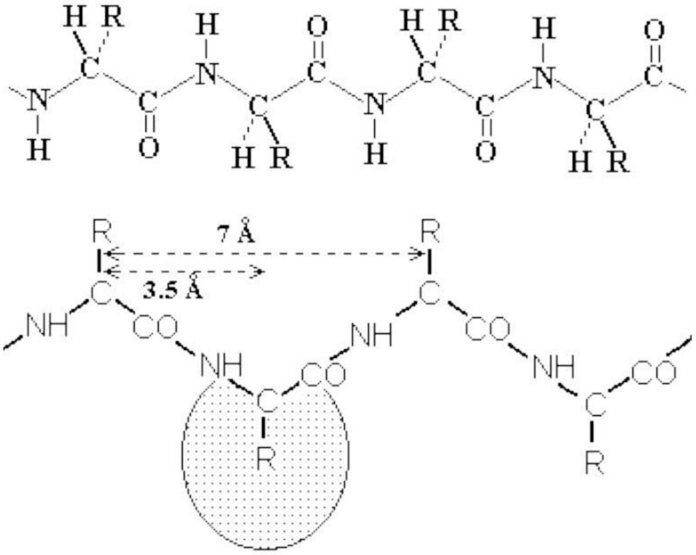


Note that residues 0, 3, 4, 7, and 8 are all located on one face of the helix

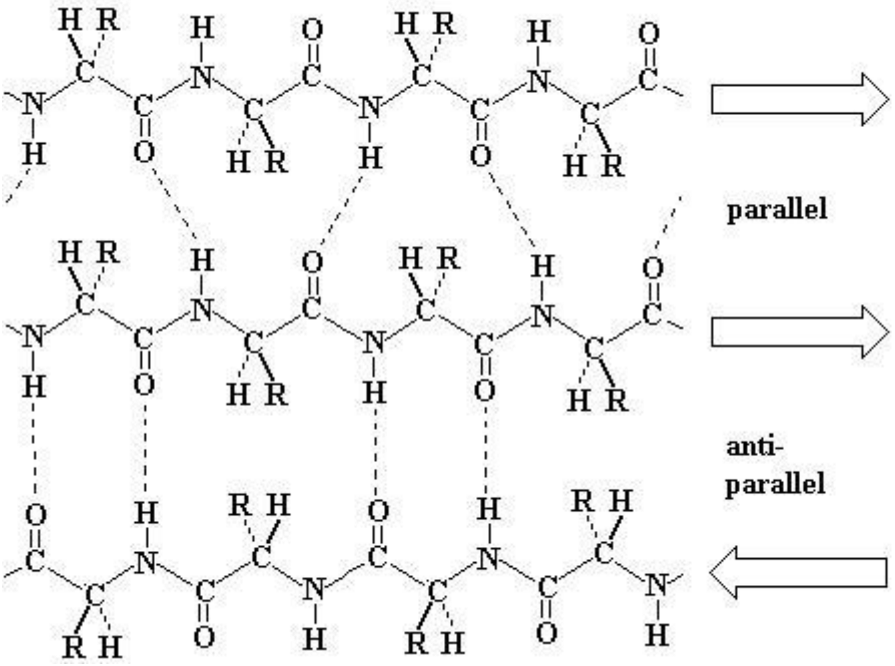
A helix that has its axis along the border of this region would be expected to have a corresponding, amphipathic, distribution of polar and non-polar residues. (Amphipathic, meaning "hating both" refers to the presence of both polar and non-polar groups in the helix.)



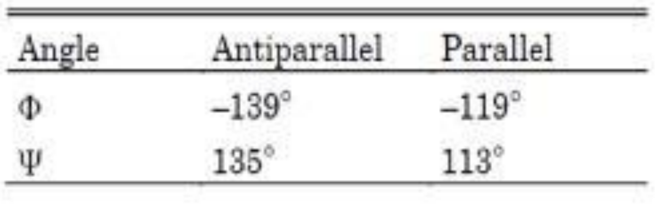
The βß Conformation Organizes Polypeptide Chains into Sheets



Pauling and Corey predicted a second type of repetitive structure, the β conformation. an **extended** state for which angles phi = -135^{o} and psi = +135^{o}; the polypeptide chain **alternates** in direction, resulting in a zig-zag structure for the peptide chain. Note the shaded circle around R; the extended strand arrangement also allows the **maximum space and freedom of movement for a side chain**. The repeat between identically oriented R-groups is 7.0 Å, with 3.5 Å per amino acid, matching the fiber diffraction data for beta-keratins.



Pauling's extended state model matched the spacing of fibroin exactly (3.5 and 7.0 Å). In the extended state, H-bonding NH and CO groups point out at 90^{o} to the strand. If extended strands are lined up side by side, H-bonds bridge from strand to strand. Identical or opposed strand alignments make up parallel or antiparallel beta sheets (named for beta keratin). Antiparallel beta-sheet is significantly more stable due to the well aligned H-bonds.



Angle	Antiparallel	Parallel
Φ	–139°	–119°
Ψ	135°	113°

Amino acid preferences for different secondary structure

Alpha helix may be considered the default state for secondary structure. Although the potential energy is not as low as for beta sheet, H-bond formation is intra-strand, so there is an entropic advantage over beta sheet, where H-bonds must form from strand to strand, with strand segments that may be quite distant in the polypeptide sequence.

The main criterion for alpha helix preference is that the amino acid side chain should **cover and protect the backbone H-bonds** in the core of the helix. Most amino acids do this with some key exceptions:

alpha-helix preference: **Ala,Leu,Met,Phe,Glu,Gln,His,Lys,Arg**

The extended structure leaves the **maximum space free** for the amino acid side chains: as a result, those amino acids with **large bulky side chains prefer to form beta sheet structures**:

just plain large:	**Tyr, Trp, (Phe, Met)**
bulky and awkward due to branched beta carbon:	**Ile, Val, Thr**
large S atom on beta carbon:	**Cys**

The remaining amino acids have side chains which **disrupt secondary structure**, and are known as **secondary structure breakers**:

side chain H is too small to protect backbone H-bond:	**Gly**
side chain linked to alpha N, **has no N-H** to H-bond; **rigid structure** due to ring restricts to phi = -60^{o};	**Pro**
H-bonding side chains compete directly with backbone H-bonds	**Asp, Asn, Ser**

Clusters of breakers give rise to regions known as **loops or turns** which mark the boundaries of regular secondary structure, and serve to link up secondary structure segments.

β-turn

Turns are the third of the three "classical" secondary structures that serve to reverse the direction of the polypeptide chain.

They are located primarily on the protein surface and accordingly contain polar and charged residues.

Turns were first recognised from a theoretical conformational analysis by Venkatachalam (1968). He considered what conformations were available to a system of three linked peptide units (or four successive residues) that could be stabilised by a backbone hydrogen bond between the CO of residue n and the NH of residue n+3.

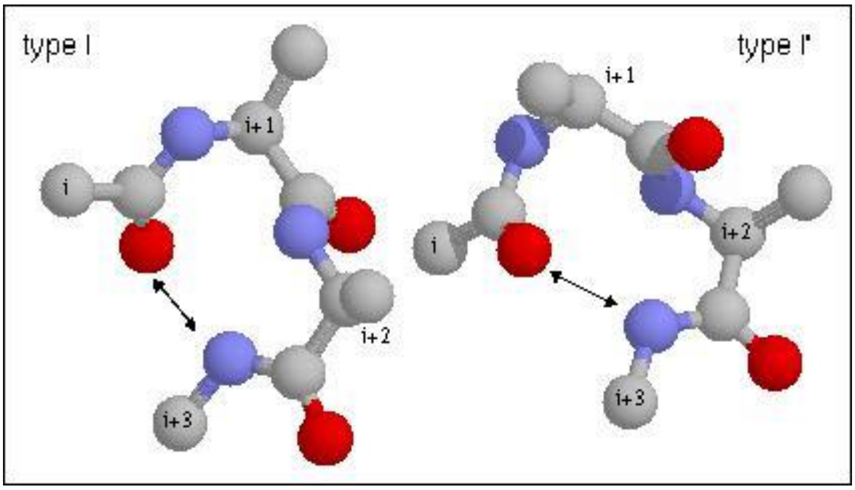
He found three general types, one of which

(type III) actually has repeating ϕ, ψ values of -60deg, -30deg and is identical with the 3_{10}-helix. The three types each contain a hydrogen bond between the carbonyl oxygen of residue i and the amide nitrogen of i+3. These three types of turns are designated I, II, and III. Many have speculated on the role of this type of secondary structure in globular proteins.

Turns may be viewed as a weak link in the polypeptide chain, allowing the other secondary structures (helix and sheet) to determine the conformational outcome. In contrast (based on the recent experimental finding of "turn-like" structures in short peptides in aqueous solutions, turns are considered to be structure-nucleating segments, formed early in the folding process.

Type I turns occur 2-3 times more frequently than type II. There are position dependent amino acid preferences for residues in turn conformations.

Type I can tolerate all residues in position i to i+3 with the exception of Pro at position i+2. Proline is favoured at position i+1 and Gly is favoured at i+3 in type I and type II turns. The polar sidechains of Asn, Asp, Ser, and Cys often populate position i where they can hydrogen bond to the backbone NH of residue i+2.



Other secondary structures

Random coil

most proteins have regions in which the Φ and Ψ angles are not repeating. These regions are sometimes referred to as “random coil” although their structures are not actually “random”. The non-repeating structures may be considered “secondary structure”, in spite of their irregular nature.

Fibrous Proteins Are Adapted for a Structural Function

α-Keratin, collagen, and elastin provide clear examples of the relationship between protein structure and biological function (Table 1).

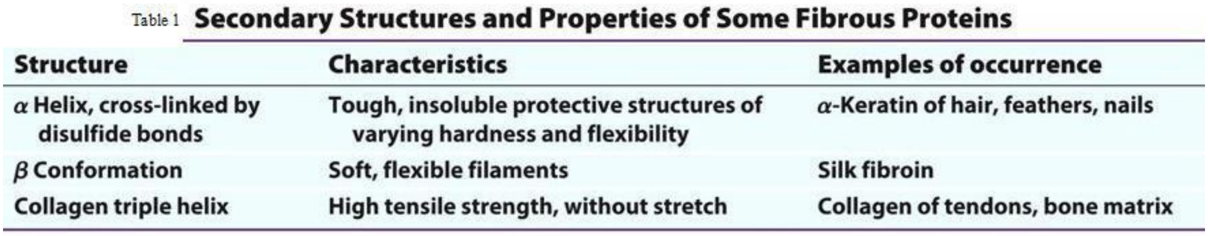


Table 1 **Secondary Structures and Properties of Some Fibrous Proteins**

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, nails
β Conformation	Soft, flexible filaments	Silk fibroin
Collagen triple helix	High tensile strength, without stretch	Collagen of tendons, bone matrix

These proteins share properties that give strength and/or elasticity to structures in which they occur. They have relatively simple structures, and all are insoluble in water, a property conferred by a high concentration of hydrophobic amino acids both in the interior of the protein and on the surface. These proteins represent an exception to the rule that hydrophobic groups must be buried. The hydrophobic core of the molecule therefore contributes less to structural stability, and covalent bonds assume an especially important role.

α-Keratin and **collagen** have evolved for strength.

In vertebrates, α-keratins constitute almost the entire dry weight of hair, wool, feathers, nails, claws, quills, scales, horns, hooves, tortoise shell, and much of the outer layer of skin.

Collagen is found in connective tissue such as tendons, cartilage, the organic matrix of bones, and the cornea of the eye.

The polypeptide chains of both proteins have simple helical structures. The α-keratin helix is the right-handed α helix found in many other proteins (Fig. 2). However, the collagen helix is unique. It is left-handed (see Fig 3) and has three amino acid residues per turn (Fig. 3).

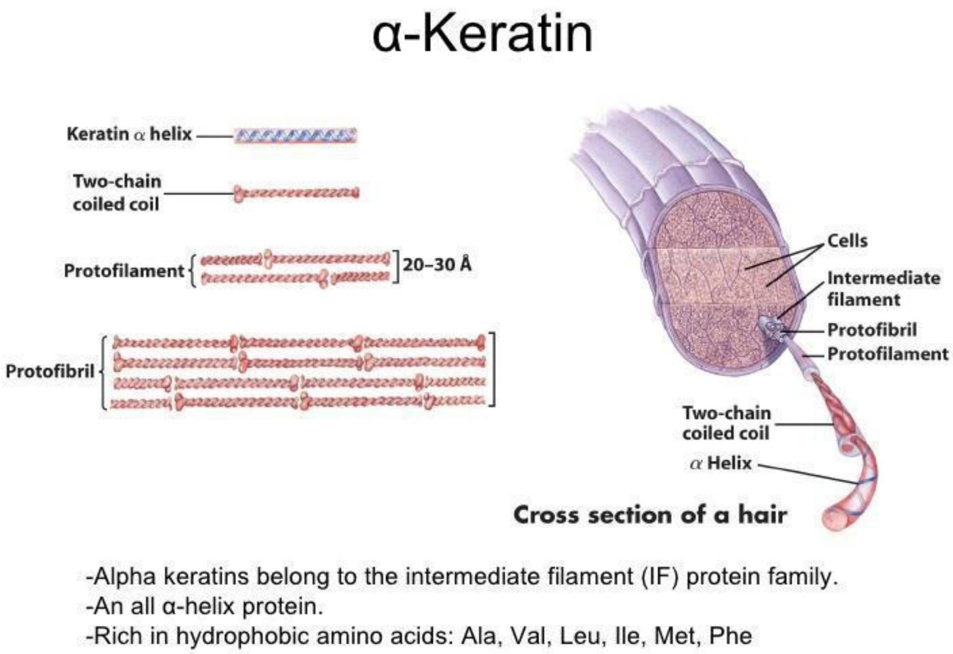
In both α-keratin and collagen, a few amino acids predominate.

α-Keratin is rich in the hydrophobic residues Phe, Ile, Val, Met, and Ala.

Collagen is 35% Gly, 11% Ala, and 21% Pro and Hyp (hydroxyproline; see Fig. 3). The unusual amino acid content of collagen is imposed by structural constraints unique to the collagen helix. The amino acid sequence in collagen is generally a repeating tripeptide unit, Gly-X-Pro or Gly-X-Hyp, where X can be any amino acid. The food product gelatin is derived from collagen. Although it is protein, it has little nutritional value because collagen lacks significant amounts of many amino acids that are essential in the human diet.

In both α-keratin and collagen, strength is amplified by wrapping multiple helical strands together in a superhelix, much the way strings are twisted to make a strong rope (Figs. 2,3). In both proteins the helical path of the supertwists is opposite in sense to the twisting of the individual polypeptide helices, a conformation that permits the closest possible packing of the multiple polypeptide chains.

The superhelical twisting is probably left-handed in α-keratin (Fig.2) and right-handed in collagen (Fig.3). The tight wrapping of the collagen triple helix provides great tensile strength with no capacity to stretch: Collagen fibers can support up to 10,000 times their own weight and are said to have greater tensile strength than a steel wire of equal cross section.



α-Keratin

-Alpha keratins belong to the intermediate filament (IF) protein family.
-An all α-helix protein.
-Rich in hydrophobic amino acids: Ala, Val, Leu, Ile, Met, Phe

Fig 2 α-keratin

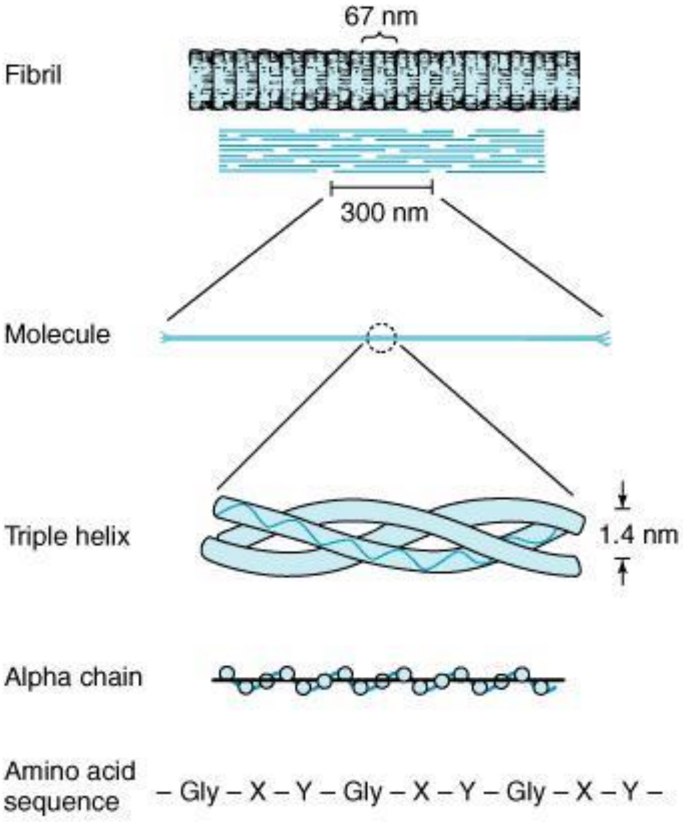


Fig 3 collagen

The fibroin protein consists of layers of antiparallel beta sheets (Fig 4). Its primary structure mainly consists of the recurrent amino acid sequence (Gly-Ser-Gly-Ala-Gly-Ala)*n*. The high glycine (and, to a lesser extent, alanine) content allows for tight packing of the sheets, which contributes to silk's rigid structure and tensile strength. A combination of stiffness and toughness make it a material with applications in several areas, including biomedicine and textile manufacture.

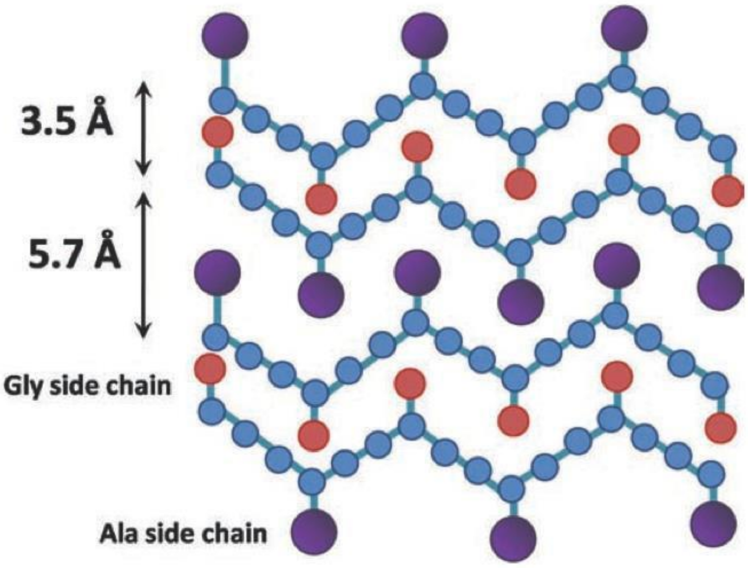
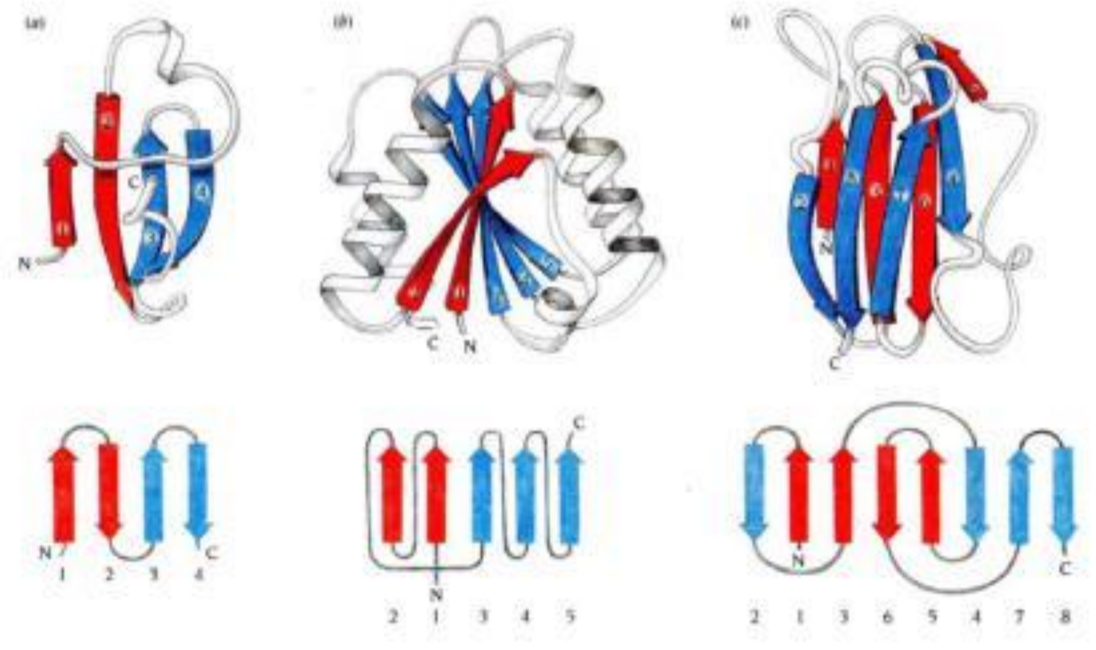


Fig 4 Silk Fibroin

Topology diagrams

The most characteristic features of a β sheet are the number of strands, their relative directions (parallel or antiparallel), and how the strands are connected. This information can be represented by topology diagrams. They are useful to compare β structures.



1.2.Protein Tertiary Structure

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.

Examination of protein structures resolved by X-ray diffraction and NMR has revealed a variety of folding patterns common to many different proteins. However, even within these folds, distinct substructures or structural **motifs**, i.e. distinctive arrangements of elements of secondary structure, have been described. The term **supersecondary structure** has been coined to describe this level of organisation, which is intermediate between secondary and tertiary.

Motifs or folds, are particularly stable arrangements of several elements of the secondary structure. • Supersecondary structures are usually produced by packing side chains from adjacent secondary structural elements close to each other.

Rules for secondary structure.

- Hydrophobic side groups must be buried inside the folds, therefore, layers must be created (β−α−β; α− α).
- α-helix and β-sheet, if occur together, are found in different structural layers.
- Adjacent polypeptide segments are stacked together.
- Connections between secondary structures do not form knots.
- The β-sheet is the most stable.

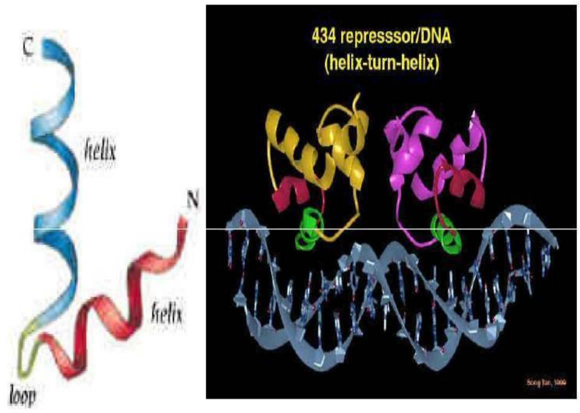
Motif

- Secondary structure composition, e.g. all α, all β, segregated α+β, mixed α/β
- Motif = small, specific combinations of secondary structure elements, e.g. β-α-β loop

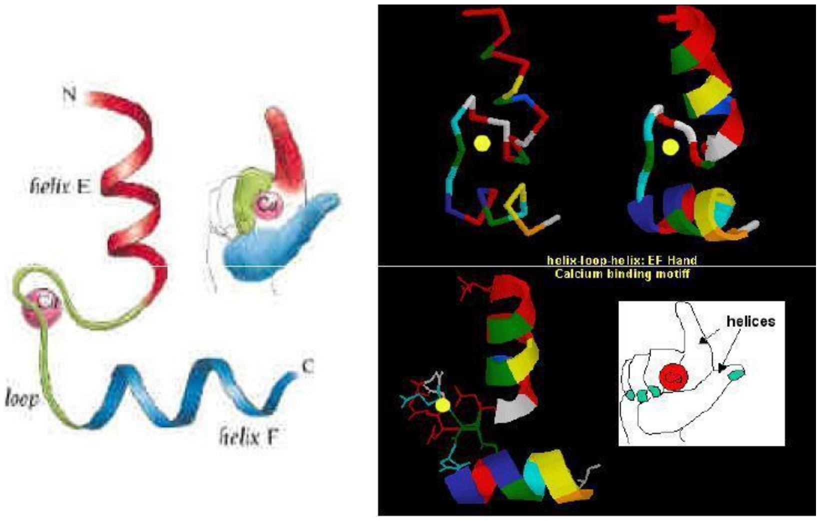
1. Helix super secondary structures

Helix-Turn-Helix Motif

Also called the alpha-alpha type (αα-type). The motif is compromised of two antiparallel helices connected by a turn. The helix-turn-helix is a functional motif and is usually identified in proteins that bind to DNA minor and major grooves, and Calcium-binding proteins.

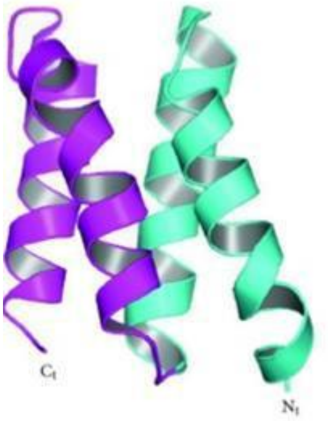


DNA binding Helix-turn-Helix motif



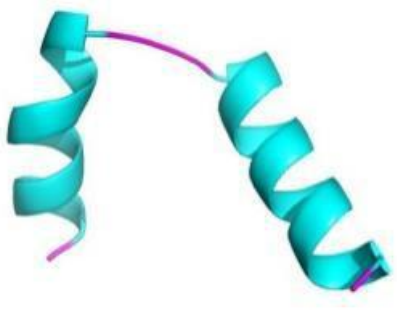
Calcium binding (EF Hand- Calcium binding) motif

Helix-hairpin-helix: Involved in DNA binding



Alpha-alpha corner

Short loop regions connecting helices which are roughly perpendicular to one another



2. Sheet super secondary structures

All beta tertiary structural domains can occur in proteins with one domain (eg. concanavalin A, superoxide dismutase), and occurs at least once in proteins with two domains (eg. chymotrypsin), or three domains (eg. OmpF).

The beta strands making up these domains are all essentially antiparallel and form structures to achieve stable packing arrangements within the protein.

There are presently (as of version 1.39) about 70 subclasses listed in SCOP for this domain, and some examples of these are outlined below.

Beta barrels

This is the most abundant beta-domain structure and as the name suggests the domain forms a 'barrel-like' structure. The beta barrels are not geometrically perfect and can be rather distorted.

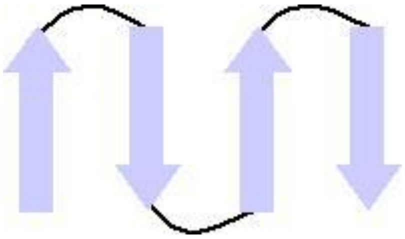
There are three main types:

1. Up-and-down barrels

2. Greek key barrels

3. Jelly roll (Swiss roll) barrels

Up-and-down beta-sheets or beta-barrels



The simple topology of an up-and-down barrel (named because the beta strands follow each other in sequence in an up-and-down fashion).

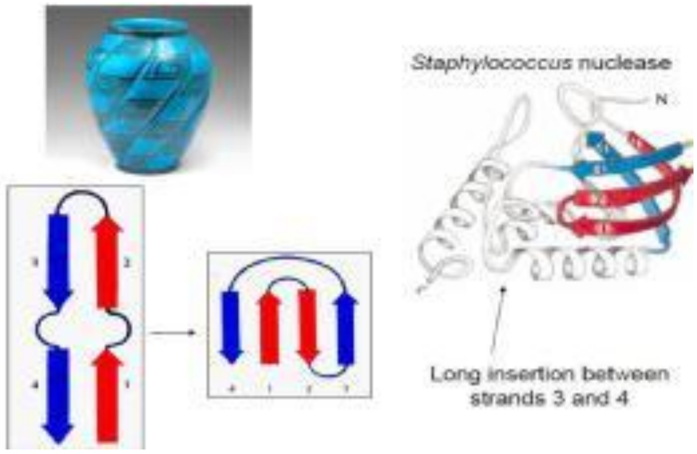
Usually, the loops joining the beta strands do not crossover the 'ends' of the barrel.

Greek key barrels

These are barrels formed from two, or more, Greek Key motifs.

It is a stable structure

The Greek key barrel consists of four anti-parallel Beta strands where one strand changes the topology direction. Hydrogen bonding occurs between strands 1:4, and strands 2:3. Strand 2 then folds over to form the structural motif.



Jelly roll barrels

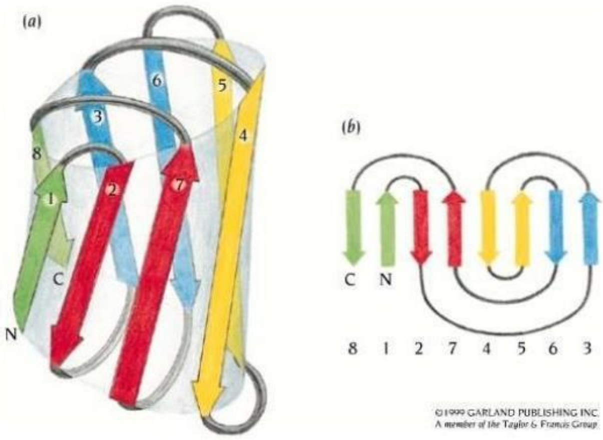
These barrels are formed from a 'Greek Key-like' structure called a jelly roll. Supposedly named because the polypeptide chain is wrapped around a barrel core like a jelly roll (swiss roll).

It is a stable structure

This structure is found in coat proteins of spherical viruses, plant lectin concanavalin A, and hemagglutinin protein from influenza virus.

The essential features of a jelly roll barrel are that:

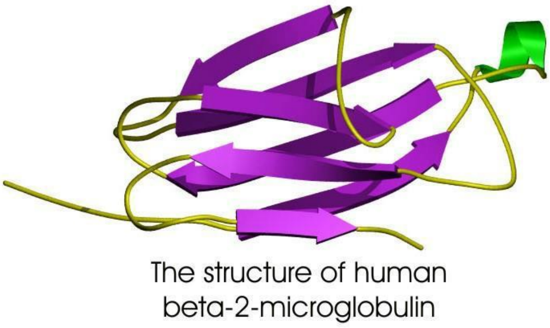
- it is like an inverted 'U' (which is often seen twisted and distorted in proteins)
- it is usually divided into two beta sheets which are packed against each other
- most jelly roll barrels have eight strands although any even number greater than 8 can form a jelly roll barrel
- it folds such that hydrogen bonds exist between strands 1 and 8; 2 and 7; 3 and 6; and 4 and 5



©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

Beta sandwich

A beta sandwich is essentially a 'flattened' beta barrel with the two sheets packing closely together (like a sandwich!). The first and last strands of the sandwich do not hydrogen bond to each other to complete a 'barrel' structure.

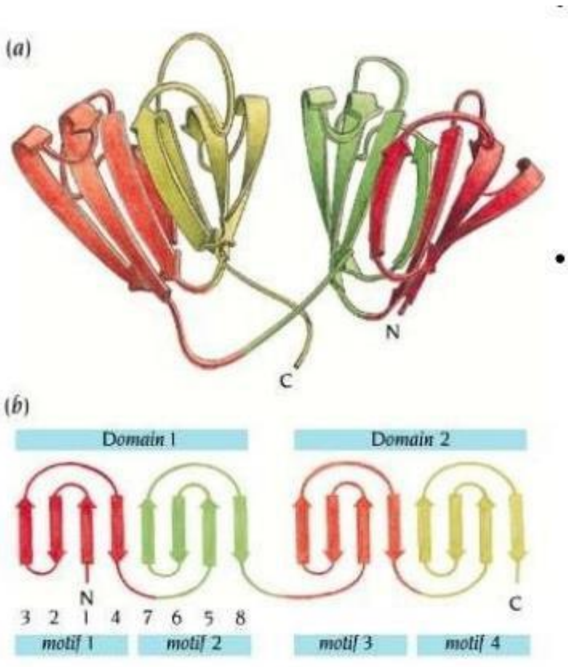


Beta sandwich in beta 2 microglobulin.

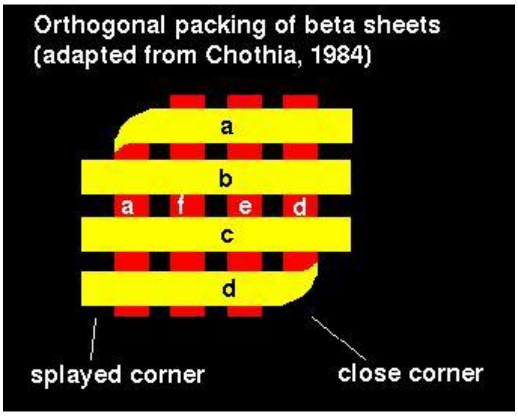
Aligned or Orthogonal beta strands

Beta strands in barrels or sandwich structures can be orientated in two general ways:

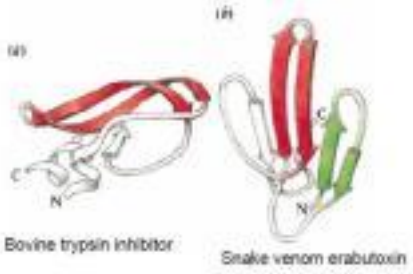
- where the strands in two sheets are almost aligned, and in the same orientation, to each other and form an 'aligned beta' structure (eg. gamma crystallin)



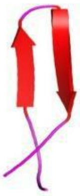
- where the strands, in at least two sheets, are roughly perpendicular to each other and form an 'orthogonal beta' structure.


Orthogonal packing of beta sheets
(adapted from Chothia, 1984)
a
b
a
f
e
d
c
d
splayed corner
close corner

Beta-hairpin: two antiparallel beta strands connected by a “hairpin” bend, i.e. beta-turn 2 x antiparallel beta-strands + beta-turn = beta hairpin



Beta-beta corner

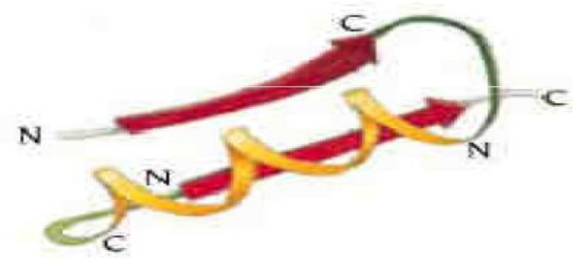


- Two antiparallel beta strands which form a beta hairpin can change direction abruptly. The angle of the change of direction is about 90 degrees and so the structure is known as a 'beta corner'
- □ The abrupt angle change is achieved by one strand having a glycine residue (so there is no steric hindrance from a side chain) and the other strand having a beta bulge (where the hydrogen bond is broken).
- no known function

α/β Topologies

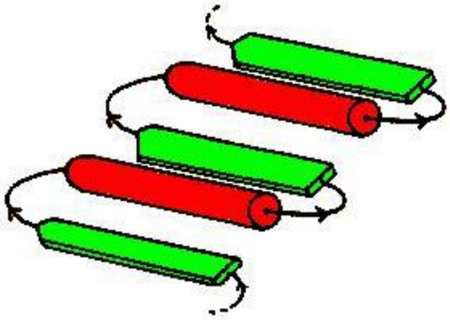
Beta-Helix-Beta Motif

An important and widespread supersecondary structural motif in proteins is known as the β-α-β motif (Beta-Alpha-Beta motif). The motif consists of two parallel Beta strands that is connected via an alpha helix (with two turns). The motif is found in most proteins that contain parallel beta strands, and the axis of the Helix and the Strands are roughly parallel to each other with all three elements forming a hydrophobic core due to shielding. The β-α-β motif may be structurally or functionally involved. The Loop that connects the C-terminal of first Beta strand and N-terminal of Helix is frequently involved in ligand binding functions, and the motif itself is frequently found in ion channels.



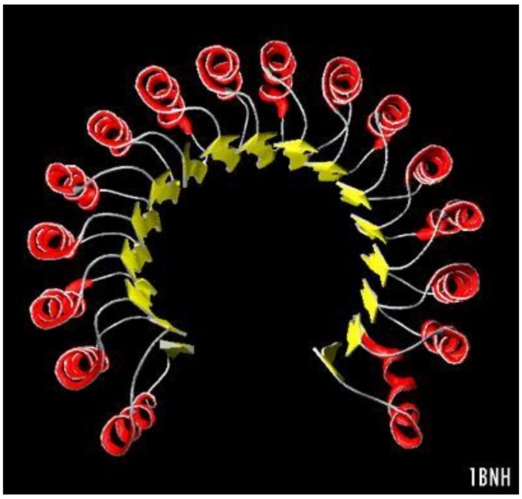
The β - α - β - α - β subunit, often present in nucleotide-binding proteins, is named the **Rossman Fold**, after Michael Rossman

The Rossman fold



α/β horseshoe

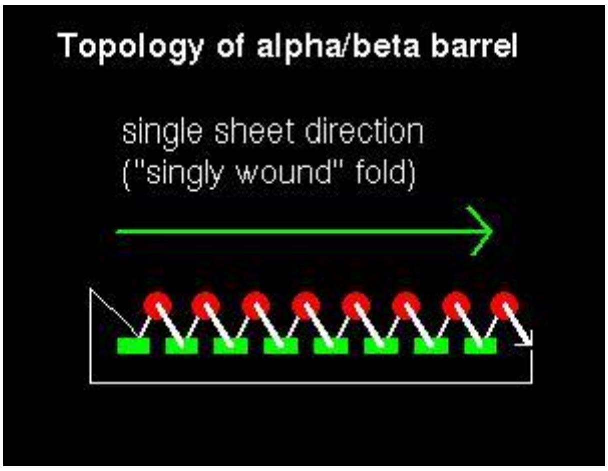
17-stranded parallel b sheet curved into an open horseshoe shape, with 16 a-helices packed against the outer surface. It doesn't form a barrel although it looks as though it should. The strands are only very slightly slanted, being nearly parallel to the central `axis'.



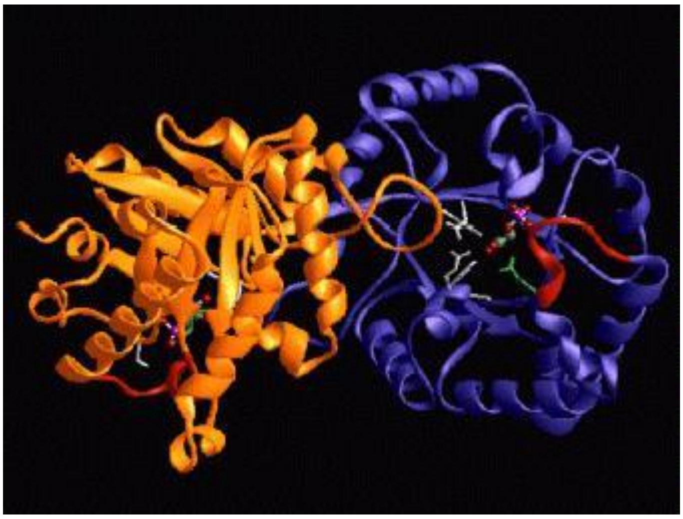
placental ribonuclease inhibitor takes the concept of the repeating α/β unit to extremes.

α/β barrels

Consider a sequence of eight α/β motifs:

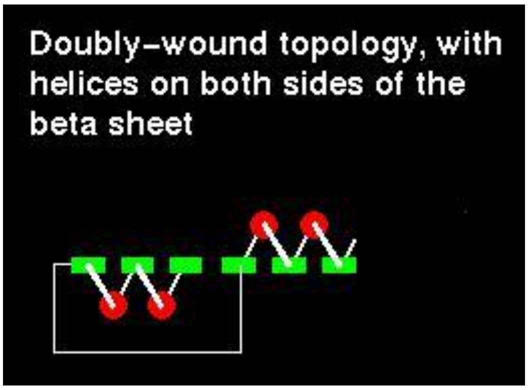


If the first strand hydrogen bonds to the last, then the structure closes on itself forming a barrel-like structure. This is shown in the picture of triose phosphate isomerase.



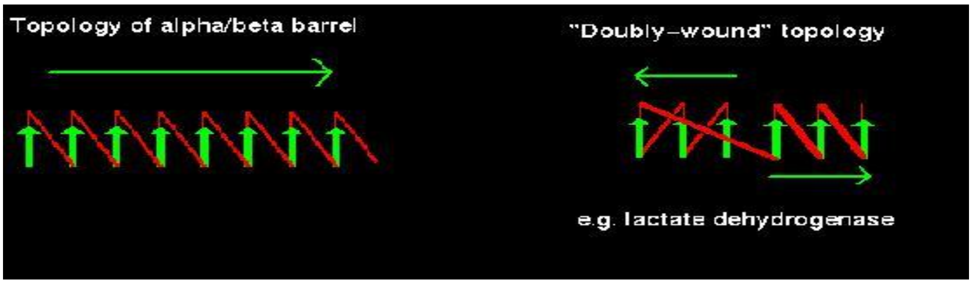
Note that the "staves" of the barrel are slanted, due to the twist of the b sheet. Also notice that there are effectively four layers to this structure. The direction of the sheet does not change (it is anticlockwise in the diagram). Such a structure may therefore be described as **singly wound**.

In a structure which is open rather than closed like the barrel, helices would be situated on only one side of the b sheet if the sheet direction did not reverse. Therefore open a/b structures must be **doubly wound** to cover both sides of the sheet.



Doubly–wound topology, with helices on both sides of the beta sheet

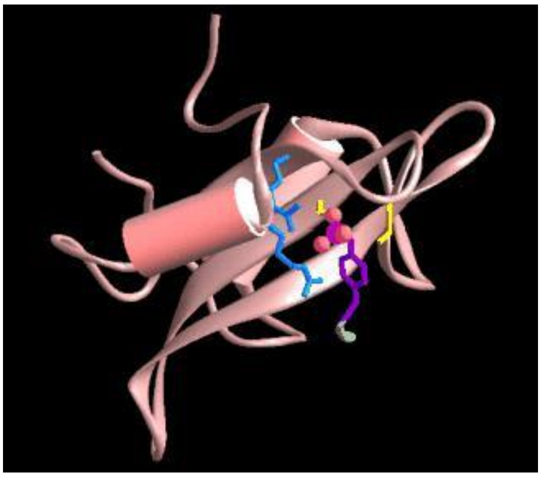
The chain starts in the middle of the sheet and travels outwards, then returns to the centre via a loop and travels outwards to the opposite edge:



Doubly-wound topologies where the sheet begins at the edge and works inwards are rarely observed.

Alpha+Beta Topologies

This is where we collect together all those folds which include significant alpha and beta secondary structural elements, but for which those elements are `**mixed**', in the sense that they do NOT exhibit the wound alpha-beta topology. This class of folds is therefore referred to as **α+ β**



Domains

stable, independently folded, globular units, often consisting of combinations of motifs

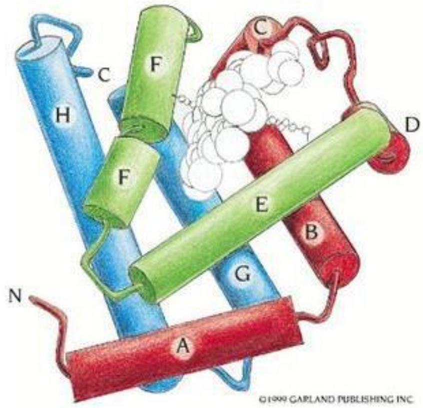
vary from 25 to 300 amino acids, average length – 100.

large globular proteins may consist of several domains linked by stretches of polypeptide. Separate domain may have distinct functions (eg G3P dehydrogenase). In many cases binding site formed by cleft between 2 domains

frequently correspond to exon in gene

- Some examples of domains:

1. in volving α-
helix 4-helix bundle
globin fold

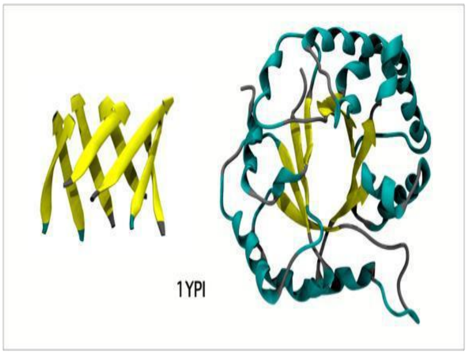


The globin fold is found in its namesake globin protein families: hemoglobins and myoglobins, as well as in phycocyanins. Because myoglobin was the first protein whose structure was solved, the globin fold was thus the first protein fold discovered.

2. parallel β-sheets

hydrophobic residues on both sides, therefore must be buried.

□□ barrel: 8 β strands each flanked by an antiparallel α-helix eg triose phosphate isomerase.)



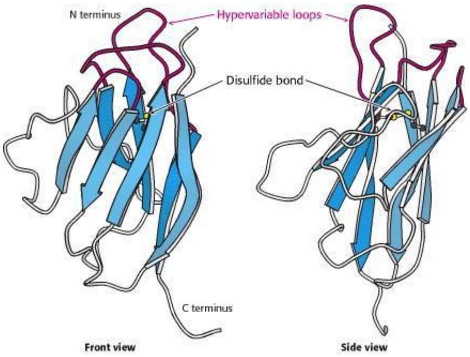
3. antiparallel β -sheet

hydrophobic residues on one side, one side can be exposed to environment, minimum structure 2 layers

Sheets arranged in a barrel shape

More common than parallel β -barrels

eg. immunoglobulin



The **immunoglobulin domain** is a type of protein domain that consists of a 2-layer sandwich of 7-9 antiparallel β-strands arranged in two β-sheets with a Greek keytopology, consisting of about 80 amino acids.

The backbone switches repeatedly between the two β-sheets. Typically, the pattern is (N-terminal β-hairpin in sheet 1)-(β-hairpin in sheet 2)-(β-strand in sheet 1)-(C-terminal β-hairpin in sheet 2). The cross-overs between sheets form an "X", so that the N- and C-terminal hairpins are facing each other.

Members of the immunoglobulin superfamily are found in hundreds of proteins of different functions. Examples include antibodies, the giant muscle kinase titin, andreceptor tyrosine kinases. Immunoglobulin-like domains may be involved in protein–protein and protein–ligand interactions.

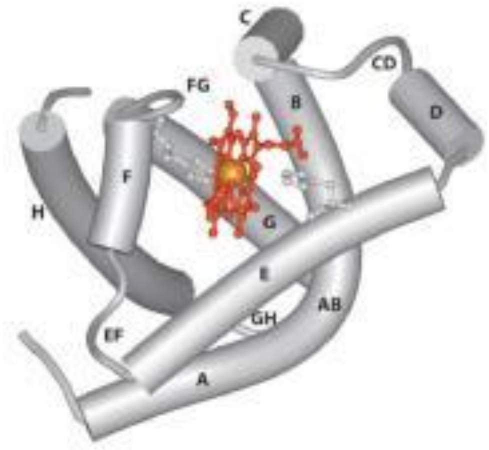
Example of Tertiary Structure: Myoglobin and Hemoglobin

Myoglobin and hemoglobin are hemeproteins whose physiological importance is principally related to their ability to bind molecular oxygen.

Myoglobin

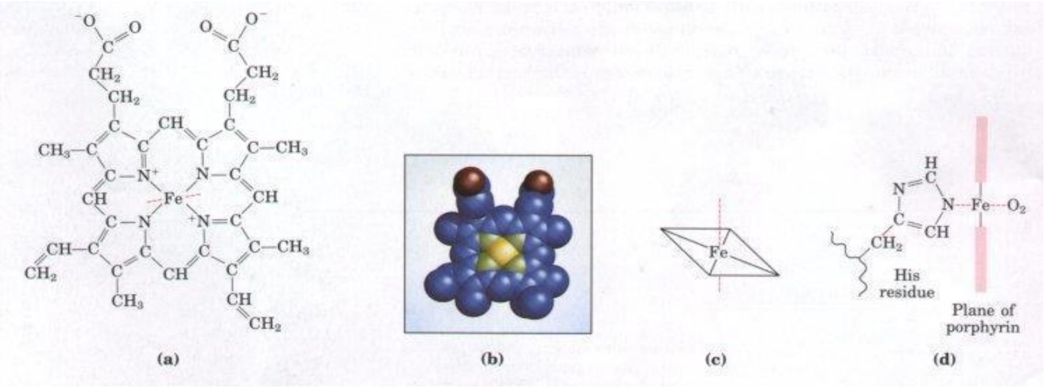
Single polypeptide chain (153 amino acids)

No disulfide bonds 8 right handed alpha helices form a hydrophobic pocket which contains heme molecule protective sheath for a heme group



Myoglobin is a monomeric heme protein found mainly in muscle tissue where it serves as an intracellular storage site for oxygen During periods of oxygen deprivation oxymyoglobin releases its bound oxygen which is then used for metabolic purposes The tertiary structure of myoglobin is that of a typical water soluble globular protein Its secondary structure is unusual in that it contains a very high proportion (75%) of α-helical secondary structure A myoglobin polypeptide is comprised of 8 separate right handed a-helices, designated A through H, that are connected by short non helical regions Amino acid R-groups packed into the interior of the molecule are predominantly hydrophobic in character while those exposed on the surface of the molecule are generally hydrophilic, thus making the molecule relatively water soluble

Each myoglobin molecule contains one heme prosthetic group inserted into a hydrophobic cleft in the protein Each heme residue contains one central coordinately bound iron atom that is normally in the Fe 2+ , or ferrous, oxidation state The oxygen carried by hemeproteins is bound directly to the ferrous iron atom of the heme prosthetic group



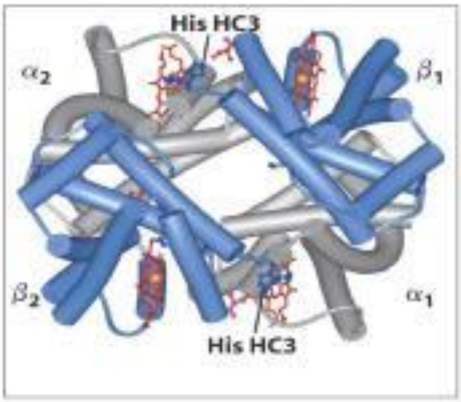
The heme group is located in a crevice Except for one edge, non polar side chains surround the heme Fe 2+ is octahedrally coordinated Fe 2+ covalently bonded to the imidazole group of histidine 93 (F8) O 2 held on the other side by histidine 64 (E7)

Hydrophobic interactions between the tetrapyrrole ring and hydrophobic amino acid R groups on the interior of the cleft in the protein strongly stabilize the heme protein conjugate. In addition a nitrogen atom from a histidine R group located above the plane of the heme ring is coordinated with the iron atom further stabilizing the interaction between the heme and the protein. In oxymyoglobin the remaining bonding site on the iron atom (the 6th coordinate position) is occupied by the oxygen, whose binding is stabilized by a second histidine residue Carbon monoxide also binds coordinately to heme iron atoms in a manner similar to that of oxygen, but the binding of carbon monoxide to heme is much stronger than that of oxygen.

The preferential binding of carbon monoxide to heme iron is largely responsible for the asphyxiation that results from carbon monoxide poisoning.

Hemoglobin

Oxygen transporter Four polypeptide chains Tetramer Each chain has a heme group Hence four O 2 can bind to each Hb Two alpha (141 amino acids) and two beta (146 amino acids) chains



Hemoglobin is an [α(2):β(2)] tetrameric hemeprotein found in erythrocytes where it is responsible for binding oxygen in the lung and transporting the bound oxygen throughout the body where it is used in aerobic metabolic pathways Each subunit of a hemoglobin tetramer has a heme prosthetic group identical to that described for myoglobin. Although the secondary and tertiary structure of various hemoglobin subunits are similar, reflecting extensive homology in amino acid composition, the variations in amino acid composition that do exist impart marked differences in hemoglobin's oxygen carrying properties In addition, the quaternary structure of hemoglobin leads to physiologically important allosteric interactions between the subunits, a property lacking in monomeric myoglobin which is otherwise very similar to the α-subunit of haemoglobin

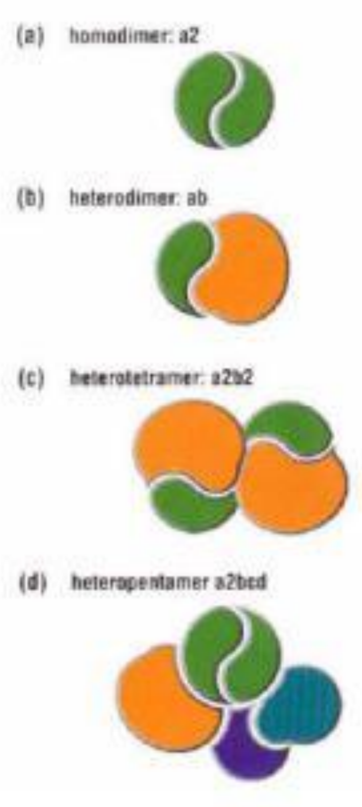
1.3. Quaternary structure

- 3-dimensional relationship of the different polypeptide chains (subunits) in a multimeric protein, the way the subunits fit together and their symmetry relationships

• only in proteins with more than one polypeptide chain; proteins with only one chain have no quaternary structure.)

Terminology

• Each polypeptide chain in a multichain protein = a subunit • 2-subunit protein = a dimer, 3 subunits = trimeric protein, 4 = tetrameric • homo(dimer or trimer etc.): identical subunits • hetero(dimer or trimer etc.): more than one kind of subunit (chains with different amino acid sequences) • different subunits designated with Greek letters – e.g., subunits of a heterodimeric protein = the "α subunit" and the "β subunit".



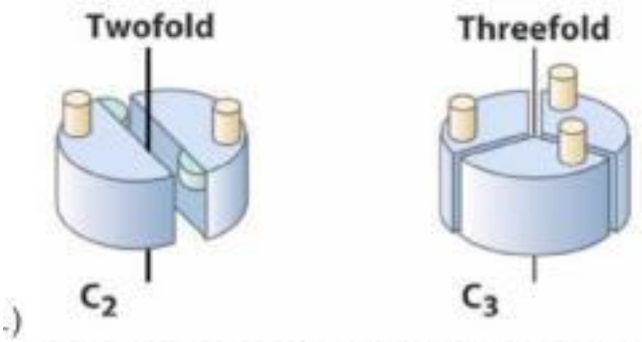
– NOTE: This use of the Greek letters to differentiate different polypeptide chains in a multimeric protein has nothing to do with the names for the secondary structures α helix and β conformation.

• Some protein structures have very complex quaternary arrangements; e.g., mitochondrial ATP synthase, viral capsids….

Symmetry in quaternary structures

- simplest kind of symmetry = rotational symmetry

- Individual subunits can be superimposed on other identical subunits (brought into coincidence) by rotation about one or more rotational axes.



Two types of cyclic symmetry

- If the required rotation = 180° (360°/2), protein has a 2-fold axis of symmetry (e.g., Cro repressor protein above).

- If the rotation = 120° (360°/3), e.g., for a homotrimer, the protein has a 3-fold symmetry axis. Rotational symmetry in proteins: Cyclic symmetry: all subunits are related by rotation about a single n-fold rotation axis (C2 symmetry has a 2-fold axis, 2 identical subunits; C3 symmetry has a 3-fold axis, 3 identical subunits, etc.)

Example: Protein Capsid

Viral genomes are surrounded by protein shells known as capsids. One interesting question is how capsid proteins recognize viral, but not cellular RNA or DNA. The answer is that there is often some type of "packaging" signal (sequence) on the viral genome that is recognized by the capsid proteins. A capsid is almost always made up of repeating structural subunits that are arranged in one of two symmetrical structures, a **helix** or an **icosahedron**. In the simplest case, these "subunits" consist of a single polypeptide. In many cases, however, these **structural subunits (also called protomers)** are made up of several polypeptides. Both helical and icosahedral structures are described in more detail below.

1) **Helical Capsids**: The first and best studied example is the plant tobacco mosaic virus (TMV), which contains a SS RNA genome and a protein coat made up of a single, 17.5 kd protein. This protein is arranged in a helix around the viral RNA, with 3 nt of RNA fitting into a groove in each subunit. Helical capsids can also be more complex, and involve more than one protein subunit.

A helix can be defined by two parameters, its amplitude (diameter) and pitch, where pitch is defined as the distance covered by each turn of the helix. **P = m x p**, where m is the number of subunits per turn and p is the axial rise per subunit. For TMV, m = 16.3 and p= 0.14 nm, so P=2.28 nm. This structure is very stable, and can be dissociated and re-associated readily by changing ionic strength, pH, temperature, etc. The interactions that hold these molecules together are non-covalent, and involve H-bonds, salt bridges, hydrophobic interactions, and vander Waals forces.

Several families of animal virus contain helical nucleocapsids, including the *Orthomyxoviridae* (influenza), the *Paramyxoviridae* (bovine respiratory syncytial virus), and the *Rhabdoviridae* (rabies). All of these are enveloped viruses (see below).

2) **Icosahedral Capsids**: In these structures, the subunits are arranged in the form of a hollow, quasi spherical structure, with the genome within. An icosahedron is defined as being made up of **20 equilateral triangular faces** arranged around the surface of a sphere. They display **2-3-5 fold symmetry** as follows:

- an axis of 2 fold rotational symmetry through the center of each edge.

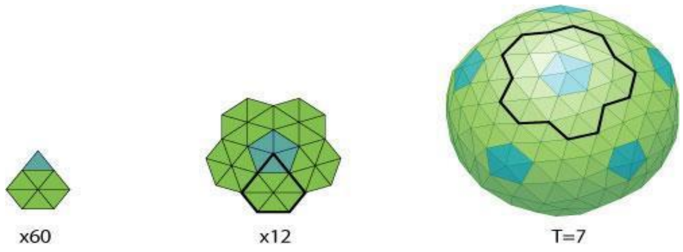
- an axis of 3 fold rotational symmetry through the center of each face.

- an axis of 5 fold rotational symmetry through the center of each corner.

These corners are also called Vertices, and each icosahedron has 12.

Since proteins are not equilateral triangles, each face of an icosahedron contains more than one protein subunit. The simplest icosahedron is made by using 3 identical subunits to form each face, so the minimum # of subunits is 60 (20 x 3). Remember, that each of these subunits could be a single protein or, more likely, a complex of several polypeptides.

Many viruses have too large a genome to be packaged inside an icosahedron made up of only 60 polypeptides (or even 60 subunits), so many are more complicated. In these cases, each of the 20 triangular faces is divided into smaller triangles; and each of these smaller triangles is defined by 3 subunits. However, the total number of subunits is always a multiple of 60. The total number of subunits can be defined as 60 X N, where N is sometimes called the **Triangulation Number**, or T. Values for T of 1,3,4,7,9, 12 and more are permitted.

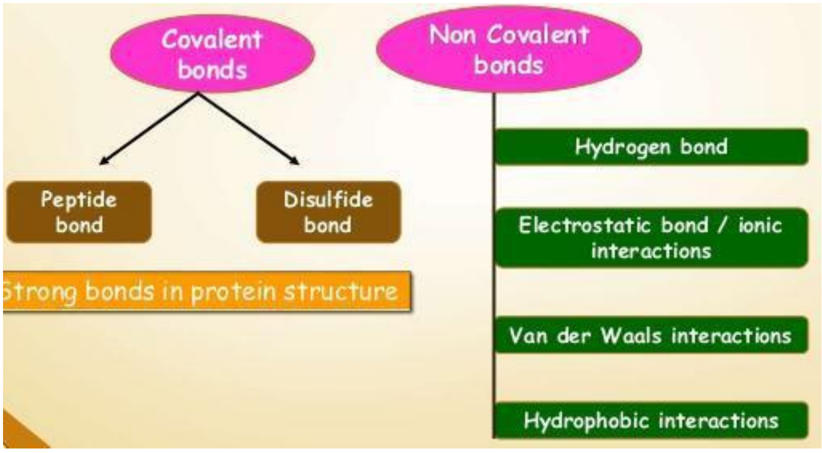


© 2010
Swiss Institute of Bioinformatics

When virus nucleocapsids are observed in the electron microscope, one often sees apparent "lumps" or clusters on the surface of the particle. These are usually protein subunits clustered around an axis of symmetry, and have been called "morphological units" or **capsomers**.

Forces that stabilize Protein Structure

Proteins are formed of amino acids linked together by the following types of bonds


Covalent bonds

Peptide bond

Disulfide bond

Strong bonds in protein structure

Non Covalent bonds

Hydrogen bond

Electrostatic bond / ionic interactions

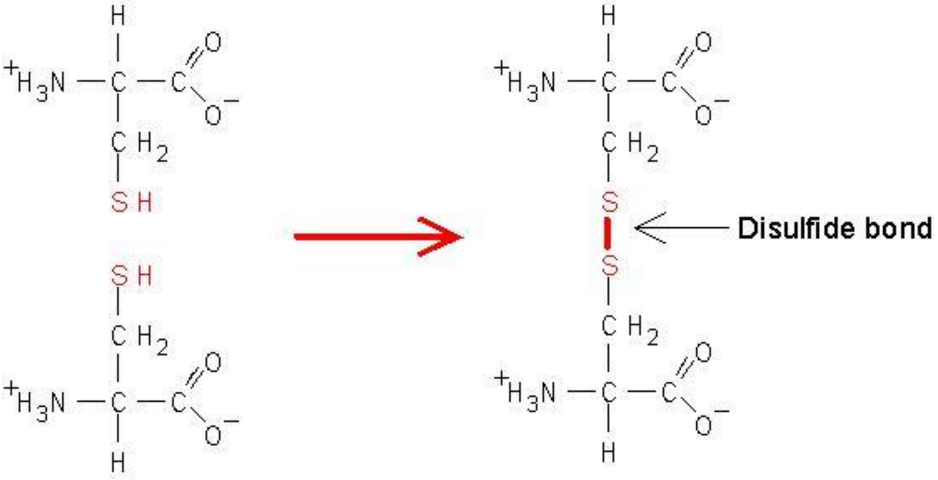
Van der Waals interactions

Hydrophobic interactions

Covalent Bonds - Disulfide Bridges

Covalent bonds are the strongest chemical bonds contributing to protein structure. Covalent bonds arise when two atoms share electrons.

In addition to the covalent bonds that connect the atoms of a single amino acid and the covalent peptide bond that links amino acids in a protein chain, covalent bonds between cysteine side chains can be important determinants of protein structure. Cysteine is the sole amino acid whose side chain can form covalent bonds, yielding disulfide bridges with other cysteine side chains: --CH_2-S-S-CH_2 . A disulfide bridge is shown here:



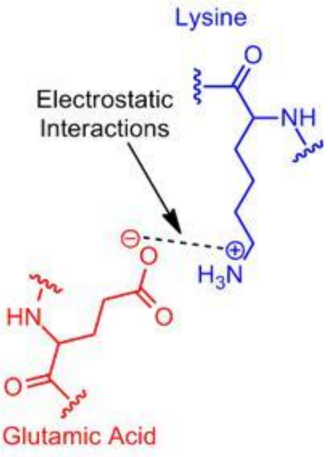
T

Non-covalent bonds

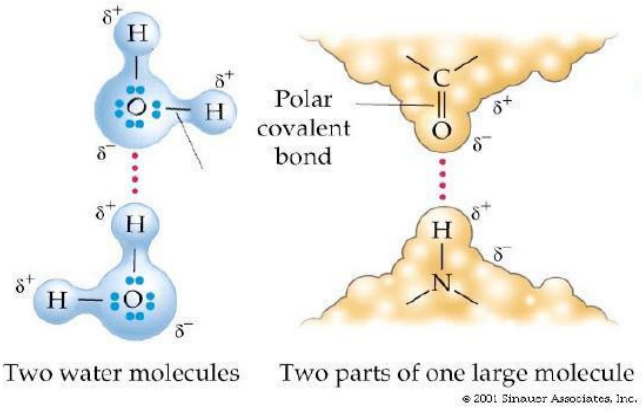
Electrostatic Interactions

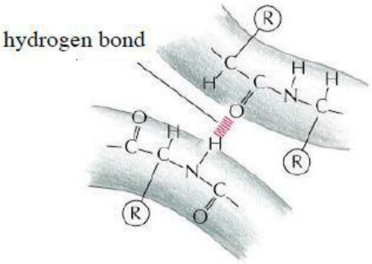
A. Ionic Bonds - Salt Bridges

Ionic bonds are formed as amino acids bearing opposite electrical charges are juxtaposed in the hydrophobic core of proteins. Ionic bonding in the interior is rare because most charged amino acids lie on the protein surface. Although rare, ionic bonds can be important to protein structure because they are potent electrostatic attractions that can approach the strength of covalent bonds. A ionic bond-salt bridge between a negatively charged O on the sidechain of glutamic acid lies 2.8 Å from the positively charged N on the amino terminus (lysine) is shown here .



B. Hydrogen Bonds



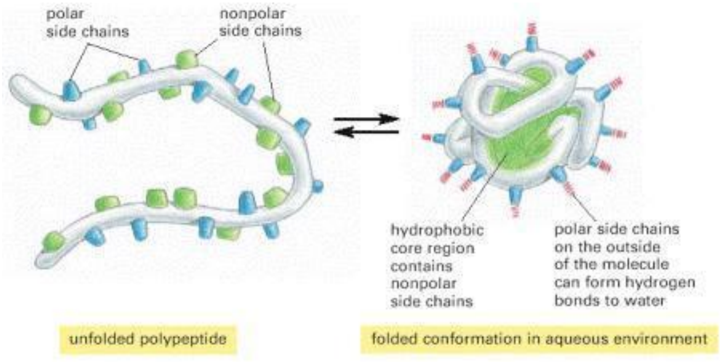


Hydrogen bonds are a particularly strong form of dipole-dipole interaction. Because atoms of different elements differ in their tendencies to hold onto electrons -- that is, because they have different electronegativities -- all bonds between unlike atoms are polarized, with more electron density residing on the more electronegative atom of the bonded pair. Separation of partial charges creates a dipole, which you can think of as a mini-magnet with a positive and a negative end. In any system, dipoles will tend to align so that the positive end of one dipole and the negative end of another dipole are in close proximity. This alignment is favorable. Hydrogen bonds are dipole-dipole interactions that form between heteroatoms in which one heteroatom (e.g. nitrogen) contains a bond to hydrogen and the other(e.g. oxygen) contains an available lone pair of electrons. You can think of the hydrogen in a hydrogen bond as being shared between the two heteroatoms, which is highly favorable. Hydrogen bonds have an ideal X-H-X angle of 180°, and the shorter they are, the stronger they are. Hydrogen bonds play an important role in the formation of secondary structure. Alpha helices are hydrogen bonded internally along the backbone whereas beta strands are hydrogen bonded to other beta strands. Side chains can also participate in hydrogen bonding interactions. You should be able to list the side chains that can participate in hydrogen bonds now that you know the structures of the side chains. Because hydrogen bonds are directional, meaning the participating dipoles must be aligned properly for a hydrogen bond to form (another w ay of saying it is that the hydrogen bonding angle must be larger than about 135°, with an optimum of 180°), and because unfavorable alignment of participating dipoles is repulsive, hydrogen bonds between side chains play key roles in determining the unique structures that different proteins form.

Hydrophobic Bonds

Hydrophobic bonds are a major force driving proper protein folding. Burying the nonpolar surfaces in the interior of a protein creates a situation where the water molecules can hydrogen bond with each other without becoming excessively ordered. Thus, the energy of the system goes down.

Therefore, an important factor governing the folding of any protein is the distribution of its polar and nonpolar amino acids. The nonpolar (hydrophobic) side chains in a protein such as those belonging to phenylalanine, leucine, isoleucine, valine, methionine and tryptophan tend to cluster in the interior of the molecule (just as hydrophobic oil droplets coalesce in water to form one large droplet). In contrast, polar side chains such as those belonging to arginine, glutamine, glutamate, lysine, etc. tend to arrange themselves near the outside of the molecule, where they can form hydrogen bonds with water and with other polar molecules. There are some polar amino acids in protein interiors, however, and these are very important in defining the precise shape adopted by the protein because the pairing of opposite poles is even more significant than it is in water.


polar side chains
nonpolar side chains
hydrophobic core region contains nonpolar side chains
polar side chains on the outside of the molecule can form hydrogen bonds to water
unfolded polypeptide
folded conformation in aqueous environment

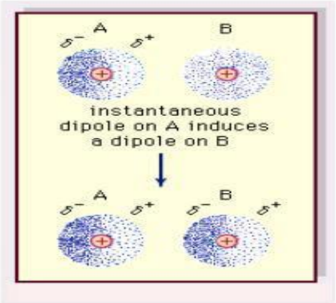
.

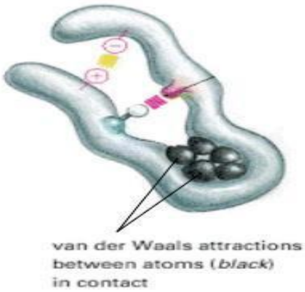
Van der Waals Forces

The Van der Waals force is a transient, weak electrical attraction of one atom for another. Van der Waals attractions exist because every atom has an electron cloud that can fluctuate, yielding a temporary electric dipole. The transient dipole in one atom can induce a complementary dipole in another atom, provided the two atoms are quite close. These short-lived, complementary dipoles provide a weak electrostatic attraction, the Van der Waals force. Of course, if the two electron clouds of adjacent atoms are too close, repulsive forces come into play because of the negatively-charged electrons. The appropriate distance required for Van der Waals attractions differs from atom to atom, based on the size of each electron cloud, and is referred to as the Van der Waals radius. The dots around atoms in this and other displays represent Van der Waals radii.

Van der Waals attractions, although transient and weak, can provide an important component of protein structure because of their sheer number. Most atoms of a protein are packed sufficiently close to others to be involved in transient Van der Waals attractions.

Van der Waals forces can play important roles in protein-protein recognition when complementary shapes are involved. This is the case in antibody-antigen recognition, where a "lock and key" fit of the two molecules yields extensive Van der Waals attractions.





Thermodynamics of protein folding

In contemplating protein folding, it is necessary to consider different types of amino acid side-chains separately. For each situation, the reaction involved will be assumed to be:

$$\text{Protein}_{unfolded} \rightleftharpoons \text{Protein}_{folded}$$

Note that this formalism means that a negative ΔG implies that the folding process is spontaneous.

First we will look at polar groups in an aqueous solvent. For polar groups, the ΔH_{chain} favors the unfolded structure because the backbone and polar groups interact form stronger interactions with water than with themselves. More hydrogen bonds and electrostatic interactions can be formed in unfolded state than in the folded state. This is true because many hydrogen bonding groups can form more than a single hydrogen bond. These groups form multiple hydrogen bonds if exposed to water, but frequently can form only single hydrogen bonds in the folded structure of a protein.

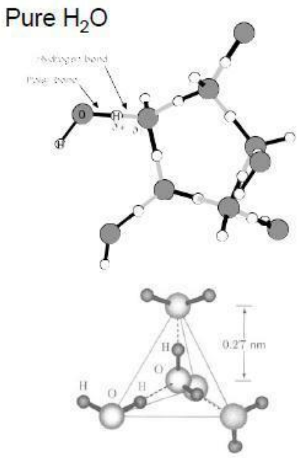
For similar reasons, the $\Delta H_{solvent}$ favors the folded protein because water interacts more strongly with itself than with the polar groups in the protein. More hydrogen bonds can form in the absence of an extended protein, and therefore the number of bonds in the solvent increases when the protein folds.

The sum of the ΔH_{polar} contributions is close to zero, but usually favors the folded structure for the protein slightly. The chain ΔH contributions are positive, while the solvent ΔH contributions are negative. The sum is slightly negative in most cases, and therefore slightly favors folding.

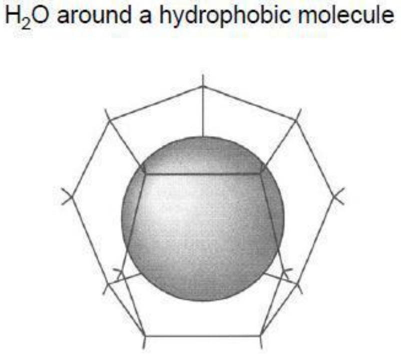
The ΔS_{chain} of the polar groups favors the unfolded state, because the chain is much more disordered in the unfolded state. In contrast, the $\Delta S_{solvent}$ favors the folded state, because the solvent is more disordered with the protein in the folded state. In most cases, the sum of the ΔS_{polar} favors the unfolded state slightly. In other words, the ordering of the chain during the folding process outweighs the other entropic factors.

The ΔG_{polar} that is obtained from the values of ΔH_{polar} and ΔS_{polar} for the polar groups varies somewhat, but usually tends to favor the unfolded protein. In other words, the folding of proteins comprised of polar residues is usually a nonspontaneous process.

Next, we will consider a chain constructed from non-polar groups in aqueous solvent. Once again, the ΔH_{chain} usually favors the unfolded state slightly. Once again, the reason is that the backbone can interact with water in the unfolded state. However, the effect is smaller for non-polar groups, due to the greater number of favorable van der Waals interactions in the folded state. This is a result of the fact that non-polar atoms form better van der Waals contacts with other non-polar groups than with water; in some cases, these effects mean that the ΔH_{chain} for nonpolar residues is slightly negative.



Water molecules have more degrees of freedom as the H-bonds are in the true tetrahedral arrangement.



Water molecules have less degrees of freedom in the clathrate cage arrangements because some H-bonds cannot point inside toward the hydrophobic sphere

As with the polar groups, the $\Delta H_{solvent}$ for non-polar groups favors the folded state. In the case of non-polar residues, $\Delta H_{solvent}$ favors folding more than it does for polar groups, because water interacts much more strongly with itself than it does with non-polar groups.

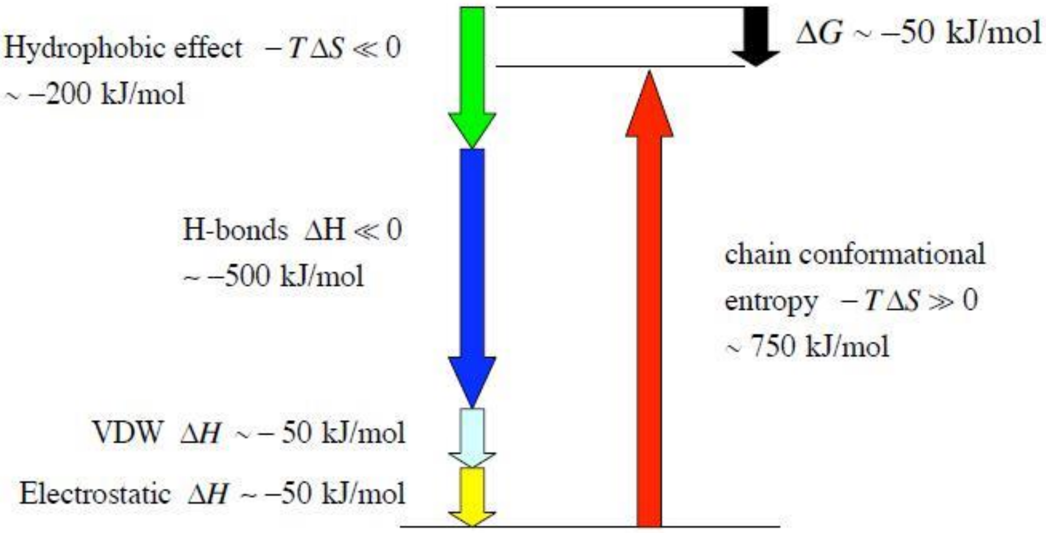
The sum of the $\Delta H_{non\text{-}polar}$ favors folding somewhat. The magnitude of the $\Delta H_{nonpolar}$ is not very large, but is larger than the magnitude of the ΔH_{polar}, which also tends to slightly favor folding.

The ΔS_{chain} of the non-polar groups favors the less ordered unfolded state. However, the $\Delta S_{solvent}$ highly favors the folded state, due to the hydrophobic effect. During the burying of the non-polar side chains, the solvent becomes more disordered. The ΔSsolvent is a major driving force for protein folding which is called conformational entropy.

The $\Delta G_{non\text{-}polar}$ is therefore negative, due largely to the powerful contribution of the $\Delta S_{solvent}$. Adding together the terms for ΔG_{polar} and $\Delta G_{non\text{-}polar}$ gives a slightly negative overall ΔG for protein folding, and therefore, proteins generally fold spontaneously.

$$\text{Unfolded} \overset{\Delta G}{\rightleftharpoons} \text{Folded}$$

$$\Delta G = \Delta H - T\Delta S < 0, \;\; \Delta G =\sim -50 \text{ kJ/mol}$$


Hydrophobic effect $-T\Delta S \ll 0$
~ -200 kJ/mol
H-bonds $\Delta H \ll 0$
~ -500 kJ/mol
VDW $\Delta H \sim -50$ kJ/mol
Electrostatic $\Delta H \sim -50$ kJ/mol
$\Delta G \sim -50$ kJ/mol
chain conformational
entropy $-T\Delta S \gg 0$
~ 750 kJ/mol

Raising the temperature, however, tends to greatly increase the magnitude of the $T\Delta S_{chain}$ term, and therefore to result in unfolding of the protein.

The folded state is the sum of many interactions. Some favor folding, and some favor the unfolded state. The qualitative discussion above did not include the magnitudes of the effects. For real proteins, the various ΔH and ΔS values are difficult to measure accurately. However, for many proteins it is possible to estimate the overall ΔG of folding. Measurements of this value have shown that the overall ΔG for protein folding is very small: only about –10 to –50 kJoules/mol. This corresponds to a few salt bridges or hydrogen bonds.

Studies of protein folding have revealed one other important point: the hydrophobic effect is very important, but it is relatively non-specific. Any hydrophobic group will interact with essentially any other hydrophobic group. While the hydrophobic effect is a major driving force for protein folding, it is the constrains imposed by the more geometrically specific hydrogen bonding and electrostatic interactions in conjunction with the hydrophobic interactions that largely determine the overall folded structure of the protein.

PROTEIN FOLDING MECHANISM

Protein Folding

Protein folding is a process in which a polypeptide folds into a specific, stable, functional, three-dimensional structure. It is the process by which a protein structure assumes its functional shape or conformation Proteins are formed from long chains of amino acids; they exist in an array of different structures which often dictate their functions. Proteins follow energetically favorable pathways to form stable, orderly, structures; this is known as the proteins‘ native structure. Most proteins can only perform their various functions when they are folded. The proteins‘ folding pathway, or mechanism, is the typical sequence of structural changes the protein undergoes in order to reach its native structure. Protein folding takes place in a highly crowded, complex, molecular environment within the cell, and often requires the assistance of molecular chaperones, in order to avoid aggregation or mis folding. Proteins are comprised of amino acids with various types of side chains, which may be hydrophobic, hydrophilic, or electrically charged. The characteristics of these side chains affect what shape the protein will form because they will interact differently intra molecularly and with the surrounding environment, favoring certain conformations nd structures over others. Scientists believe that the instructions for folding a protein are encoded in the sequence. Researchers and scientists can easily determine the sequence of a protein, but have not cracked the code that governs folding.

Protein Folding theory and experiment

Early scientists who studied proteomics and its structure speculated that proteins had templates that resulted in their native conformations. This theory resulted in a search for how proteins fold to attain their complex structure. It is now well known that under physiological conditions, proteins normally spontaneously fold into their native conformations. As a result, a protein's primary structure is valuable since it determines the three-dimensional structure of a protein. Normally, most biological structures do not have the need for external templates to help with their formation and are thus called self-assembling.

Protein Renaturation

Protein renaturation known since the 1930s. However, it was not until 1957 when Christian Anfinsen performed an experiment on bovine pancreatic RNase A that protein renaturation was quantified. RNase A is a single chain protein consisting of 124 residues. In 8M urea solution of 2-mercaptoethanol, the RNase A is completely unfolded and has its four disulfide bonds cleaved through reduction. Through dialysisof urea and introducing the solution to O2 at pH 8, the enzymatically active protein is physically incapable of being recognized from RNase A. As a result, this experiment demonstrated that the protein spontaneously renatured.

One criteria for the renaturation of RNase A is for its four disulfide bonds to reform. The likelihood of one of the eight Cys residues from RNase A reforming a disulfide bond with its native residue compared to the other seven Cys residues is 1/7. Futhermore, the next one of remaining six Cys residues randomly forming the next disulfide bond is 1/5 and etc. As a result, the probability of RNase A reforming four native disulfide links at random is (1/7 * 1/5 * 1/3 * 1/1 = 1/105). The result of this probability demonstrates that forming the disulfide bonds from RNase A is not a random activity.

When RNase A is reoxidized utilizing 8M urea, allowing the disulfide bonds to reform when the polypeptide chain is a random coil, then RNase A will only be around 1 percent enzymatically active after urea is removed. However, by using 2-mercaptoethanol, the protein can be made fully active once again when disulfide bond interchange reactions occur and the protein is back to its native state. The native state of the RNase A is thermodynamically stable under physiological conditions, especially since a more stable protein that is more stable than that of the native state requires a larger activation barrier, and is kinetically inaccessible.By using the enzyme protein disulfide isomerase (PDI), the time it takes for randomized RNase A is minimized to about 2 minutes. This enzyme helps facilitate the disulfide interchange reactions. In order for PDI to be active, its two active site Cys residues needs to be in the -SH form. Furthermore, PDI helps with random cleavage and the reformation of the disulfide bonds of the protein as it attain thermodynamically favorable conformations.

Post translationally Modified Proteins Might Not Renature

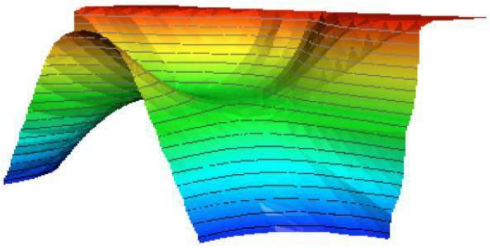
Proteins in a "scrambled" state go through PDI to renature, and their native state does not utilize PDI because native proteins are in their stable conformations. However, proteins that are posttranslationally modified need the disulfide bonds to stabilize their rather unstable native form. One example of this is insulin, a polypeptide hormone. This 51 residue polypeptide has two disulfide bonds that is inactivated by PDI. The following link is an image showing insulin with its two disulfide bonds. Through observation of this phenomena, scientists were able to find that insulin is made from proinsulin, an 84-residue single chain. This link provides more information on the structure of proinsulin and its progression on becoming insulin. The disulfide bonds of proinsulin need to be intact before conversion of becoming insulin through proteolytic excision of its C chain which is an internal 33-residue segment. However according to two findings, the C chain is not what dictates the folding of the A and B chains, but instead holds them together to allow formation of the disulfide bonds. For one, with the right renaturing conditions in place, scrambled insulin can become its native form with a 30% yield. This yield can be increased if the A and B chains are cross-linked. Secondly, through analysis of sequences of proinsulin from many species, mutations are permitted at the C chain eight times more than if it were for A and B chains.

The Protein Folding Process

Considerable evidence suggests that all of the information to describe the three dimensional conformation of a protein is contained within the primary structure. However, for the most part, we cannot fully interpret the information contained within the sequence. To understand why this is true, we need to take a more careful look at proteins and how they fold.

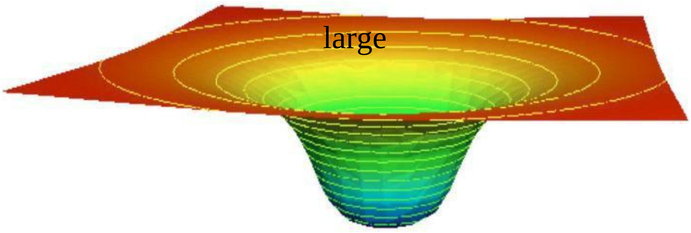
The polypeptide chain for most proteins is quite long. It therefore has ***many*** possible conformations. If you assume that all residues could have 2 possible combinations of ϕ and ψ angles (real peptides can have many more than this), a 100 amino acid peptide could have 2^{100} (~10^{30}) possible conformations. If the polypeptide tested a billion conformations/second, it would still take over 10^{13} years to find the correct conformation. (Note that the universe is only ~10^{10} years old, and that a 100 residue polypeptide is a relatively small protein.) The observation that proteins cannot fold by random tests of all possible conformations is referred to as the Levinthal paradox.

Folding pathways



In classical transition state theory, the reaction diagram for a spontaneous two state system is considered to have a high-energy starting material, a lower energy product, and an energy barrier between them. While the typical diagram that describes the process (such as the one shown at right) is useful, it is incomplete. The process for the conversion of S to P could actually take many pathways; the pathway shown is merely the minimum energy route from one state to another. The true situation is described by an energy landscape, with the minimum energy route being the equivalent of a pass between two mountains. Thus, although the pathway involves an energy barrier, other pathways require passing through even higher energy states.

A large part of the reason that single pathways (or small numbers of pathways) exist for chemical reactions is that most reactions involve the cleavage and reformation of covalent bonds. The energy barrier for breaking a covalent bond is usually quite high. In protein folding, however, the interactions involved are weak. Because the thermal energy of a protein molecule is comparable to the typical non covalent interaction strength, an unfolded polypeptide is present in a large variety of rapidly changing conformations. This realization led to the Levinthal paradox: because the unfolded protein should be constantly changing its shape due to thermal motions of the different parts of the polypeptide, it seemed unlikely that the protein would be able to find the correct state to begin transiting a fixed folding pathway.



An alternate hypothesis has been proposed, in which *portions* of the protein self-organize, followed by folding into the final structure. Because the different parts of the protein begin the folding process independently, the shape of the partially folded protein can be very variable. In this model, the protein folds by a variety of different paths on an energy landscape. The folding energy landscape has the general shape of a **funnel**. In the folding process, as long as the overall process results in progressively lower energies, there can be a large variety of different pathways to the final folded state.

The folding funnel shown above has a smooth surface. Actual folding funnels may be fairly smooth, or may have irregularities in the surface that can act to trap the polypeptide chain in misfolded states. Alternatively, the folding funnel may direct the polypeptide into a ***metastable*** state. Metastable states are local minima in the landscape; if the energy barriers that surround the state are high enough, the metastable state may exist for a long time – metastable states are stable for **kinetic** rather than **thermodynamic** reasons.

The difficulty in refolding many proteins *in vitro* suggests that the folded state of at least some complex proteins may be in a metastable state rather than a global energy minimum.

Folding process

The lower energies observed toward the depression in the folding funnel are thought to be largely due to the collapse of an extended polypeptide due to the hydrophobic effect. In addition to the hydrophobic effect, de solvation of the backbone is necessary for protein folding, at least for portions of the backbone that will become buried. One method for desolvation of the backbone is the formation of secondary structure. This is especially true for helical structures, which can form tightly organized regions of hydrogen bonding while excluding water from the backbone structure. A general outline for the process experienced by a folding protein seems to look like this:

A general outline for the process experienced by a folding protein seems to look like this:

1. Some segments of a polypeptide may rapidly attain a relatively stable, organized structure (largely due to organization of secondary structural Elements).

2. These structures provide nuclei for further folding.

3. During the folding process, the protein is proposed to form a state called a **Molten globule**. This state readily rearranges to allow interactions between different parts of the protein.

4. These nucleated, partially folded domains then coalesce into the folded protein. If this general pathway is correct, it seems likely that at least some of the residues within the sequence of most proteins function to guide the protein into the proper folding pathway, and prevent the ―trapping‖ of the polypeptide in unproductive Partially folded states.

Folding inside cells

Real cells contain **many** proteins at a high overall protein concentration. The protein concentration inside a cell is ~150 mg/ml. folding inside cells differs from most experiments used to study folding *in vitro*:

- Proteins are synthesized on ribosomes. The entire chain is not available to fold at once, as is the case for an experimentally unfolded protein in a test tube.
- Within cells, the optimum ionic concentration, pH, and macromolecule Concentration for each protein to fold properly cannot be controlled as tightly as in an experimental system.
- Major problems could arise if unfolded or partially folded proteins encountered one another. Exposed hydrophobic regions might interact, and form potentially lethal insoluble aggregates within the cell.

One mechanism for limiting problems with folding proteins inside cells volves specialized proteins called **molecular chaperones**, which assist in folding proteins. Molecular chaperones were first observed to be involved in responses to elevated temperature (*i.e.* ―heat shock‖) to stabilize existing proteins and prevent protein aggregation and were called heat-shock proteins (abbreviated as ―hsp‖). Additional research revealed that heat shock proteins are present in all cells, and that they decrease or prevent non-specific protein aggregation and assist in protein folding.

MOLECULAR CHAPERONES

In molecular biology, **molecular chaperones** are proteins that assist the covalent folding or unfolding and the assembly or disassembly of other macromolecular structures. Chaperones are present when the macromolecules perform their normal biological functions and have correctly completed the processes of folding and/or assembly. The chaperones are concerned primarily with protein folding. The first protein to be called a chaperone assists the assembly of nucleosomes from folded histones and DNA and such assembly chaperones, especially in the nucleus, are concerned with the assembly of folded subunits into oligomeric structures.

One major function of chaperones is to prevent both newly synthesised polypeptide chains and assembled subunits from aggregating into nonfunctional structures. It is for this reason that many chaperones, but by no means all, are heat shock proteins because the tendency to aggregate increases as proteins are denatured by stress. In this case, chaperones do not convey any additional stericinformation required for proteins to fold. However, some highly specific 'steric chaperones' do convey unique structural (steric) information onto proteins, which cannot be folded spontaneously. Such proteins violate Anfinsen's dogma.

Various approaches have been applied to study the structure, dynamics and functioning of chaperones. Bulk biochemical measurements have informed us on the protein folding efficiency, and prevention of aggregation when chaperones are present during protein folding. Recent advances in single-molecule analysis have brought insights into structural heterogeneity of chaperones, folding intermediates and affinity of chaperones for unstructured and structured protein chains.

Properties

- Molecular chaperones interact with unfolded or partially folded protein subunits, e.g. nascent chains emerging from the ribosome, or extended chains being translocated across subcellular membranes.
- They stabilize non-native conformation and facilitate correct folding of protein subunits.
- They do not interact with native proteins, nor do they form part of the final folded structures.
- Some chaperones are non-specific, and interact with a wide variety of polypeptide chains, but others are restricted to specific targets.
- They often couple ATP binding/hydrolysis to the folding process.
- Essential for viability, their expression is often increased by cellular stress.

Main role: They prevent inappropriate association or aggregation of exposed hydrophobic surfaces and direct their substrates into productive folding, transport or degradation pathways.

Location and Function

Many chaperones are heat shock proteins, that is, proteins expressed in response to elevated temperatures or other cellular stresses. The reason for this behaviour is thatprotein folding is severely affected by heat and, therefore, some chaperones act to prevent or correct damage caused by misfolding. Other chaperones are involved in folding newly made proteins as they are extruded from the ribosome. Although most newly synthesized proteins can fold in absence of chaperones, a minority strictly requires them for the same.

Some chaperone systems work as foldases: they support the folding of proteins in an ATP-dependent manner (for example, the GroEL/GroES or the DnaK/DnaJ/GrpE system). Other chaperones work as holdases: they bind folding intermediates to prevent their aggregation, for example DnaJ or Hsp33.

Macromolecular crowding may be important in chaperone function. The crowded environment of the cytosol can accelerate the folding process, since a compact folded protein will occupy less volume than an unfolded protein chain. However, crowding can reduce the yield of correctly folded protein by increasing protein aggregation. Crowding may also increase the effectiveness of the chaperone proteins such as GroEL, which could counteract this reduction in folding efficiency.

More information on the various types and mechanisms of a subset of chaperones that encapsulate their folding substrates (e.g. GroES) can be found in the chaperonins. Chaperonins are characterized by a stacked double-ring structure and are found in prokaryotes, in the cytosol of eukaryotes, and in mitochondria.

Other types of chaperones are involved in transport across membranes, for example membranes of the mitochondria and endoplasmic reticulum (ER) in eukaryotes. Bacterial translocation—specific chaperone maintains newly synthesized precursor polypeptide chains in a translocation-competent (generally unfolded) state and guides them to the translocon.

New functions for chaperones continue to be discovered, such as assistance in protein degradation, bacterial adhesin activity, and in responding to diseases linked to protein aggregation (e.g. see prion) and cancer maintenance.

CHEPARONINE

Chaperonins are proteins that provide favourable conditions for the correct folding of other proteins, thus preventing aggregation. Newly made proteins usually must fold from a linear chain of amino acids into a three-dimensional form. Chaperonins belong to a large class of molecules that assist protein folding, called molecular chaperones. The energy to fold proteins is supplied by adenosine triphosphate

GroupI Chaperonins

GroupI Chaperonins are found in bacteria as welas organelles of endosymbiotic origin: chloroplasts and mitochondria. The GroEL/GroES complex in *E. coli* is a Group I chaperonin and the best characterized large (~ 1 MDa) chaperonin complex.

1.GroEL is a double-ring 14mer with a greasy hydrophobic patch at its opening and can accommodate the native folding of substrates 15-60 kDa in size.

2.GroES is a single-ring heptamer that binds to GroEL in the presence of ATP or transition state analogues of ATP hydrolysis, such as ADP-AlF_3. It's like a cover that covers GroEL (box/bottle).

GroEL/GroES may not be able to undo protein aggregates, but kinetically it competes in the pathway of misfolding and aggregation, thereby preventing aggregate formation.

Group II Chaperonins

Group II chaperonins, found in the eukaryotic cytosol and in archaea, are more poorly characterized. TRiC (TCP-1 Ring Complex, also called CCT for chaperonin containing TCP-1), the eukaryotic chaperonin, is composed of two rings of eight different though related subunits, each thought to be represented once per eight-membered ring. TRiC was originally thought to fold only the cytoskeletal proteins actin and tubulin but is now known to fold dozens of substrates.

Mm cpn (Methanococcus maripaludis chaperonin), found in the archaea Methanococcus maripaludis, is composed of sixteen identical subunits (eight per ring). It has been shown to fold the mitochondrial protein rhodanese; however, no natural substrates have yet been identified.

Group II chaperonins are not thought to utilize a GroES-type cofactor to fold their substrates. They instead contain a "built-in" lid that closes in an ATP-dependent manner to encapsulate its substrates, a process that is required for optimal protein folding activity.

Mechanism of action

Chaperonins undergo large conformational changes during a folding reaction as a function of the enzymatic hydrolysis of ATP as well as binding of substrate proteins and cochaperonins, such as GroES. These conformational changes allow the chaperonin to bind an unfolded or misfolded protein, encapsulate that protein within one of the cavities formed by the two rings, and release the protein back into solution. Upon release, the substrate protein will either be folded or will require further rounds of folding, in which case it can again be bound by a chaperonin.

The exact mechanism by which chaperonins facilitate folding of substrate proteins is unknown. According to recent analyses by different experimental techniques, GroEL-bound substrate proteins populate an ensemble of compact and locally expanded states that lack stable tertiary interactions. A number of models of chaperonin action have been proposed, which generally focus on two (not mutually exclusive) roles of chaperonin interior: passive and active. Passive models treat the chaperonin cage as an inert form, exerting influence by reducing the conformational space accessible to a protein substrate or preventing intermolecular interactions e.g. by aggregation prevention. The active chaperonin role is in turn involved with specific chaperonin–substrate interactions that may be coupled to conformational rearrangements of the chaperonin.

Probably the most popular model of the chaperonin active role is the iterative annealing mechanism (IAM), which focus on the effect of iterative, and hydrophobic in nature, binding of the protein substrate to the chaperonin. According to computational simulation studies, the IAM leads to more productive folding by unfolding the substrate from misfolded conformations or by prevention from protein misfolding through changing the folding pathway.

HUMAN CHAPERONE PROTEINS

Chaperones are found in, for example, the endoplasmic reticulum (ER), since protein synthesis often occurs in this area.

Endoplasmic reticulum

In the endoplasmic reticulum (ER) there are general, lectin- and non-classical molecular chaperones helping to fold proteins.

- General chaperones: GRP78/BiP, GRP94, GRP170.
- Lectin chaperones: calnexin and calreticulin
- Non-classical molecular chaperones: HSP47 and ERp29
- Folding chaperones:

  Protein disulfide isomerase (PDI),

  Peptidyl prolyl cis-trans-isomerase (PPI)

  ERp57

Nomenclature and examples of bacterial and archael chaperons.

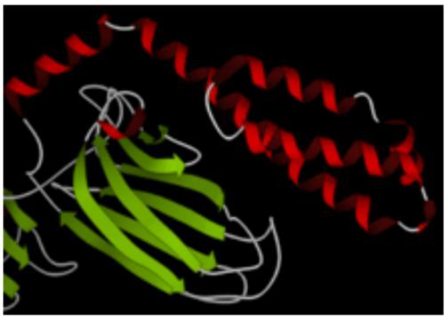
There are many different families of chaperones; each family acts to aid protein folding in a different way. In bacteria like *E. coli*, many of these proteins are highly expressed under conditions of high stress, for example, when the bacterium is placed in high temperatures. For this reason, the term "heat shock protein" has historically been used to name these chaperones. The prefix "Hsp" designates that the protein is a heat shock protein.

Hsp60

Hsp60 (GroEL/GroES complex in *E. coli*) is the best characterized large (~ 1 MDa) chaperone complex. GroEL is a double-ring 14mer with a hydrophobic patch at its opening; it is so large it can accommodate native folding of 54-kDa GFP in its lumen. GroES is a single-ring heptamer that binds to GroEL in the presence of ATP or ADP. GroEL/GroES may not be able to undo previous aggregation, but it does compete in the pathway of misfolding and aggregation.[19] Also acts in mitochondrial matrix as molecular chaperone.

Hsp70

Hsp70 (DnaK in *E. coli*) is perhaps the best characterized small (~ 70 kDa) chaperone.



The Hsp70 proteins are aided by Hsp40 proteins (DnaJ in *E. coli*), which increase the ATP consumption rate and activity of the Hsp70s.

It has been noted that increased expression of Hsp70 proteins in the cell results in a decreased tendency toward apoptosis.Although a precise mechanistic understanding has yet to be determined, it is known that Hsp70s have a high-affinity bound state to unfolded proteins when bound to ADP, and a low-affinity state when bound to ATP. It is thought that many Hsp70s crowd around an unfolded substrate, stabilizing it and preventing aggregation until the unfolded molecule folds properly, at which time the Hsp70s lose affinity for the molecule and diffuse away. Hsp70 also acts as a mitochondrial and chloroplastic molecular chaperone in eukaryotes.

Hsp90

Hsp90 (HtpG in *E. coli*) may be the least understood chaperone. Its molecular weight is about 90 kDa, and it is necessary for viability in eukaryotes (possibly for prokaryotes as well).Heat shock protein 90 (Hsp90) is a molecular chaperone essential for activating many signaling proteins in the eukaryotic cell.Each Hsp90 has an ATP-binding domain, a middle domain, and a dimerization domain.

Hsp100

Hsp100 (Clp family in *E. coli*) proteins have been studied *in vivo* and *in vitro* for their ability to target and unfold tagged and mis folded proteins. Proteins in the Hsp100/Clp family form large hexameric structures with unfoldase activity in the presence of ATP. These proteins are thought to function as chaperones by processively threading client proteins through a small 20 Å (2 nm) pore, thereby giving each client protein a second chance to fold. Some of these Hsp100 chaperones, like ClpA and ClpX, associate with the double-ringed tetradecameric serine protease ClpP; instead of catalyzing the refolding of client proteins, these complexes are responsible for the targeted destruction of tagged and misfolded proteins. Hsp104, the Hsp100 of Saccharomyces cerevisiae, is essential for the propagation of many yeast prions. Deletion of the HSP104 gene results in cells that are unable to propagate certain prions.