LECTURE 05: 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.

PROTEIN ENGINEERING

There Are Four Levels of Architecture in Proteins

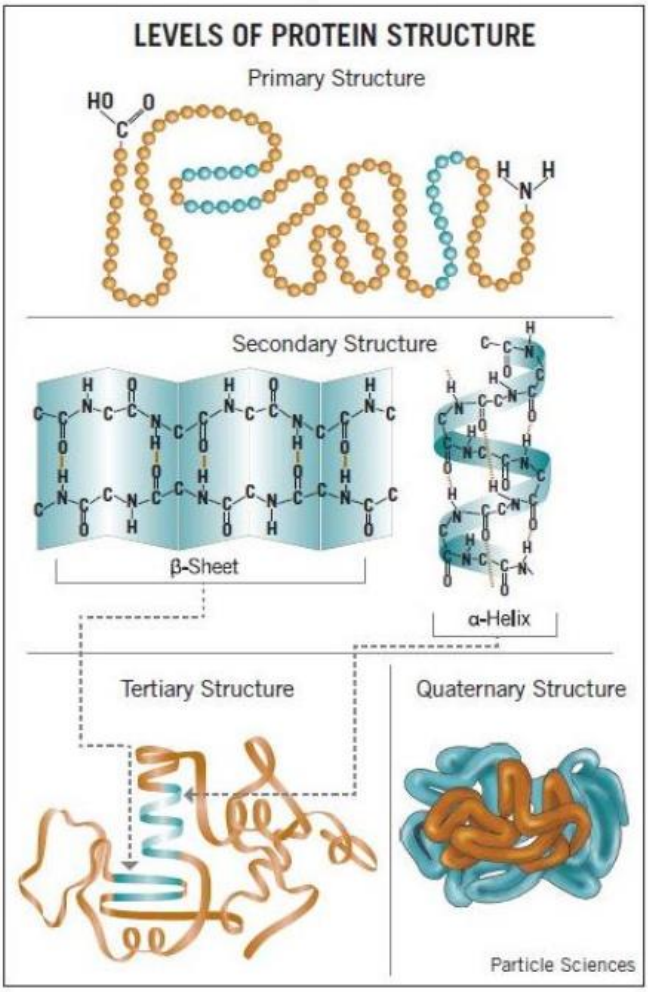


Figure 1 Levels of structure in proteins

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.

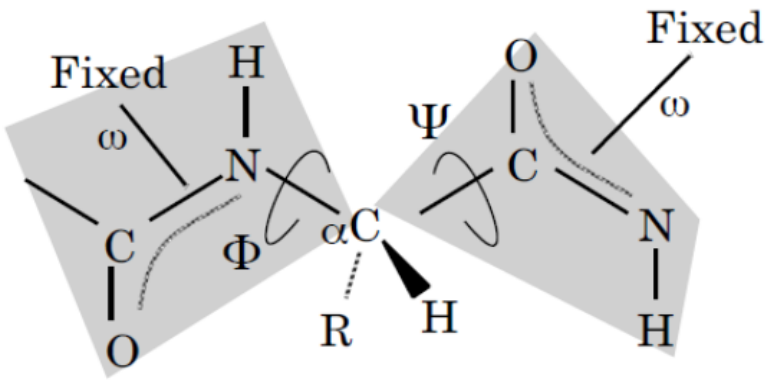
1.1.Protein Secondary Structure

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.

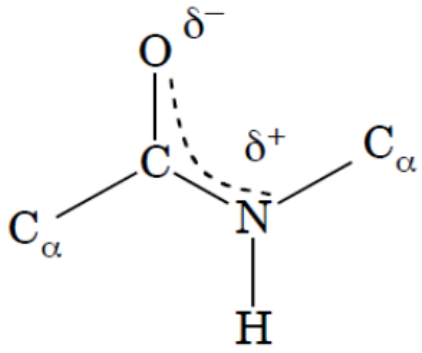
In considering secondary structure, it is useful to classify proteins into two major groups: fibrous proteins, having polypeptide chains arranged in long strands or sheets, and globular proteins, with polypeptide chains folded into a spherical or globular shape. Fibrous proteins play important structural 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.

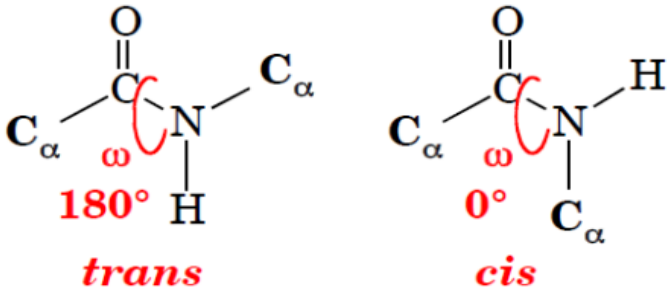


This allows the peptide bond to participate in electrostatic interactions, and contributes to the hydrogen bond strength between the backbone carbonyl and the Namide proton.

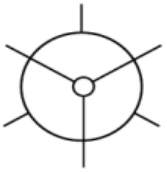
Peptide bond and protein structure

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°.

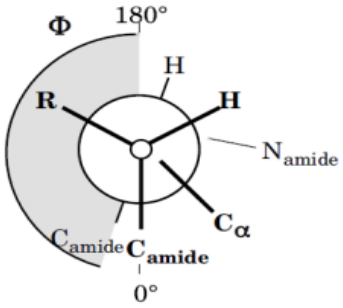
Observed ω angles for the peptide bond

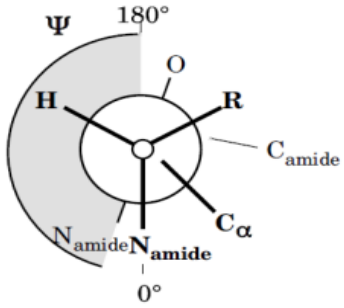


In considering peptide structures, it is usually much more important to look at the backbone angles that can vary more widely. These angles are the Φ (= phi, Cα–Namide) and Ψ (= psi, Cα–Camide) angles. By definition, the fully extended conformation corresponds to 180° for both Φ and Ψ. (Note that 180°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



Tetrahedral Carbon–Carbon bond





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.

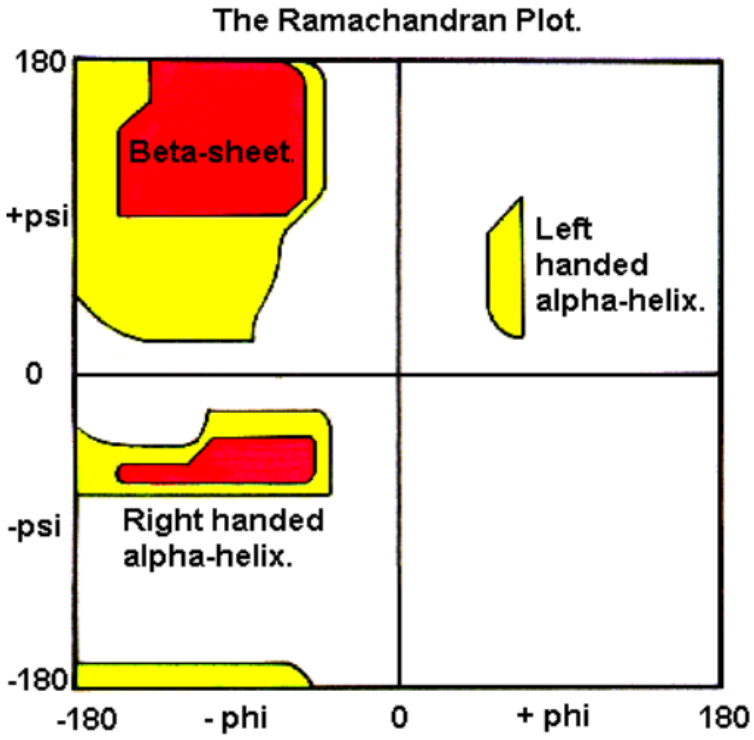
A Ψ angle of 180° results in an alignment of the Namide with the carbonyl oxygen from the same residue. This is allowed, although not especially favored. A Ψ angle of 0° places the

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.

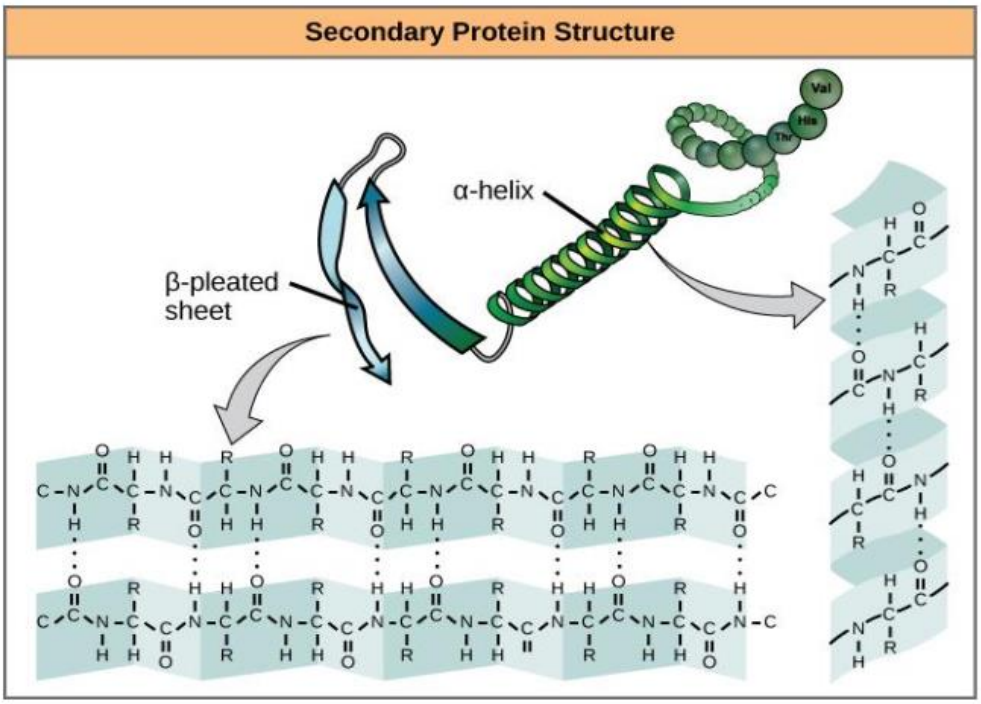
L-amino acids cannot form extended regions of left-handed helix but occassionally individual residues adopt this conformation. These residues are usually glycine but can also be asparagine or aspartate where the side chain forms a hydrogen bond with the main chain and therefore stabilises this otherwise 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.

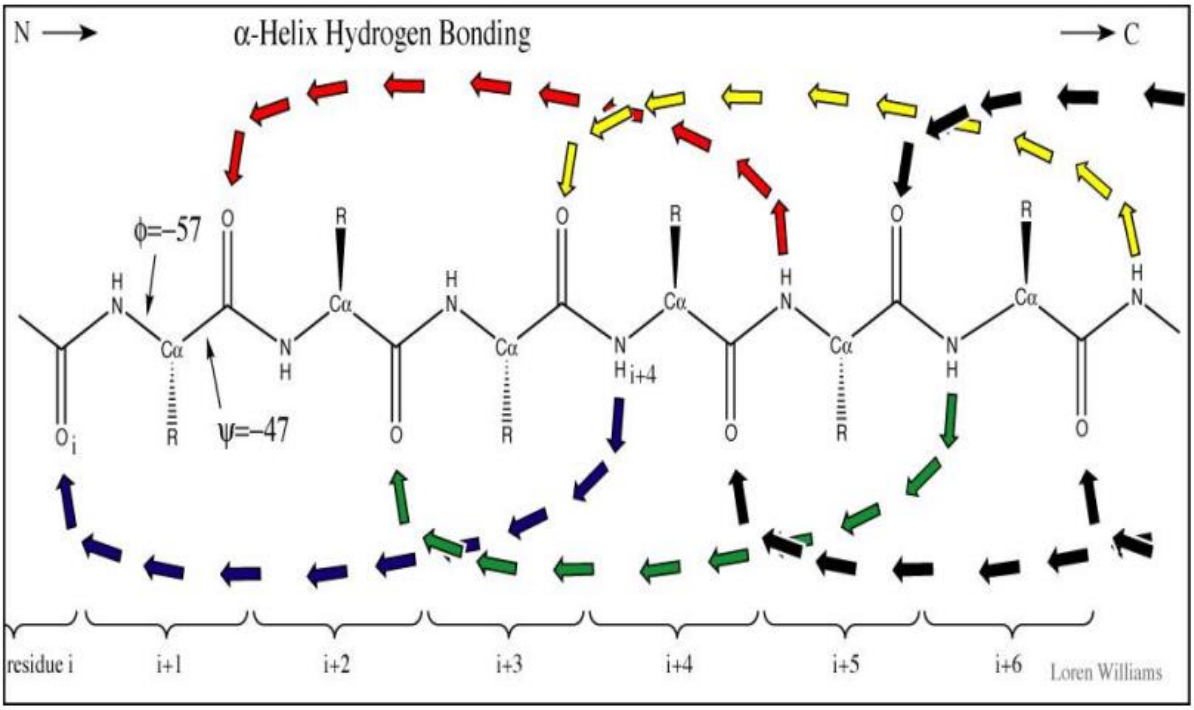
Secondary structure

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.



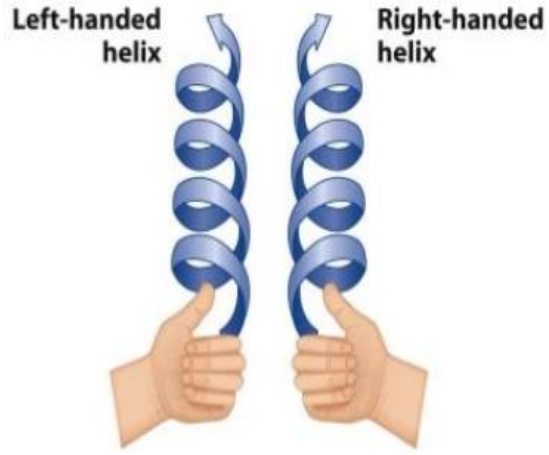
Alpha helix (α-helix)

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.

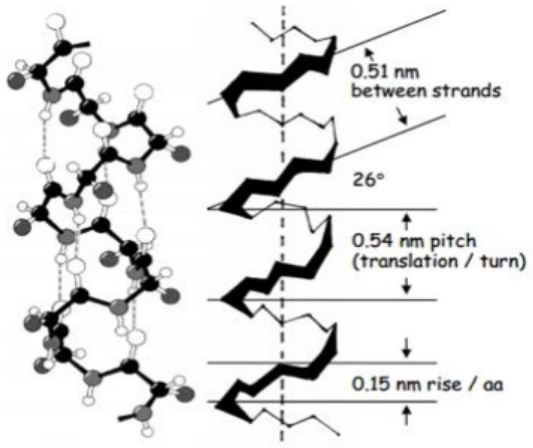


PROPERTIES

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.

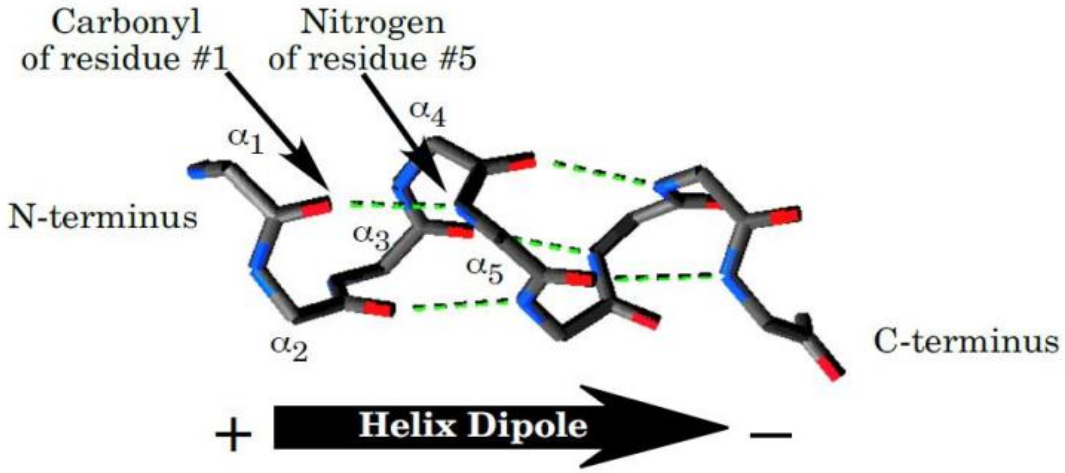
Similar structures include the 3_{10} helix ($i + 3 \rightarrow i$ hydrogen bonding) and the π-helix ($i + 5 \rightarrow i$ hydrogen bonding). The α helix can be described as a 3.6_{13} helix, since the i + 4 spacing adds 3 more atoms to the H-bonded loop compared to the tighter 3_{10} helix, and on 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

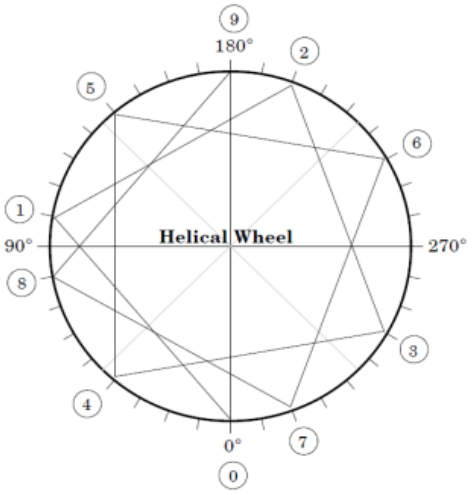
Drawing a three-dimensional helix on paper is difficult. Two types of two dimensional representations (helical wheel and helical net diagrams) are commonly used to simplify the analysis of helical segments of proteins. The two-dimensional representations are somewhat stylized, but show the major features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

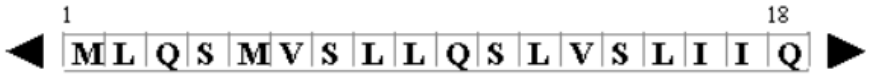
Residue #0 = 0° (by definition)
#1 = 100°
#2 = 200°
#3 = 300°
#4 = 400° = 40°
#5 = 140°
#6 = 240°
#7 = 340°
#8 = 440° = 80°
9# = 900° (from first) = 180°
These angles can be plotted on a circle.

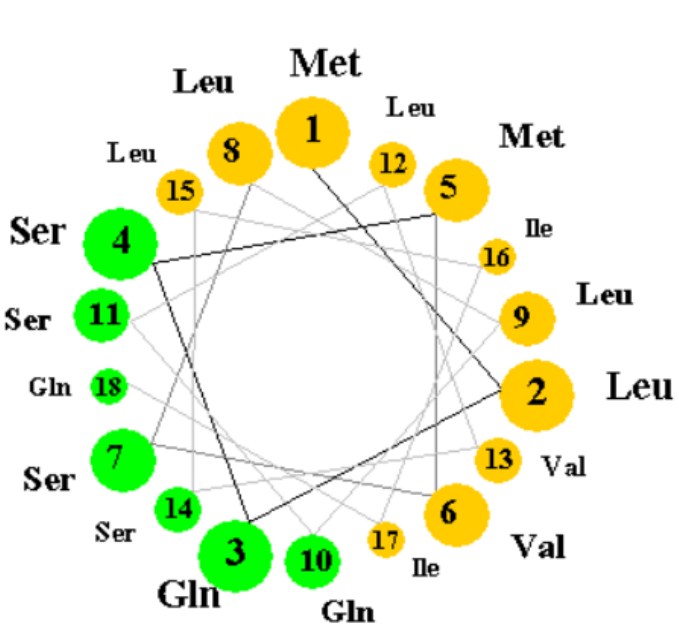
Doing so results in a representation that corresponds to the view looking down the long axis of the helix. (**Note that the rotation is clockwise as the residue number increases**.)

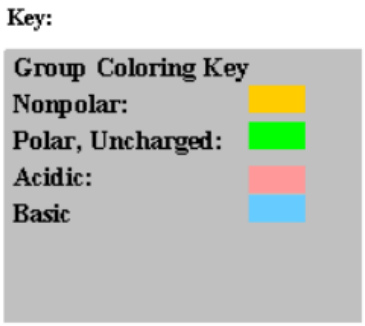


Note that residues 0, 3, 4, 7, and 8 are all located on one face of the helix

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.)

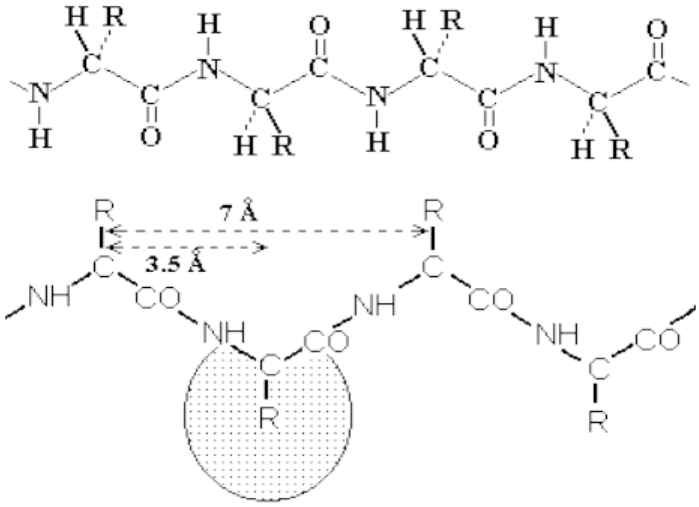




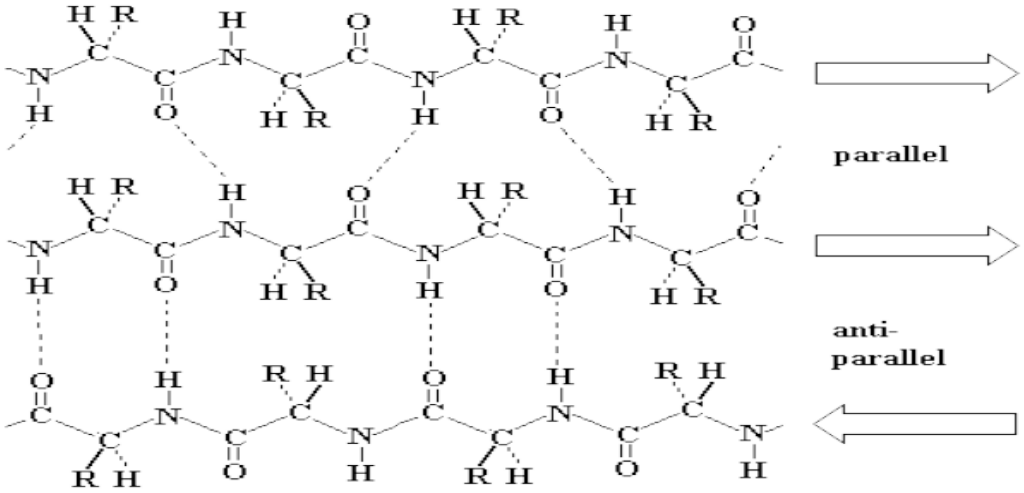


PROTEIN ENGINEERING

The ββ 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° and psi = +135°; 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 90º 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°;	**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.

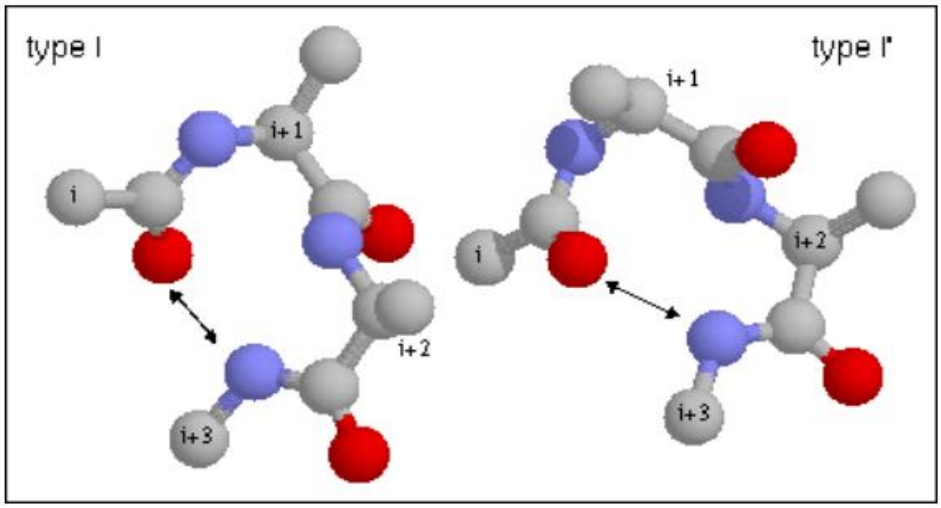
He found three general types, one of which

(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.