

Introduction to Protein Structure

Proteins are large heteropolymers usually comprised of 50 – 2500 monomer units, although larger proteins are observed⁷. The monomer units of proteins are amino acids. The **primary structure** is the sequence of monomer units in the protein. **Secondary structure** is made up from regular structures within proteins. The **tertiary structure** is the overall three-dimensional structure of the protein. **Quaternary structure** is the organization of multisubunit proteins (proteins comprised of single polypeptide chains do not have quaternary structure).

A **peptide** is a molecule containing two or more amino acid residues linked by peptide bonds. The term peptide typically refers to molecules comprised of relatively small numbers of amino acid residues. This is sometimes made explicit in the term **oligopeptide**: “oligo-” means “short”, and thus an oligo peptide is a short polymer of amino acids.

A **polypeptide** is a longer polymer of amino acids. The term “**protein**” is often used interchangeably with “polypeptide”. In some cases, however, the term “polypeptide” is used to refer to a single chain peptide part of a molecule that either also contains non-amino acid groups as part of the overall “protein” or which is not folded into a defined structure.

You may have noticed the term amino acid **residue** in the preceding paragraphs. In a strict sense, proteins do not contain actual amino acids; as will be seen, the amino and carboxylic acid functional groups of amino acids incorporated into proteins are transformed into amide linkages, which lack the ionizable groups. Therefore the monomer units within proteins are properly referred to as “amino acid residues” or simply “residues”.

Primary Structure

The primary structure of a protein is the sequence of amino acids that comprises the protein. The three-dimensional structure of a protein (and therefore, all of its functions) is dependent on the primary structure.

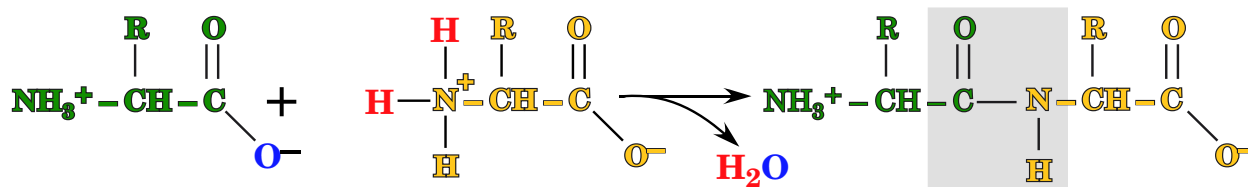
The primary structure is always written from the amino-terminus (the residue with the free α -amino group) to the carboxy-terminus. This convention has several reasons, including the fact that the protein biosynthetic machinery produces proteins from N-terminus to C-terminus. Thus, anyone seeing a peptide written as GRSKKNS (or the equivalent sequence: Gly-Arg-Ser-Lys-Lys-Asn-Ser) immediately knows that this peptide has an N-terminal glycine and a C-terminal serine.

The peptide bond

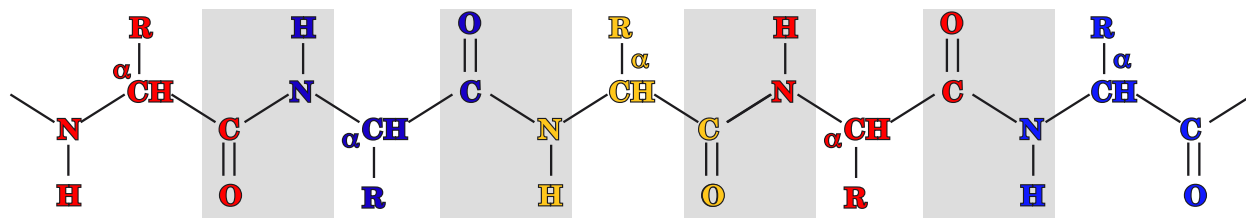
Peptides and proteins are **linear** polymers of amino acids. The amino acids are

⁷ The largest known polypeptide, the full length version of the mouse muscle protein called titin, has 35,213 amino acid residues (the human version of titin is smaller, with “only” 34,350 residues in the full length protein).

connected by an amide linkage called a peptide bond. In principle, the peptide bond is formed by removing an H₂O molecule while linking the α-amino group of one amino acid to the α-carboxylic acid of another.



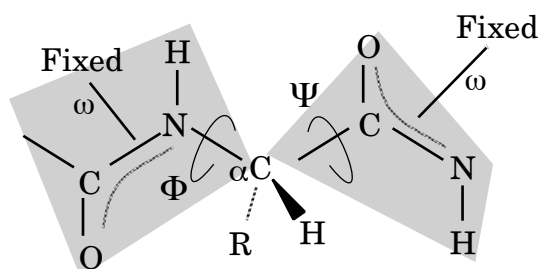
Proteins are polymers of amino acids linked by these peptide bonds:



Structural Features of the Peptide Bond

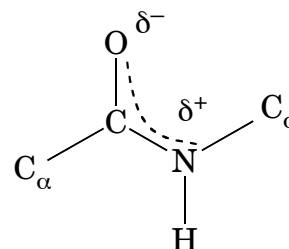
In a peptide bond, the amide oxygen and amide proton in the bond are almost invariably in the *trans* configuration (as shown below). This decreases steric conflicts between these atoms, and between the α-carbons and their substituents.

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**.

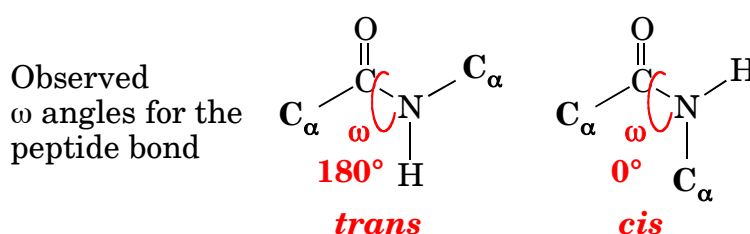
Finally, the peptide bond has a dipole, with the O having a partial negative charge, and the N_{amide} having a partial



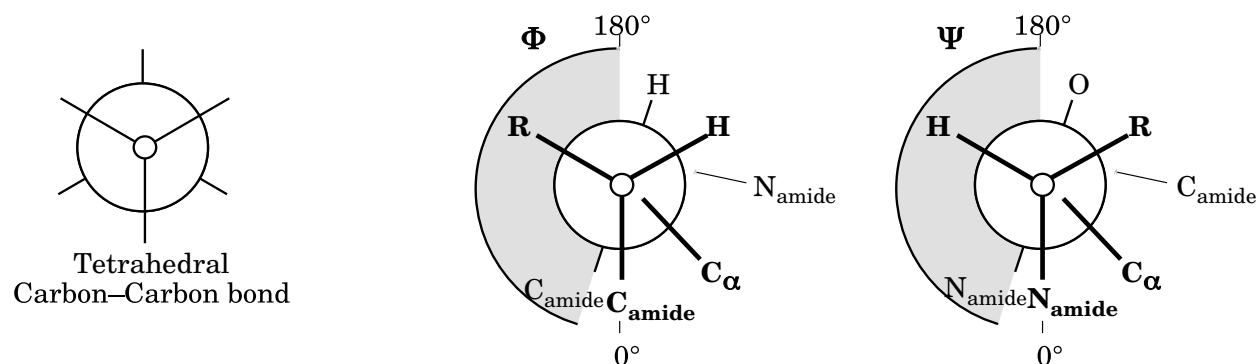
positive charge. This allows the peptide bond to participate in electrostatic interactions, and contributes to the hydrogen bond strength between the backbone carbonyl and the N_{amide} proton.

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 N_{amide} 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.⁸ 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, C_{α} - N_{amide}) and Ψ (= psi, C_{α} - C_{amide}) angles. By definition, the fully extended conformation corresponds to 180° for both Φ and Ψ . (Note that $180^\circ = -180^\circ$). Numeric values of angles increase in the clockwise direction when looking away from the α -carbon.



By definition, $\Phi = 0^\circ$ when the $C_{\text{amide}}-N_{\text{amide}}$ and $C_{\text{amide}}-C_{\alpha}$ bonds are in the same plane, and $\Psi = 0^\circ$ when the $N_{\text{amide}}-C_{\text{amide}}$ and $N_{\text{amide}}-C_{\alpha}$ bonds are in the same plane. The (+) direction is clockwise while looking away from the C_{α} .

⁸ Proline is an exception; the tertiary amide formed by an X-Pro peptide bond does not have strong energetic preference for *cis* or *trans* configurations. As a result, both are observed in proteins. Interconversion between *cis* and *trans* X-Pro peptide bonds requires an enzyme (*proline isomerase*), which plays a role in folding of many proteins.

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: $0^\circ:0^\circ$, $180^\circ:0^\circ$, and $0^\circ:180^\circ$ 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 N_{amide} with the carbonyl oxygen from the same residue. This is allowed, although not especially favored. A Ψ angle of 0° places the N_{amide} from one residue very close to the N_{amide} from the previous residue; this results in a steric clash (as well as an unfavorable electrostatic interaction, because both N_{amide} have partial positive charges).

The residue side-chains also impose steric constraints. Glycine, because of its small side chain, has a much larger range 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 N_{amide} . 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.

Secondary Structure (α -helix and β -sheet)

Secondary structural elements are repeated regular structures within proteins. In these repeated structures, several successive residues exhibit the same Φ / Ψ pairs. The two most common types of secondary structural elements are α -helices and β -sheets.

The α -helix

Examining an α -helix is easiest with a molecular model of a short polypeptide. With a model: start at top with amino terminus, and rotate the chain clockwise as you move downward. The carbonyl oxygens point downward. The carbonyl oxygen from α -carbon #1 forms an H-bond to the amide H-N that follows three intervening carbonyl oxygens; this is the amide proton from the 5th residue. More generally, the **carbonyl_i accepts an H-bond from H-N_{i+4}**.

In looking at a model, removing the side chains after the β -carbon (*i.e.* constructing polyalanine) makes the demonstration easier, because the extra atoms may be heavy enough to distort chain in gravity field (which is not a problem with real proteins). As it happens, alanine is frequently found in α -helices.

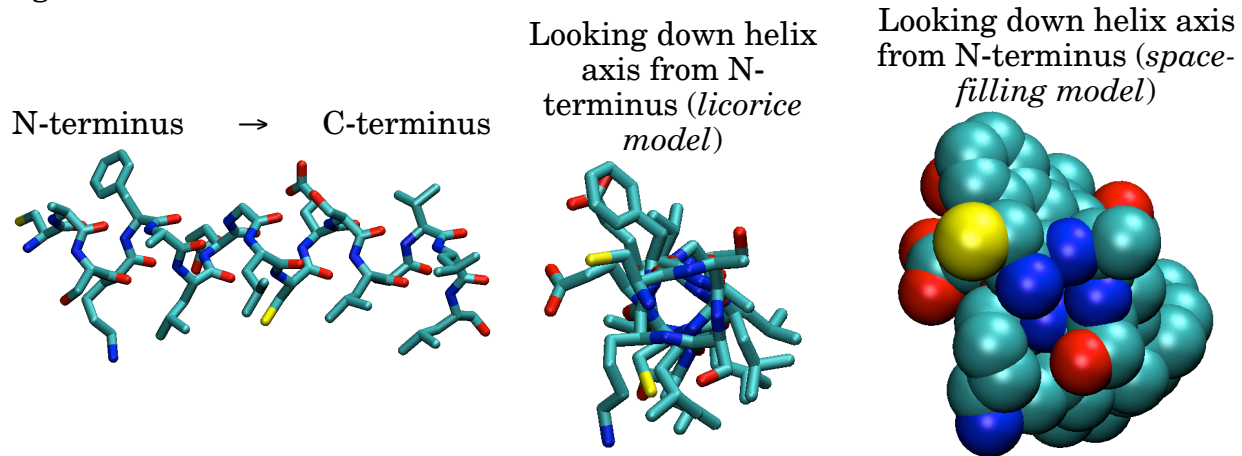
An α -helix is a **right-handed** helix: looking end on, the rotation is clockwise while moving away from observer. This is true looking at the helix from either N-terminal or C-terminal end.

Note that the first 4 nitrogens (counting the one from the initial residue that contains the $i=1$ carbonyl) cannot form H-bonds within the helix.

Properties of an α -helix

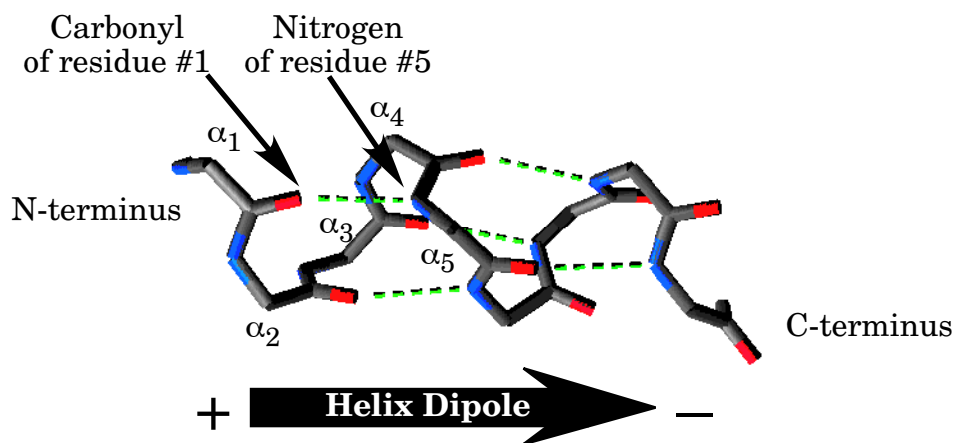
1) The α -helix satisfies the backbone H-bonds (except for the first 4 N_{amide} , and last 4 carbonyl groups).

2) There is no “hole” in the middle of the helix: there is no room. If you look down the axis of a space-filling model of a helix, you will not see any empty space. The figures below show the same α -helix.



3) The **side chains extend outwards** and angled slightly ($\sim 30^\circ$) toward the N-terminus of the helix.

4) 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.



5) An α -helix has **3.6 residues/turn**. Since a circle contains 360° , each residue is rotated by $360/3.6 = 100^\circ$.

6) An α -helix has a **5.4 Å pitch/turn**. This means that as the chain moves along the axis of the helix, the helix becomes 5.4 Å longer for each full turn. This corresponds to **1.5 Å/residue**. (A C–C single bond length is ~ 1.5 Å; since each residue represents

two C–C bonds and a C–N bond, 1.5 Å per residue is a very compact structure.)

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

8) The backbone bond angles are $\Phi = -57^\circ$ and $\Psi = -47^\circ$. Note that the Φ and Ψ angles in real helices are not exact, and may vary by up to 10° from these ideal values and occasionally vary by even larger amounts.

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.⁹

Helical Wheel

Residue #0 = 0° (by definition)

#1 = 100°

#2 = 200°

#3 = 300°

#4 = $400^\circ = 40^\circ$

#5 = 140°

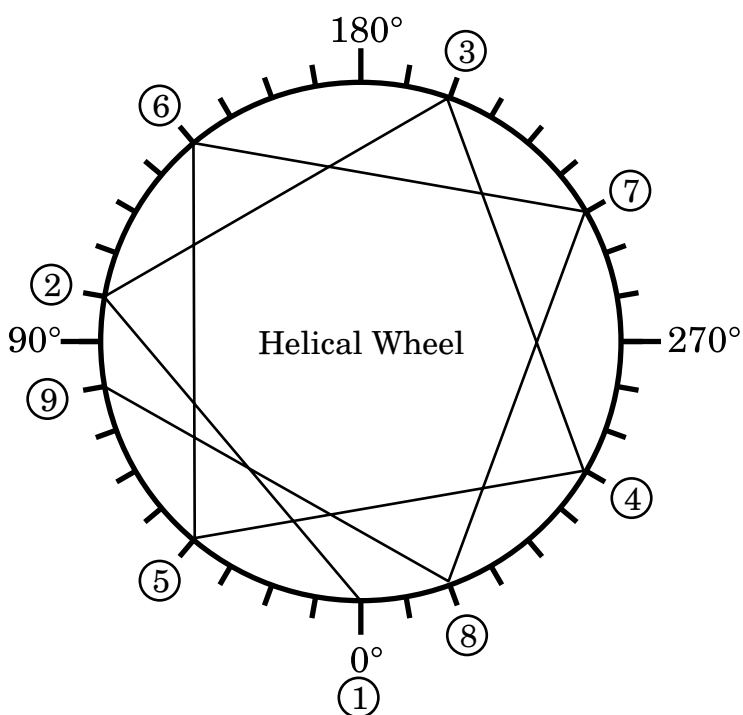
#6 = 240°

#7 = 340°

#8 = $440^\circ = 80^\circ$

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.**)



⁹ Numbering the helix beginning with residue #0 simplifies some calculations. Note that distances along helices are measured from the first residue, but, in effect, do not count the first residue. Thus, the length of a four-residue helix is 4.5 Å, not 6 Å.

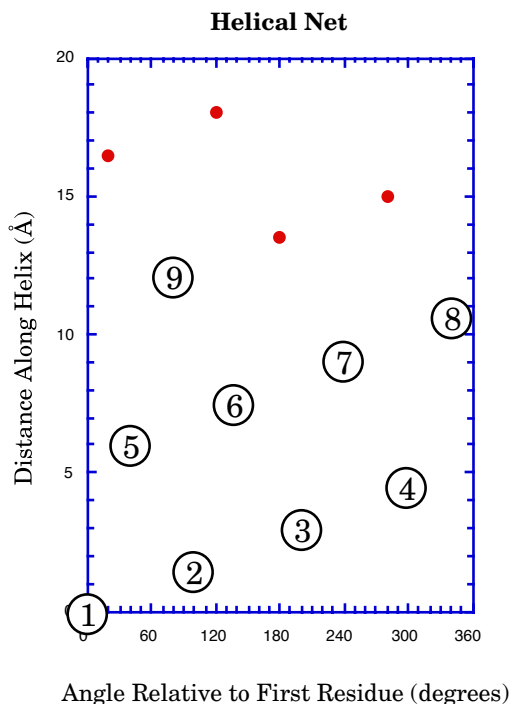
Note that residues 0, 3, 4, 7, and 8 are all located on one face of the helix.

Helical Net

A separate representation involves a hypothetical razor cut lengthwise along the helix axis, followed by unpeeling the surface of the helix. This results in a plot of distance along the helix *versus* distance around the helix called a **Helical Net**.

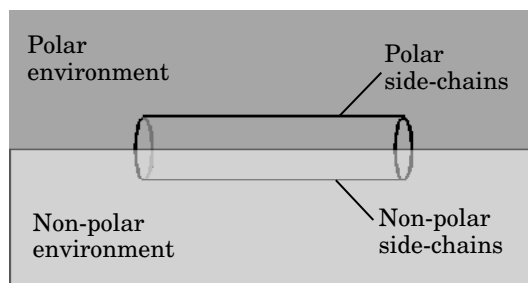
While it is sometimes more difficult to see the helix “faces” in a helical net (where vertical bands within net are on same face of helix), the helical net does emphasize the distance **along** the helix, which is ignored in the helical wheel. Compare the two representations, and note the relative positions of different residues.

In the helical wheel, residues 1 and 8 are quite close together; the helical net shows that these residues exhibit a 10.5 Å vertical separation (although they are still present on the same face of the helix).



Both helical wheel and helical net analysis types are useful for analyzing the helix for patterns in residue properties around the circumference of the helix. The helical wheel is somewhat easier to interpret, but leaves out the sometimes-crucial distance information.

A common reason for using the helical wheel or helical net is to look for helical regions that exist at protein surfaces. The interior of a protein is typically non-polar, while the surrounding solvent is polar. 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.) Note, however, that residues 1 and 8 are much closer in position around the helix than are residues 1 and 2. A simple examination of the linear sequence does not reveal amphipathic tendencies; in contrast, helical net, and especially helical wheel diagrams, show amphipathic helical character clearly.



The 3-10 helix

In some proteins, the backbone forms variant on the α -helix instead. The 3-10 helix (for “3 residues/turn, 10 backbone atoms/turn”) has a pitch of 6 Å (2 Å/residue). The residues in a 3-10 helix have slightly different Φ and Ψ angles to yield a helix that is less tightly wound and more extended than an α -helix.

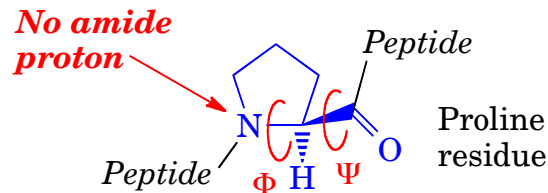
Feature	α -helix	3-10 Helix
Φ	-57°	-49°
Ψ	-47°	-26°
Rise per residue	1.5 Å	2Å

Amino acids in helices

Most amino acids are observed to be present in helices within proteins. The side-chains of helical residues can interact, at least to some degree, but with rare exceptions (see below) the side-chains do not participate in the hydrogen-bonding pattern of the helical backbone.

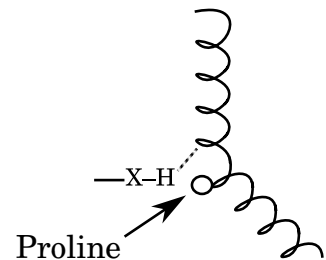
There are a few amino acids that are *relatively* rare in helical structures.

- 1) While proline has limited conformational flexibility, it is capable of forming Φ angles close to -57° . However, **proline lacks an amide proton**, and therefore cannot participate in helix H-bonds. Proline is often found at end of a helix (especially at the N-terminus, when the lack of the amide proton is not a problem, but also at the C-terminus, where it helps terminate the helix.



- 2) **Glycine** has a very small side-chain, and therefore has a **large conformational flexibility**. This means that the ordered structure of helix is **entropically disfavored** for glycine. Glycine is capable of forming the appropriate Φ / Ψ angles for an α -helix, but tends to allow too much flexibility to maintain the stable helical structure.

Proline and glycine are sometimes referred to as “helix breakers” because they tend to have destabilizing effects on helices. Proline and glycine also allow “kinks” in a helix, where the direction of the helix axis changes. Note the hydrogen bond donor in the drawing. This hydrogen bond donor from outside the helix is necessary to satisfy the hydrogen bonding requirements of the P_{i-4} residue (*i.e.* the position four residues closer to the N-terminus than the proline). The P_{i-4} residue has a backbone carbonyl that



cannot hydrogen bond to the proline (as shown above, proline lacks an amide proton). The kink in the helix is a consequence of the missing H-bond from the proline and the likely distortion of the helix to allow room for a hydrogen bond donor from outside the helix.

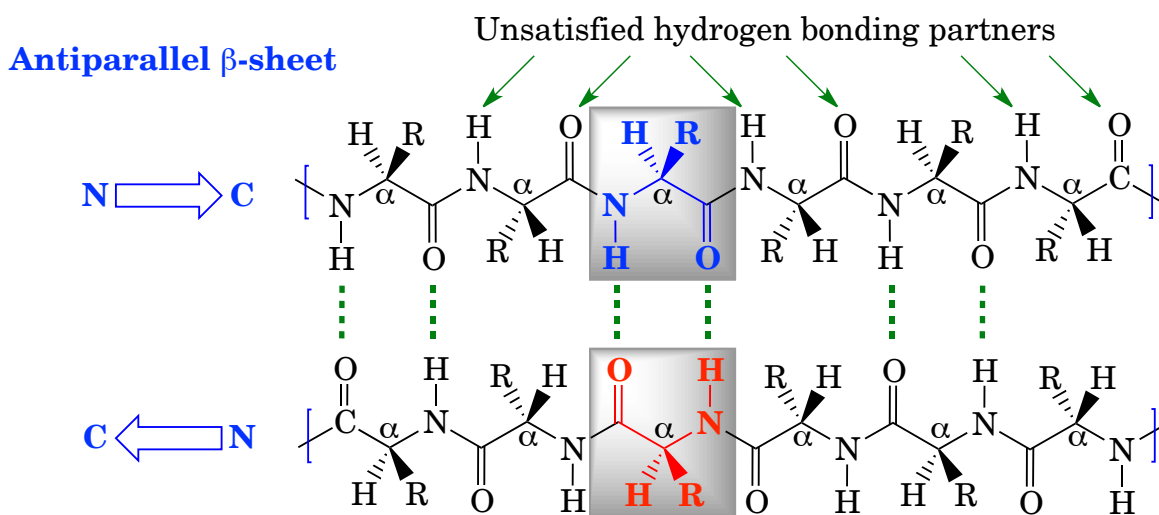
- 3) Charged residues can either stabilize or destabilize a helical structure. For example, glutamate residues usually favor helix formation. However, too many glutamates in the absence of positively charged residues can result in sufficient electrostatic repulsion to disrupt helix formation.

The β -sheet

A second widely observed method for satisfying backbone atom hydrogen bonding requirements is called a β -sheet. A β -sheet is a much more extended form of secondary structure than an α -helix. A β -sheet has the **backbone atoms in an essentially planar structure**, with the **side chains alternately pointing above and below** the sheet plane.

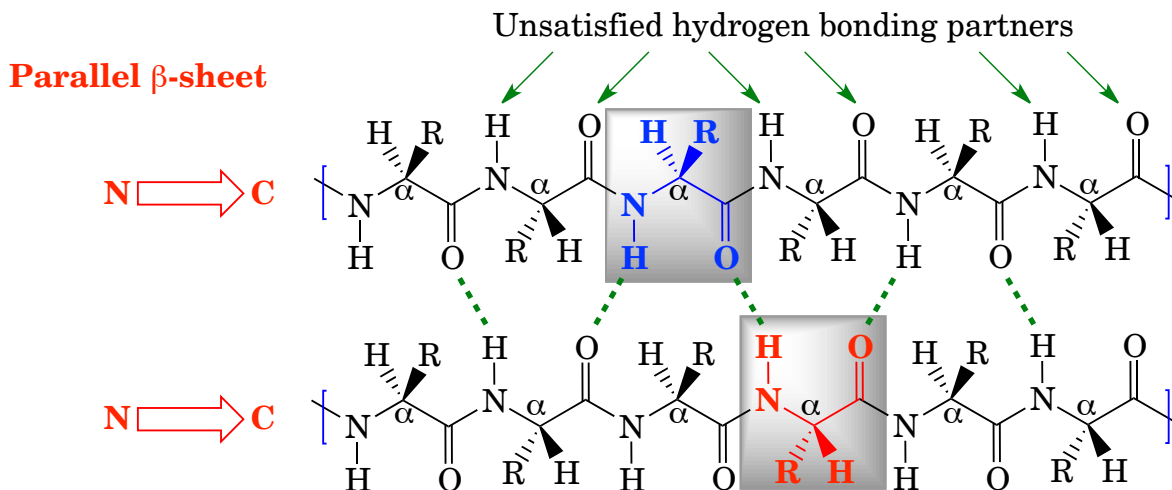
An **antiparallel β -sheet** is a secondary structural element comprised of two (or more) strands leading in **opposite** directions. In an antiparallel β -sheet, the backbone amide proton and carbonyl from one residue form hydrogen bonds with the corresponding groups from a single residue in another strand.

Note the unsatisfied hydrogen bonds on each edge of the two-stranded sheet shown in the figure below. These could be satisfied by the presence of additional β -sheet strands (which allows β -sheets to form rather large structures). However, half of the backbone hydrogen bonds at the edge of any planar β -sheet must be formed using groups from outside the secondary structure.

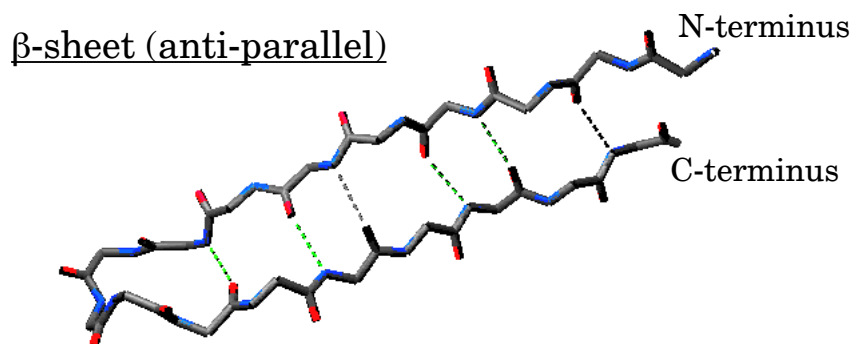


A **parallel β -sheet** is comprised of two (or more) strands leading in the **same** direction. Note that a residue from one strand forms hydrogen bonds to groups from

two residues of the other strand. Note also that the hydrogen bonds are somewhat distorted (this distortion is somewhat exaggerated in the diagram). Parallel β -sheets tend to be less stable than their antiparallel counterparts, and are less commonly observed. Most parallel β -sheets are formed from multiple strands.



Note also that forming a parallel β -sheet requires strands from different parts of the protein (an antiparallel sheet can form from a single sequence assuming a turn in the middle (as shown in the diagram below)).



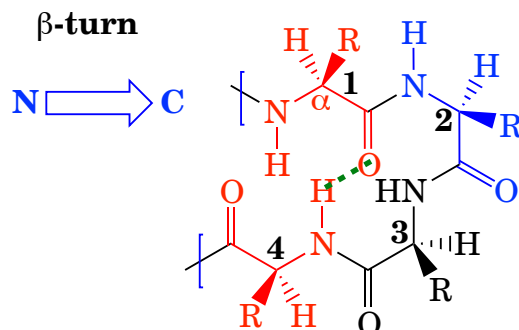
The Φ and Ψ angles required to form a β -sheet are quite different from those used for forming an α -helix. These bond angles result in a much more extended structure (the rise/residue for a β -sheet is about 3.5Å).

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

Both parallel and antiparallel β -sheets put the residue side-chains on alternating sides of H-bond/backbone plane. Generating a β -sheet with specific properties requires taking this alternating side-chain pattern into account.

β -turn

A third type of secondary structural element observed frequently in proteins is a β -turn (also called β -bend, or tight turn, or merely turn). A β -turn is a motif that allows a rapid reversal of the peptide strand direction.



In a β -turn, the carbonyl oxygen of residue #1 H-bonds to the amide proton of residue #4. In these structures, residue #2 is often glycine (because the glycine side-chain is small enough to avoid significant steric hindrance at the angles required for the structure). Residue #3 is often proline (which readily assumes the required Φ and Ψ angles, and which lacks the amide proton that can form a steric clash with residue #1).

Other types of secondary structure

Some other regular, repeating sets of Φ and Ψ angles are observed in a few types of proteins. Some fibrous proteins such as α -keratin have α -helices wrapped about one another in a structure called a coiled-coil.

An even more prevalent protein is a structural protein called collagen; collagen is comprised of three intertwined chains arranged in a helix. The collagen triple helix residues exhibit Φ and Ψ angles of -51° and 153° , respectively. Most animals have large amounts of a few types of collagen.

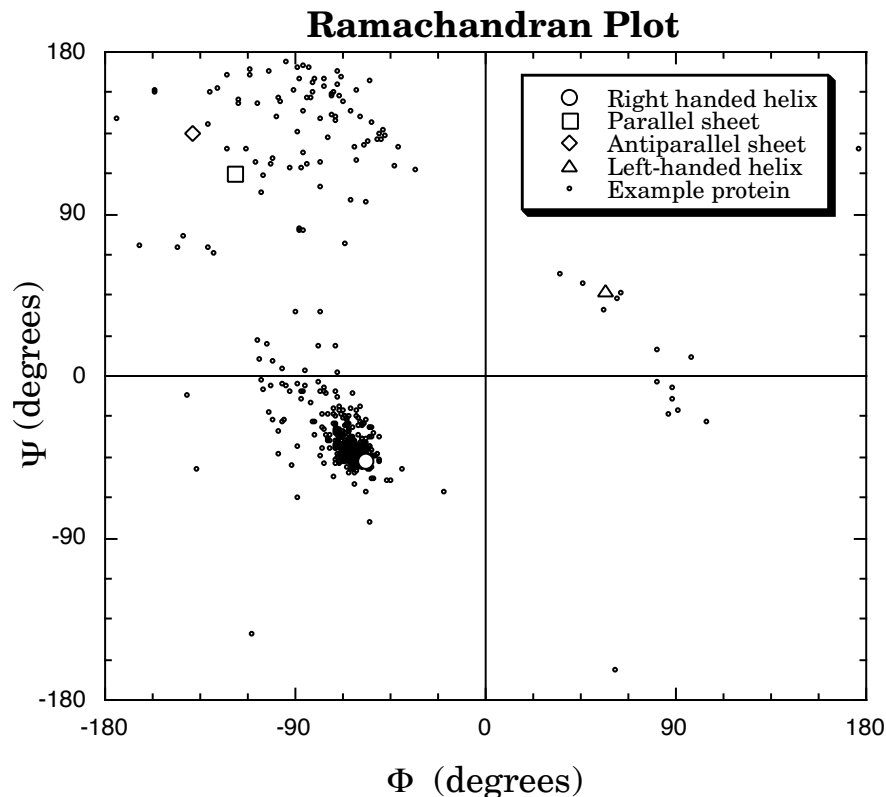
Secondary structure	Φ	Ψ	Rise per residue
α -helix	-57°	-47°	1.5 Å (3.6 residues per turn)
Coiled-coil	-64°	-42°	1.5 Å
β -sheet (antiparallel)	-139°	135°	3.5 Å
β -sheet (parallel)	-119°	113°	3.5 Å
Collagen	-51°	153°	3 Å

A few proteins have a left-handed helix. This is a much less stable structure than the standard α helix.

Finally, 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.

Ramachandran diagram

A Ramachandran diagram is a plot of Ψ versus Φ angles for each residue in a protein. The diagram shows several features. One obvious feature is the limited Φ / Ψ angle pairs present; many of the pairs are sterically impossible, or so strained as to be extremely rare.



Another obvious feature is the presence of many points in small regions of the diagram. These correspond to many residues with very similar Φ / Ψ angle pairs, and are usually indications that secondary structural elements are present. The example protein shown contains large amounts of α -helix; the region near $\Phi = -57^\circ$; $\Psi = -47^\circ$ is therefore heavily populated. Note, however, that all of the residues do not have identical angles (hence the scatter in this region). This is a general phenomenon: few residues have the exact angles of a perfect secondary structural element.

A Ramachandran diagram can be generated for any protein for which structural data are available. The collagen triple helix, the α -helix, β -sheets, and less regular structures all have specific locations on the diagram.