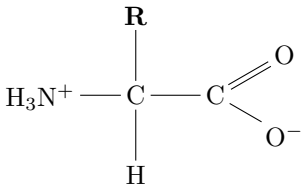


Figure 1: A generic amino acid. The "R" group (side chain) is a variable. It is the only difference between the 20 common amino acids. The rest of the amino acid is called the backbone. This form is called a zwitterion, because it has both positive and negatively charged atoms. The zwitterionic state results from the amine group (H_3N^+) being positively charged and the carboxylic group (COO^-) being negatively charged.

Part I
Protein Structure

Here we give the basic introduction to proteins. What are they made of, how is their sequence and structure determined.

1 The Basics

Proteins are large polymers made of a linear chain of building blocks called *amino acids*. An amino acid is a simple organic molecule consisting of a basic (hydrogen-accepting), amine group bound to an acidic (hydrogen-donating) carboxyl group via a single intermediate carbon atom, called the alpha carbon, or C-α. Figure 1 shows a generic amino acid. The C-α, along with the amine and carboxyl groups, form the amino acid *backbone*. All amino acids have the same backbone. The difference between amino acids lies only in their *side chains*, denoted **R** in Figure 1. The R group can be anything and the possibilities are endless. There are hundreds of naturally occurring amino acids, only 20 amino acids are generated by the universal genetic code. The side chains of these amino acids are shown in Table 1. Two other rare amino acids, selenocysteine (SEC) and pyrrolysine (PYL), are incorporated into proteins by unique synthetic mechanisms. The amino acids in proteins are also called alpha (α) amino acids.

All the amino acids with the exception of glycine are *chiral*, since their C-α is bound to four distinct chemical groups. Chiral molecules have two *stereoisomers*. This means that they can come in two forms (denoted D and L), which are identical in their chemical formula and atomic sequence, but their three-dimensional orientation is different. In fact, stereoisomers are mirror images of each other. Imagine your right hand and left hand. Your right hand is a mirror image of your left hand. No matter how you try, you cannot superpose your right hand onto your left hand! The difference becomes obvious if you try to wear a left glove on your right hand. Similarly, two stereoisomers cannot be superposed onto one another by any rotation or translation. This means that they are two distinct molecules. You cannot simply convert one into another without forming and breaking bonds (see Figure 2). For some unknown reason, all the amino acids in proteins are L ("Left-handed") amino acids only. D ("Right handed") amino acids can by synthesized, but they are only seen very rarely in nature.

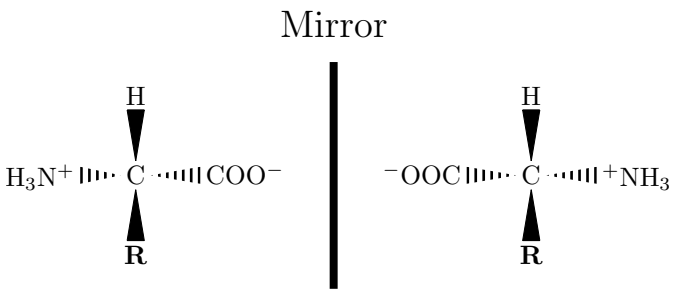
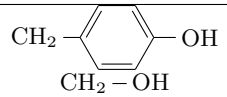
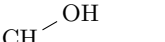
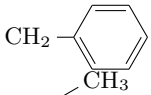
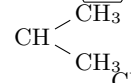
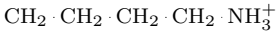
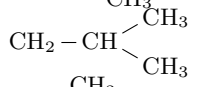
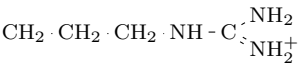
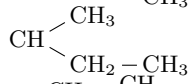
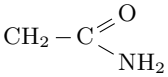
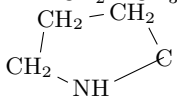
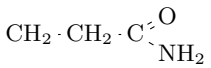
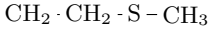
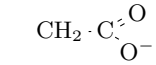
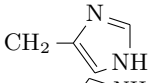
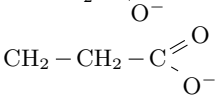
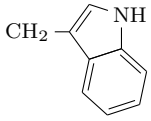


Figure 2: L (left) and D (right) amino acids are like mirror image of each other. The dashed lines are pointing away from you, and the bold lines are pointing towards you.

Table 1: The side chains of the 20 standard amino acids.

Name	3-letter	1-letter	Side chain	Name	3-letter	1-letter	Side chain
Glycine	GLY	G	H	Tyrosine	TYR	Y	
Alanine	ALA	A		Serine	SER	S	
Phenylalanine	PHE	F		Threonine	THR	T	
Valine	VAL	V		Lysine	LYS	K	
Leucine	LEU	L		Arginine	ARG	R	
Isoleucine	ILE	I		Asparagine	ASN	N	
Proline	PRO	P		Glutamine	GLN	Q	
Methionine	MET	M		Aspartic acid	ASP	D	
Histidine	HIS	H		Glutamic acid	GLU	E	
Tryptophan	TRP	W		Cysteine	CYS	C	

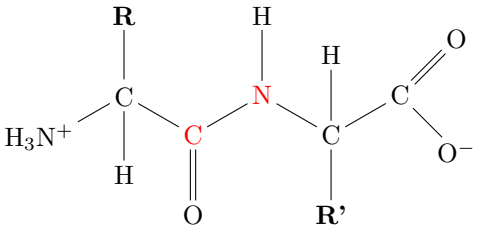


Figure 3: Two amino acids linked by a peptide bond. The C and N atoms involved in the peptide bond are marked in red. The side chains R and R' may or may not be the same.

There are several ways to describe the structure of a protein molecule in a hierarchical manner from the amino acids sequence **(primary structure)**, through small, stable structural elements in the protein **(secondary structure)**, to the 3D structure of a protein chain **(tertiary structure)** and finally multiple protein chains bound together **(quaternary structure)**, as detailed below.

2 Primary Structure

During the translation of a protein from a gene, amino acids form an end-to-end bond with each other through a reaction of their amine and carboxyl groups. This reaction forms an amide bond, also known as peptide bond in the case of amino acids. (see Figure 3). Such a chain (polymer) is also referred to as a polypeptide. A medium sized protein may contain hundreds of amino acids connected through peptide bonds. The *Primary Structure* of a protein refers to its amino acid sequence. The primary structure of a protein is easily obtainable from its corresponding gene sequence, as well as by experimental techniques. Unfortunately, the primary structure is only indirectly related to the protein's function. In order to work properly, a protein must fold to form a specific three-dimensional shape, called its **native structure** or **native conformation**.

3 Spatial Positioning

Before learn more about the structural properties of proteins, it is important to know how atoms are spatially positioned. The peptide bond imposes some restrictions on the three-dimensional structure. The C-N bond, marked in red in Figure 3 is conformationally restricted and the C=O and N-H atoms involved in the peptide bond are approximately on the one plane. The N-C-α bond and the C-α-Carbonyl bond, on the other hand, are more flexible and the ability of the polypeptide chain to rotate around these bonds is what allows the protein to fold into a 3D structure. Let us first define what is a dihedral angle in general: A rotatable bond in 3D is called torsion or dihedral angle. In general, a dihedral angle is an angle between two hyperplanes in space. In 3D, a dihedral angle is uniquely defined by four ordered points in space – p_1, p_2, p_3, p_4. One plane is defined by p_1, p_2, p_3 and the other plane is defined by p_2, p_3, p_4. In molecules, a dihedral angle is therefore uniquely defined by four atoms: a, b, c, d. The two planes are the plane defined by a, b, c and the plane defined by b, c, d. One way to look at the dihedral angle is as the angle between these two planes. An equivalent way to look at it is the clockwise rotation angle, in 3D, between the two middle atoms, a_2 and a_3. Figure 4 shows the two alternative ways to think of molecular dihedral angles. By convention, the dihedral angle range is not $[0, 360]$ but $[-180, 180]$ or $[-\pi, \pi]$ in radians.

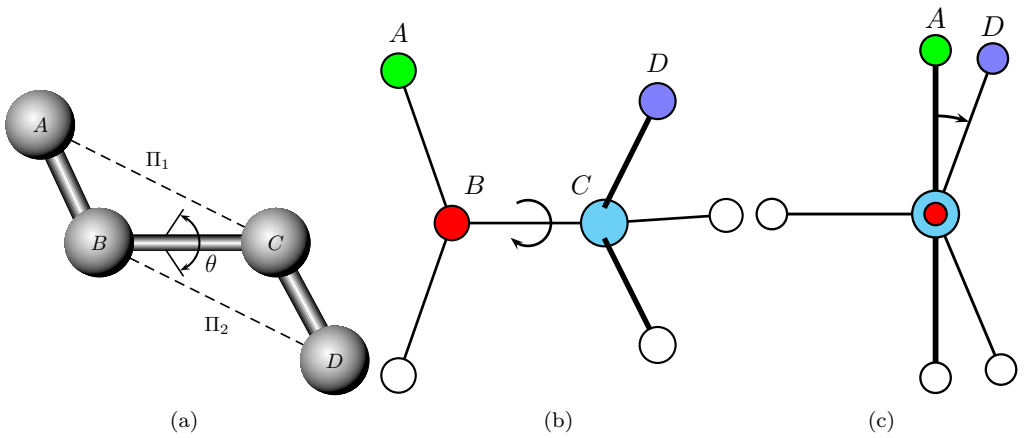


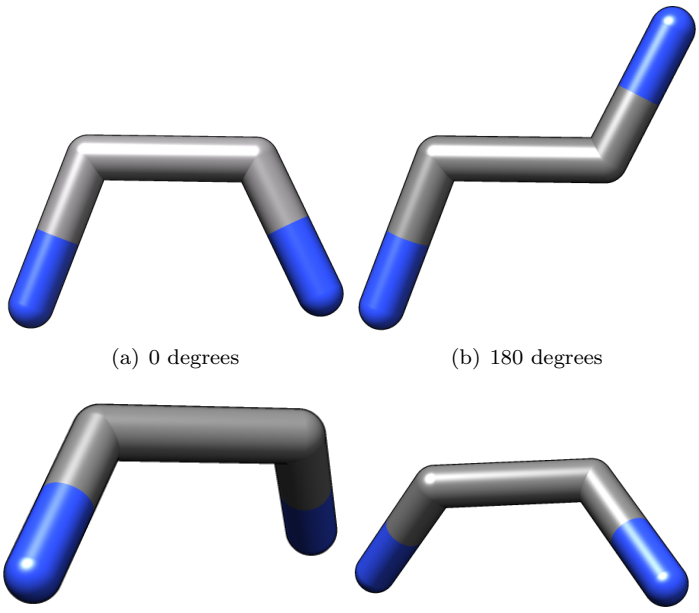
Figure 4: A dihedral angle is defined as the angle between two planes. In molecules, it is defined by four consecutive atoms. (a) It can be viewed as the spatial bond around the two middle atoms or alternatively, as the rotation angle around the two planes Π_1 and Π_2, defined by A,B,C and B,C,D respectively. If you rotate the image in (b) by 90° so that B and C co-incide, the dihedral angle is the angle from A to D (c). The sign of the angle is positive if D is on the right side of A, and negative otherwise.

There are three possible backbone dihedral angles. They are called phi (ϕ), involving the backbone atoms C_{i-1}-N-Cα-C (the i-1 refers to the C atom of the previous amino acid), psi (ψ), involving the backbone atoms N-Cα-C-$N_i + 1$ (the i+1 refers to the N atom of the next amino acid) and omega ω, involving the backbone atoms Cα-C'-N_{i+1}-Cα_{i+1}. See Figure 6. As mentioned before, the ω angle, which is the angle around the peptide bond between two consecutive amino acid, is rigid and planar – always approximately 180°.

Figure 5 shows the position of the four atoms involved in a dihedral angle for various values of the ψ. When ψ is 0°, the four atoms are on the same plane and the first and last atoms are on the same side of the plane (also called a *cis* configuration). When ψ is 180°, the four atoms are on the same plane and the first and last atoms are on opposite sides of the plane (also called a *trans* configuration). For any other values, the four atoms are not on the same plane. For example, when ψ is 90°, if you place a, b and c on the same plane, d will "stick out" towards you at a 90° angle. when ψ is $-90°$, if you place a, b and c on the same plane, a_4 will point away from you at a 90° angle.

4 Secondary Structures

Secondary structure refers to folding in a small part of the protein that forms a characteristic shape. These are small regular structural elements, stabilized by hydrogen bonds between the carbonyl oxygen atom and the amine group in the backbone of the protein. The most common secondary structure elements are alpha helices and beta sheets.



(a) 0 degrees (b) 180 degrees

(c) 90 degrees (d) -90 degrees

Figure 5: The spatial position of four atoms as a function of several dihedral angle values.

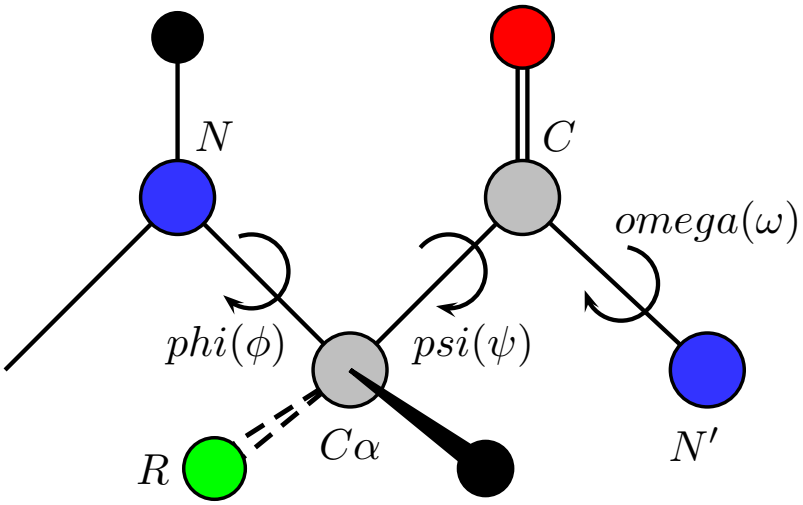


Figure 6: The backbone dihedral angles, ϕ, ψ and ω.

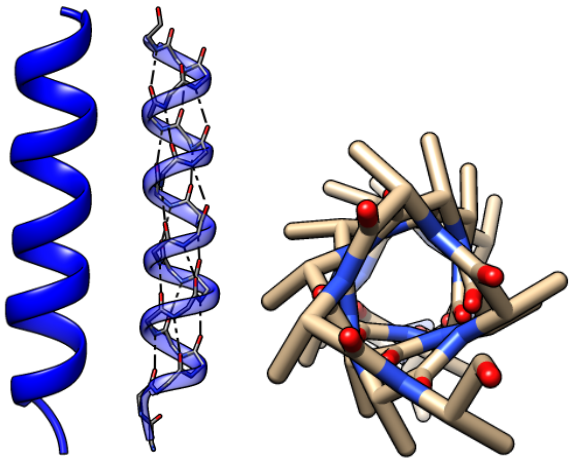


Figure 7: From left to right: A side view of an alpha helix as ribbon diagram; An alpha helix with the backbone hydrogen bonds displayed in dashed lines; A front view of an alpha helix.

4.1 Alpha Helices

An alpha-helix (or α-helix) is a right hand-coiled or spiral conformation (helix) in which every backbone amine group donates a hydrogen bond to the backbone carboxyl group of the amino acid four residues earlier (i+4 $\rightarrow$ i hydrogen bonding). There are 3.6 amino acids in each helical turn, which means that each amino acid residue corresponds to a 100° turn in the helix. The alpha helix is the most common secondary structure in proteins. Helices observed in proteins can range from four to over 40 amino acids, but a typical helix contains about ten amino acids (about three turns). Residues in α-helices typically adopt (ϕ, ψ) backbone angles around (-60°, -45°). These values are approximate, and several sources give slightly different values. Typically, the sum of the ψ angle of a residue and the ϕ angle of the next residue is around -105°. Figure 7 shows an α-helix from the front and a side view. Notice the regular hydrogen bond pattern.

4.2 Beta Sheets

The β-sheet (also β-pleated sheet) is the second form of regular secondary structure in proteins. Beta sheets consist of beta strands connected laterally by at least two or three backbone hydrogen bonds, forming a generally twisted, pleated sheet. A β strand is a stretch of polypeptide chain typically 3 to 10 amino acids long with backbone.

The majority of β strands are arranged adjacent to other strands and form an extensive hydrogen bond network with their neighbors in which the N-H groups in the backbone of one strand establish hydrogen bonds with the C=O groups in the backbone of the adjacent strands. In the fully extended β strand, successive side chains point straight up, then straight down, then straight up, etc. Adjacent β strands in a β sheet are aligned so that their C-α atoms are adjacent and their side chains point in the same direction. The "pleated" appearance of β strands arises from tetrahedral chemical bonding at the C-α atom; for example, if a side chain points straight up, then the bond to the C$'$ must point slightly downwards, since its bond angle is approximately 109.5°. The pleating causes the distance between $C^{\alpha}{}_i$ and C^{α}_{i+2} to be approximately 6 Å, rather than the

7.6 Å(2 3.8 Å) expected from two fully extended trans peptides. The "sideways" distance between adjacent Cα atoms in hydrogen-bonded β strands is roughly 5Å.

Because peptide chains have a directionality conferred by their N-terminus and C-terminus, β strands too can be said to be directional. They are usually represented in protein topology diagrams by an arrow pointing toward the C-terminus. Adjacent β strands can form hydrogen bonds in antiparallel, parallel, or mixed arrangements.

In an antiparallel arrangement, the successive β strands alternate directions so that the N-terminus of one strand is adjacent to the C-terminus of the next. This is the arrangement that produces the strongest inter-strand stability because it allows the inter-strand hydrogen bonds between carbonyls and amines to be planar, which is their preferred orientation. The peptide backbone dihedral angles (ϕ, psi) are about (140°, 135°) in antiparallel sheets. In this case, if two atoms C_i^α and C_j^α are adjacent in two hydrogen-bonded βstrands, then they form two mutual backbone hydrogen bonds to each other's flanking peptide groups; this is known as a close pair of hydrogen bonds.

In a parallel arrangement, all of the N-termini of successive strands are oriented in the same direction; this orientation may be slightly less stable because it introduces nonplanarity in the inter-strand hydrogen bonding pattern. The dihedral angles (ϕ, ψ) are about (120°, 115°) in parallel sheets. It is rare to find less than five interacting parallel strands in a motif, suggesting that a smaller number of strands may be unstable, however it is also fundamentally more difficult for parallel β-sheets to form because strands with N and C termini aligned necessarily must be very distant in sequence.

In parallel β-sheet structure, if two atoms C_i^α and C_j^α are adjacent in two hydrogen-bonded β strands, then they do not hydrogen bond to each other; rather, one residue forms hydrogen bonds to the residues that flank the other (but not vice versa). For example, residue i may form hydrogen bonds to residues j-1 and j+1; this is known as a wide pair of hydrogen bonds. By contrast, residue j may hydrogen-bond to different residues altogether, or to none at all.

Finally, an individual strand may exhibit a mixed bonding pattern, with a parallel strand on one side and an antiparallel strand on the other. Such arrangements are less common than a random distribution of orientations would suggest, suggesting that this pattern is less stable than the anti-parallel arrangement, however bioinformatic analysis always struggles with extracting structural thermodynamics since there are always numerous other structural features present in whole proteins. Also proteins are inherently constrained by folding kinetics as well as folding thermodynamics, so one must always be careful in concluding stability from bioinformatic analysis. The hydrogen bonding of β strands need not be perfect, but can exhibit localized disruptions known as beta bulges.

The hydrogen bonds lie roughly in the plane of the sheet, with the peptide carbonyl groups pointing in alternating directions with successive residues; for comparison, successive carbonyls point in the same direction in the alpha helix.

4.3 Other Secondary Structure Elements

Several other, less common secondary structure elements exist:

3_{10} **helix:** This is a rare form of helix, found in approximately 10-15% of all helices, in particular as an extension of α helices near the N- or C-termini. 3_{10} helices are usually quite short. The amino acids are arranged in a right-handed helical structure, similar to α helices. However, there

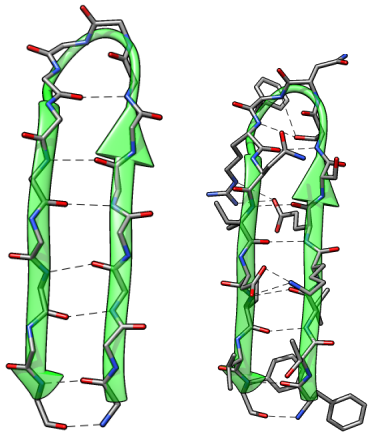


Figure 8: From left to right: An anti-parallel β sheet made of two strands. The same sheet with bonds.

are three residues per turn as each amino acid corresponds to a 120° turn in the helix. There is a translation of 2Åalong the helical axis, and the ring formed by making the hydrogen bonds has 10 residues (hence the name 3_{10}. Most importantly, the N-H group of an amino acid forms a hydrogen bond with the C=O group of the amino acid three residues earlier (a repeated i + 3 $\rightarrow$ i hydrogen bonding pattern). 3_{10} helices adopt (ϕ, ψ) values near (-49, -26). Typically, the sum of the ψ angle of a residue and the ϕ angle of the next residue is around -75°.

π helix: This is another rare form of helix, found in approximately 15% of known protein structures. The amino acids are arranged in a right-handed helical structure. Each amino acid corresponds to an 87° turn in the helix (i.e., the helix has 4.1 residues per turn), and a translation of 1.15Åalong the helical axis. Most importantly, the N-H group of an amino acid forms a hydrogen bond with the C=O group of the amino acid five residues earlier (a repeated i+5$\rightarrow$ i hydrogen bond pattern). π helices are generated by inserting an additional amino acid into a standard α helix. This is destabilizing and most π helices are short – at most 7 amino acid long. Due to the short length of π helices, there are no single dihedral angle values for them. However, the sum of the ψ angle of a residue and the ϕ angle of the next residue is roughly -125°.

Left-handed α helix: Most naturally occurring α helices are right-handed. That is, if you hold the helix pointing away from you, it twists clockwise moving away. Left handed α helices exist too, but they are much less common. As far as dihedral angles are concerned, a left handed helix is a mirror image of a right-handed helix, so that the (ϕ, ψ) backbone angles are around (+60°, +45°). The hydrogen bonding patterns and the length of the helical twist are the same, but left-handed helices are less favorable in nature. The reason lies in the fact that naturally occurring amino acids are L-amino acids (see above). The left-handed configuration positions the amino acid side chains next to the C=O group, making the structure over-crowded and causes steric clashes. The right-handed configuration places the side chains near the smaller N-H group, which is much more favorable.

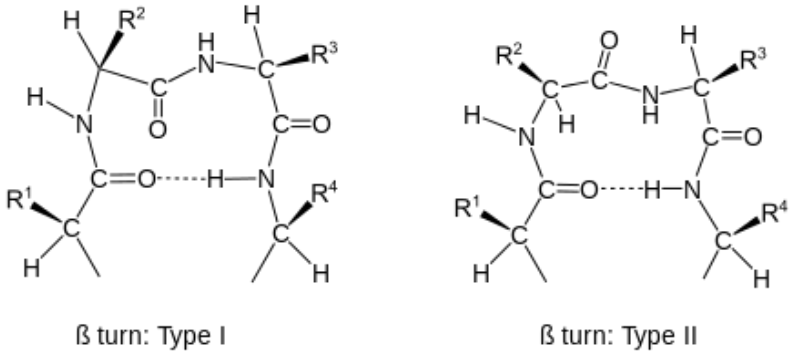


Figure 9: Examples of Type I (left) and Type II (right) β turns.

Table 2: Different β turn types and their characteristic backbone dihedral angles.

Type	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
I	-60	-30	-90	0
II	-60	120	80	0
I'	60	30	90	0
II'	60	-120	-80	0
VIa1	-60	120	-90	0
VIa2	-120	120	-60	0
VIb	-135	135	-75	160
VIII	-60	-30	-120	120
IV	Turns not included in the above categories			

4.4 Loops, Coils and Turns

Loops, coils and turns are parts of the polypeptide chain that do not belong to any of the regular secondary structures but they still play a very important role in protein structure. Loops vary in size. Some can be as short as two amino acids, but some can be very long – 10 amino acids or more. Longer loops are sometimes called "random coils" even though they are not coiled and rarely random. They have amino acid and structural preferences, and they tend to contain binding sites and enzyme activity sites. For this reason they often reside on protein surfaces and are evolutionarily conserved.

Turns (or β turns) connect β strands and as the name suggests, they reverse the direction of the chain and allow the protein chain to fold. Each turn consists of four amino acid residues (labeled $i, i+1, i+2$ and $i+3$), where the C=O of residue i and the N-H of residue $i+3$; the distance between the C-α atoms of i and $i+3$ is less than 7Å [**?**]. The turns are classified into several types – I, II, I', II', VIa, VIb, VIII and IV, distinguished by the ϕ, ψ angles of residues $i+1$ and $i+2$. Type I is the most common. The average values of the ϕ, ψ angles for the two middle residues, $i+1, i+2$, are shown in Table 2. Notice that Types I' and II' are simply mirror images of Types I and II, respectively. Types I and II are shown in Figure 9

4.5 Super-Secondary Structures, Motifs and Domains

4.6 Ramachandran Plots

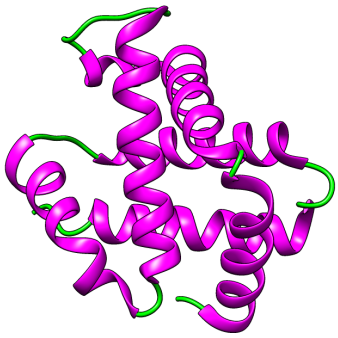
The Ramachandran plot provides an easy way to view the distribution of torsion angles of a protein structure [**?**]. It is named after the Indian physicist, G. N. Ramachandran who described them in 1963 along with C. Ramakrishnan, and V. Sasisekharan. The plot provides an overview of allowed (energetically favorable) and disallowed (unfavorable) regions of torsion angle values. It can be used to measure the quality of protein structures and severs as a very convenient way to visualize the dihedral angles.

The plot is simply a two-dimensional scatter-plot. The horizontal axis shows the ϕ value and the vertical axis shows the *psi* values. Each dot on the Ramachandran plot provides the ϕ and ψ values for an amino acid in a protein. Remember that by convention the ϕ and ψ values are in the range [-180,+180]. Therefore, the left hand corner starts from -180 and extends to +180 for both axes. This is a convenient presentation and allows clear distinction of the characteristic regions of α-helices and β-sheets. The typical secondary structure regions can clearly be seen as clusters of points on the Ramachnadran plot in Figure 10. The area on the right side of the plots, with a smaller cluster of dots, represents a left-handed α helix. The regions are "allowed" regions, where dihedral angles are commonly seen in proteins.

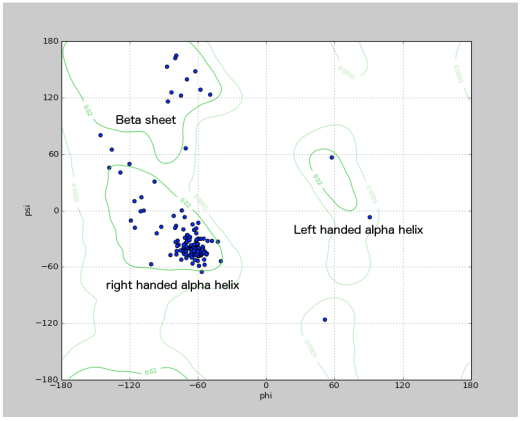
Loops and coils can vary in their ϕ and ψ values and wander outside the secondary structure characetristic region, but even in loops, dihedral angles are restricted to a small subset of the possible values, since most angles will result in sterical clashes between the backbone and side-chain atoms which will be too close to one another. These regions are called "forbidden". (You probably know that when two atoms are too close to each other the energy of the system gets too high). For a high-quality structure the forbidden regions are usually empty or almost empty – very few amino acids have their dihedral angles within these regions. Sometimes such values can be found and they most probably will result in some strain in the protein chain. In such cases additional interactions will be present to stabilize such structures. They may have functional significance and may be conserved within a protein family.

Two amino acids, Glycine and Proline, do not follow the above rules. Glycine does not have a side chain and therefore it is highly flexible and can reach dihedral angles which are normally not allowed for other amino acids. That is why glycine is often found in loop regions, where the polypeptide chain makes a sharp turn. Glycine is rarely in secondary structure elements since it tends to break them. This is the reason for the high conservation of glycine residues in protein families, since the presence of turns at certain positions is a characteristic of a particular fold of a protein structure. Figure 11 (a) shows the Ramachandran plot for glycine.

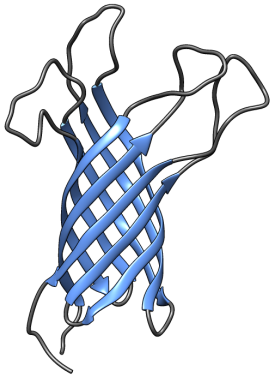
Another residue with special properties is proline. Proline, in contrast to glycine, has a very constrained side chain which forms a five-membered ring with the backbone. Proline is often found at the end of helices and functions as a helix disruptor. This makes the allowable regions for Proline considerably more restricted than other amino acids (Figure 11 (b)). The constrained dihedral angle space of Proline also restricts the allowed values of ϕ and ψ of the amino acid preceding it on the polypeptide chain (Figure 11 (c)).



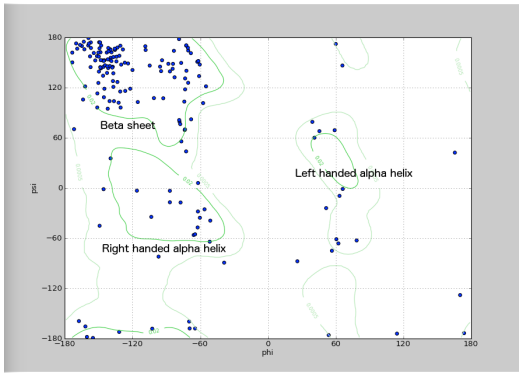
(a)



(b)



(c)



(d)

Figure 10: An example of a Ramachandran plot of an α helical protein (a-b) and a mostly β sheet protein (c-d). Notice the clusters of amino acids around the secondary structure regions.

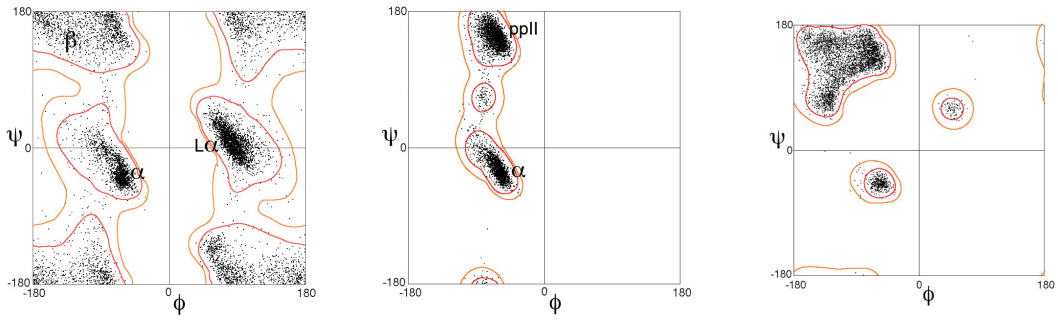


Figure 11: (a) Ramachandran plot or Glycine residues. Data from 500 proteins at 1.7Å resolution or better. (b) Ramachandran plot for Proline residues. (c) Ramachandran plot for the residue preceding Proline on the polypeptide chain.

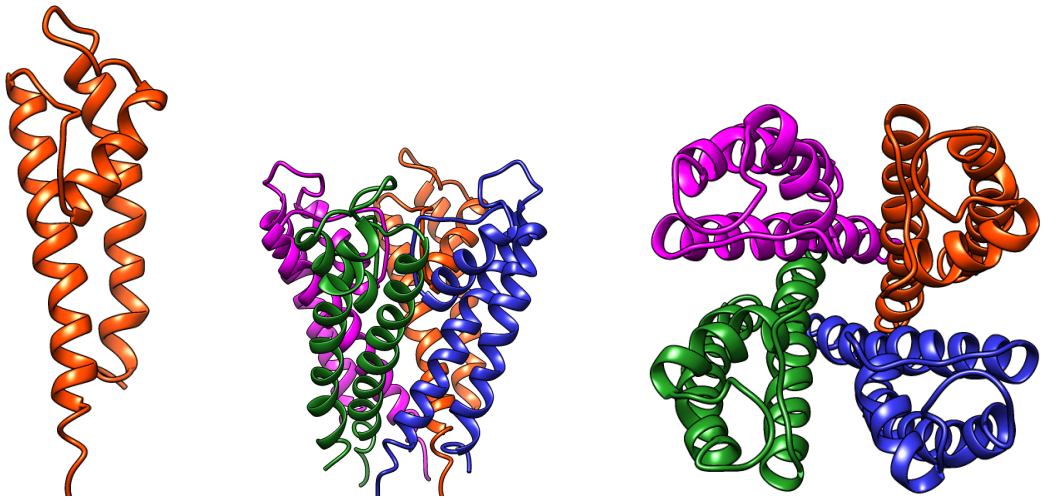
5 Tertiary Structures

The overall three-dimensional shape of an entire protein molecule is the tertiary structure. The protein molecule will bend and twist in such a way as to achieve maximum stability or lowest energy state. Although the three-dimensional shape of a protein may seem irregular and random, it is fashioned by many stabilizing forces due to bonding interactions between the side-chain groups of the amino acids. See Figure 12 (a).

Protein folding is the process by which a protein structure assumes its functional shape or conformation. It is the physical process by which a polypeptide folds into its characteristic and functional three-dimensional structure from random coil.[1] Each protein exists as an unfolded polypeptide or random coil when translated from a sequence of mRNA to a linear chain of amino acids. This polypeptide lacks any stable (long-lasting) three-dimensional structure (the left hand side of the first figure). Amino acids interact with each other to produce a well-defined three-dimensional structure, the folded protein (the right hand side of the figure), known as the native state. The resulting three-dimensional structure is determined by the amino acid sequence (Anfinsen's dogma).[2] Experiments [3] beginning in the 1980s indicate the codon for an amino acid can also influence protein structure.

6 Quaternary Structures

Many proteins are made up of multiple polypeptide chains, often referred to as protein subunits. These subunits may be the same (as in a homodimer) or different (as in a heterodimer). The quaternary structure refers to how these protein subunits interact with each other and arrange themselves to form a larger aggregate protein complex. The final shape of the protein complex is once again stabilized by various interactions, including hydrogen-bonding, disulfide-bridges and salt bridges. For example, see Figure 12 (b) and (c).



(a) (b) (c)

Figure 12: (a) A tertiary structure example (a) single chain from Potassium channel (KCSA) from Streptomycis Lividans, PDB code 1BL8). (b) A side view of the quaternary structure of the entire complex comprising four chains of the protein shown in (a). (c) A front view.

7 Protein Structure Determination

7.1 X-ray Crystallography

The most commonly used and usually highest-resolution method of structure determination is x-ray crystallography. To obtain structures by this method, laboratory biochemists obtain a very pure, crystalline sample of a protein. X-rays are then passed through the sample, in which they are diffracted by the electrons of each atom of the protein. The diffraction pattern is recorded, and can be used to reconstruct the three-dimensional pattern of electron density, and therefore, within some error, the location of each atom. A high-resolution crystal structure has a resolution on the order of 1 to 2 Angstroms (Å) One Angstrom is the diameter of a hydrogen atom (10^{-10} meter), or one hundred-millionth of a centimeter). Unlike other structure determination methods, with x-ray crystallography, there is no fundamental limit on the size of the molecule or complex to be studied. However, in order for the method to work, a pure, crystalline sample of the protein must be obtained. For many proteins, including many membrane-bound receptors, this is not possible. In addition, a single x-ray diffraction experiment provides only static information - that is, it provides only information about the native structure of the protein under the particular experimental conditions used. As we will see later, proteins are often flexible, dynamic objects when in their natural state in solution, so a single structure, while useful, may not tell the full story. More information on X-ray Crystallography is available at Crystallography 1015 and in the Wikipedia6 .

7.2 NMR

Nuclear Magnetic Resonance (NMR) spectroscopy has recently come into its own as a protein structure determination method. In an NMR experiment, a very strong magnetic field is transiently applied to a sample of the protein being studied, forcing any magnetic atomic nuclei into alignment. The signal given off by a nucleus as it returns to an unaligned state is characteristic of its chemical environment. Information about the atoms within two chemical bonds of the resonating nucleus can be deduced, and, more importantly, information about which atoms are spatially near each other can also be found. The latter information leads to a large system of distance constraints between the atoms of the protein, which can then be solved to find a three-dimensional structure. Resolution of NMR structures is variable and depends strongly on the flexibility of the protein. Because NMR is performed on proteins in solution, they are free to undergo spatial rearrangements, so for flexible parts of the protein, there may be many more than one detectable structures. In fact, NMR structures are generally reported as ensembles of 20-50 distinct structures. This makes NMR the only structure determination technique suited to elucidating the behavior of intrinsically unstructured proteins, that is, proteins that lack a well-defined tertiary structure. The reported ensemble may also provide insight into the dynamics of the protein, that is, the ways in which it tends to move. NMR structure determination is generally limited to proteins smaller than 25-30 kilodaltons (kDa), because the signals from different atoms start to overlap and become difficult to resolve in that range. Additionally, the proteins must be soluble in concentrations of 0.2-0.5 mM without aggregation or precipitation. For more information on how NMR is used to find molecular structures, see here: World of NMR: Magnets, Radio Waves, and Detective Work at .

7.3 Electron Diffraction

Electron diffraction (ED) uses a transmission electron microscope (TEM) to probe the structure instead of x-rays. Because of difficulties in obtaining and interpreting electron diffraction data, it is rarely used for protein structure determination. However, it is useful in cases where X-rays cannot form a diffraction, because electrons interact more strongly with atoms than X-rays do. For example, X-rays will travel through a thin 2-dimensional crystal (say a secondary structure) without diffracting significantly, whereas electrons can be used to form an image.

Electron diffraction can also be useful in determining the structures of very large polyhedral macromolecules such as virus capsids with icosahedral symmetry [**?**]. Large macromolecular complexes and molecular machines present a particular challenge in structure determination. Generally too large to be crystallized, and too complex to solve by NMR, determining the structure of these objects usually requires the combination of high-resolution microscopy combined with computational refinement and analysis. The main techniques used are cryo-electron microscopy (Cryo-EM) and standard light microscopy [**?**]. The sample is studied at cryogenic temperatures (generally liquid nitrogen temperatures) and an electron beam bombards the sample. The image that is viewed is a result of the interaction of the sample with this beam. One useful method of Cryo-electron microscopy for determining the structure of proteins is Cyro-Electron tomography (CET). CET is performed at cryogenic temperatures as is cryo-electron microscopy, and it constructs a 3D sample from 2D images. Multiple 2D images are taken from various projections and a 3D image is reconstructed using software and visualization tools. Cryo-EM techniques have improved steadily, and in 2014 some structures at near atomic resolution had been obtained [**?**].