Chapter 2

Overview of Protein Structure

2.1 Synthesis and Information Contents

It has been long recognized that life is based on morphological units known as **cells**. The formulation of this concept is generally attributed to an 1838 paper by Matthias Schleiden and Theodor Schwann, but its origins may be traced to the seventeenth century observations of early microscopists such as Robert Hooke. Most of the molecular constituents of living systems are composed of carbon atoms covalently joined with other carbon atoms and with hydrogen, oxygen, or nitrogen. The special bonding properties of carbon permit the formation of a great variety of molecules. Organic compounds of molecular weight (M_r) less than about 500, such as nucleotides, amino acids and monosaccharides, serve as monomeric subunits of nucleic acids, proteins and polysaccharides, respectively. The Deoxyribonucleic acids (DNA) are polymers of nucleotides Adenine (A), Guanine (G) Cytosine (C) and Thymine (T) while in Ribonucleic acids (RNA) Thymine is replaced by Uracil (U). DNA (and sometimes RNA) is the cell's master repository of genetic "**information**". The expression of the genetic information is a two-stage process. In the first stage, which is termed **transcription**, a DNA strand serves as a template for the synthesis of a complementary strand of RNA. In the second stage of genetic expression, which is known as **translation**, ribosomes enzymatically link together amino acids to form proteins. The order in which the amino acids are linked together is prescribed by RNA's sequence of bases, since proteins are self-assembling, the genetic information encoded by DNA serves, through the intermediacy of RNA, to specify protein structure and function. Proteins are composed of 20 different kind of amino acids. The nucleotides from which nucleic acids are built and the amino acids from which proteins are built are identical in all leaving organisms. Consider following example. One can make a 8 unit word out of 26 letters of English alphabet, 4 different deoxyribonucletides and 20 different amino acids in 26^8 (2.1x 10^{11}), 4^8 (65,536) and 20^8 (2.56x10^{10}) ways, respectively. It is clear from the above example that such monomeric subsist in linear sequences can spell infinitely complex messages depends upon its length and as the information flows from DNA to protein the complexity increases rapidly (Voet and Voet, 1995).

Proteins are important as structural, functional and information career molecules. Talking biologically, proteins store and transport a variety of particles ranging from macromolecules to electrons. They guide the flow of electrons in the vital process of photosynthesis; as hormones, they transmit the information between specific cells and organs in complex organisms. Some proteins control the passage of molecules across the membranes that compartmentalize cells and organelles; proteins function in the immune systems of the complex organisms to defend against intruders; and proteins control gene expression by binding to the specific sequence of nucleic acids, thereby turning genes on and off. Proteins are the crucial components of muscles and other systems for converting chemical energy in to mechanical energy. They are also necessary for sight, hearing, and other senses. Many proteins are simply structural providing the filamentous architecture within cells and materials that are used in hair, nails, tendons, and bones of animals (Creighton, 1993).

2.2 Structural Hierarchy in Proteins

All proteins, in all species, regardless of the their function or biological activity, are polymers of the same set of 20 amino acids which are linked by covalent bonds. Amino acid sequences of the proteins can be deduced form the direct sequencing or from the DNA sequences of the related gene. Conceptually, protein structure can be considered at four levels.

Primary structure includes all the covalent bonds between amino acids and is normally defined by the peptide-bonded amino acids and location of disulfide bonds. The relative spatial arrangement of the linked amino acids is unspecified.

Secondary structure refers to regular, recurring arrangements in the space of adjacent amino acids in a polypeptide chain. There are a few common types of secondary structure, most prominent being the α helix and 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 structure 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. Understanding of protein structure, folding, and evolution has made it necessary to define two additional level between secondary structure and tertiary structure. A stable clustering of several elements of secondary structure is sometimes referred to as

supersecondary structure. The term is used to describe particularly stable arrangements that occur in many different proteins and sometimes many times in a single protein.

A somewhat higher level of structure is the **domain**. This refers to a compact region, including perhaps 40 to 400 amino acids, that is a distinct structural unit within a larger polypeptide chain. Many domains fold independently into thermodynamically stable structures. A large polypeptide chain can contain several domains that are readily distinguishable within overall structure. In some cases the individual domains have separate functions. However, important patterns exist at each of these levels of structure that provide clues to understanding the overall structure and function of large proteins.

2.3 Amino Acids and the Peptide Bond

Of the 20 amino acids usually found in proteins, 19 have the general structure as shown in Figure 2.1a-d.

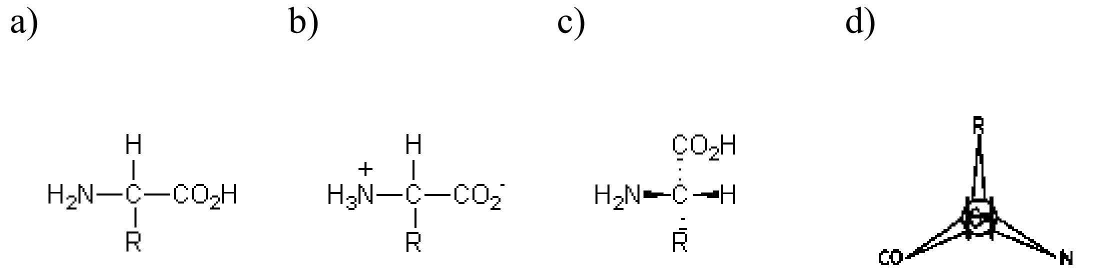

Figure 2.1: Different representations of general structure of amino acids. (a) Normal representation (b) Zwitter ionic structure (c) Geometric representation (d) The “CORN crib” for determining the handedness of an amino acid. Looking at the α carbon from the direction of hydrogen, the other substituents should read CO (carbonyls), R (side chain), and N (backbone NH) in clockwise order for a biologically appropriate L-amino acid.

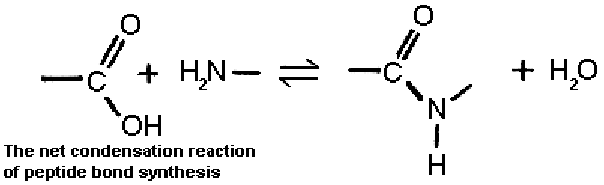

Figure 2.2 Showing net condensation reaction that results in formation of peptide bond. In the process a water molecule gets liberated.

The amino acids are linked in to proteins by the peptide bond, as in Figure 2.2, by the condensation of two amino acids. Generally, between 50 and 3000 such amino acids are linked in this way to form a typical linear polypeptide chain

2.4 Properties of Polypeptide Backbone and the Ramachandaran Plot

The backbone of the linear polypeptide chain consists of three atoms of each residue in the chain, the amide N_i, the C_i^{α}, and the carbonyl C_i', where *i* is the number of the residue, starting from the amino end of the chain.

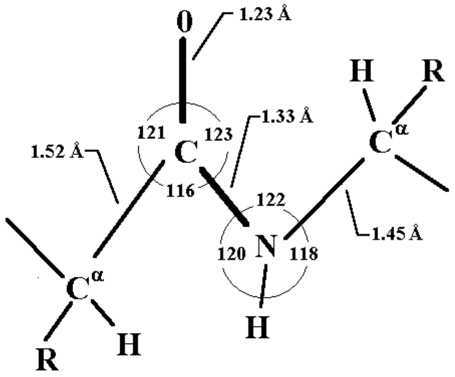

Figure 2.3: Showing the dimension of protein backbone derived from Ramachandran *et al.*, 1974.

The dimension of the peptide group of a residue is given in Figure 2.3, have been derived from three-dimensional structure analysis of small peptides (Ramachandran *et al.*, 1974). The presence of an asymmetric center at the C_α carbon atom, and only **L** amino acid residues, results in an inherent asymmetry of the polypeptide chain, that is important for spectral and conformational properties of polypeptides and proteins. The convention used to recognize correct L-amino acid handedness when dealing with physical models, stereo figures. Or molecular graphical displays: if one looks down on the α carbon from the direction of the hydrogen, other substituents should read "CO-R-N" in the clockwise order as shown in Figure 2.1d. In all the structures the central carbon or α carbon is bonded to an amino group, a carbonyl group, a hydrogen and an **R** group, that acts as

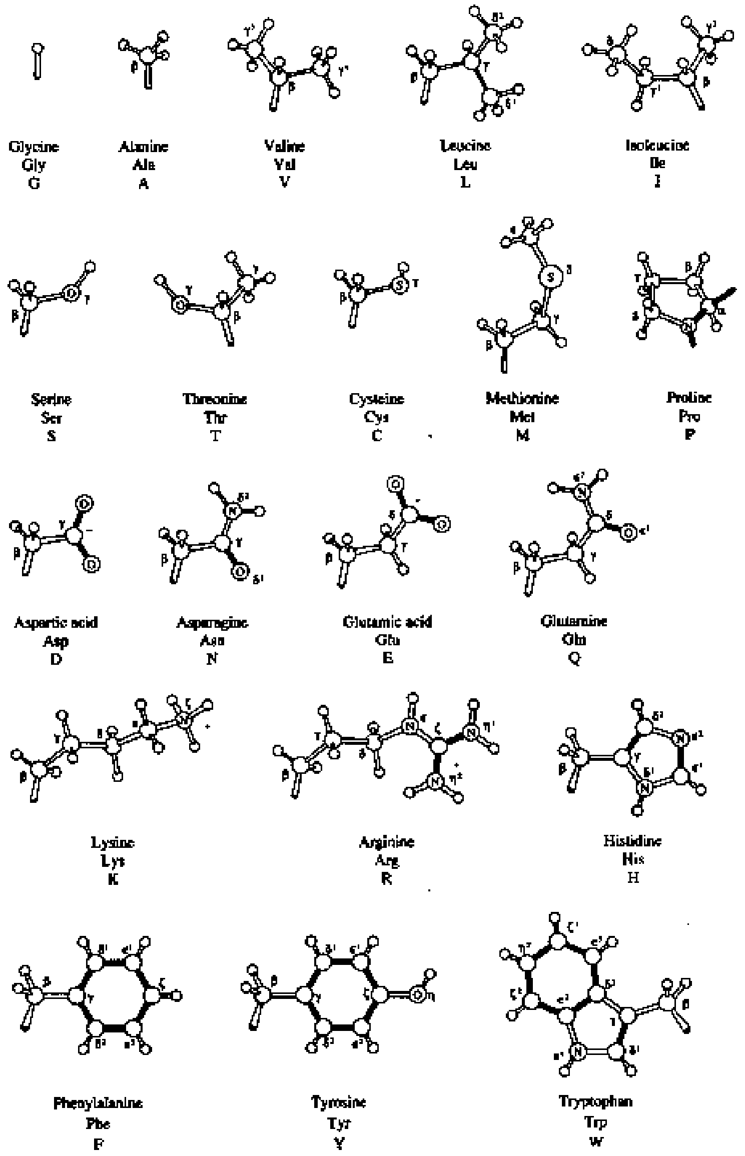

Figure 2.4 Side chains of the 20 amino acids that occur naturally in proteins. Small-unlabeled spheres are hydrogen atoms, and large unlabeled atoms are carbon atoms; other atoms are labeled. Double bonds are black, and partial double bonds are shaded. In the case of Pro, the bonds of the polypeptide backbone are included and are black. Below the name of the amino acid are the three-letter and the one-letter abbreviations commonly used. Note that isoleucine and threonine have asymmetric centers in the side chains, and only isomer illustrated is used biologically.

"side chain". The amino acids differ only in the chemical structures of the side chain **R**. The 20th natural amino acid, proline, is similar, but its side chain is bonded to the nitrogen atom to give the imino acid. Except in glycine, where the side chain is only a hydrogen atom, the central carbon atom is asymmetric and is always the **L** isomer. The side chain structure of each amino acid is shown in the Figure *2.4* with its full name, three- and one-letter codes. The central atom is designated as α, and the atoms of the side chains are commonly designated β,γ, δ, ε, and ζ, in order away from the α carbon

In principle, rotation could occur about any of the three bonds of each residue of the polypeptide backbone, but the peptide bond appears to have partial double-bonded character due to resonance. Consequently, the peptide bond length is only 1.33 Å, shorter than the usual C−N bond length of 1.45 Å, as in the C_α−N bond. It is however, longer than the value of 1.25 Å for the average C=N double bond. The peptide bond appears to have approximately 40% double-bonded character. As a result, rotation of this bond is restricted, and residues shown in Figure 2.3 have a strong tendency to be coplanar.

Resonance of the peptide bond tends to redistribute its electrons, and peptide backbone is correspondingly polar. The H and N atoms appear to have, respectively, positive and negative equivalent charges of 0.20 electron, where as C and O, respectively, have positive and negative equivalent charges of 0.42 electron. This gives the peptide bond a substantial permanent dipole moment of about 3.5 Debye units. The polypeptide backbone of the each residue contains one potent hydrogen bond donor, −NH−, and a hydrogen bond acceptor, carbonyl −CO−. This property is crucial for the polypeptide chain for three-dimensional architecture of proteins.

Two configurations of the planar peptide bonds are possible, one in which the C^α atoms are *trans*, and the other in which they are *cis* in conformation. The *trans* form is intrinsically favored energetically, probably owing to fewer repulsions between non-bonded atoms. If the residue that follows the peptide bond is Pro, how ever, its cyclic side chain diminishes the repulsions between atoms, and the intrinsic stability of the *cis* isomer is comparable to that of the *trans* isomer.

The above description indicates that the backbone of the protein is a linked sequence of rigid planar peptide groups. It is possible, therefore specify a polypeptide's backbone conformation by the dihedral angles (the angle formed between two planes) or rotation angles about C_α–N and C_α–C' bonds of each amino acid residues. A dihedral angle involves four successive atoms −A, B, C, and D−and three bonds joining them. If one look directly down the length of the central bond joining atoms B and C (the answer is the same as viewed from either end of this bond) and put the atom A at 12 on the clock face, then clock position of the far atom D reads out the dihedral angle for B−C bond. By convention, dihedral angles are assigned in the range of -180° to +180° with the clockwise direction being positive. The dihedral angle formed by $C_i–N_i–C_{\alpha i}–C'_{i+1}$ is denoted as φ and generally referred as rotation angle of $N_i–C_\alpha$ bond (Figure 2.5). The dihedral angle formed by $N_i–C_{\alpha\ i}–C'_{i+1}–N_{i+1}$ is denoted as ψ and generally referred as rotation angle of $C_{\alpha i}–C'_{i+1}$ bond (Figure 2.5). The third dihedral angle is formed by $C_{\alpha\ i-1}–C'_i–N_i–C_{\alpha i}$ is denoted as ω or and generally referred as rotation angle of $C'_i–N_i$ or the peptide bond (Figure 2.5).

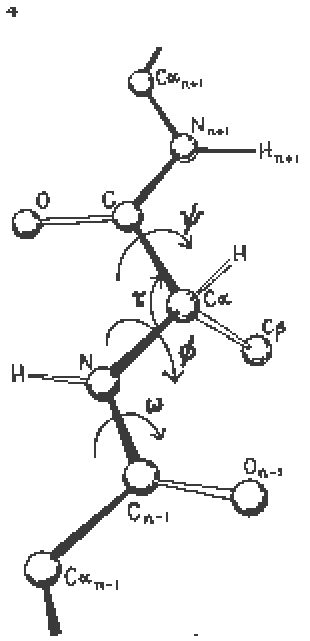

Figure 2.5 Nomenclature for the atoms of the polypeptide chain, the tetrahedral bond angle τ, and backbone dihedral angles φ, ψ, and ω.

Torisional angles of side chains are designated by χ_j, where *j* is the number of the bond counting outward from the C_α atom of the main chain. Assuming the ideality for the rest of the geometry, then three backbone dihedral angles per residue (φ,ψ, and ω) plus the side chain dihedral angles χ_j provides complete description of the local conformation. In practice, just φ and ψ suffice for the main chain, because the partial double bond character of the peptide bond keeps ω very close to flat. ω has a monomodel distribution with a mean of 180° and a small standard deviation of approximately 6°, which is the fully extended or *trans* conformation (Creighton, 1993). The curled up *cis* conformation of ω at or near 0° is observed about 10% of the time of proline and extremely rare for any other kind of amino acid. Hence, proline is the only exception in where ω distribution is bimodal.

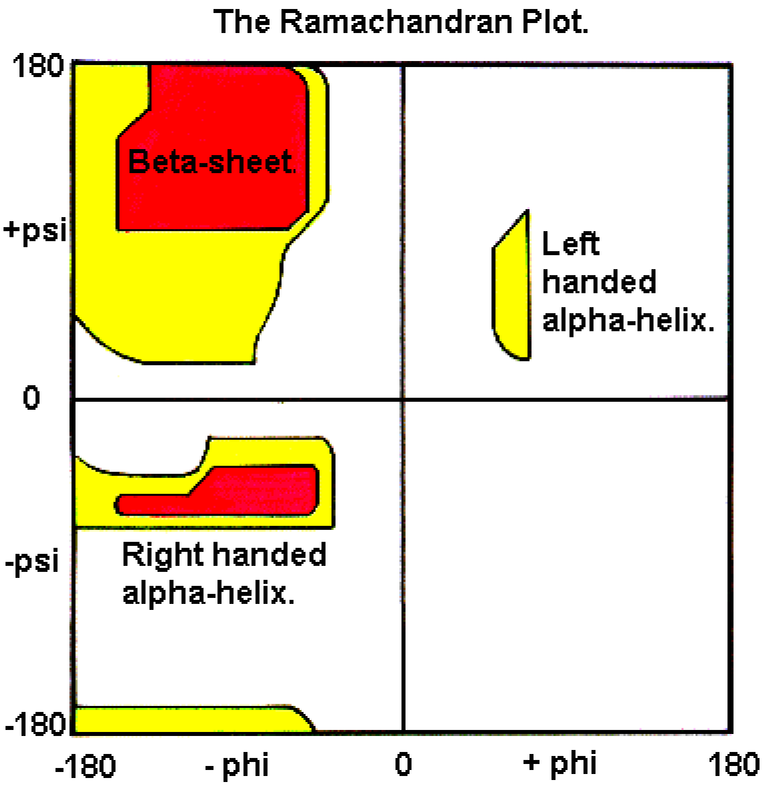

Figure 2.6 The positions of commonly found regular conformations of the proteins on a Ramachandran plot. The shown conformations are right and left-handed α helices and β sheet.

Since φ and ψ form a virtually complete description of the backbone conformation, a two dimensional plot of them is an important type of representation (Ramachandran and Sasiekharan, 1968). The plot is know as Ramachandran plot (Figure 2.6). Ramachandran plot can be used to illustrate properties of repeating conformations, single residues, or two successive residues and in general for studying the conformational properties. The regions of φ,ψ space, are generally named after the conformation the conformation that results, if they are repeated.

2.5 Definition of Secondary Structures

Main chain conformation can be classified in to secondary structures using Kabsch and Sander definitions (Kabsch and Sander, 1983). The distribution of φ,ψ pairs obtained from the protein structures in the protein structures in the protein databank shows six different picks. They correspond to right-handed α-helix (A), idealised β-strand (B), polyproline conformation (P), the ε region accessible primarily to Gly residues with positive φ angle (G), left-handed α-helix (L) and extended conformation. (E). These six peaks represent six different conformation states of the main chain of a particular residue.

The major conformations on the Ramachandran plot are the right-handed α helical cluster in the lower left near -60°, -40°; the broad region of extended β strands in the upper left quadrant (centered around -120°, 140°); and sparsely populated left-handed α-helical region in the upper right around +60°, +40° (Figure 2.6). Other regular conformations, like 3_{10}-helix (-49°, -26°), π-helix (-57°, -70°), polyproline1 (-83°, +158°), polyproline2 (-78°, +149°) and polyglycine2 (-80°, +150°) however do occur in proteins. The approximate mean values and standard deviations of the main chain dihedral angles in the classes are listed in Table1. Vacant areas in the Ramachandran plot (Figure 2.6) are the conformations that place the atoms unfavorably close together within the dipeptide unit. The asymmetry of the plot results from the collisions of the C_β.

	Mean (°)		Standard deviation (°)		Residue
					per
	φ_i	ψ_i	$\sigma_i(\varphi)$	$\sigma_i(\psi)$	turn
A (α-helix, R)	-65	-41	15	15	3.6
B (β-strand)	-130	135	15	20	2.0
P (polyproline)	-65	140	15	15	3.0
G (Gly with $+\varphi$)	60	40	10	10	NA
L (α-helix, L)	90	-10	15	10	3.6
E (extended)	130	180	25	25	NA

Table 1 Showing the mean dihedral angles for defining a secondary structure. The table is modified from Sali *et al.*, 1993.

2.6 Hydrogen Bonding

One of the more remarkable properties of the repetitive secondary structures observed in proteins is that the optimum φ, ψ values and the permissible range for good long-range H-bonding and steric fit are close to the optimum and range favorable for dipeptide conformations.

The dual hydrogen bonding capacity of the backbone peptide group is a persuasive influence on the protein structure. Although H-bonds are weak, non-covalent interactions, they are fairly directional and specific. Since each peptide can form a bond in both the directions, the co-operative effect of a network of such interactions can hold the polypeptide together in a strong and specific network.

Hydrogen bond involves an electrostatic attraction, either between two actual or between dipoles and they also involve the sharing of a proton. The group on one side of the H-bond is the "donor" D (usually, in proteins, a nitrogen or a water but sometimes an OH), which has a hydrogen it can contribute to the bond. The other group is the "acceptor", A, with accessible pair of electrons (usually a CO or water, but sometimes an unprotonated N or the backside of an OH). The optimum distance for a strong H-bond is about 3Å between D and A or 2Å between H and A. Angular criterion is important for hydrogen bonding.

2.7 Secondary Structures or Repetitive Structures

Patterns of main-chain hydrogen bonding, combined with repeating values of φ, ψ angles define secondary structures in proteins. The β-structures involves repeating patterns of H-bonds between distant part of the backbone, whereas helices involve repeating patterns of local H-bonding.

2.7.1 Helices

Helices are predominant, recurring form of secondary structure. The number of residues (*i)* and atoms (*x*) per single turn defines each type of helix. Two of the first helices hypothesized by Pauling and Corey in 1951, to occur in proteins were the α-helix or 3.6_{13}-helix, where $i = 3.6$ and $x = 13$, and the γ-helix or 5.1_{17}-helix, where $i = 5.1$ and $x = 17$. In conjunction with α- and γ- helices of Pauling and Corey, Donohue hypothesized the 2.2_7-helix, the 3_{10}-helix, the 4.3_{14}-helix and the 4.4_{16}-helix (Donohue, 1953). Of these hypothesized structures α-helix and 3_{10}-helices are reported in numerous reported protein structures, with α-helix being most abundant. Of other helices hypothesized by Donohue only π-helix has been reported and the occurrence is rare. The main features of the helical structures reported by known protein structures are as follows.

The right-handed α-helix (Figure 2.7b) is the best known and most easily recognized of the polypeptide regular structures, formed by repeated H-bonds between the CO of residue *i* and NH of residue *i+4*, with repeated φ, ψ values near -60º, -40º. Though the preferred values for φ and ψ angles differs with different analysis. The α-helices observed in actual protein structures are nearly always right-handed both because of the cumulative effect of a moderate energy difference for each residue and even more because each C_{β} would collide with the following turn of a left-handed α-helix.

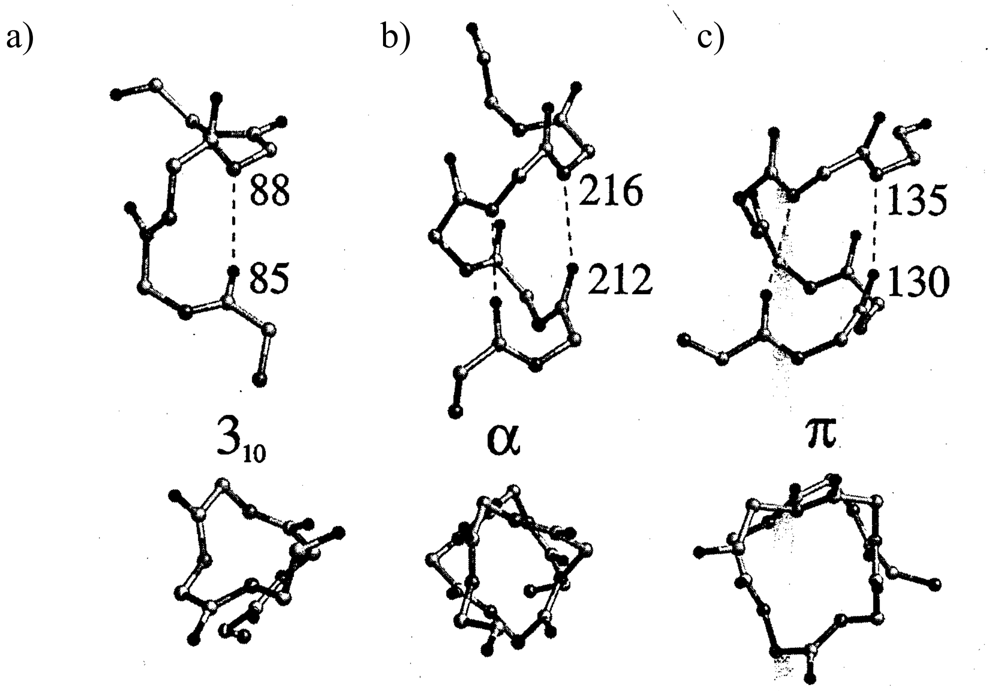

Figure 2.7: Helical structures witnessed within protein crystal structures. They are (a) 3_{10}-helix (b) α helix and (c) π helix. Figures on the top show the side view and that on the bottom shoes the top view of the helices. Residue numbering is random and shows the number of residues per turn in case of each helix. Hydrogen bonds between the backbone carbonyl oxygen and the backbone nitrogen are represented as dashed lines in the figures.

All atoms have been colored by type, where light gray = carbon, dark gray = nitrogen and black = oxygen.

The H-bonds in a α-helix are nearly parallel to the helix axis, with the CO all pointing towards the C-terminal end. Each peptide is tilted slightly, however, so that all the oxygen atoms point a bit outward. The β carbons do not extend radially out from the α carbons but make a clockwise pinwheel shape with Cβ nearly in the plane of the preceding peptide. The pitch, or repeat, of an ideal α helix is 3.6 residue per turn. For that pitch, the rise per residue along the helix axis is 1.5 Å, or 5.4 Å per turn. Real helices match this value quite well; however, a difference in average pitch of 5% (between, say, 3.5 and 3.7 residues per turn, which is well within the common range of variation) produces an offset of an entire residue by the end of a typical four or five turn helix. That 5% difference makes a trivial change in φ, ψ angles but has a substantial effect on side chain packing.

Variations on the helices come when the chain is either more tightly or more loosely coiled. Helices with hydrogen bonds to residues *i+3* and φ, ψ values near to -70°, -5° are designated as 3_{10} helix (Figure 2.7a). The 3_{10} helix is more tightly wound than α-helix and it has very distinctive triangular appearance in the end view. In the 3_{10} helix the α carbons on the successive turns are exactly in line with one another since there are an integral number of residues per turn; this makes the H-bond quite tilted relative to the helix axis. In contrast, the non-integral pitch of a α-helix lines up a CO on one turn with NH on the next to make parallel H-bonds, and α-carbons does not line up. The H-bond geometry and van der Waals interactions between successive turns are not quite as favorable in3_{10} helix, and long stretches are rare. The major importance of 3_{10} helix is that it very frequently forms the last turn at the C terminus of a α-helix and it is fourth most common structure found in proteins. Helices with hydrogen bond to *i+5* and φ, ψ values -57°, -70° are termed as π helices (Figure 2.7c). Their occurrence in structures is rare and till date only ten π helices are reported in the literature (Weaver, 2000). In each case the occurrence of the π helix was correlated with function. The conformation of π helix has been postulated to be disfavored for three reasons: (1) the dihedral angles are unfavorable (Low and Greenville-Wells, 1953; Ramachandran and Sasiekharan, 1968); (2) the 1Å hole at the center is wide enough to create a loss of van der Waals interactions, but too narrow to accommodate the water molecule for compensation, and (3) four residues need to be correctly aligned to allow collinear *i* to *i+5* hydrogen bond (Rohl and Doig, 1996). In the case of α and 3_{10} helices all main-chain H-bonding groups within the body of the helix are satisfied by the secondary structure formation. Each end produces three unsatisfied groups that often H-bonds to solvent, especially the open carbonyls at the C terminus. Very frequently, one of the free NHs nears the N-terminus H-bonds to the side chain of N-cap residue.

Pro residues are not ideally suited for either α-helix or β-sheet conformations. Poly(Pro) forms other regular conformations known as poly(Pro) I and II. Proline residues are special in permitting both *cis* and *trans* peptide bonds, and the two forms of poly(Pro) differ in this respect. Poly(Pro) I contains all *cis* peptide bonds whereas form II has all *trans* (Sasiekharan, 1959; Creighton, 1993). The values of φ and ψ are very similar for both, but form I is a right-handed helix with 3.3 residues per turn, whereas form II is a left-handed helix with 3.0 residues per turn. The values of φ (-83° and -78° for forms I and II, respectively) are compatible with that dictated by cyclic Pro side chain. The values for ψ are (+158 and +149 for forms I and II, respectively) constrained by steric repulsions and very similar in both the cases. Gly residue, owing to the fact that it lack the side chain, have unique conformational flexibility, and poly(Gly) likewise forms two regular conformations, designated as I and II. The former has a β-sheet conformation; the later is a helix with three residues per turn like that of poly(Pro) I.

2.7.2 β-sheet

After the α-helix the second most regular and identifiable secondary structure is the extended β strand (Pauling and Corey, 1951) with φ, ψ values in the upper left quadrant of the plot near -120°, 140°. In the extended β strand, the polypeptide backbone is fully extended and it has 2.0 residues per turn and a translation of 3.4 Å per residue. The backbone H-bonding groups are again completely satisfied within the body of β-sheet, but since the H-bonds go from one strand to another, β structure is inherently less local and modular than helices (Chou *et al.*, 1983). As a result, the primitive unit of β structure is not the individual β strand but the β strand pair, which can be hydrogen bonded in either parallel or anti-parallel arrangement with close to optimal geometry and dipole

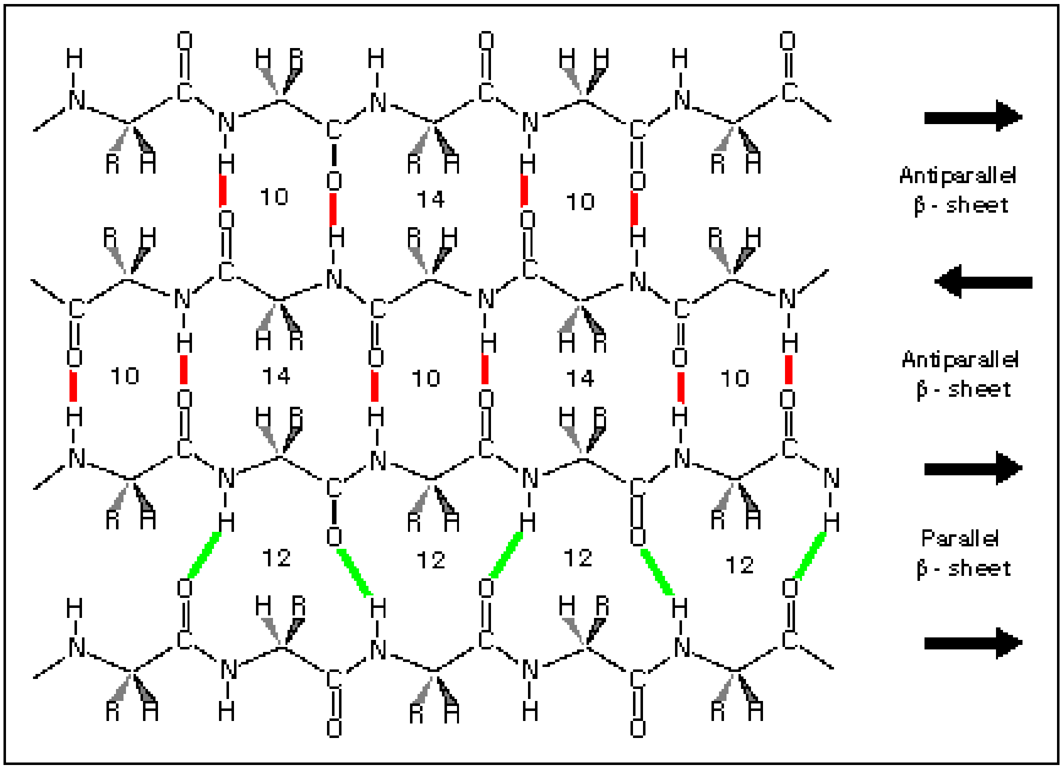

Figure 2.8 Showing the hydrogen bonding patterns in parallel and anti-parallel β- sheet structures. The direction of polypeptide backbone is marked with arrows and hydrogen bonds are shown with solid lines. The figure illustrates narrow and wide pairing of H-bonds and the side-chain alternation above and below the plane in anti-parallel β-sheet and evenly spaced but alternately slanting H-bonds in parallel β-sheet.

moments of the strands aligned favorably. Adjacent strands can be either parallel or anti-parallel (figure 2.8), and the stereochemistries of the strand in both the cases are slightly different. For anti-parallel β sheet, the relationship between adjacent strands is a two-fold axis perpendicular to the sheet, with the H-bonds perpendicular to the strands and alternating between a closely spaced pair and a widely spaced pair. In parallel β sheet, the H-bonds are evenly spaced but alternatively slant forward and backward, and the relationship between adjacent strand is a translation. The side chains on β strands extend approximately perpendicular to the plane of H-bonding. Along the strand they alternate from one side to the other, but on adjacent strand they are in register. For anti-parallel β sheets typically one side is buried in the interior and the other side is exposed to solvent, so that the amino acid types tend to alternate hydrophobic and hydrophilic. Parallel sheets, on the other hand, are usually buried on both sides, so their central sequences are highly hydrophobic, and hydrophilics concentrate at the ends. For both types of β structure, edge strands can be much more hydrophilic than central strands (Fasman, 1989).

Distinguishing these characteristic patterns can be of some help in secondary structure prediction and is clearly important for working towards probable tertiary structures. The usefulness of this results from a strong tendency for β sheets to be either pure parallel or pure anti-parallel. Mixed sheets occur, but not at anything like random expectation. More efforts are usually needed for prediction of the sheets.

The most prevalent local disruption in a sheet is the β bulge (Richardson *et al*., 1978). A β bulge can be thought of as an insertion of an extra residue into one strand, so that between a pair of H-bonds there is one residue on the normal strand but two residues on the bulged strand. Bulges are common in anti-parallel β structure but rare in parallel β. Usually they are located between a close pair of H-bonds rather than a wide pair. The extra residue puts the hydrophobic-hydrophilic side-chain alternation out of register across the bulge, an effect that is sometimes recognizable in the sequence. To accommodate the surrounding H-bond pattern, usually one of the two bulge- strand residues stays close to normal β-conformation while the other is close either to α-helical conformation (a "classical" bulge) or close to left-handed 3_{10} conformation (a "G-1" bulge). The single residue on the opposite strand is usually near polyproline conformation in order to match greatly accentuated right-handed twist produced by a β bulge. Bulges can mitigate the damage done by single residue insertion or deletions in β strands, at least when they occur near an end or an edge of the β sheet (Chan *et al*., 1993).

2.8 Non-repetitive Structure: Turns, Connections and Compact Loops

The secondary structures (described above) are one in which the φ,ψ angles repeats for each consecutive residues. Large portions of protein structure, however, are made up of well-ordered but nonrepeating conformations. These have often been referred to as "coil" or even "random coil", which unfortunately has connotation of disordered, mobile, unfolded chain. Nearly one third of the residues of globular proteins are involved in tight turns that reverse the direction of polypeptide chains at the surfaces of the molecules and make possible overall globular structure. Turns have also been implicated in molecular recognition (Rose *et al*., 1985) and in protein folding. Because of their prevalence, these reverse turns or loops are frequently classified as a third type of secondary structure.

Various types of reverse turns occur, involving different numbers of residues and depending upon which type of secondary structure they link. The best characterized are the β hairpins that link adjacent strands in antiparallel β-sheet. If only one residue is not involved in the H-bonding pattern of the sheet, there is a γ –turn, of which two types are possible. This very tight turn requires unfavorable geometry for the adjacent hydrogen bond of the β-sheet and unusual values of φ,ψ in the central residue of the turn. More common are β turns, in which two residues are not involved in the hydrogen bonding of the β-sheet; the two residues on either side of the non-hydrogen-bonded residues are included in the β turn, which, therefore, defined by four residues at the positions designated *i* to *i+3*. The existence of three ideal β turns, designated as types I, II and III, was predicted by Venkatachalam (1968) on the basis of allowed polypeptide geometry with planar *trans* peptide bonds. Mirror images of backbone -but not the side chains, occur in variants I', II', III'. There have been a lot of efforts to classify the β turns and loops in general. Themean dihedral angles for γ turns and β-turn types are tabulated in Table 2. Loops have been analyzed and classified according to various structural properties and relationships, amog them main chain conformation, size, inter C_α distances, hydrogen bonding patterns, orientation, and type of secondary structure flanking the loop (Donate *et al.*, 1996 and references within). A recent automated classification of conformational clusters and consensus sequences for the protein loops have been derived from a non-redundant data set by computational analysis (Oliva *et al.*, 1997).

Turn	Ramachandran	Mean dihedral angles[b]			
type	nomenclature[a]	φ(*i*+1)	ψ(*i*+1)	φ(*i*+2)	ψ(*i*+2)
γ turn[c]					
Classical		70 to 85	-60 to -70		
Inverse		-70 to -85	60 to 70		
β turns					
I	$\alpha_R\alpha_R$	-64(-60)	-27(-30)	-90(-90)	-7(0)
I'	$\alpha_L\alpha_L$	55(60)	38(30)	78(90)	6(0)
II	$\beta_{\gamma L}$	-60(-60)	131(120)	84(80)	1(0)
II'	$\varepsilon\alpha_R$	60(60)	-126 (-120)	-91(-80)	1(0)
III[c]		-60	-30	-60	-30
III'[c]		60	30	60	30
IV		-61	10	-53	17
VIa1	$\beta\alpha_R$	-64 (-60)	142(120)	-93(-90)	5(0)
VIa2	$\beta\alpha_R$	-132 (-120)	139 (120)	-80(-60)	-10(0)
VIb	ββ	-135(-135)	131(135)	-76(-75)	157(160)

Table 2 Mean dihedral angles for γ turns and β-turn types derived from Crighton (1993) and Hutchinson *et al*., (1994).

[a] Ramachandran nomenclature for turn types as in Wilmot and Thornton (1990). The nomenclature describes the region of the Ramachandran plot occupied by residues i+1 and i+2 of the turn.

[b] The idealized φ, ψ values as determined by Lewis *et al*.(1973) are given in the parenthesis after the averaged values determined from dataset of Thornton *et al*., (1994).

[c] Values taken from Crighton T.E. (1983).

Loops have been classified into five types (α-α, β-β links, β-β hairpins, α-β and β-α) according to the secondary structures they embrace and in to total 56 classes (9 α-α, 11 β-β links, 14 β-β hairpins, 13 α-β and 9 β-α) were identified with consensus Ramachandran angles in the loops and consensus sequence patterns for each class. However, still the amino acid sequences of the loop region do not provide a fingerprint that can be used to identify the presence of a loop of a particular conformation anywhere in a protein sequence. However, if the position and nature of the neighboring β-strands and helices are known or suspected on the basis of the known three-dimensional structure of a homologue, or form a reliable secondary structure prediction, then the particular conformation of the connecting peptide may be identified by comparison with sequence templates or patterns that characterize loop classes. These can be useful in comparative modeling (Greer, 1980; Thornton *et al*., 1988; Sibanda *et al*, 1989; Overington *et al*., 1990; Topham *et al*., 1993) as well as in suggesting conformations of super-secondary motifs where a satisfactory structure prediction has been performed (Donate *et al*., 1996).

2.9 Amino acid Residues

The 20 different amino acids possess a variety of chemical properties. This variety is greatly enhanced when the various groups are combined in various sequences in a single molecule, which gives a protein properties far beyond those of simpler molecules. The chemical properties of a protein molecule are far more complex than the sum of the properties of its constituent amino acids but understanding side chain properties can be a good beginning towards it. Side chains are divided and discussed briefly according its various properties in **Appendix A**. However, it should be mentioned that residues in biologically active proteins may have chemical and physical properties very different from those described. Amino acids can be classification as shown in the Venn diagram of Figure 2.9 (Taylor, 1986a,b). Thus we have seen that the amino acid properties and their preferences of staying in a particular kind of environment is the one that determine the protein secondary structures and may be the tertiary structure.

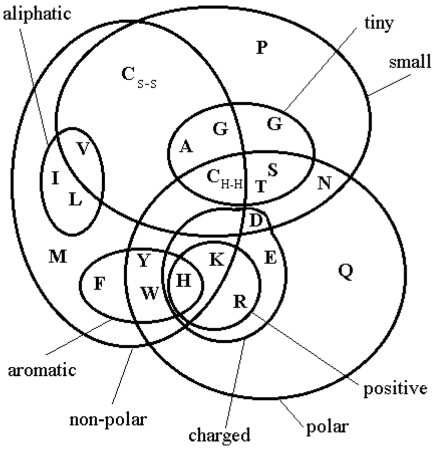

Figure 2.9 Venn diagram showing the classification of amino acids according to physical properties of their side chains (Taylor, 1986a,b).

Assemblies of a number of secondary structure elements, including the connecting loops, that have been observed often enough that they are becoming recognized as another level of structure, termed as super-secondary structures or motifs. These structures are a higher level of structure than secondary structure but does not constitute entire structural domains. However, description of super-secondary structures and structural domains is out of the scope of this thesis. For the description of recurring super-secondary structures please see Rossmann and Argos, (1981); Branden and Tooze., (1991) and Sowdhamini *et al*., (1992) etc. For description of protein structural domains and their classification please see Sowdhamini *et al*., (1995); Sternberg *et al*., (1995); Sowdhamini *et al*., (1996) Orengo *et al*., (1997); and Holm and Sander, (1998).