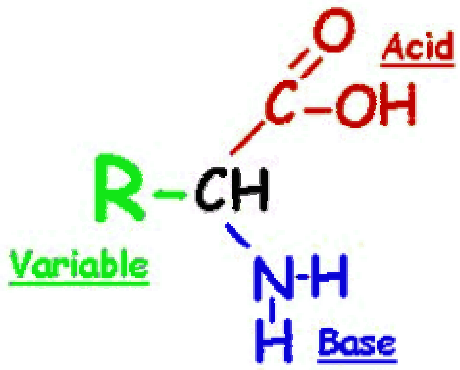
What are the intermolecular forces that lead to protein's compact folding?

We will go over these in much more detail, but for now, let's list them, with a very short explanation.

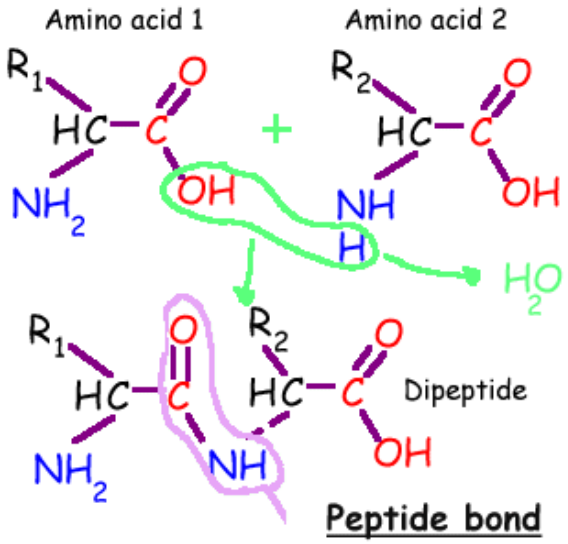
- **<u>Remember</u>** this is an amino acid



A theoretical amino acid

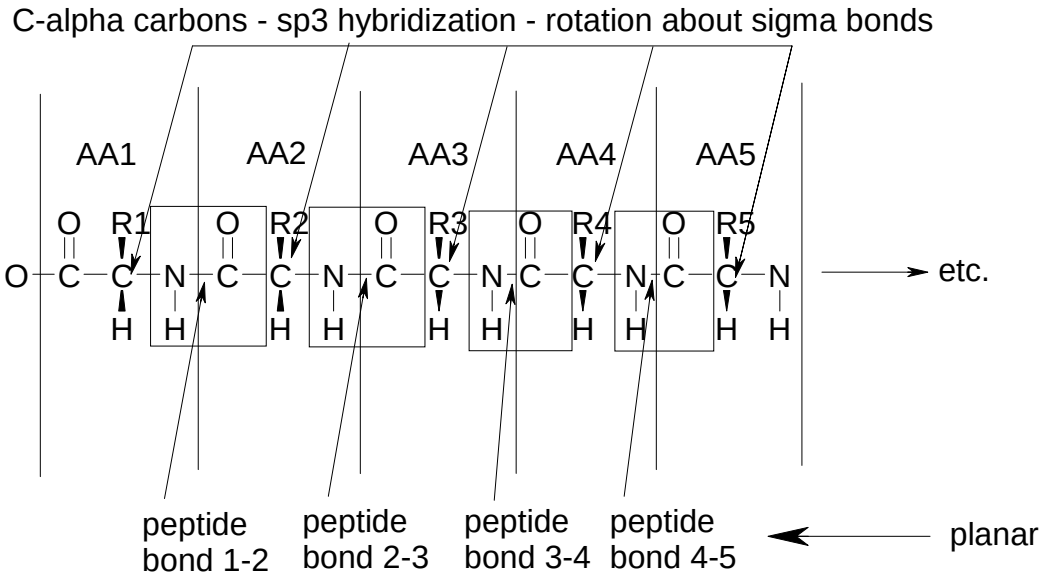
-

- This is the way that the peptide bond is made from two amino acids



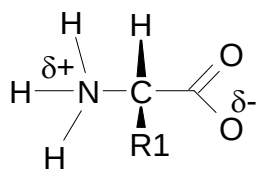
•

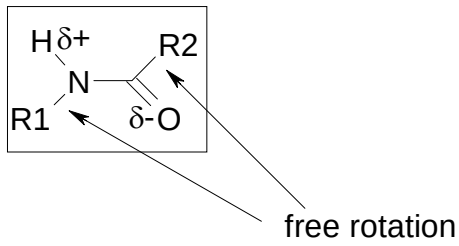
From: http://www.iacr.bbsrc.ac.uk/notebook/courses/guide/aa.htm
Beginner's Guide to Molecular Biology



Protein or polypeptide

Remember, the amino acids (AAs) and the peptide bonds have the structures:





L-Amino Acid — Peptide Bond
Central carbon asymmetric — Planar

pK α-Carboxyl group = 2 – 2.4
pK α–Amino group = 9 – 9.7 (10.4-10.8 for Proline, threonine and cysteine)

- acid-base dissociation and pH

$HA <-> H^+ + A^-$ $Ka = [H^+][A^-]/[HA]$ $pH = pKa + \log([A^-]/[HA])$

where $pX = -\log(X)$

- **<u>The amino acid sequence of a polypeptide determines its conformation in solution</u>**

- If proteins are unfolded (denatured) by temperature, pH, ionic strength or certain solvents, they often return to their native structure.

– THIS IS NOT ALWAYS THE CASE!

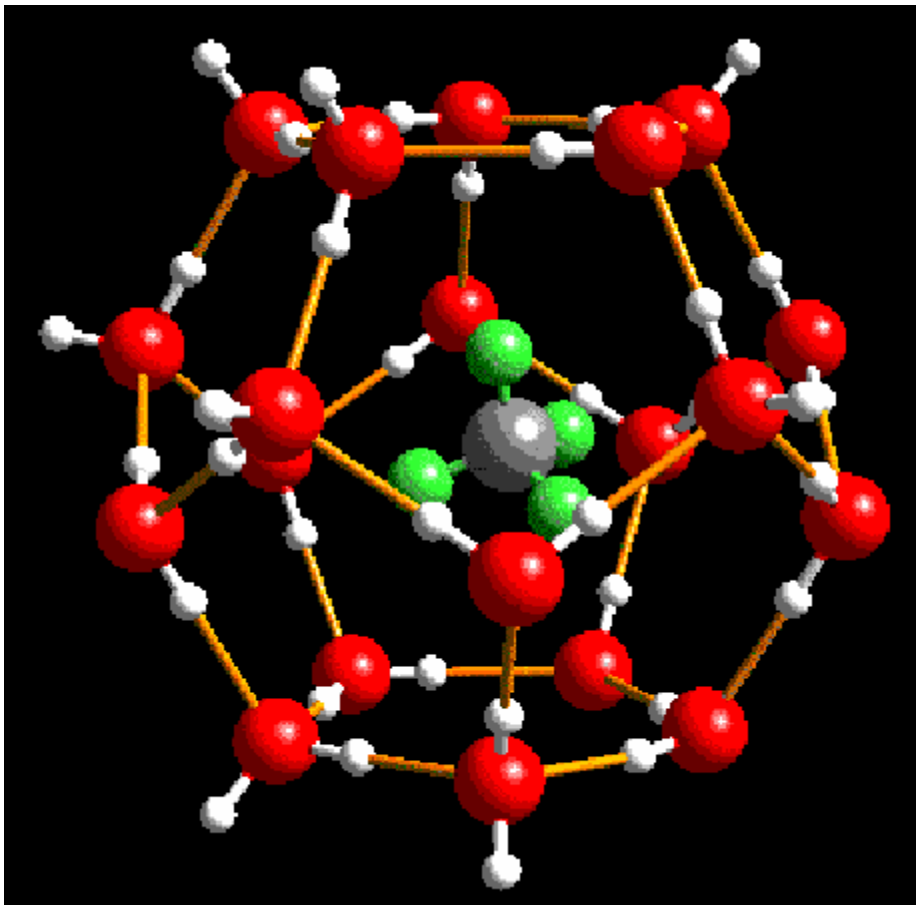
water1.pdf

solubH2Or.pdf

Noncovalent interactions are primarily responsible for maintaining the protein's conformation (minimize free energy: $\Delta G = \Delta H -$

TΔS). Three principle interactions are (we will discuss them more later):

Hydrophobic interactions – remove nonpolar groups from water to interior; primarily entropy driven – structured water around nonpolar surfaces (?) – This interaction increases as the temperature increases.

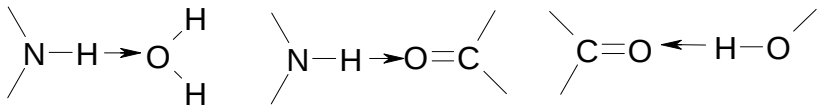


[Hydrophobic Effect_all.pdf]

van der Waals interactions between hydrophobic groups

VanDerWaals.pdf

H-bonds especially between C=O and N-H groups of each peptide bond, and the electronegative groups of polar side-chains. Other examples:



residue - water residue-residue residue-residue
residue-water

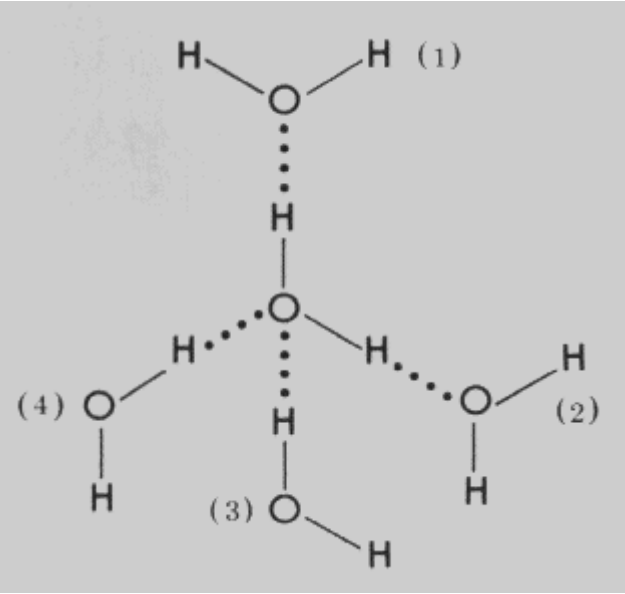
internal H-bonds and ionic bonds do not contribute to stability of folding, because these interactions also take place with water. – Insures correct folding - But if residues fold into the interior of the protein, excluding water, and these interactions are not present with these groups in the protein, this destabilizes the folded conformation.

What are "hydrogen bonds"?

Hydrogen Bonds

Polar molecules, such as water molecules, have a weak, partial **negative charge** at one region of the molecule (the oxygen atom in water) and a partial **positive charge** elsewhere (the hydrogen atoms in water).

Thus when water molecules are close together, their positive and negative regions are attracted to the oppositely-charged regions of nearby molecules. The force of attraction, shown here as a dotted line, is called a



hydrogen bond. Each water molecule is hydrogen bonded to four others.

The hydrogen bonds that form between water molecules account for some of the essential — and unique — properties of water.

- The attraction created by hydrogen bonds keeps water liquid over a wider range of temperature than is found for any other molecule its size.
- The energy required to break multiple hydrogen bonds causes water to have a high heat of vaporization; that is, a large amount of energy is needed to convert liquid water, where the molecules are attracted through their hydrogen bonds, to water vapor, where they are not.

Two outcomes of this:

- The evaporation of sweat, used by many mammals to cool themselves, achieves this by the large amount of heat needed to break the hydrogen bonds between water molecules.
- Moderating temperature shifts in the ecosystem (which is why the climate is more moderate near large bodies of water like the ocean)

The hydrogen bond has only 5% or so of the strength of a covalent bond. However, when many hydrogen bonds can form between two molecules (or parts of the same molecule), the resulting union can be sufficiently strong as to be quite stable.

Multiple hydrogen bonds

- hold the two strands of the DNA double helix together
- hold polypeptides together in such secondary structures as the alpha helix and the beta conformation;

- help enzymes bind to their substrate;
- help antibodies bind to their antigen
- help transcription factors bind to each other;
- help transcription factors bind to DNA

Ionic bonds: between ionized side-chains of opposite charge

IonsandDebyeHuck.pdf

Comparison of Covalent and non-covalent bonds found in proteins:

Bond	Bond Energy (kcal/mol)	
Covalent:		
C-C	83	
S-S	50	
Noncovalent:		
H-bond	3-7	
Ionic bond	3-7	
Hydrophobic association	3-5	free energy to move group into water (T sensitive)
Van der Waals bond	1-2	

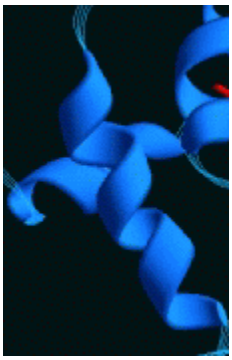
kcal/mol = 4.18 Joule/mol =>

- Most non-polar groups inside – most polar groups outside (contact with water) of folded protein structure – (trans-membrane sections of proteins is the opposite)
- Disulfide bonds (-S-S-) form between Cysteine residues
- The two most common structures in proteins are: **α helix** and **β sheet**.

Secondary Structure

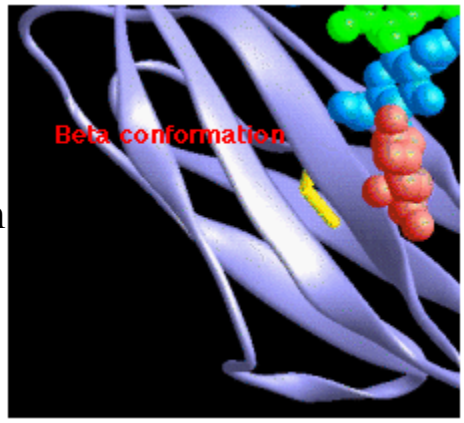
Most proteins contain one or more stretches of amino acids that take on a characteristic structure in 3-D space. The most common of these are the **alpha helix** and the **beta conformation**.

Alpha Helix



- The R groups of the amino acids all extend to the outside.
- The helix makes a complete turn every 3.6 amino acids.
- The helix is right-handed; it twists in a clockwise direction.
- The carbonyl group (-C=O) of each peptide bond extends parallel to the axis of the helix and points directly at the -N-H group of the peptide bond 4 amino acids below it in the helix. A hydrogen bond forms between them [**-N-H·····O=C-**] .

Beta Conformation



- consists of pairs of chains lying side-by-side and
- stabilized by hydrogen bonds between the carbonyl oxygen atom on one chain and the -NH group on the adjacent chain.
- The chains are often "anti-parallel"; the N-terminal to C-terminal direction of one being the reverse of the other.

Properties of globular protein sections

❖ **α helix**:

<u>Some facts about the α helix</u>:

- A **helical turn** spans **0.54 nanometers** (5.4 Angstroms)
- These **H-bonds** are **linear** – therefore very stable
- α helix is prevented by presence of **proline** sidechains, and by consecutive residues with branching sidechains (e.g. Val, Ilu and Thr), or by two consecutive residues with ionized sidechains with the same charge.
- α helix is **flexible** and **elastic**

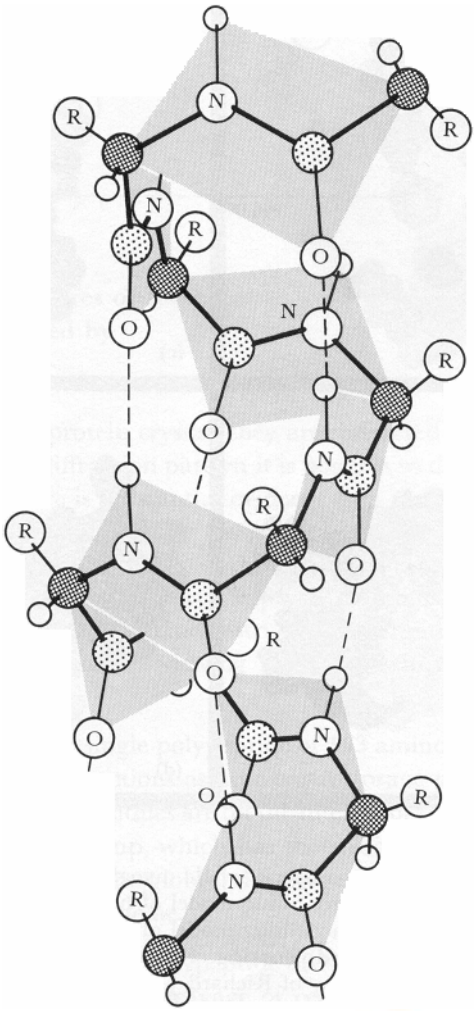


Figure 2-31
Models of a right-handed α helix. (A) Only the α carbon atoms are shown on a helical ribbon; (B) only the backbone nitrogen (N), α carbon (C_α), and carbonyl carbon (C) atoms are shown; (C) entire helix. Hydrogen bonds (denoted in part C by red dots) between NH and CO groups stabilize the helix.

- **31%** of AAs in all globular proteins are in an α helix;
- the helix is ***right-handed***
- ***intrastrand*** H-bonding **parallel** to helical axis
- AA residues ***outside*** of helix, and tilt toward the N-H end of the helix
- Classical α helix has ϕ=-62 ° and ψ=-41°.
- But in proteins ϕ=-57 ° and ψ=-47°; this is more **favorable** for N-H to C=O internal H-bonding **and** external H-bonding to H_2O or other donors.
- The **helix forming propensity** of different AAs is very **hard to determine**. It appears that the **variance** for different AAs is greater than previously thought (helix-coil measurements of polypeptides).
- Essentially only the "**trans**" conformation of the **peptide bond** structure is allowed.
- The only AA that is **not compatible** with the α helix, is "**proline**"
- Polar AA residues, such as serine, threonine, Aspartic acid and Asparagine can H-bond with the α helix, interfering with the α helix structure.
- **Amphipathic** α helices have hydrophobic groups on one side of the helix, and hydrophilic regions on the other side (see the *helix wheel representation*). This is a common structure for a helix to "fit" together with other protein regions (hydrophobic side forming the "inside" of the structure), and yet offer a hydrophilic surface to the "solvent" - ***or visa-versa***. for membrane proteins
- The **stability** of the α helix is **marginal** at normal conditions (can be "denatured" easily)
- The **kinetics** of forming an α helix is **rapid** (10^{-5} to 10^{-7} seconds)
- **Difficult** to form the **starting** α helical structure with the first few AA residues.

- no intramolecular H-bonding until at least 4 residues have the helical structure
- entropy loss is therefore high without an energetic compensation
- dipoles of the peptide bonds are parallel, and until the head-to-tail positioning of the peptide dipoles are formed, it is energetically unfavorable

Helical wheel representation for an α helix: projection down the helical axis to show the positions of the side groups on the helical structure.

A useful method for visualizing the properties of a protein helix is to display the sequence on a helical wheel diagram. This diagram is based on a five-turn helix (18 residues). Adjacent residues in the sequence are 100^0 apart on the wheel and a residue occurs every 20°. The sequence is written onto the wheel in the order shown. As an exercise, display the following sequence on a helical wheel and explain what information this representation provides. Ala-Phe-Asp-Lys-Met- Ile-Glu-Asn-Leu-Gln-Arg-Leu-Trp-Ser-Glu-Phe-Leu-Gln.

Analysis of individual sequences

Helical wheels, amphipatic helices and coiled-coils

For many years experimental work on protein structure was concerned almost exclusively with globular proteins. Consequently much of our knowledge now is biased towards this class of proteins and our knowledge on structural proteins or transmembrane proteins is only slowly increasing. Many structural proteins contain **amphipatic helices**, which consist of hydrophobic, non-polar residues on one side of the helical cylinder and hydrophilic and polar residues on the other side, resulting in a hydrophobic moment. In this way, they aggregate with other hydrophobic surfaces and serve for example as pores or channels in the cell membrane. A **helical wheel** (see Figure 2) provides a possibility to visualize an amphipatic helix.

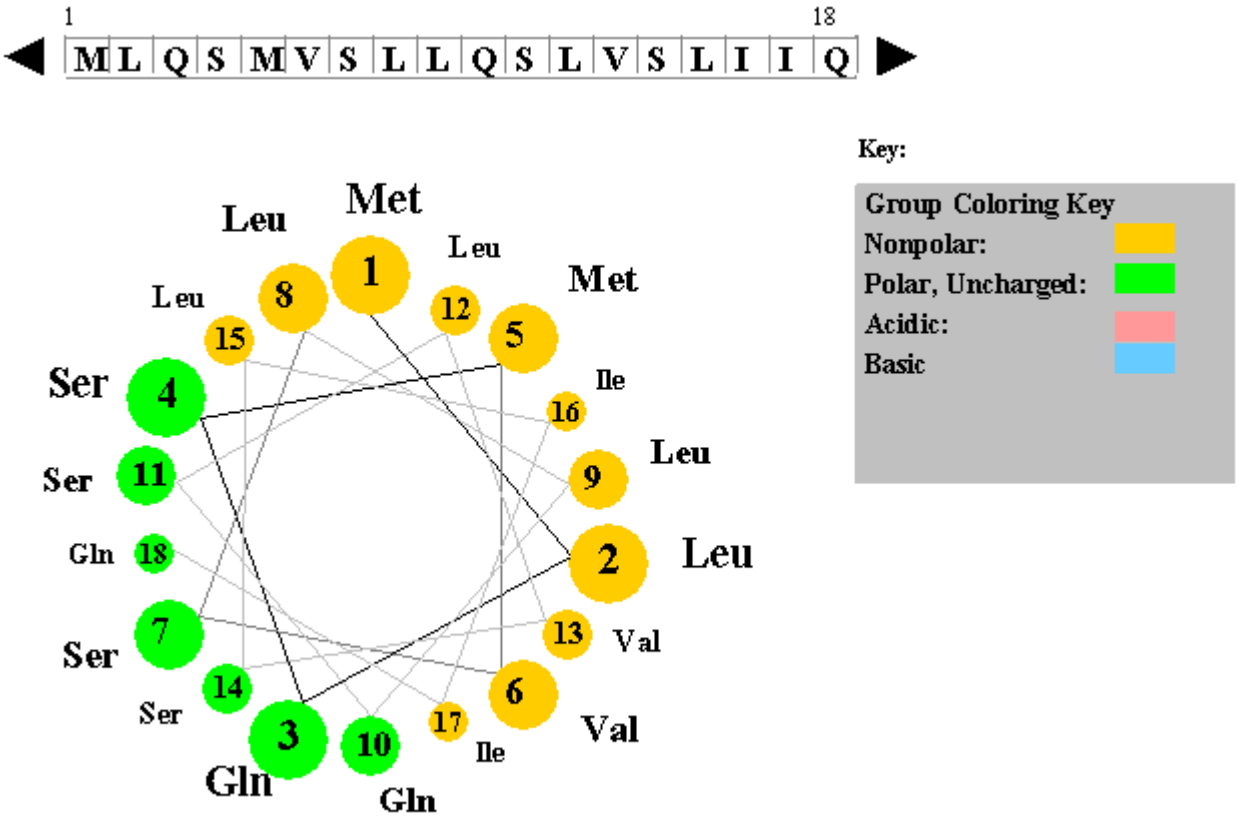


Figure 2: Helical wheel representation of an amino acid sequence. The amino acid side chains are projected down the axis (the axis of an alpha helix, orthogonal to the paper plane). As an ideal alpha helix consists of 3.6 residues per complete turn, the angle between two residues is chosen to be 100 degrees and thus there exists a periodicity after five turns and 18 residues. The figure is a snapshot of a Java Applet written by Edward K. O'Neil and Charles M. Grisham (University of Virginia in Charlottesville, Virginia). The applet is accessible at http://cti.itc.Virginia.EDU/~cmg/Demo/wheel/wheelApp.html.

Some amphipatic helices are arranged as intertwined helices that are termed a **coiled-coils** or **super-helices**. Generally, the sequence of an alpha helix that participates in a coiled-coil region will display a periodicity with a repeated unit of length 7 amino acids, which is called a heptad repeat. Denote those 7 positions by a through g, then position a and d are hydrophobic and define an apolar stripe, while there exist electrostatic interactions between residues at positions e and g. Prediction methods for coiled-coil regions ([3][4]) are making use of these preferences. The so-called **Leucine Zipper** dimerization domains often occur together with DNA-binding domains in regulatory proteins, e.g. eukariotic transcription factors. They form a parallel coiled coil of alpha-helices from two polypeptides chains holding them together. There are only few heptads, usually three to six, whereat the first residue in a heptad is Leucine.

Comments are very welcome.
luz@molgen.mpg.de

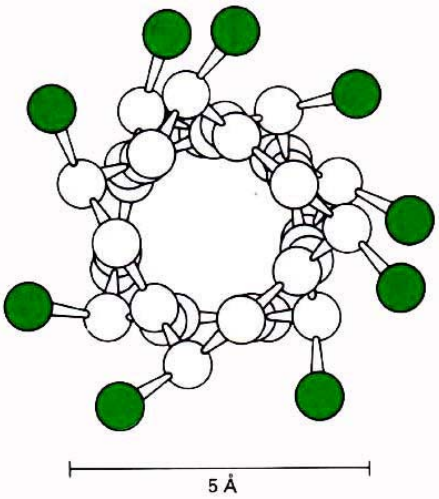


Figure 2-32
Cross-sectional view of an α helix. Note that the side chains (shown in green) are on the outside of the helix. The van der Waals radii of the atoms are larger than shown here; hence, there is actually almost no free space inside the helix.

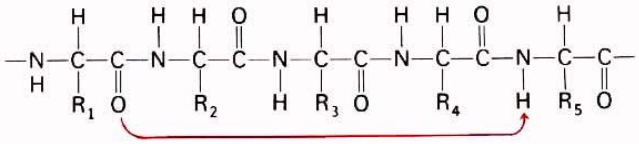


Figure 2-33
In the α helix, the CO group of residue n is hydrogen bonded to the NH group of residue $(n + 4)$.

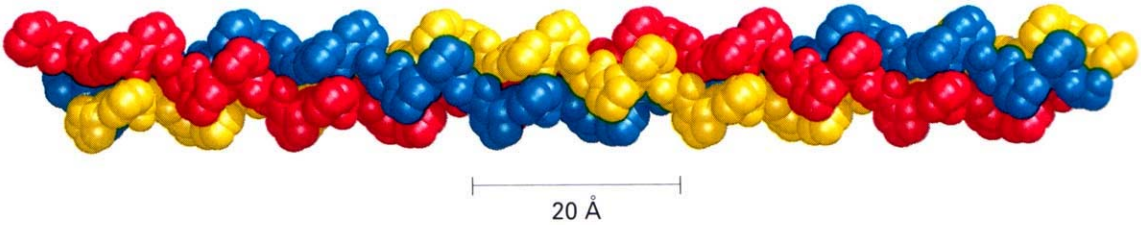


Figure 2-40
Space-filling model of the collagen triple helix. The three strands are shown in different colors. [Drawn from coordinates reported in K. Okuyama, K. Okuyama, S. Arnott, M. Takayanagi, and M. Kakudo. *J. Mol. Biol.* 152(1981):427.]

β-strands:

<u>Some facts about the β sheet</u>:

- Side-by-side polypeptides connected by H-bonds
- Two- residue repeat distance is 0.70 nm
- Adjacent polypeptides generally run antiparallel to each other.
- Readily forms with polypeptides with repeating sequences of AAs with compact sidechains (e.g. Gly-Ser-Gly-Ala-Gly-Ala)n.
- β sheet is **flexible** but **inelastic**.
-

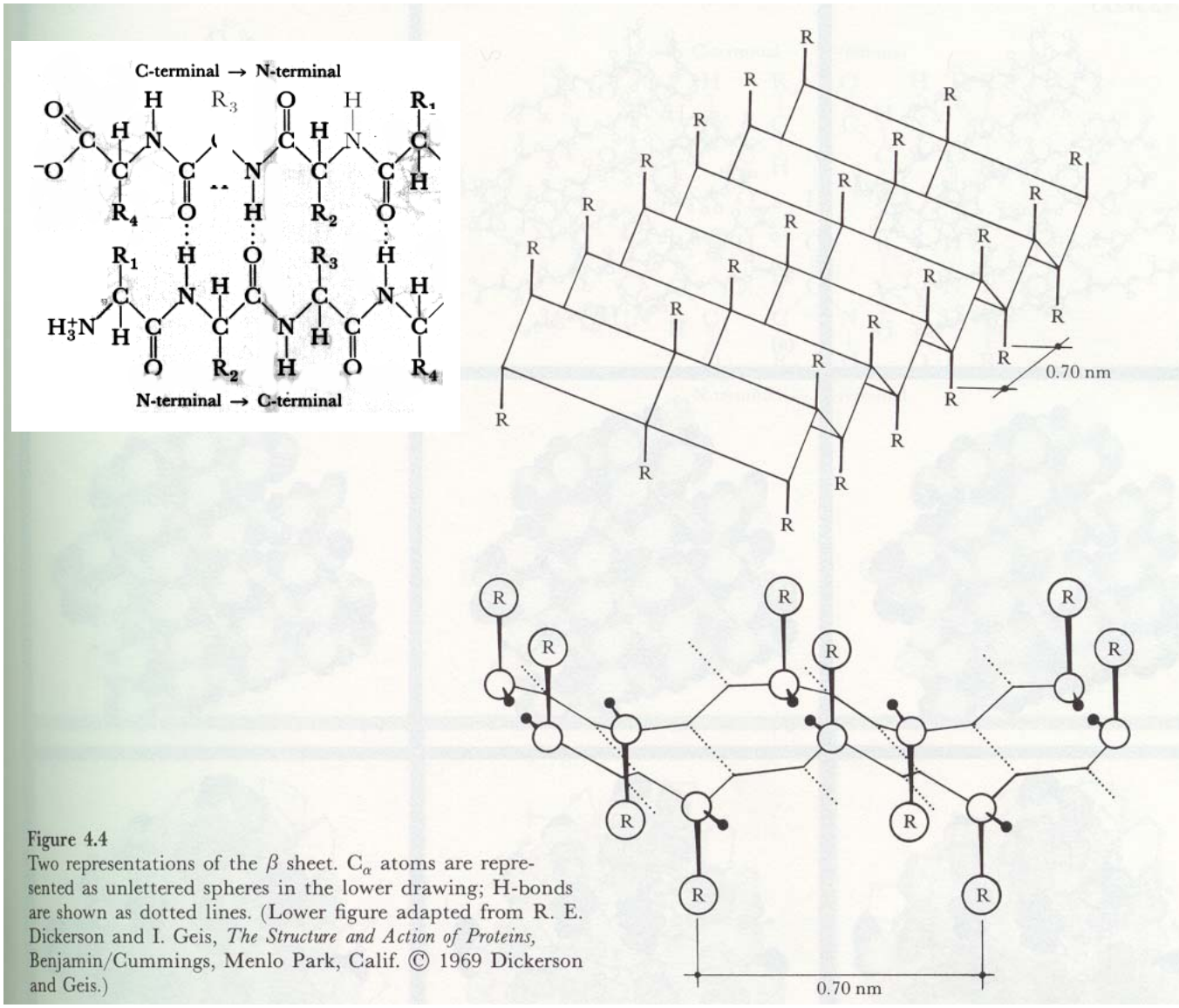


Figure 4.4
Two representations of the β sheet. C_α atoms are represented as unlettered spheres in the lower drawing; H-bonds are shown as dotted lines. (Lower figure adapted from R. E. Dickerson and I. Geis, *The Structure and Action of Proteins*, Benjamin/Cummings, Menlo Park, Calif. © 1969 Dickerson and Geis.)

- **28%** of AAs in all globular proteins are in an α helix
- ***interstrand* H-bonding laterally** between N-H and C=O groups of the peptide backbone
- usually slightly **twisted** (0 – 30 %) between the β strands – the twisting depends on the values of ϕ and ψ
- parameters often **deviate from ideality** – vary widely
- β strands can run **parallel** or **antiparallel**.

Antiparallel β-strand

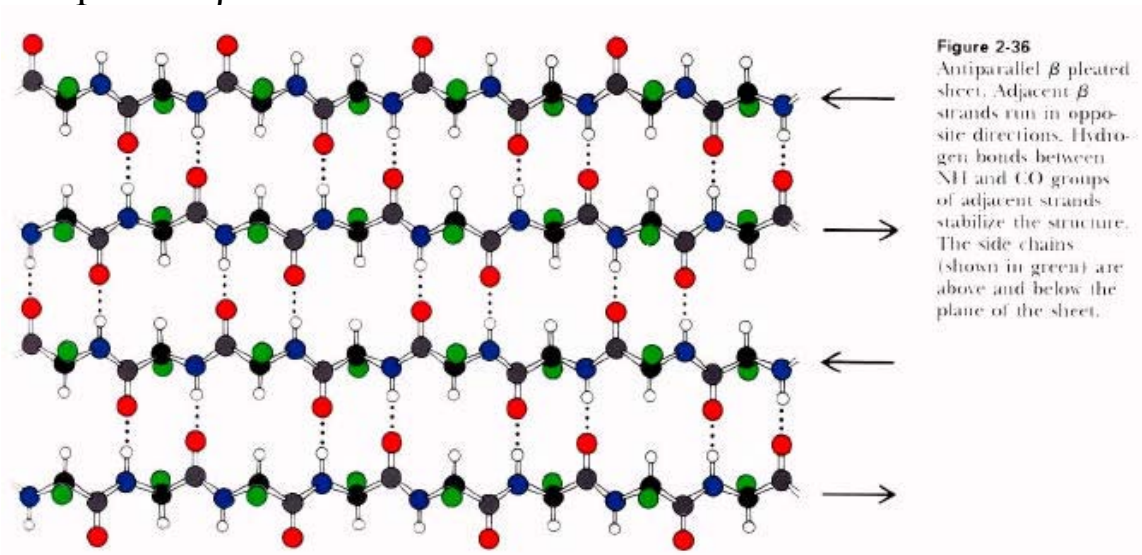


Figure 2-36
Antiparallel β pleated sheet. Adjacent β strands run in opposite directions. Hydrogen bonds between NH and CO groups of adjacent strands stabilize the structure. The side chains (shown in green) are above and below the plane of the sheet.

- AA residues protrude alternatively up and down from the "plane" of the pleated along the peptide sequence – no interference between these neighboring groups.
- Little is known about the thermodynamics or kinetics of formation of β sheets – no good peptide models, because they tend to aggregate. The only kinetic information is that intramolecular β sheet formation takes place in about 10^{-2} seconds – much slower than a helical regions..

❖ Reverse turns

- Reverse the direction of the strands

The importance of beta-turns in proteins

Beta-turns are estimated to comprise between a quarter and a third of all residues in proteins and the formation of these turns has long been thought to be an important early step in the protein folding pathway. Beta-turns are commonly found to link two strands of anti-parallel beta-sheet, forming a beta-hairpin, and such secondary structure elements have been postulated as possible nucleation sites for the protein folding pathway. A basic nomenclature for hairpin turns is shown in **figure 3.1**, where the loop residues are labeled L1, L2 and so on, and residues in the N-terminal strand are labeled -B1, -B2, etc. from the turn, and the C-terminal residues +B1, +B2, etc.

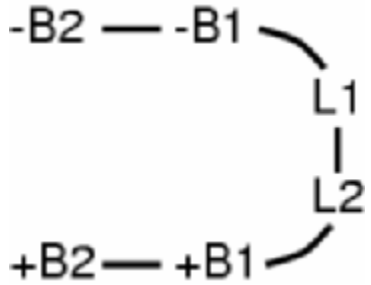
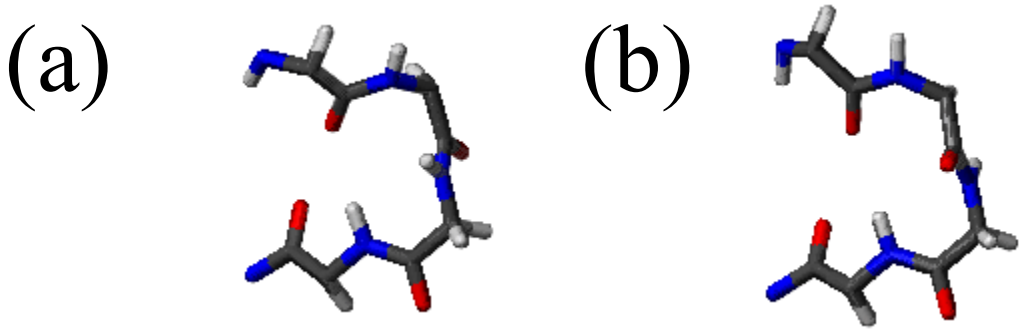


Figure 3.1: Nomenclature for residues in hairpin beta-turns



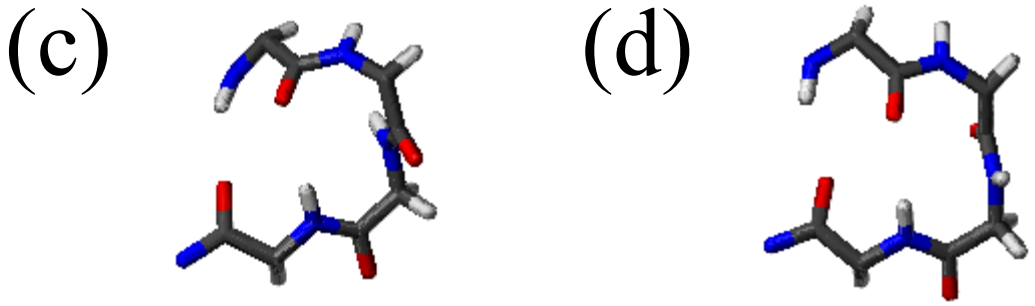


Figure 3.2: Peptide backbone conformation for (a) type I, (b) type II, (c) type I', and (d) type II' turns.

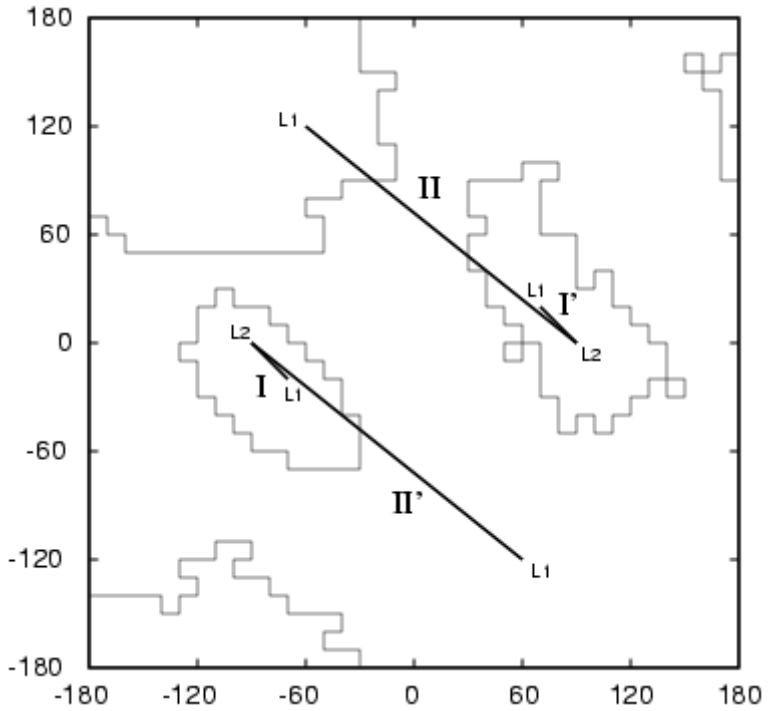


Figure 3.3: Ramachandran plot showing phi and psi angles for residues L1 and L2 in type I, II, I' and II' turns.

Beta-turns can be classified according to the number of residues in the loop and by far the most common is the two residue turn, followed in frequency by three, four and five residue turns. Turns are subdivided into a number of groups based on the backbone angles phi and psi of the residues in the loop. The most common two residue turns in globular proteins are type I and type II turns (**figure 3.2(a)** and **(b)**). In contrast, the preferred turns in beta-hairpins are their mirror images, types I' and II' (**figure 3.2(c)** and

(d)). The phi,psi angles associated with these four types of turn are shown in **figure 3.3**. Sibanda and Thornton concluded that the hairpin preference for types I' and II' is a consequence of compatibility of the twist of the turn with the natural right-handed twist of the beta-sheet. In contrast, types I and II favour rapid divergence of the strands. Therefore type I and II turns in beta-hairpins generally tend to be more than two residues or bulge turns such as in model systems derived from tendamistat and ubiquitin, as described below.

The number of residues in a turn is defined as the number of residues not in beta-strand or alpha-helix and there are two ways to determine this. Under IUPAC convention rule 6.3, a residue is said to be in alpha-helix or beta-strand if either its NH or CO are involved in the appropriate hydrogen bond. Alternatively, IUPAC rule 6.2 determines that a residue is in alpha-helix or beta-sheet if its phi,psi angles correspond with those for regular secondary structure. In practice, the latter definition generally means that both NH and CO are involved in the appropriate hydrogen bonds (the Kabsch and Sander definition). Sibanda and Thornton therefore propose that two numbers should be used to represent turns based on these two definitions, so if both hydrogen bonds across the turn are formed, we have the set of turns 2:2, 3:3 and so on, and if only one hydrogen bond is formed we have the set of turns 2:4, 3:5 etc. The most common hairpin turns are 2:2 (two residues in the loop, and two hydrogen bonds between the immediately flanking residues); next come three residue loops, with only one inter-strand hydrogen bond (3:5); and then those with four loop residues and two inter-strand hydrogen bonds (4:4).

- **General**:
  - **α helix and β strand** structures are **generally short**
    - α helix => 10-15 AAs

- β strand => 3-10 AAs

- The β strand is more extended than the α helix
  - **antiparallel β sheet:** 2 residues per turn & 3.4 Å per residue
  - **parallel β sheet:** 2 residues per turn & 3.4 Å per residue
  - **α helix**: 3.6 residues per turn & 1.5 Å per residue
- Where the AA-side chains are located in a protein structure depends on the natural environment of the protein.
- The **protein folding problem** (will cover later) is compounded by the enormously large conformational space available.