Problem Set 4 due today.

Reading for Lectures 22—24: PKT Chapter 8

DNA Packing for Eukaryotes:

The packing problem for the larger **eukaryotic genomes** is even more serious than for bacterial genomes. Take the human genome at $3\text{x}10^9$ bp, implying $\sqrt{\langle R^2 \rangle} \approx 6\sqrt{3\times10^9} = 3.3\times10^5 \text{ nm} = 0.3 \text{ mm}$ vs cell size at ~ 1--10μm.

Comment:

Part of the reason for the unusual stiffness of DNA is that, under physiological conditions, DNA carries a negative charge of -2e/bp due to ionizing of phosphate groups along the backbone:

$HM \leftrightarrow H^+ + M^-$, where M is the macromolecule and H^+ is the hydrogen ion (acidic).

Thus, law of mass action: $\dfrac{[H^+][M^-]}{[HM]} = K_d$.

If K_d is large enough, then essentially all the phosphates are ionized.

This equation is often written in logarithmic form ($\log_{10}$!!):

$$\log_{10} K_d = \log_{10}[H^+] + \log_{10}[M^-] - \log_{10}[HM].$$

$$pK \equiv -\log_{10}[K_d]$$

But,

$$pH \equiv -\log_{10}[H^+],$$

so $\log_{10}\dfrac{[M^-]}{[HM]} = pH - pK$, which is called the **Henderson-Hasselbalch** equation.

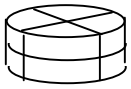
Anything below pH 7 is acidic; normal "physiological pH" (e.g., blood) is 7.3.

And, pK~1 for the phosphates.

It follows that $\dfrac{[M^-]}{[HM]} \sim 10^6$, i.e., these groups are fully ionized. (DNA is a strong acid.)

Here, the strategy is a heiarchical compactification: (see p. 26.2)

The DNA strand is wound twice around a cylindrical protein octamer called **histone**, which is coated with positive charges to bond the negatively charged DNA. The complex made up of the histone with its windings is called a **nucleosome**.

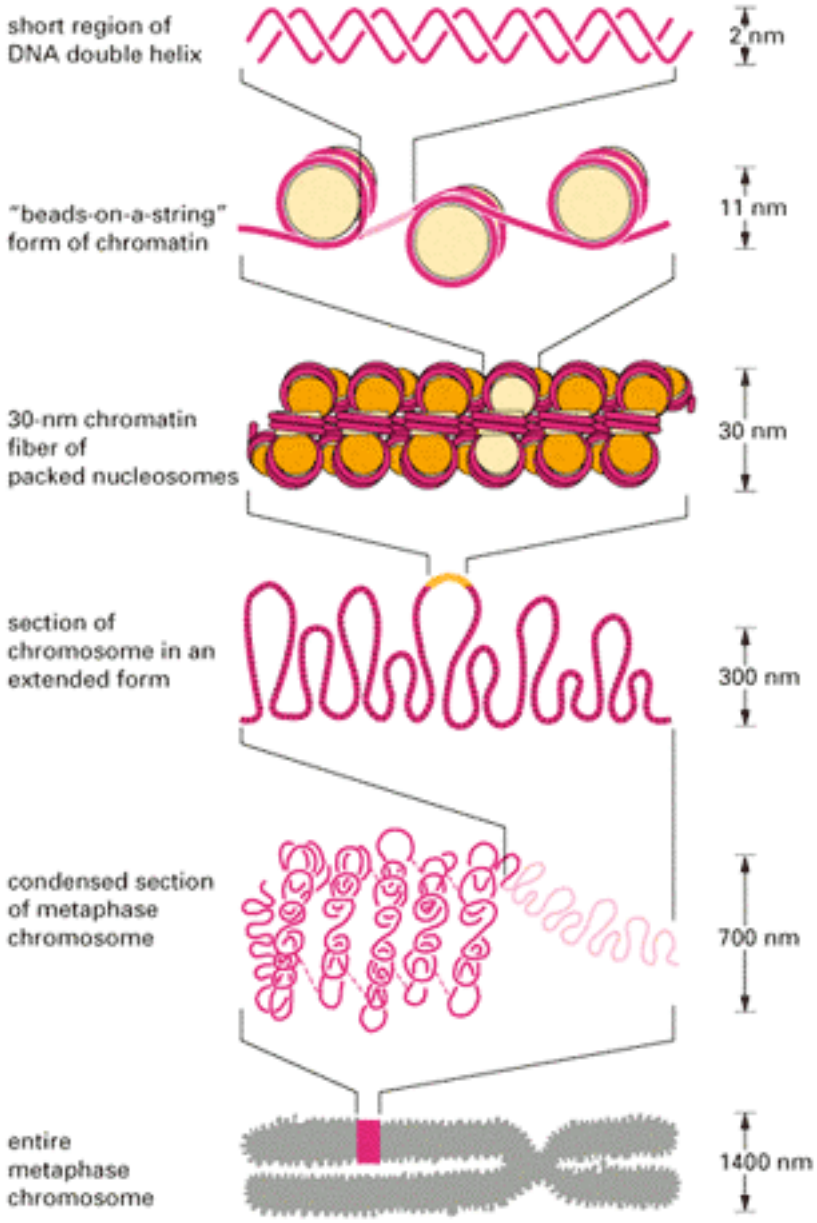


The diameter of the histone cylinder is about 7 nm, so the centers of the DNA helix are about 9 nm apart, and the circumference is ~28 nm. Thus, 1.8 turns constitutes 28x1.8x3~146 bp per histone. With a spacer between, the full repeat distance is about 200 bp.

DNA transcription to provide mRNA synthesis proceeds without unwrapping the histones.

At the next level of organization, the wound histones with their spacers are packed into chromatin fibers 30 nm in diameter, which form looped domains.

The looped domains are then organized into chromosomes. The human genome is packed into 23 chromosome pairs, each with the characteristic (asymmetrical) "X" structure.



http://library.thinkquest.org/C004535/chromosomes.html

2. Protein Folding: 1D information leads to 3D structure

The volume of an amino acid is roughly 0.2 nm^3 (see Problem 5), so the size is roughly 0.6 nm.
Take our typical protein of 300 aa's: Length~300x0.6=200 nm; Volume 60 nm^3.

Persistence length is 1—2 aa's, so $\left\langle R^2 \right\rangle = 2\xi_p L \sim 2\cdot(0.6)\cdot 200 \sim 240$ nm^2, i.e., $\sqrt{\left\langle R^2 \right\rangle} \sim 15$ nm.

However, the radius of a typical globular protein is about 2 nm (PKT, Table 1.1).

Compare with volume at $V = 300\cdot 0.2 = 60 = \frac{4\pi}{3}R^3 \Rightarrow R \sim 2.4\,\text{nm}$.

Conclusion: Typical protein in its functional ("native") state is a compact object, not a random coil.
Go back at this point and do the end of Lect. 24.
Much shorter objects than DNA (~300 aa's vs 10^6—10^9 bp's).
Completely different problem: You want to fold *reliably* into a *unique* structure.
Basic driving force to compactification is hydrophobicity.

Biologically relevant proteins are heteropolymers; however, the amino-acid sequence is anything but random. Indeed, the ordering of the amino acids determines the 3 structure.

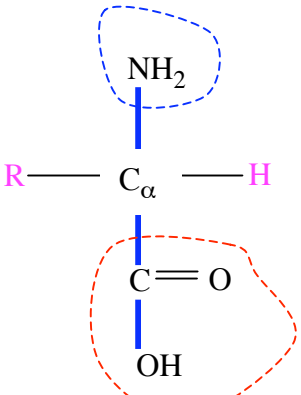
Heierarchical organization:

Primary structure: the amino-acid sequence (forces: covalent bonding)

- Secondary structure: α-helices and β-sheets: local organization (forces mostly H-bonds)
- Tertiary structure: Complete globular structures (charge interactions, sulfur bridges (cysteine), etc.)
- Quaternary structure: Associations of globular structures (distinct globular domains of a single long protein or associations of two or more distinct globular proteins)

Let's start at the beginning:

All but one amino acid (proline = Pro = P) has the generic form (Fischer):



Carboxyl group (or terminus)

--OH is ionized at pH 7 ($--O^- + H^+$) making a weak acid ("amino acid")

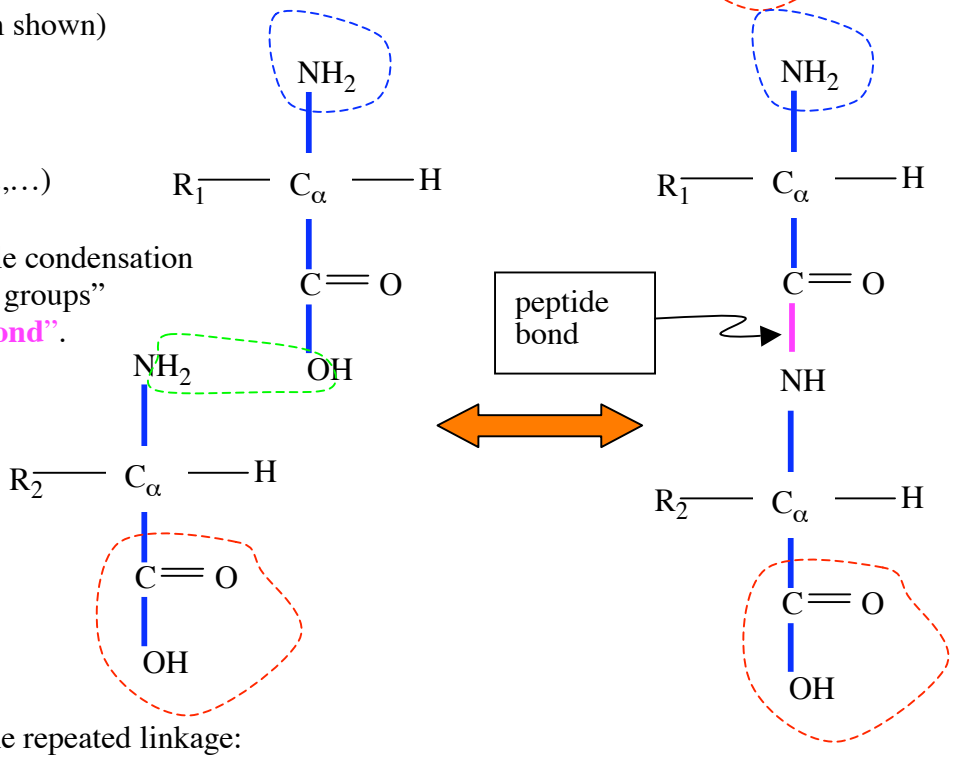
Amino group (or terminus) becomes $-NH_3^+ + OH^-$ at pH 7)

Central "alpha" carbon is generally chiral (ex: R=H Gly),
i.e., there is another form with R ⟷ H.
Nature uses exclsively the L-form shown)

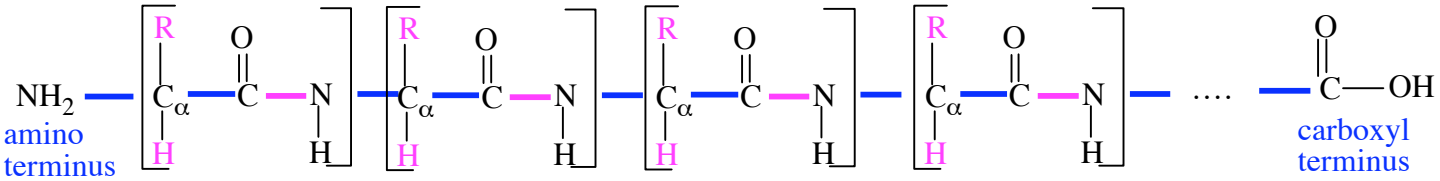
"backbone" follows **blue** bonds

R is "side chain" (contains C_β, C_γ,…)

Amino acids polymerize by simple condensation between the amino- and carboxyl groups" forming the so-called "**peptide bond**".

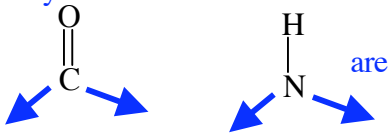
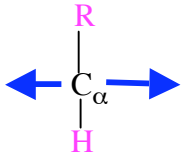


Thus, "backbone of polymer in the repeated linkage:



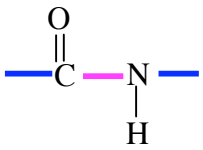
Geometry of the chain:

Bond lengths are fixed by covalent radii.

The local structures:  are planar. The  are tetrahedral.

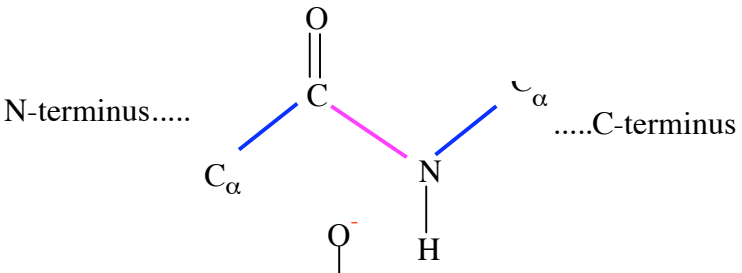
Therefore, the analysis of protein folding focuses on rotation about the N—C_α, C_α—C, and C—N bonds. **26.4**

However, there is an additional subtlety: The peptide bond

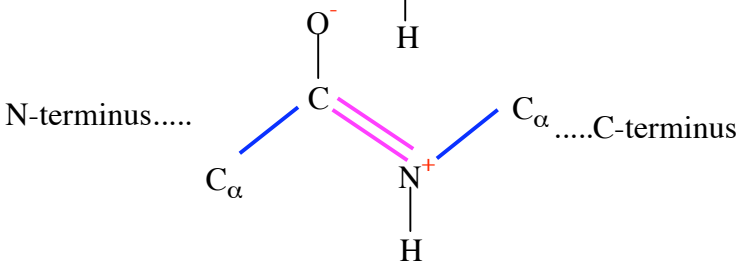


is partially a double bond due to a resonance:

(~60%)



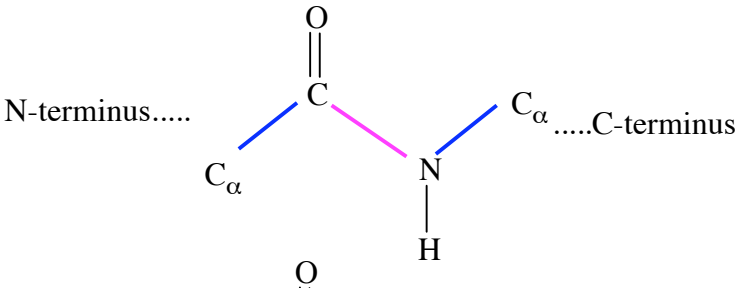
(~40%)



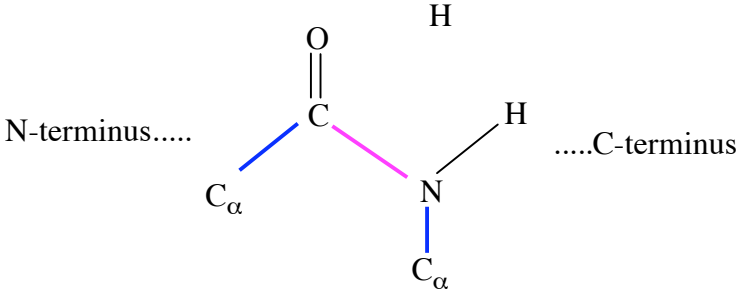
Thus,

- The length of the peptide bond (1.33 Å) becomes intermediate between the normal (single-bond) C—N (1.45 Å) and the normal (double-bond) C=N (1.25 Å).
- The entire group –CO—NH—becomes strongly planar (as it would be for the double bond).
- The –CO—NH—group generally lies close to either a "trans" or a "cis" form with "trans" being by far the more likely:

"trans"
common:



"cis"
uncommon: 1-10 to 1:1000

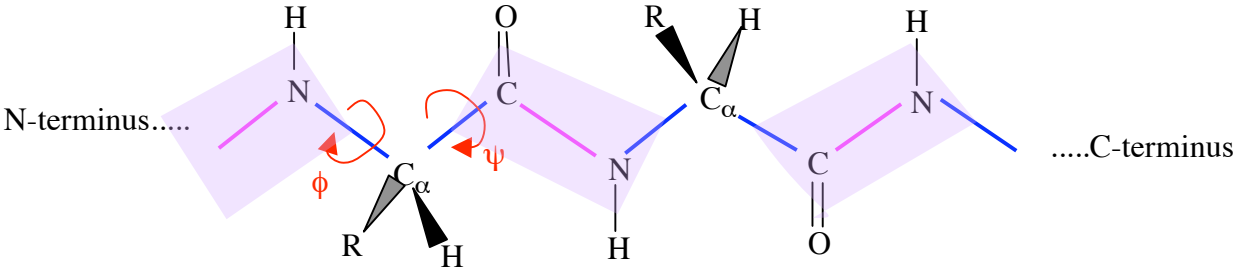


Note: Exception when Pro follows on the C-terminal side, in which case the "cis" is considerably more likely (but still a minority). (1:4 to 3:7).

Upshot: For practical purposes, we may regard the peptide configuration as rigid-planar-trans.

It follows that the dominant configurational degrees of freedom available to the protein are rotations about the (blue) C_α—C (ψ) and N—C_α (ϕ) bonds. Thus, a configuration of the whole protein chain is given by a specification of all these rotation angle, starting from the N-terminus all the way down to the C-terminus.

You might imagine that the "best" configuration of the protein would tend to be one in which the "big" (except for Gly) R residues are maximally separated to reduce steric interference. This leads to the configuration shown below, in which the R's alternate between out-of-page and into-page. The peptide planes are shown shaded.



This is, indeed, in the range of normally observed configurations; however, the actual configuration depends in detail on the specific amino acids and amino-acid sequences.

There are many contributing factors:

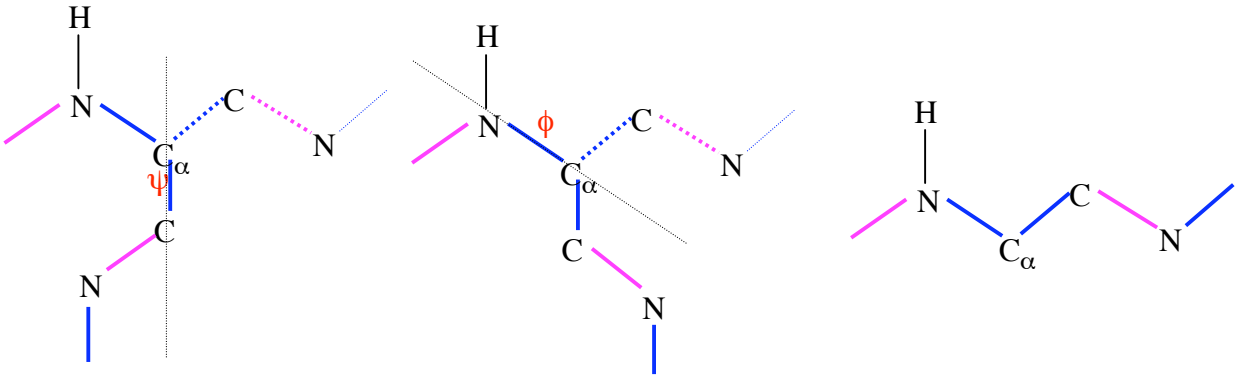
- steric hinderance
- hydrophobic/hydrophilic interactions of aa groups with water
- charge groups on amino acids
- van der Waals interactions
- thermal fluctuations

The Ramachandran plot and secondary stuctures:

Consider rotations about a particular C_α—C (ϕ) and N—C_α (ψ) pair.

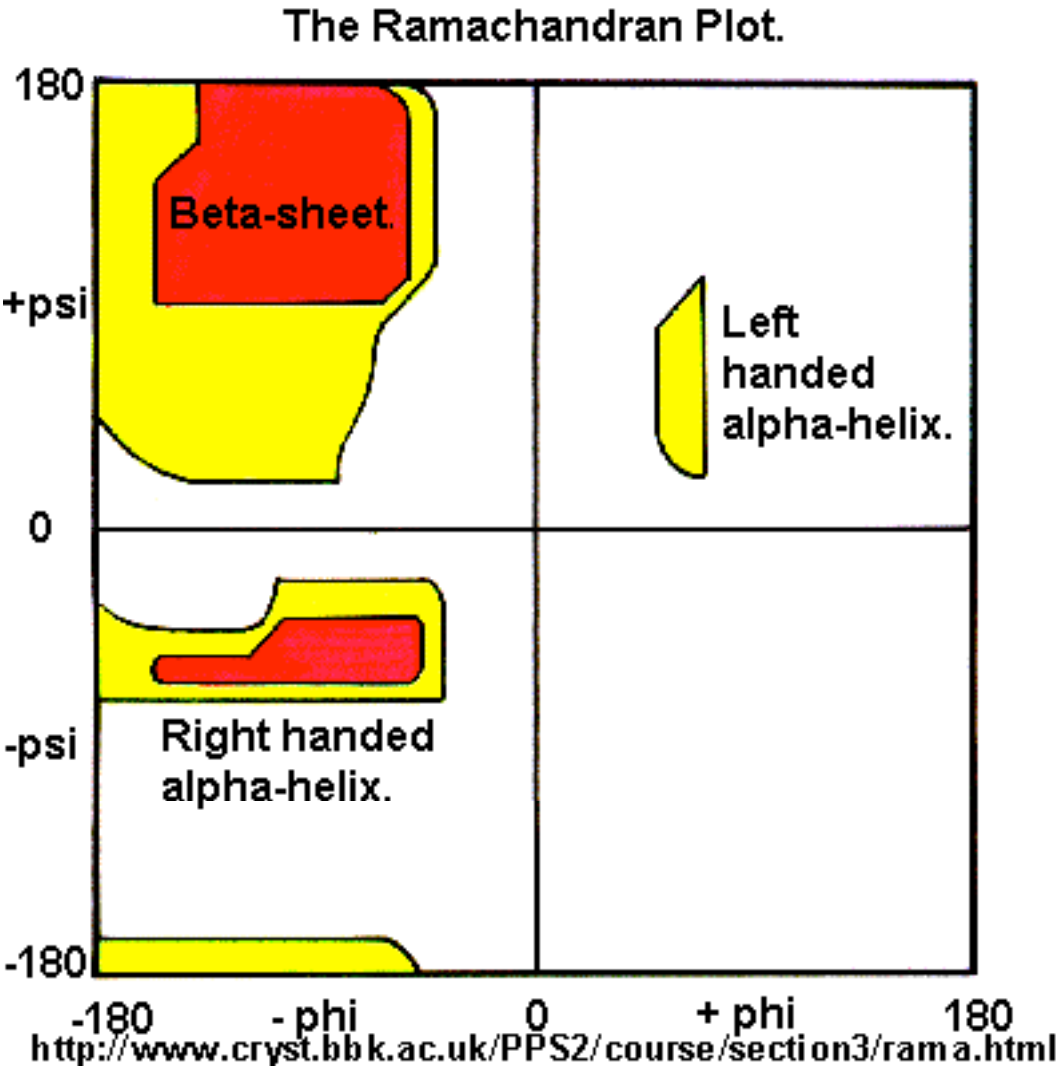
You need to specify a "zero" for each angle.

Oddly, it seems to me, the ϕ=ψ=0 state is chosen as one in which the chain is "wound up" into a sterically impossible configuration:



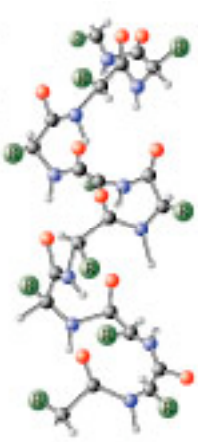
It is useful to look at the plot of allowed (ϕ,ψ) pairs which are allowed:

Sterically allowed angles fall in certain regions of the Ramchandran plot (depending somewhat on the specific sequence of three amino acids). Ala-Ala-Ala **26.6**



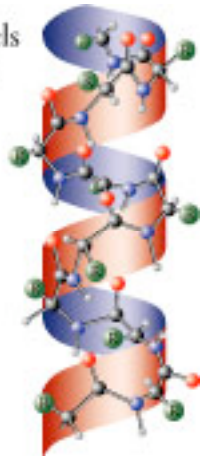
When these plots are made for real proteins, there are concentrations of points in certain regions of the Ramchandran plots which correspond to specific "secondary structures": α-helices (left- and right-handed and β-sheets (ribbons)(parallel and antiparallel).

Ball-and-stick model of a portion of the α-helical secondary structure of a protein molecule

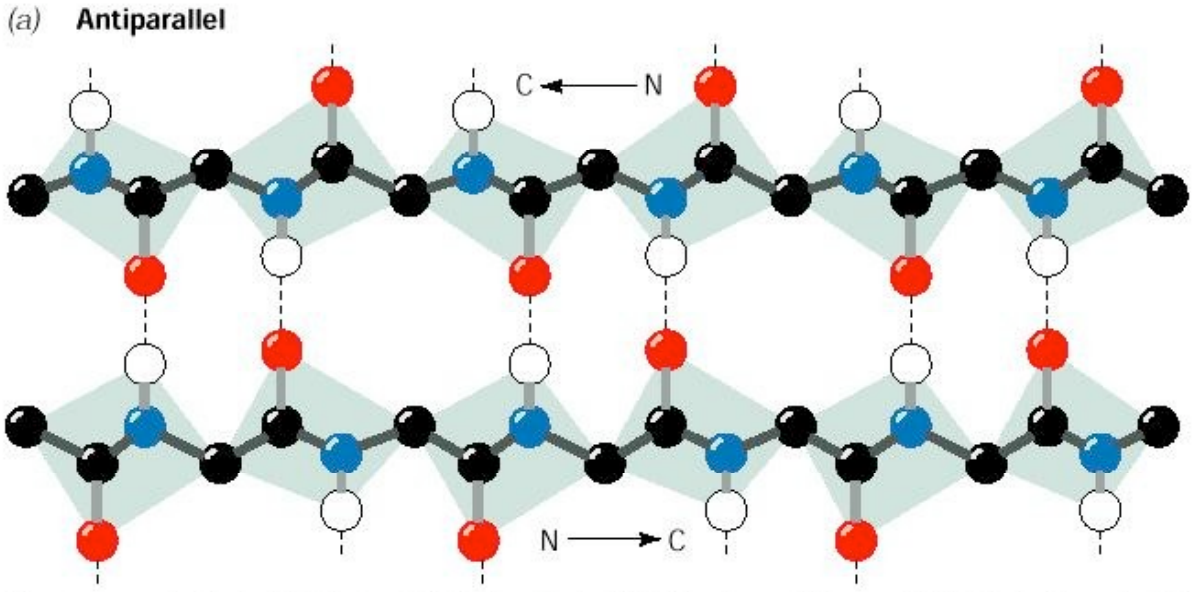


This ribbon model shows the general arrangement of atoms in a portion of the α-helical secondary structure of a protein molecule.

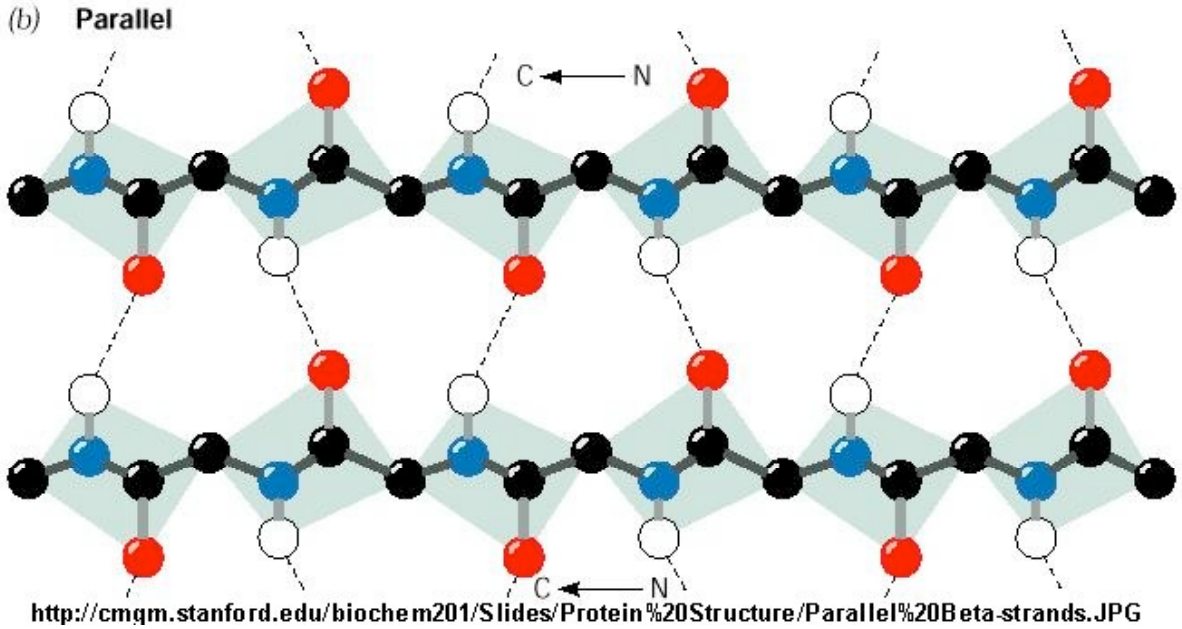
The two models superimposed



Beta strands or sheets: **26.7**



http://cmgm.stanford.edu/biochem201/Slides/Protein%20Structure/Anti_parallel%20Beta-Strands.JPG



http://cmgm.stanford.edu/biochem201/Slides/Protein%20Structure/Parallel%20Beta-strands.JPG

All these considerations still give us what is, basically, only local structure: What local angles are preferred?