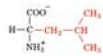
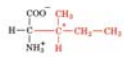
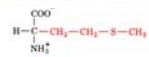
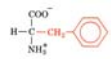
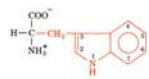
Reading for lecture 4

1. Protein Structure

- Voet and Voet, Chapters 4, 7
- Alberts et al, Chapter 3

Name	Structural Formula [a]	Residue Mass (D)	Average Occurrence in Proteins (%) [b]
Glycine		57.0	7.5
Alanine		71.0	9.0
Valine		99.1	6.9
Leucine		113.1	7.5
Isoleucine		113.1	4.6
Methionine		131.1	1.7
Proline		97.1	4.6
Phenylalanine		147.1	3.5
Tryptophan		186.2	1.1

Protein Structure

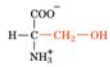
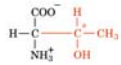
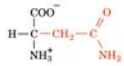
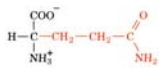
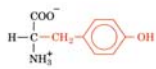
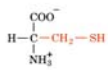
Amino acid side chains

[non-polar side amino acids] (2)

Proline is non-standard

Aromatics – brief explanation

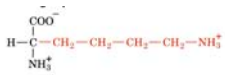
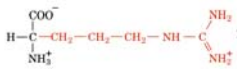
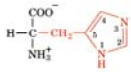
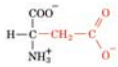
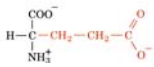
Glycine is the simplest possible amino-acid

Name	Structural Formula [a]	Residue Mass (D)	Average Occurrence in Proteins (%) [b]
Serine		87.0	7.1
Threonine		101.1	6.0
Asparagine		114.1	4.4
Glutamine		128.1	3.9
Tyrosine		163.1	3.5
Cysteine		103.1	2.8

[polar side-chains] (3)

OH, NH2, =O, SH in cysteine

NH2 not NH3+ in these molecules.- predominant form at neutral pH (more on acid-base equilibrium in lecture 5).

Name	Structural Formula [a]	Residue Mass (D)	Average Occurrence in Proteins (%) [b]
Lysine		129.1	7.0
Arginine		157.2	4.7
Histidine		137.1	2.1
Aspartic acid		114.0	5.5
Glutamic acid		128.1	6.2

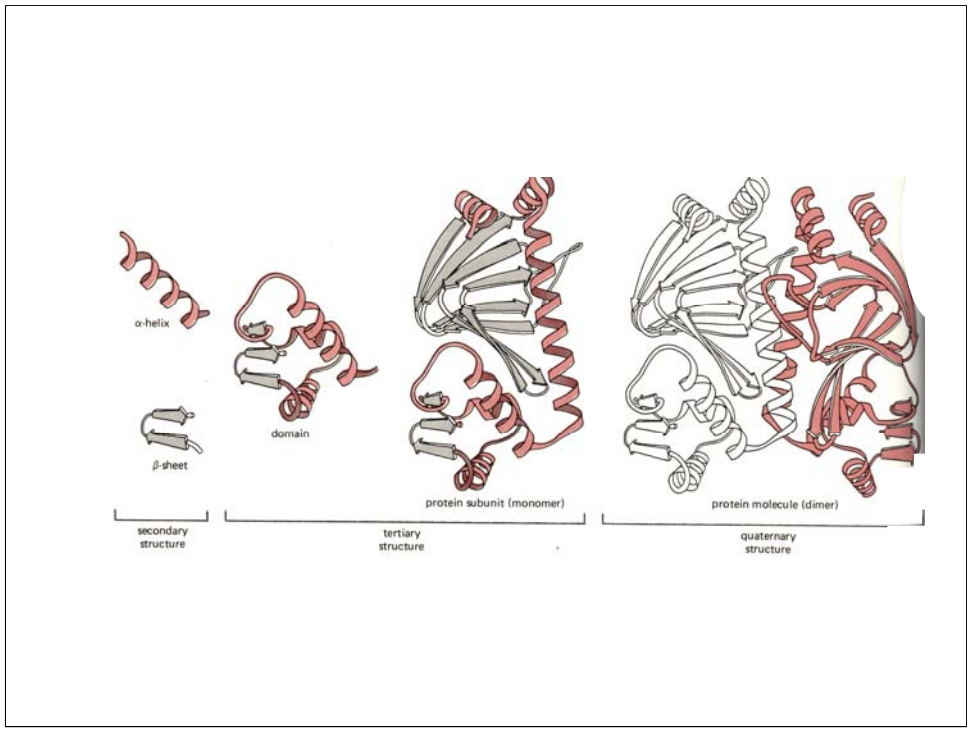
[charged side chains] (4)

Histidine actually has roughly equal probability of being charged (C1NH2+) or neutral (as shown here)

There are also ~ dozen non-standard amino-acids found in proteins.

Only one (Seleno-cysteine) has a tRNA. Very low occurrence.

Others, side-chains are modified after protein chain is made.



Protein folding

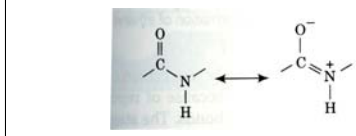
1-D polypeptide folds into precise 3-D structure that performs a function

[Structural heirarchy of proteins] (5)

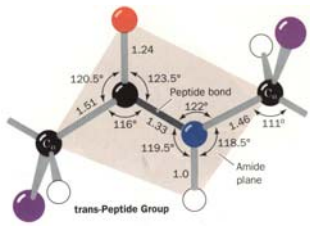
Self-assembly.

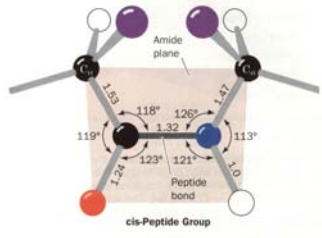
Not very well understood…

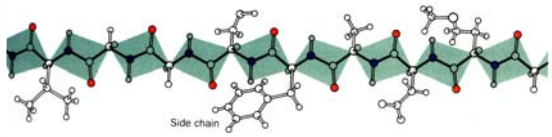
This lecture will describe elements at different levels of protein structure…



Double bonds are planar







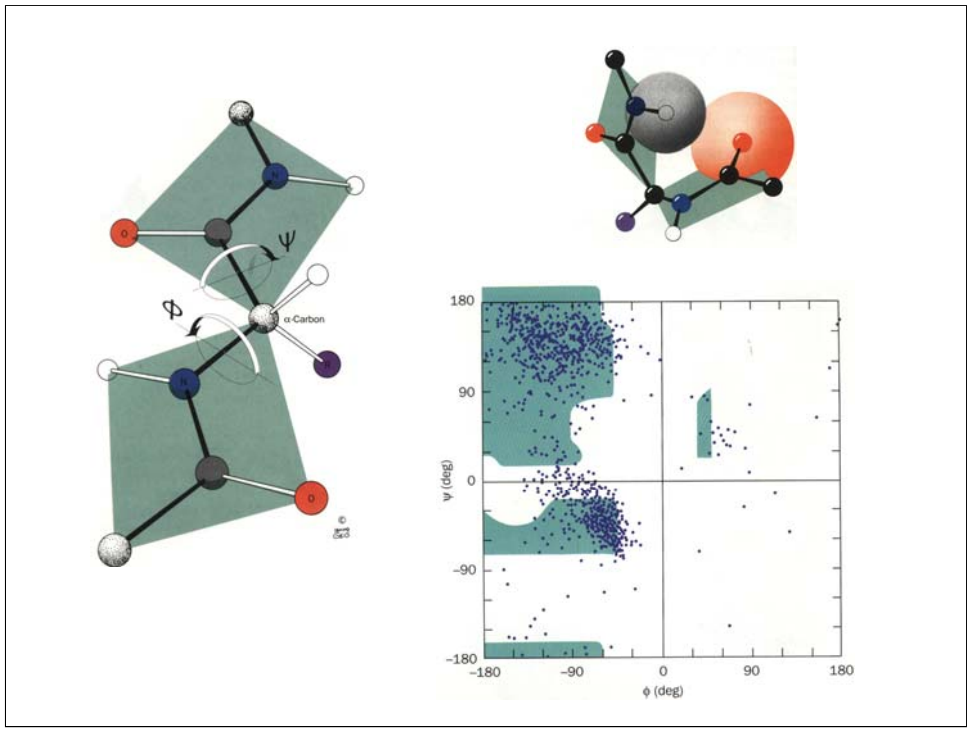
Secondary structure

[peptide bond] (6)

Planar due to partial double-bond nature of C-N peptide bond. (no rotation about this bond – hybrid bonding electron orbitals are different than single bonds)

Cis- due to steric hindrance. Proline is exception

“Backbone”



[Ramachandran plot] (7)

Only 2 degrees of freedom left after constraints in previous slide

Even these are not fully free, steric hindrance

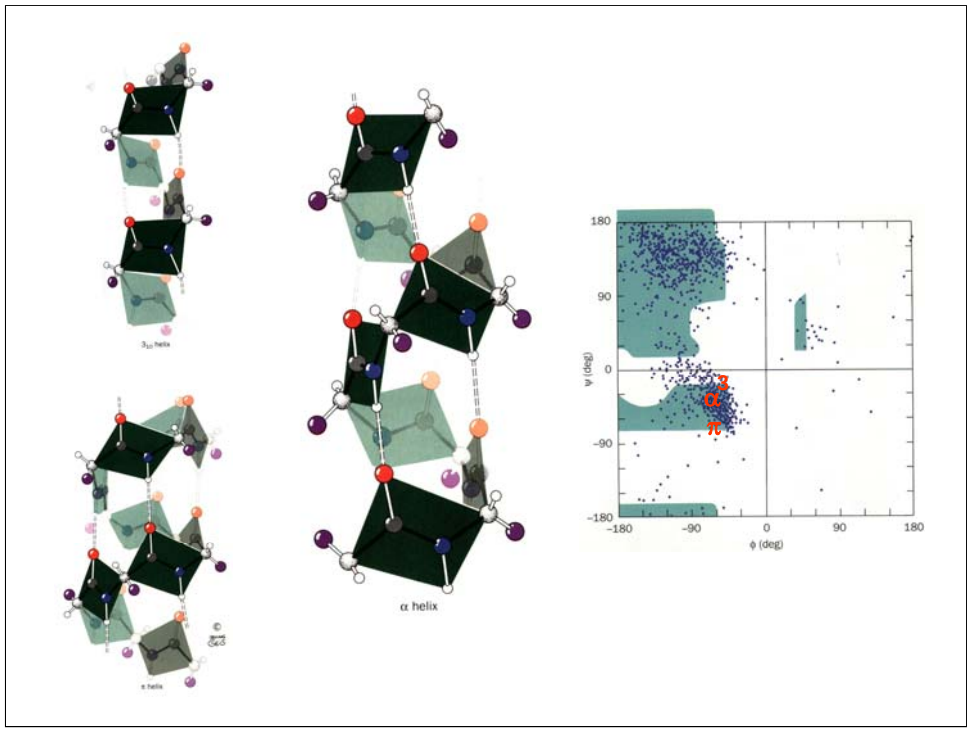
Ramachandran plot is for alanine-alanine (R = CH3)

- Based on minimal overlap of atoms calculated from Van der Waals radius
- Glycine is MUCH freer
- Points are for non-glycine residues in 1000 proteins – fit quite well to Alanine plot
- Bigger amino acids are more restricted
- Proline is much less free
- “forbidden” points due to accommodation by twisting planar peptide bond.

Ramachandran plot restricts allowed folds – other factors detrmine which allowed folds will be stable…

Two particularly stable (therfore common) repeating structural elements. **a-helix** and **b-sheet**…

Stabilized because H-bonds form between different parts of the backbone



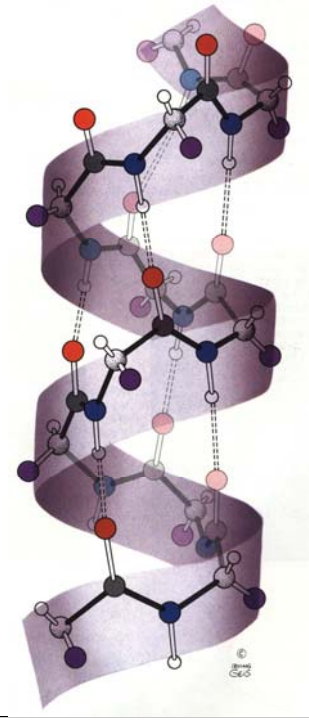
[protein helix motifs] (8)

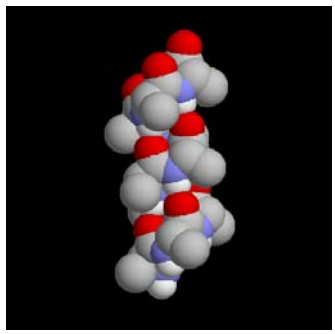
Helix is common – forms with simple rotation+translation from one element to the next (eg DNA double helix).

H-bonds between each backbone C=O and the N-H at 2,3,4 residue along the chain

Alpha Helix is the most stable by far because…

- In fully allowed region of RC plot.
- H-bonds are optimum length and in a striaght line
- Core is tightly packed – no gaps, no overlap
- Side-chains all point out – no steric hindrance



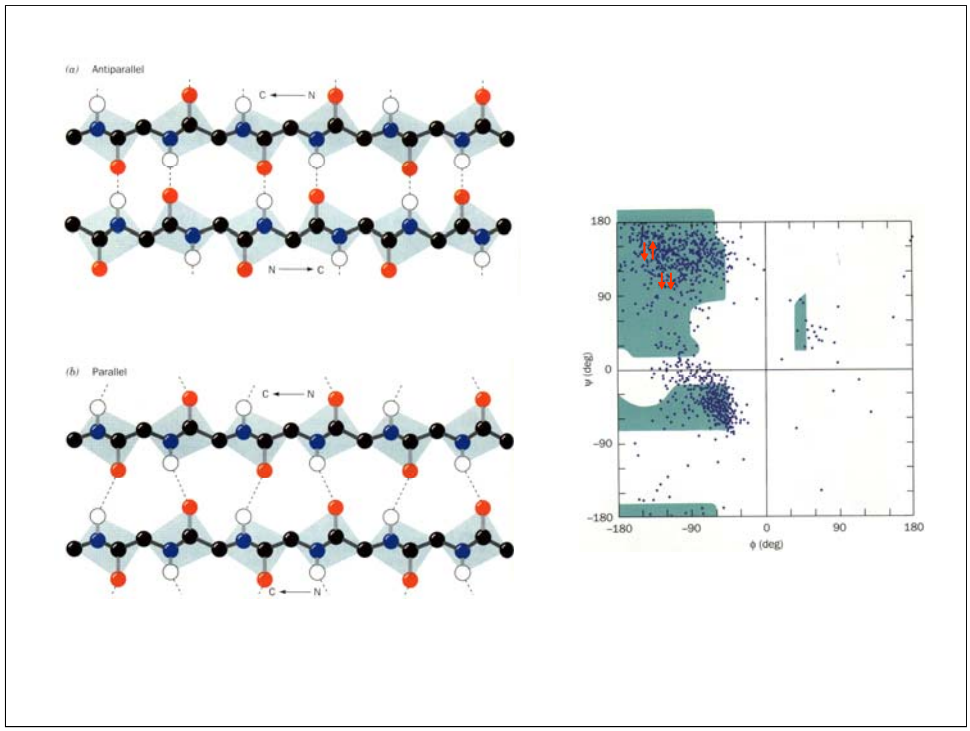


(Windows key rtclk to view RASMOL)

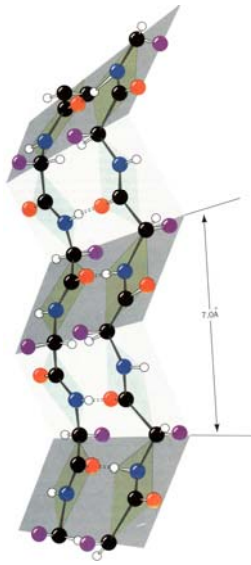
[alpha helix representations] (9)

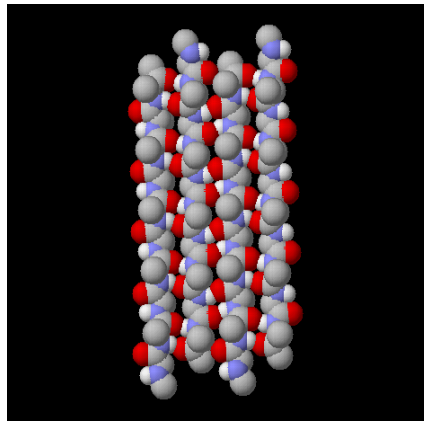
[PDB STRUCTURE]

Rigid-rods as building-blocks in the next level of structure.



[beta sheets] (10)





(RASMOL)

[beta sheets] (11)

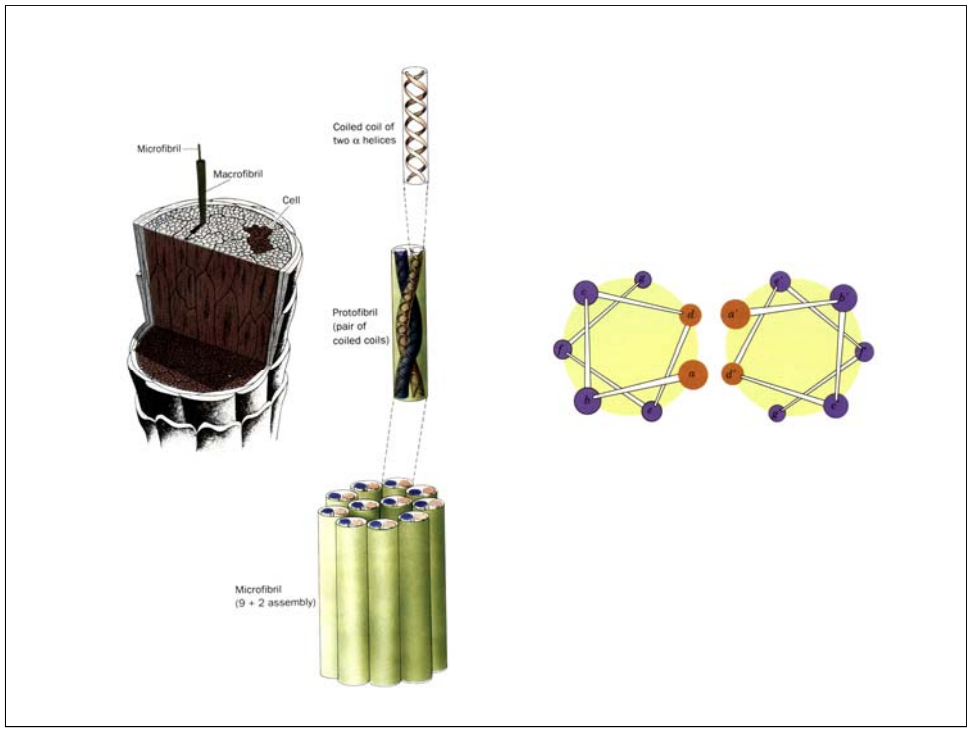
Anti-parallel are more stable – H-bonds are linear.

Side-chains point outwards from sheet

Typically 6 long and 6 wide, though up to 15.

Mixed parallel/anti-parallel are less common than if it were random

Often have a RH twist – not perfect planar as shown



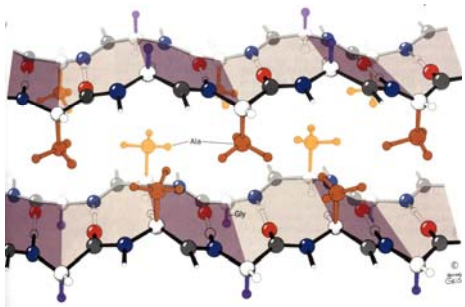
Tertiary and Quaternary Structure

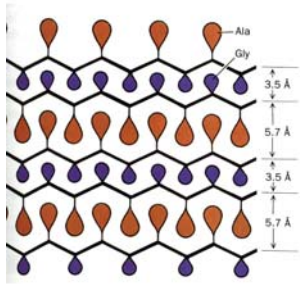
Repetitive structures do exist

Strong and chemically inactive

[keratin] (12)

Coiled coil motif is due to mutually sticky strips on alpha helices. Not quite periodic, so helices coil rather than stacking adjacent.



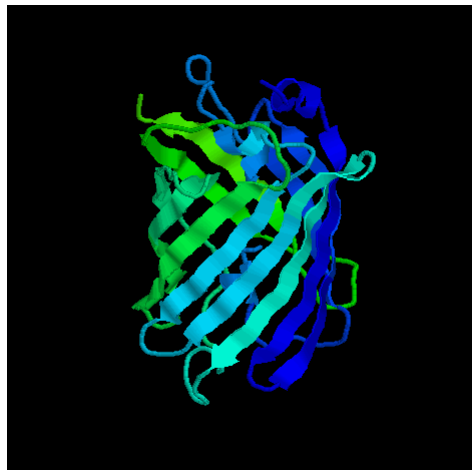


[silk] (13)

More flexible

More typically, alpha and beta elements are linked by non-repetivive “loops”

“Globular Proteins”

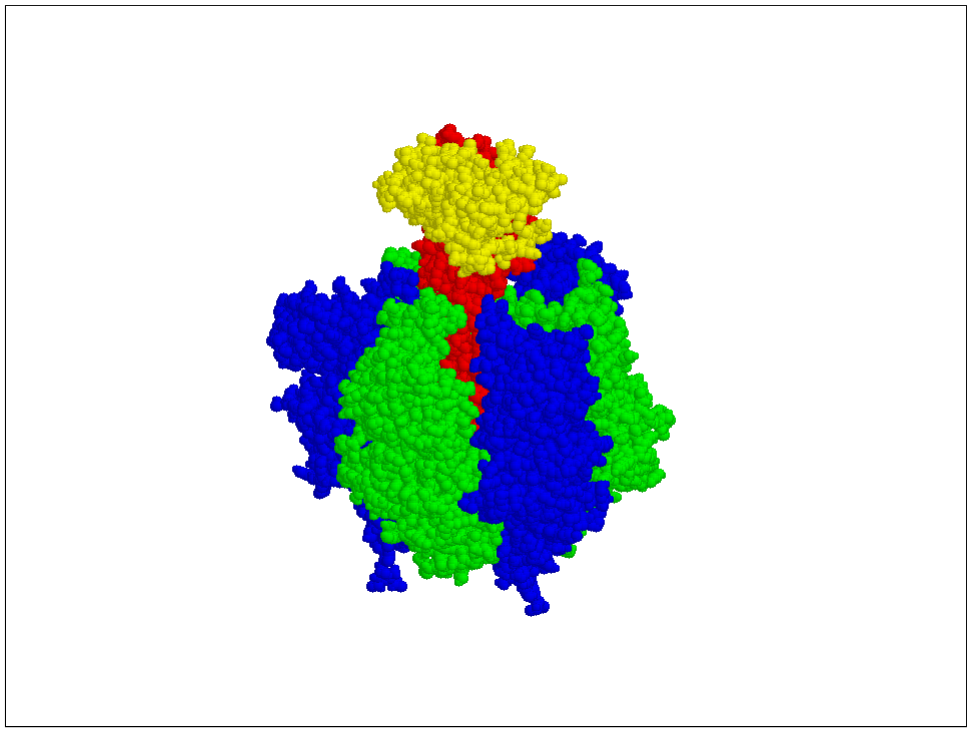


(RASMOL)

[RASMOL – GFP] (14)

different representations: cartoons, spacefill

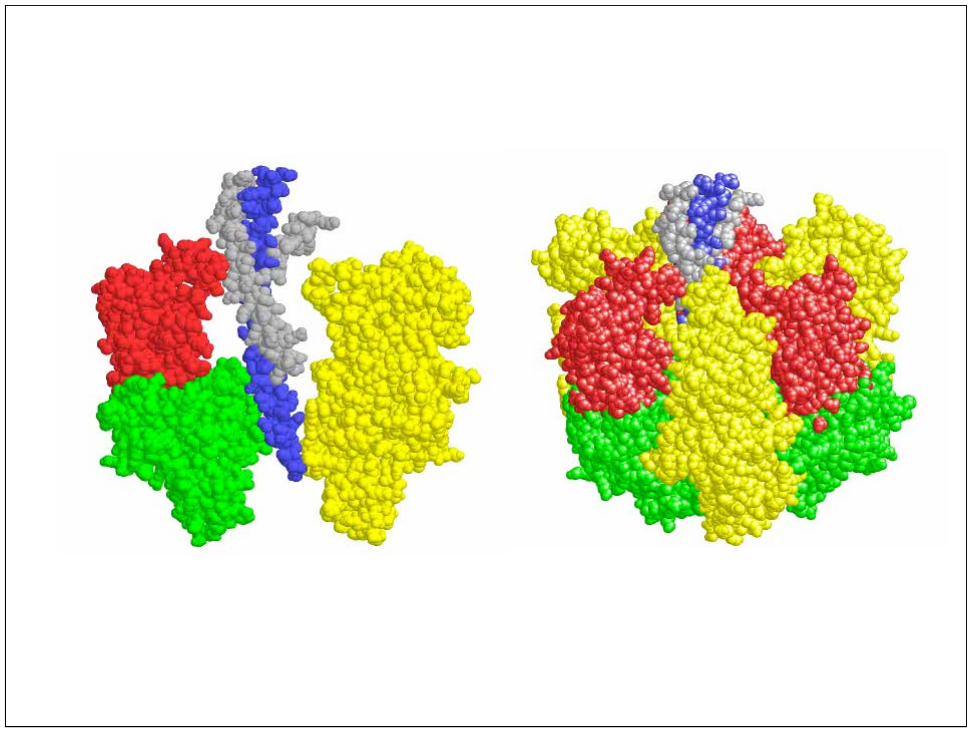
beta-barrel



[RASMOL – F1-ATPase] (15)

Gamma – alpha helices to form the axle

Beta – domains, typical globular protein, eg beta barrel at the bottom.

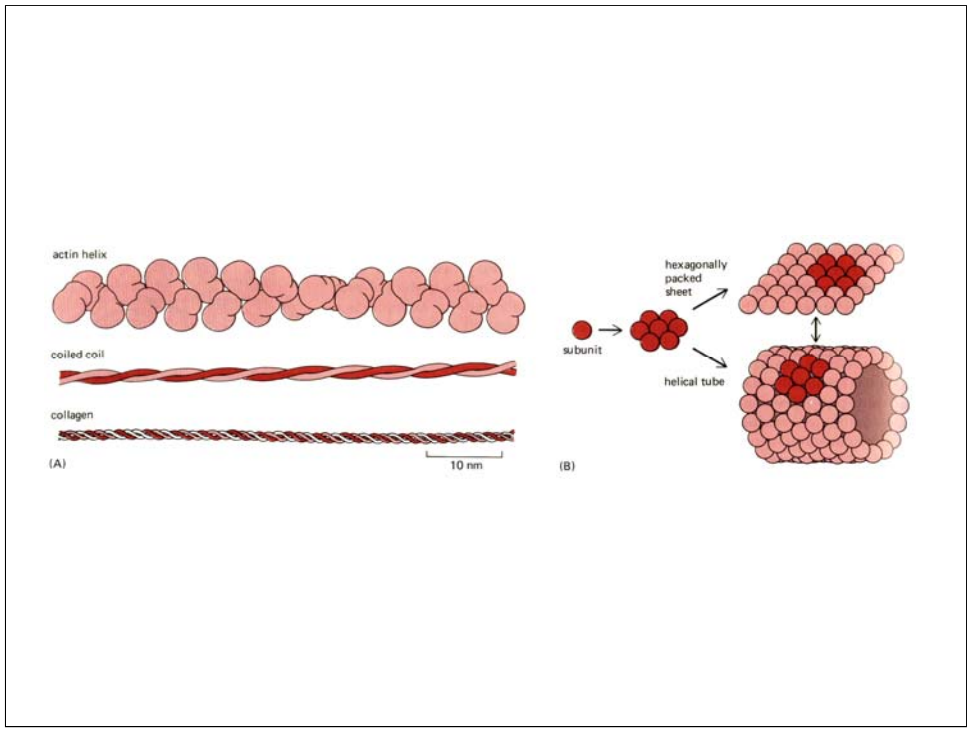


[F1 3-D simulation] (16)

Domains moving relative to each other in a protein machine

Caveats: All based on one crystal structure with pseudo 3-fold symmetry.

More on F1 later



Globular Proteins can also be building blocks in lareger periodic structures [actin, microtubules] (17)

Important Forces In Protein Folding

1. Hydrogen Bonds
   - Between backbone atoms
   - Between side-chain atoms
2. Disulphide Bonds
3. Van der Waals and dipole forces
4. **Hydrophobic Forces**

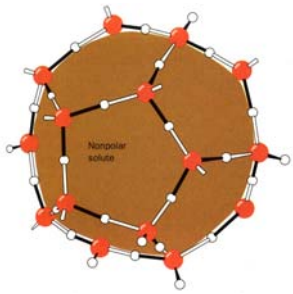
Forces involved in protein folding

[Summary of important forces] (18)

Hydrogen bonds **do not** stabilize folded structure compared to unfolded – because the same atoms would form H-bonds to water in unfolded structure.

BUT, H-bonds DO discriminate between “correct” fold and other folds. For example alpha-helix and beta-sheet motifs.

Stability against un-folding is mostly due to hydrophobic force.

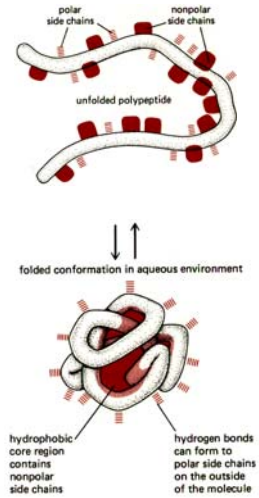


Non-polar solutes cause increased order in surrounding water molecules

For solvation of a hydrophobic molcule…

$\Delta G = \Delta H - T\Delta S$

~0 >0



[Hydrophobic force] (19)

Water forms transient H-bonded network

Non-polar (“hydrophobic”) molecules cannot participate in this

It’s **NOT simply that there are less H-bonds** if a hydrophobic molecule is in water – the water molecules can re-orient to H-bond each other

Rather it’s an increase in order as water has to accommodate the molecule

dS > 0 in spontaneous process, so hydrophobic molecule will leave water.

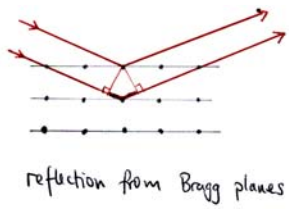
(Gibbs) free energy – more in lecture 5

Mis-folding, prions, “chaperones”

Proteins are only just stable against unfolding (“denaturation”)

DG ~ 2x H-bond (compare to kT and covalent bond)

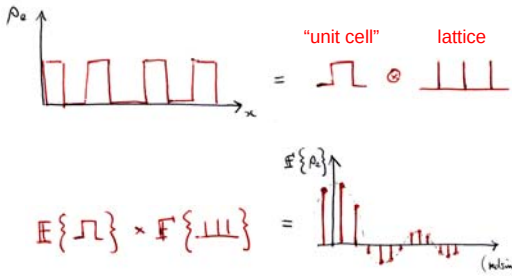
Nearly balanced – lots of H-bonds, but cost of burying parst of chain away from water.



Constructive interference when Bragg condition is met :

"Diffraction spots"

1-D illustration (eg Fraunhofer diffraction in optics)



Amplitudes of spots give

"structure factor"
= Fourier transform of unit cell

Determintation of Protein Structure

Several methods, most important is **x-ray crystallography**

Source of all atomic level structures shown

For a proper treatment see for example: "Principles of Protein X-Ray crystallography", Jan Drenth, Springer

[Bragg planes and structure factor] (20)

If we knew amplitude and phase of the diffraction spots, 3-D inverse fourier transform to obtain electron density

Phase problem – intensities only, 2-D data in diffraction pattern.

Solution : heavy-metal derivatives.

Strong pont scatterer(s) in each unit cell – (many electrons)

different interference pattern dependent upon relative phase of scattering by heavy metal and protein.

Can obtain enough information to reconstruct the 3-D protein structure **GIVEN constraints** imposed by…

known primary structure and

secondary structure predictions based on other known structures

"Threading" polypeptide into electron density map

Computerised process, no longer rate-limiting in obtaining structures

Obtaining (highly regular)crystals is the hardest part

PDB has thousands of protein crystal structures, and growing fast.