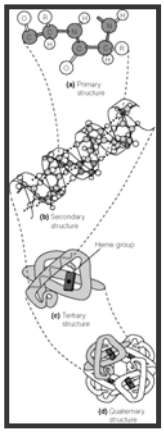
Introduction to Protein Folding

Chapter 4 Proteins: Three Dimensional Structure and Function

- **Conformation** - three dimensional shape
- *Native conformation* - each protein folds into a single stable shape (physiological conditions)
- Biological function of a protein depends completely on its native conformation
- A protein may be a single polypeptide chain or composed of several chains



There are Four Levels of Protein Structure

1. **Formation of Primary Structure**
2. **Formation of Secondary Structure**
3. **Formation of Tertiary Structure**
4. **Formation of Quaternary Structure**

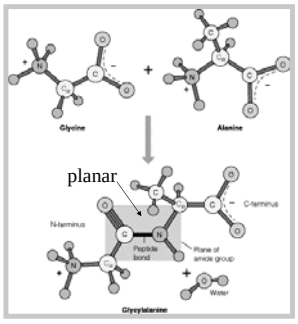
How many AA sequences are there for a typical protein 100 AA long?

20^{100}

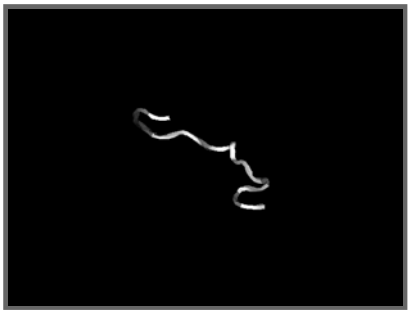
The Conformation of the Peptide Group

- **The peptide group consists of 6 atoms (next slide)**
- **Peptide bonds have some double bond properties so that their conformation is restricted to either *trans* or *cis***
- ***Cis* conformation is less favorable than *trans* due to steric interference of α-carbon side chains**
- ***Cis-trans* isomerases can catalyze the interconversion of *cis* and *trans* conformations**

How Do the Amino Acids Connect to One Another?



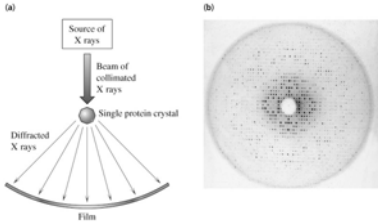
Once We Have a Long Polypeptide, Then What?

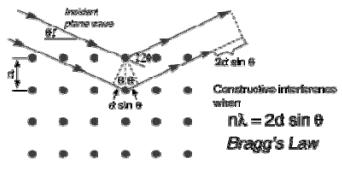


Form Secondary Structures

Methods for Determining Protein Structure

- X-ray crystallography is used to determine the three-dimensional conformation of proteins

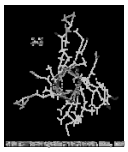




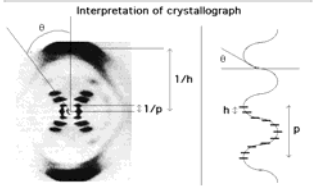
http://www.eserc.stonybrook.edu/ProjectJava/Bragg/

http://epswww.unm.edu/xrd/xrdbasics.pdf

X-ray diffraction pattern from B form of DNA



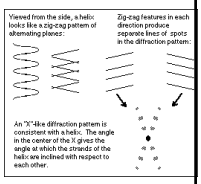
Interpretation of crystallograph



θ - tilt of helix (angle from perpendicular to long axis)

h – 3.4 Å (Distance between bases)

p – 34 Å (Distance for one complete turn of helix; Repeat unit of the helix)



Linus Pauling and collaborators used X-ray diffraction studies to postulate several principles that a structure must obey.

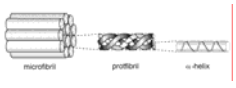
1. **The bond lengths and bonds angles should be distorted as little as possible.**
2. **No two atoms should approach one another more closely than is allowed by there van der Waal radii.**
3. **The amide group must remain planar and in the *trans* configuration. This allows only rotation about the two bonds adjacent to the α-carbon.**
4. **Some kind of noncovalent bonding is necessary to stabilize a regular folding.**

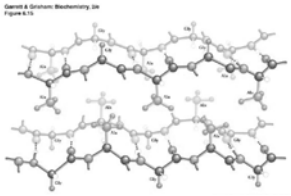
Pauling and Corey's Work

- They worked on the fibrous proteins: α-keratin (hair, wool, and skin) and β-keratin (silk and spider webs).

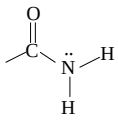
β-keratin

α-keratin



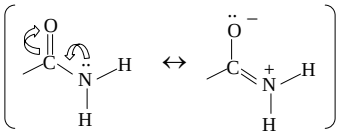


Pauling Studied Structures of Small Molecules Containing Amide Bonds. He found all amide bonds have same structure.



- **C, O, N, H are coplanar**
- **C-N bond is shorter than most C-N bonds**
- **O and H are always trans**

Rationale: partial double bond character in the C-N bond.



X-ray diffraction experiments on α and β keratin concluded:

Structural information

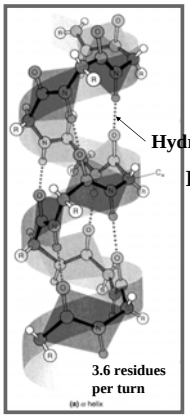
- Repeat distance – distance before folding pattern repeats

α-keratin = 0.55 nm

β-keratin = 1.3 nm-1.4 nm

The structure of a polypeptide chain can be described as amide bonds separated by tetrahedral carbon bonds

α C of Amino Acids



Alpha helix

Right-handed helix

Hydrogen Bond (axial)

Repeats itself in 18 residues and 5 turns

**18 residues/ 5 turns =
3.6 residues/per turn**

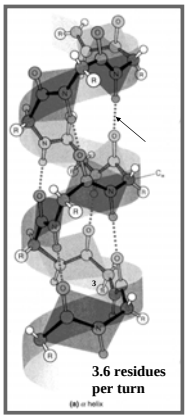
13 atoms in the hydrogen bond loop

A Common Way to Express Other Types of Helices is Using the n_N method.

**Remember n = number residues per turn and
N = atoms in H-bond network**

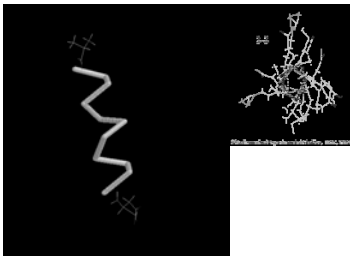
α-helix

3.6_{13}



Alpha Helix

Repeats itself in 18 residues and 5 turns



http://www.umass.edu/molvis/freichsman/protarch/page_helix/menu.html

18 residues/5 turns = 3.6 residues/turn

Describing the Structures

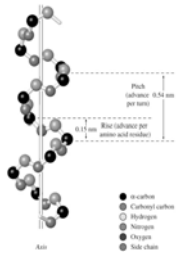
c = crystallographic repeat
p = pitch (nm/turn)
h = rise (nm/residue)
n = residues per turn
m = residues per repeat (must be an integer)
N= atoms in hydrogen bond loop

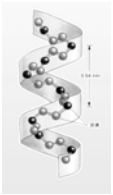
If the rise of the α-helix is 0.15 nm/residue, what is the pitch?

p = hn

p = 0.15 nm/residue * 3.6 residues/turn

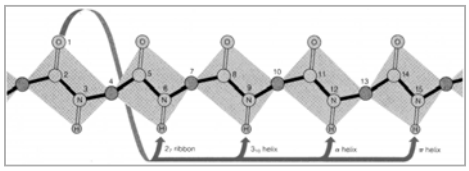
p = 0.54 nm/turn





Hydrogen Bonding Patterns for Four Helices

Tighter helices (less residues per turn)

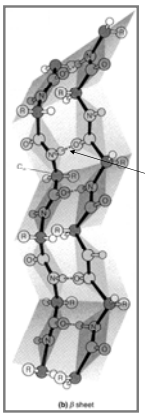


n = 2 will not form a helix

It is very difficult to have only two residues per turn and linear hydrogen bonds between residues in the same chain

↓

Therefore, structures which have only two amino acids in the turn form a beta pleated sheet



Each residue is rotated by 180° with respect to the previous

Chains are folded in an accordion-like fashion from α-carbon to α-carbon

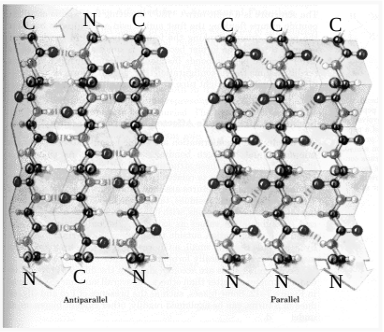
Hydrogen Bond

The hydrogen bonds occur between adjacent chains

Parallel and antiparallel β-stands

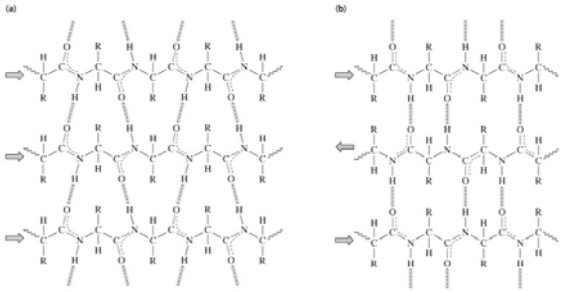
- β Strands in a sheet are parallel or antiparallel
- Parallel β sheets - strands run in the same N- to C- terminal direction
- Antiparallel β sheets - strands run in opposite N- to C- terminal directions
- In antiparallel β sheets the H-bonds are nearly perpendicular to the chains (more stable than parallel chains with distorted H-bonds)

Antiparallel vs Parallel β-sheets

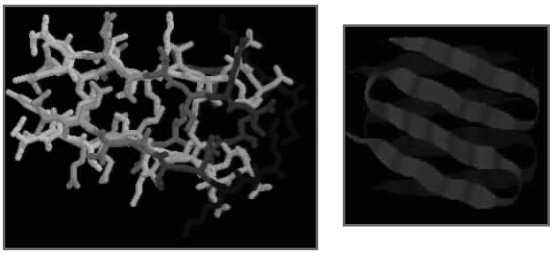


http://www.agsci.ubc.ca/courses/fnh/301/protein/protprin.htm

β-Sheets (a) parallel, (b) antiparallel



The Conformation of β-sheets

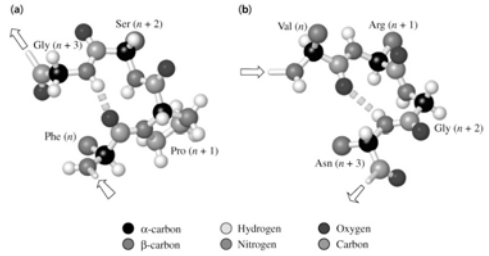


Loops and Turns

- Loops and turns connect a helices and β strands and allow a peptide chain to fold back on itself to make a compact structure
- **Loops** - often contain hydrophilic residues and are found on protein surfaces
- **Turns** - loops containing 5 residues or less
- **β Turns** (**reverse turns**) - connect different antiparallel β strands

Reverse turns

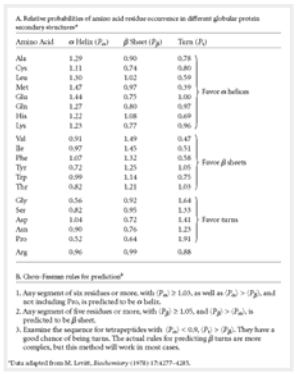
(a) Type I, and (b) Type II



What constitutes into what secondary structure the protein will fold?

1. **Amino acid sequence**
2. **Angles of rotation about ϕ and ψ**

Relative Probabilities of Amino Acid Residues Occurrence in Different Globular Protein Secondary Structure as predicted by P. Y. Chou and G. D. Fasman



A. Relative probabilities of amino acid residue occurrence in different globular protein secondary structures[a]

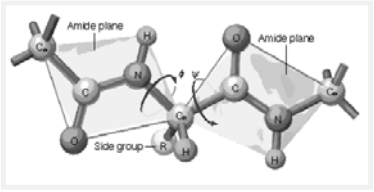
Amino Acid	α Helix (P_α)	β Sheet (P_β)	Turn (P_t)	
Ala	1.29	0.90	0.78	Favor α helices
Cys	1.11	0.74	0.80	
Leu	1.30	1.02	0.59	
Met	1.47	0.97	0.39	
Glu	1.44	0.75	1.00	
Gln	1.27	0.80	0.97	
His	1.22	1.08	0.69	
Lys	1.23	0.77	0.96	
Val	0.91	1.49	0.47	Favor β sheets
Ile	0.97	1.45	0.51	
Phe	1.07	1.32	0.58	
Tyr	0.72	1.25	1.05	
Trp	0.99	1.14	0.75	
Thr	0.82	1.21	1.03	
Gly	0.56	0.92	1.64	Favor turns
Ser	0.82	0.95	1.33	
Asp	1.04	0.72	1.41	
Asn	0.90	0.76	1.23	
Pro	0.52	0.64	1.91	
Arg	0.96	0.99	0.88	

B. Chou–Fasman rules for prediction[a]

1. Any segment of six residues or more, with $(P_\alpha) \geq 1.03$, as well as $(P_\alpha) > (P_\beta)$, and not including Pro, is predicted to be α helix.
2. Any segment of five residues or more, with $(P_\beta) \geq 1.05$, and $(P_\beta) > (P_\alpha)$, is predicted to be β sheet.
3. Examine the sequence for tetrapeptides with $(P_\alpha) < 0.8$, $(P_t) > (P_\beta)$. They have a good chance of being turns. The actual rules for predicting β turns are more complex, but this method will work in most cases.

[a]Data adapted from M. Levitt, *Biochemistry* (1978) 17:4277–4285.

Rotation Around the Bonds in a Polypeptide Bond

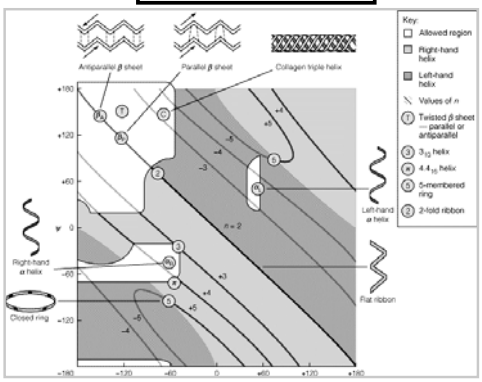


The angles of rotation about these bonds are defined as ϕ and ψ with directions defined as positive rotation with respect to the α carbon

Permissible values of ϕ and ψ

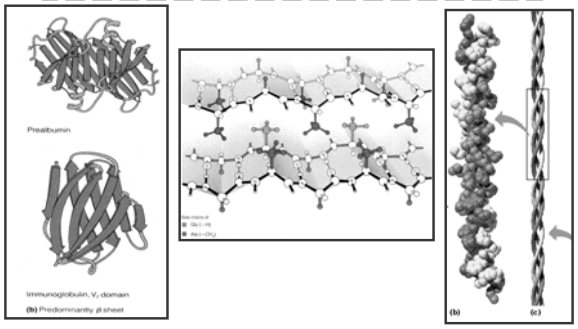
- Conformation of a polypeptide chain can be solely described by ϕ and ψ angles
- Ramachandran plots of ϕ and ψ show permissible angles for polypeptide chains
- Some ϕ and ψ angles are not allowed because of steric hindrance
- Conformations of several types of secondary structures fall within permissible areas

A Ramachandran Plot



Tertiary Structures

Now that a secondary structure has formed, what will it do?



Tertiary Structure of Proteins

- Tertiary structure results from the folding of a polypeptide chain into a closely-packed three-dimensional structure
- Amino acids far apart in the primary structure may be brought together
- Stabilized primarily by noncovalent interactions (e.g. hydrophobic effects) between side chains
- Disulfide bridges also part of tertiary structure

Supersecondary Structures (Motifs)

Motifs - recurring protein structures

(a) **Helix-loop-helix** - two helices connected by a turn

(b) **Coiled-coil** - two amphipathic α helices that interact in parallel through their hydrophobic edges

(c) **Helix bundle** - several α helices that associate in an antiparallel manner to form a bundle

(d) **βαβ Unit** - two parallel β strands linked to an intervening α helix by two loops

Supersecondary structures (cont)

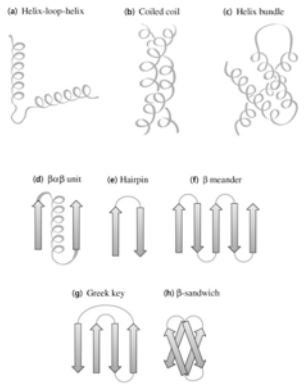
(e) **Hairpin** - two adjacent antiparallel β strands connected by a β turn

(f) **β Meander** - an antiparallel sheet composed of sequential β strands connected by loops or turns

(g) **Greek key** - 4 antiparallel strands (strands 1,2 in the middle, 3 and 4 on the outer edges)

(h) **β Sandwich** - stacked β strands or sheets

Common motifs



Domains

- Independently folded, compact units in proteins
- Domain size: ~25 to ~300 amino acid residues
- Domains are connected to each other by loops, bound by weak interactions between side chains
- Domains illustrate the evolutionary conservation of protein structure

Four categories of protein domains

(1) **All α -** domains consist almost entirely of α helices and loops

(2) **All β** - all domains contain only β sheets and non-repetitive structures that link the β strands

Protein domains (continued)

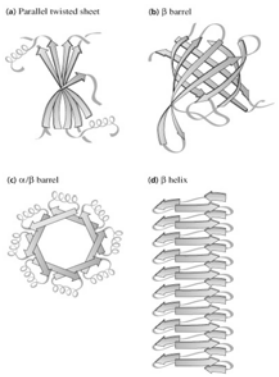
(3) **Mixed α/β** - contain supersecondary structures such as the αβα motif, where regions of α helix and β strand alternate

(4) **α + β** - domains consist of local clusters of α helices and β sheet in separate, contiguous regions of the polypeptide chain

Folds

- Within each of the four main structural categories, domains can be classified by characteristic "**folds**"
- A "fold" is a combination of secondary structures that form the core of a domain
- Some domains have simple folds, others have more complex folds

Common domain folds



What are two major factors involved in the formation of a tertiary structure?

Two Major Factors

Thermodynamics

Conformational Entropy
Internal Hydrogen Bond
Electrostatic Interactions
Hydrophobic Effect
van der Waals Interaction
Disulfide Bonds

Kinetics

Levinthal's Paradox

Conformational Entropy

To go from  **To** 

Involves a decrease in randomness (less disorder)

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

If entropy is becoming small, then ΔG is becoming more positive

∴conformational entropy change works against folding

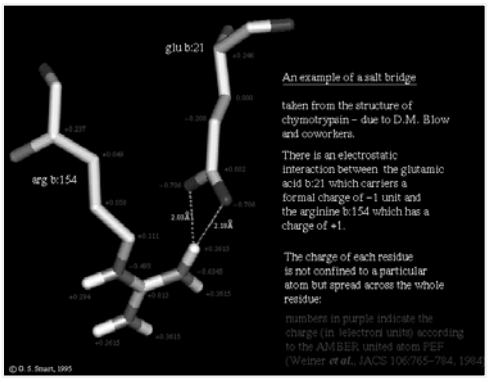
WE MUST SEEK FEATURES OF PROTEIN FOLDING THAT YIELD EITHER LARGE NEGATIVE ΔH OR SOME OTHER INCREASE IN ENTROPY ON FOLDING

Electrostatic Interactions

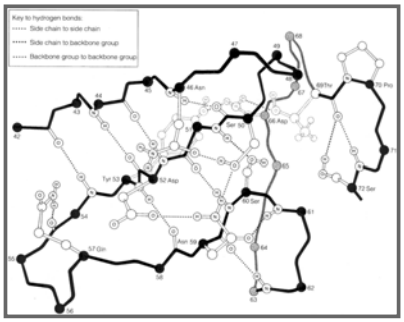
Typical charge-charge interactions that favor protein folding are those between oppositely charged R-groups such as K or R and D or E.

Another component of the energy involved in protein folding is charge-dipole interactions.

This refers to the interaction of ionized R-groups of amino acids with the dipole of the water molecule.



Internal Hydrogen Bonds

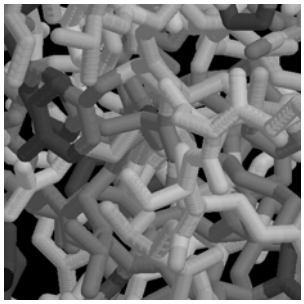


Polypeptides have many opportunities to hydrogen bond both in the backbone of the protein as well as side chains

H-bonding, therefore, occurs not only within and between polypeptide chains but with the surrounding aqueous medium.

van der Waals Forces in Proteins

van der Waal forces are considered to be weak forces, but since a protein has such a huge number of these interactions, it plays a significant role in folding

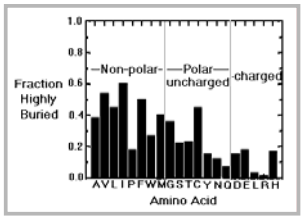


<u>Attractive</u> van der Waals-induced dipoles between adjacent atoms

<u>Repulsive</u> van der Waals-electron-electron repulsion due to the electron clouds overlapping between adjacent atoms

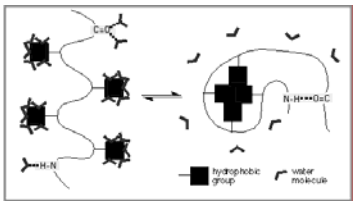
Hydrophobic vs. Hydrophilic Amino Acids

Amino acids in proteins contain either hydrophobic or hydrophilic side chains



It is the nature of the interaction of the different R-groups with the aqueous environment that plays the major role in shaping protein structure.

The Hydrophobic Effect



Increasing randomness

$\Delta S°_{universe} = \Delta S°_{system} + \Delta S°_{surroundings}$

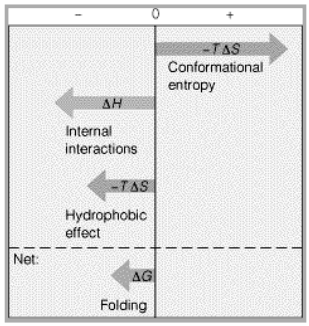
Thermodynamic parameters for folding of some globular Proteins at 25 °C in aqueous solution

Protein	ΔG (kJ/mol)	ΔH (kJ/mol)	ΔS (J/K·mol)
Ribonuclease	−46	−280	−790
Chymotrypsin	−55	−270	−720
Lysozyme	−62	−220	−530
Cytochrome *c*	−44	−52	−27
Myoglobin	−50	0	+170

Note: Data adapted from P. L. Privalov and N. N. Khechinashvili, *J. Mol. Biol.* (1974) 86:665–684. Each data set has been taken at the pH value where the protein is maximally stable; all are near physiological pH. Data are for the folding reaction: Denatured ⟺ native.

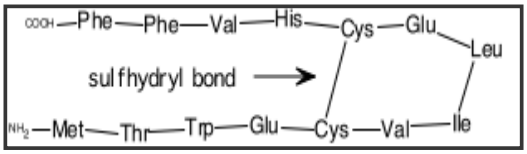
The small negative for cytochrome c and the positive value for myoglobin are a consequence for the hydrophobic effect

Contributions to the free energy of folding of globular proteins



Disulfide Bonds

Once the folding has occurred, the three dimensional structure is in some cases further stabilized by the formation of disulfide bonds between cysteine residues



Once disulfide bonds are formed, they contribute greatly to the stability of the protein

Kinetics of Protein Folding

The folding of globular proteins from their denatured state is a rapid process, often complete in less that a second

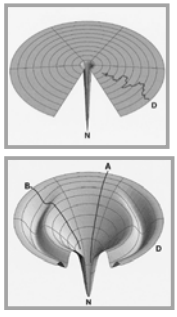
There are about 10^{50} different conformations for the polypeptide ribonuclease

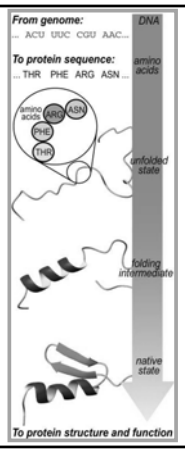
Let's say it tries a new conformation every 10^{-13} second, it would take about 10^{30} years to try significant fraction of them

↓

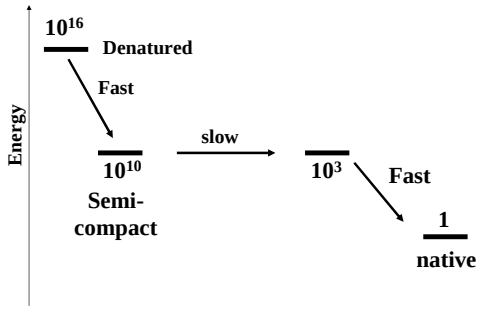
Yet we know that is folds in about 1 minute???????

This dilemma is known As *Levinthal's Paradox*





Kinetic Diagram of Protein Folding

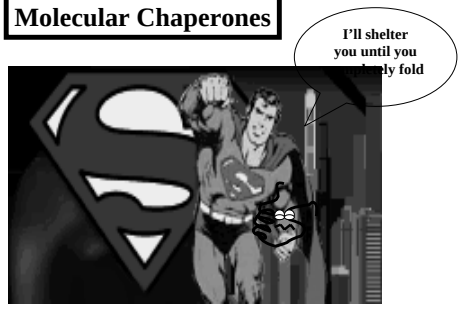


Nature (1994) **369**, 248-251

Quaternary Structure

- Refers to the organization of subunits in a protein with multiple subunits (an "oligomer")
- Subunits (may be identical or different) have a defined stoichiometry and arrangement
- Subunits are held together by many weak, noncovalent interactions (hydrophobic, electrostatic)

Chaperonins or Molecular Chaperones

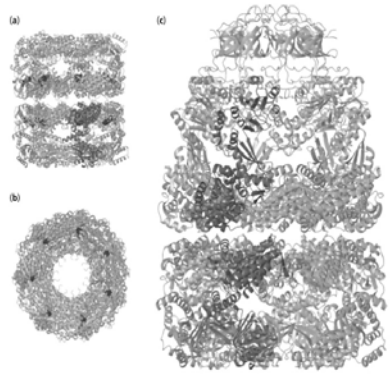


Chaperonins either help prevent protein misfolding or aggregation

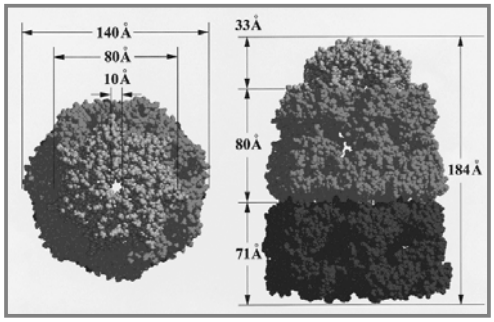
E. coli chaperonin

(a) (b) Core consists of 2 identical rings (7 GroE subunits in each ring)

(c) Protein folding takes place inside the central cavity



Molecular Chaperones



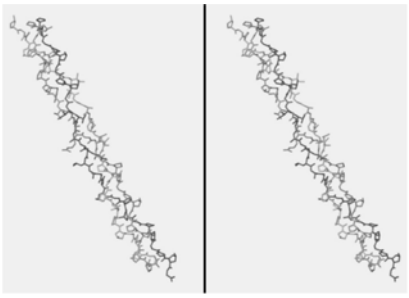
Fibrous proteins

- Provide mechanical support
- Often assembled into large cables or threads
- **α-Keratins**: major components of hair and nails
- **Collagen**: major component of tendons, skin, bones and teeth

Collagen, a Fibrous Protein

- Collagen is a major protein in connective tissue of vertebrates (25-35% of total protein in mammals)
- Diverse forms include tendons (ropelike fibers) and skin (loosely woven fibers)
- Collagen consists of three left-handed helical chains coiled around each other in a right-handed supercoil
- Three amino acids per turn, rise 0.31 nm per residue (collagen is more extended than an a helix)

Stereo view of human Type III collagen triple helix



Collagen triple helix

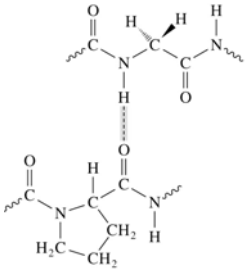
- Multiple repeats of **-Gly-X-Y-** where X is often **proline** and Y is often **4-hydroxyproline**
- Glycine residues are located along central axis of a triple helix (other residues cannot fit)
- For each -Gly-X-Y- triplet, one interchain H bond forms between amide H of Gly in one chain and -C=O of residue X in an adjacent chain
- No intrachain H bonds exist in the collagen helix

4-Hydroxyproline and 5-hydroxylysine

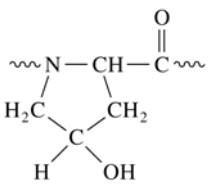
- Formed by enzyme hydroxylation reactions (require vitamin C) <u>after</u> incorporation into collagen
- Vitamin C deficiency (scurvy) leads to lack of proper hydroxylation and defective triple helix (skin lesions, fragile blood vessels, bleeding gums)
- Unlike most mammals, humans cannot synthesize vitamin C

Interchain H bonding in collagen

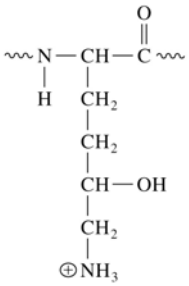
- Amide H of Gly in one chain is H-bonded to C=O in another chain



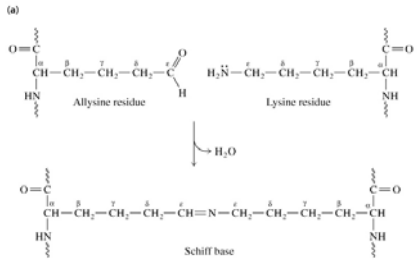
4-Hydroxyproline



5-Hydroxylysine



Covalent cross-links in collagen



Globular proteins

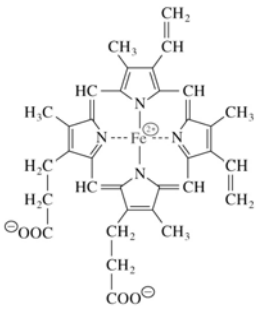
- Usually water soluble, compact, roughly spherical
- Hydrophobic interior, hydrophilic surface
- Globular proteins include enzymes,carrier and regulatory proteins

Structures of Myoglobin and Hemoglobin

- **Myoglobin** (**Mb**) - monomeric protein that facilitates the diffusion of oxygen in vertebrates
- **Hemoglobin** (**Hb**) - tetrameric protein that carries oxygen in the blood
- **Heme** consists of a tetrapyrrole ring system called **protoporphyrin IX** complexed with iron
- Heme of Mb and Hb binds oxygen for transport

Heme Fe(II)-protoporphyrin IX

- Porphyrin ring provides four of the six ligands surrounding iron atom

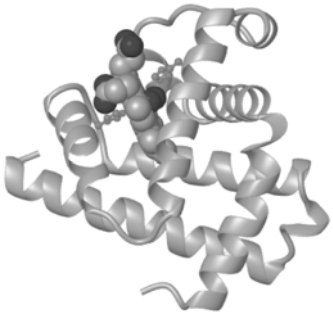


Protein component of Mb and Hb is globin

- Myoglobin is composed of 8 α helices
- Heme prosthetic group binds oxygen
- **His-93** is complexed to the iron atom, and **His-64** forms a hydrogen bond with oxygen
- Interior of Mb almost all hydrophobic amino acids
- Heme occupies a hydrophobic cleft formed by three a helices and two loops

Sperm whale oxymyoglobin

- Oxygen (red)
- His-93 and His-64 (green)

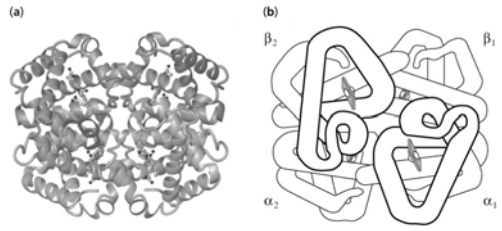


Hemoglobin (Hb)

- Hb is an $\alpha_2\beta_2$ tetramer (2 α globin subunits, 2 β globin subunits)
- Each globin subunit is similar in structure to myoglobin
- Each subunit has a heme group
- The α chain has 7 α helices, β chain has 8 α helices

Hemoglobin tetramer

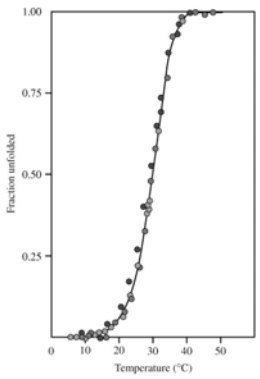
(a) Human oxyhemoglobin (b) Tetramer schematic



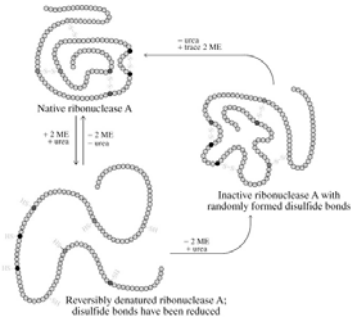
Protein Denaturation and Renaturation

- **Denaturation** - disruption of native conformation of a protein, with loss of biological activity
- Energy required is small, perhaps only equivalent to 3-4 hydrogen bonds
- Proteins denatured by heating or chemicals
- Some proteins can be refolded or renatured

- Heat denaturation of ribonuclease A
- Unfolding monitored by changes in ultraviolet (blue), viscosity (red), optical rotation (green)



Denaturation and renaturation of ribonuclease A



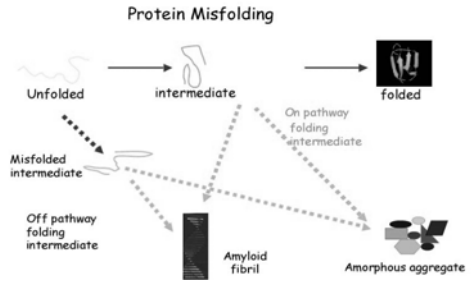
Why Worry About Protein Misfolding or Aggregation????

Alzheimer's

What are the plaques that form that cause cell death?

They are proteins that MISFOLD and begin to aggregate into β-sheets.

The once small protein that usually is excreted is now too large to move through membrane and begins to kill the cells



Mad Cow Disease

The Prion Protein

Misfolds and becomes a scrapie prion

Bumps into normal folding proteins and causes them to go toward a scrapie prion