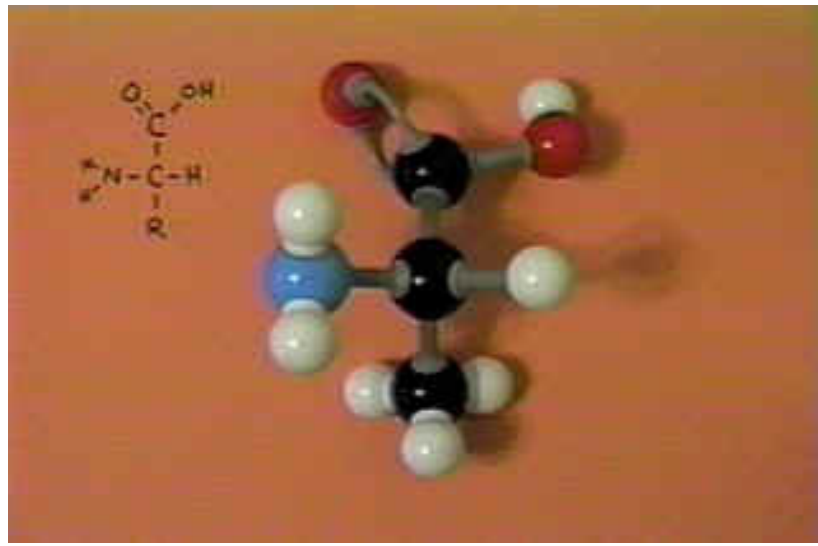
Basic Building Blocks: Proteins

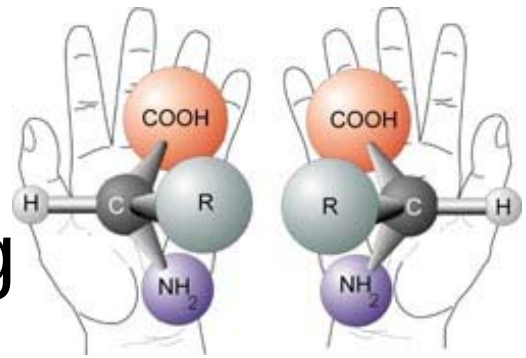
- Largest variety of biomolecules
- Most of the weight of cells, aside from water
- Basic unit is amino acid
- Form of amino acid
- Simplest is glycine

  with R = H
- All others are asymmetric

two stereoisomers L & D

with mainly L naturally occuring





Human Genome Project Facts

- Human DNA codes about 30,000 genes

  (vs. fruit flies:13,500 and C. elegans: 19,000)
- These genes represent only ~ 1% of DNA – lots of coding for control & transcription factors
- Average human protein has ~450 amino acids
- One of the largest proteins is titin (27,000 amino acids in a single chain)

Protein Functions

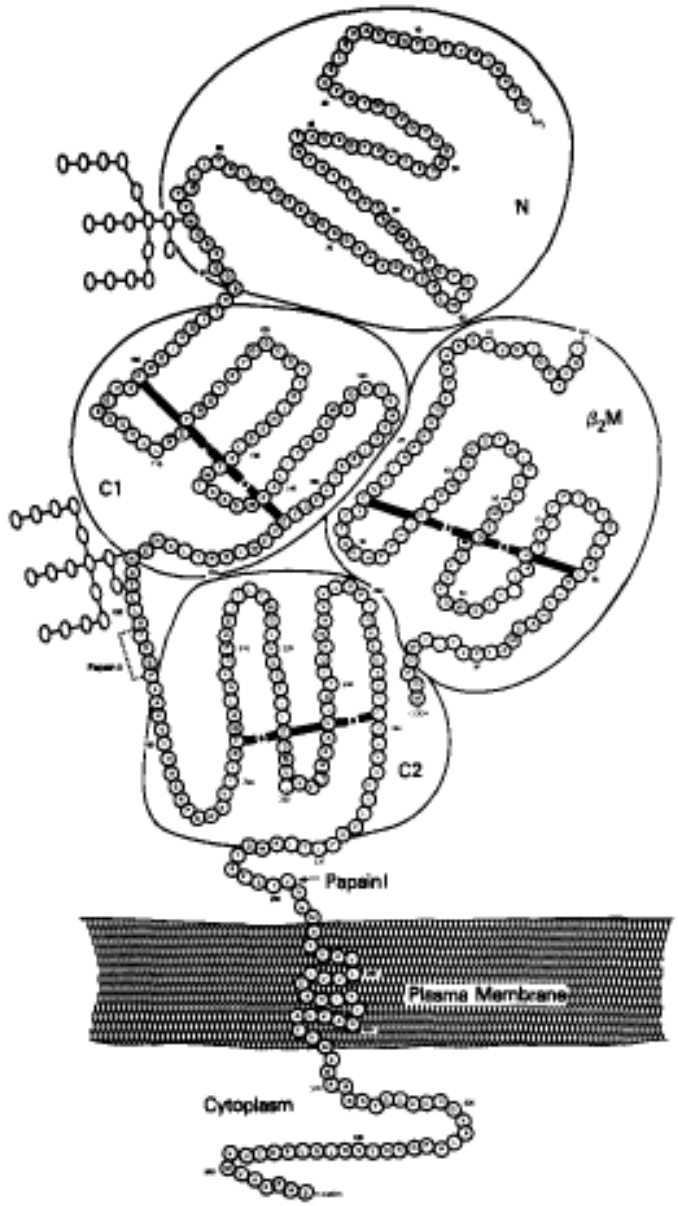
- Motion & locomotion of cells/organism (contractile proteins)
- Catalysis of all biochemical reactions (enzymes)
- Structure of cells and extracellular matrix (e.g. collagens)
- Receptors for hormones/ signaling molecules
- Transcription factors
- Etc.

Example Protein (H-2K) - Structure

- This antigen displays many features of proteins
  - Two polypeptide chains
  - Longer heavy chain has 5 domains – 3 extracellular, one transmembrane, and one cytoplasmic – it is called an integral membrane protein
  - Smaller polypeptide chain is attached to heavy chain by H bonds (no covalent bonds) – it is a peripheral membrane protein
  - The dark bars are disulfide bridges (S-S)
  - Two short branched sugars are on the left making this a glycoprotein (sugar + protein compex)

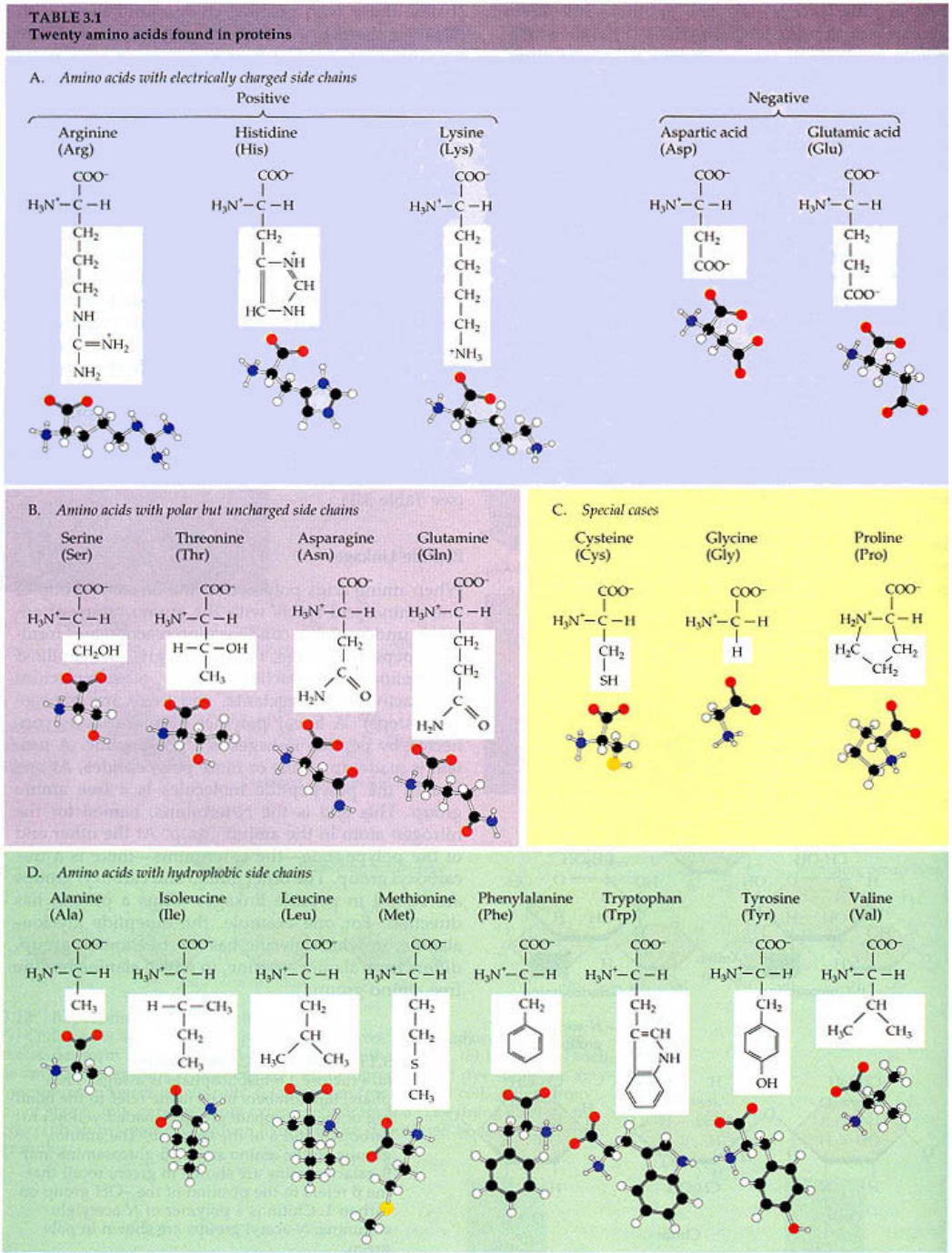
  The view seen here does not show its real 3D arrangement

  Look in PDB



Types of amino acids

- Classify aa by various criteria – each has 3 letter or 1 letter code
- 3 have ring-structures – important in fluorescence
- All are ampholytes (+/- charge depending on pH)



Amino Acids

Amino Acid Codes

Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Aspartate	Asp	D	Methionine	Met	M
Asparagine	Asn	N	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamate	Glu	E	Serine	Ser	S
Glutamine	Gln	Q	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

Digression: pH ideas

- pH = -log[H^+]
- Neutrality when [H^+]=[OH^-]=10^{-7} M
- Higher pH – basic; lower – acidic
- Simple idea: $H_2O \longleftrightarrow OH^- + H^+$
- Dissociation constant K

$$K = \frac{[H^+][OH^-]}{[H_2O]} = e^{-\Delta G / kT}$$

where G = free energy per mole of bond formation; with [H_2O] = 55 M ~ constant

So K' = [H^+][OH^-] = 10^{-14} and pK = -log K in general

pH and pK

- Each charged group has a pK
- For proteins, e.g.,
  - $COOH \longrightarrow COO^- + H^+$ pK 2.34
  - $NH_3^+ \longrightarrow NH_2 + H^+$ pK 9.69
  - R group dissociation also

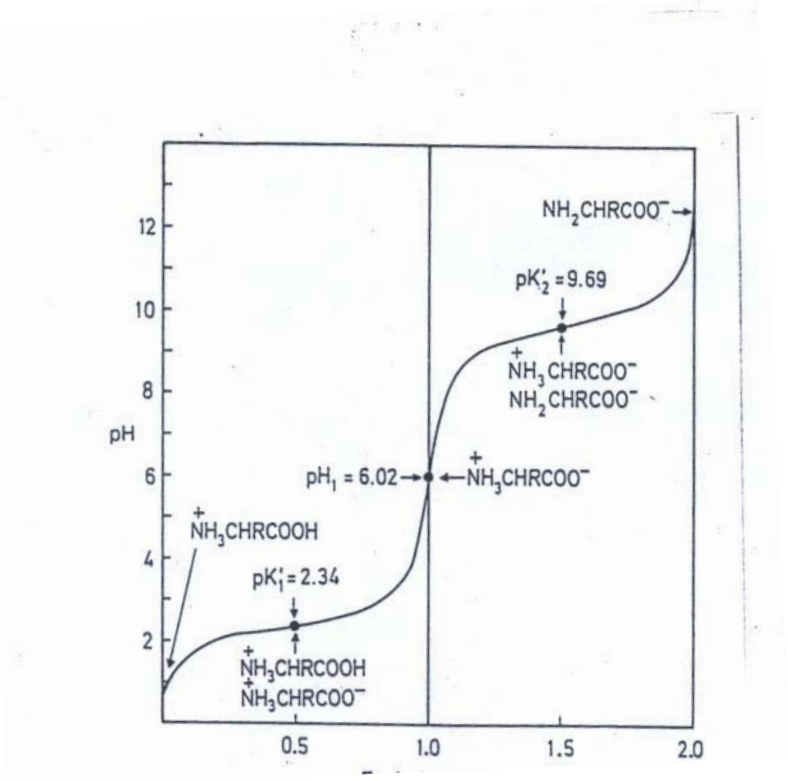
  If pH > pK $\longrightarrow$ more basic form

  If pH < pK $\longrightarrow$ more acidic form

  Different forms predominate at different pH - polyelectrolyte

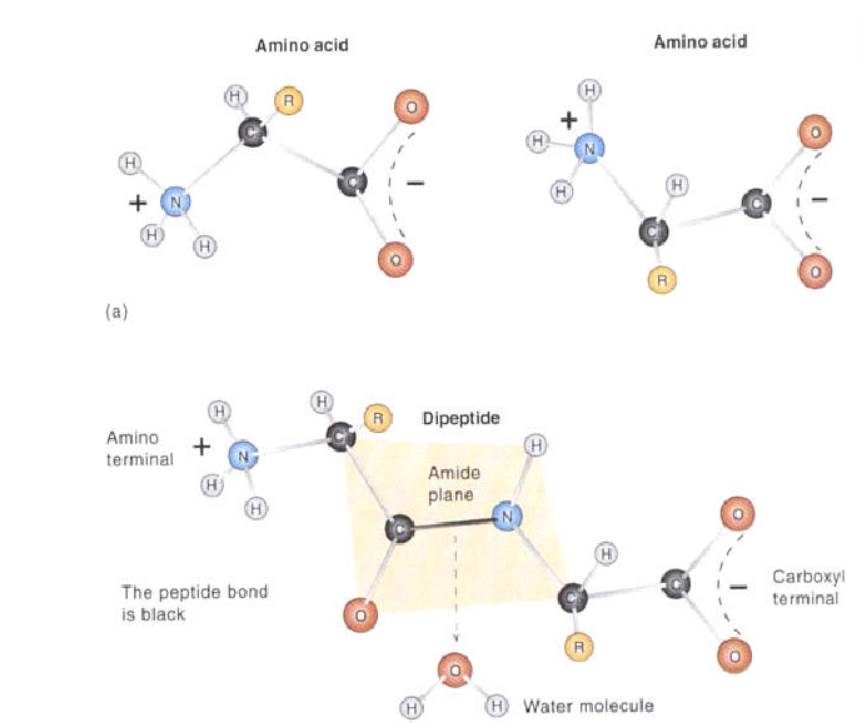
Example: Titration of alanine

- Different forms at different pH
- Alanine has R = CH_3
- pI = isoelectric point = pH at which neutral



Peptide bond

- Amino acids link together to form a continuous linear chain = backbone of protein



Primary Structure

- With even only 10 a.a. long – number of possible polypeptides (decamers) = $20^{10} = 10^{10}x2^{10} \sim 10^{13}$
- Amino acid composition – not sequence – can be automatically determined by aa analyzer to give % composition
- General features of 1^{o} structure:
  - Most polypeptide chains are 100 – 500 aa; smallest 25 – 100, largest 3000
  - Some proteins have more than 1 chain – held together by weaker non-covalent bonds
  - Protein data bank – on-line

Facts about 1º structure

- Wide variation in composition
- Certain aa are fairly rare (methionine, Tryptophan)
- Ala, Leu very common
- Many proteins contain other molecules, including carbohydrates, metal ions (Ca, Fe, Zn, Cu)

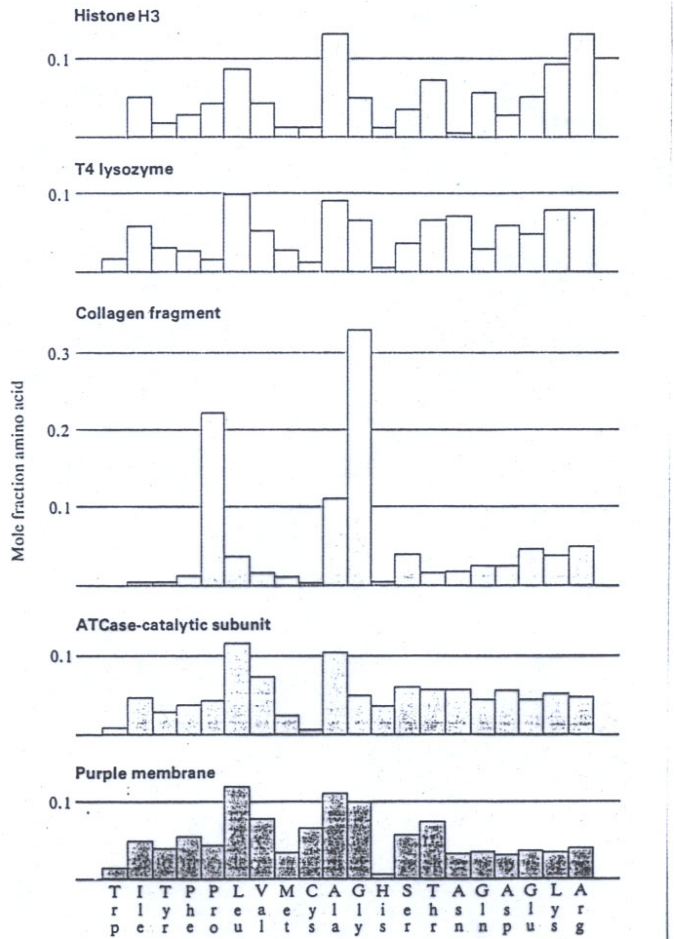
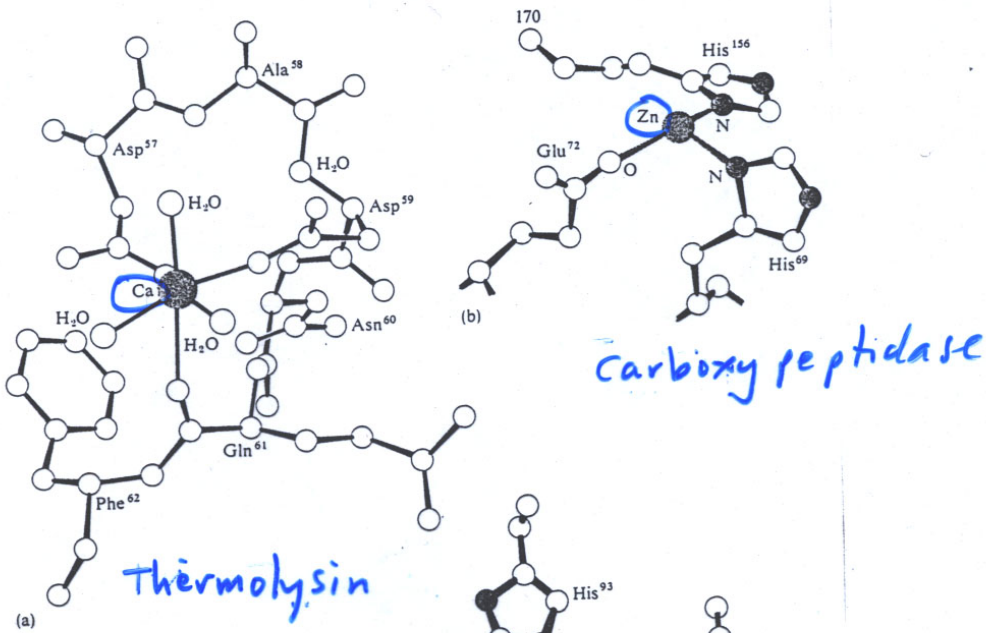


Figure 2-3

Amino acid composition of five proteins. The mole fraction of each amino acid type is plotted in approximate order of decreasing hydrophobicity.

Metal Ions in Proteins



Secondary Structure (2º) of Proteins

- Backbone of protein chain has series of rotatable bonds. Two angles describe possible rotations of each peptide
- Rotations about these bonds lead to certain allowed structures – or stable conformations

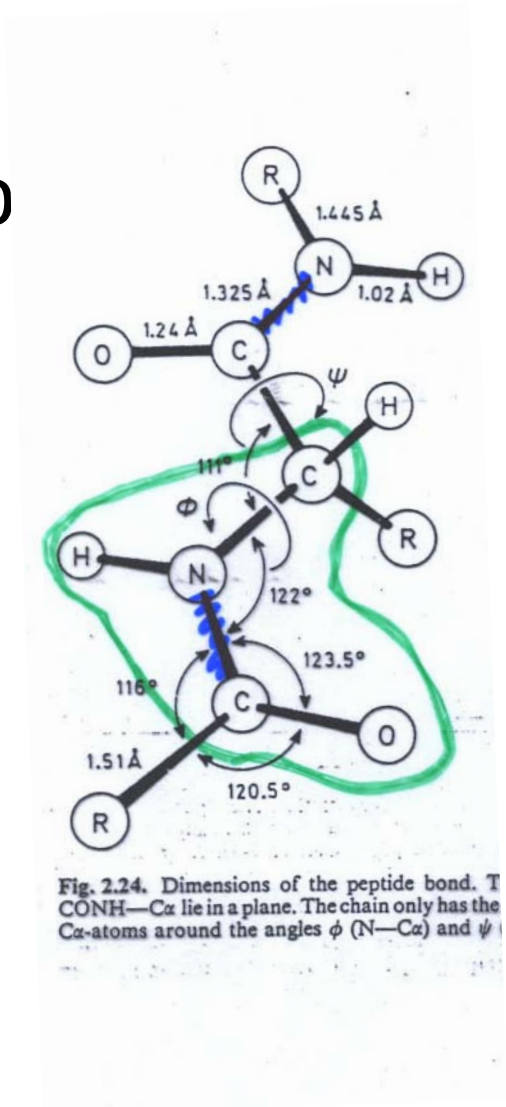
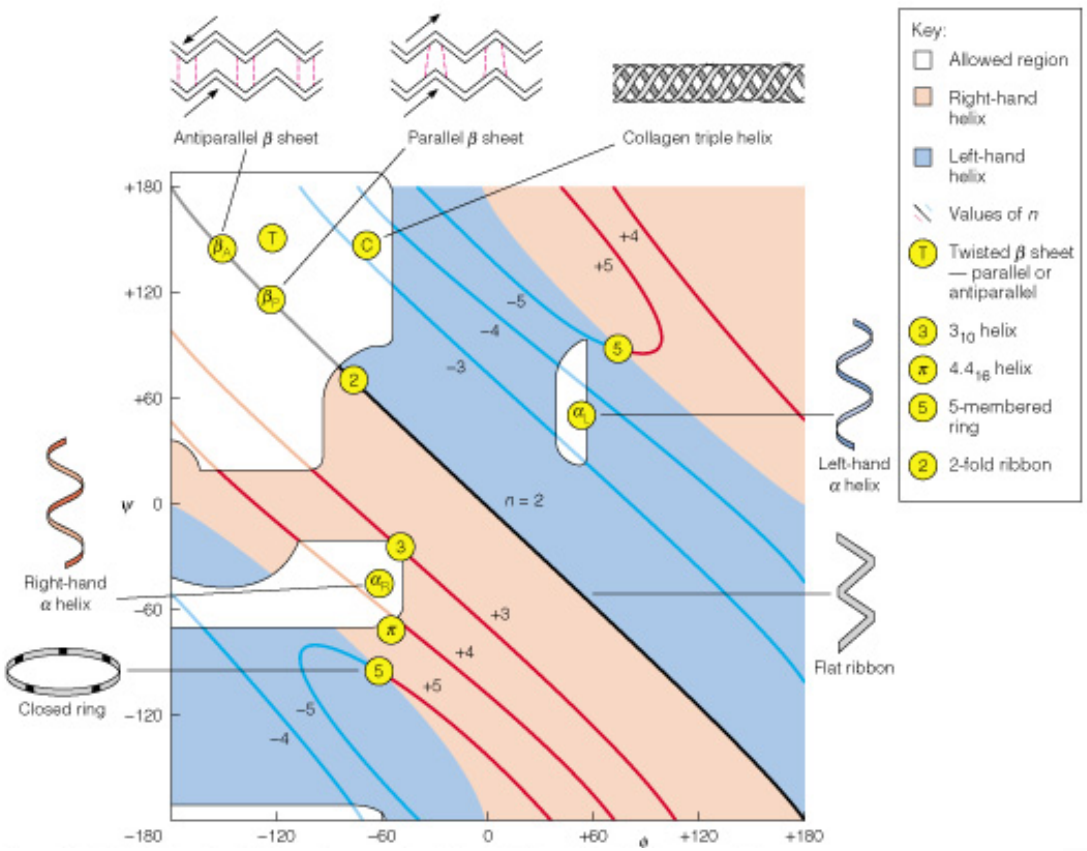


Fig. 2.24. Dimensions of the peptide bond. T
CONH—Cα lie in a plane. The chain only has the
Cα-atoms around the angles φ (N—Cα) and ψ

Ramachandran Diagram

- A number of helices and β sheets are possible



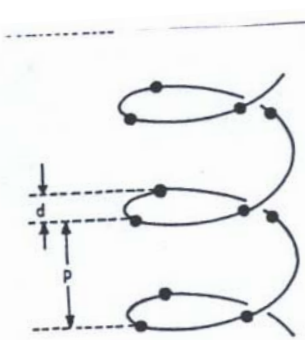


Fig. 2.26. Structural parameters of a helix. The parameters for stable helices are given in Table 2.5

Table 2.5. Structural parameters of important polypeptide conformations

	α-helix	3_{10}-helix	2_7-band	Poly-prolin helix	Antiparallel β-pleated sheet structure
	132°(113°)	131°(106°)	105°	103°	40°
	123°(136°)	154°(176°)	250°	326°	215°
n	3.61	3.00	2.00	−3.00	2.00
d[nm]	1.50	2.00	2.80	3.12	3.47
p[nm]	5.41	6.00	5.60	9.36	6.95

n = Number of repeating units per helix turn.
d = The raise along the helix axis per repeating unit.
p = Pitch of a helix.

α-helix + β-sheet

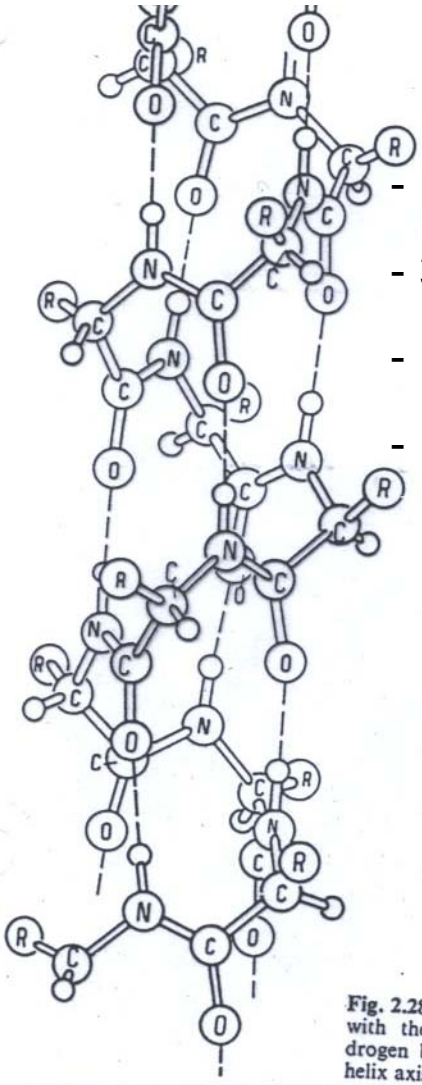


Fig. 2.28. Model of an α-helix with the intramolecular hydrogen bonds parallel to the helix axis

- Right-handed
- 3.6 aa per turn
- R groups outside
- H bond between
  - -NH and –C=O
  - 4 aa apart pointing along axis

-Pairs of chains lying side by side

- Stabilized by H bonds

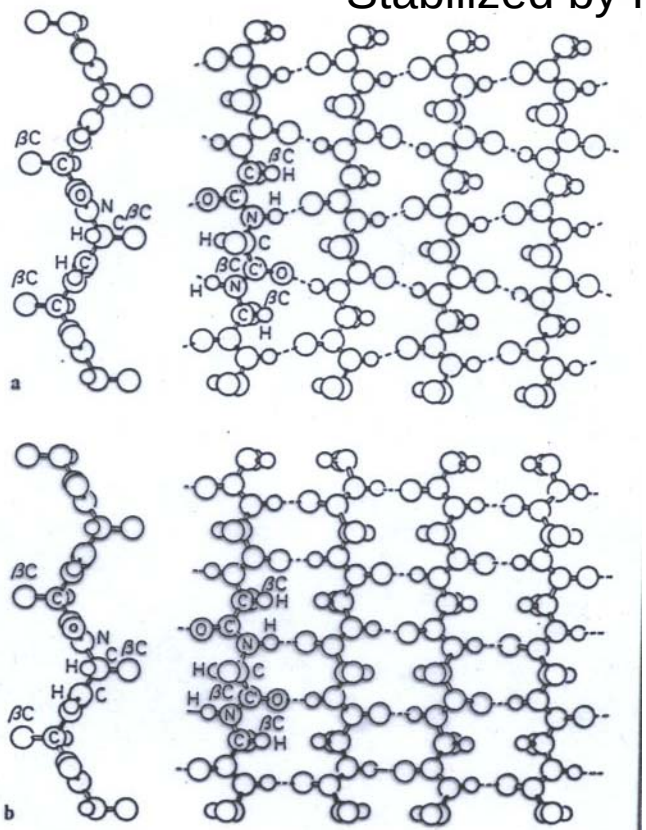
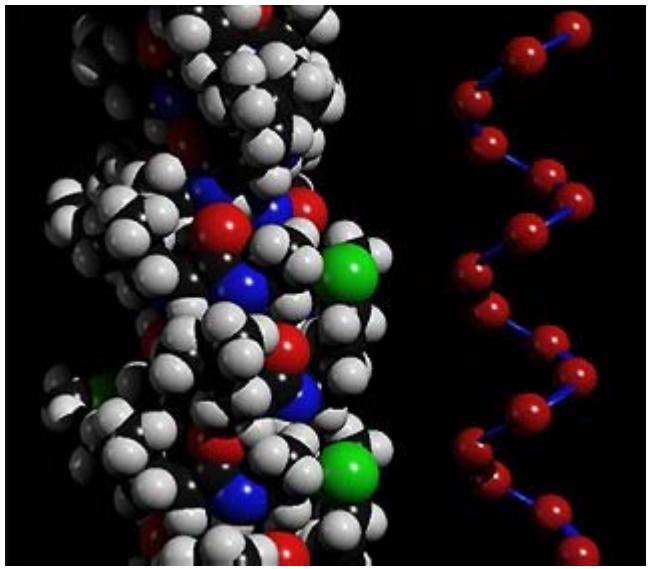
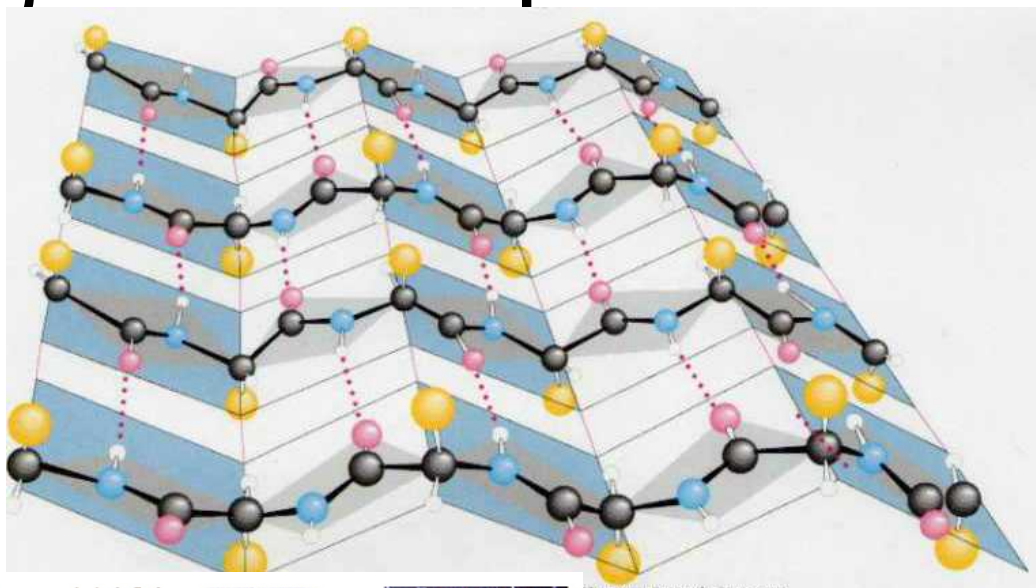


Fig. 2.30a, b. Representation of a the parallel, and b the antiparallel pleated-sheet structure. [After Pauling, L. and Corey, R. B.: Proc. Natl. Acad. Sci. U.S. 37, 729 (1951)]

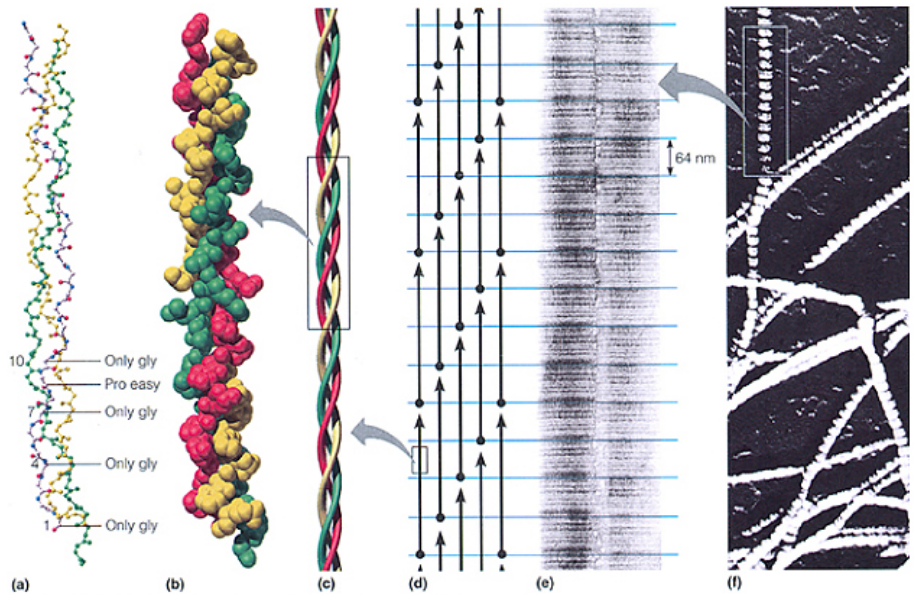
More α-helix, β sheet, triple helix





Harcourt, Brace & Company)

Collagen triple helix



All proteins consist of regions of 2nd structure w/ random coil connections

Prediction of structure

- Based on knowing aa sequence, we are able to predict α-helix, β-sheet regions
- For example:

  residues 1-36 in histone have 12 + charges – able to bind to neg. charges on d-s DNA
- For example: glycophorin from human RB cells spans membrane from 73 – 95 non-polar region

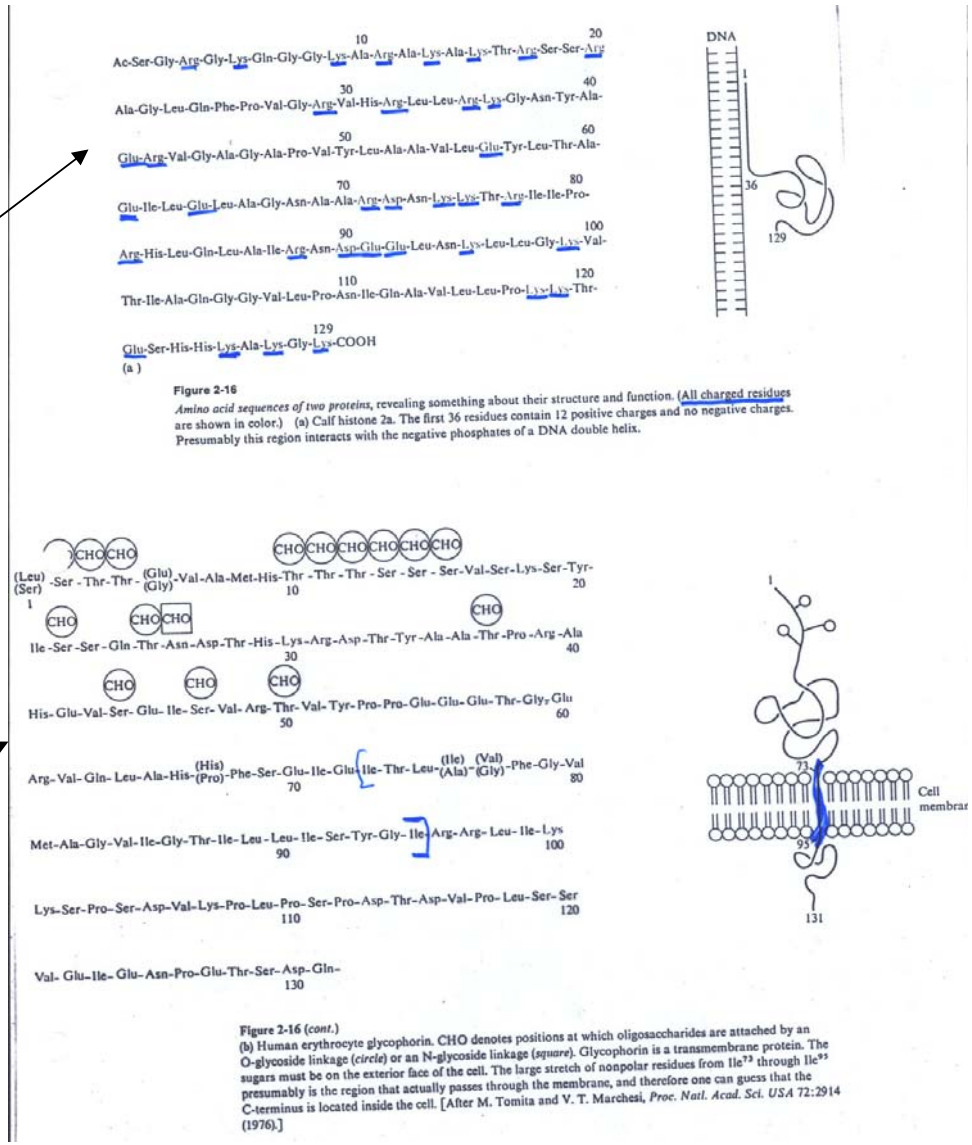


Figure 2-16
Amino acid sequences of two proteins, revealing something about their structure and function. (All charged residues are shown in color.) (a) Calf histone 2a. The first 36 residues contain 12 positive charges and no negative charges. Presumably this region interacts with the negative phosphates of a DNA double helix.

Figure 2-16 (*cont.*)
(b) Human erythrocyte glycophorin. CHO denotes positions at which oligosaccharides are attached by an O-glycoside linkage (*circle*) or an N-glycoside linkage (*square*). Glycophorin is a transmembrane protein. The sugars must be on the exterior face of the cell. The large stretch of nonpolar residues from Ile^{73} through Ile^{95} presumably is the region that actually passes through the membrane, and therefore one can guess that the C-terminus is located inside the cell. [After M. Tomita and V. T. Marchesi, *Proc. Natl. Acad. Sci. USA* 72:2914 (1976).]

Prediction of Structure II

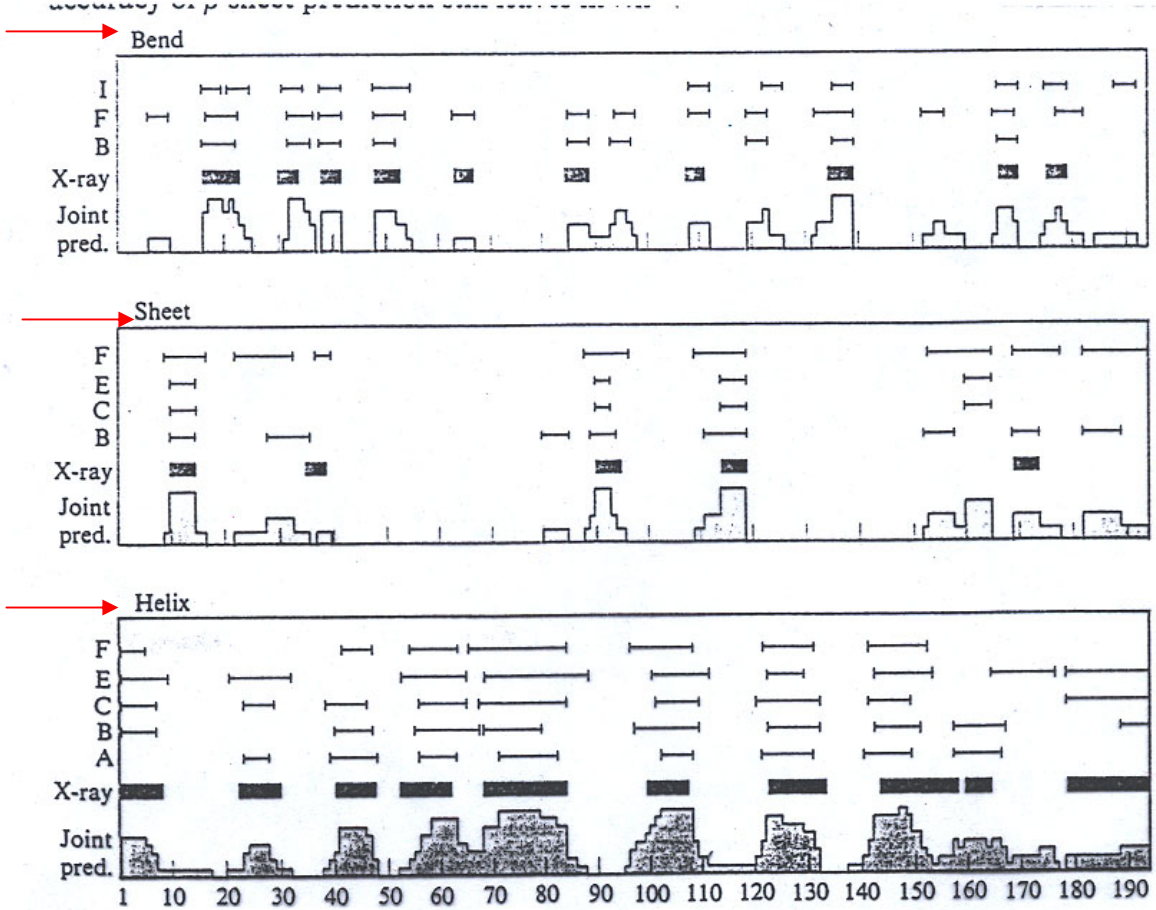


Figure 2-13

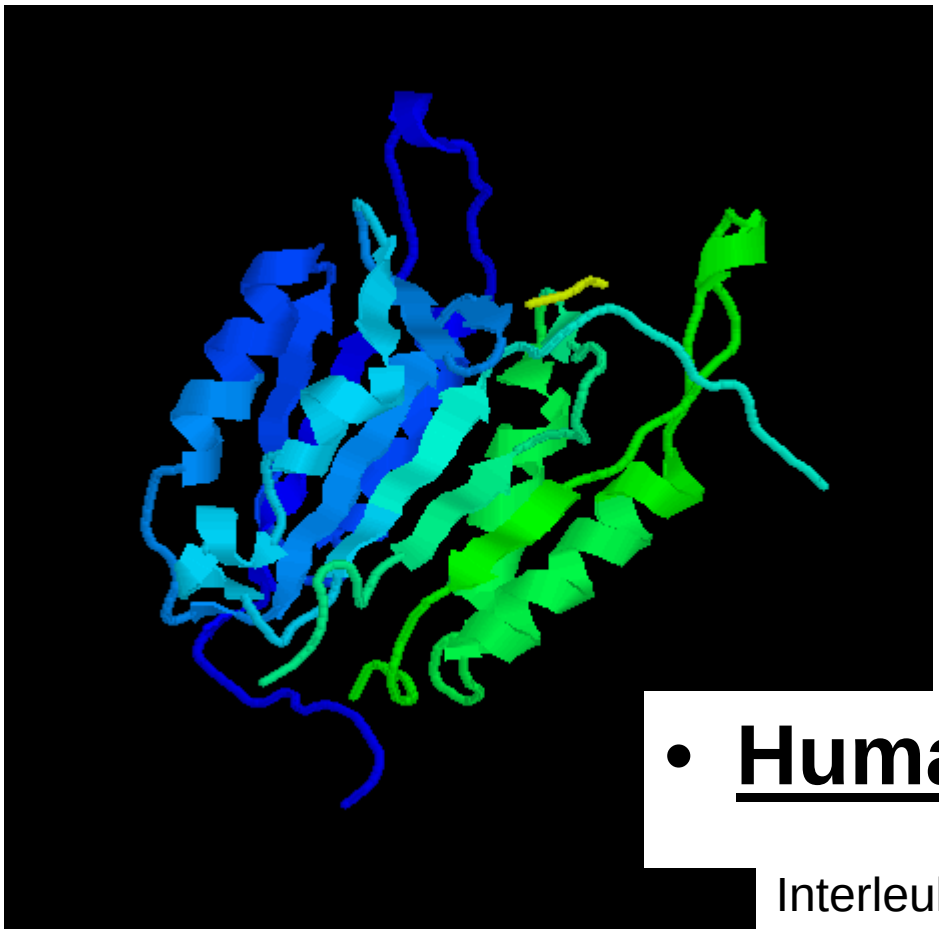
Distribution of secondary structure types in adenylate kinase. Regions predicted by several different theoretical methods to be in β bends, β sheets, or α helices are shown. Actual secondary structure regions found by x-ray crystallography are indicated, as is the average predicted distribution considering the various theories jointly. The predictions were made before the crystal structure was known. [After G. E. Schultz et al., *Nature* 250:140 (1974).]

Protein Folding Problem

- Big Question is: If you know the primary sequence of aa can you predict the 3-D structure of a protein? [Protein-folding problem – one of challenges]
- Can occur spontaneously – involves basic electrical interactions that we'll study soon
  - Co-valent bonds along backbone
  - H-bonds – weaker, directional
  - Van der Waals – non-specific attractive
  - Hydrophobic/ hydrophilic – entropy driven forces

Tertiary Structure (3º)

- All proteins consist of 2º structure regions connected by random coil



- **<u>Human ICE-protease</u>**

Interleukin-1ß-converting enzyme

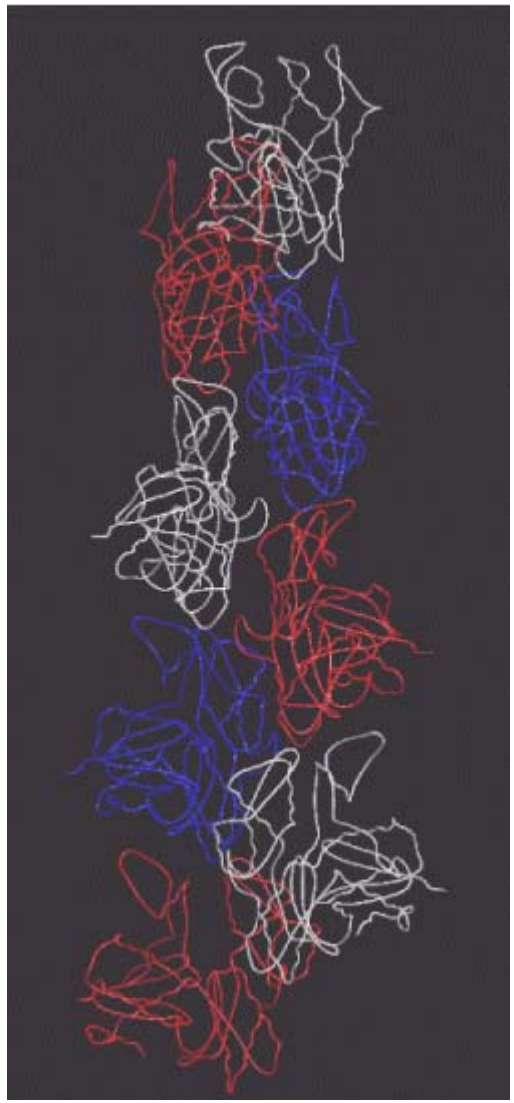
Protein Domains

- Tertiary structure of proteins is built up from domains
- Each domain has a separate function to perform – for example:
  - Binding a small ligand
  - Spanning the plasma membrane
  - Containing a catalytic site
  - DNA binding (transcription factors)
  - Providing a binding surface for another protein
- Often each domain is encoded by a separate exon in the gene encoding that protein – this correspondence is most likely to occur in recently-evolved proteins (exon shuffling idea to generate new proteins using established domains – like Lego pieces)

Fibrous Proteins

- Two major classes of proteins based on 3^{o} Structure
  - Fibrous – fiber-like, includes
    - Keratins – in hair, horns, feathers, wool
    - Actin – muscle thin filaments, cells
    - Collagen – connective tissue

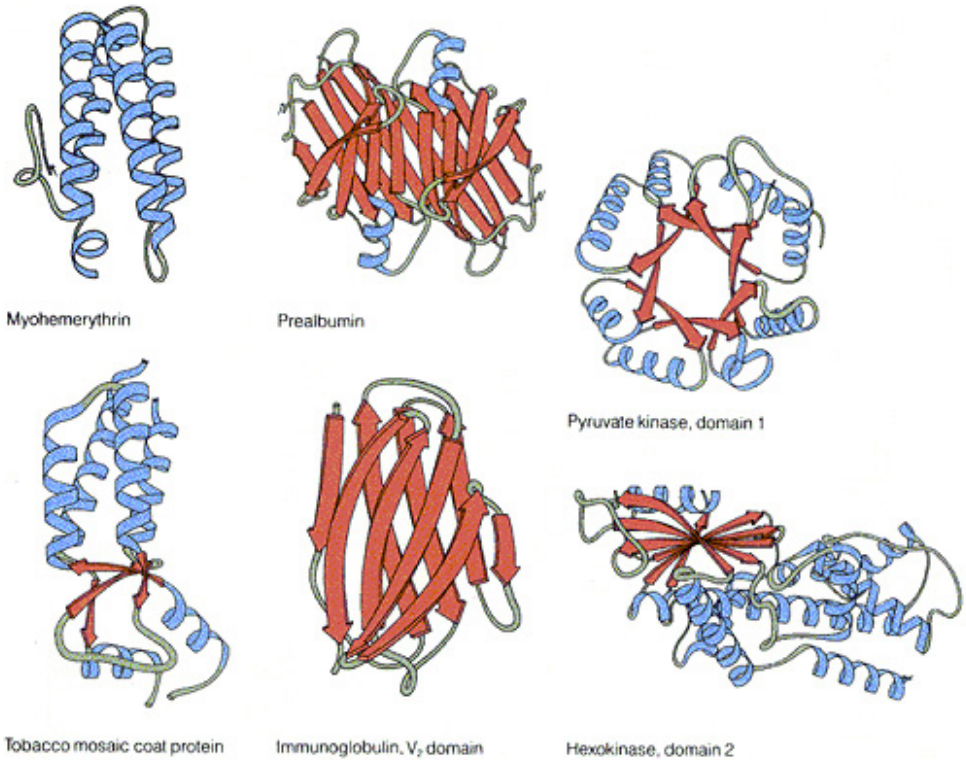
    Often these are polymers made up from monomer subunits and form all α helices and/or all β sheets (e.g. silk)



Actin filament made from monomers

Globular Proteins

- Second class is globular – most enzymes, hormones, transport proteins – folded up structure

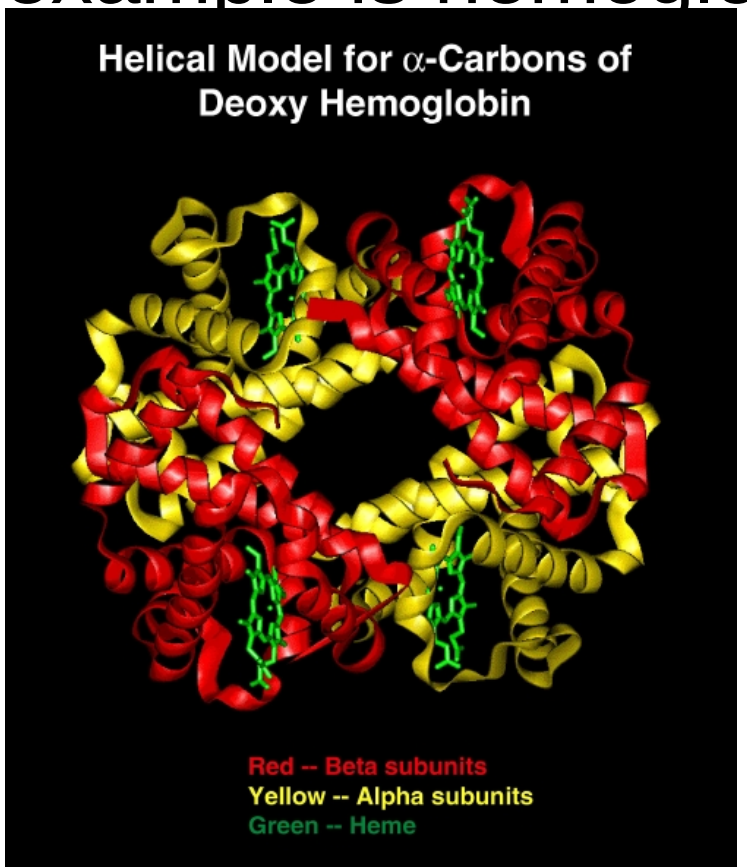


General Properties of 3º Structure

1. Lowest energy states are most stable 3º structures
2. Charged residues are on surface or exposed clefts
3. Non-polar (hydrophobic) residues are internal
4. Nearly all possible H-bonds form

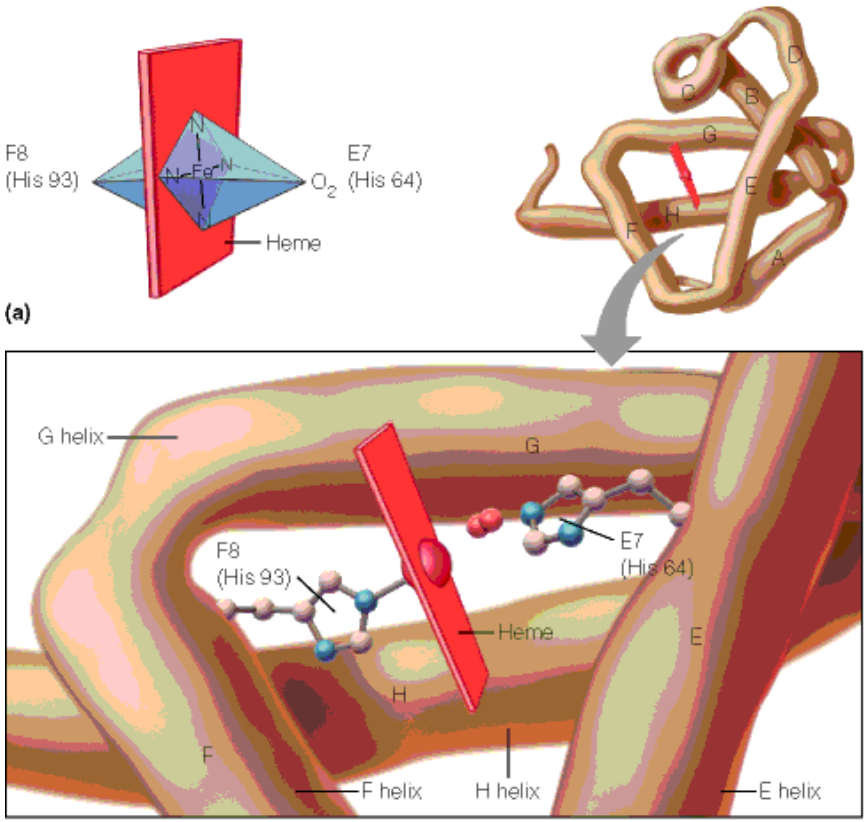
Quaternary (4^o) Structure

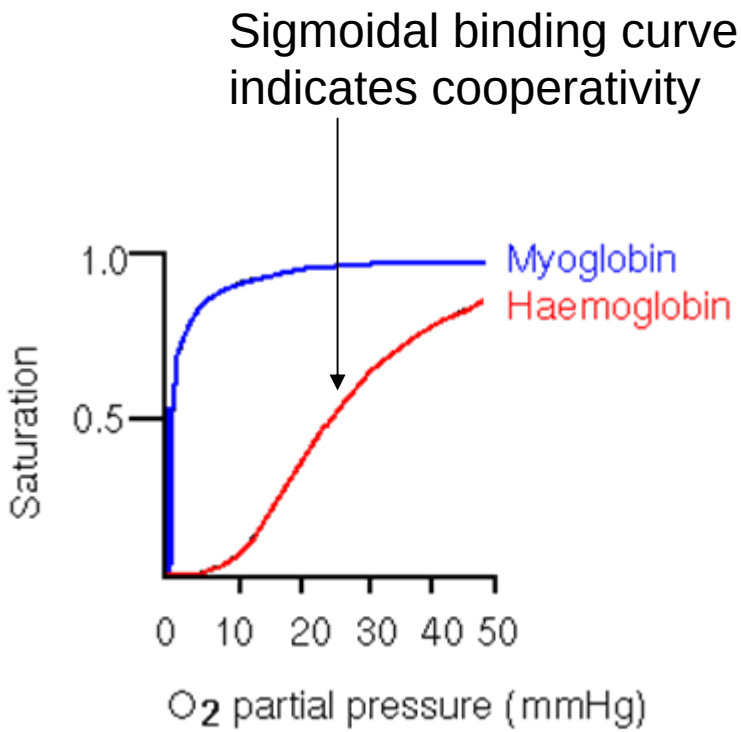
- Multiple sub-units bound together non-covalently
- Canonical example is hemoglobin:



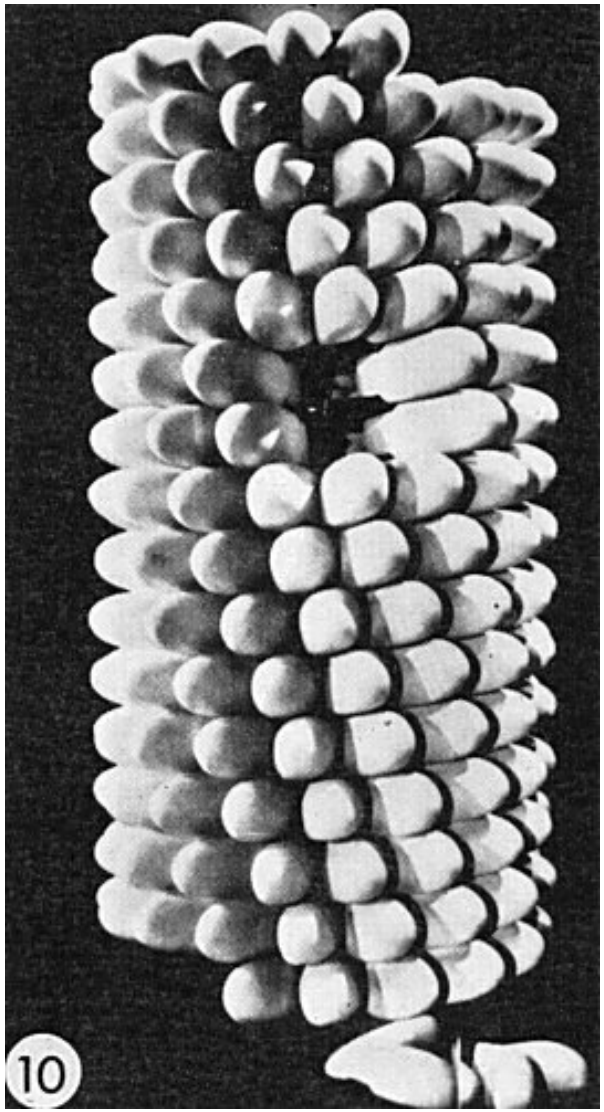
Cooperative Binding by Hemoglobin

- Fe in the heme group binds oxygen – separately, each of 4 hemes binds O_2 as in myoglobin – 4 together bind O_2 cooperatively – Allosteric conformational change





TMV – 4º structure



Packing Density of Proteins

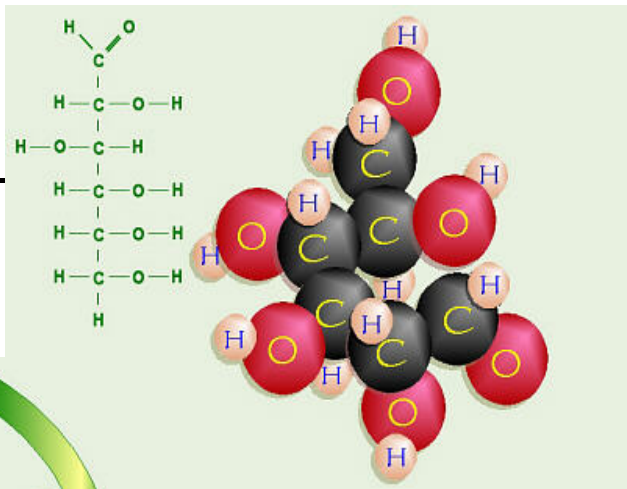
- How filled is volume of protein?
- Quantitative measure = packing density =

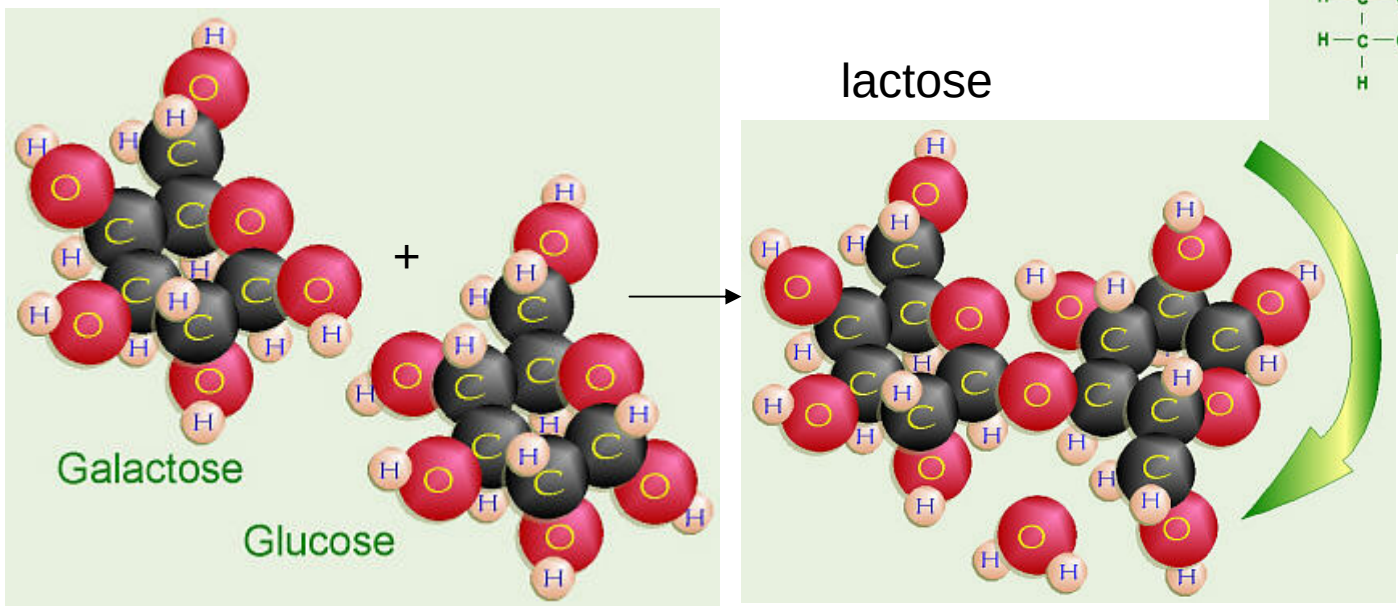
$$PD = \frac{\text{volume enclosed by all van der Waals R}}{total \quad volume}$$

- For continuous solid PD = 1
- For close packed spheres PD = 0.74
- For close packed cylinders PD = 0.91
- For ribonuclease S, PD = 0.75

Two Other Classes of Biomolecules- Polysaccharides + Lipids

- Polysaccharides (carbohydrates)
  - Monosaccharide – eg glucose
  - Disaccharide – eg lactose
  - Polymers of sugars – M ~ 10^4 -

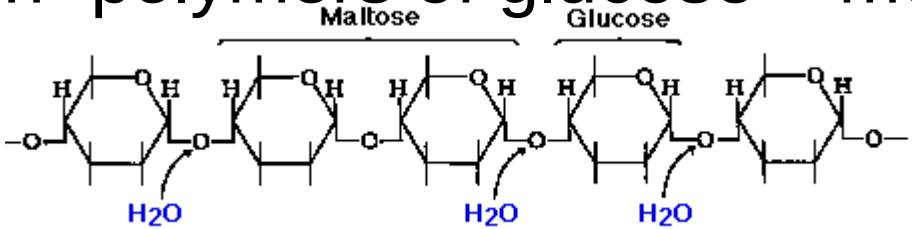




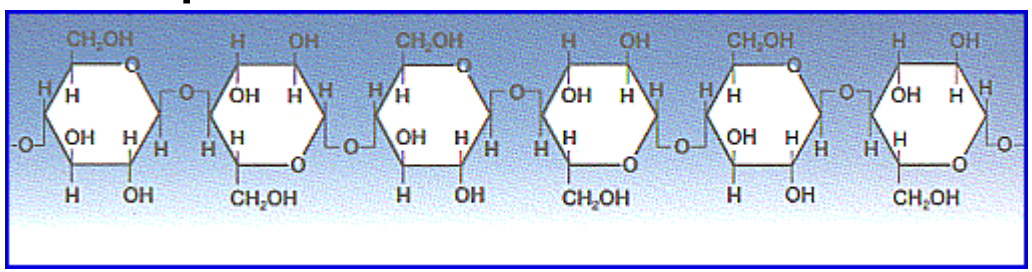
lactose

Polysaccharides - con't

- Glucose can polymerize into 3 types of polymers
  - Starch- polymers of glucose – metabolic

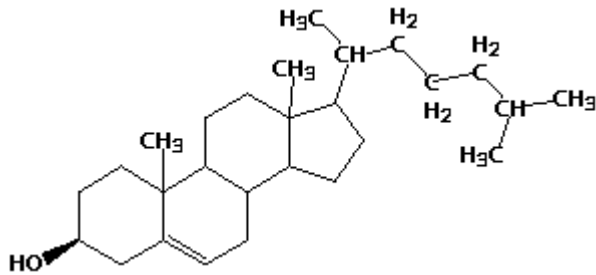


  - Glycogen- ditto, but with more shorter branching –also metabolic- stores glucose
  - Cellulose – most prevalent biomolecule - structural

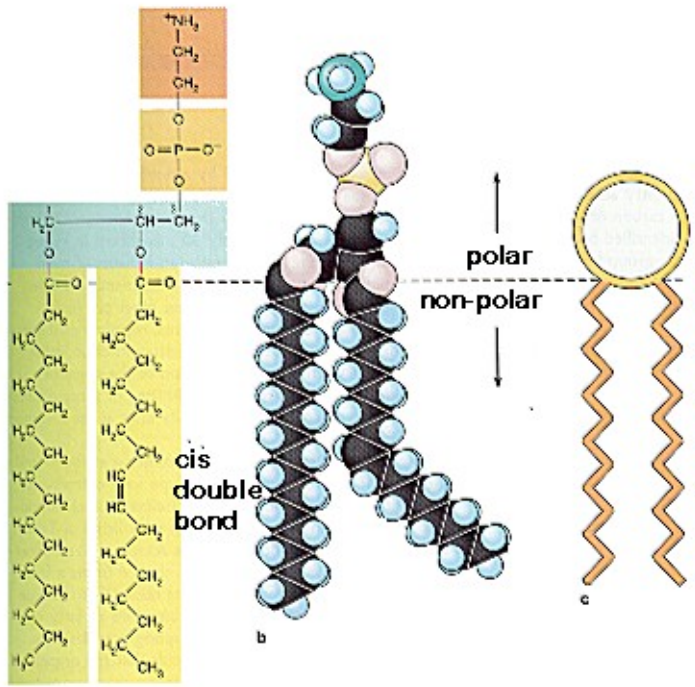


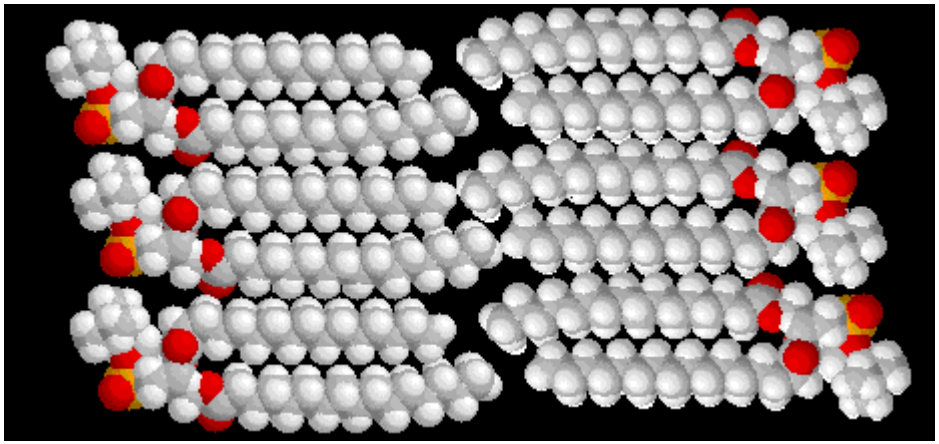
Lipids

- Very diverse family – all insoluble in water/ rich in hydrocarbons
- Includes fatty acids, steroids, phosphoglycerides/phospholipids in membranes
- Polar head group = fatty acid tail with 12 – 24 C's in tail

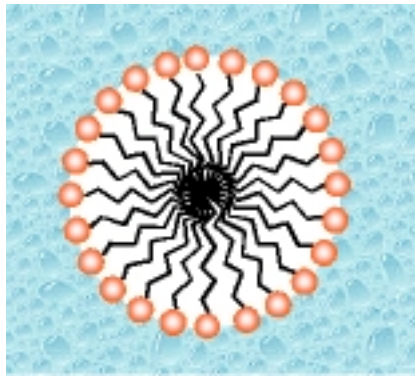


cholesterol

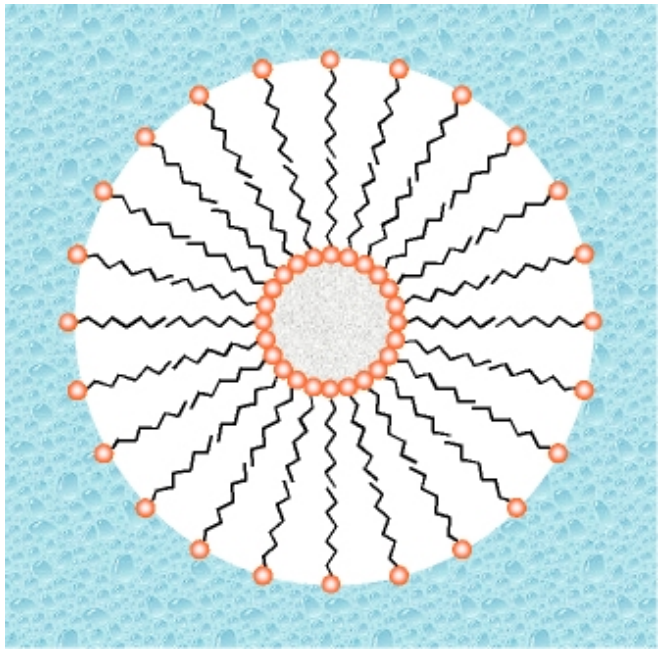




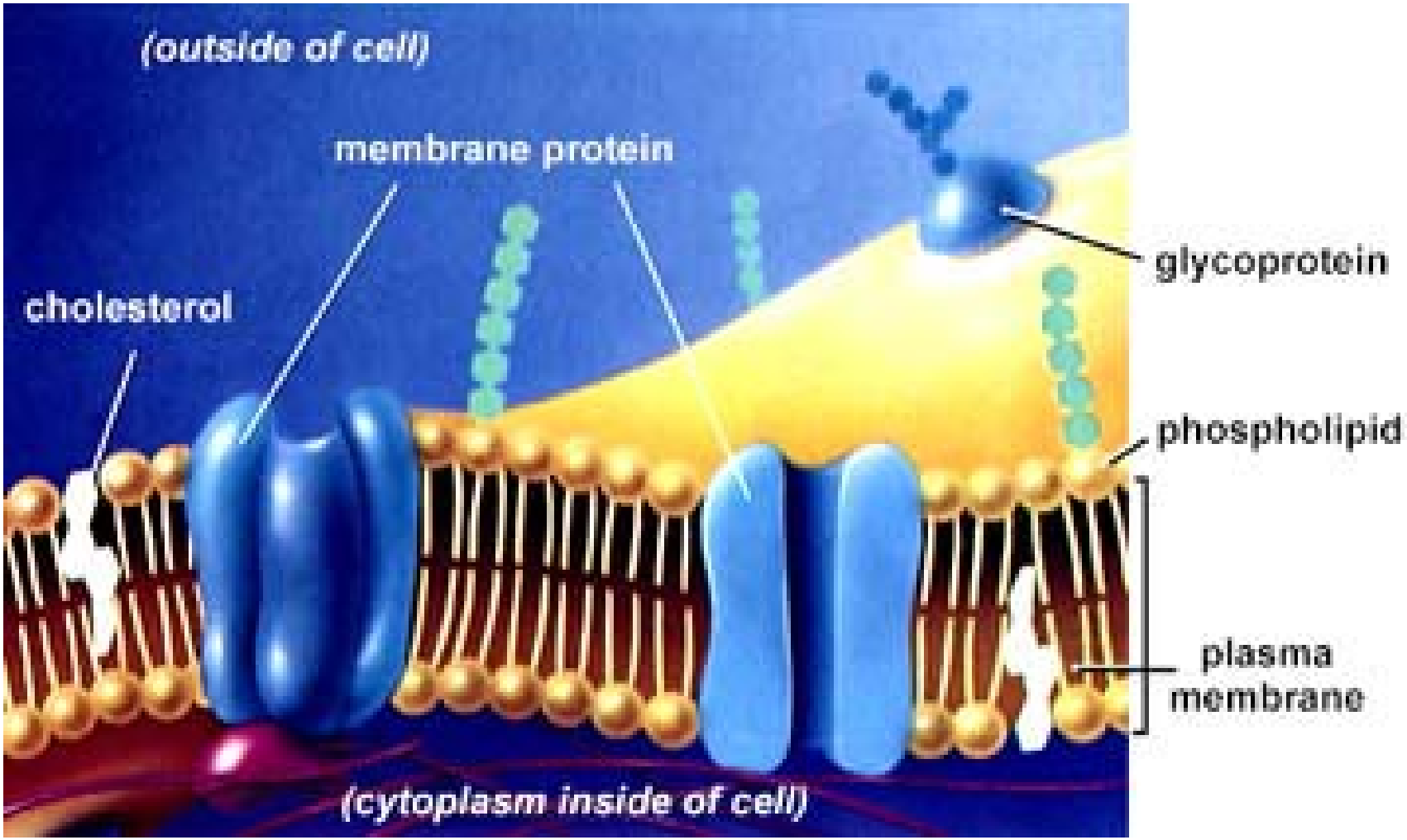
bilayer



micelle



Vesicle (unilamellar)


(outside of cell)
membrane protein
glycoprotein
cholesterol
phospholipid
plasma membrane
(cytoplasm inside of cell)