• No permission is granted for distribution of the course or any use for profit.
• Copyright for any of the supplementary material is that of the respective Copyright owners and those materials may not be used or distributed without their prior permission.
• Please note that only some of the programs examples are currently available online.

Contact address:
Department of Chemistry, UCD,
Belfield, Dublin 4, Ireland,
E-mail: Edward.Timoshenko@ucd.ie
http://chemistry.ucd.ie/timoshenko

Biopolymers

4 lectures

Dr Edward G. Timoshenko

This course is available
at URL: http://darkstar.ucd.ie/courses/bp

Outline:

1. Introduction. Physical Chemical approach to modelling biological molecules. Cell organisation and biopolymers involved.
2. Protein structure, conformation and function.
3. Nucleic acids: DNA and RNA conformation and complexes.
4. Inter–molecular interactions and self–assembly. Biomaterials and bio nano science.

References:

(1) H. Lodish, D. Baltimore, A. Berk, S.L. Zipursky, P. Matsudaira, J. Darnell. *Molecular Cell Biology.* **Scientific American Books, N.Y. (1995).**
(2) C. Branden, J. Tooze. *Introduction to Protein Structure.* **Garland Publishing, N.Y. (1999).**
(3) L. Stryer. *Biochemistry.* **Freeman and Co. N.Y. (1995).**
(4) J.R. Holum. *Fundamentals of General, Organic and Biological Chemistry.* **Wiley. N.Y. (1998).**
(5) P.W. Atkins. Physical Chemistry. **5th Ed. Oxford Univ. Press. (1994).**
(6) Alan Hinchliffe. *Modelling molecular structures.* **Wiley. Chichester (2000).**
(7) Frank Jensen. *Introduction to Computational Chemistry.* **Wiley. Chichester (1999).**

Lecture 1

Born–Oppenheimer approximation
The full Schrödinger equation for a molecule is,

$$-\frac{\hbar^2}{2}\sum_n \frac{\partial_n^2\psi}{M_n} - \frac{\hbar^2}{2m_e}\sum_i \partial_i^2\psi - \sum_{i,n}\frac{Z_n e^2}{4\pi\epsilon_0\, r_{in}}\psi +$$

$$+\sum_{i>j}\frac{e^2}{4\pi\epsilon_0\, r_{ij}}\psi + \sum_{m>n}\frac{e^2\, Z_m Z_n}{4\pi\epsilon_0\, r_{nm}}\psi = E\,\psi, \tag{1}$$

where i, j refers to e and n, m to nuclei.
Taking into account that nuclei are much heavier than e

$$-\frac{\hbar^2}{2m_e}\sum_i \partial_i^2\psi_e - \sum_{i,n}\frac{Z_n e^2}{4\pi\epsilon_0\, r_{in}}\psi_e + \sum_{i>j}\frac{e^2}{4\pi\epsilon_0\, r_{ij}}\psi_e = V_e(\vec{r}_n)\,\psi_e, \tag{2}$$

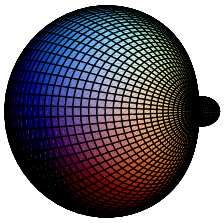
we obtain motion of e on potential energy surface $V_e(\vec{r}_n)$.
If Coulomb repulsion in Eq. (2) is further neglected the total electronic function becomes product of single e wave functions **moving in some central field. This is known as** orbital approximation.
Although the repulsion term can be important, still assuming existence of an averaged central field **the electronic wave function would have the same angular parts. Thus, approximately** we can consider each electron as being assigned to different orbitals, **and also we have to remember** two possible spin states.
Atomic Orbitals
Shells: $n = 1\,(K)\ 2\,(L)\ 3\,(M)\ 4\,(N)\ \ldots$
Sub–shells: $l = 0\,(s)\ 1\,(p)\ 2\,(d)\ 3\,(f)\ \ldots$

Build-up (Aufbau) principle:
$1s\ 2s\ 2p\ 3s\ 4s\ 3d\ 4p\ 5s\ 4d\ 5p\ 6s\ \ldots$
Because of e repulsion: e occupy different orbitals of a sub–shell before doubly occupying any one of them.
Spin correlations: Hund's rule: an atom in the ground state adopts configuration with the greatest number of unpaired e.

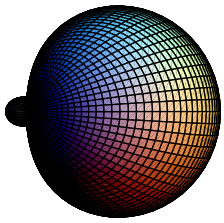
Electronic configurations:
${}_1H\ \ 1s^1\ \uparrow$
${}_2He\ 1s^2 \equiv [He]\ \uparrow\downarrow$
${}_3Li\ 1s^2\,2s^1 \equiv [He]\,2s^1\ \uparrow$
${}_4Be\ 1s^2\,2s^2 \equiv [He]\,2s^2\ \uparrow\downarrow$
${}_5B\ \ [He]\,2s^2\,2p^1\ \uparrow|\ |$

${}_6C\ \ [He]\,2s^2\,2p^2\ \uparrow|\uparrow|$
${}_7N\ \ [He]\,2s^2\,2p^3\ \uparrow|\uparrow|\uparrow$
${}_8O\ \ [He]\,2s^2\,2p^4\ \uparrow\downarrow|\uparrow|\uparrow$
${}_9F\ \ [He]\,2s^2\,2p^5\ \uparrow\downarrow|\uparrow\downarrow|\uparrow$
${}_8Ne\ [He]\,2s^2\,2p^6\ \uparrow\downarrow|\uparrow\downarrow|\uparrow\downarrow$
Note that there are some irregularities for very close levels e.g. for $Z = 24$ chromium, $Z = 29$ copper etc.
Hybrid orbitals **obtained as linear combinations of different lm orbitals extend further in some directions. They form a new ortho–normal basis.**
sp hybrid orbitals $(sp_z)_\pm = (1/\sqrt{2})(s \pm p_z)$. These are connected by a rotation on $\pi = 180^o$.

sp plus

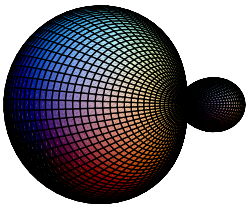


sp minus

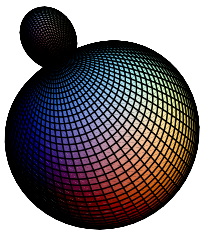


sp^2 hybrid orbitals $(sp_x p_x) = (1/\sqrt{3})(s + \sqrt{2}p_x)$, $(sp_x p_y)_\pm = (1/\sqrt{3})(s - (1/\sqrt{2})p_x \pm \sqrt{3}p_y)$. Rotations on $2\pi/3 = 120^o$.

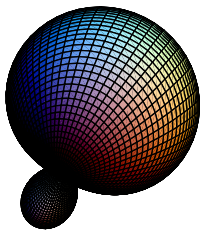
sp2 plus



sp2 minus minus

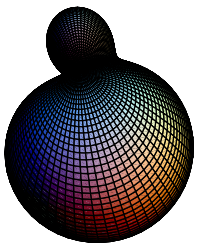


sp2 minus plus

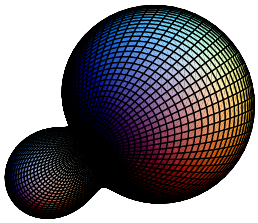


sp^3 hybrid orbitals $(sp^3 3)_{\pm\pm\pm} = (1/2)(s \pm p_x \pm p_y \pm p_z)$. Tetrahedron (center/corners angle is $109^o\,28'$).

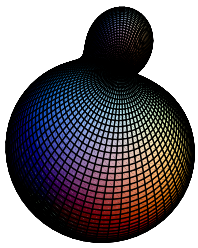
sp3 --+



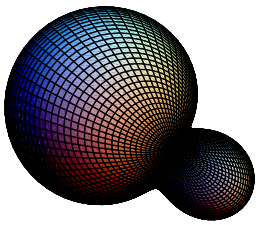
sp3 -+-



sp3 +--



sp3 +++



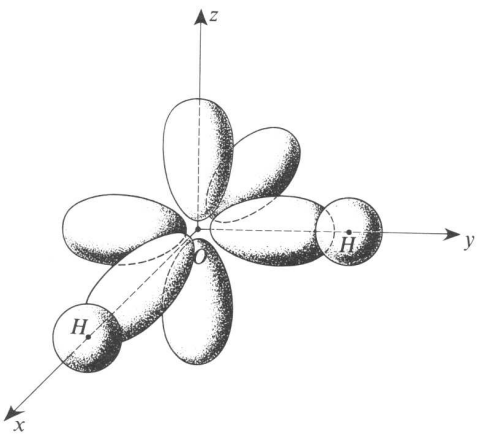


FIGURE 5

Schematic structure of the water molecule H_2O. The $2p_x$ and $2p_y$ orbitals yields bonds making an angle of approximately 90° (the real angle is 104° because of the electrostatic repulsion between the two protons).

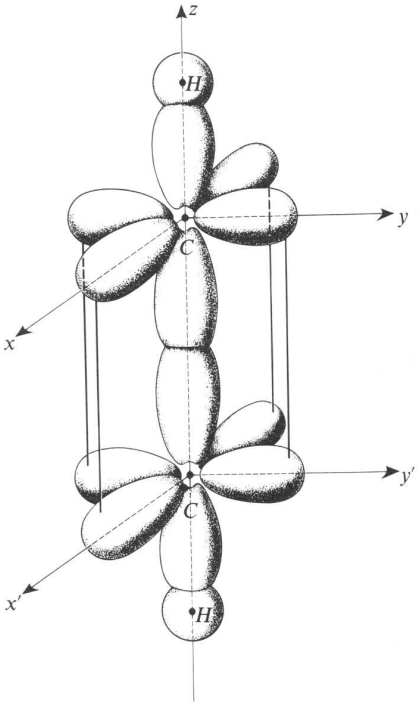


FIGURE 7

Schematic structure of the acetylene molecule C_2H_2. For each carbon atom, two electrons are in the sp_z hybrid orbitals (*cf.* fig. 6) and contribute to the C – H and C – C bonds (σ bonds). In addition, two electrons are in the p_x and p_y orbitals and form additional bonds between the two carbon atoms (π bonds, weaker than σ bonds), shown by the vertical lines in the figure. The C – C bond is therefore a "triple bond".

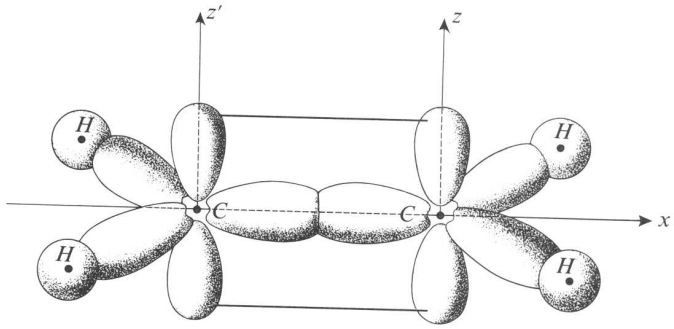


FIGURE 9

Schematic structure of the ethylene molecule C_2H_4. The two carbon atoms form a double bond with each other : one σ bond due to sp^2 orbitals of the type of those shown in figure 8 (the other two sp^2 hybrid orbitals at 120° with this one form the C – H bonds), and one π bond, due to the overlapping of the p_z orbitals.

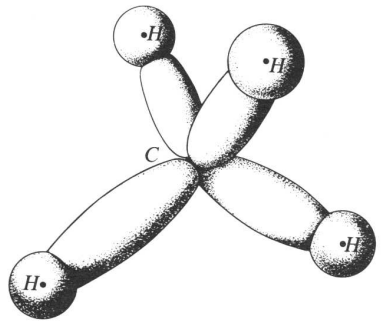


FIGURE 10

Schematic structure of the methane molecule. The sp^3 orbitals produce bonds arranged like the straight lines joining the center of a tetrahedron to its four corners (angles of 109°28′).

LCAO (linear combination of atomic orbitals) MO (molecular orbitals). **See P.W. Atkins textbook for more details.**

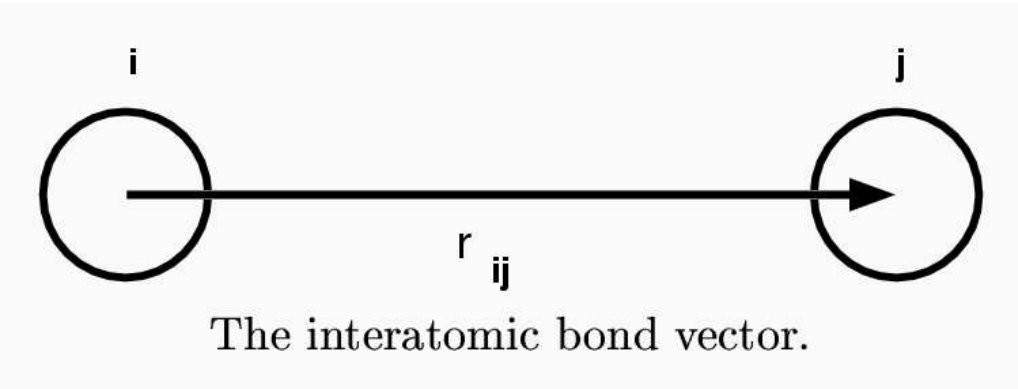
Potential energy surfaces and Force Fields

Having solved QM Eq. and obtained PES $V_e(\vec{r}_n)$ **we can further apply the** classical treatment **for the motion of nuclei based on the** Newton's equations:

$$m_n \frac{d^2\vec{r}_n}{dt^2} = \vec{F}_n, \quad \vec{p}_n = m_n \frac{d\vec{r}_n}{dt}, \quad \vec{F}_n = -\frac{\partial V_e}{\partial \vec{r}_n}. \tag{3}$$

Bonded interactions

Bond stretching energies



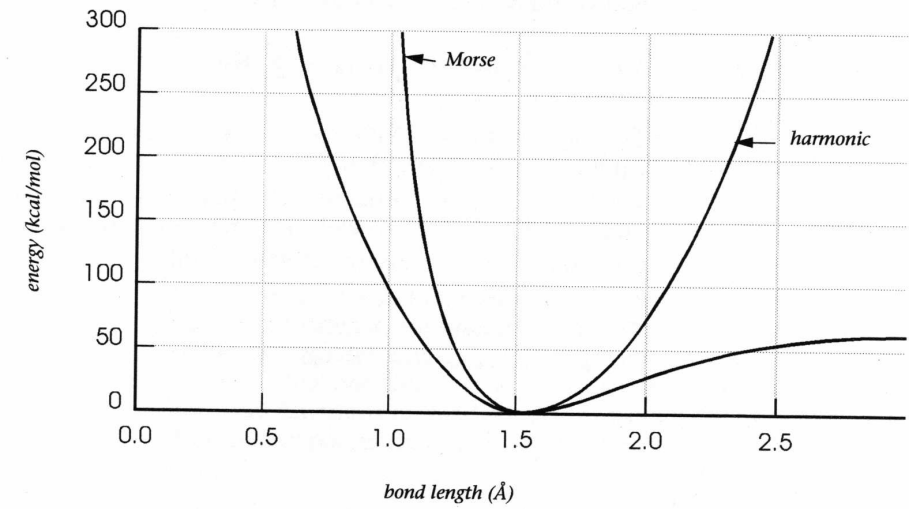
$$r_{ij} = |\vec{r}_i - \vec{r}_j|. \tag{4}$$

Harmonic bond potential

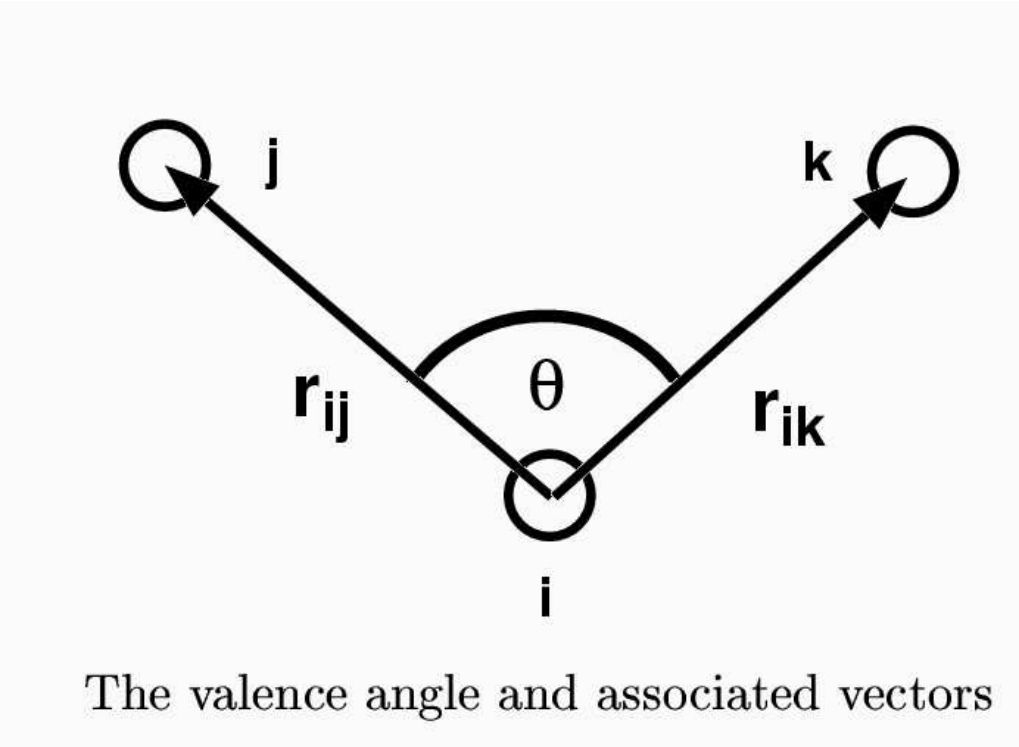
$$V(r_{ij}) = \frac{k_r}{2}(r_{ij} - r_0)^2. \tag{5}$$

Morse potential

$$V(r_{ij}) = E_0\Big(\Big(1 - \exp\left[-k_r(r_{ij} - r_0)\right]\Big)^2 - 1\Big). \tag{6}$$



Angular energies **describe rotation between two bonds**



$$\theta_{jik} = \arccos\left(\frac{(\vec{r}_{ij} \bullet \vec{r}_{ik})}{r_{ij}\, r_{ik}}\right). \tag{7}$$

Harmonic angle potential

$$V(\theta_{jik}) = \frac{k_\theta}{2}(\theta_{jik} - \theta_0)^2. \tag{8}$$

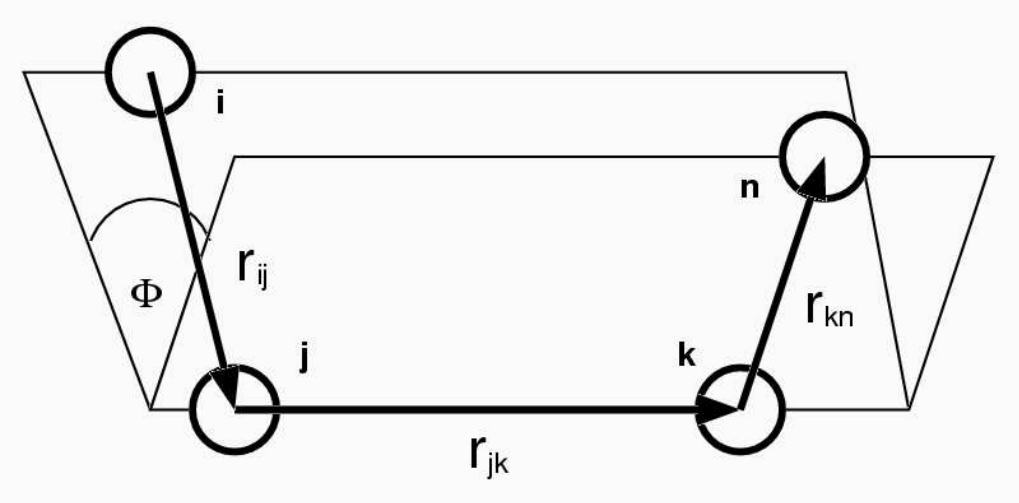
Harmonic cosine angle potential

$$V(\theta_{jik}) = \frac{k_\theta}{2}(\cos(\theta_{jik}) - \cos(\theta_0))^2. \tag{9}$$

Cosine angle potential

$$V(\theta_{jik}) = \sum_m A_m \Big(1 + \cos(m\theta_{ijk} - \delta_m)\Big). \tag{10}$$

Torsional energies **describe rotation between two planes built from three consequent bonds**



$$\varphi_{ijkn} = \arccos\left(\frac{(\vec{r}_{ij} \times \vec{r}_{jk} \bullet \vec{r}_{jk} \times \vec{r}_{kn})}{|\vec{r}_{ij} \times \vec{r}_{jk}|\,|\vec{r}_{jk} \times \vec{r}_{kn}|}\right). \tag{11}$$

Harmonic angle potential

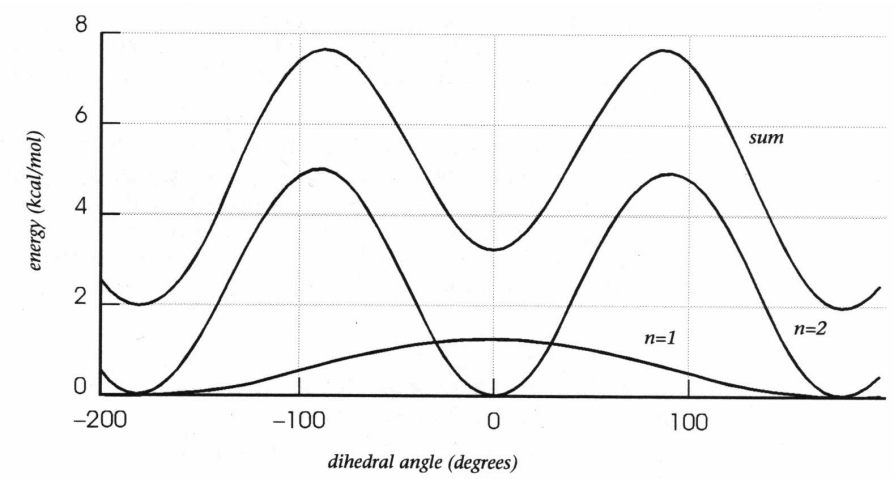
$$V(\varphi_{ijkn}) = \frac{k_\varphi}{2}(\varphi_{ijkn} - \varphi_0)^2. \tag{12}$$

Harmonic cosine angle potential

$$V(\varphi_{ijkn}) = \frac{k_\varphi}{2}(\cos(\varphi_{ijkn}) - \cos(\varphi_0))^2. \tag{13}$$

Cosine angle potential

$$V(\varphi_{ijkn}) = \sum_m B_m \Big(1 + \cos(m\varphi_{ijk} - \Delta_m)\Big). \tag{14}$$



Non–bonded interactions

Van der Waals forces (dispersion and repulsion)

$\mathbf{12-6}$ potential

$$V(r_{ij}) = \frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}}. \tag{15}$$

Lennard–Jones potential

$$V(r_{ij}) = 4\epsilon \left(\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} \right). \tag{16}$$

Buckingham potential

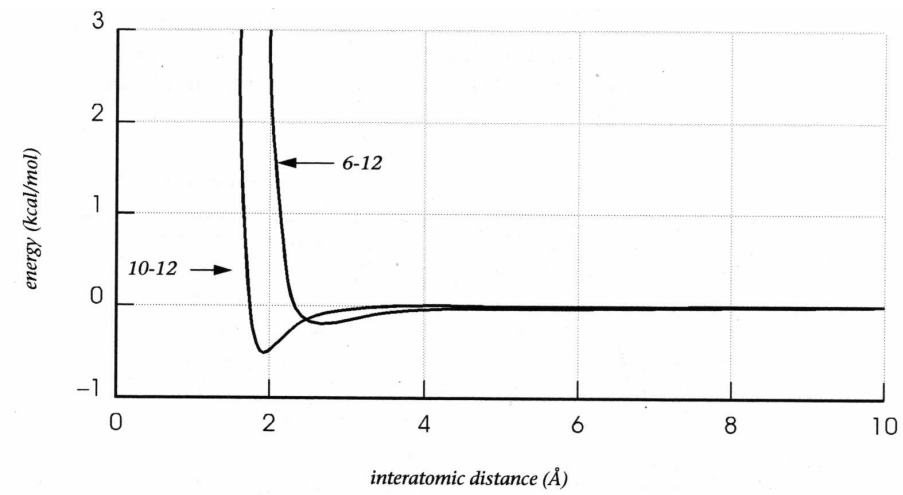
$$V(r_{ij}) = C \exp\left(-\frac{r_{ij}}{\rho}\right) - \frac{D}{r_{ij}^{6}}. \tag{17}$$

Coulomb potential **is rather long–ranged**

$$V(r_{ij}) = \frac{1}{4\pi\epsilon_0} \frac{q_i\, q_j}{r_{ij}}. \tag{18}$$

H-bonding specific potential

$$V(r_{ij}) = \frac{A'}{r_{ij}^{12}} - \frac{B'}{r_{ij}^{10}}. \tag{19}$$



Typical scales and values

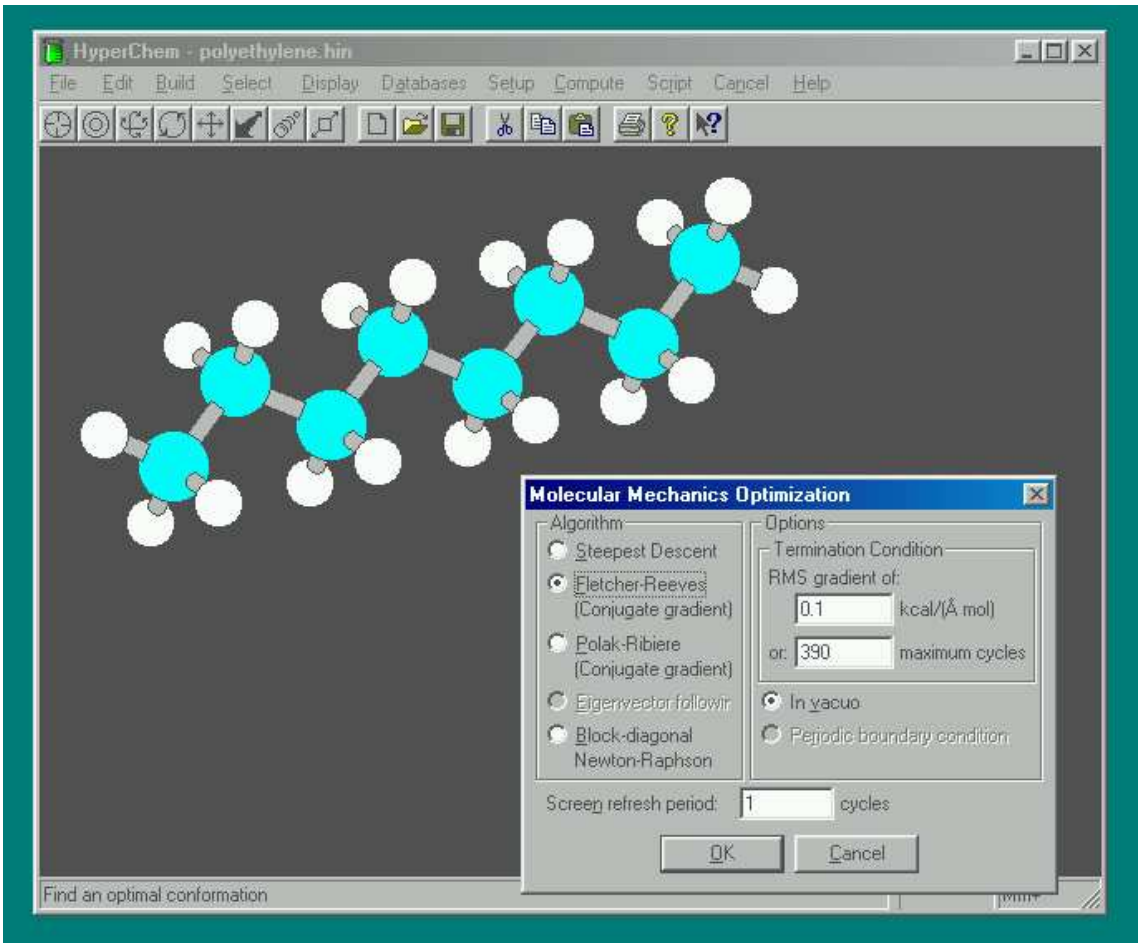
Time	$t_0 = 10^{-15}$ **s = 1 femtoseconds**
Length	$l_0 = 10^{-10}$ **m = 1 Å = 0.1 nanometers**
Mass	$m_0 \simeq 1.66 \cdot 10^{-27}$ **kg = 1 a.m.u.**
Charge	$q_0 \simeq 1.6 \cdot 10^{-19}$ **Cl (charge of proton)**

Lennard–Jones **for He** $\epsilon \simeq 85$ **J/mol** $\sigma \simeq 2.6$**Å**

Geometry optimisation at 0 K temperature: Steepest Descent or Newton–Raphson **method**

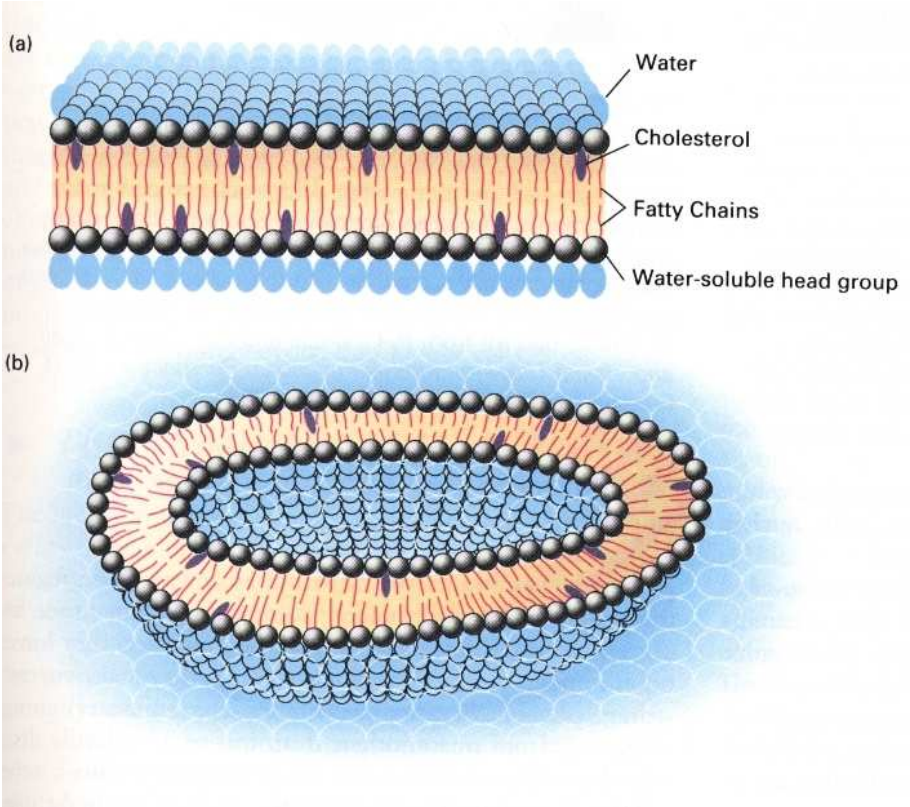
$$g_i^a = -\frac{\partial V}{\partial X_i^a}, \quad X_{i+1}^a = X_i^a + \lambda_i \frac{g_i^a}{|g_i|}. \tag{20}$$

Equilibrium structure: Monte Carlo Metropolis **method for NVT ensemble**
Dynamics: Newtonian or Langevin **equations are integrated numerically with the use of thermostats/barostats**
Periodic boundary conditions: Ewald, Lekner and wavelet **summation techniques**
Water molecules and ions: Implicit treatment **methods and integral equation techniques for small molecules**

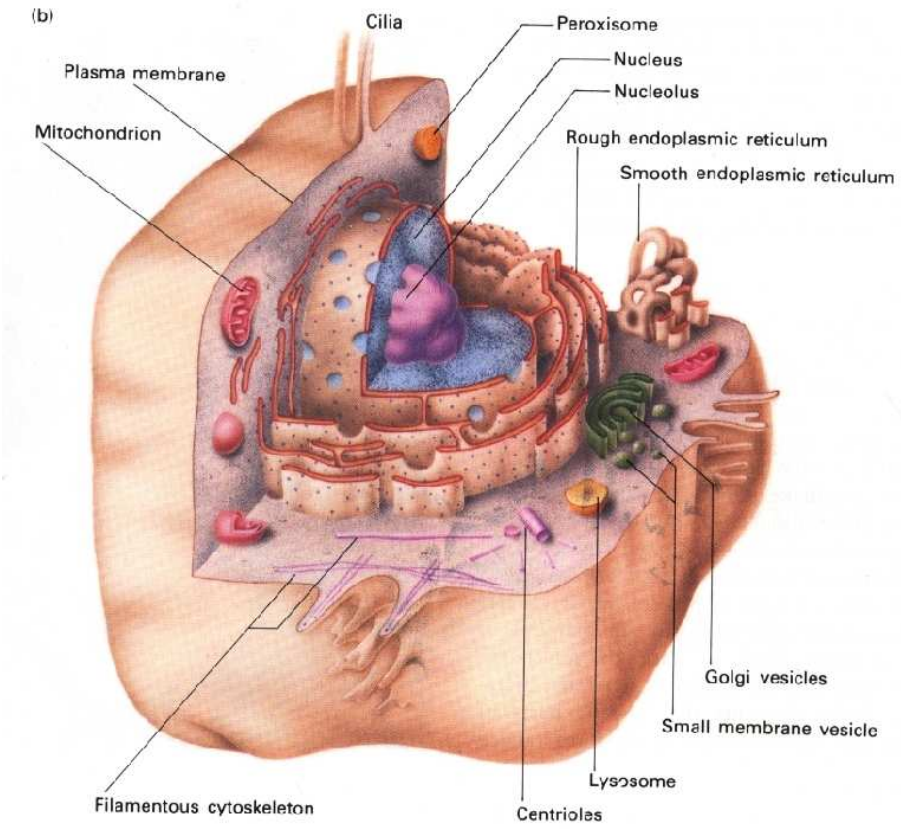


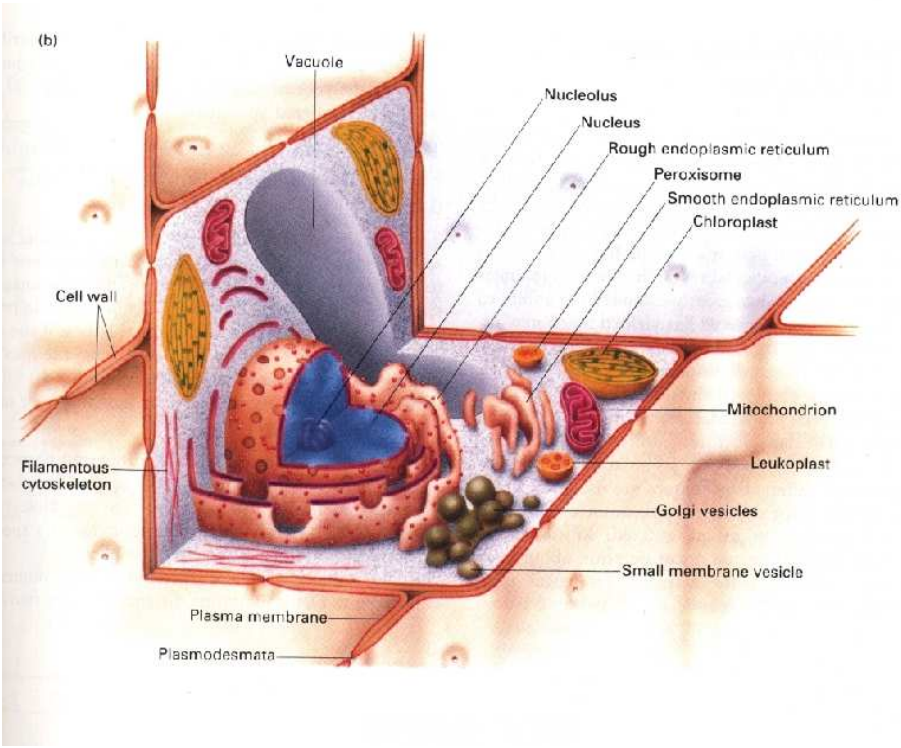
Cell organisation and biopolymers involved

Common processes in all cells: replication of DNA, synthesis of proteins and production of chemical energy by conversion of glucose to carbon dioxide.
Types of cells: eukaryotic **(with a nucleus): from ferns to flowering plants, from amoeba to mammals;** prokaryotic **(no nucleus):** eubacteria **(e.g. cyanobacteria) and** archaebacteria **(e.g. methanogens in swamps and thermoacidophiles in sulfur springs).**
Prokaryotes **have just the plasma membrane based on a** phospholipid bilayer, **through which O_2, CO_2 and H_2O can diffuse freely, but** channels **are used to transport other larger molecules.**



Eukaryotes **have also internal membranes separating the** organelles **from the rest of the** cytoplasm.
Nucleus **contains most of the cellular DNA and is the site of cellular RNA synthesis.**
Mitochondria **generates most of cellular ATP (ATP is adenosine triphosphate, ADP is diphosphate, and AMP is monophosphate) by oxidation of small molecules in aerobic cells. Mitochondria have their own** DNA **needed for synthesis of mitochondrial proteins.**
Rough and smooth ER (endoplasmic reticula) **are networks of membranes where glycoproteins, secreted proteins and lipids are synthesised.**
Golgi vesicles **process secretary and membrane proteins and direct membrane constituents to appropriate places.**
Lysosomes **in animals degrade proteins and** Chloroplasts **in plants are photosynthesis sites.**
Vacuoles **store nutrients and degrade some molecules.**
Cytoskeleton **is an array of fibrous proteins for stability.**





Macromolecular contents of cells: **70%** is water, **7% is other small molecules (salts, lipids, amino acids, nucleotides), 20%** is protein, **2%** is RNA (ribonucleic acid) **and less than 1%** is DNA (deoxyribonucleic acid), **the genetic material.**

Prokaryotes **have a single** chromosome **containing full length of DNA near the centre of the cell.**

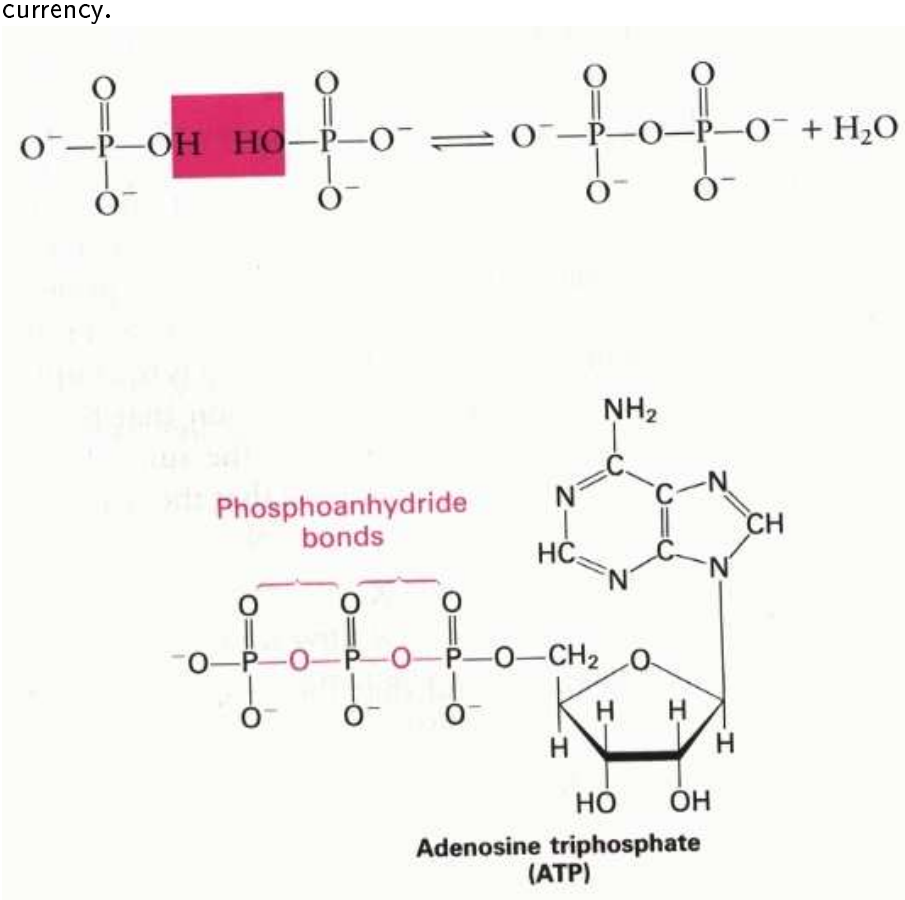
Eukaryotes **have** several chromosomes **each containing a different single double–stranded DNA molecule within the** nucleus.
The nucleus **contains the** nucleolus **(fibrous matrix), where most of RNA is synthesised, and DNA-protein complexes.**

Energetics of the cell

Processes in an NpT system occur spontaneously **if**

$$\Delta G = \Delta H - T\Delta S = \Delta E + p\Delta V - T\Delta S \leq 0.$$

In most of organisms the most important molecule for capturing and transferring free energy **is** ATP (adenosine triphosphate), **which is produced in huge numbers and could be viewed as a** biological currency.



An ATP molecule has two phosphoanhydride **bonds as is symbolically written $Ap\sim p\sim p$, breaking such bonds in** hydrolysis **produces a lot of free energy,**

$$\begin{aligned} Ap\sim p\sim p + H_2O &\rightarrow Ap\sim p + P_i + H^+ \\ Ap\sim p\sim p + H_2O &\rightarrow Ap + PP_i + H^+ \\ Ap\sim p + H_2O &\rightarrow Ap + P_i + H^+ \end{aligned}$$

ATP is used to fuel many cellular processes which otherwise would not take place energetically **(e.g. protein and polysacharide synthesis, transport of molecules into cells, generation of electric potential, muscle contraction etc).**
Photosynthesis **uses ATP to form** glucose **and sugars**

$$\begin{aligned} ATT &\rightarrow ADP + P_i \\ 6CO_2 + 6H_2O &\rightarrow C_6H_{12}O_6 + 6O_2 \end{aligned}$$

Aerobic catabolism **of** glucose **produces lots of ATP**

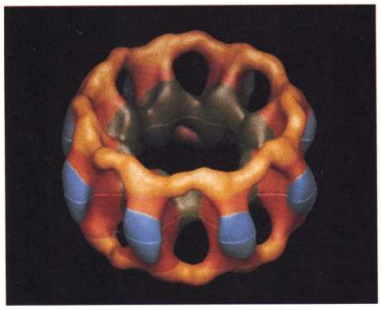
$$C_6H_{12}O_6 + 6O_2 + 36P_i + 36ATP \rightarrow 6CO_2 + 6H_2O + 36ATP$$

Lecture 2

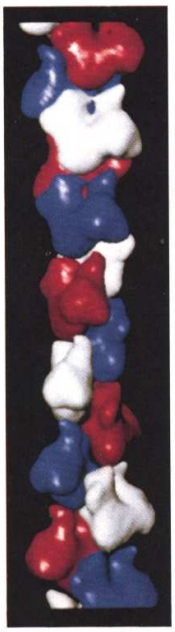
Protein structure, conformation and function

Versatility: Proteins **play building blocks, structural, functional, catalytic and many other functions in cells (e.g. muscle, motors, ion channels, pumps).**
Examples: (a) nuclear pore; (b) actin filament, (c) reverse transcriptase enzyme for copying viral RNA into DNA; (d) DNA polymerase participates in DNA synthesis.

(a)

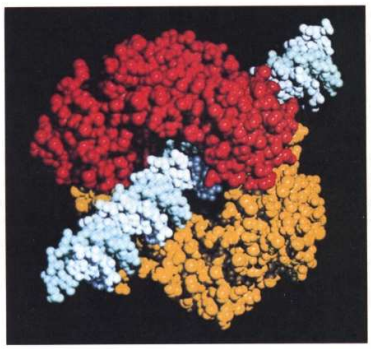


(b)



(c)

(d)

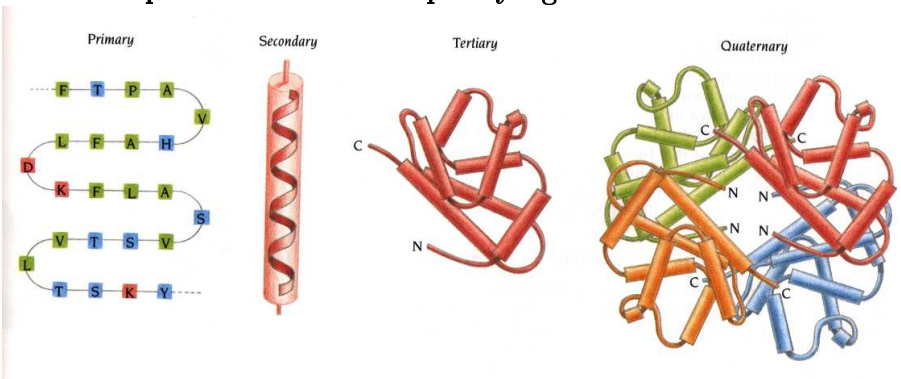


A FIGURE 3-1 Gallery of protein structural models. (a) The nuclear pore, a complex of proteins with a total molecular weight (MW) of 1.2×10^8, an outer diameter of 133 nm, a height of 70 nm, and a central hole 42 nm in diameter. Transport of molecules into and out of the nucleus occurs through such pores located in the nuclear membrane (Figure 5-33). (b) An actin filament, a polymer built by end-to-end association of identical actin subunits. Each subunit measures 5.5 nm on a side and 3.5 nm thick. This figure shows 14 subunits, alternately colored red, white, and blue. The cell cytoplasm is criss-crossed by actin filaments that are as long as 50–100 μm and have various functions in the cell (Chapter 22). (c) Reverse transcriptase, an enzyme present in RNA viruses that copies the viral RNA genome into DNA. Reverse transcriptase from the AIDS-causing virus HIV measures $11 \times 3.0 \times 4.5$ nm. RNA to be copied lies in a groove on the surface of the enzyme. (d) Two β subunits of DNA polymerase (red and yellow) from *E. coli* clamped around a DNA molecule (light blue). The ring has an outer diameter of 8.0 nm and an inner diameter of 3.5 nm. DNA polymerase is required for DNA synthesis; this association of the β subunits and DNA assures that polymerase remains complexed with DNA during DNA replication (Chapter 10). [Part (a) from J. E. Hinshaw, B. O. Carragher, and R. A. Milligan, 1992, *Cell* **69:**1133; part (b) from M. F. Schmid et al., 1994, *J. Cell Biol.* **124:**341; part (c) from T. A. Steitz et al., 1992, *Science* **256:**1783; part (d) from Kuriyan et al., 1992, *Cell* **69:**425]

▲ FIGURE 3-8 Various graphical representations of the structure of Ras, a guanine nucleotide–binding protein. GDP is shown as a white stick figure in parts a–c and as a red stick figure in part d. (a) The C_α trace of Ras highlighting the course of the backbone. Evident from this view is how the polypeptide is packed into the smallest possible volume. (b) Ball-and-stick model of Ras showing the location of every carbon (green), nitrogen (blue), oxygen (orange), and sulfur (yellow). (c) A schematic diagram of Ras showing how β strands (arrows) and α helices (cylinders) are organized in the protein. Note the turns and loops connecting pairs of helices and strands. (d) The water-accessible surface of Ras. Painted on the surface are regions of positive charge (blue) and negative charge (red). Here we see that the surface of a protein is not smooth but has lumps, bumps, and crevices. The molecular basis for specific binding interactions lies in the uneven distribution of charge over the surface of the protein. [Courtesy of Gareth White. Adapted from L. Tong, et al., 1991, J. Mol. Biol. **217**:503.]

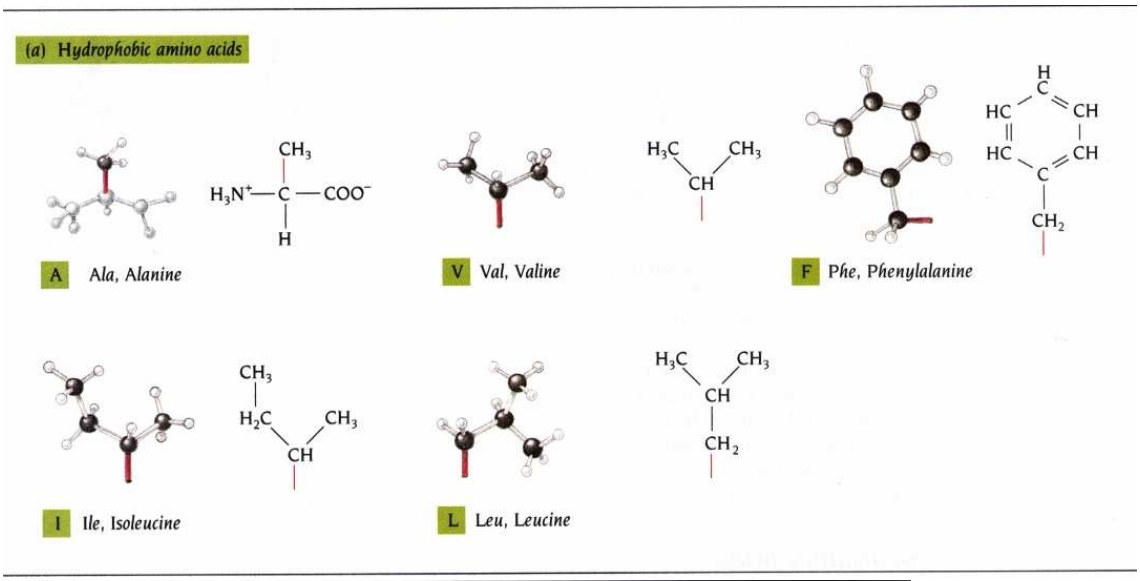
Four different levels of protein conformation.

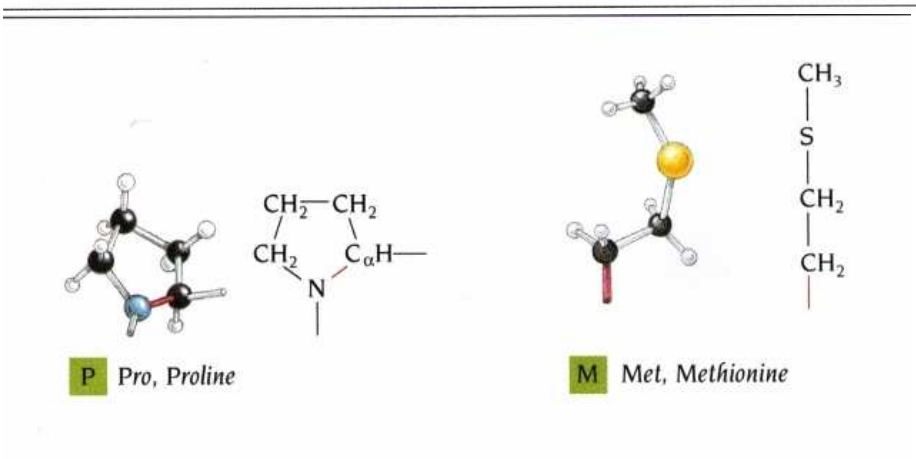
Conformation **specifies 3D coordinates of all atoms (nuclei) in a molecule, whereas** Configuration **refers to quantum numbers specifying states of electrons.**



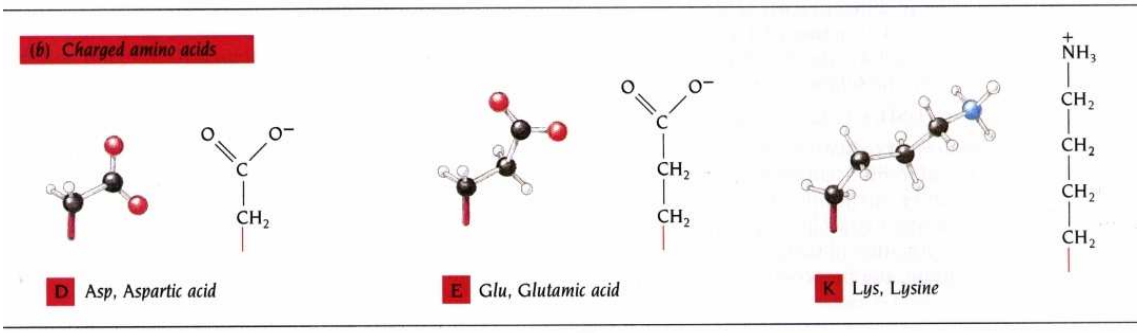
There are 20 **different naturally occurring** amino acid residues **the sequence of which specifies the** primary structure **of a protein.**

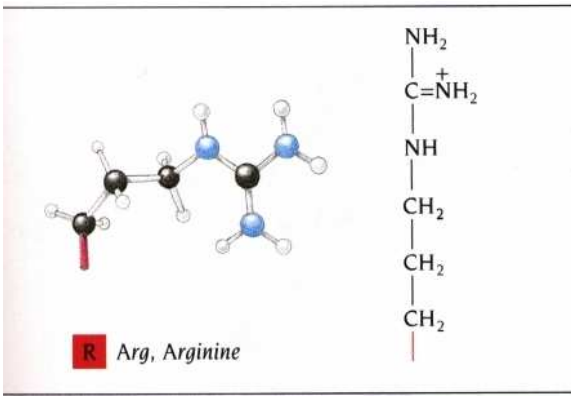
Hydrophobic residues:



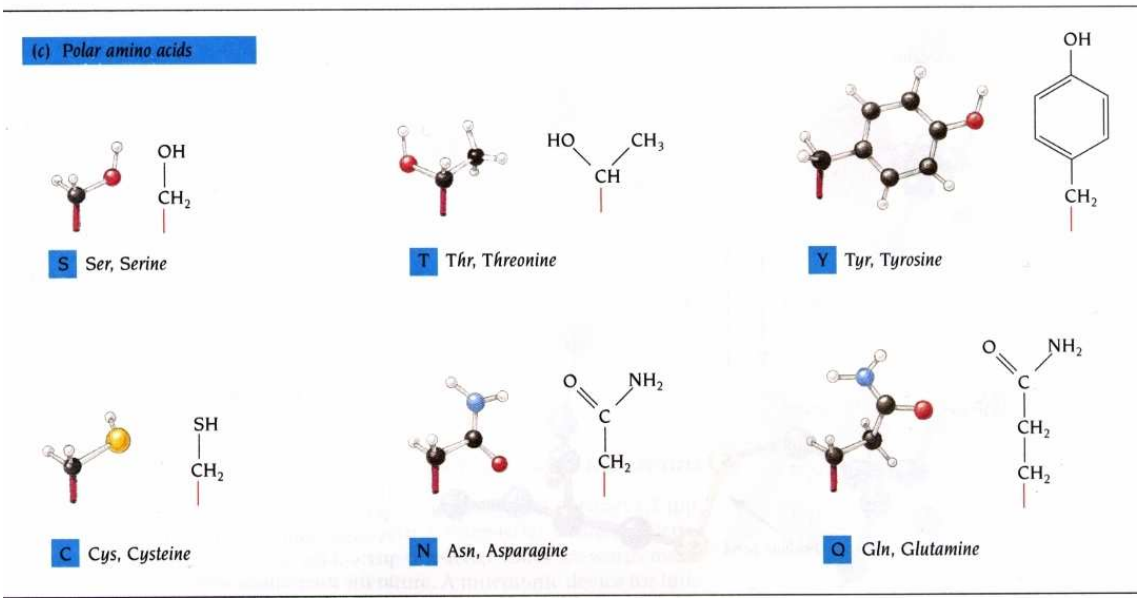


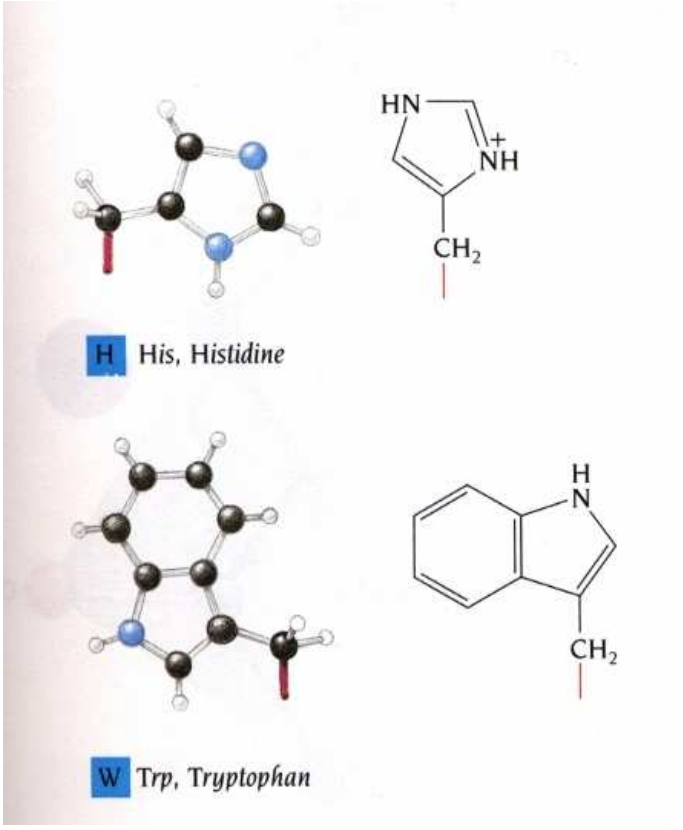
Charged residues:



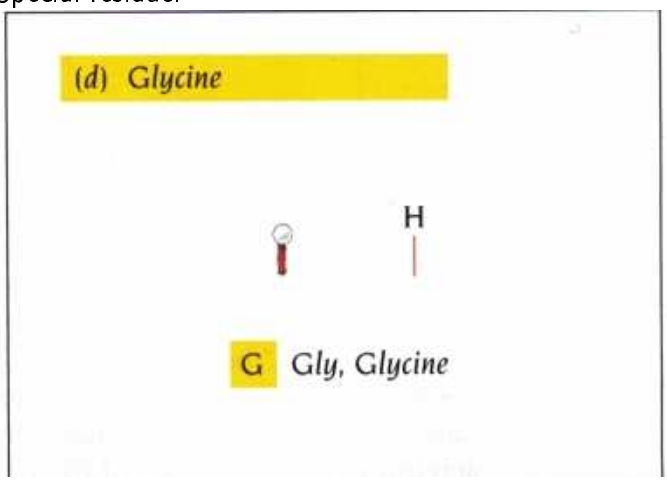


Polar residues:

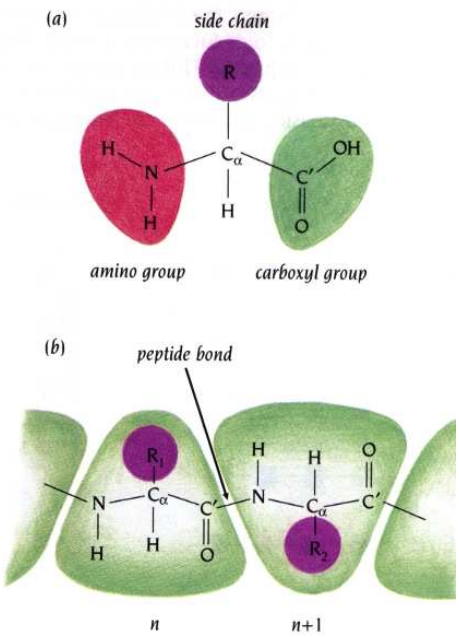




Special residue:



Basic residues: **Lys (K), Arg (R), His (H).**
Acidic residues: **Asp (D), Glu (E).**
Disulfide bridges: **Cys (C),** Backlooping: **Pro (P).**
Amino acids **polymerise with a release of a water molecule forming a** peptide bond **into a polypeptide chain (protein).**



Chirality **of R groups (except Glycine) means that the mirror–images of proteins are not super–imposable.** L form **occurs naturally.**

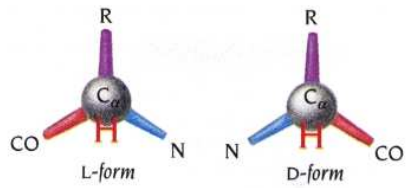


Figure 1.3 The "handedness" of amino acids. Looking down the H–C_α bond from the hydrogen atom, the L-form has CO, R, and N substituents from C_α going in a clockwise direction. There is a mnemonic to remember this; for the L-form the groups read CORN in clockwise direction.

Disulfide bridges **formed by oxidation in Cysteins help to stabilise the tertiary structure in many proteins.**

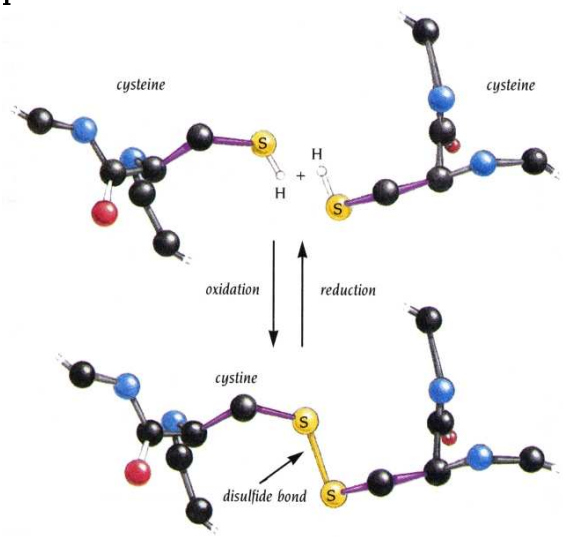


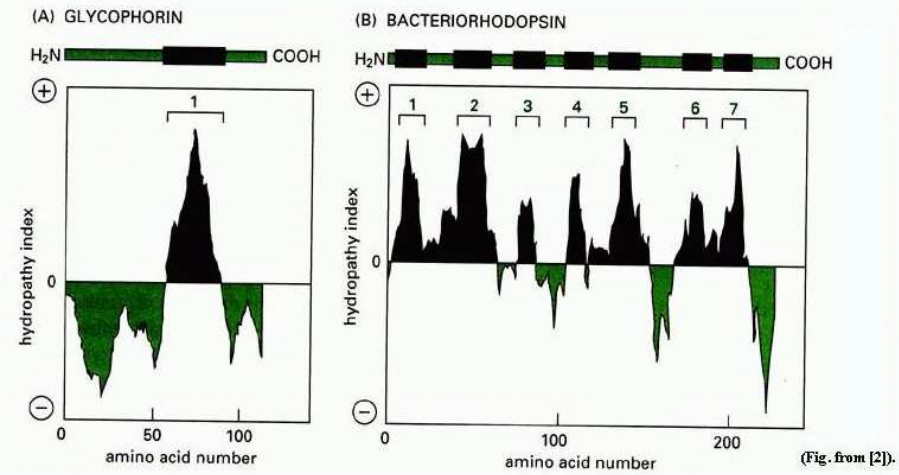
Figure 1.4 The disulfide is usually the end product of air oxidation according to the following schematic reaction scheme:

$$2\ -CH_2SH + \tfrac{1}{2}\ O_2 \rightleftharpoons -CH_2\text{–}S\text{–}S\text{–}CH_2 + H_2O$$

Disulfide bonds form between the side chains of two cysteine residues. Two SH groups from cysteine residues, which may be in different parts of the amino acid sequence but adjacent in the three-dimensional structure, are oxidized to form one S–S (disulfide) group.

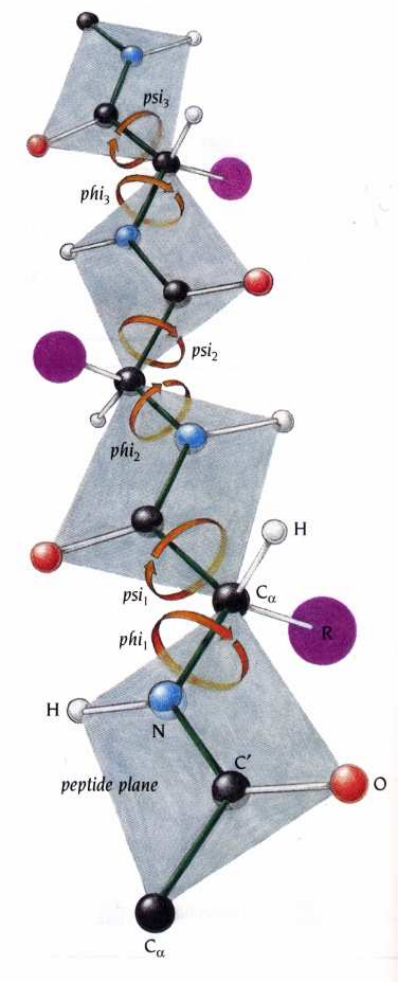
R groups can be broadly divided into hydrophobic **and** polar (hydrophilic), **as well as into** charged (+ **or** -) **and** uncharged **at physiological pH.** Glycine **is special and** Proline **loops back onto the backbone.**

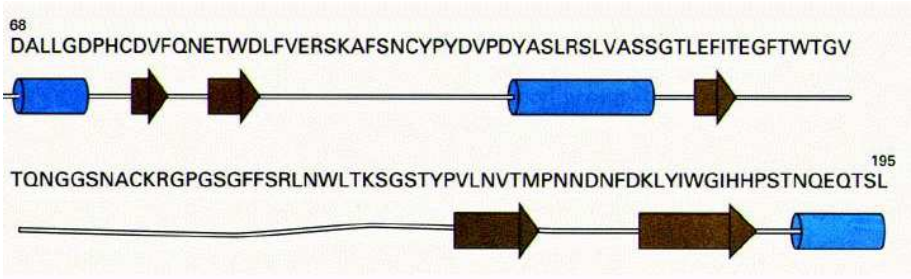
Hydropathy plots illustrate the random copolymer model of proteins.



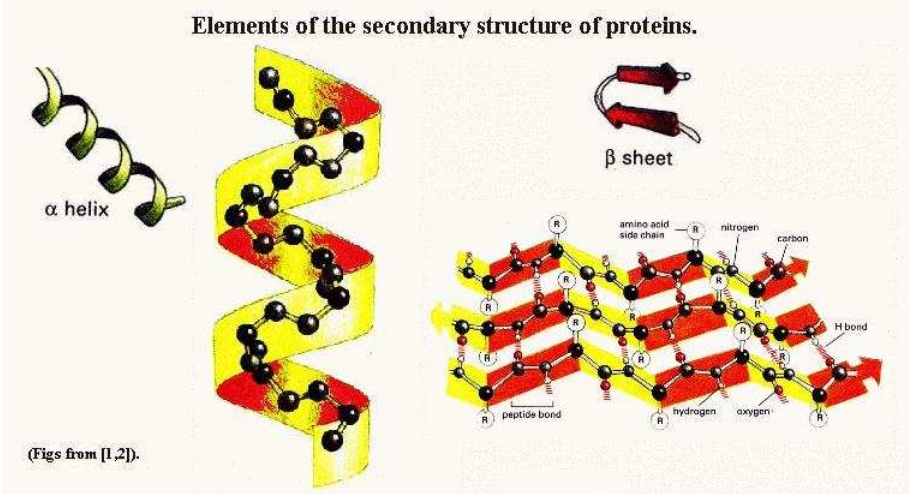
(Fig. from [2]).

Peptide units **are rigid and flat due to partial double bond character of a peptide bond (e delocalisation) with atoms positions fixed in the plane. There are** two torsional angles **describing degrees of freedom of flexibility between consecutive peptide units linked via C-alpha atoms.**

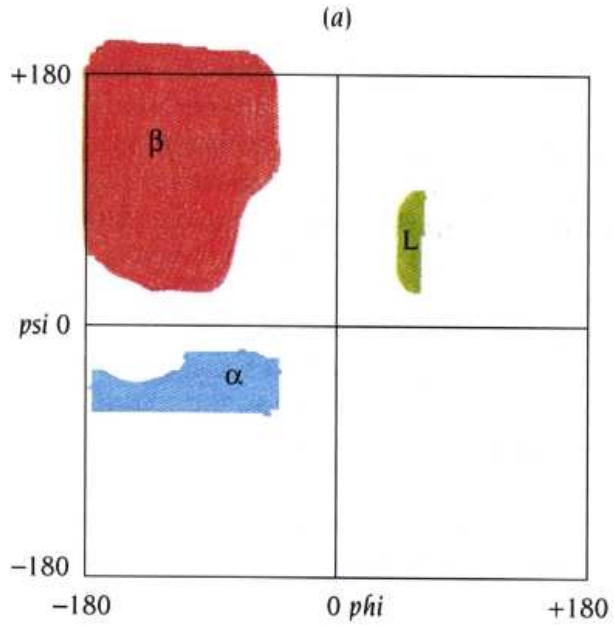


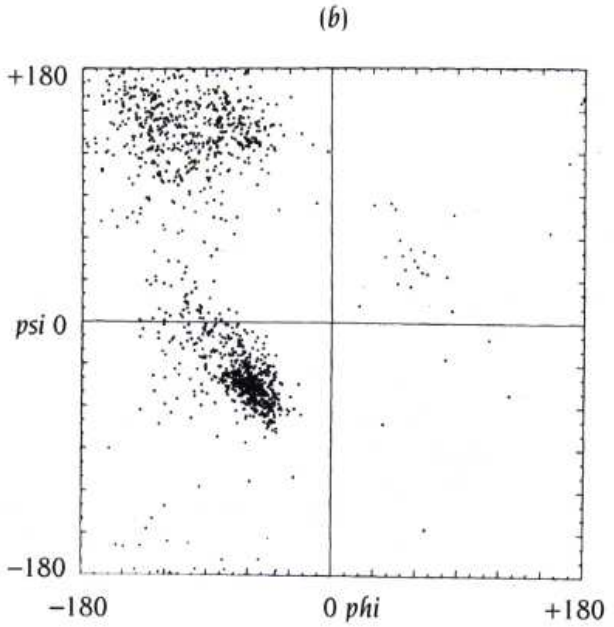


Primary structure of hemagglutinin. The region 68-195 is used by influenza virus to bind to animal cells. (Fig. from [1] and reprinted from A. Wilson et al, Nature 289, 366 (1981)).

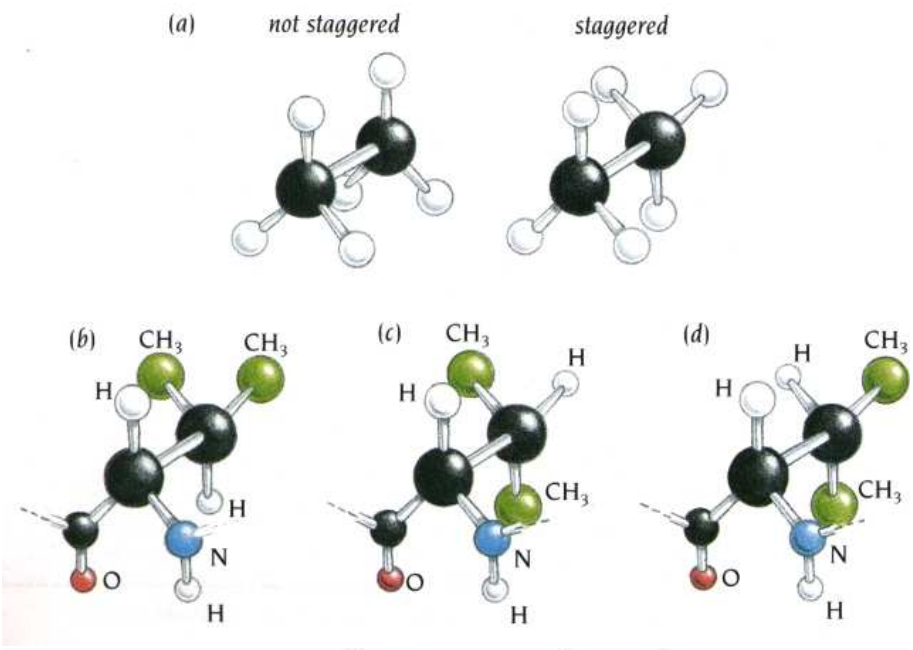


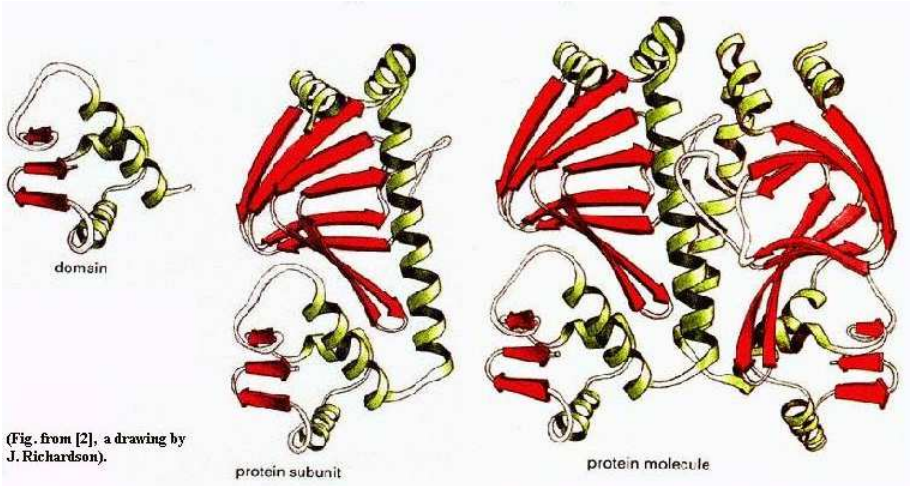
Ramachandran plot **of the two torsional angles allows one to easily characterise elements of the secondary structure.**





Certain side-chain conformations are more energetically favourable (to do with torsional potentials).

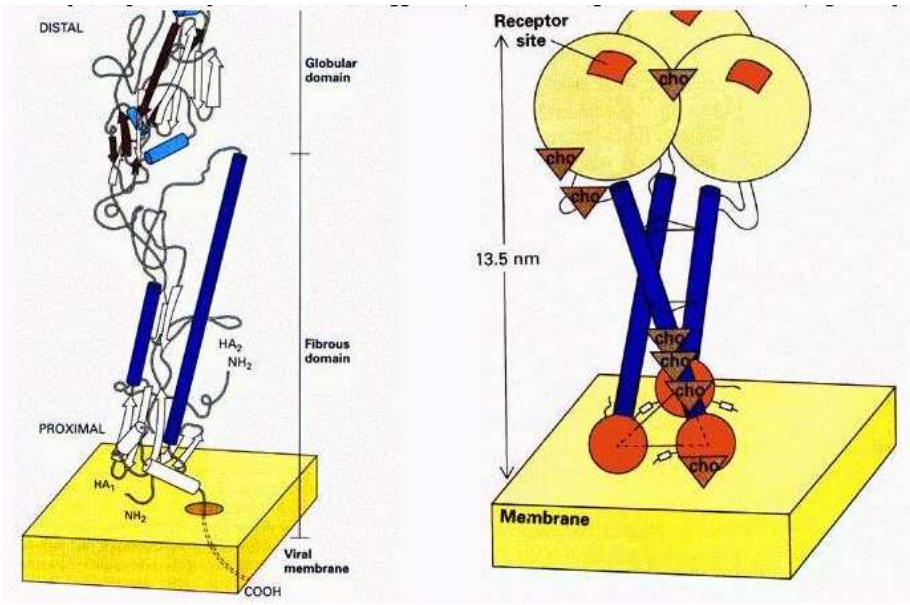




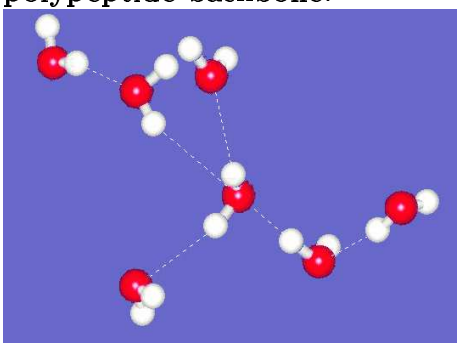
(Fig. from [2], a drawing by J. Richardson).

Catabolite activator protein (CAP).

Some proteins are multimers **of several distinct protein molecules forming a quaternary superstructure (e.g. hemagglutinin in picture and hemoglobin).**



Alpha helix **(right–handed) suggested by** Linus Pauling **is stabilised by** hydrogen bonding **between the polypeptide backbone.**



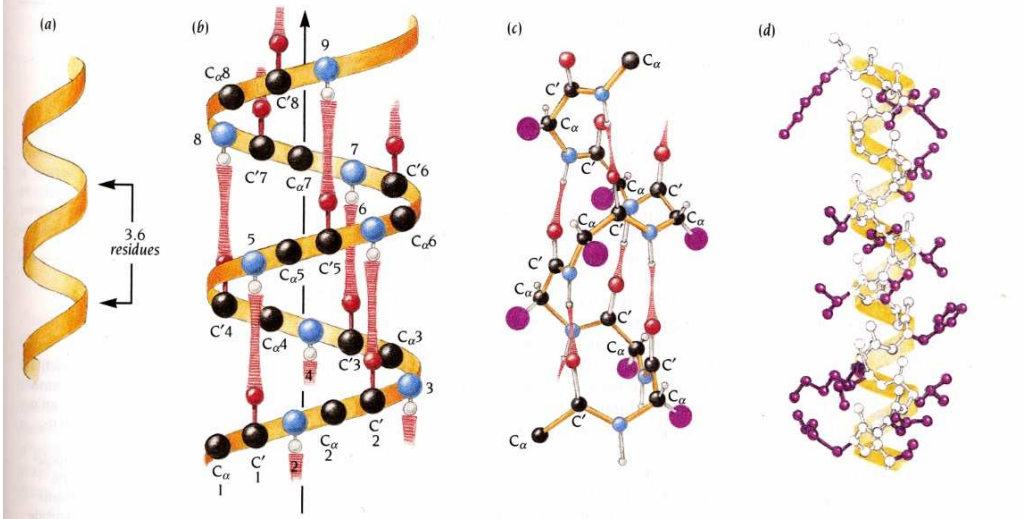
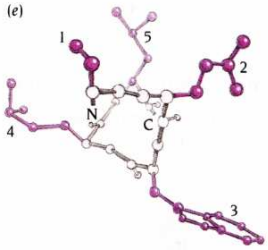
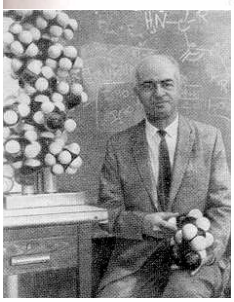
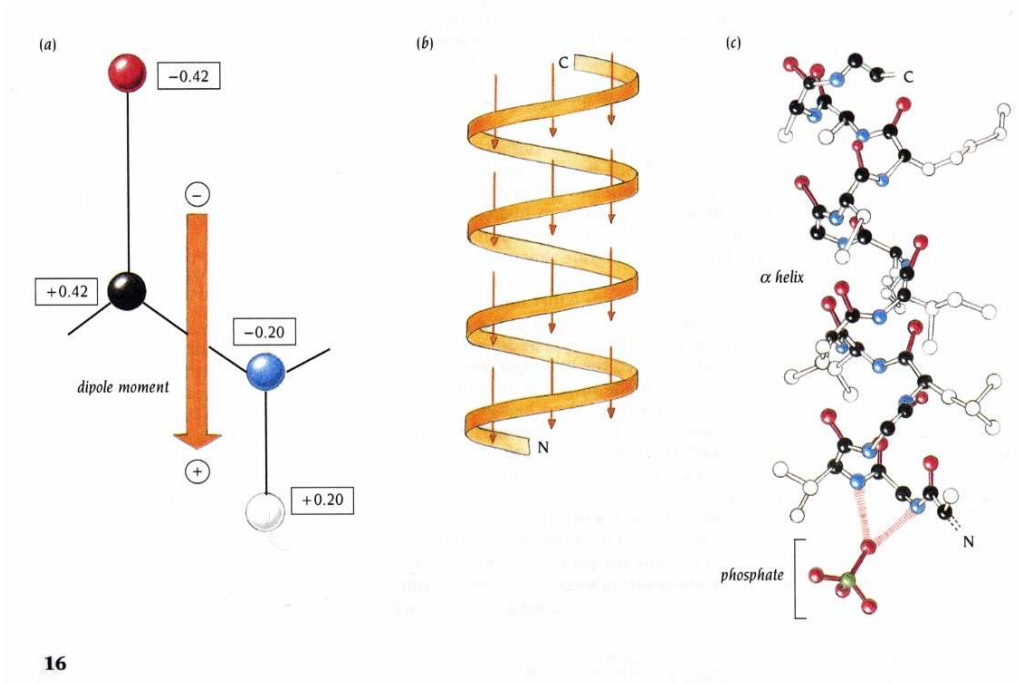


Figure 2.2 The α helix is one of the major elements of secondary structure in proteins. Main-chain N and O atoms are hydrogen-bonded to each other within α helices. (a) Idealized diagram of the path of the main chain in an α helix. Alpha helices are frequently illustrated in this way. There are 3.6 residues per turn in an α helix, which corresponds to 5.4 Å (1.5 Å per residue). (b) The same as (a) but with approximate positions for main-chain atoms and hydrogen bonds included. The arrow denotes the direction from the N-terminus to the C-terminus. (c) Schematic diagram of an α helix. Oxygen atoms are red, and N atoms are blue. Hydrogen bonds between O and N are red and striated. The side chains are represented as purple circles. (d) A ball-and-stick model of one α helix in myoglobin. The path of the main chain is outlined in yellow; side chains are purple. Main-chain atoms are not colored. (e) One turn of an α helix viewed down the helical axis. The purple side chains project out from the α helix.

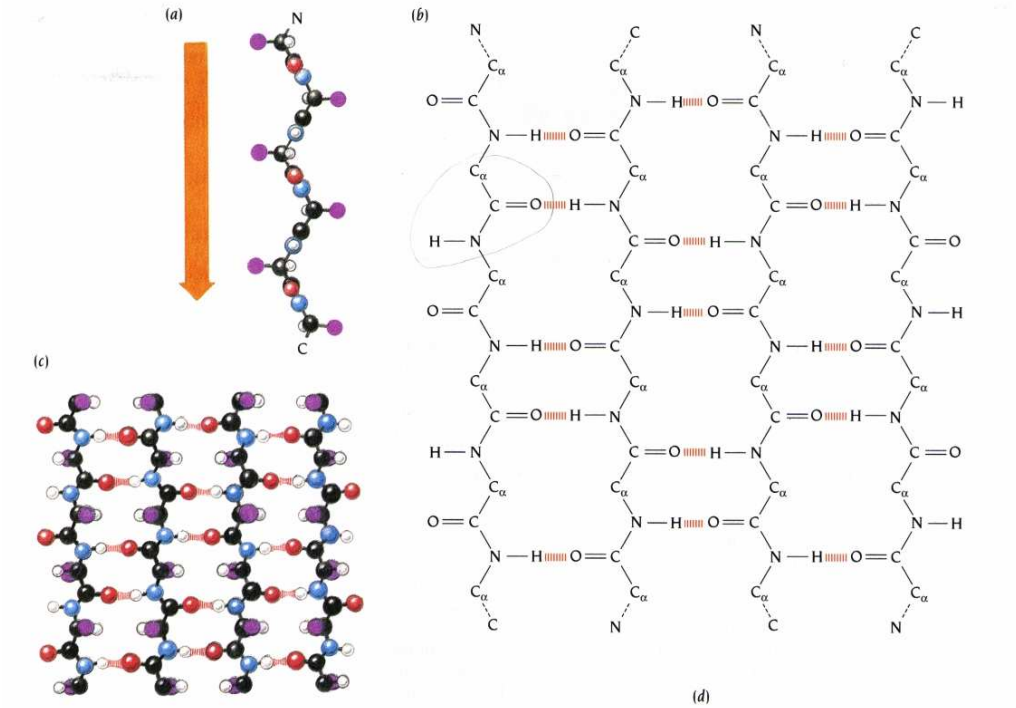




The alpha helix has a dipole moment. Some amino acids are preferred in alpha helix.



Antiparallel beta–sheet **is also stabilised by H-bonding. It may be** inter–molecular **or** intra–molecular **connected by loopy coil regions.**



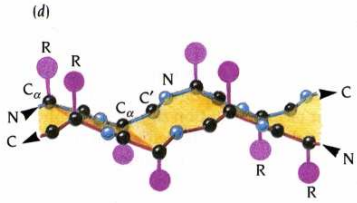
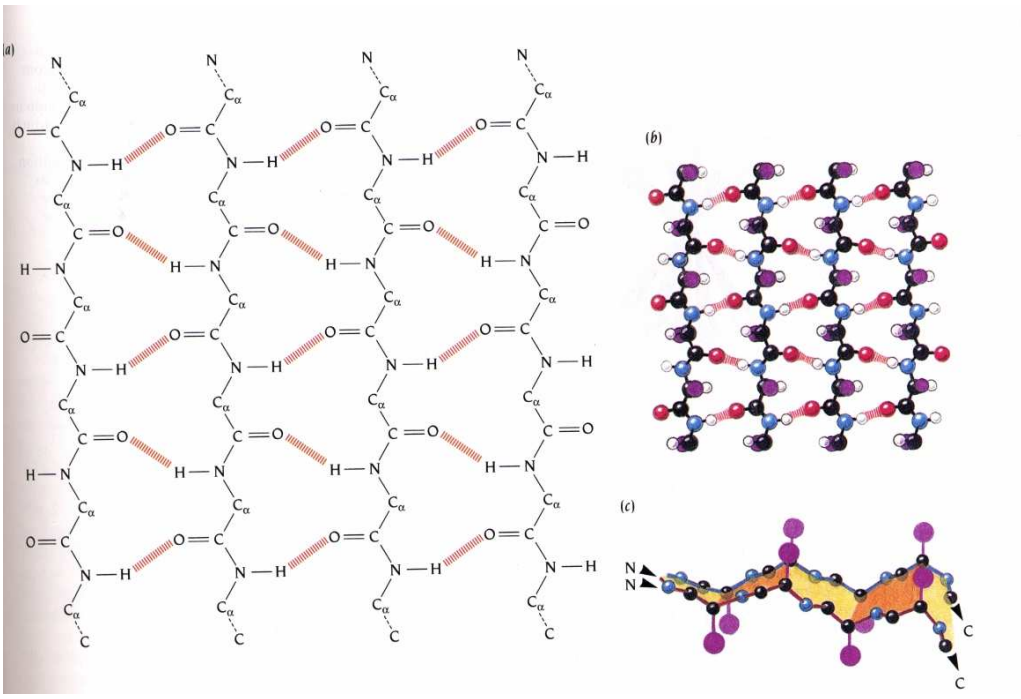


Figure 2.5 Schematic illustrations of antiparallel β sheets. Beta sheets are the second major element of secondary structure in proteins. The β strands are either all antiparallel as in this figure or all parallel or mixed as illustrated in following figures. (a) The extended conformation of a β strand. Side chains are shown as purple circles. The orientation of the β strand is at right angles to those of (b) and (c). A β strand is schematically illustrated as an arrow, from N to C terminus. (b) Schematic illustration of the hydrogen bond pattern in an antiparallel β sheet. Main-chain NH and O atoms within a β sheet are hydrogen bonded to each other. (c) A ball-and-stick version of (b). Oxygen atoms are red; nitrogen atoms are blue. The hydrogen atom in N–H···O is white. The carbon atom in the main chain, C_α, is black. Side chains are illustrated by one purple atom. The orientation of the β strands is different from that in (a). (d) Illustration of the pleat of a β sheet. Two antiparallel β strands are viewed from the side of the β sheet. Note that the directions of the side chains, R (purple), follow the pleat, which is emphasized in yellow.

There are also Parallel beta–sheets.



Beta-sheets can be twisted **(thioredoxin protein from E-coli).**

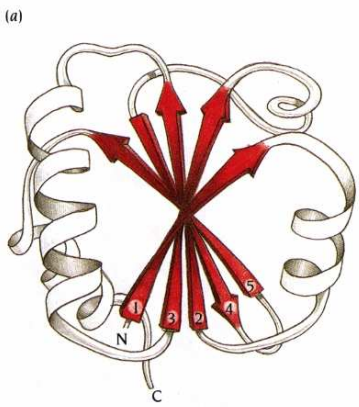
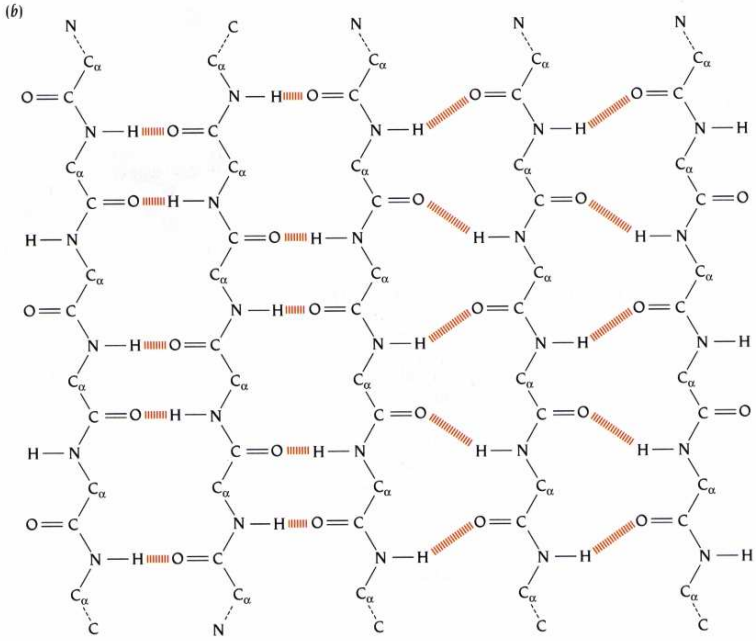


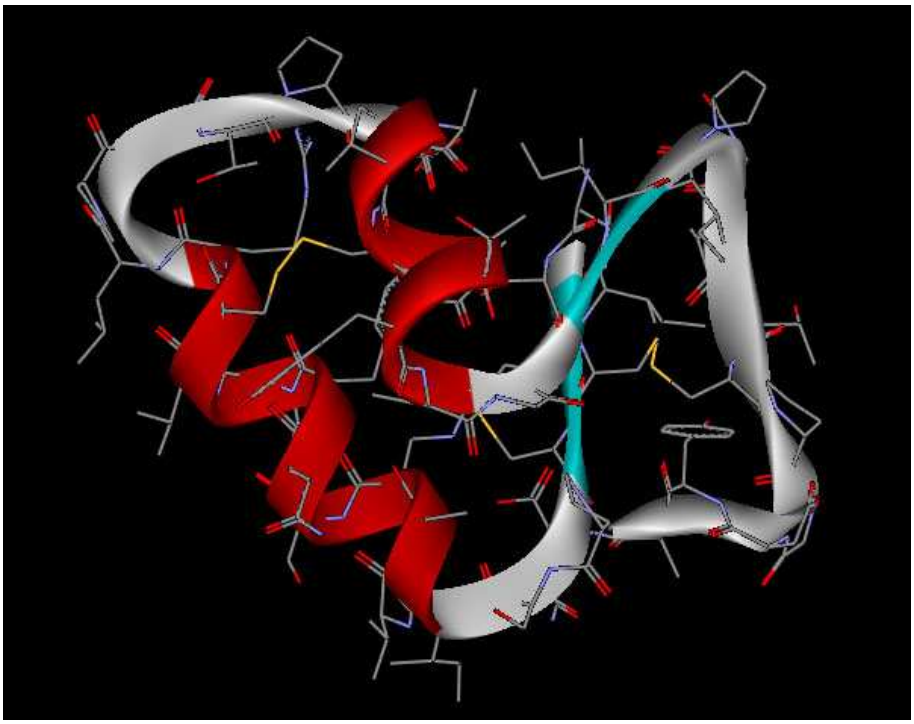
Figure 2.7 (a) Illustration of the twist of β sheets. Beta strands are drawn as arrows from the amino end to the carboxy end of the β strand in this schematic drawing of the protein thioredoxin from *E. coli*, the structure of which was determined in the laboratory of Carl Branden, Uppsala, Sweden, to 2.8 Å resolution. The mixed β sheet is viewed from one of its ends. (b) The hydrogen bonds between the β strands in the mixed β sheet of the same protein. [(a) Adapted from B. Furugren.]



Beta strands can also combine into mixed β sheets with some β strand pairs parallel and some antiparallel. There is a strong bias against mixed β sheets; only about 20% of the strands inside the β sheets of known protein structures have parallel bonding on one side and antiparallel bonding on the other. Figure 2.7 illustrates how the hydrogen bonds between the β strands are arranged in a mixed β sheet.

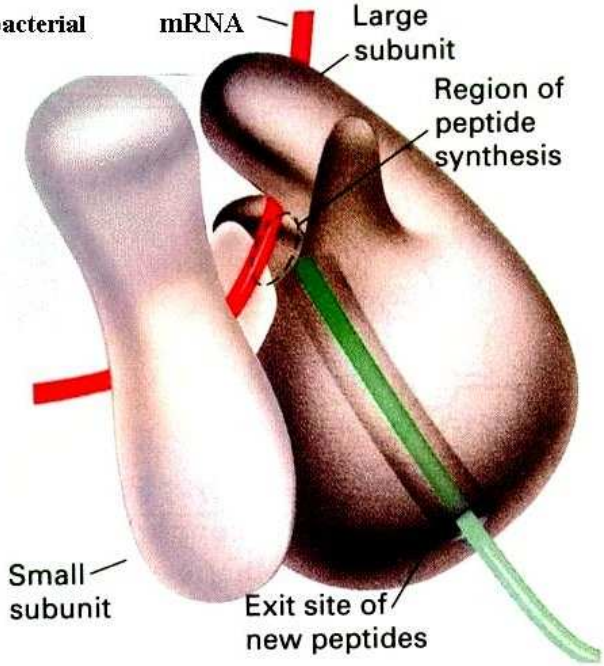
As they occur in known protein structures, almost all β sheets—parallel, antiparallel, and mixed—have twisted strands. This twist always has the same handedness as that shown in Figure 2.7, which is defined as a right-handed twist.

Cartoon **depiction of alpha-helices and beta-sheets along with wire-frame representation in crambin protein.**



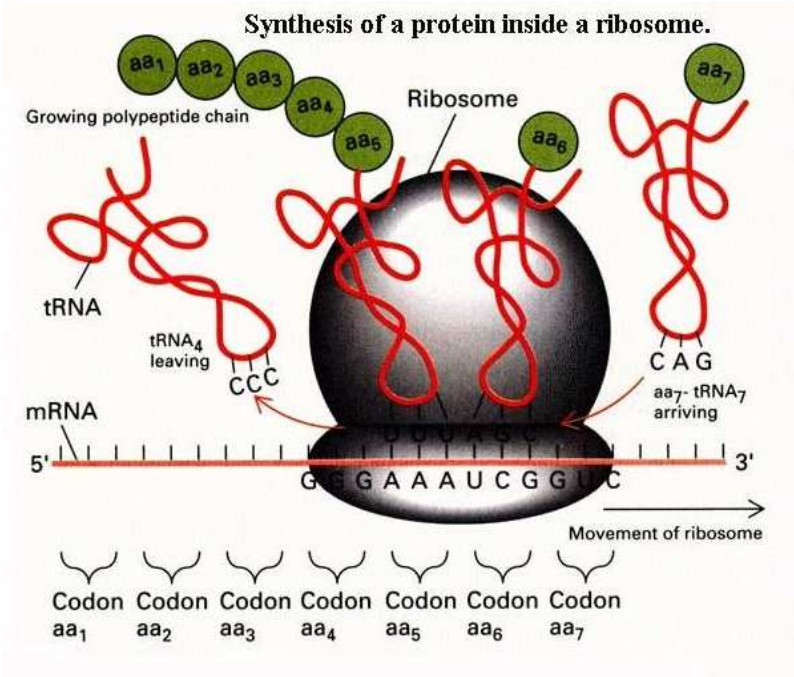
Proteins are synthesised in ribosomes from messenger–RNA **containing genetic information (from nuclear DNA).**

Overall structure of bacterial ribosomes.



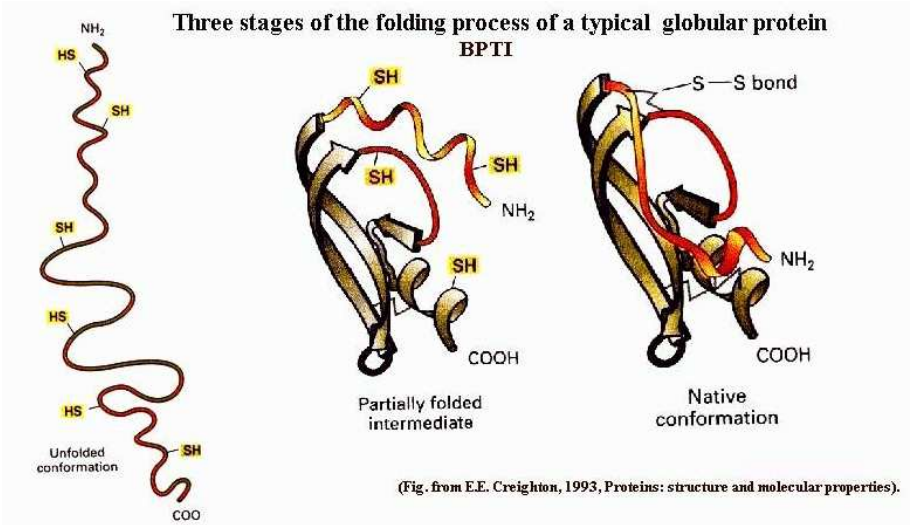
(Fig. from [1], drawing after J.A. Lake).

Nucleotide (4 letter) sequence of mRNA is translated **into (20 letter) R-sequence of a growing polypeptide.**

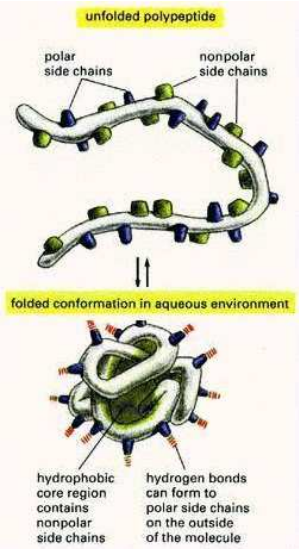


Newly synthesised proteins need to undergo folding **process to reach their** native (biologically active) **conformation.**

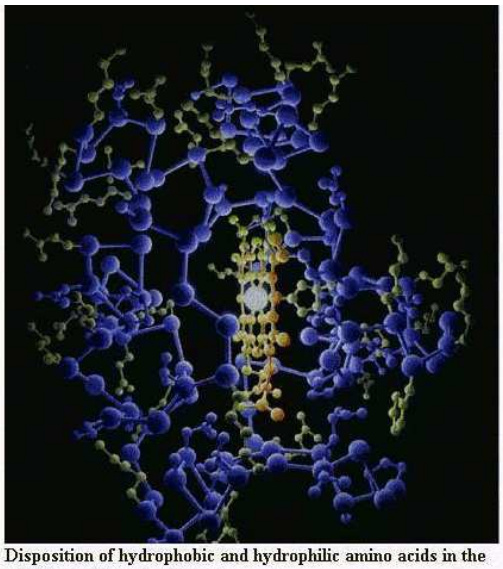
Many proteins fold into globular shape, but some are linear and there are also unstructured proteins (e.g. kinases).



Hydrophobicity **is one of the driving forces in folding in aqueous environment.**

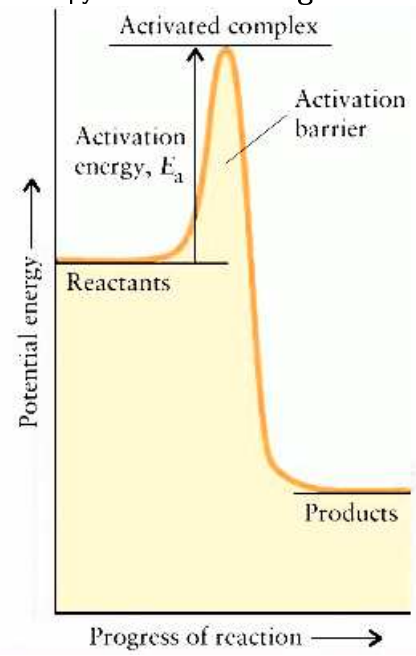


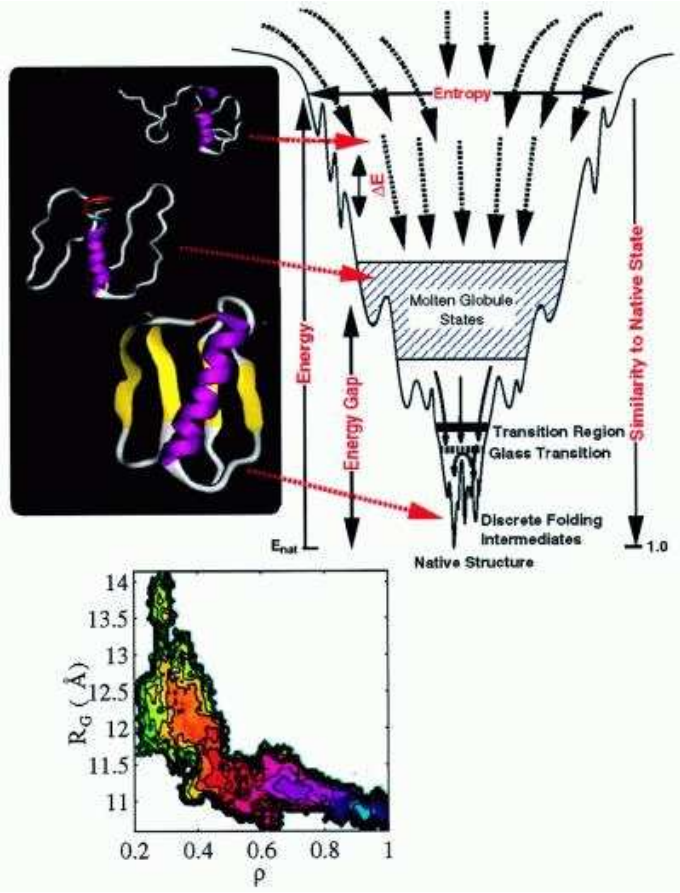
(Left: Fig. from [2] B. Alberts, et al, Molecular biology of the cell 1994; Righ: Fig. from G. Karp, Cell and molecular biology, 1996, original drawing by I.Geis).



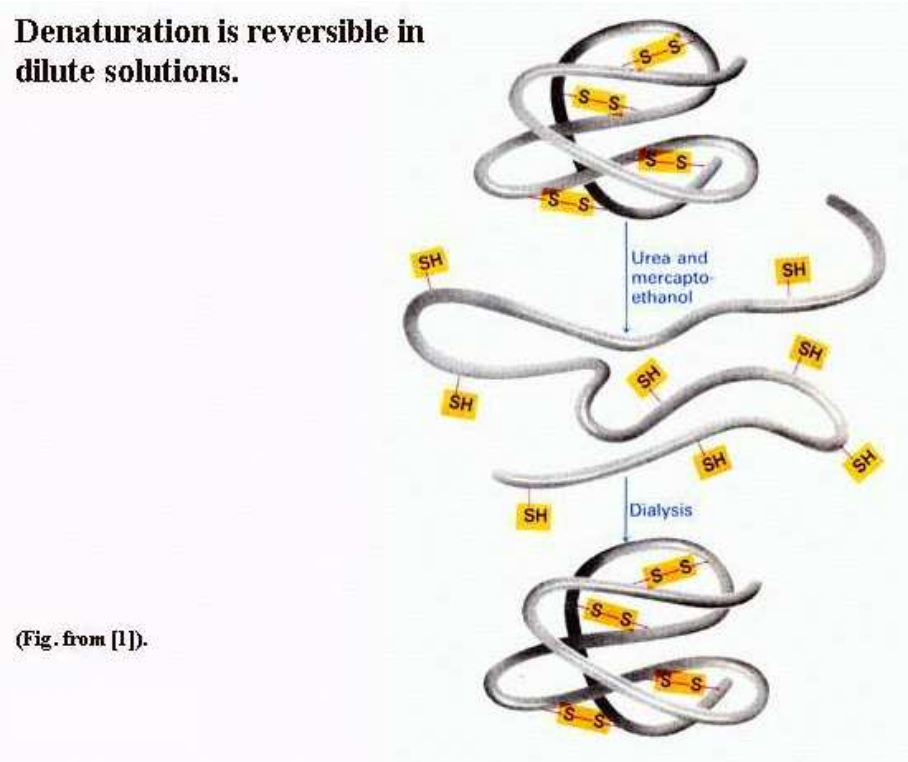
Disposition of hydrophobic and hydrophilic amino acids in the soluble protein cytochrome c.

Thermodynamics **of folding:** Rugged free energy surface landscape **for an ensemble of conformations.** Entropy funnel. Molten **globule and** native **states are separated by an activated nucleation process.**





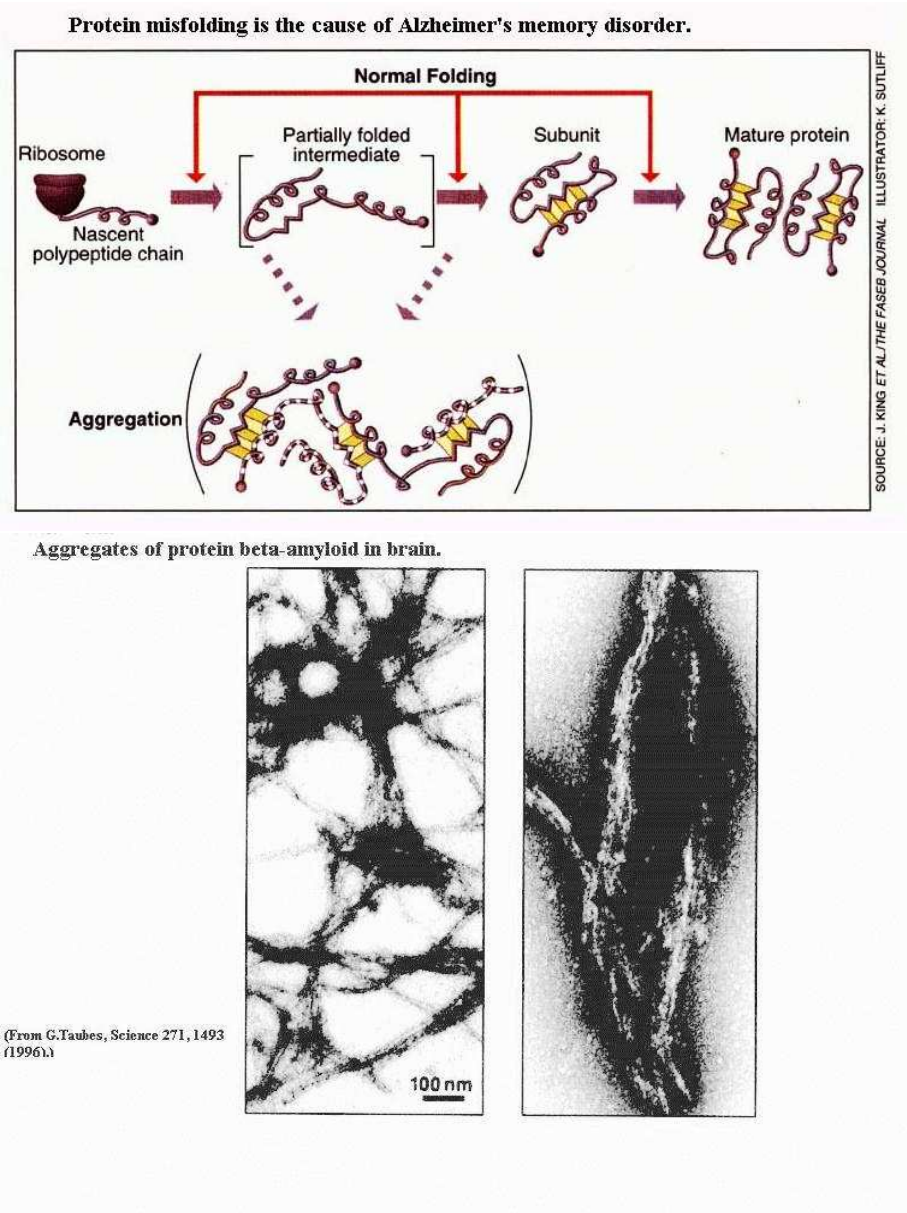
Denaturation is reversible thermodynamically for most proteins, but some utilise chaperons to aid folding kinetically by hiding some 'sticky' parts.



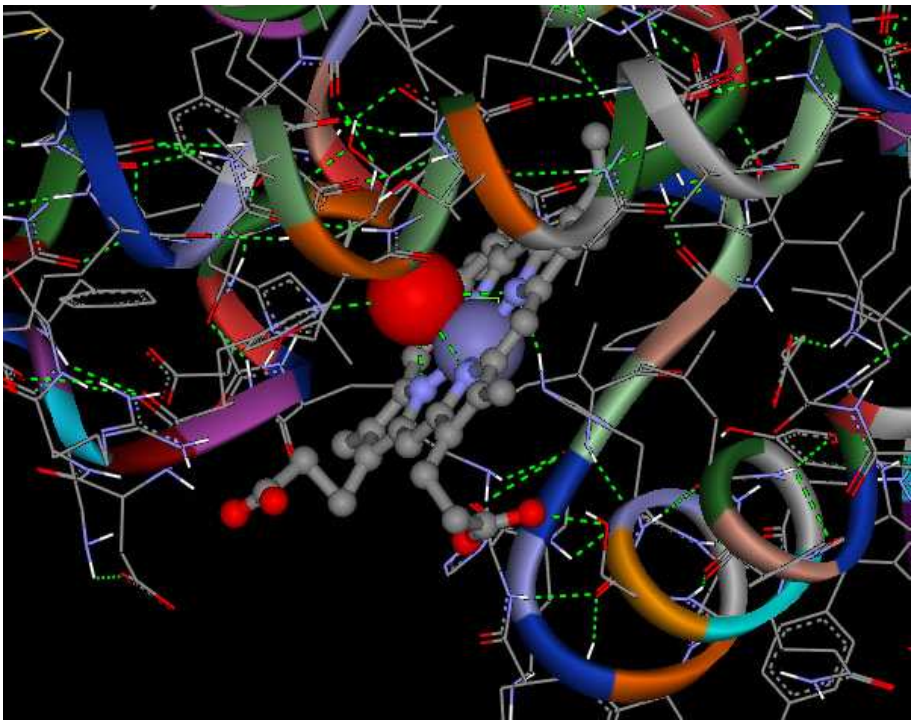
(Fig. from [1]).

Increasingly large number of conformational diseases are now understood to be related to protein misfolding (e.g. Alzheimer's, Huntington's, CJD, BSE, Kuru, cataract) resulting in plaques

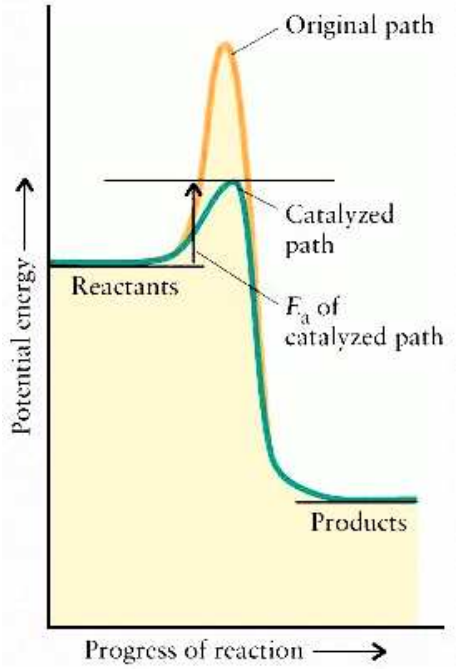
formation **akin to crystallisation process.**

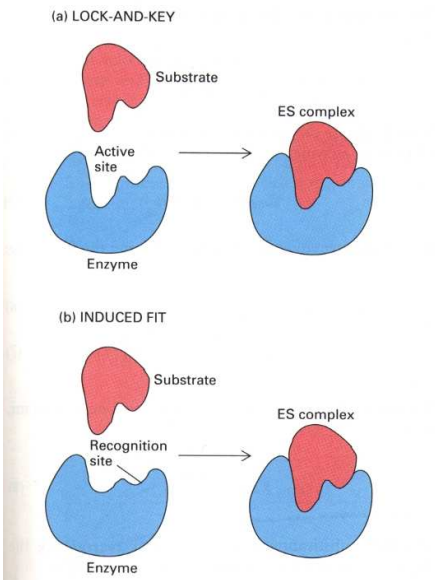


X-ray **and neutron crystallography allows one to determine the shapes of many structured proteins in their native state.**

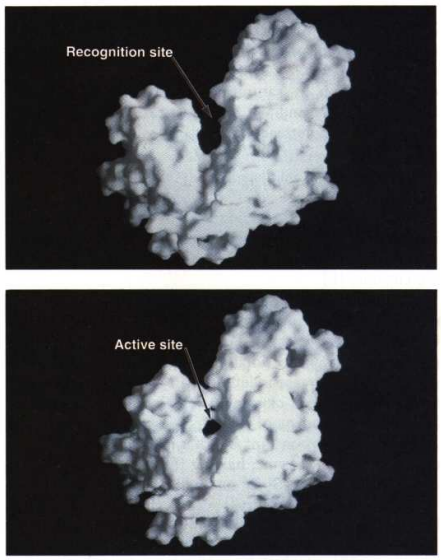


Enzymes **are amazingly efficient** catalysts **which act in a** lock–key **fashion to facilitate chemical reactions at an** active site.





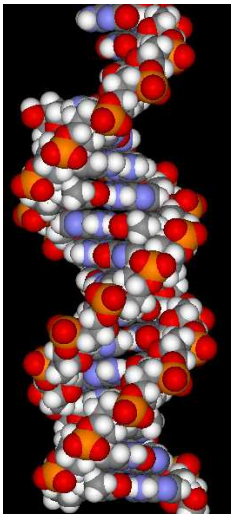
▲ FIGURE 3-27 Two mechanisms that position substrate for catalysis. (a) In enzymes that exhibit the lock-and-key mechanism, the active site contains residues involved in both the binding of substrate and catalysis. In this case, binding of substrate to the complementary active site positions it for catalysis. (b) In enzymes that exhibit the induced-fit mechanism, binding of substrate induces a conformational change in the enzyme that brings the residues involved in catalysis into proximity of the substrate.

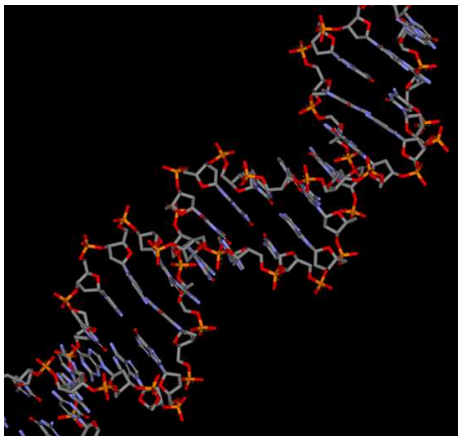


▲ FIGURE 3-28 Computer-generated images of hexokinase illustrating the induced-fit mechanism. (*top*) Hexokinase has an open conformation when glucose is not bound to it. ATP binds at the base of the deep cleft between the two domains. (*bottom*) Binding of glucose (not shown) induces a conformational change that brings the two domains closer together and positions the glucose molecule close enough to the bound ATP that the reaction can proceed. [Courtesy of Gareth White. Adapted from C. M. Anderson, R. E. Stenkamp, and T. A. Steitz, 1978, J. Mol. Biol. **123**:15.]

Lecture 3

Nucleic acids: DNA and RNA conformation and complexes

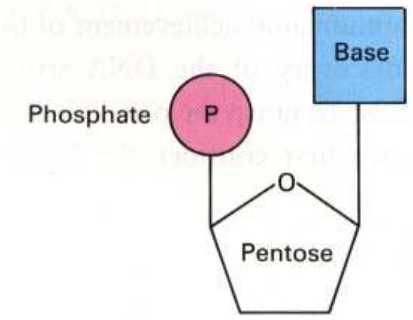




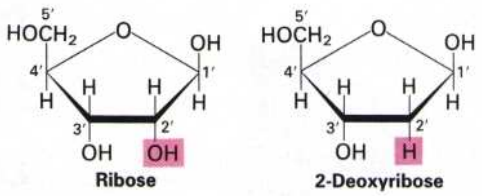
Deoxyribonucleic acid (DNA) **stores genetic information of a cell.** Types: nuclear and mitochondrial.
Ribonucleic acid (RNA) **plays a number of important functions in a cell:** transcription, translation. Types: messenger-, transfer- and ribosomal- RNA.

Both nucleic acids are linear polymers composed of nucleotide **monomers. Degree of polymerisation for RNA $N = 100 - 4,000$, for DNA $N > 100,000,000$.**

There are 5 nucleotides, but DNA (A,G,T,C) and RNA (A,G,U,C) consist of 4 of those only. A nucleotide consists of a negatively charged phosphate group (after dissociation in aqueous solution releasing H_3O^+, hence an acid)**, a pentose (5-carbon sugar molecule) an organic base.**

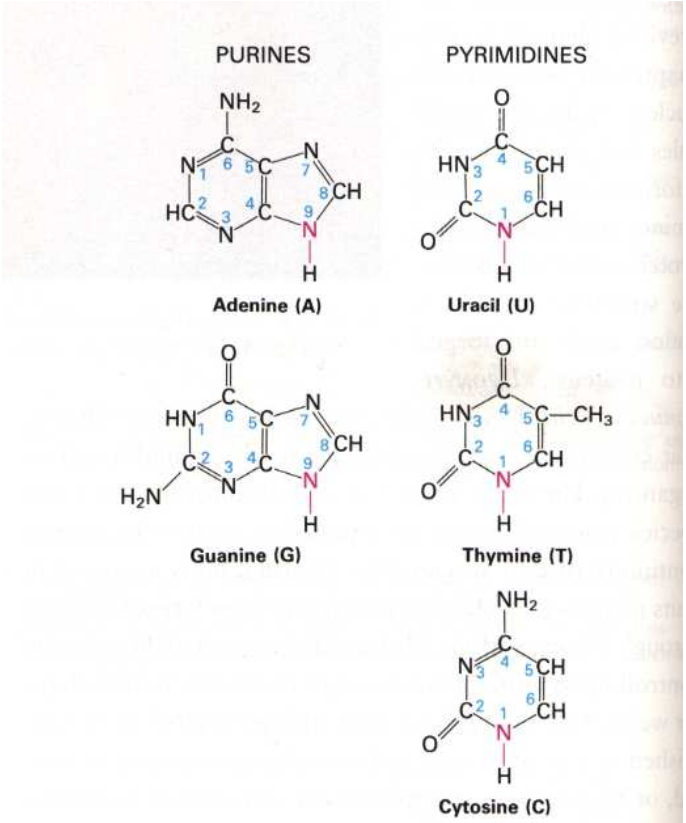


In RNA **the pentose is always** ribose **and in** DNA **it is** deoxyribose.



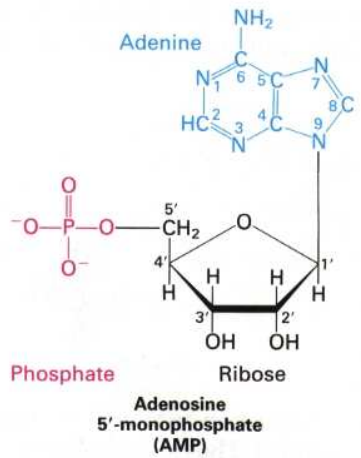
▲ FIGURE 4-2 Haworth projections of the structures of ribose and deoxyribose. By convention, the carbon atoms of the pentoses are numbered with primes. In nucleotides and nucleic acids, the 5′ carbon is linked in an ester bond to the phosphate and the 1′ carbon is linked to the base.

The bases **are divided into** Purines: Adenine (A), Guanine (G)**, which have a pair of fused rings, and** Pyrimidines: Uracil (U), Thymine (T), Cytosine (C)**, which have only one ring.**
Presence of N in these heterocyclic bases contributes to their reactivity.



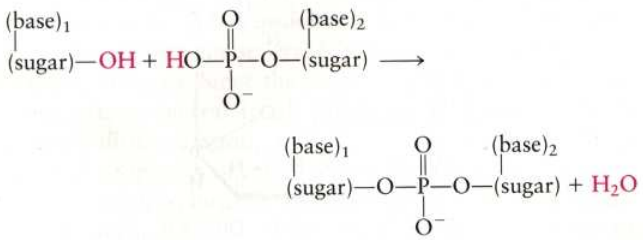
▲ FIGURE 4-3 The chemical structures of the principal bases in nucleic acids. In nucleic acids and nucleotides, the 9 nitrogen atom of purines and the 1 nitrogen atom of pyrimidines (red) are bonded to the 1′ carbon of ribose or deoxyribose.

Sugar component links the base and the phosphate: 1′ C **of a sugar is attached to** 9 N of a purine (1 N of a pyrimidine) **in a C-N bond so that the base lies above the plane of the sugar; the hydroxyl (CO) at** C 5′ **of the sugar is replaced by an** ester bond **to the phosphate group.**

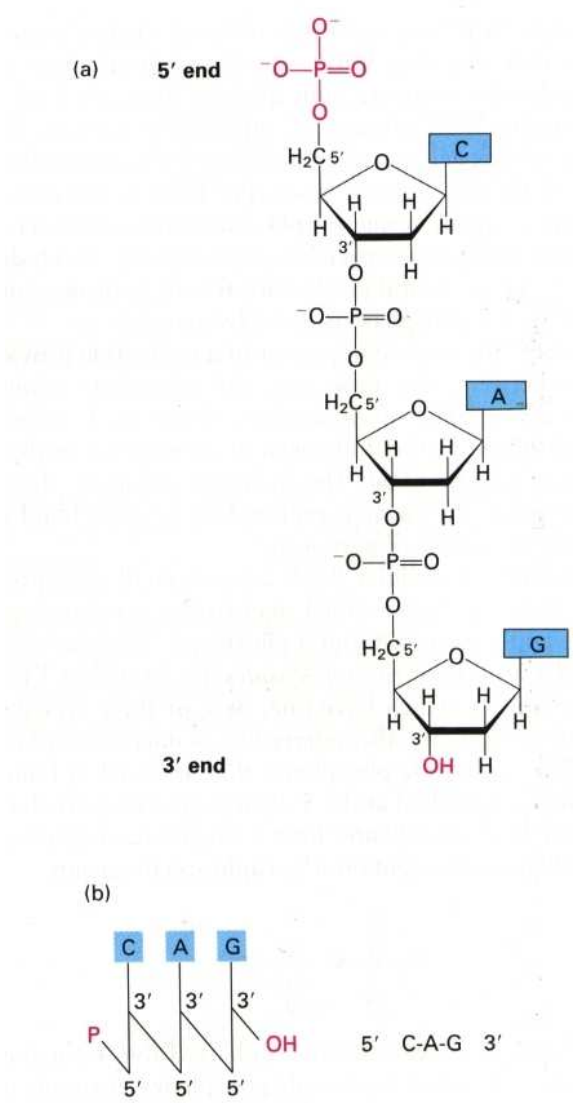


▲ FIGURE 4-4 The chemical structure of a typical nucleotide, adenosine 5′-monophosphate (AMP). The C–N linkage between the sugar and the base shown here is a β linkage, meaning that the base lies above the plane of the ribose, as conventionally drawn. (In the α linkage, the base would lie below the plane.) Only the β linkage is found in cellular nucleotides. One additional phosphate (P–O–P) in ester linkage would constitute adenosine 5′-diphosphate (ADP), and two additional phosphates (P–O–P–O–P) adenosine 5′-triphosphate (ATP).

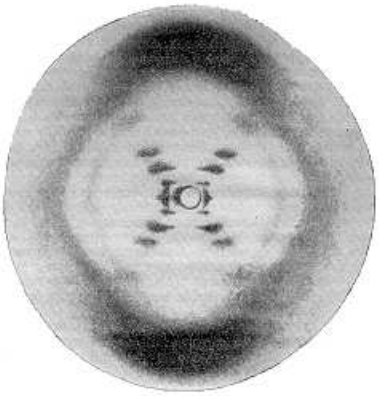
Polymerisation **of nucleic acids occurs in the manner similar to forming polypeptides with release of a water molecules. Links between nucleotides are called** phosphodiester bonds.



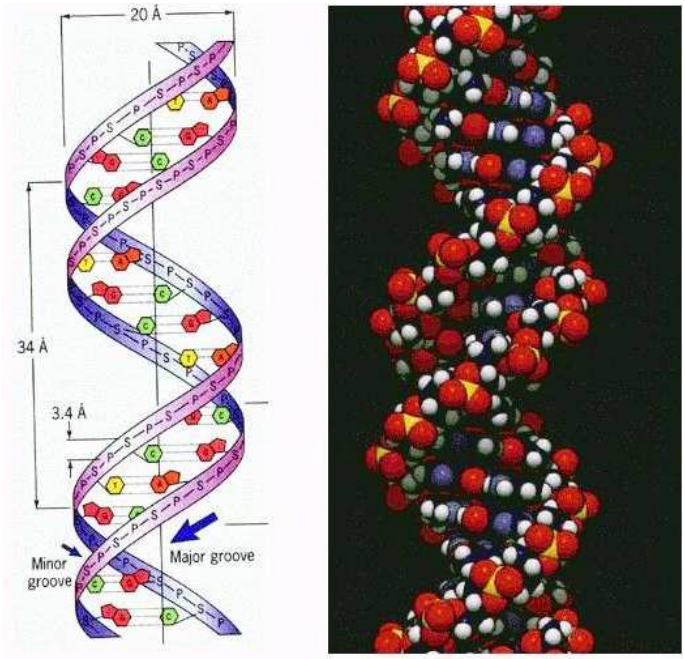
Nucleic acids chains are written in terms of base sequences from 5' to 3' end.



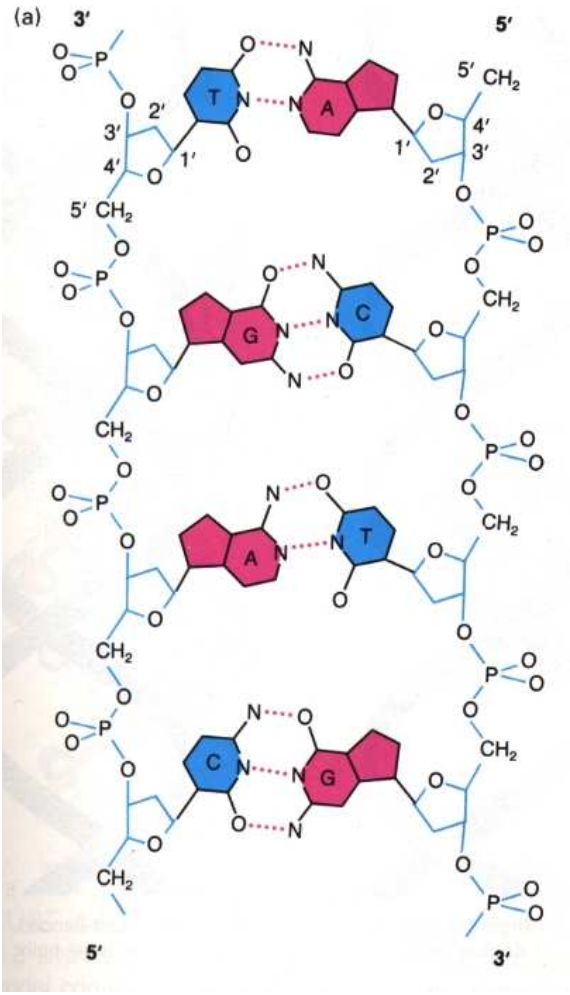
In 1953 Francis Crick and James Watson have published the famous double helix (duplex) DNA structure based on the X-ray data obtained by their colleague Rosalind Franklin.

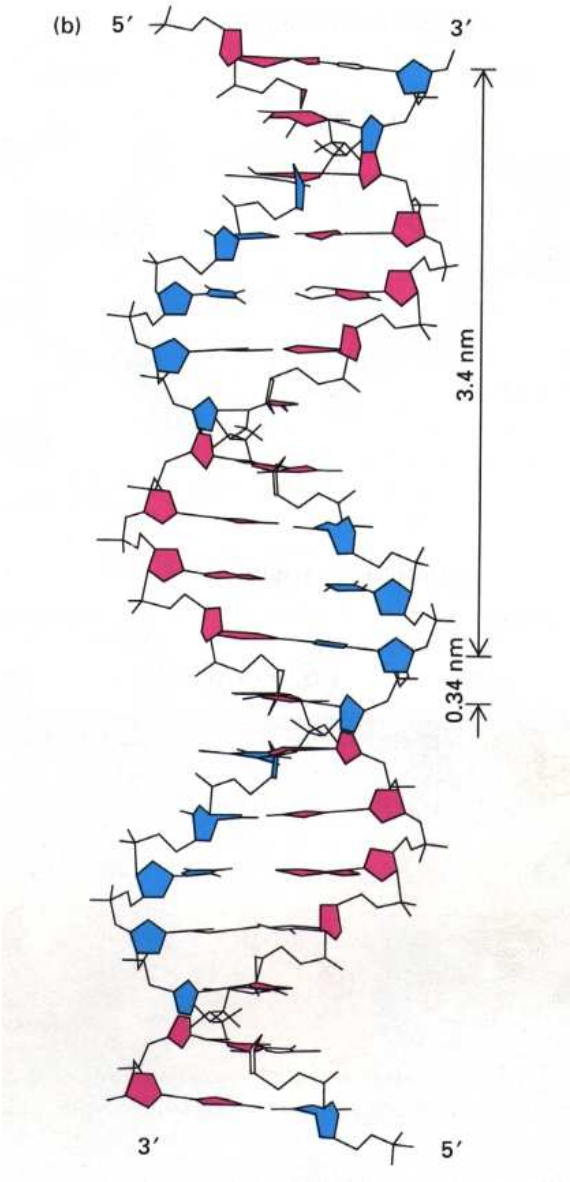


The two sugar–phosphate backbones **lie on the outside of the double helix and the bases project onto the interior. The adjoining bases in each strand stack on top of each other in parallel planes.**

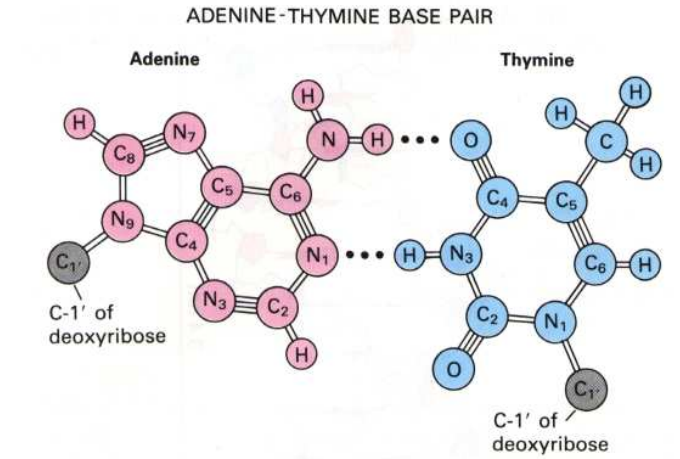


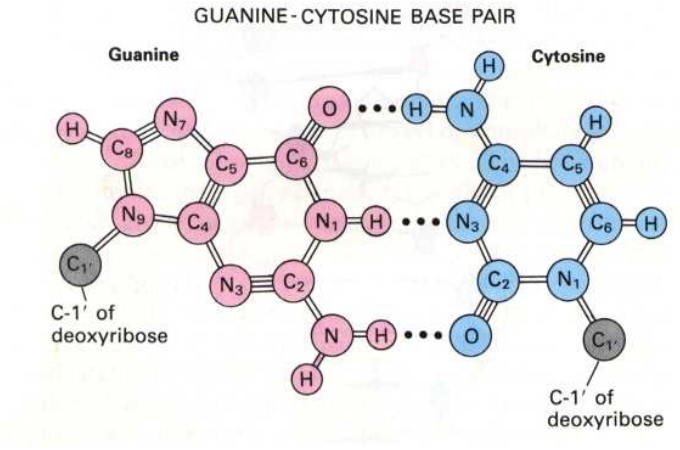
The orientation of two strands is antiparallel **(their 5'-3' directions are opposite). The two strands are held together by the cooperation of many hydrogen bonds in addition to hydrophobic interactions.**



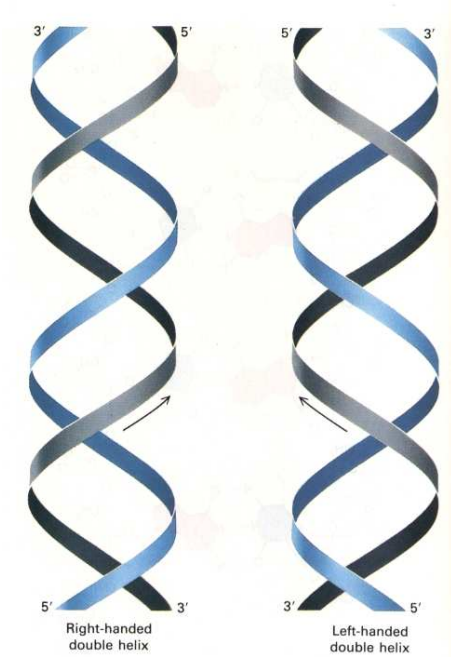


Watson–Crick pairing **of bases is regular between the two strands: A is pared with T by 2 hydrogen bonds, G is paired with C by 3 hydrogen bonds.**





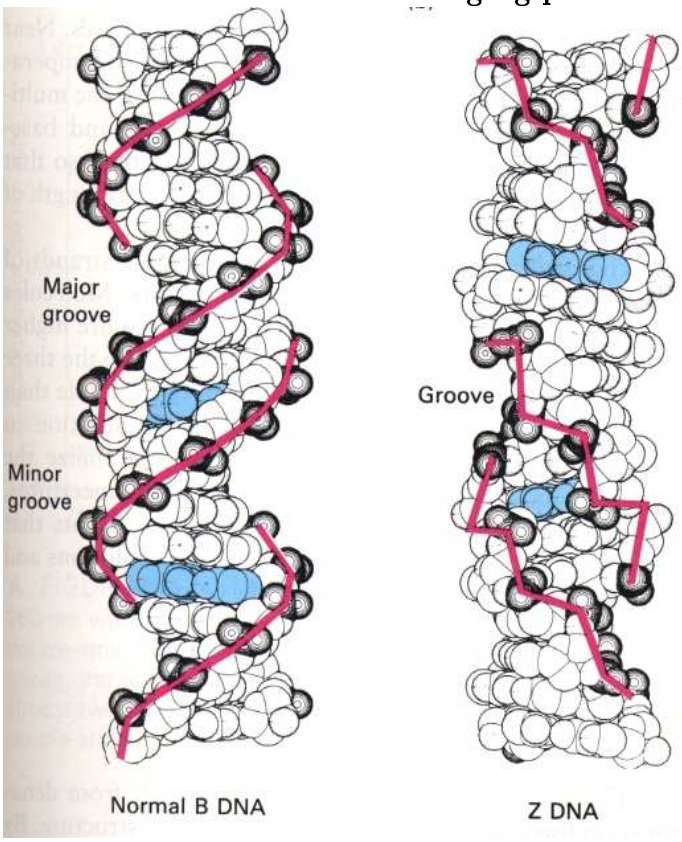
Natural DNA is always right–handed.



DNA conformations:

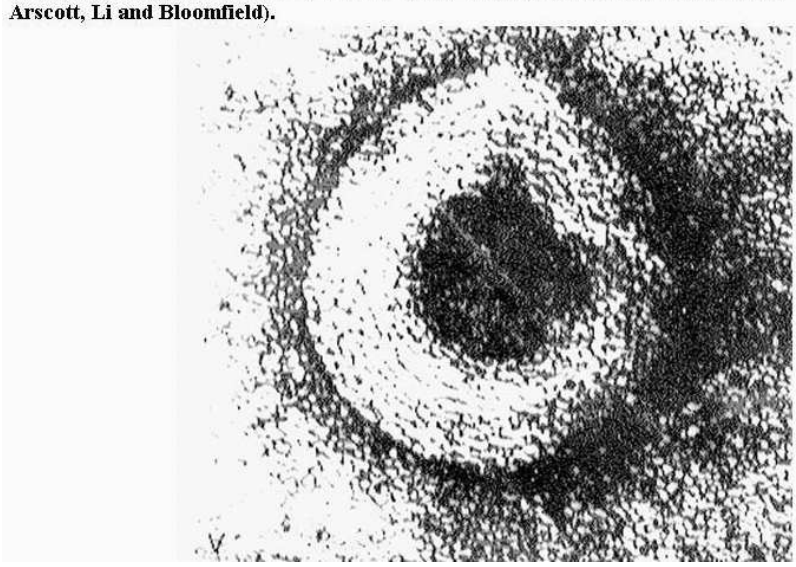
B-DNA **occurs in aqueous solution (normal form of DNA) and it has about 10 pairs per turn. This form has** two grooves: major and minor.

A-DNA is more compact and occurs in non–aqueous solutions. It has about 11 pairs per turn. Z-DNA is left–handed and has zigzag pattern of bases as viewed from the side.

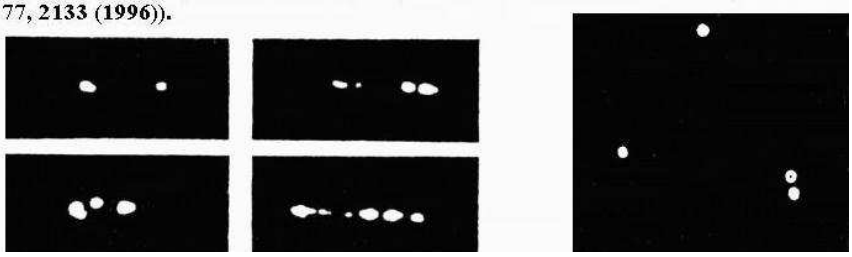


Condensation of DNA by counter–ions to compensate the repulsion between phosphates produces toroidal globules.

Toroidal DNA globule. Linear size 1000 A, chain length about 10,000 bps. (From Arscott, Li and Bloomfield).



Fluorescence images of DNA collapse. (From M.Ueda, K.Yoshikawa, Phys.Rev.Lett. 77, 2133 (1996)).



A lot of information about DNA hydration has been accumulated using experimental techniques such as X-ray crystallography, infrared, NMR and Raman spectroscopy, volumetry, ultrasonics, and Brillouin scattering.

The short–range hydrophilic interactions of electronegative DNA atoms with water give rise to a strongly sequence– and conformation–dependent organisation of water molecules forming well localised network on parts of the DNA–solvent interface. The rest of the water in the system can be thought to affect DNA via the hydrophobic interaction as well as via its bulk–medium properties (a high dielectric constant and low compressibility background).

The discovery of structured water around a B-DNA fragment further catalysed the quest for understanding DNA hydration structure. The presence of this spine of hydration, consisting of a string of water molecules in AT-rich regions of the DNA minor groove, has raised many questions.

A promising alternative for addressing the equilibrium issues is based on the methods of Statistical Mechanics and, particularly, on the integral equations Reference Interaction Site Model (RISM). The main advantage of this method is that here one explicitly deals with all quantities averaged over the equilibrium distribution. This method, first proposed by Chandler and Andersen, is based on the RISM Ornstein–Zernike integral equation for the correlation functions.

$$\rho\, h\, \rho = \rho\, w \bullet c \bullet w\, \rho + \rho\, w \bullet c \bullet \rho\, h\, \rho, \quad \rho_i^{(1)}(r) = \sum_\alpha \langle \delta(r - R_i^\alpha) \rangle\,,$$
$$\rho_{ij}^{(2)}(r, r') = \sum_{\alpha \neq \beta} \langle \delta(r - R_i^\alpha) \delta(r' - R_j^\beta) \rangle.$$

These correlation functions are further decomposed as $\rho_{ii}^{(2)} = \rho_i^{(1)} g_{ii} \rho_j^{(1)}$ and $g_{ij} = 1 + h_{ij}$. Here c_{ij} are called direct correlation functions, and the conformational matrix is $w_{ij}(r) = \delta(r - r_{ij})$, for i, j in the same molecule and zero otherwise, where the distance between these sites is fixed equal to r_{ij}. The correlation function $h(r) + 1$ has the meaning of the probability density to find an atom at the distance r from another one.

For a single N-atomic molecule in aqueous solution this equation may be written in the Fourier space as,

$$G = (\hat{V} \otimes \hat{W} - \hat{I}_{2N})\, C, \quad W_{ij} = \frac{\sin(k r_{ij})}{k r_{ij}}, \quad \hat{V}_{22} = 1 + \rho \tilde{h}_{OO}(k),$$
$$\hat{V}_{11} = 1 + 2\rho \tilde{h}_{HH}(k) + \frac{\sin(k\, r_{HH})}{k\, r_{HH}},$$
$$\hat{V}_{12} = \hat{V}_{21}/2 = \rho \tilde{h}_{OH}(k) + \frac{\sin(k\, r_{OH})}{k\, r_{OH}},$$

with G and C being columns of size 2N containing site–site correlation functions $G = \{\tilde{\gamma}_{iH}(k), \tilde{\gamma}_{iO}(k)\}$ and **$C = \{\tilde{c}_{iH}(k), \tilde{c}_{iO}(k)\}$** respectively, and **$\tilde{\gamma}_{iX} = \tilde{h}_{iX} - \tilde{c}_{iX}$ are the indirect correlation functions. Here we have also used the correlation functions of** pure water.
In addition, one also needs an additional closure relation **between the two sets of correlation functions. We shall use a version of the** Hyper Netted Chain (HNC) approximation **with the renormalised Coulomb potential for the short–range parts of the correlation functions $\gamma_{iX} = \gamma^s_{iX} - f_{iX}$, $c_{iX} = c^s_{iX} + f_{iX}$:**

$$\begin{aligned} c^s_{iX}(r) &= \exp(-U_{iX}(r) + f_{iX}(r) + \gamma^s_{iX}(r)) - \gamma^s_{iX}(r) - 1, \\ f_{iX}(r) &= -q_i q_X \frac{(1 - \mathrm{erf}(r))}{r k_B T}. \end{aligned}$$

Fig: **Stereographic representation of canonical structures of the** A-DNA (a), **the** B-DNA (b), **and the** hydrated B-DNA with the sequence **$\mathrm{d(CCGG)_3 \cdot d(CCGG)_3}$** (c), **where dark dots show points with the water density 5 times higher than that of the bulk–water.**

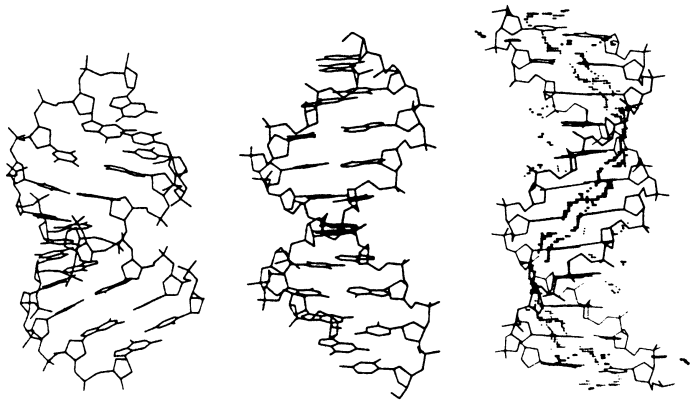
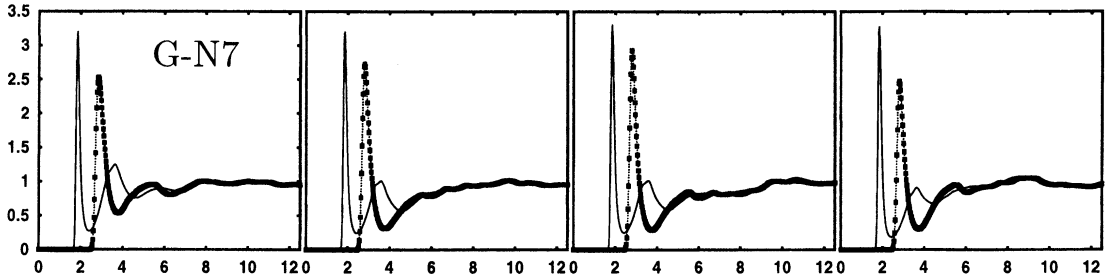
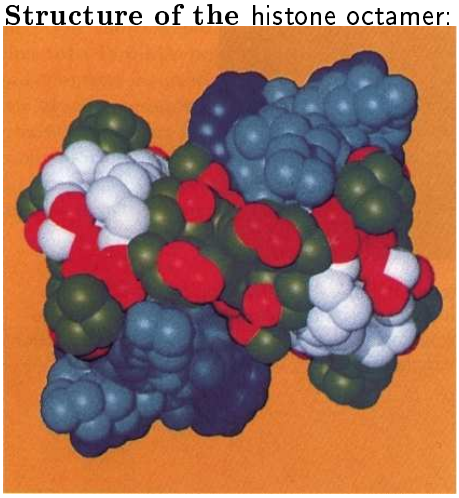


Fig: **Plot of the** radial correlation function, $g(r)$, of N7 atom of guanine with hydrogen (solid line) and oxygen (dashed line with points) atoms of water, **where the radius is measured in Angströms. Plots correspond (from left to right) to the 1st, 2nd, 3rd, and 5th nucleotide pairs of the** B-DNA duplex **d(GGGGG)· d(CCCCC) starting from the 5' end.** (From: D.A. Tikhonov, R.V. Polozov, E.G. Timoshenko Yu.A. Kuznetsov, A.V. Gorelov, K.A. Dawson. "Hydration of a B-DNA fragment in the method of atom–atom correlation functions in the RISM approximation." J. Chem. Phys., Vol. 109, No 4, pp. 1528-1539 (1998).)

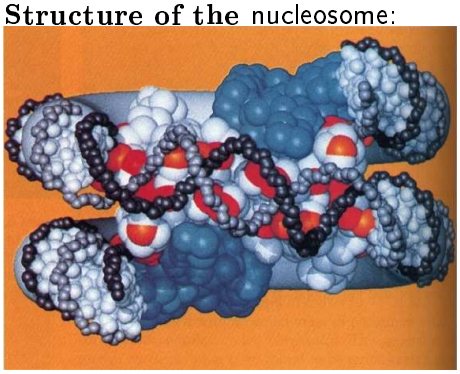


In eukaryotic cells DNA is compacted **together with** histone proteins **to form a highly complex** chromatin complex.

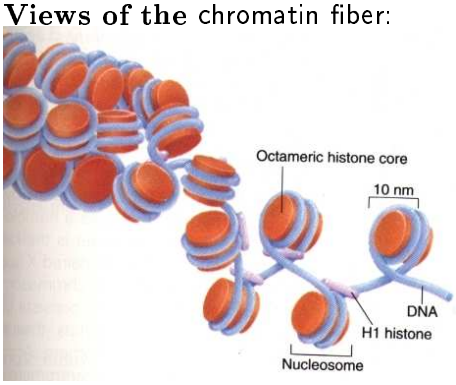
Structure of the histone octamer:

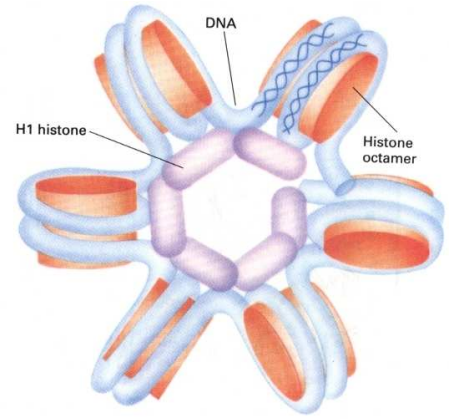


Structure of the nucleosome:

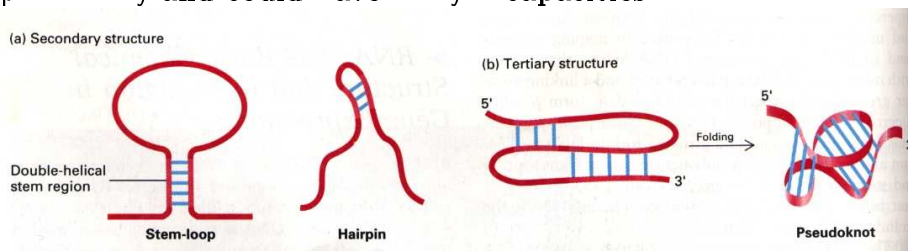


Views of the chromatin fiber:



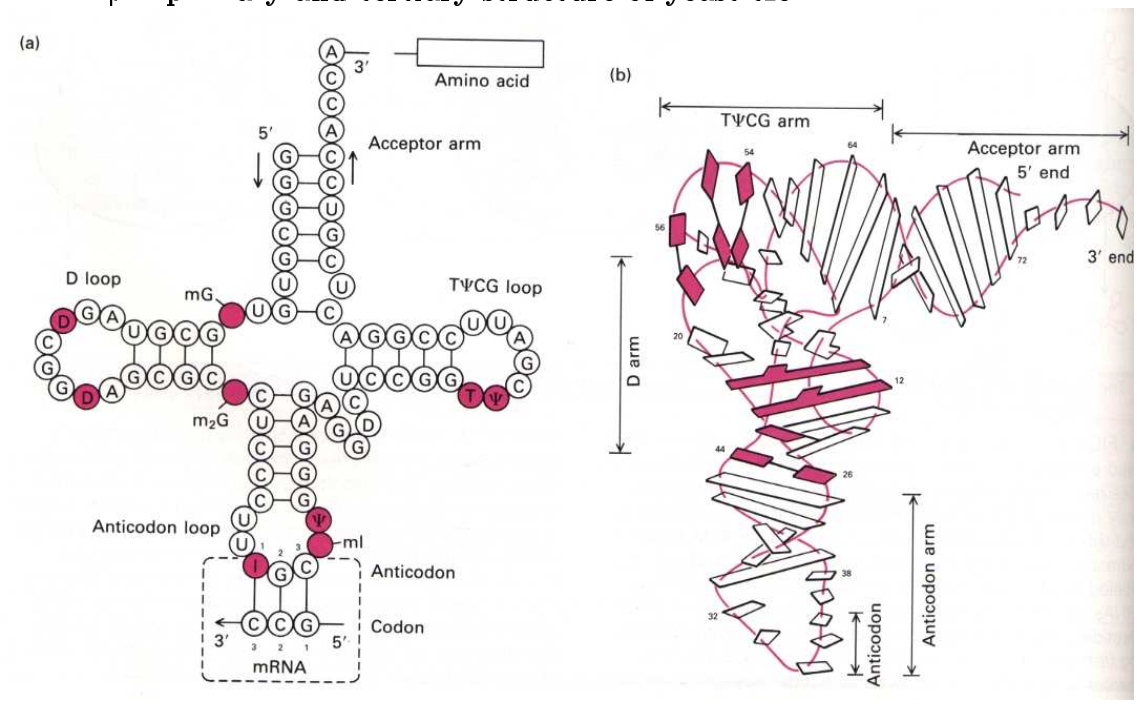


RNA **is generally similar to DNA, but an additional** hydroxyl group at C 2' **of ribose makes RNA more chemically active than DNA.**
RNA can be double stranded, single stranded, linear or circular, as well as form a DNA-RNA hybrid duplex.
Small tRNA **can adopt a well–defined** tertiary structures**, while** rRNA **have locally structured domains (secondary structures) connected by spacers. These are stabilised by** Watson–Crick complementarity **and could have** catalytic **capacities.**

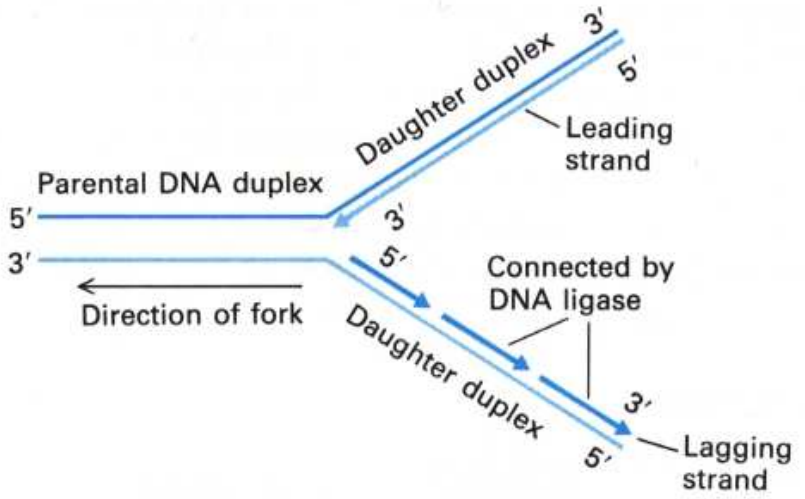


▲ FIGURE 4-16 RNA secondary and tertiary structures. (a) Base pairing between distant complementary segments of RNA molecules can form structured domains containing stem-loops, hairpins, and other secondary structures. In stem-loops, the single-stranded loop (dark red) between the base-paired helical stem (light red) may be hundreds or even thousands of nucleotides long, whereas in hairpins, the short turn may contain as few as 6–8 nucleotides. (b) Interactions between the flexible loops may result in three-dimensional folding to form tertiary structures such as the pseudoknot. This tertiary structure resembles a figure-eight knot, but the free ends do not pass through the loops, so no knot is actually formed. [Part (b) adapted from C. W. A. Pleij, K. Rietveld and L. Bosch, 1985, *Nucl. Acids Res.* **13:**1717.]

An example: **primary and tertiary structure of yeast tRNA.**

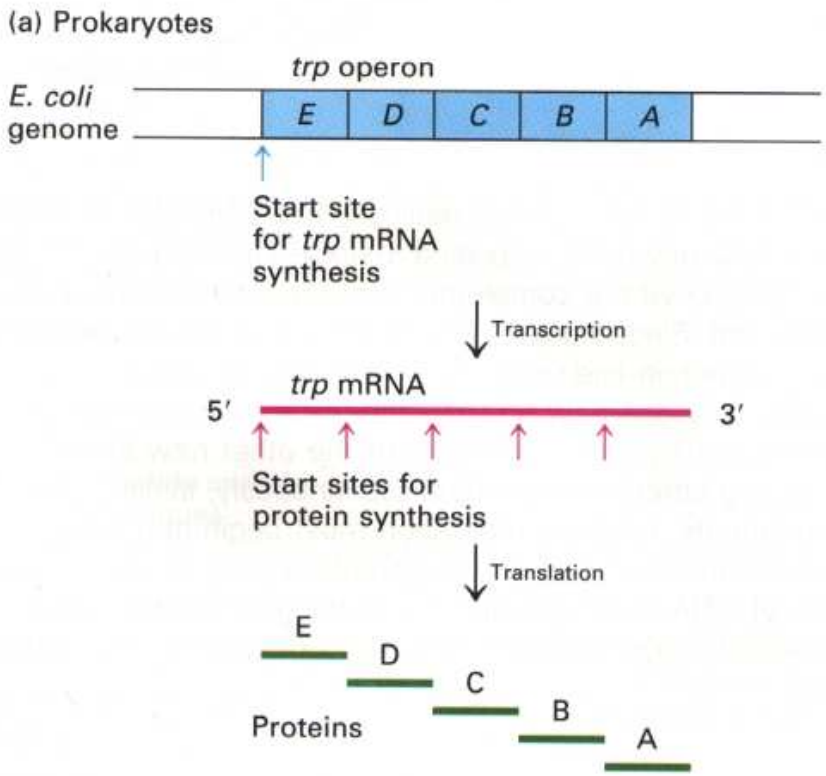


During cell division DNA replication **occurs producing two identical DNA helices.**



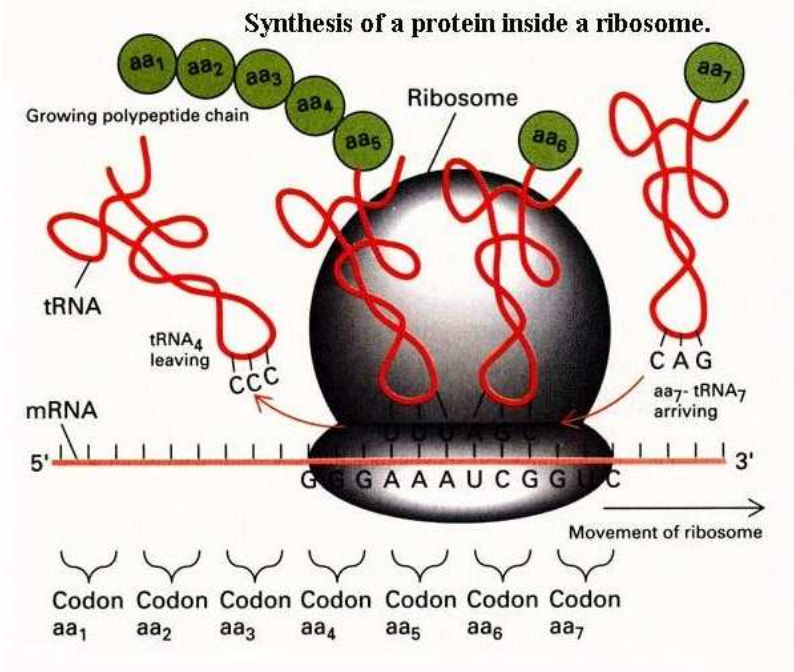
▲ FIGURE 4-22 During replication of DNA, the duplex is sequentially unwound, exposing a single-stranded region (the growing fork) where complementary daughter strands are synthesized. Since nucleic acid chains can grow only in the 5′ → 3′ direction (indicated by arrowheads on daughter strands), only one new strand—the leading strand—is synthesized continuously. Synthesis of the other new strand—the lagging strand—proceeds discontinuously, initially forming fragments, because replication must begin at a new (downstream) start site on the parental strand as each new region of DNA is exposed at the growing fork. DNA ligase connects the lagging-strand fragments to form a continuous strand.

After replication, the DNA transcription **into the mRNA and** translation **of the mRNA into proteins take place.**



RNA has three roles in protein synthesis: mRNA **(which carries genetic information from DNA in a 3-letter genetic code for each residue) is translated into proteins by a joint action of the ribosome and** tRNA. **The ribosome is a structure composed of numerous proteins and two major** rRNA **molecules.**

tRNA has two functions: **to chemically link to a particular amino acid and to recognise a codon in mRNA so that the amino acid can be added to a growing peptide chain.**

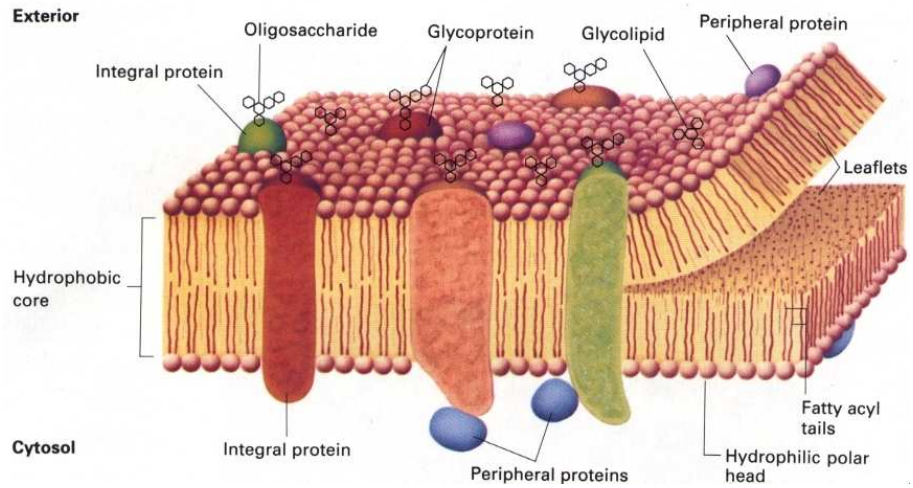


RNA is believed to be more ancient **than DNA.**
Some viruses contain only RNA (retroviruses) **as their genetic code.**
RNA is more versatile **than DNA and plays some** functions similar to proteins **and its understanding is important for elucidating the** biological evolution in the 'primordial soup'.

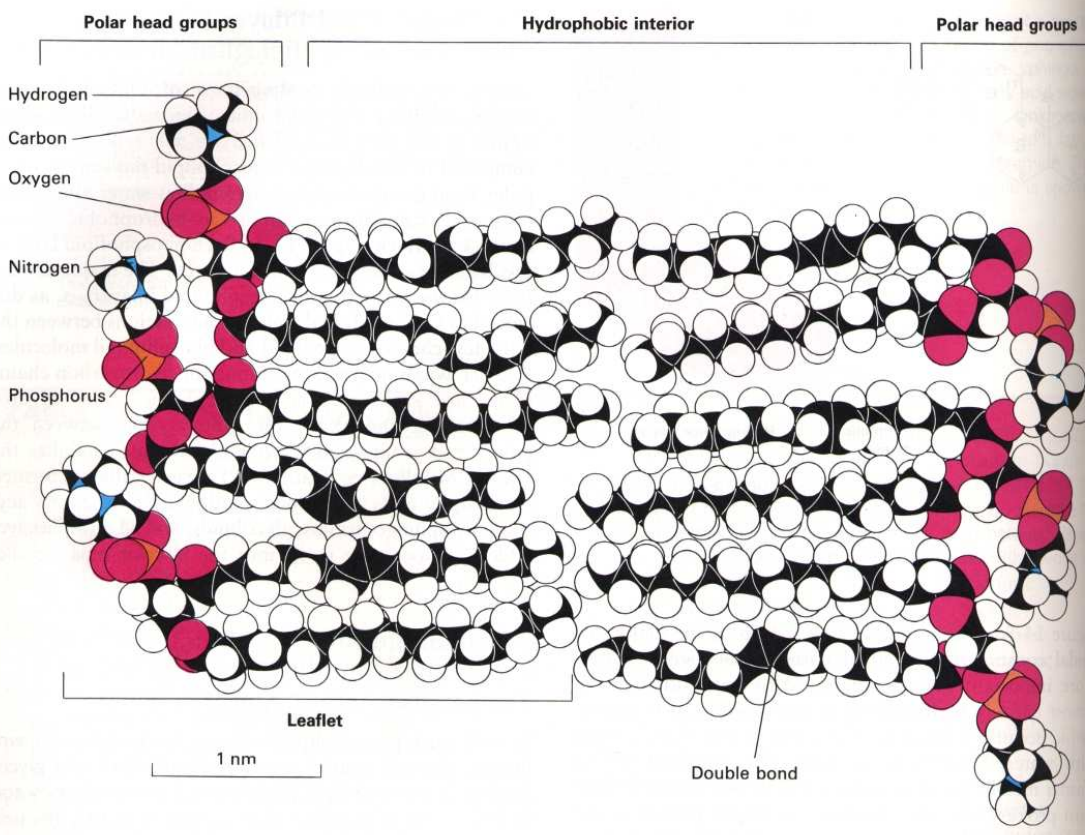
Lecture 4

Inter–molecular interactions and self–assembly.
Biomaterials and bio nano science

A cell membrane **has a** bilayer **a self–assembled structure structure consisting of** amphiphilic phospholipids **and** membrane proteins.

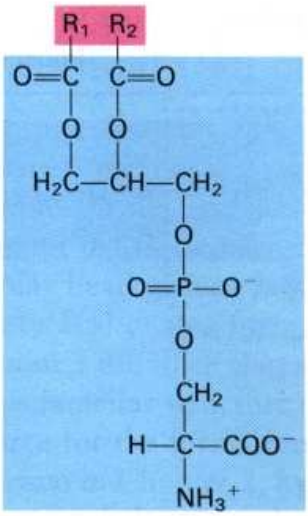


Amphiphilic **molecules contain a** polar head **and a** hydrophobic hydrocarbon tail.

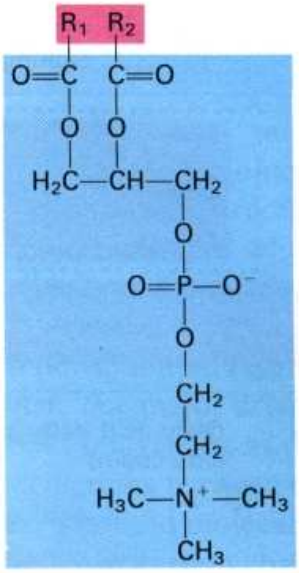


▲ FIGURE 14-7 A space-filling model of a typical phospholipid bilayer. The hydrophobic interior is generated by the fatty acyl side chains. Some of these chains have bends caused by double bonds. The different polar head groups all lie on the outer, aqueous surface of the bilayer. [From L. Stryer, 1988, *Biochemistry*, 3d ed., W. H. Freeman and Company, p. 289. Courtesy of L. Stryer.]

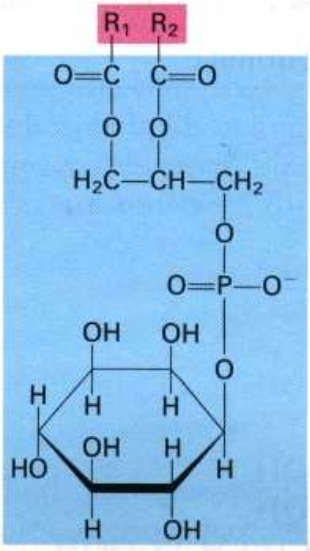
A cell membrane **has a** bilayer **structure consisting of** amphiphilic phospholipids
Phospholipids are predominantly phosphoglycerides **and** steroids **(like cholesterol).**



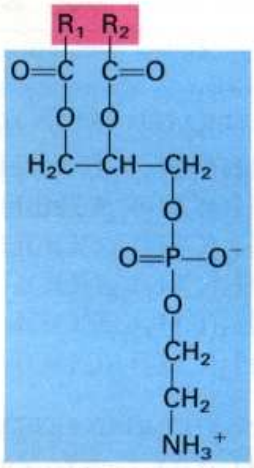
Phosphatidylserine



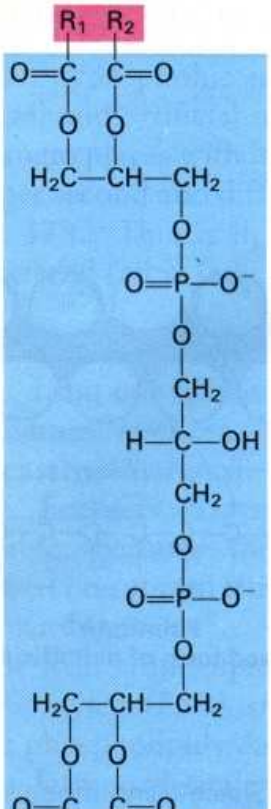
Phosphatidylcholine

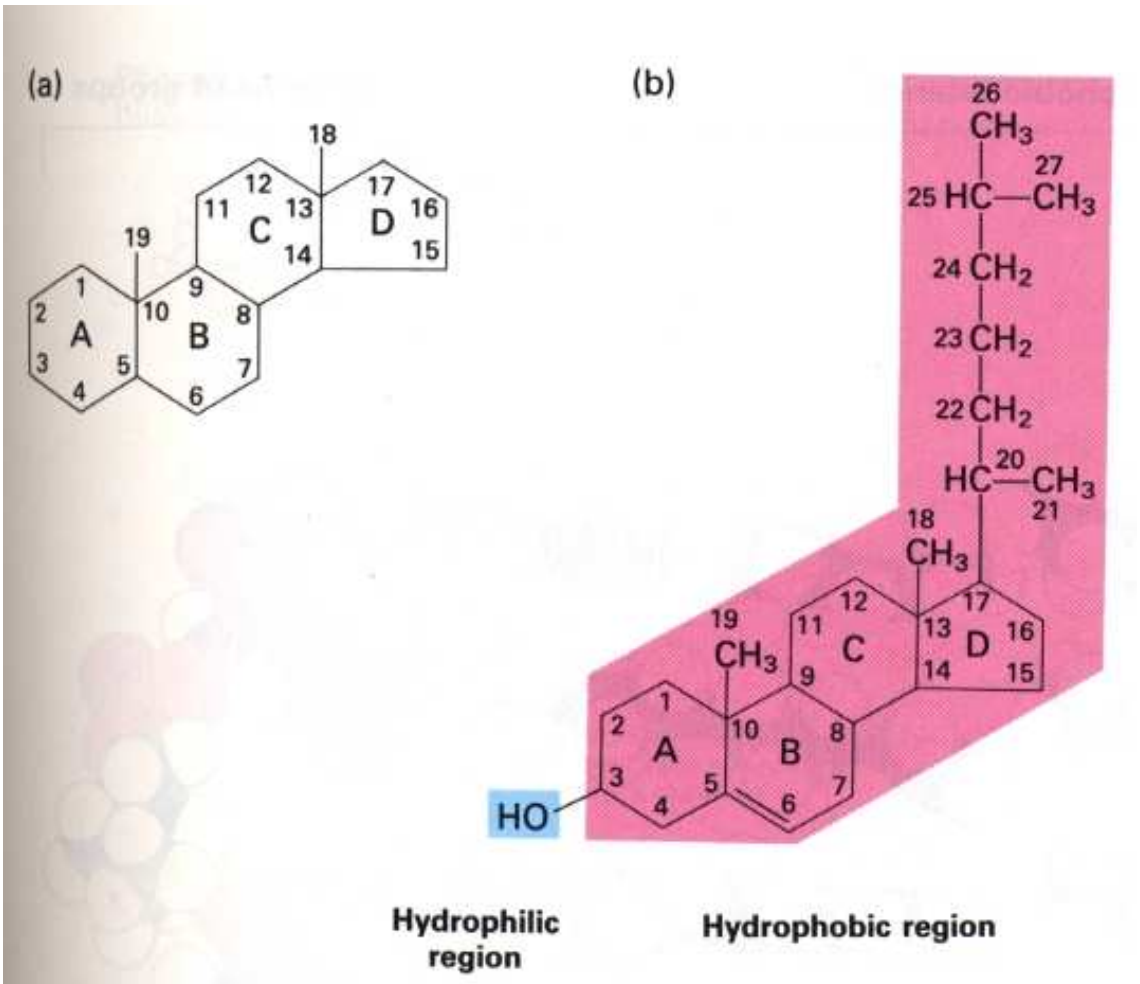


Phosphatidylinositol



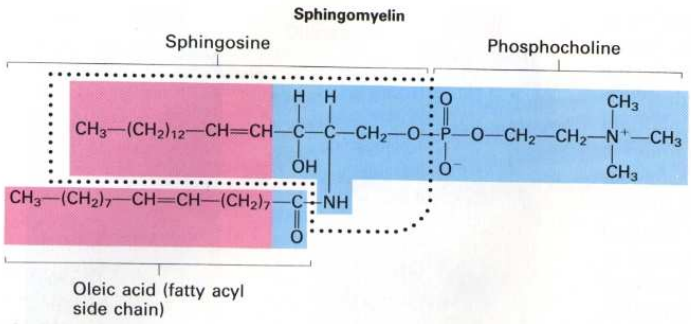
Phosphatidylethanolamine



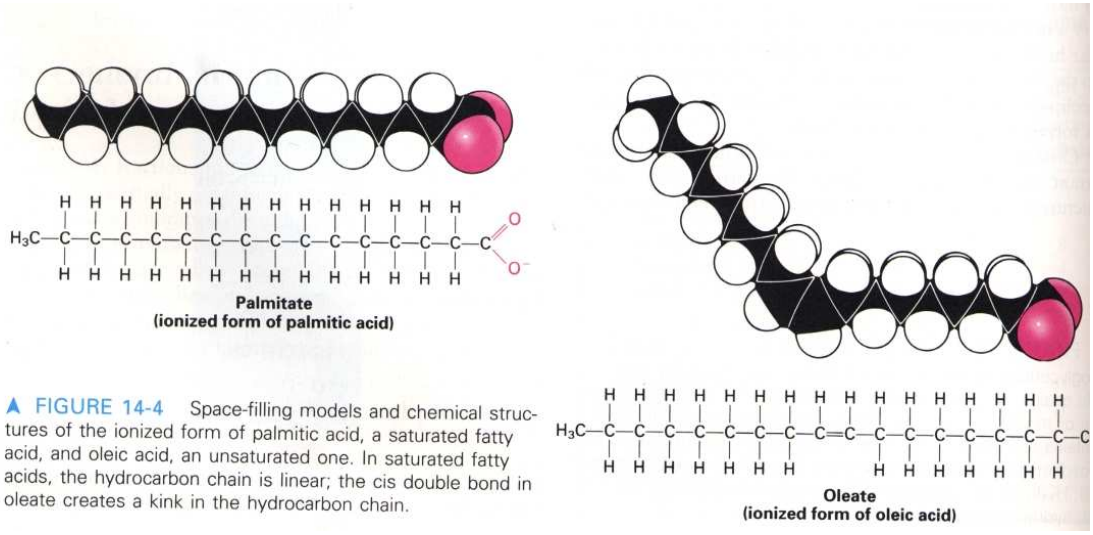


▲ FIGURE 14-5 (a) The general structure of a steroid. All steroids contain the same four hydrocarbon rings, conventionally labeled A, B, C, and D, with the carbons numbered as shown. (b) The structure of cholesterol.

These contain amphiphilic fatty acids.

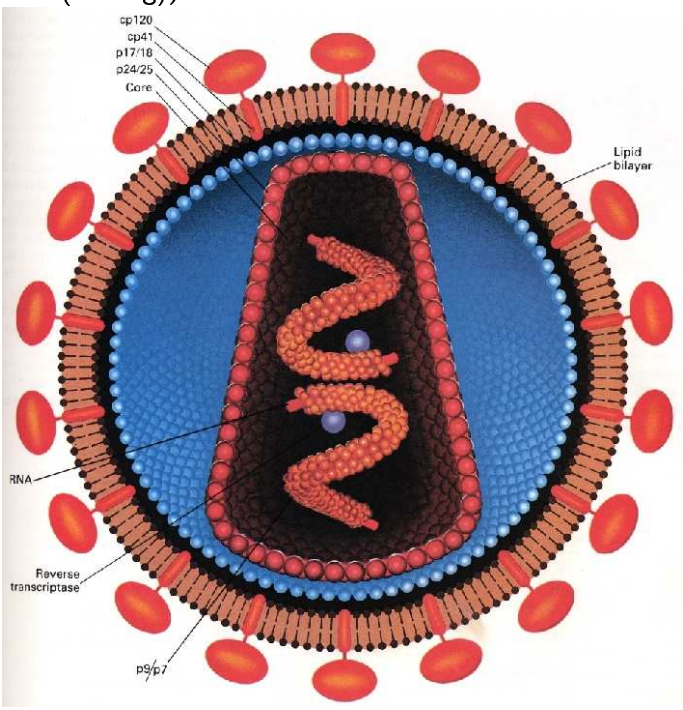


◀ FIGURE 14-3 The structure of a typic sphingomyelin molecule. A group of phospho lipids that lack a glycerol backbone, sphingo- myelin may contain a different fatty acyl side chain than oleic acid (shown here). Linkage c sphingosine (outlined by black dots) to a fatt acid via an amide bond forms a ceramide. Th hydrophobic portion of the molecule is show in red; the hydrophilic, in blue.



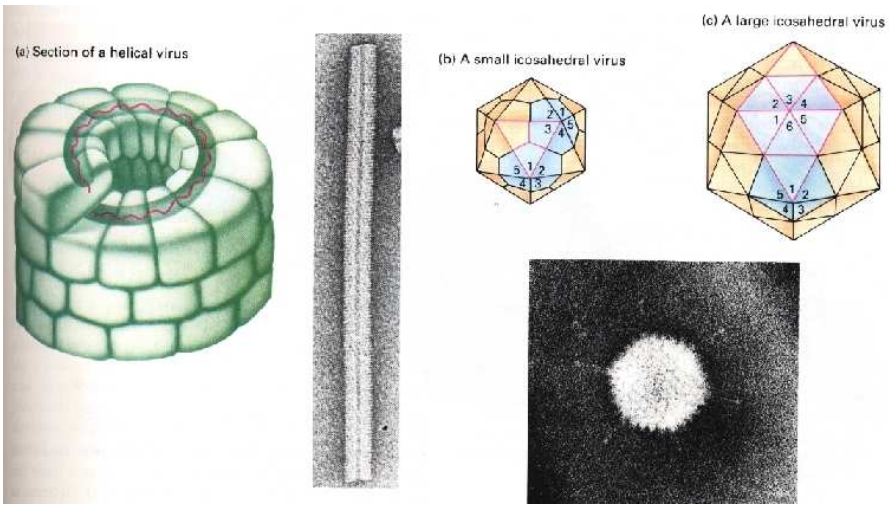
▲ FIGURE 14-4 Space-filling models and chemical structures of the ionized form of palmitic acid, a saturated fatty acid, and oleic acid, an unsaturated one. In saturated fatty acids, the hydrocarbon chain is linear; the cis double bond in oleate creates a kink in the hydrocarbon chain.

A virus **is a small parasite which cannot reproduce by itself but needs a host cell for that. Typically it consists of a** phospholipid bilayer **(or just protein shell), some** membrane proteins **and a** genetic material inside **(single- or double–stranded DNA, or RNA for retroviruses, such as the** HIV (see Fig)).

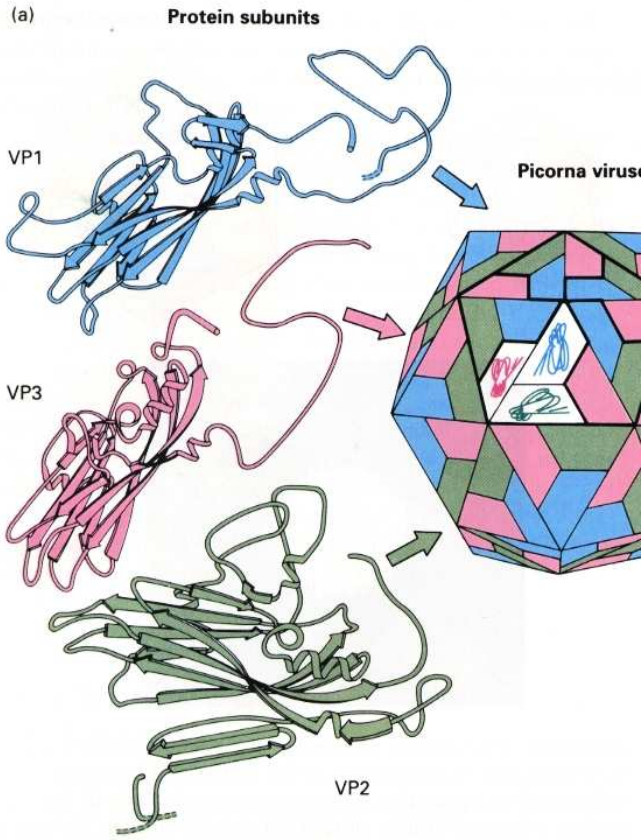


In viruses such as TMV (tobacco mosaic virus) **the protein subunits form self–organised** helical arrays **around RNA or DNA (shown in red), which runs in a groove.** These would self–assemble spontaneously if they are mixed in solution.

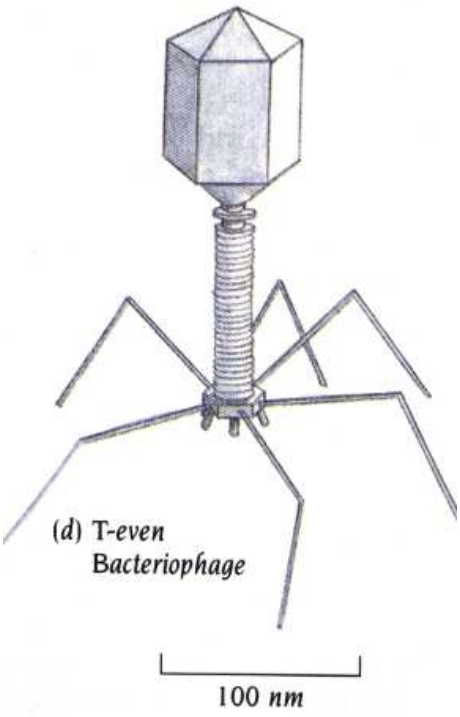
In other viruses the capsid (outer) proteins **self–assemble to form** polyhedra with icosahedral (20-sided) symmetry. **An electron micrograph of** adenovirus is shown in Fig.



Picornaviruses **include** poliovirus and rhinovirus (the common cold). **These have a** capsid composed of three proteins (VP1, VP2, VP3). The symmetry of the shape is to do with too little of genetic code and thus an ability to encode only for the most basic protein shapes.

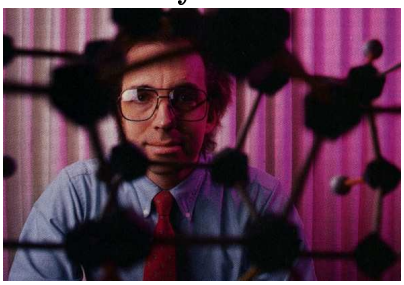


A virus which infects only bacteria is called a (bacterio)phage.



Bio nano Science and Intelligent Biomaterials

Term Nano Science **was suggested by** Eric Drexler **of MIT in his work 'Engines of creation'.** $Nano = 10^{-9}$**, so we are talking about systems of** $10 - 1000$ **of atoms.**
Many exaggerated claims have been made in this area.
However, Bio Nano Science **actually works by using biomolecules designed by Nature through the billions years of evolution.**



As was discovered in 1996 by Shuguang Zhang (MIT) **and independently by** Amalia Aggeli (Univ. of Leeds) **certain short** oligopeptides **can undergo** spontaneous self–assembly **in solution to form** beta–sheet type aggregates and further super–structures.
Fig: **an example of a** rationally designed **oligopeptide sequence.**

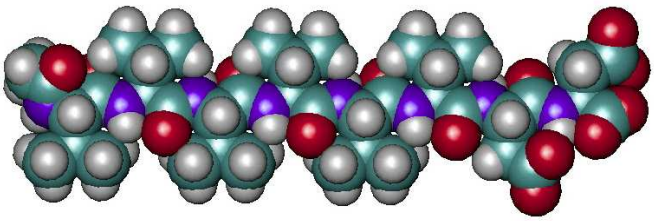
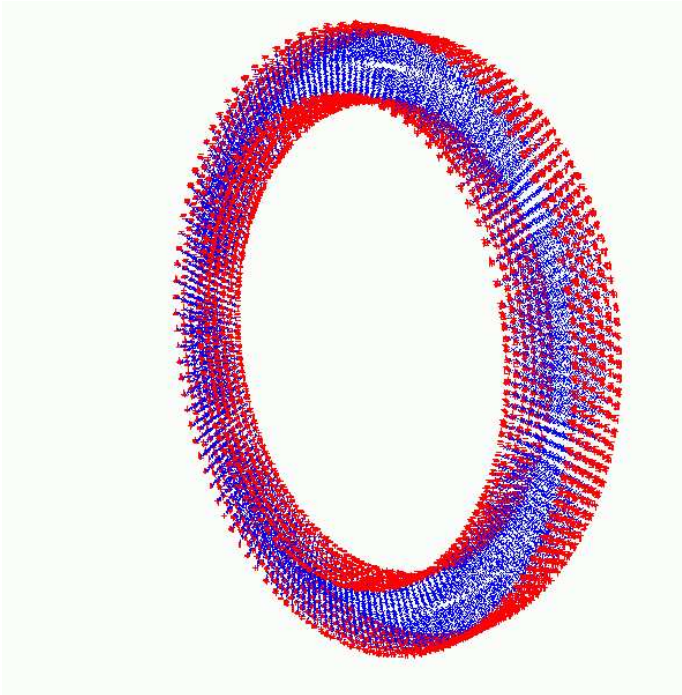
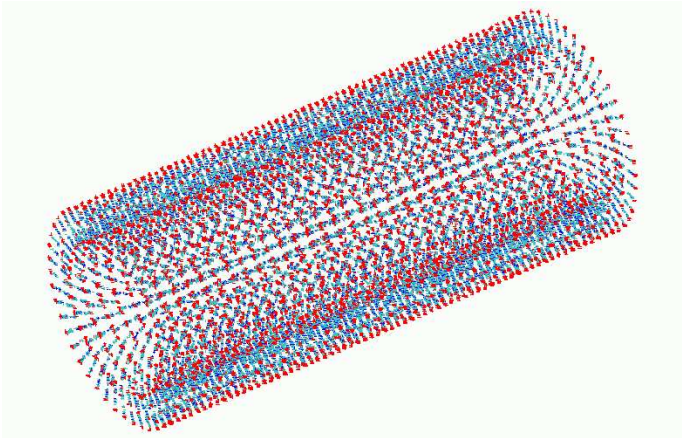
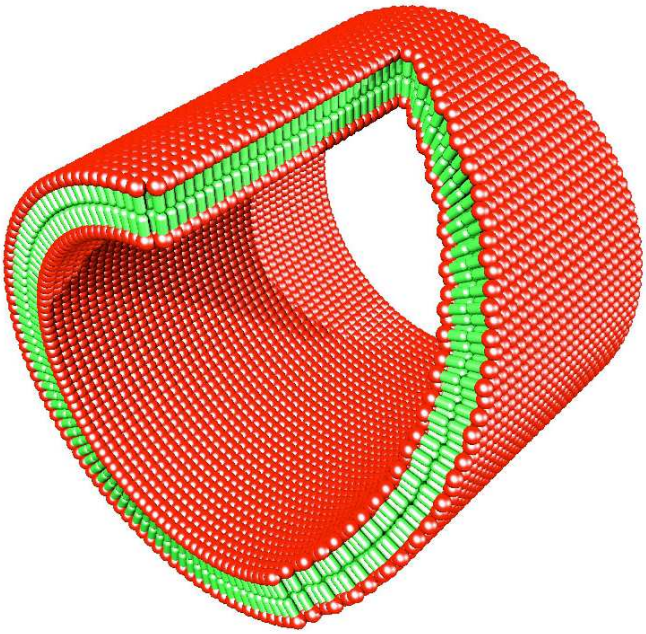


Fig: **peptide closed or open** nanotubes. **produced in the Laboratory of Shuguang Zhang at MIT.**







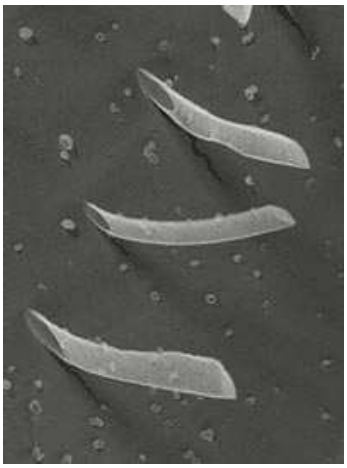
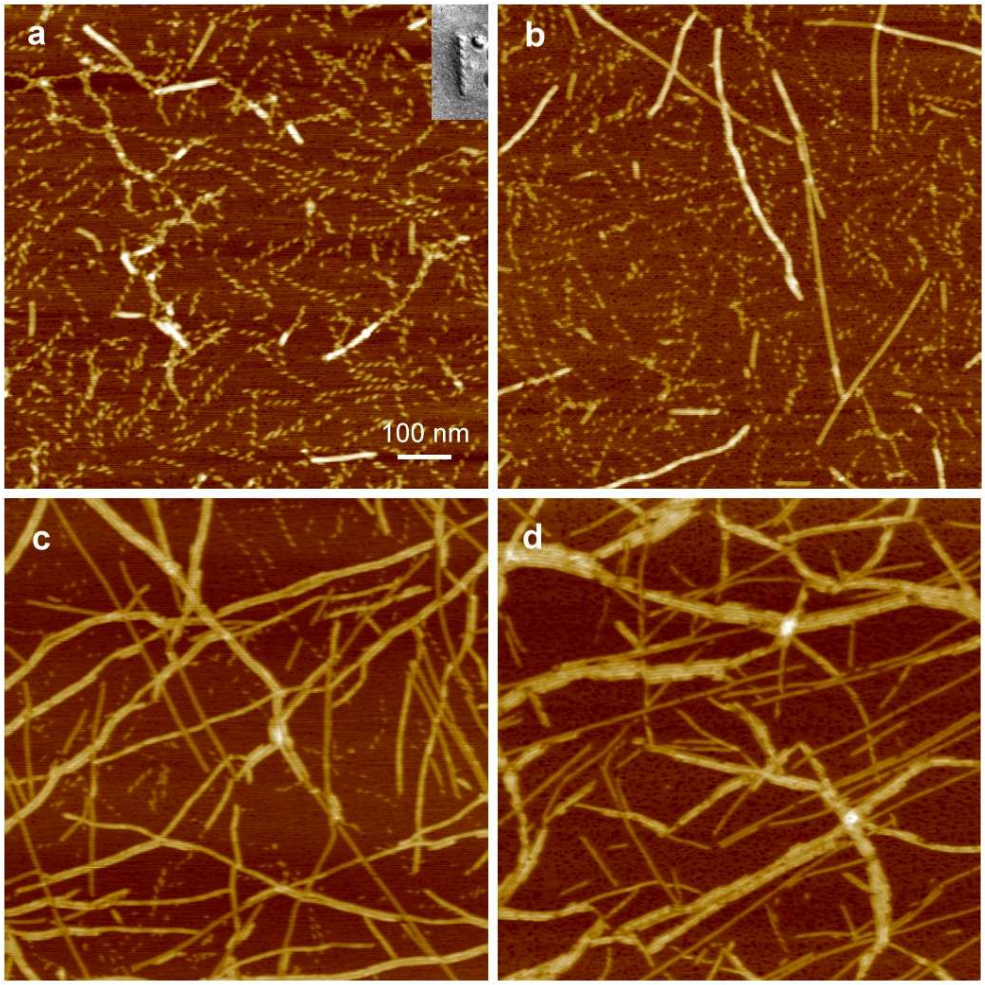


Fig: **peptide** vesicle.

Fig: **helical single and double** tapes.

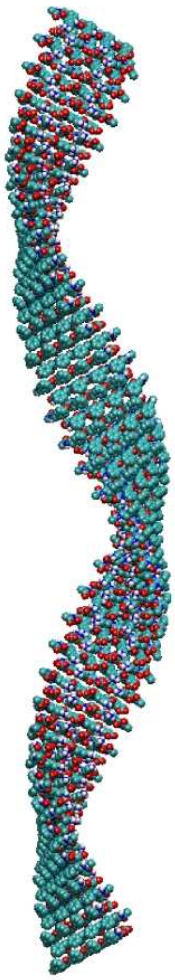
Marini_fig1



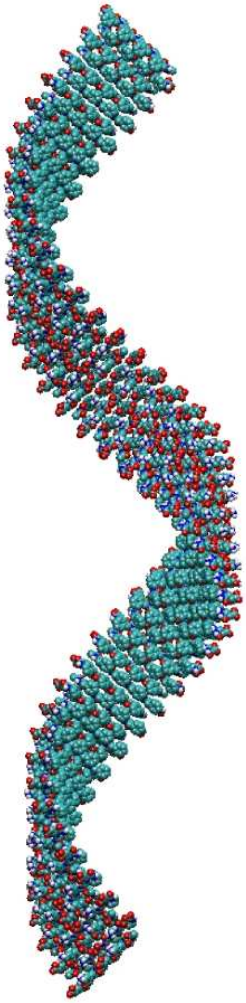
marini_fig3

a

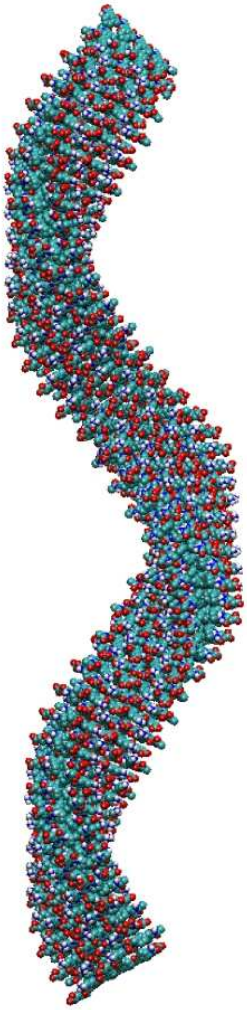
Inner helix



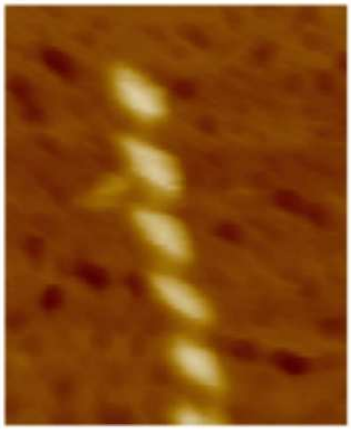
Outer helix



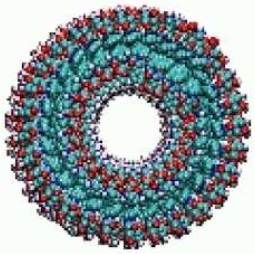
Double helix



AFM image



Axis view



b

Initial conditions

Fig: **rationally designed** DN1 sequence.

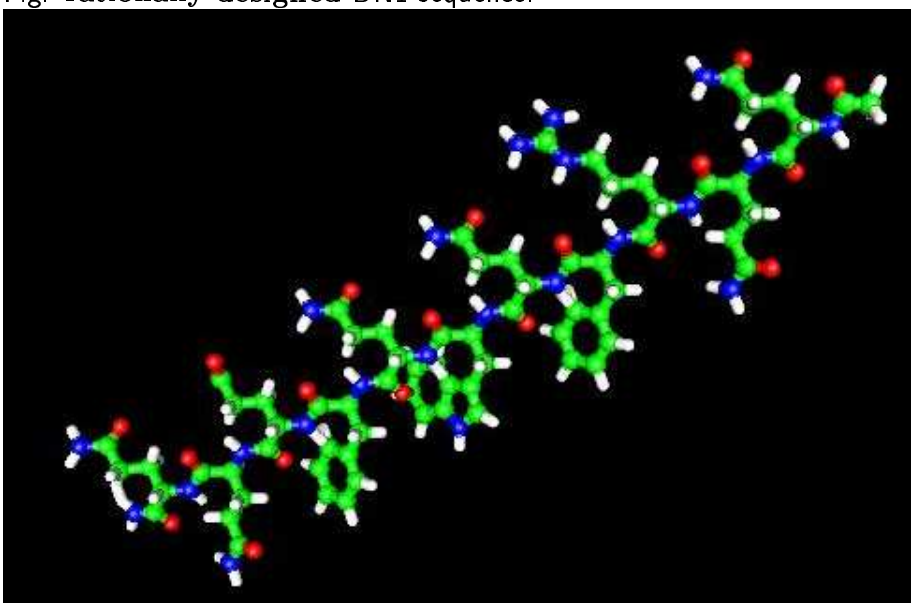


Fig: **concentration dependence and** hierarchy **of self- -assembled structures proposed by our collaborators in SOMS at University of Leeds.**

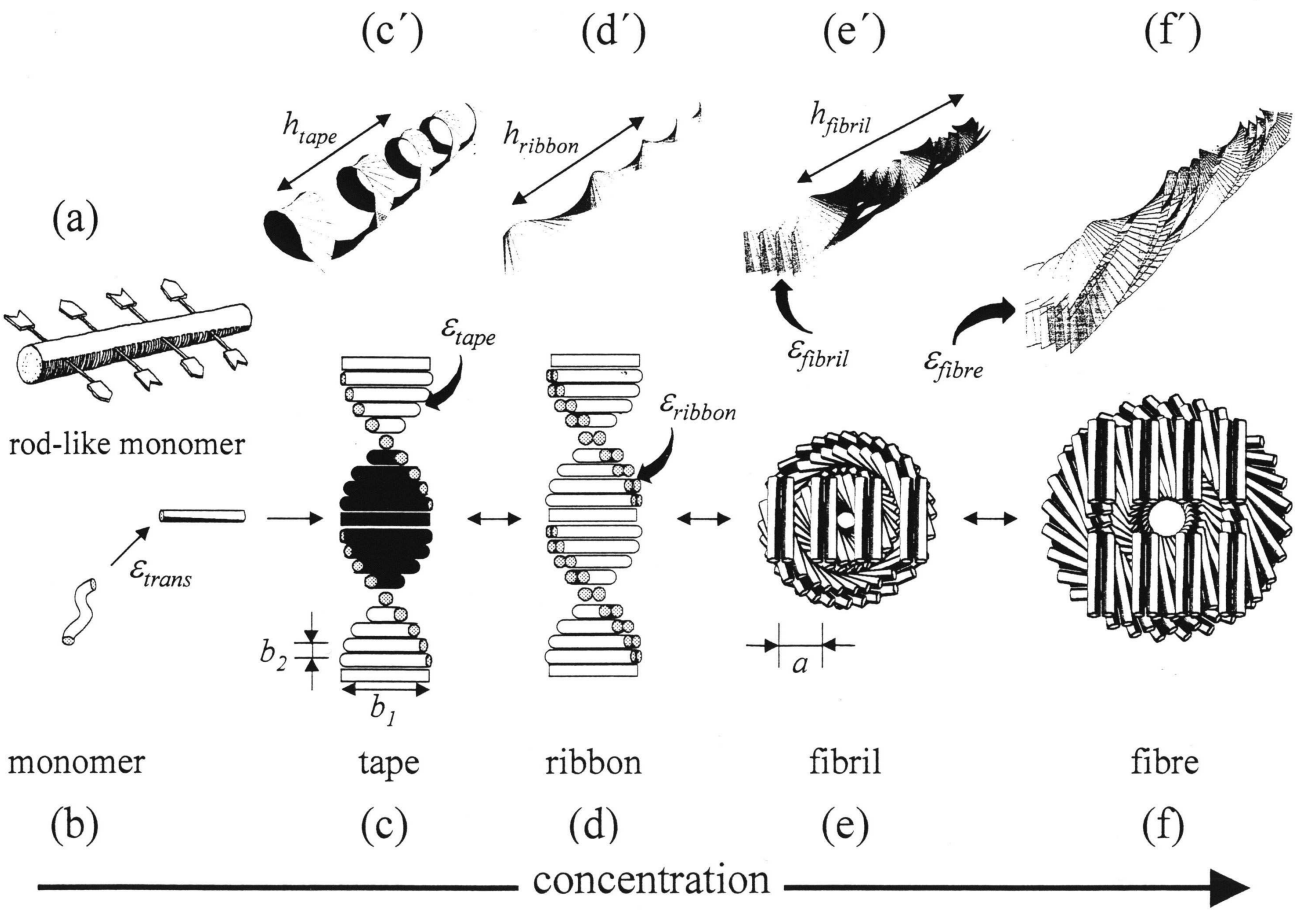


Fig: **Multiple** hydrogen bonding **stabilises such structures.**

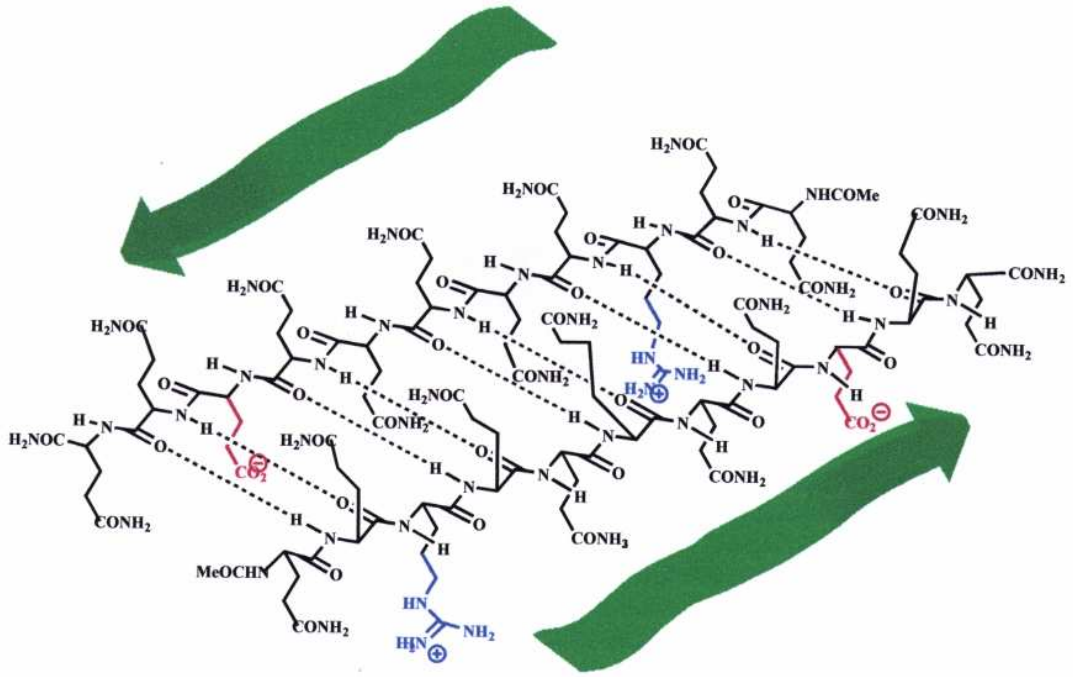


Figure 4.2.2: Molecular schematic edge-view model of pre-simulated P_{11}-1 tape. The completely polar nature of the P_{11}-1 peptide allows the tape aggregate to stably exist in water.

Fig: **shapes of self–assemblies for different sequences.**

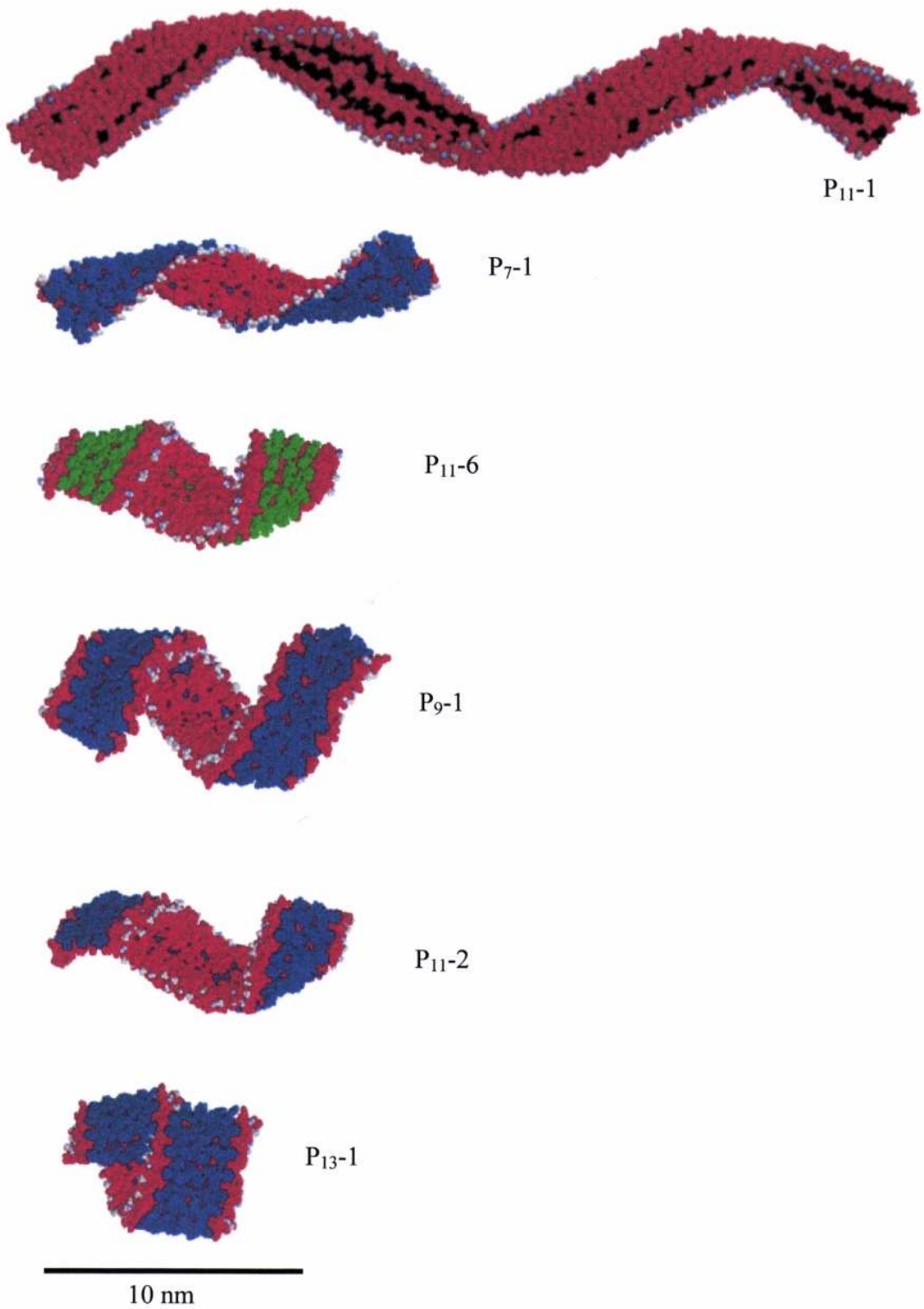


Figure 4.10.10: The increase in tape face chemical difference appears to increase the degree of cylindrical curvature, resulting in a shorter pitch.

Fig: **helical ribbon shape of** single tapes.

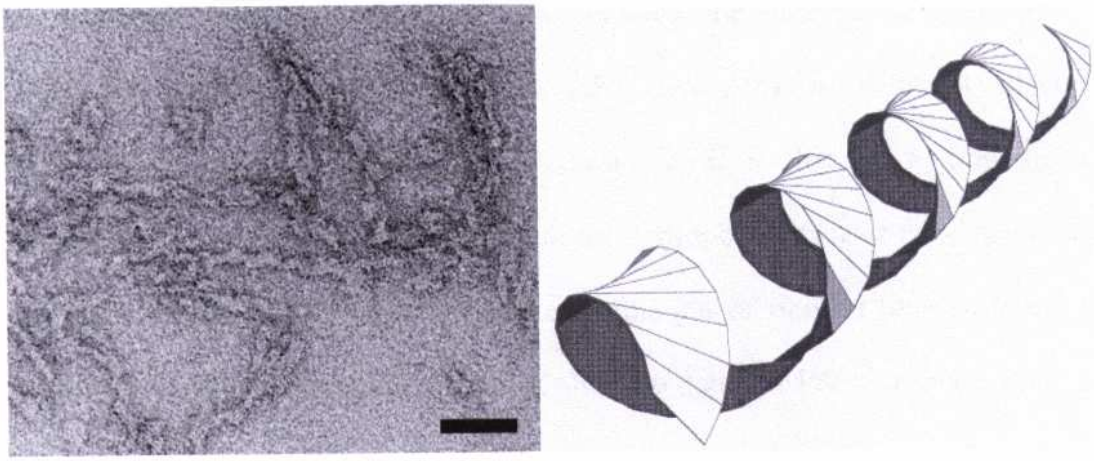


Figure 4.1.1 a: (Left) Helical P_{11}-1 tape in solution – scale bar = 50 nm (Bell 1999)

(Right) Predicted helical tape structure – h_{tape} is the helical pitch (Aggeli 2001)

Fig: **shape of** ribbons (double tapes).

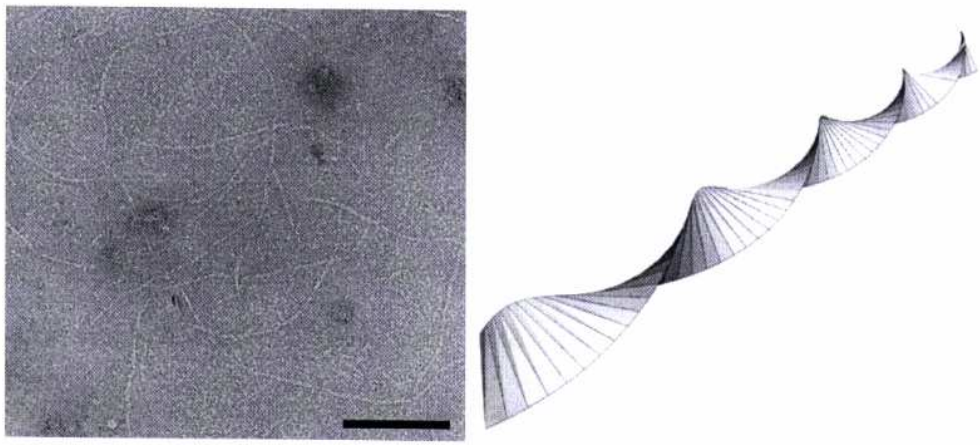


Figure 5.1.1: (Left) TEM image of self-assembled peptide ribbon in solution – scale bar = 100 nm (Carrick 2002).

(Right) theoretically predicted self-assembled peptide ribbon schematic (Aggeli 2001)

Fig: **electron micrograph of** peptide fibrils.

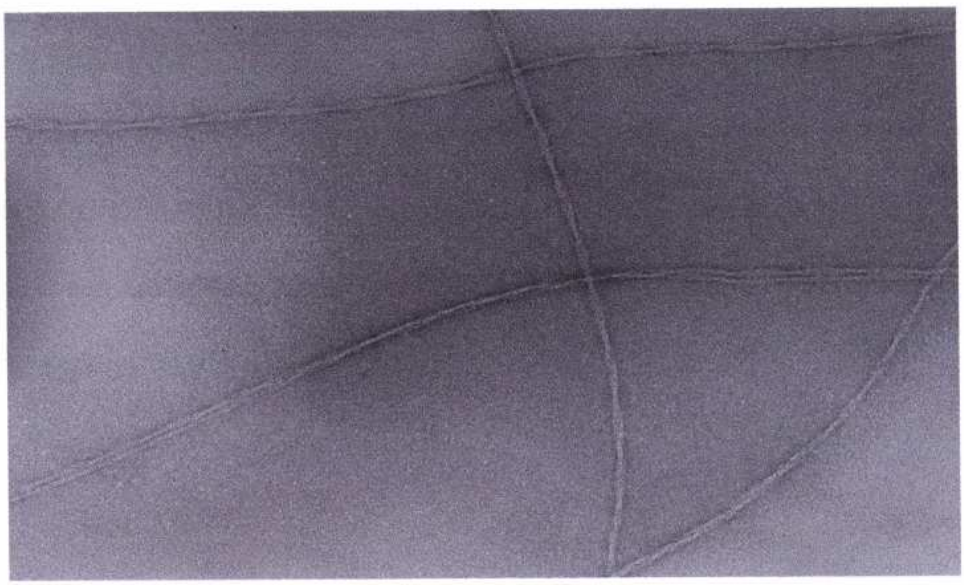
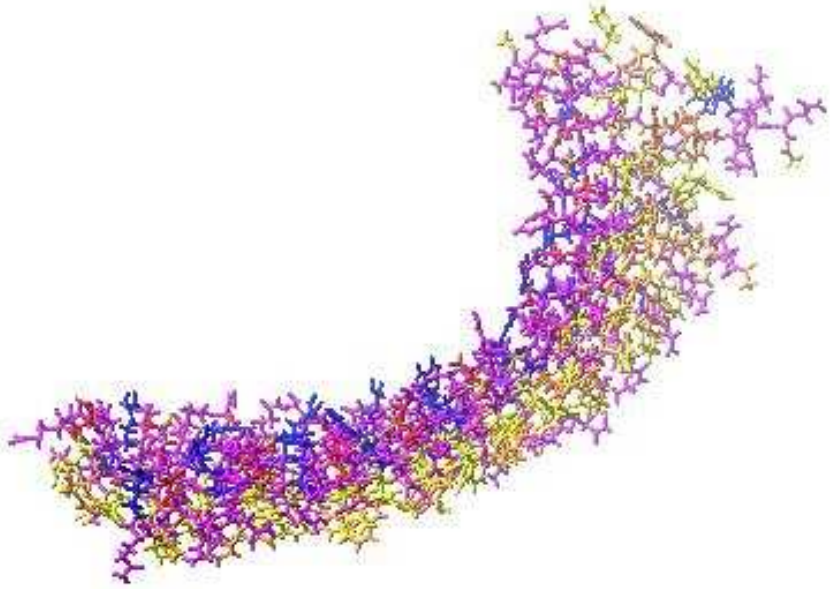
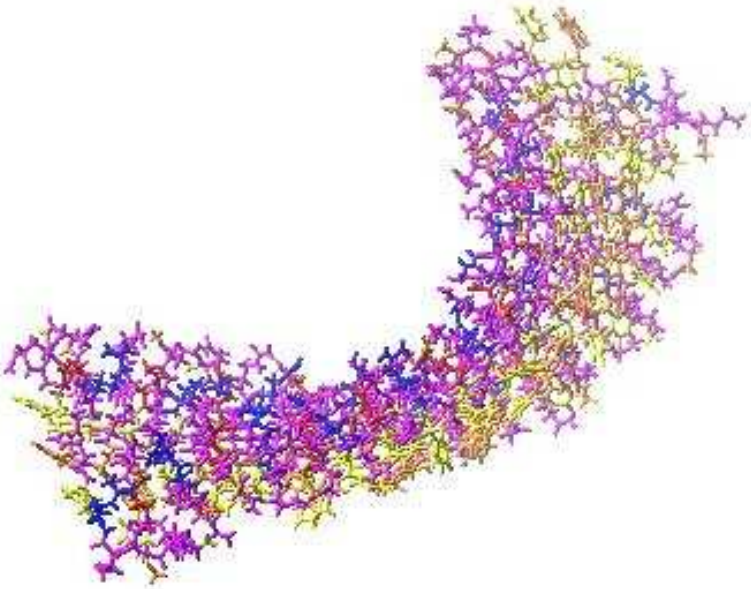
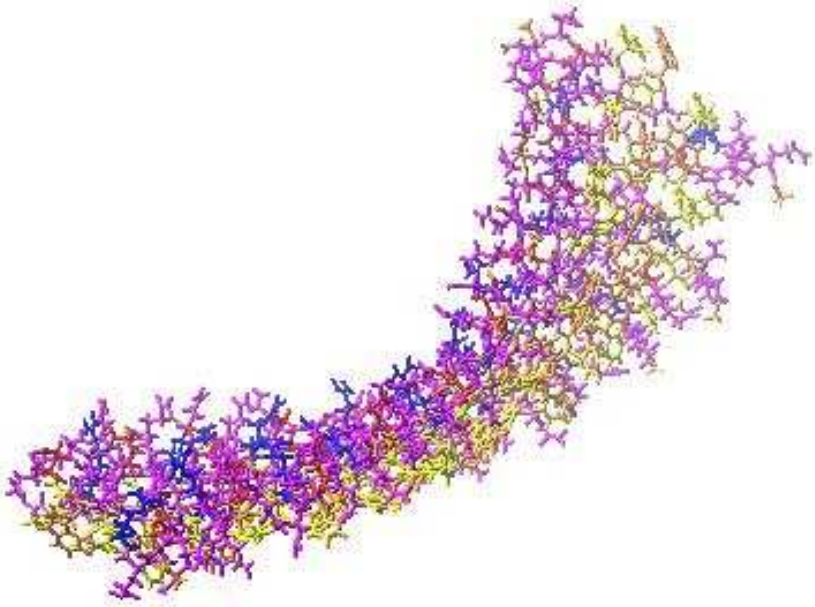


Figure 6.1.1 a: TEM image of P_{11}-2 fibril (Carrick 2002)

Fig: **animation from long–time computer simulation of the** tape motion **obtained in our Laboratory.**







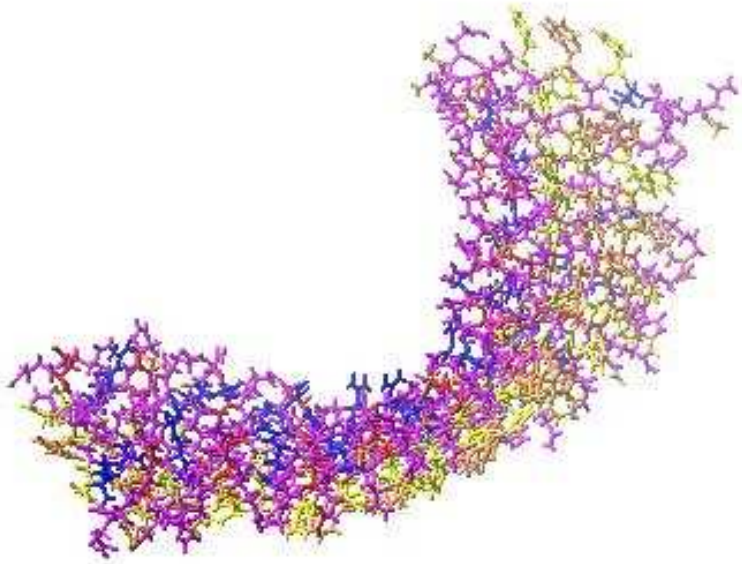
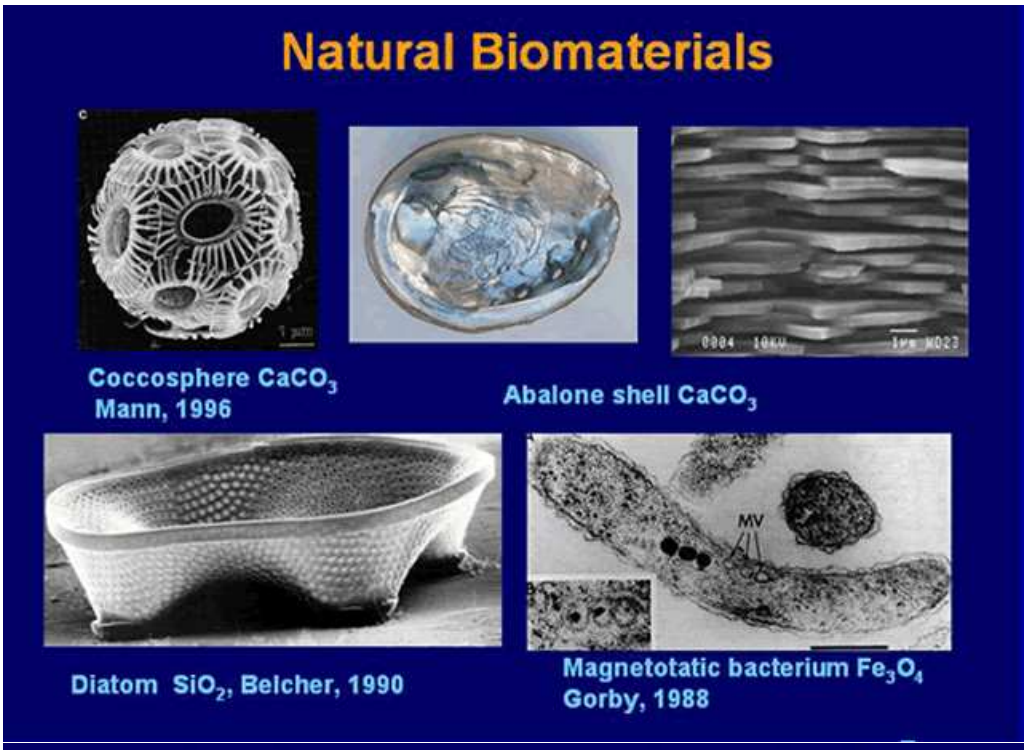
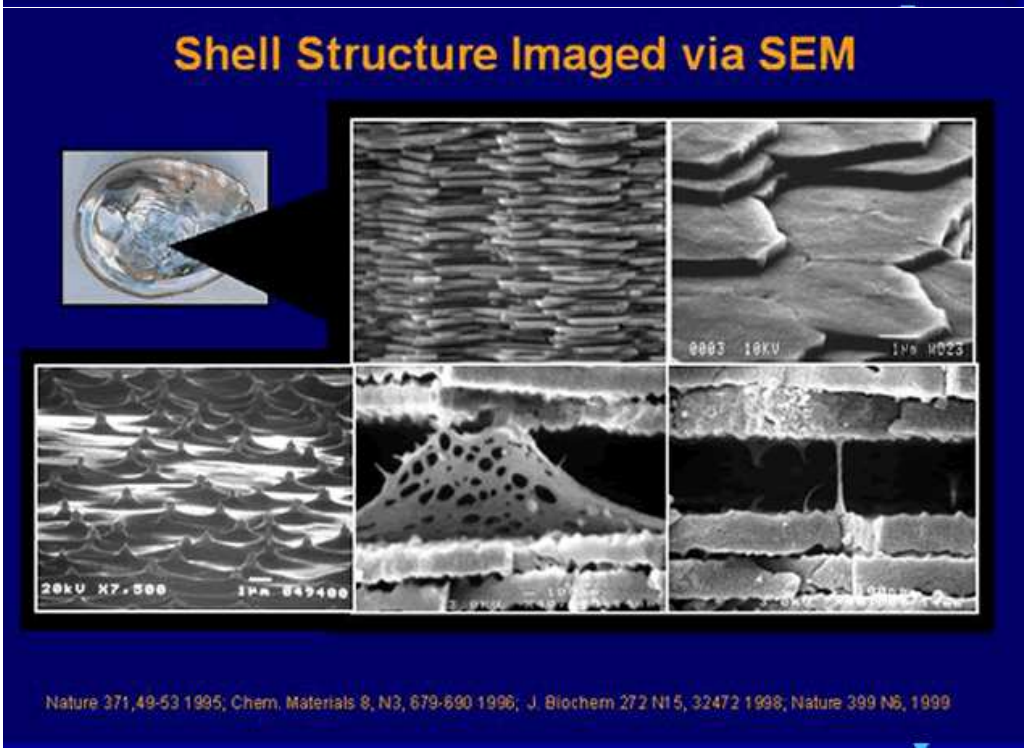
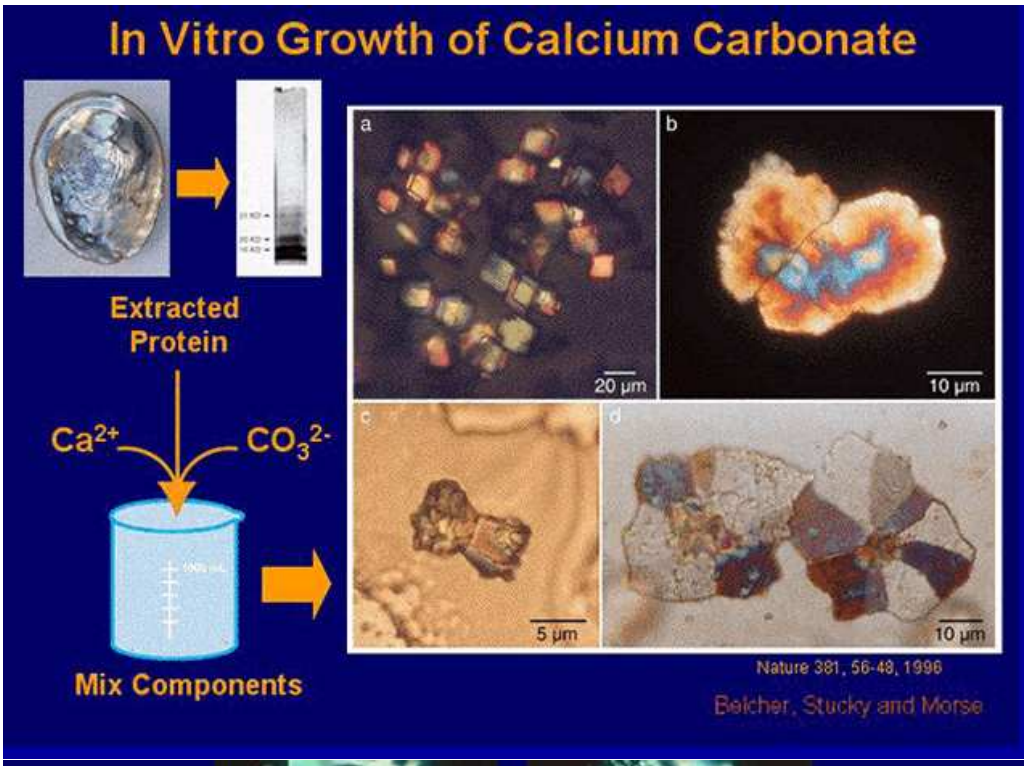
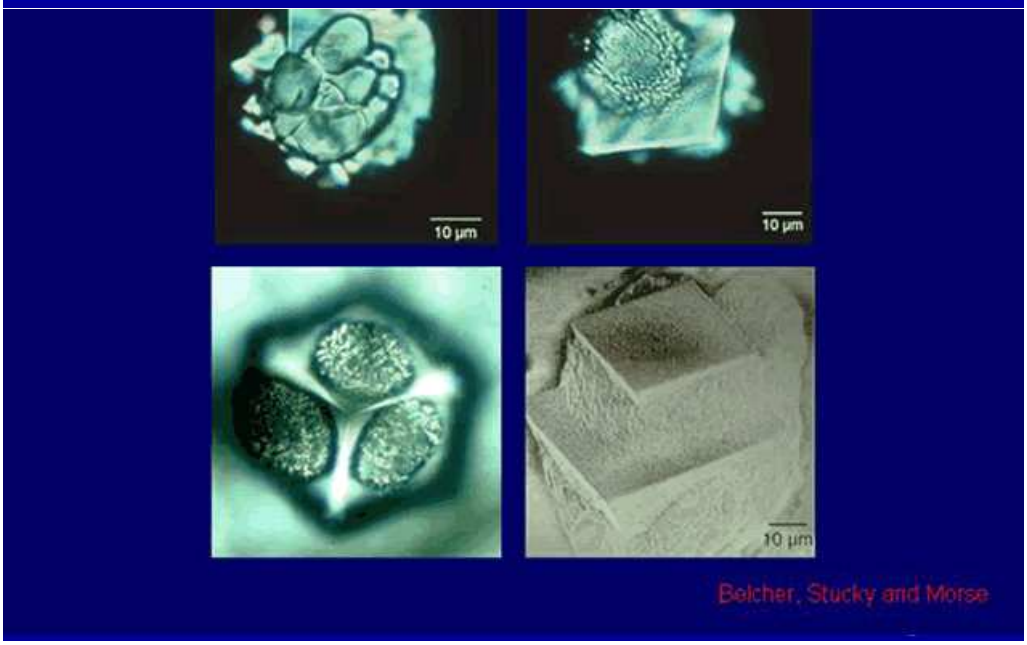


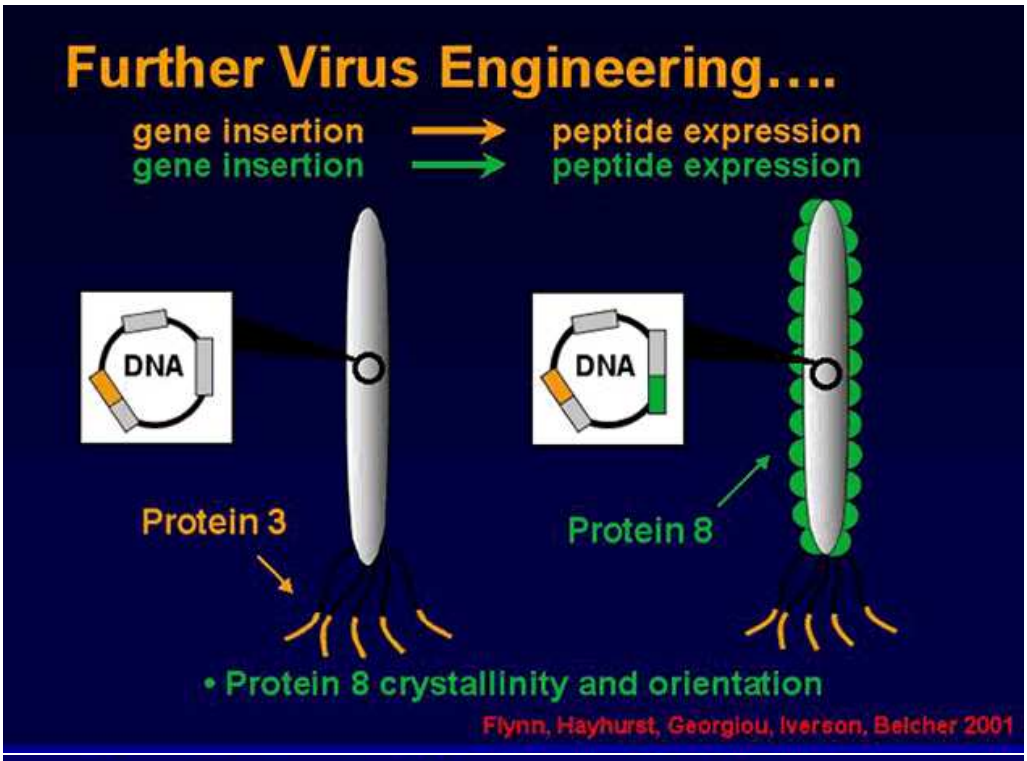
Fig: biomaterials (tissues, scaffolds, etc) **and** devices **produced in the Laboratory of Angie Belcher at MIT.**

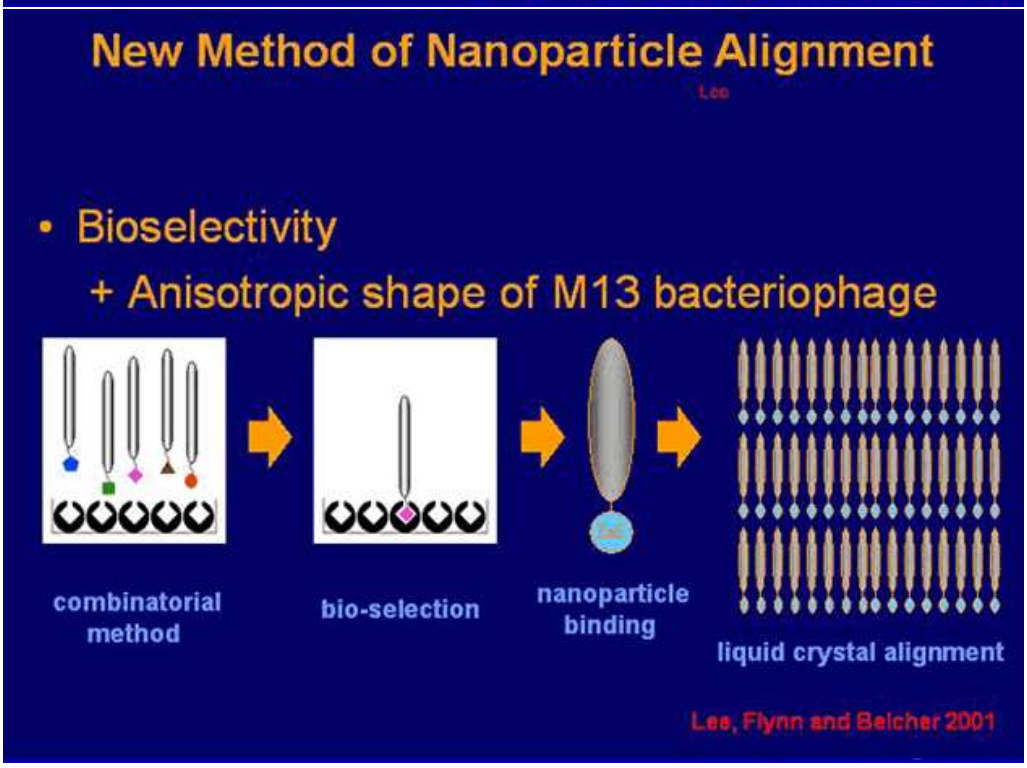

Natural Biomaterials
Coccosphere $CaCO_3$
Mann, 1996
Abalone shell $CaCO_3$
Diatom SiO_2, Belcher, 1990
MV
Magnetotatic bacterium Fe_3O_4
Gorby, 1988

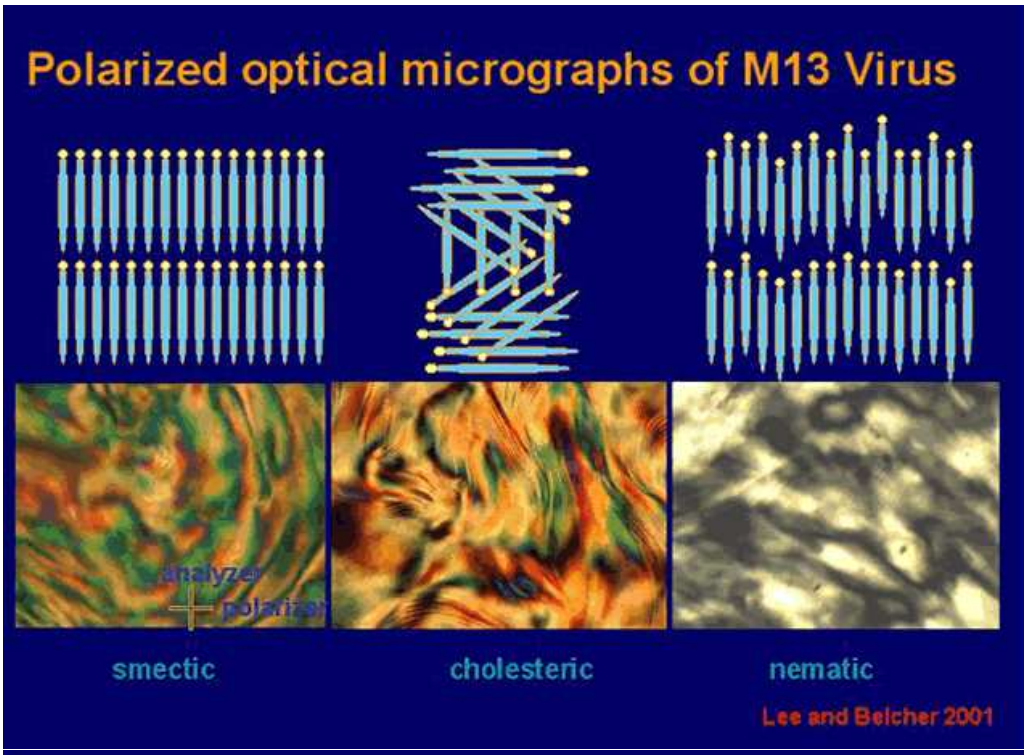

Shell Structure Imaged via SEM
Nature 371,49-53 1995; Chem. Materials 8, N3, 679-690 1996; J. Biochem 272 N15, 32472 1998; Nature 399 N6, 1999

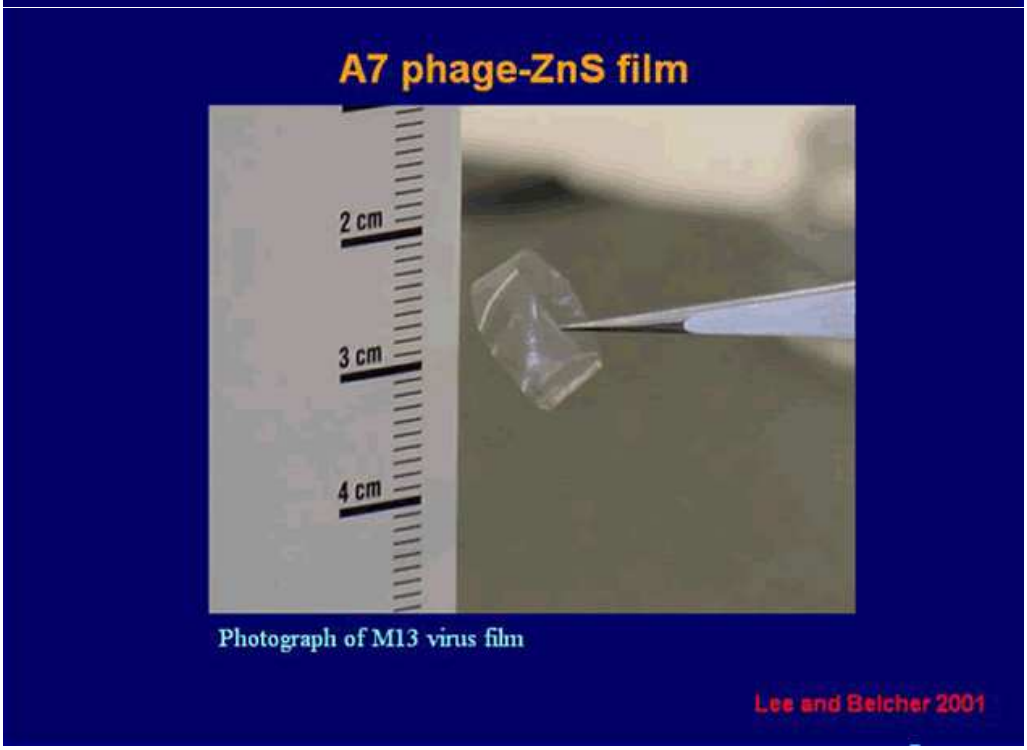

In Vitro Growth of Calcium Carbonate
Extracted Protein
Ca^{2+}
CO_3^{2-}
Mix Components
a
20 µm
b
10 µm
c
5 µm
d
10 µm
Nature 381, 56-48, 1996
Belcher, Stucky and Morse

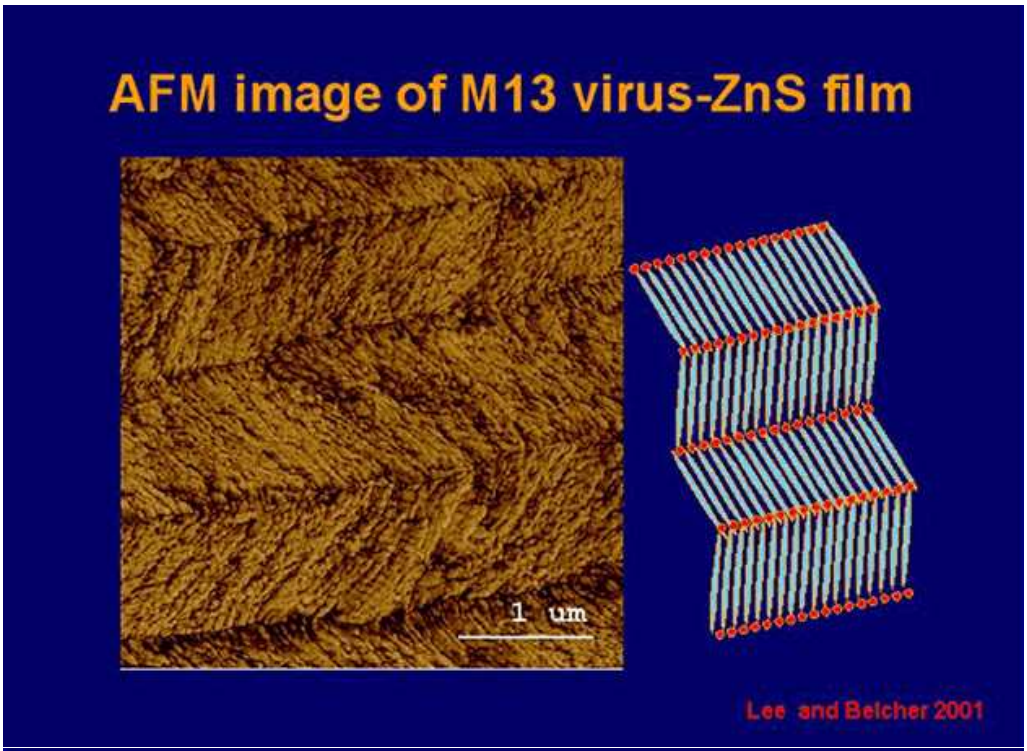

10 µm
10 µm
10 µm
Belcher, Stucky and Morse

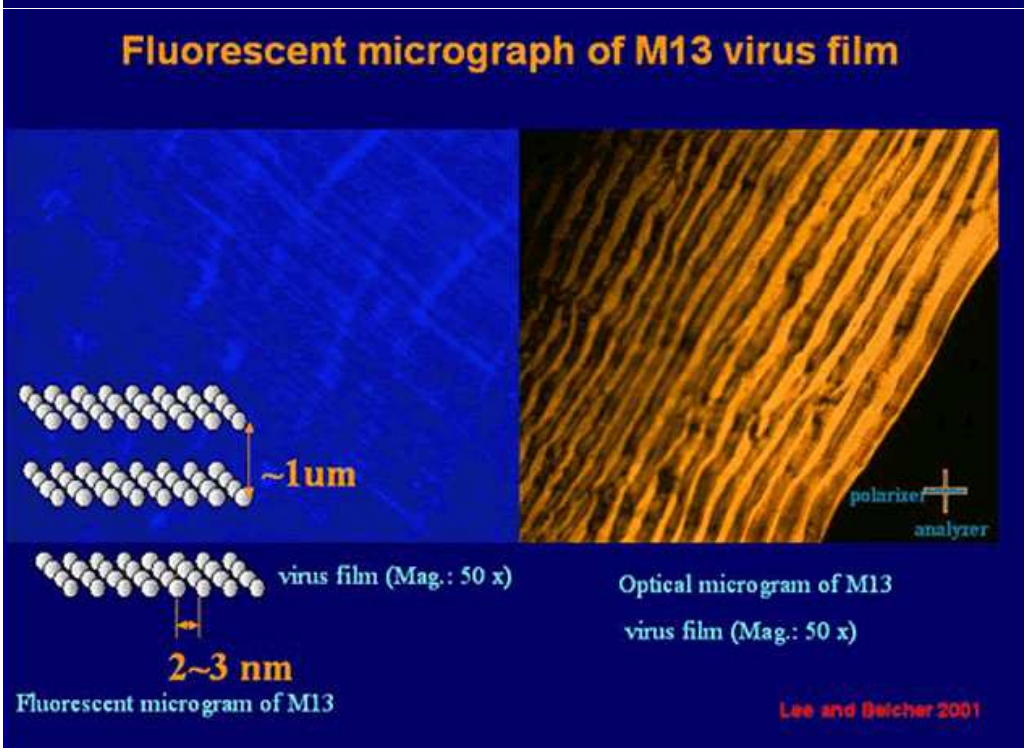

Further Virus Engineering….
gene insertion
peptide expression
gene insertion
peptide expression
DNA
DNA
Protein 3
Protein 8
• Protein 8 crystallinity and orientation
Flynn, Hayhurst, Georgiou, Iverson, Belcher 2001


New Method of Nanoparticle Alignment
• Bioselectivity
+ Anisotropic shape of M13 bacteriophage
combinatorial method
bio-selection
nanoparticle binding
liquid crystal alignment
Lee, Flynn and Belcher 2001


Polarized optical micrographs of M13 Virus
analyzer
polarizer
smectic
cholesteric
nematic
Lee and Belcher 2001


A7 phage-ZnS film
2 cm
3 cm
4 cm
Photograph of M13 virus film
Lee and Belcher 2001


AFM image of M13 virus-ZnS film
1 um
Lee and Belcher 2001


Fluorescent micrograph of M13 virus film
~1um
virus film (Mag.: 50 x)
2~3 nm
Fluorescent microgram of M13
polarizer
analyzer
Optical microgram of M13
virus film (Mag.: 50 x)
Lee and Belcher 2001