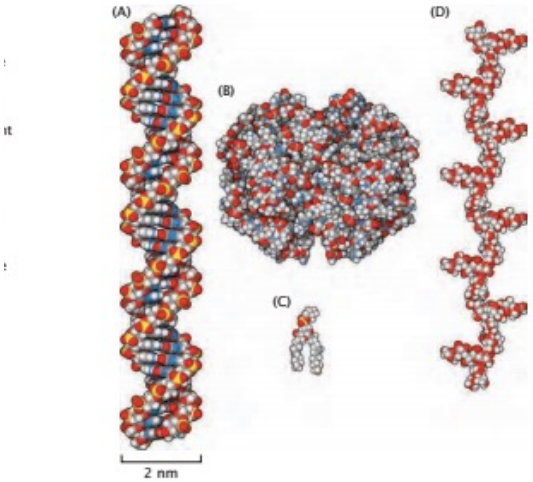
Corso di Biofisica Sperimentale

Modulo 1. Loredana Casalis

loredana.casalis@elettra.eu

The course will be largely devoted to exploring the special properties of biological macromolecules and the ways that they are used by cells, with the large-scale goal of understanding how the special properties of life may emerge from fundamental physical and chemical interactions.



(A) Atomic structure of a small fragment of the **nucleic acid** DNA in the B form
(B) atomic structure of the oxygen-carrying **protein** hemoglobin (PDB 1hho)
(C) phosphatidylcholine **lipid** molecule from a cell membrane
(D) branched complex **carbohydrate**

What is biophysics?

Biophysics is the field that applies the theories and methods of physics to understand how biological systems work, i.e. the mechanics of:

- how the molecules of life are made
- how different parts of a cell move and function
- how complex systems in our bodies—the brain, circulation, immune system, and others— work.

(ref. Biophysical Society)

What is biophysics?

Biophysics is the field that applies the theories and methods of physics to understand how biological systems work

Scientists from math, chemistry, physics, engineering, pharmacology, biology, biotechnology and materials sciences explore and develop new tools to understand how biology—all life—works. They design cutting-edge technologies and develop methods to overcome disease, but also to eradicate global hunger and produce renewable energy sources.

What do Biophysicists do

Therefore biophysicists work on:

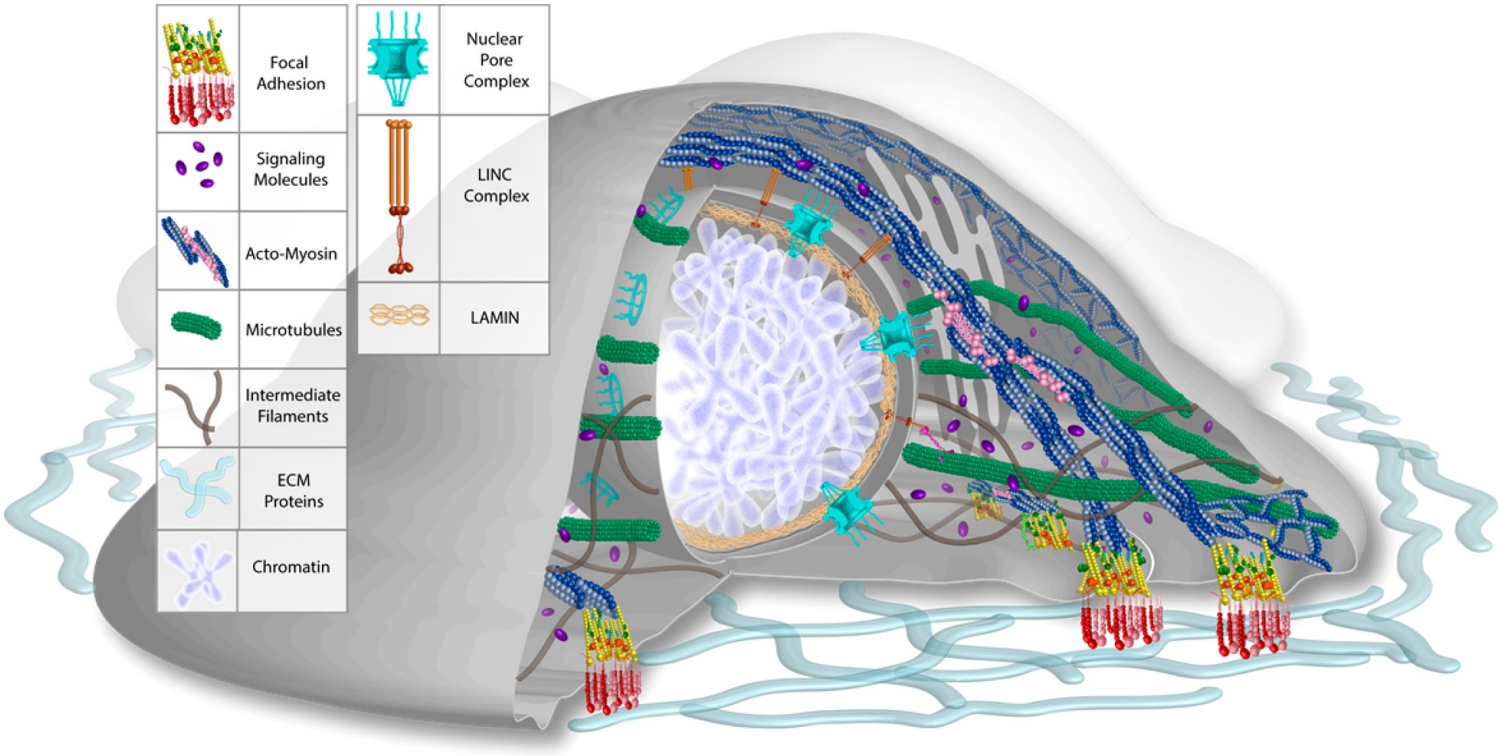
- Data Analysis and Structure (DNA sequencing and correlation with diseases, protein structure, analysis of huge quantity of data)
- Computer Modelling (see and manipulate the shapes and structures of proteins, viruses, and other complex molecules to develop new drug targets, or understand how proteins mutate and cause tumours to grow)
- Molecules in Motion, Cell-Cell Interactions (understand how molecules move inside the cells, how cells interacts with other cells and extra-cellular environment adapting theit structure/function)
- Bioengineering, Nanotechnologies, Biomaterials (biomechanics applied to understanding of diseases; design of functional nanomaterials for drug delivery and prosthetic applications)

What do Biophysicists do

Therefore biophysicists work on:

- Data Analysis and Structure (DNA sequencing and correlation with diseases, protein structure, analysis of huge quantity of data)
- Computer Modelling (see and manipulate the shapes and structures of proteins, viruses, and other complex molecules to develop new drug targets, or understand how proteins mutate and cause tumours to grow)
- Molecules in Motion, Cell-Cell Interactions **(understand how molecules move inside the cells, how cells interacts with other cells and extra-cellular environment adapting their structure/function)**
- Bioengineering, Nanotechnologies, Biomaterials (biomechanics applied to understanding of diseases; design of functional nanomaterials for drug delivery and prosthetic applications)

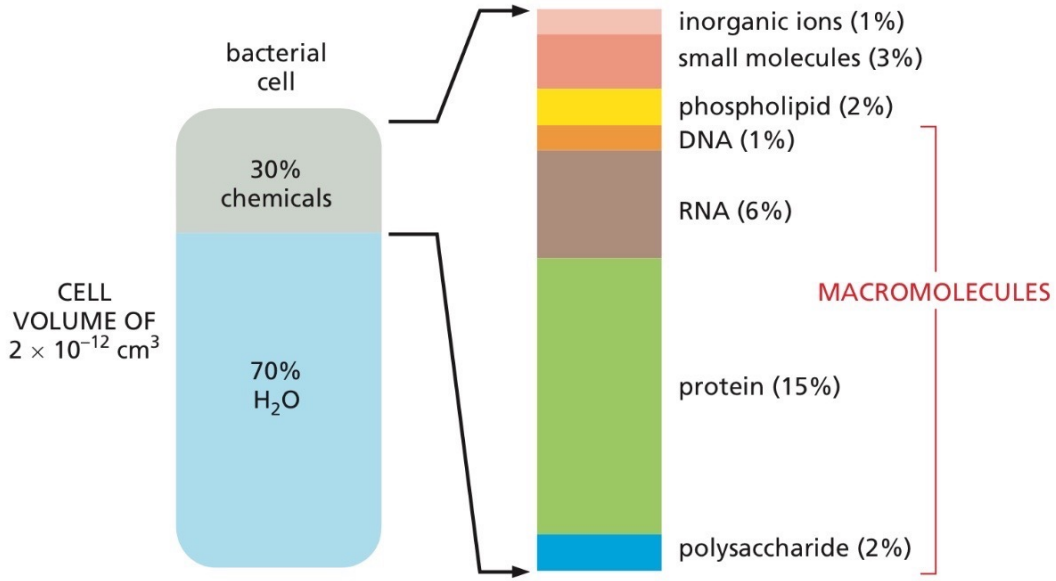
Example: mechanotransduction by cells



Cells respond to extracellular matrix (ECM) cues generating and transducing mechanical forces into biochemical signals and genomic pathways which affect cell properties.
Such forces define tissue architecture and drive specific cell differentiation programs. In adults perturbation of ECM (stiffness, mutations) cause pathologies in different organs, including ageing and malignant progression.

Inside the cell: macromolecules

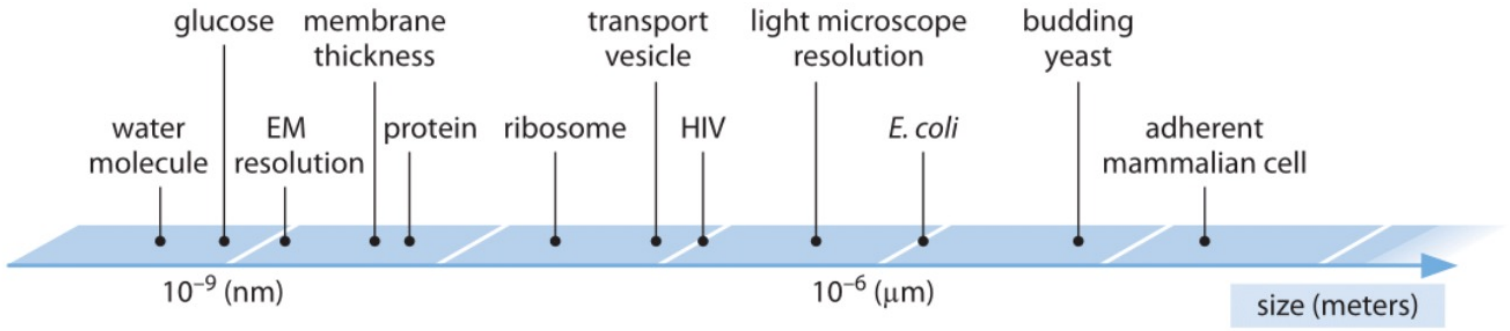
Many of the molecules that make up living organisms tend to be relatively large and structurally complex, and are hence called macromolecules. Living organisms also contain a large number of small, simple molecules that are critical to their function, ranging from water and metal ions to sugars such as glucose.



amount of total protein in some mammalian cells : one nanogram

Inside the cell: macromolecules

Many of the molecules that make up living organisms tend to be relatively large and structurally complex, and are hence called macromolecules. Living organisms also contain a large number of small, simple molecules that are critical to their function, ranging from water and metal ions to sugars such as glucose.



Biology spans 15 order of magnitude in length scale!

Central Dogma

CENTRAL DOGMA OF STRUCTURAL BIOLOGY

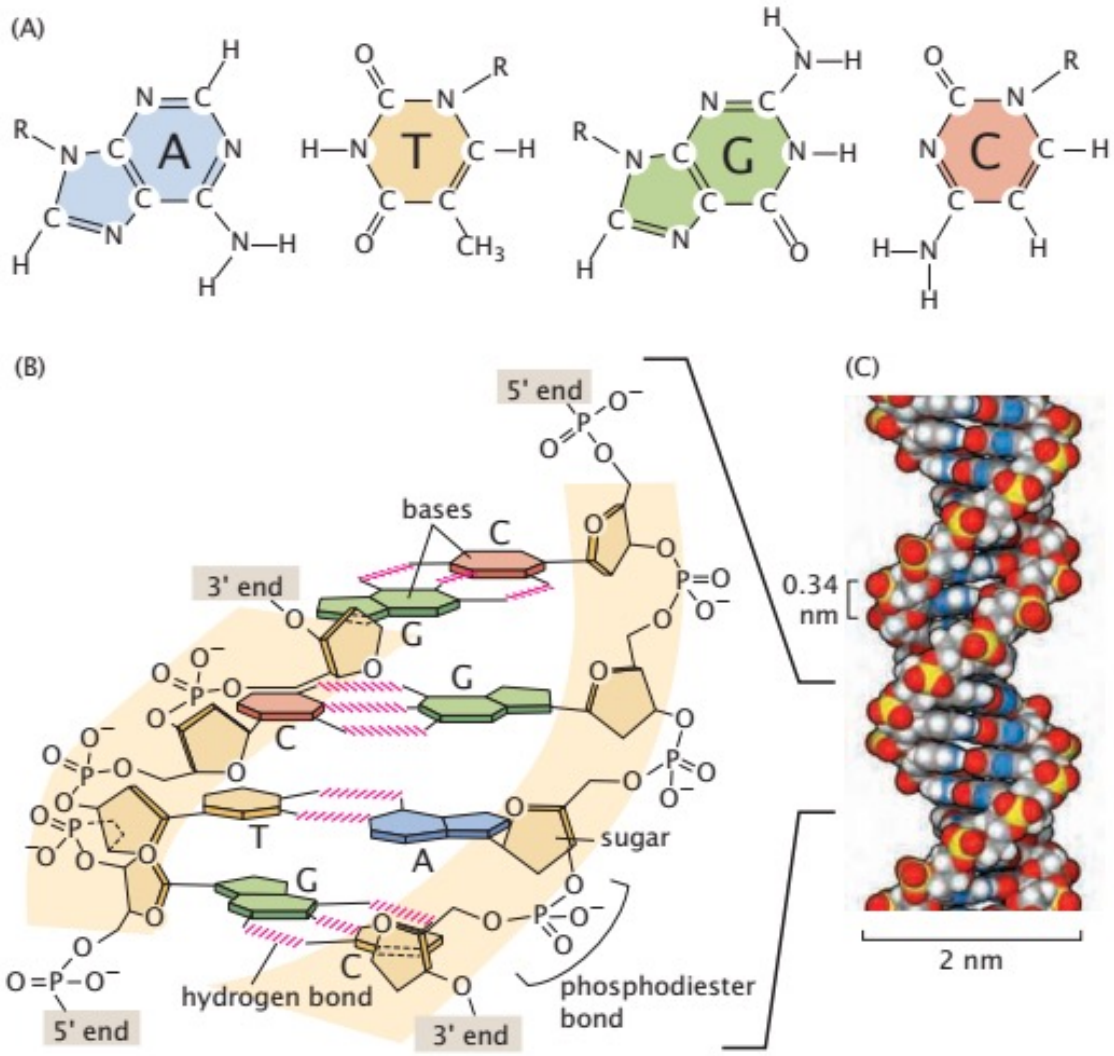
DNA ------- PROTEIN ------FUNCTION

Also the big variety of sizes and structures in the biologiocal world, responds to the function! The central dogma can be applied to other length scales.

Biophysics: underline macromolecules/cells structure/function/structural transition from simple principles. Explain complex processes **macromolecular structure; cell adaptation; fluctuactions.**

Make Models: simplify as much as possible, never more!

DNA: alphabet with 4 letters
Proteins: alphabet with 20 letters



How can just four nucleotide bases be translated into protein sequences containing 20 different amino acids?

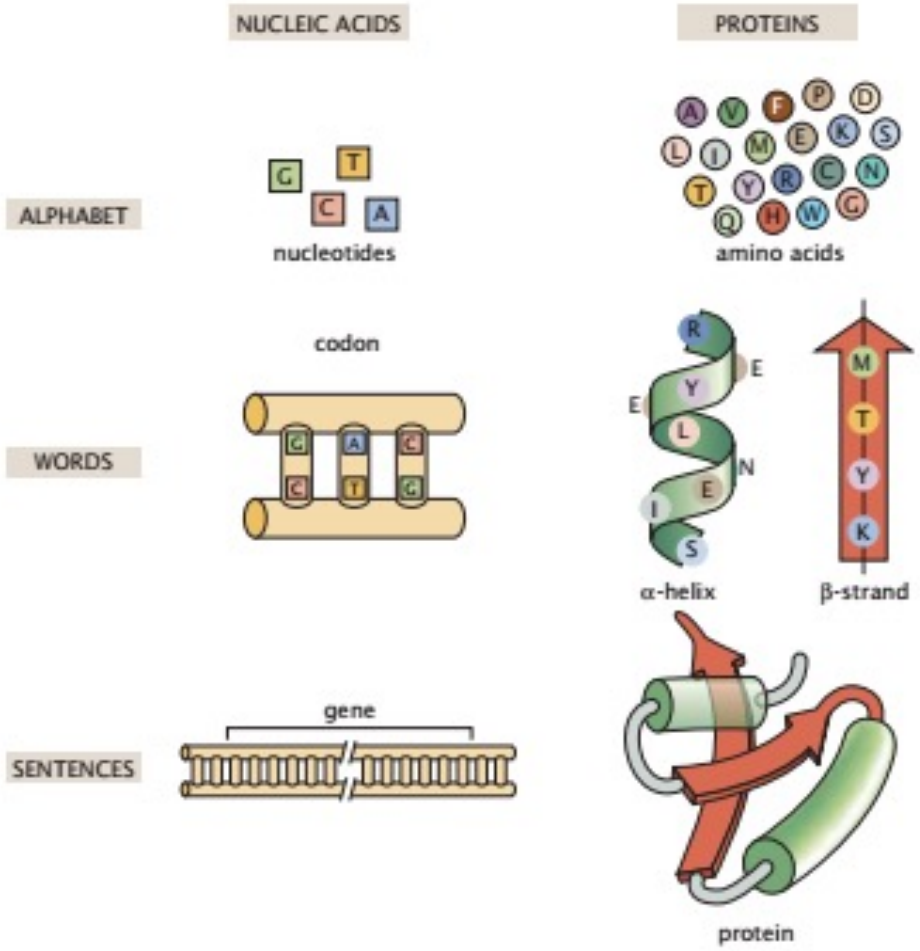
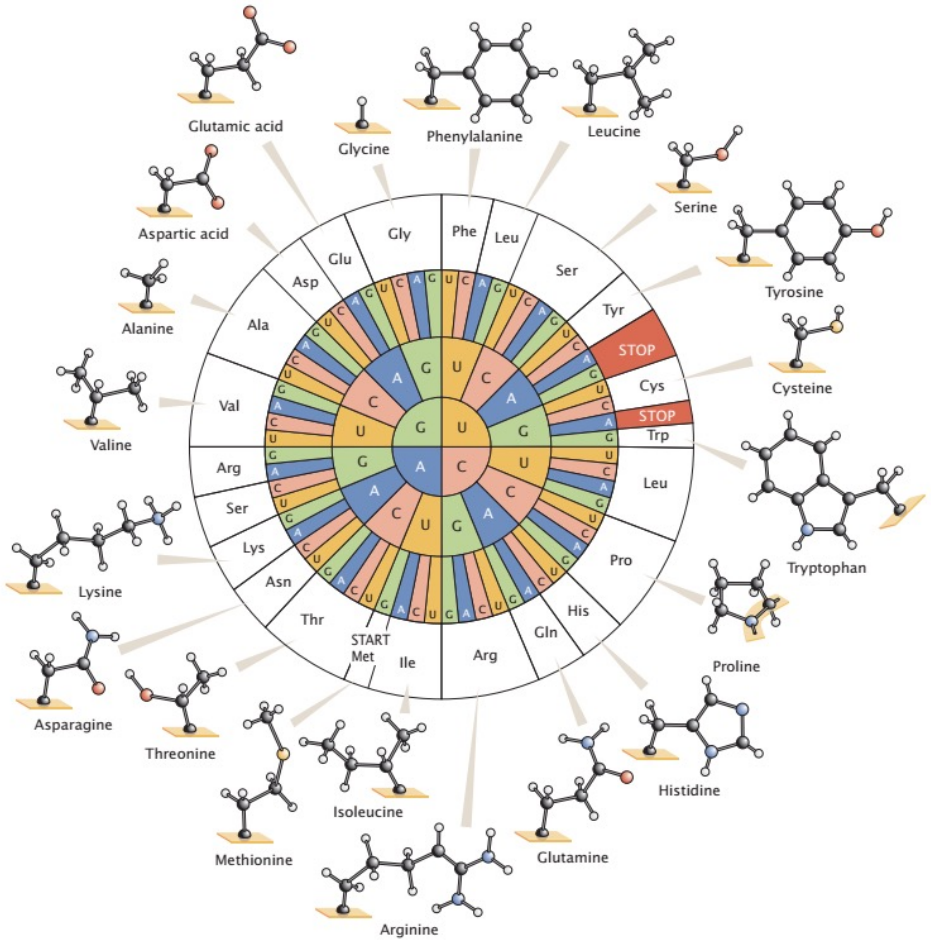


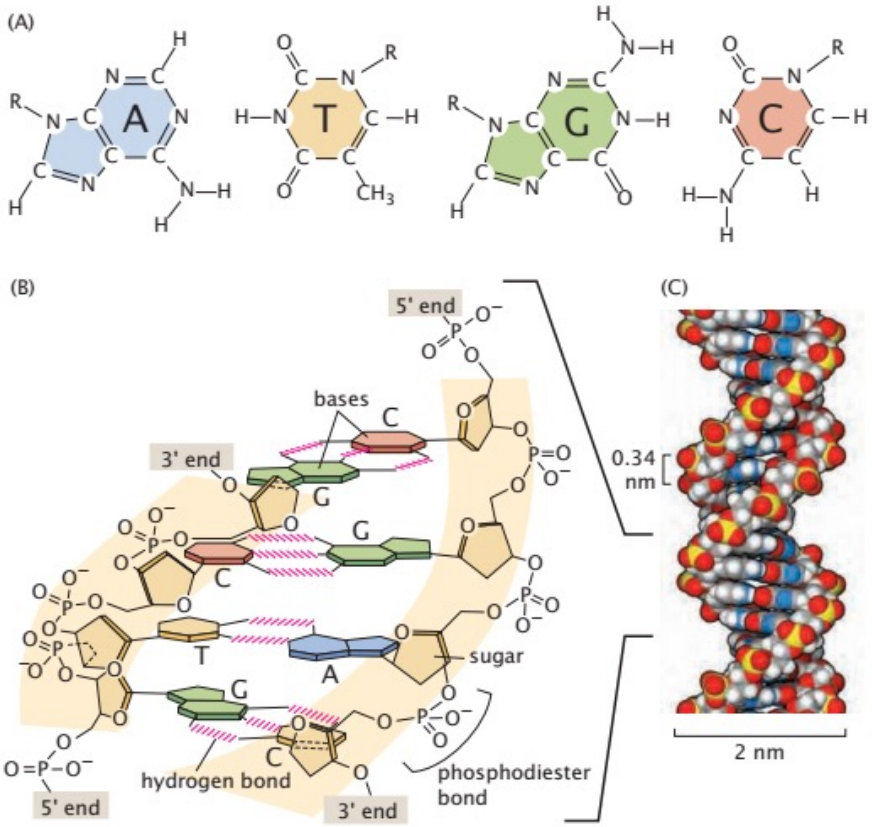
Figure 1.2: Illustration of Crick's "two great polymer languages." The left-hand column shows how nucleic acids can be thought of in terms of letters (nucleotides), words (codons), and sentences (genes). The right-hand column illustrates a similar idea for proteins, where the letters correspond to amino acids, the words to elements of secondary structure such as α helices and β strands, and the sentences to fully folded functional proteins.

The sequences associated with nucleic acids and proteins are linked mechanistically (Genetic Code) through the ribosome which takes nucleic acid sequences (in the form of messenger RNA (mRNA)) and converts them into amino acid sequences (in the form of proteins).

How can just four nucleotide bases be translated into protein sequences containing 20 different amino acids?



The sequences associated with nucleic acids and proteins are linked mechanistically (Genetic Code) through the ribosome which takes nucleic acid sequences (in the form of messenger RNA (mRNA)) and converts them into amino acid sequences (in the form of proteins).



Biological Stuff Can Be Idealized Using Many Different Physical Models: DNA

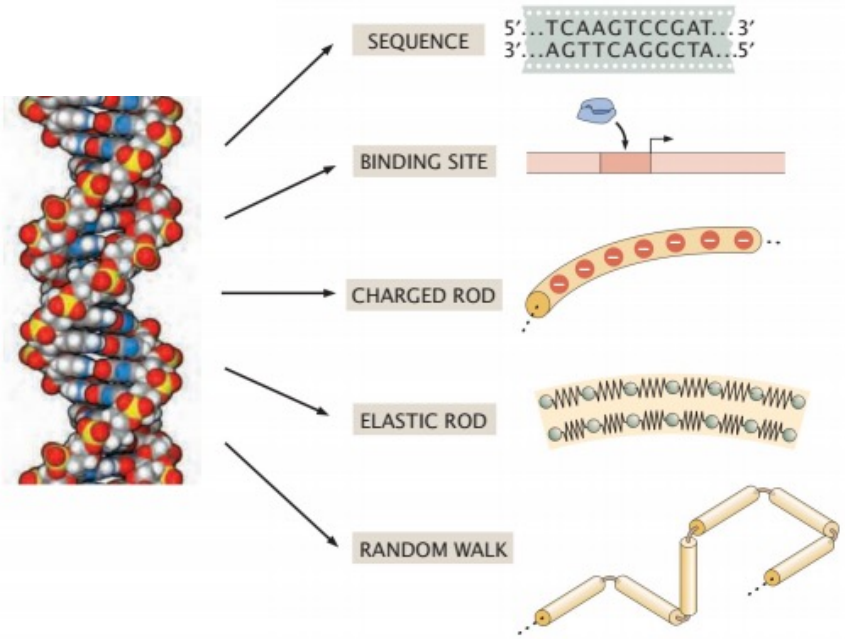


Figure 1.5: Idealizations of DNA. DNA can be thought of as a sequence of base pairs, as a series of binding sites, as a charged rod, as an elastic rod, or as a freely jointed polymer arranged in a random walk, depending upon the problem of interest.

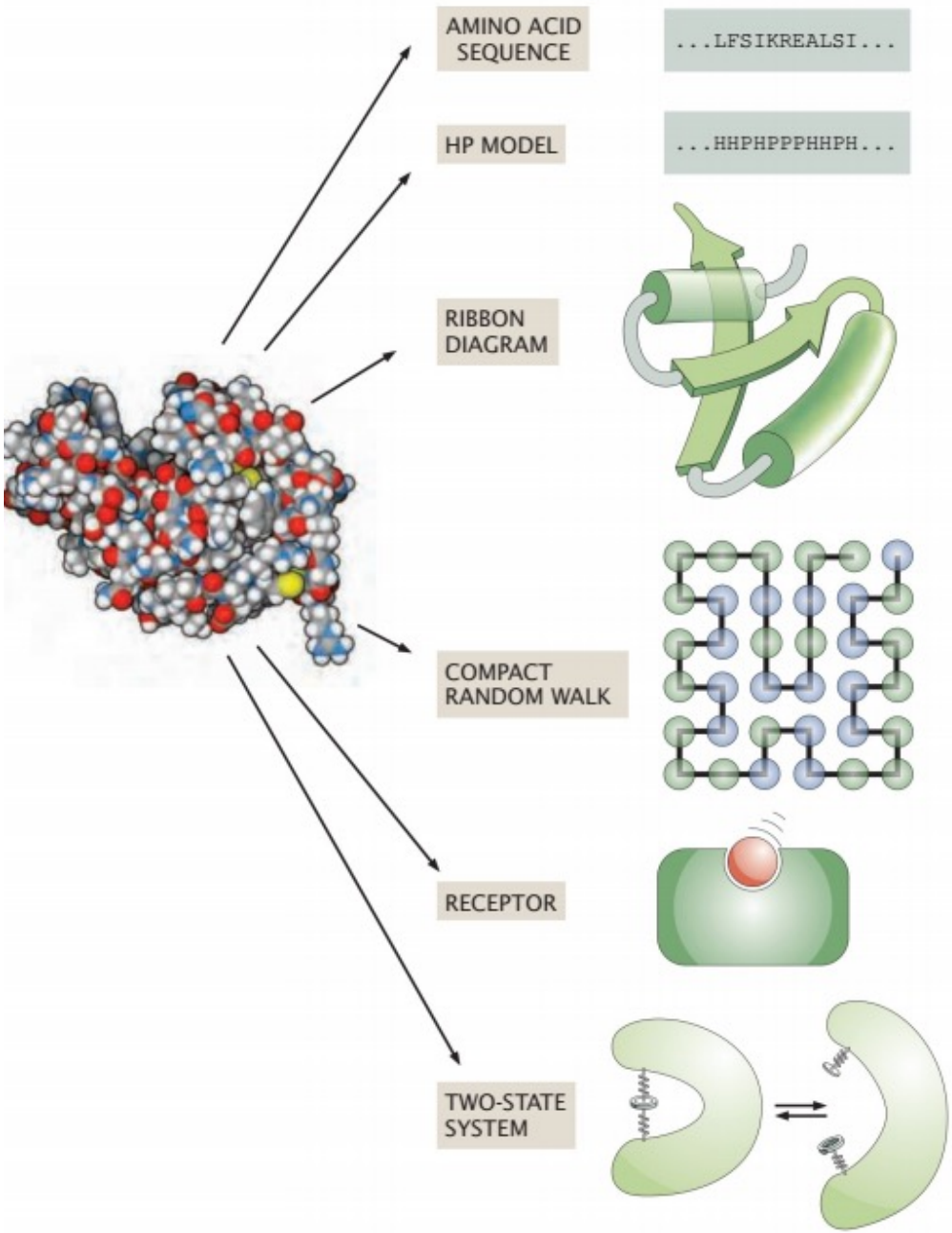
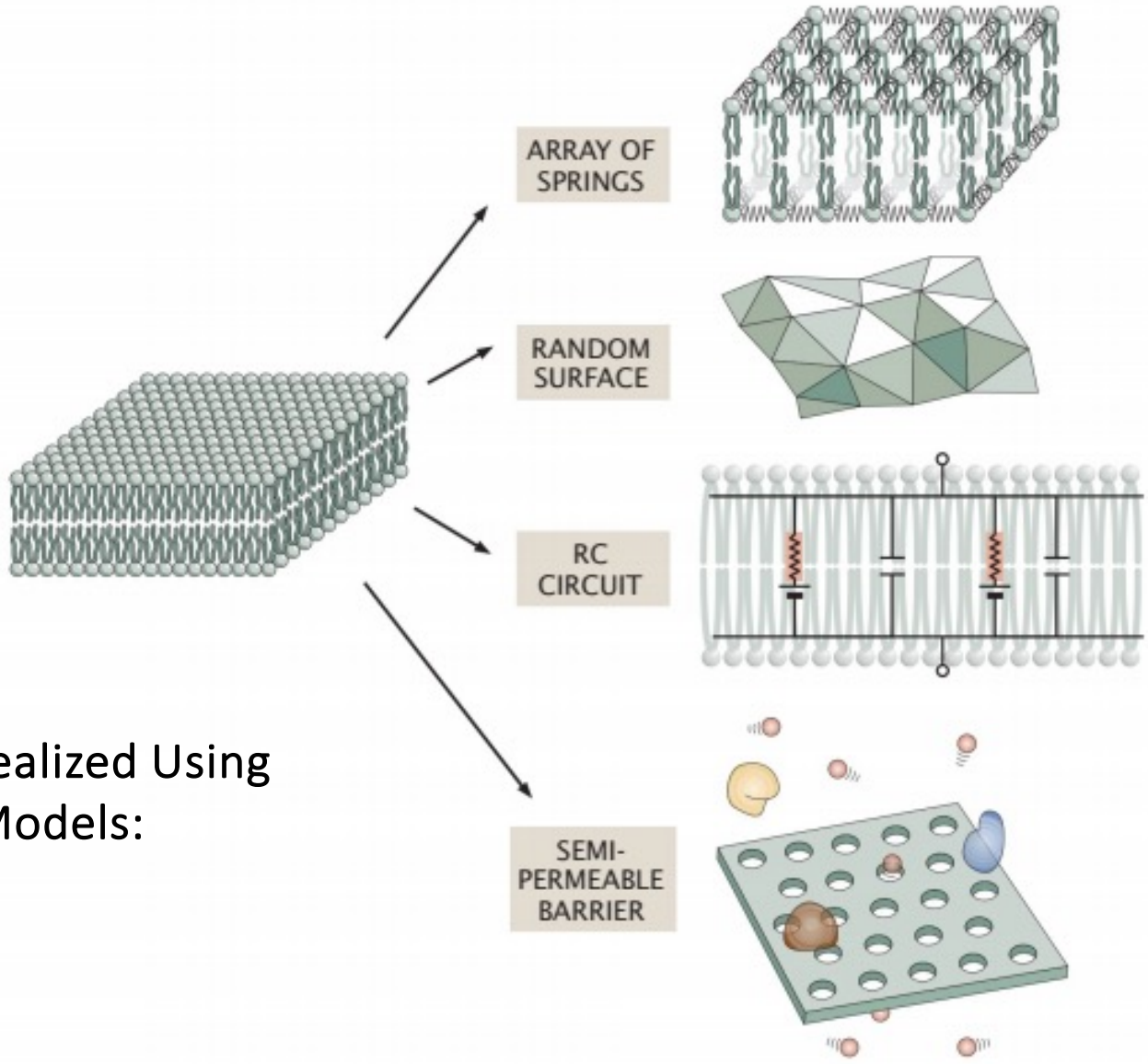


Figure 1.6: Idealizations of protein. Proteins can be thought of as a particular sequence of amino acids, as a simplified sequence reporting only the hydrophobic (H, oil-like) or polar (P, water-like) chemical character of the amino acids, as a collection of connected ribbons and cylinders, as a compact polymer on a lattice, as a binding platform for ligands, or as a two-state system capable of interconverting between different functional forms.

Biological Stuff Can Be Idealized Using Many Different Physical Models: proteins

The global conformation of a protein can be seen as a black box, without worrying about the internal machinations

Figure 1.7: Idealization of membranes. A membrane can be modeled as an elastic object that deforms in response to force, as a random surface fluctuating as a result of collisions with the molecules in the surrounding medium, as an electrical circuit element, or as a barrier with selective permeability.



Biological Stuff Can Be Idealized Using Many Different Physical Models:
Lipid membranes

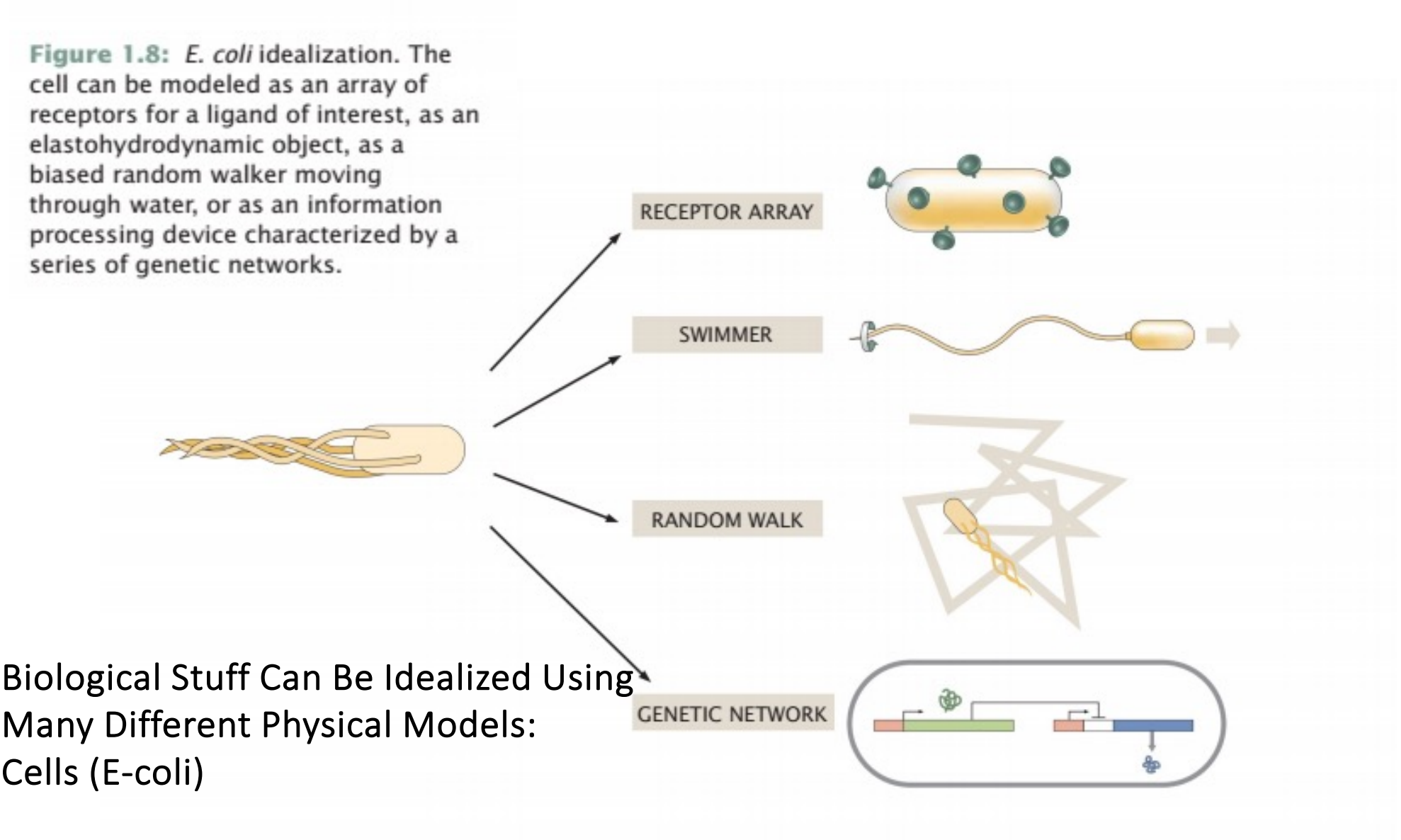


Figure 1.8: *E. coli* idealization. The cell can be modeled as an array of receptors for a ligand of interest, as an elastohydrodynamic object, as a biased random walker moving through water, or as an information processing device characterized by a series of genetic networks.

Biological Stuff Can Be Idealized Using Many Different Physical Models:
Cells (E-coli)

History of Biophysics:

In 1943 Schrödinger gave a few lectures at Trinity College, Dublin, on

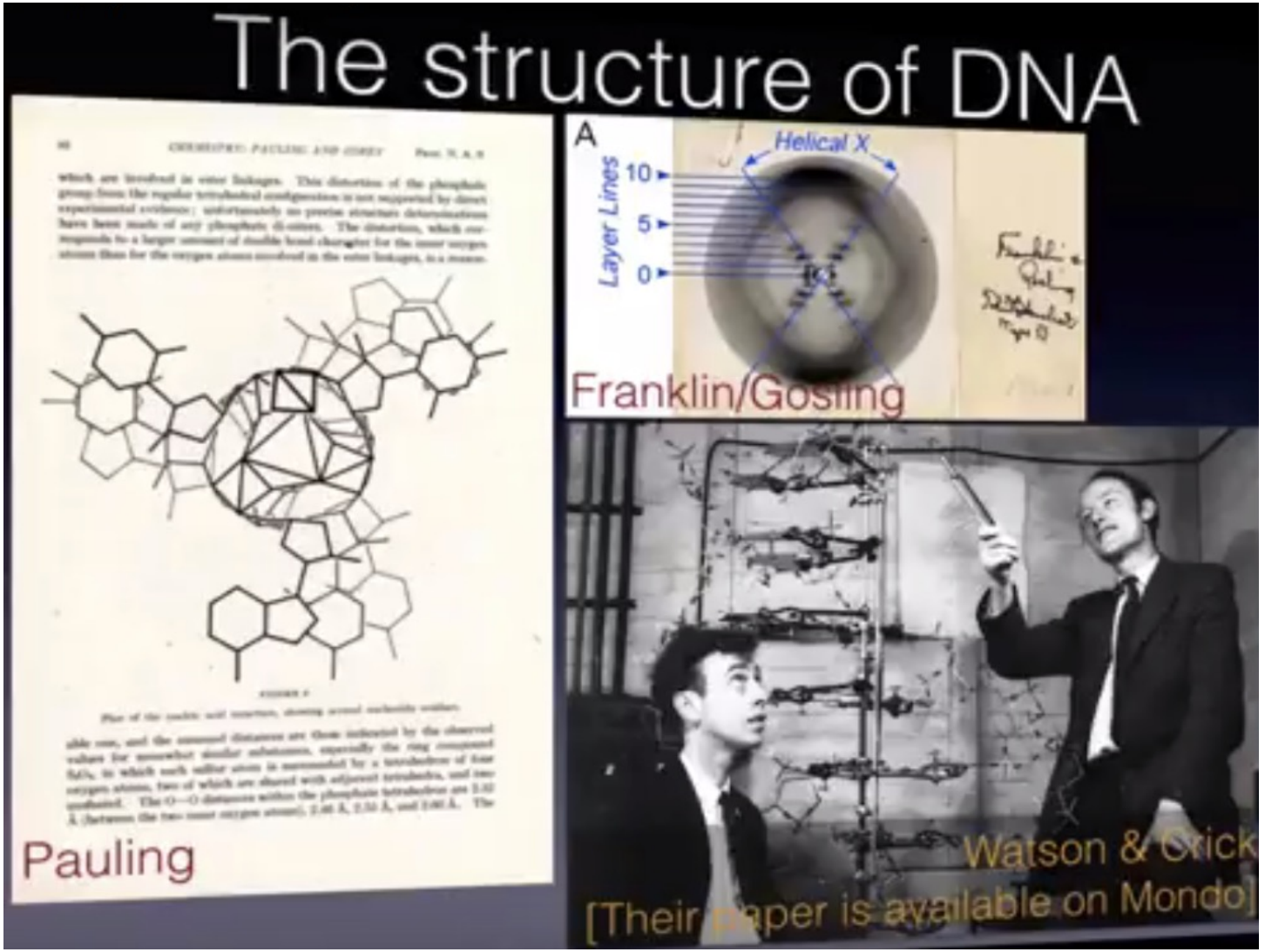
"What is life: the physical aspects of the living cell"

These lectures generated an enormous interest for biology between physicists and chemical physicists, which led to the discovery of DNA and protein structure and the development of molecular biology.

High resolution needed!

History of (Molecular) Biophysics:

Breakthrough is DNA double-helix model obtained by **J.D. Watson** (ornithologist) and **F.H.C. Crick** (physicist) following **Rosalind Franklin**, R.G. Gosling, **M.H.F. Wilkins**, A.R. Stokes, H. R. Wilson fibre diffraction/biochemical studies published on Nature in 1953

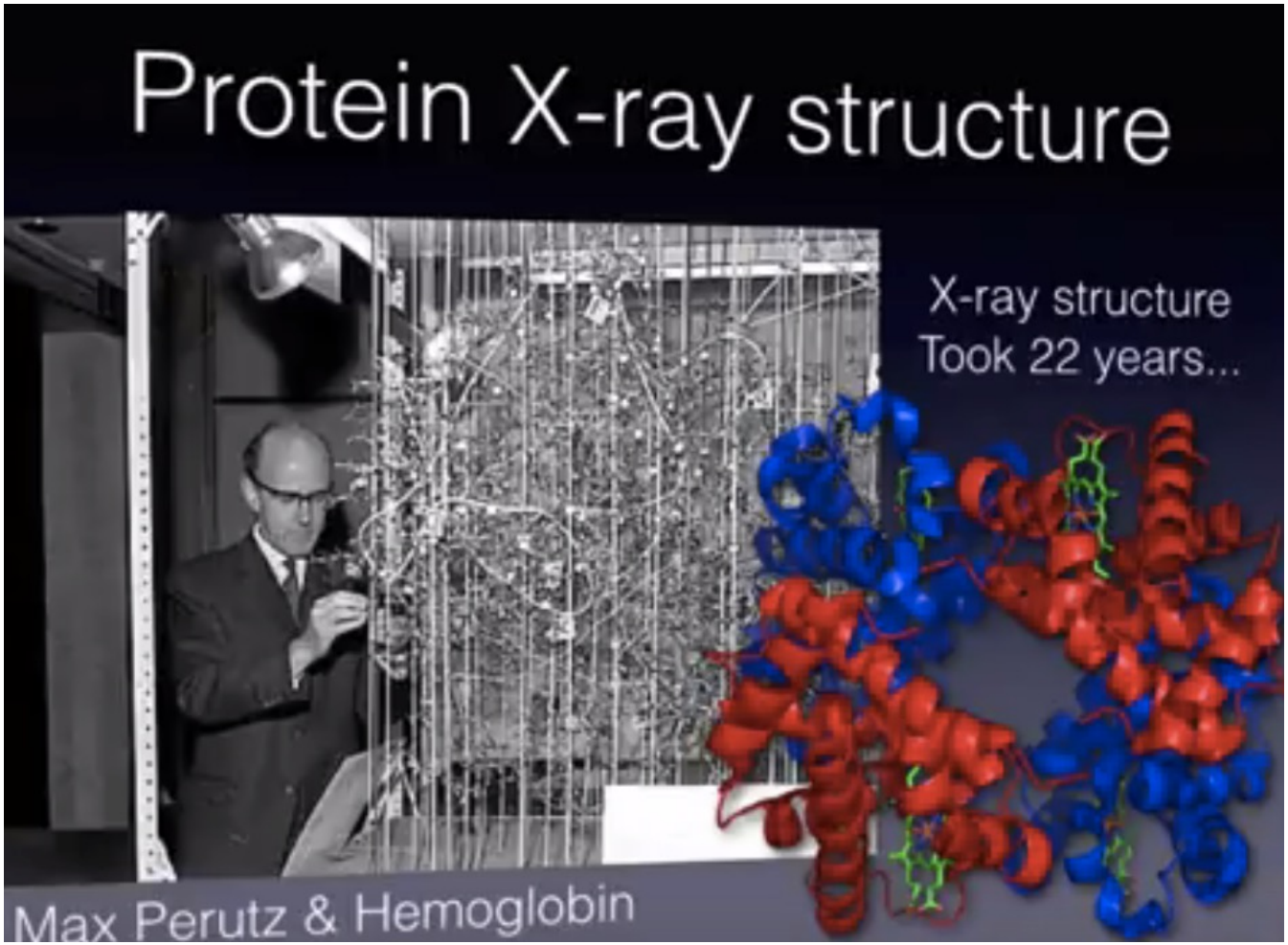


High resolution needed!

History of (Molecular) Biophysics:

The first protein crystal was obtained in 1930, the first X-ray structure by **M. Perutz** and **J. Kendrew** in 1957 (myoglobin). Complex problem, needed complementary biochemical and thermodynamical tools to be solved.

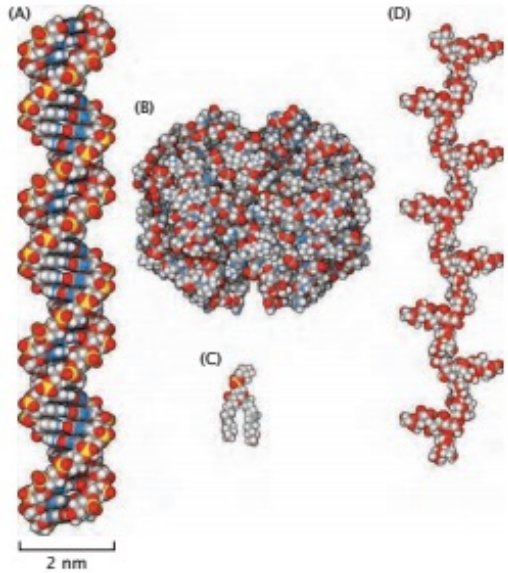
Application of physical concepts and methods to biology --- biophysics!



Macromolecules

Macromolecules can be assembled by the cell from a small number of simpler subunit or precursor molecules. It is the combinatorial assembly of these simple subunits that gives rise to their tremendous structural diversity.

A cell needs only a few chemical reactions to be able to synthesize these sets of subunits from the food in its environment.



Atoms: O, C, N, H, sometime S, P

(A) Atomic structure of a small fragment of the nucleic acid DNA in the B form
(B) atomic structure of the oxygen-carrying protein hemoglobin (PDB 1hho)
(C) phosphatidylcholine lipid molecule from a cell membrane
(D) branched complex carbohydrate

Macromolecules

Proteins:
Structural elements that catalyze reactions fundamental to life
(TENS OF THOUSANDS OF DIFFERENT PROTEINS IN A SINGLE ORGANISM!!)

Carbohydrates:
Energy storage, surface properties on cell membranes, cell walls

Lipids:
Cell membranes, separate organelles in cells (HUNDREDS OF DIFFERENT LIPIDS EXIST)

Nucleic Acids:
Memory, operating instructions, constitute the operation mechanism to generate macromolecules

Macromolecules

Proteins:

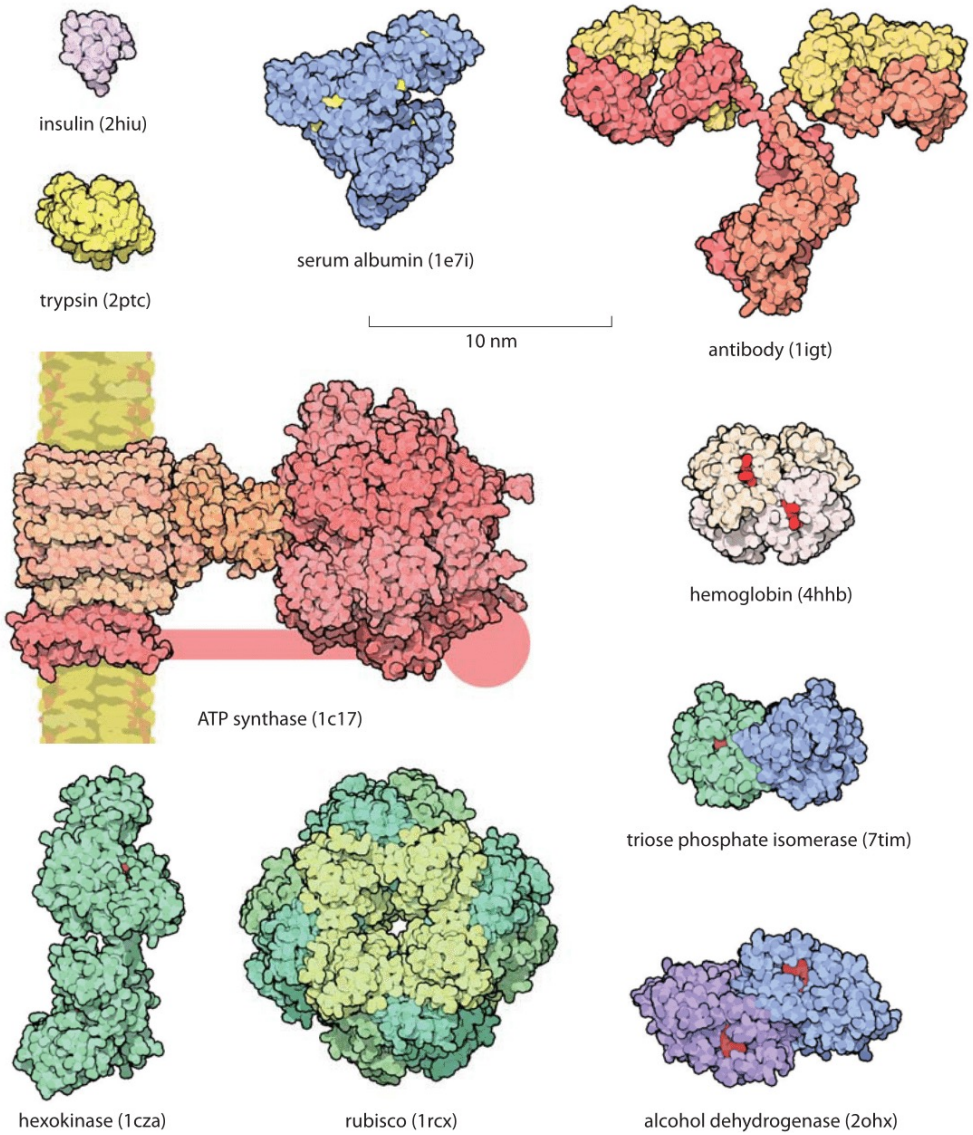
molecular machines, display a wide variety of 3D shapes and of biological functions

- catalyse small molecules synthesis and degradation
- allow cells to move and do work
- maintain cell rigidity
- control genes, switching them on/off
- direct their own synthesis
- move molecules across membranes

Nucleic acids (DNA and RNA):

- contain a coded representation of all proteins of a cell
- contain a coded set of instruction about when proteins have to be made and in which quantities

Proteins



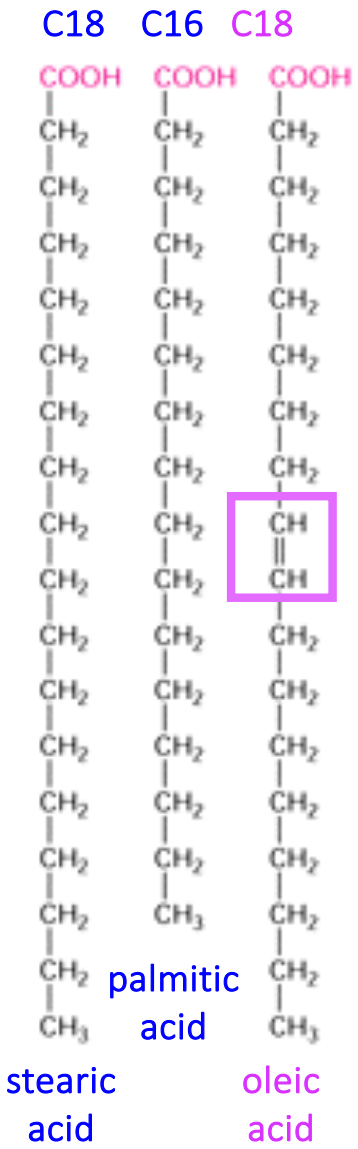
organism	median protein length (amino acids)
H. sapiens	375
D. melanogaster	373
C. elegans	344
S. cerevisiae	379
A. thaliana	356
5 eukaryotes (above)	361
67 bacteria	267
15 archaea	247

http://book.bionumbers.org/how-big-is-the-average-protein/

Lipids

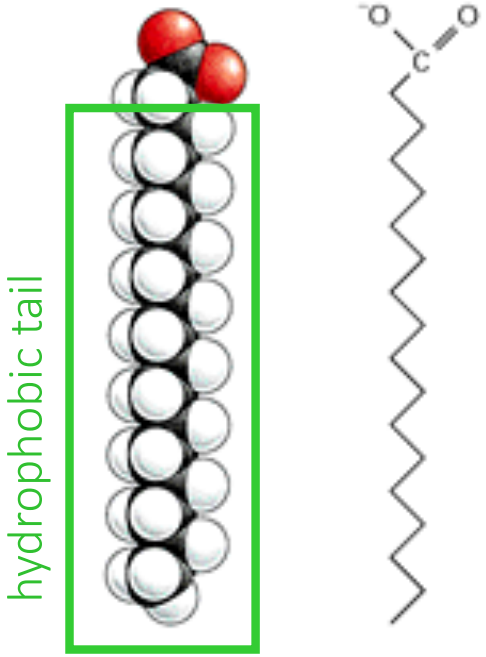
Fatty acids

Carboxylic acids with long hydrocarbon chains (12-24 $-CH_2-$ units)

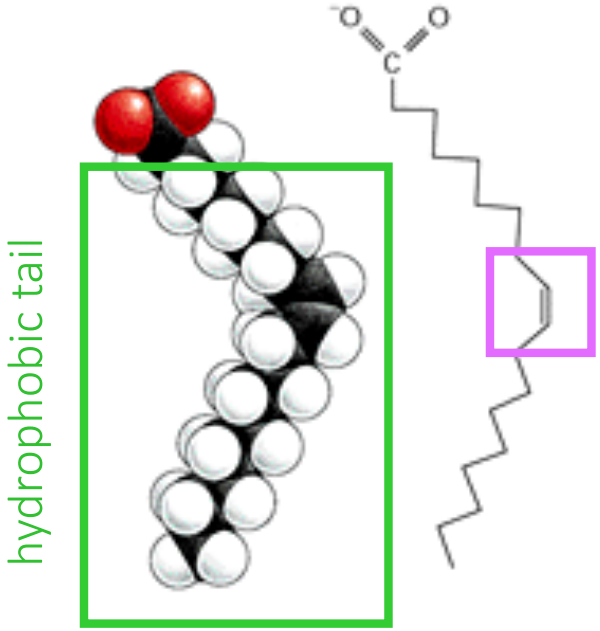


Some have one or more double bonds and are called unsaturated. The double bond is rigid and creates a kink in the chain; the rest of the chain is free to rotate

Stearic acid - saturated



Oleic acid - unsaturated



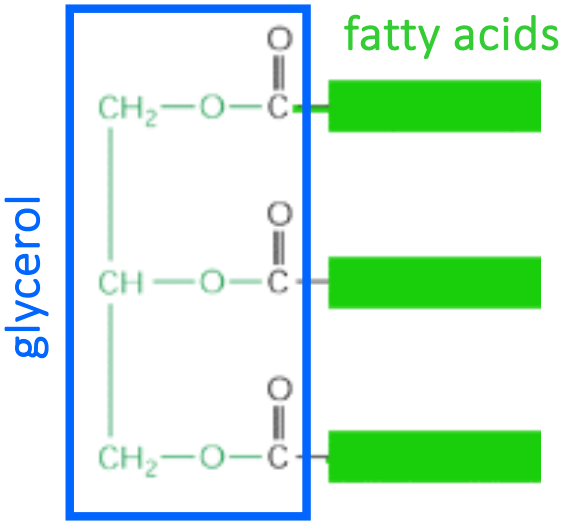
Fatty acids are used as E storage

To ensure a continuous supply of fuel for oxidative metabolism, animal cells store glucose in the form of glycogen and fatty acids in the form of fats.

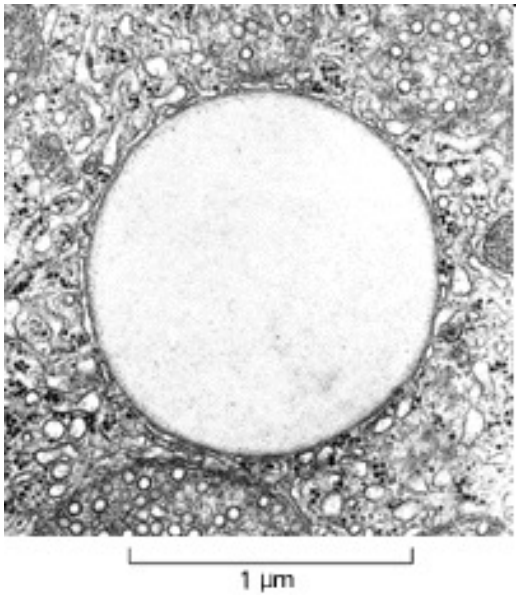
A fat molecule is composed of three molecules of fatty acid linked to glycerol: triacylglycerols (*triglycerides*).

Fat is a far more important storage form than glycogen, because its oxidation releases more than six times as much energy.

Triglycerides have no charge and are virtually insoluble in water, coalescing into droplets in the cytosol of adipose cells.

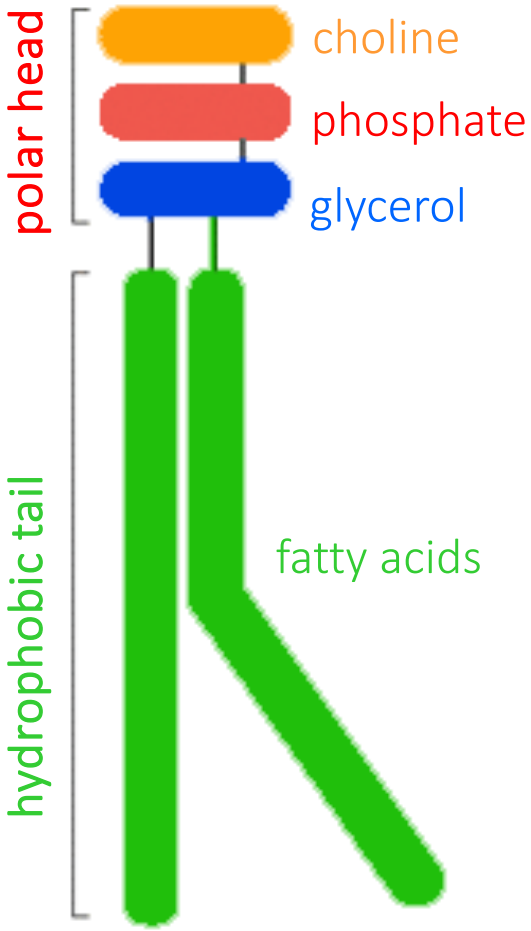


fat droplet

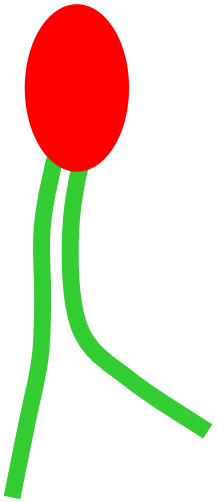


Phospholipids

In phospholipids, two of the OH groups of glycerol are linked to fatty acids, while the third is linked to a phosphate group, which can be further linked to a polar group such as choline, serine, inositol, etc...



Very asymmetric molecule:
- hydrophilic HEAD
- hydrophobic TAIL



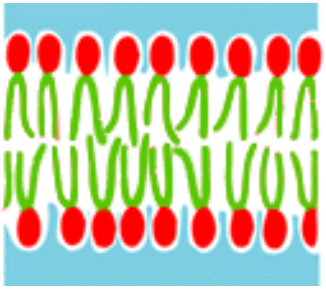
Phospholipids and membranes

Phospholipids are the major constituent of cell membranes.

When in aqueous environment the heads have affinity for the water molecules, while the tails tend to avoid water by sticking together.

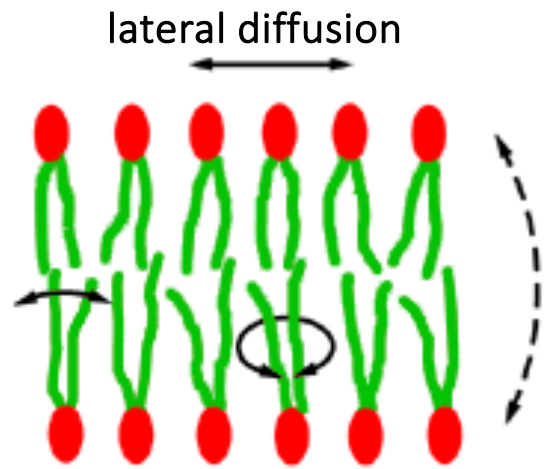
Cellular membranes are essentially made up by phospholipid bilayers.

micelle

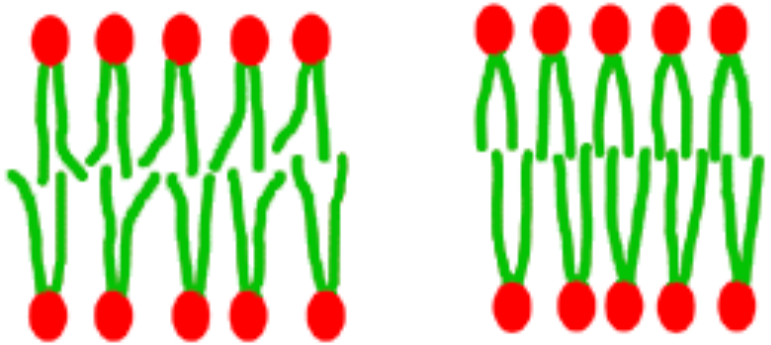


lipid bilayer

Phospholipid mobility



Saturated vs unsaturated lipids

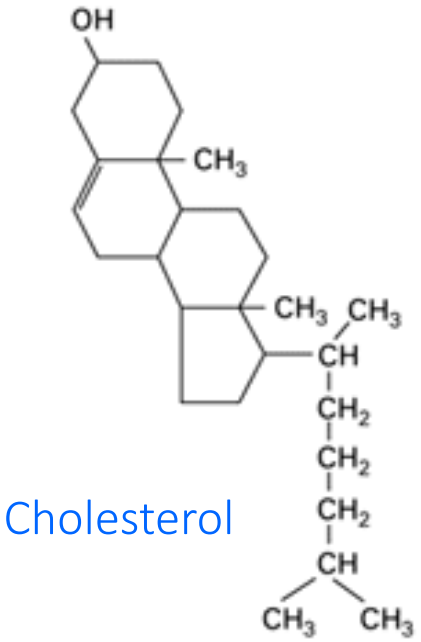


unsaturated, more difficult to pack, more fluid

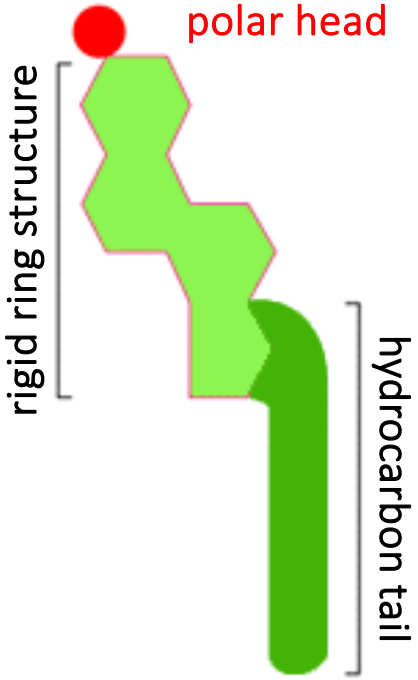
saturated, easier to pack, less fluid

Cholesterol and steroids

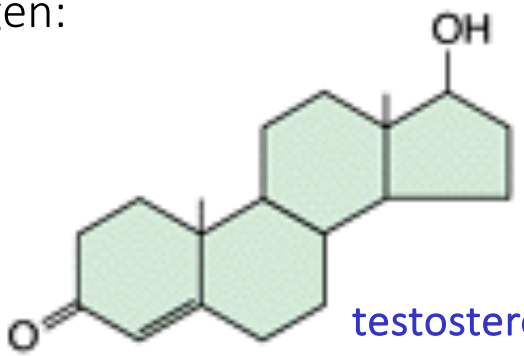
Steroids (such as cholesterol) have a rigid structure made up by 4 rings.



Cholesterol

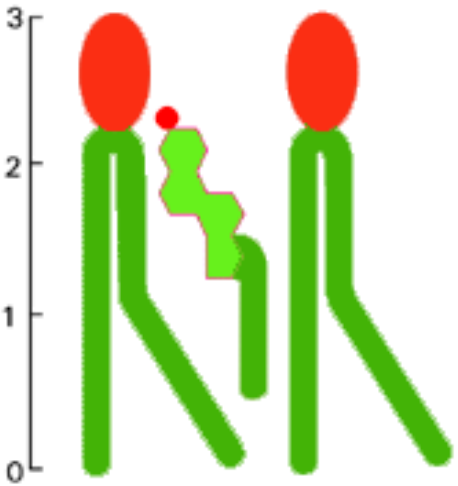


Other important steroid are the sex hormones, such as testosterone and estrogen:



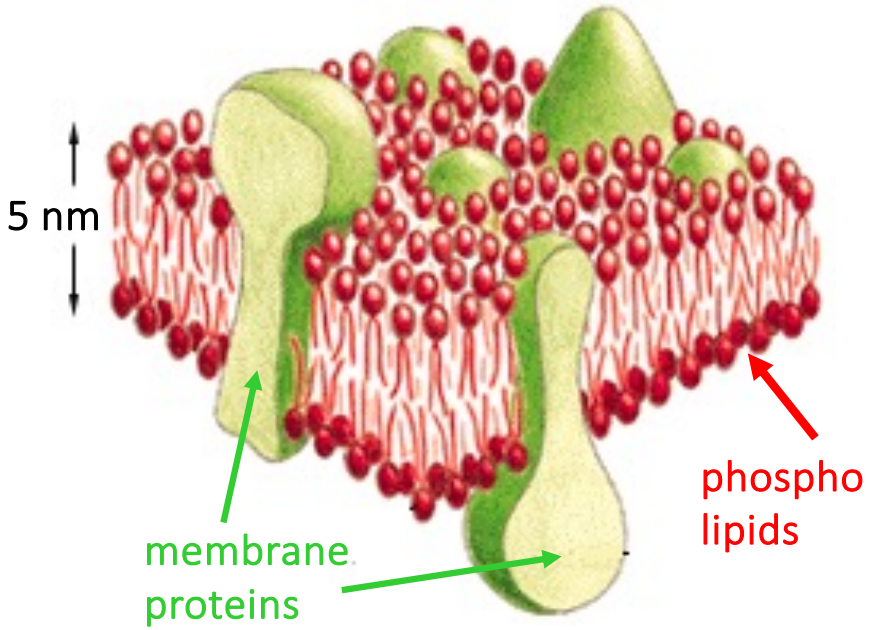
testosterone

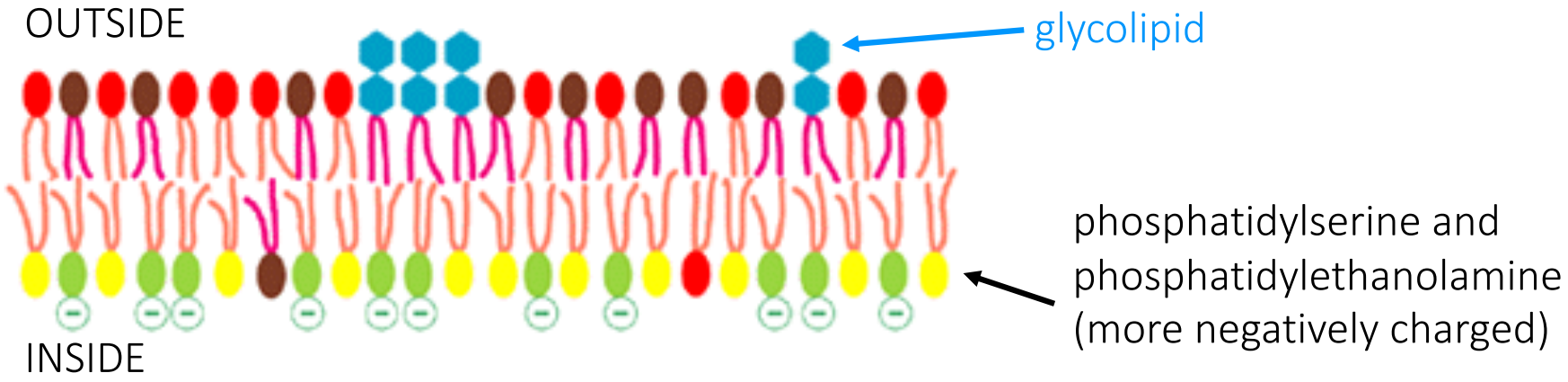
Cholesterol is an important component of the eukaryotic membranes and has a key role in controlling the membrane fluidity.



Cell membranes

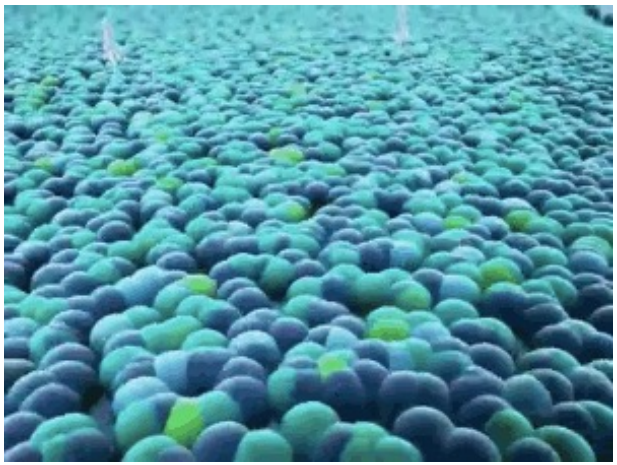
- biological membranes are fluid
- the fluidity is controlled by the % of saturated/unsaturated fatty acid and the % of cholesterol
- membranes are impermeable to ions and most polar molecules (H_2O is actively transported in)
- many proteins are embedded in the membrane
- the membrane is highly asymmetric

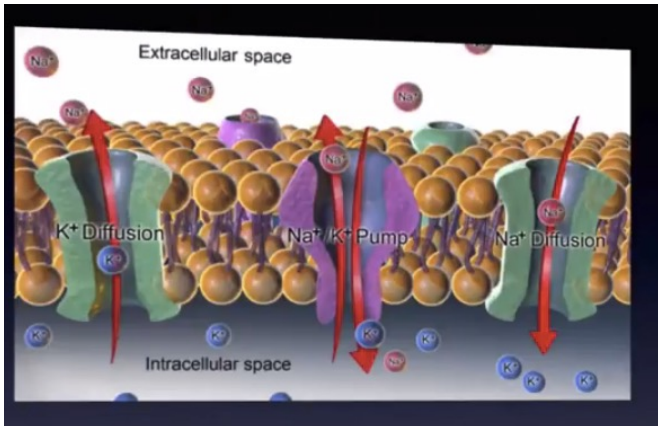


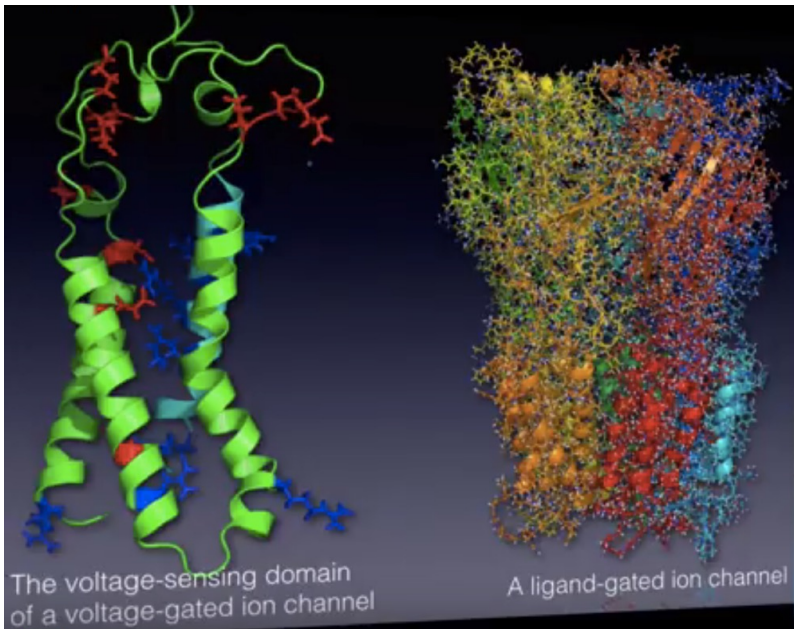


Cell membranes

Also, we know that cell membranes play a crucial role in cell-cell, cell-environment communication. How much do we know about molecular interactions at the cell membrane?



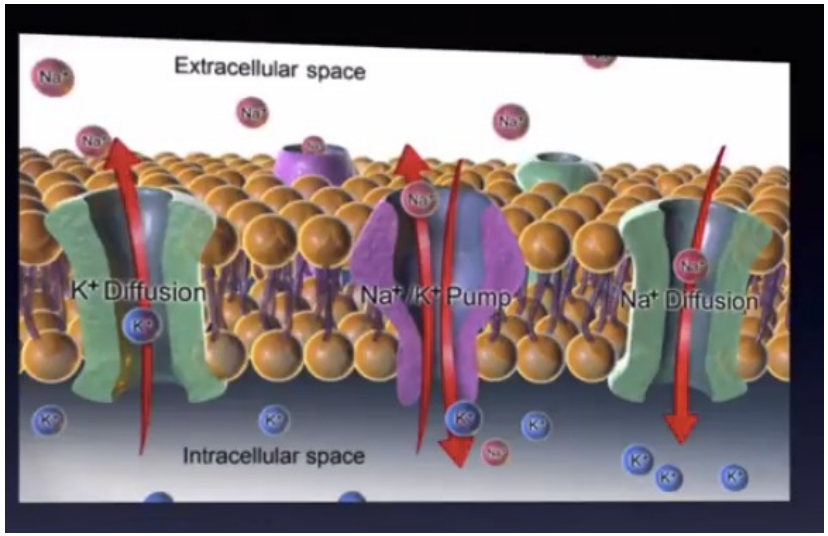


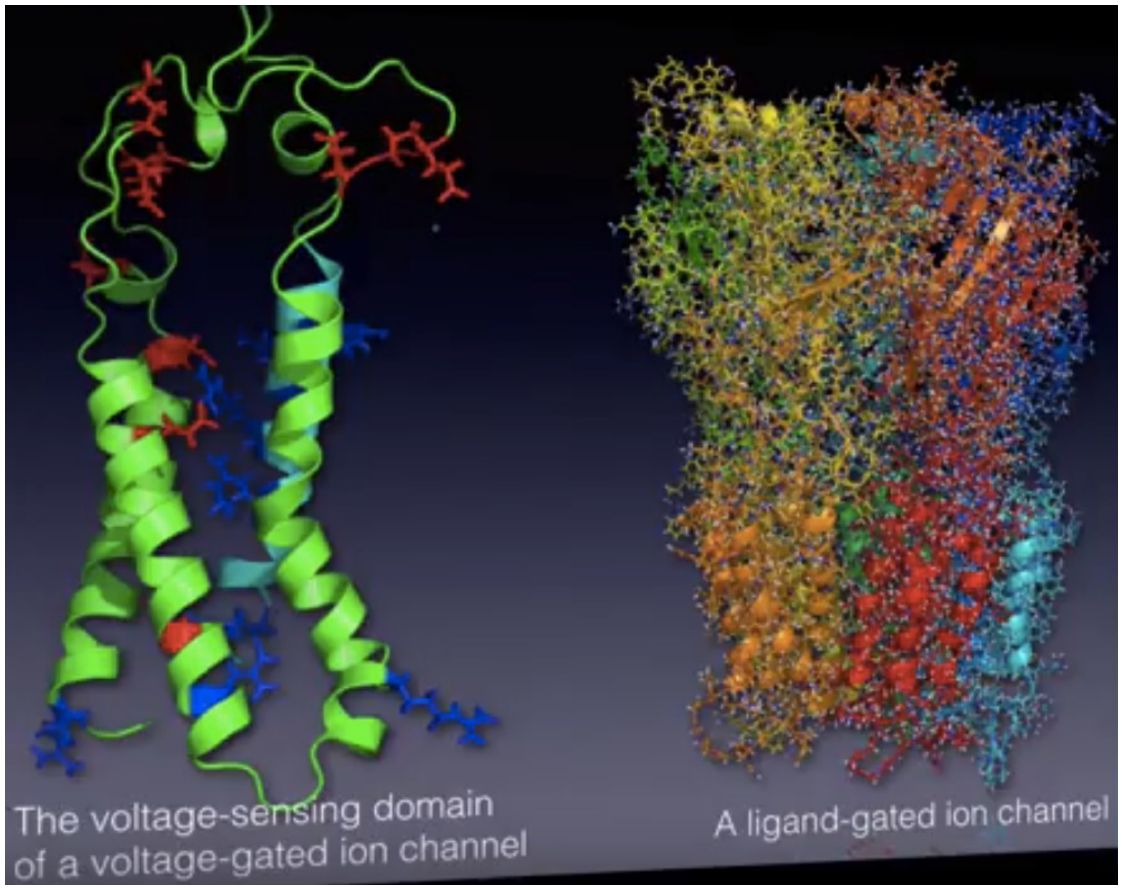


Cell is a highly organized and orderly structure: does it not obey the second law of thermodynamics?

In reality the cell is not an isolated system: takes in energy from the environment and uses this energy to generate order through chemical reaction. From chemical reactions, heat is generated towards the environment inducing disorder outside (thermal motion). The “controlled burning” of food molecules generates biological order.

Membranes are regulating as timer for such control.

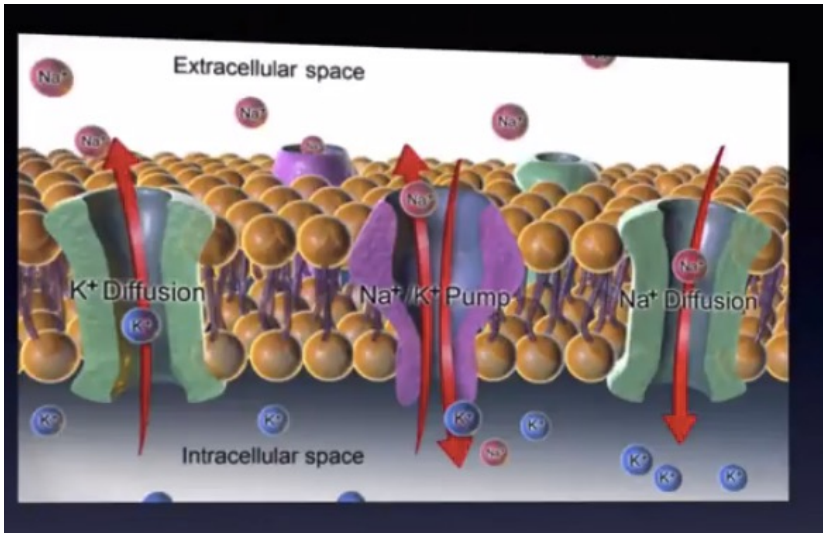


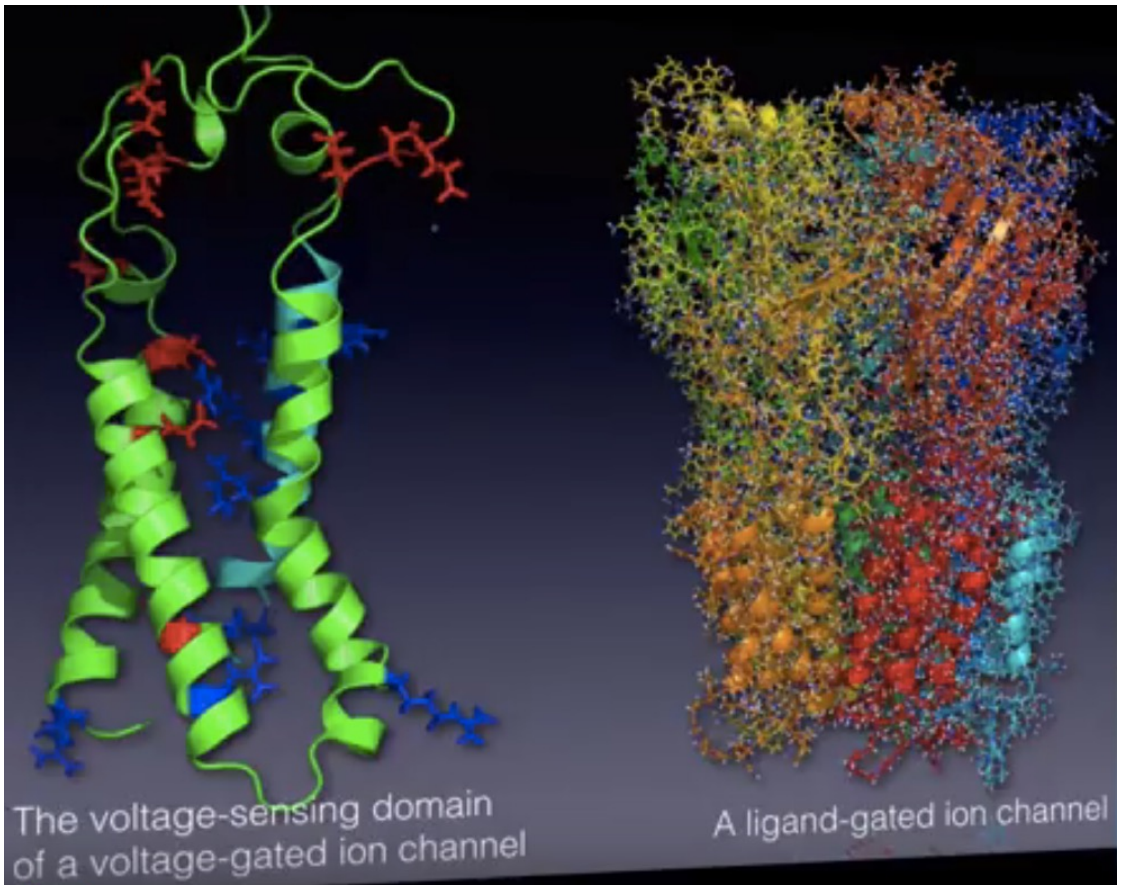


Biology is about distribute/generate/consume energy.

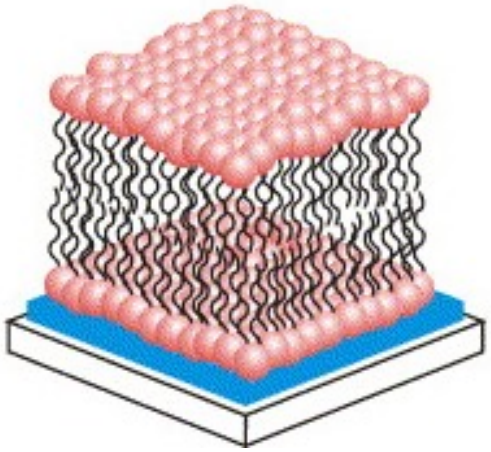
Ion channels are the voltage-sensing domains of a voltage-gated access (heart beats, nerve impulse), which make the cell working as a battery (selective opening of the channel).
Pumps are enzymes in the membrane which move ions in counter direction (ATPase).

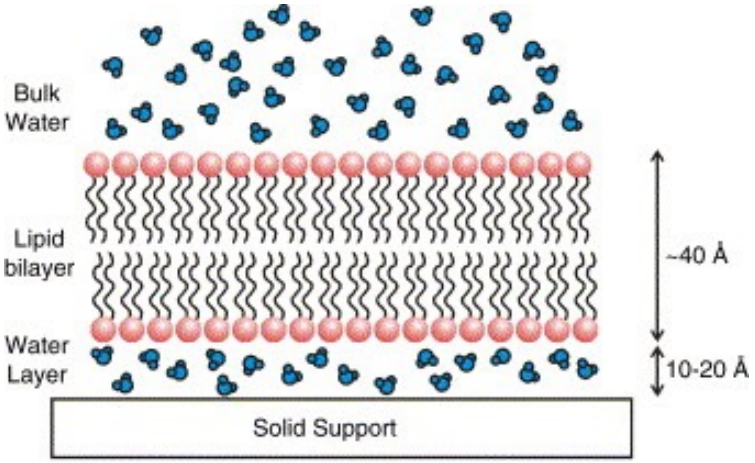
Ligand-gated channels are part of the nerve system (receiving side of the synapsis) and are amazingly specific! One mistake every billion (disease!)





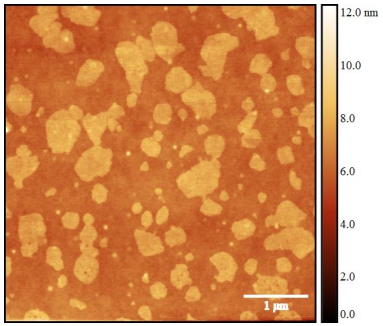
Artificial lipid bilayers



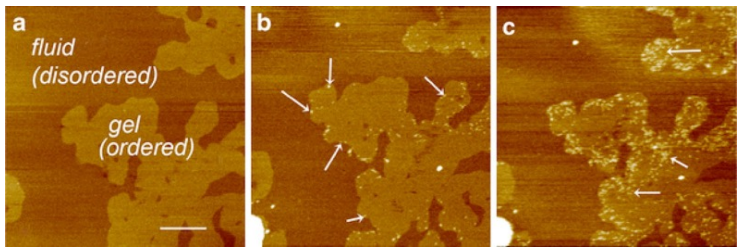


✓ Multicomponent model membranes for mimicking cellular membranes

Lipids behaviour

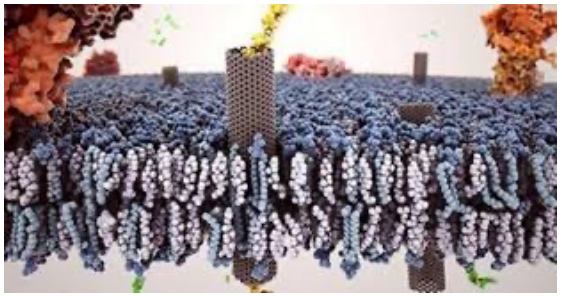


Protein-lipid interactions



Alessandrini, Facci. *J Mol Recognit*. 2011;24(3):387-96.

Carbon nanotubes interactions



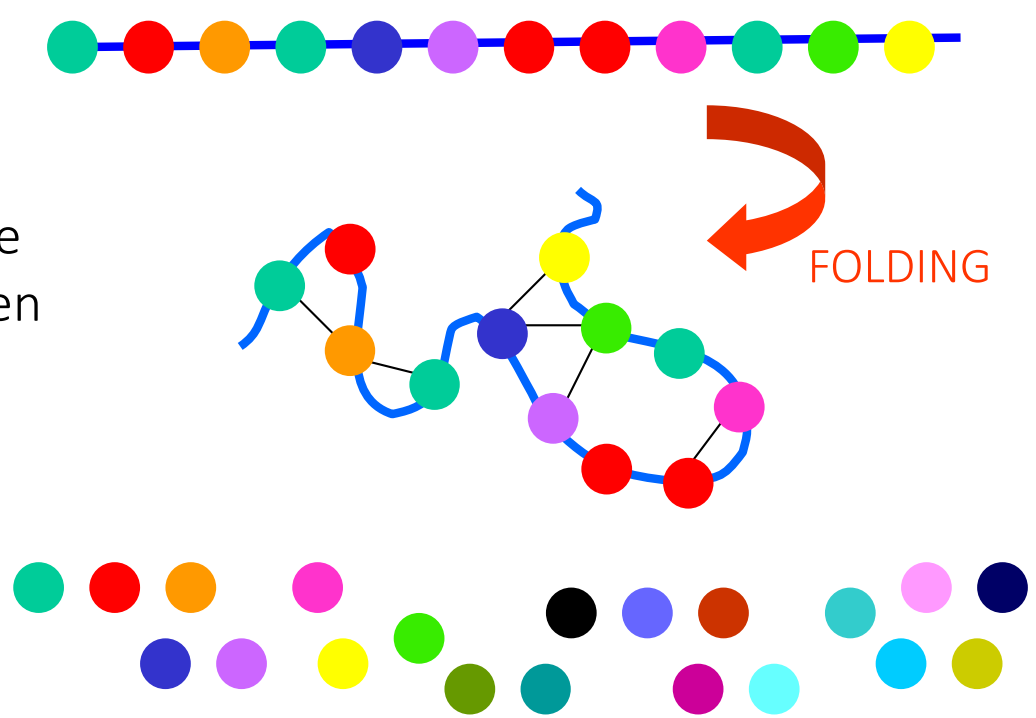
Amino acids and proteins

Proteins

Proteins are linear chains of amino acids.

These chains fold in 3D due to the non-covalent interactions between regions of the linear sequence

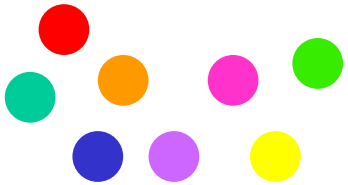
There are 20 different types of amino acid, each with different physico-chemical properties.



- FUNCTION DEPENDS ON 3D STRUCTURE
- 3D STRUCTURE DEPENDS ON SEQUENCE
- SEQUENCE IS DETERMINED GENETICALLY

Overview of protein architecture

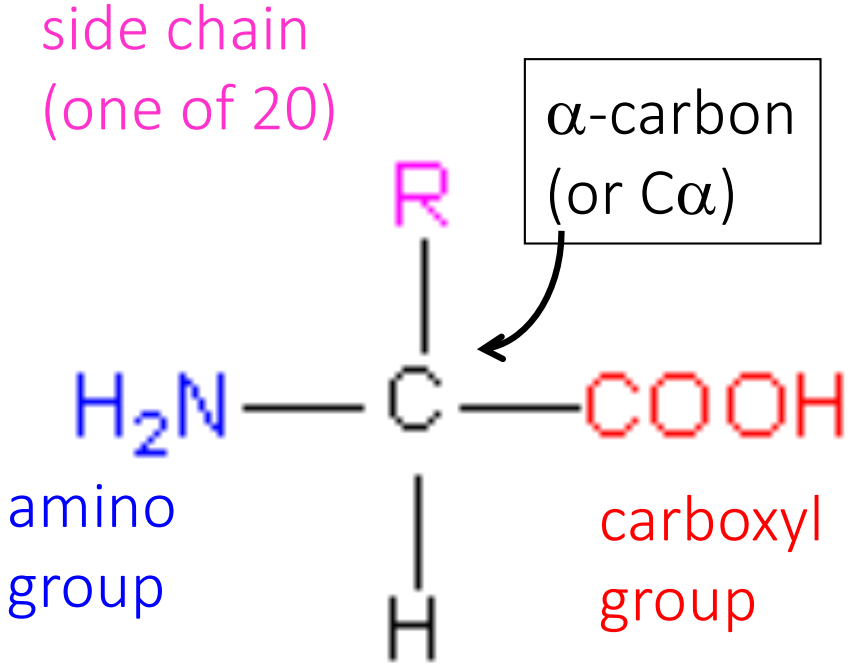
1) structure and chemistry of amino acids



2) how amino acids are linked together through peptide bonds to form a polypeptide chain

3) how the polypeptide chain folds in 3D
   - secondary structure elements (α-helix and β-sheet)
   - how secondary structure elements pack together

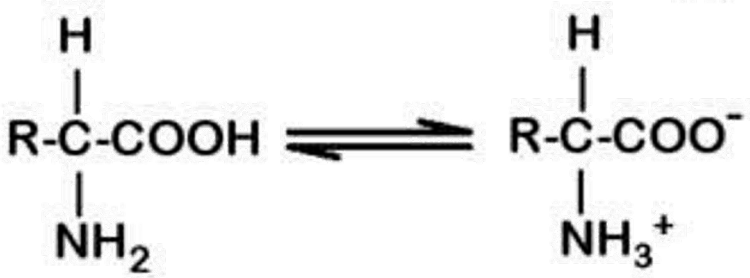
Structure of amino acids



At neutral pH:

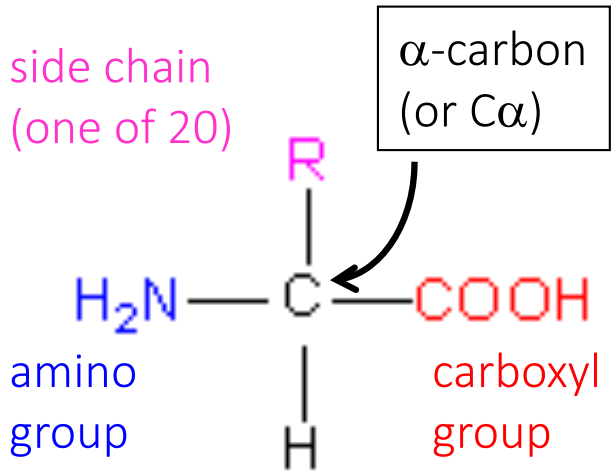
NH_2 is a base ⟶ NH_3^+

COOH is an acid ⟶ COO^-

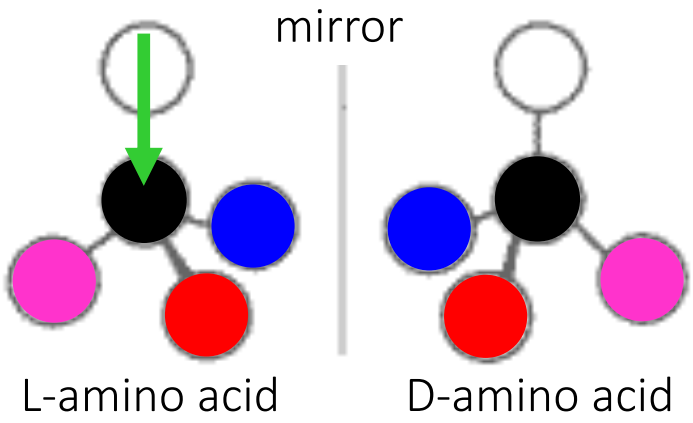


zwitterion
(dipolar form)

Structure of amino acids

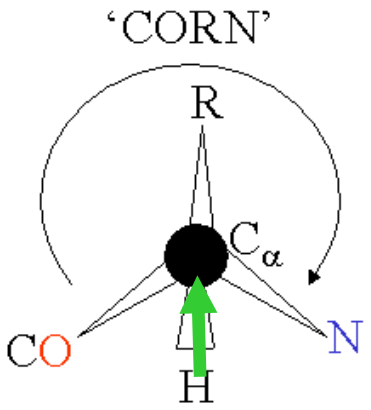


The Cα is an asymmetric carbon (bound to 4 different groups) and therefore is a chiral centre.
Two configurations (stereoisomers) are possible, which are one the mirror image of the other:

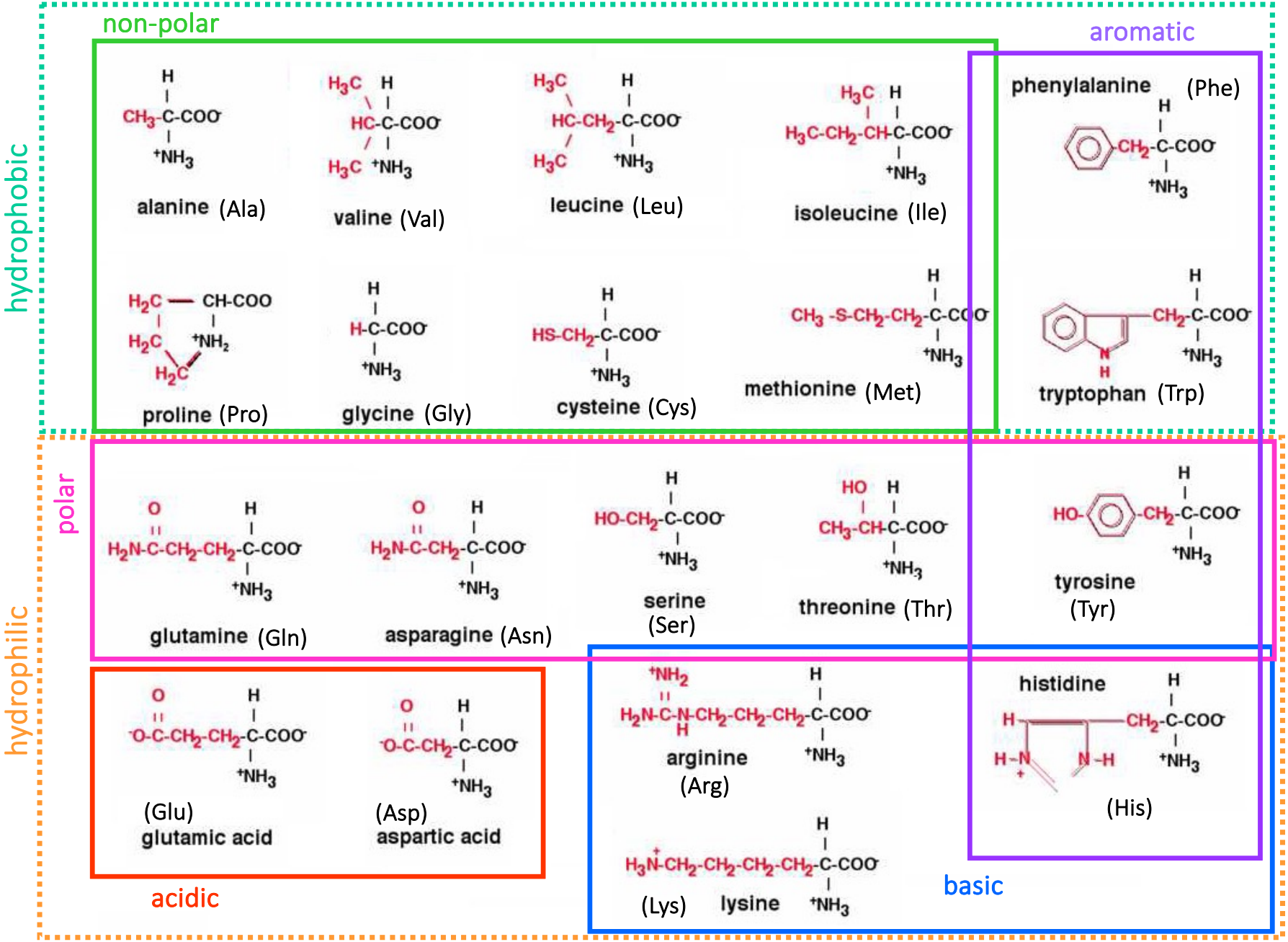


all amino acids in proteins are L!!

CORN rule:
looking down the H-Cα bond for an L amino acid we read the groups CO-R-N clockwise



The 20 amino acids:



Properties of amino-acid side chains

R varies in

- shape
- size
- charge
- hydrophobicity
- reactivity

Hydrophobic amino acids: insoluble or slightly soluble in water

(side chains made of C, H, S - atoms with similar electronegativity)

avoid water by coalescing into oily droplets - the same forces causes hydrophobic aa to pack together in the interior of proteins, away from acqueous solutions.

Hydrophylic amino acids: soluble in water

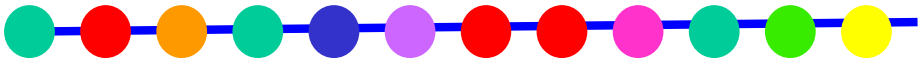
(side chains contains atoms such as N and O, which can make HB)

- polar
- basic
- acidic

Overview of protein architecture

1) structure and chemistry of amino acids

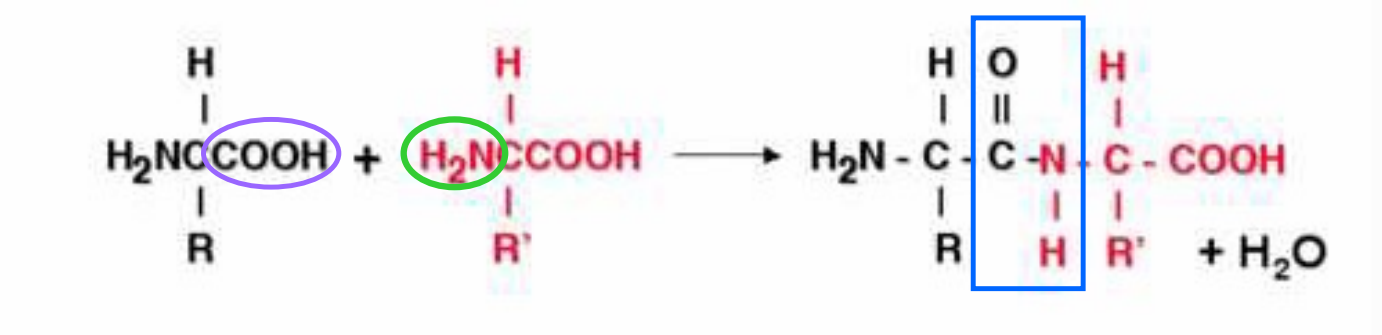
2) how amino acids are linked together through peptide bonds to form a polypeptide chain



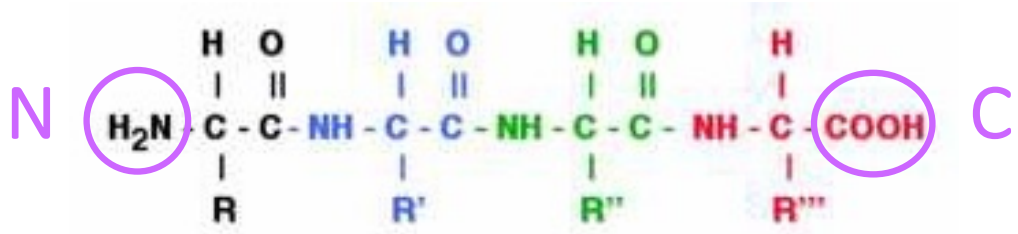
3) how the polypeptide chain folds in 3D
   - secondary structure elements (α-helix and β-sheet)
   - how secondary structure elements pack together

The peptide bond

The amino acids of a protein are joined together through a covalent bond between the carboxyl group of one aa and the amino group of the next aa (peptide bond).



This produce a chain of amino acids which is asymmetric: on one end there is a free NH_2 group (N terminus) and at the other end a free COOH (C terminus).



arg ala phe gly

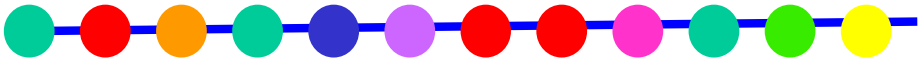
R A F G

A peptide/protein sequence is always given from the N to the C terminus (here RAFG).

Primary structure

the linear sequence of amino acids

- the sequence is always written N⊢→C
- each protein has a unique and defined sequence, which is genetically determined
- a typical protein contains 100-1000 aa
- sequencing=determining the number and order of the aa in the chain



In 1953 Saenger sequenced insulin (Nobel price); now it is more common to sequence the corresponding gene.

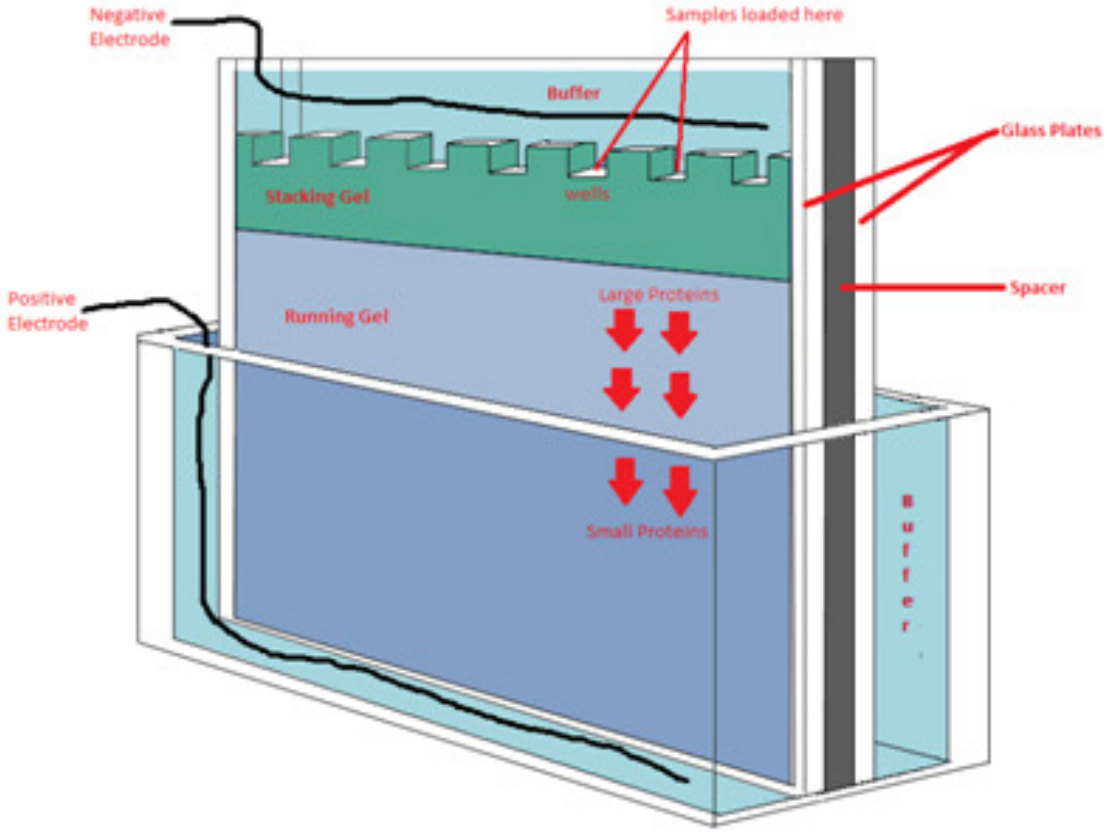
We can guess the function of an unknown protein if it shows sequence similarity to a protein of known function.

Often we know the sequence of the same protein from different organisms: these are more and more different the more the organisms have diverged in evolution.
Proteins evolve by changing (little by little) their aminoacid sequence

Mass of a protein

1D –SDS-PAGE

Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis



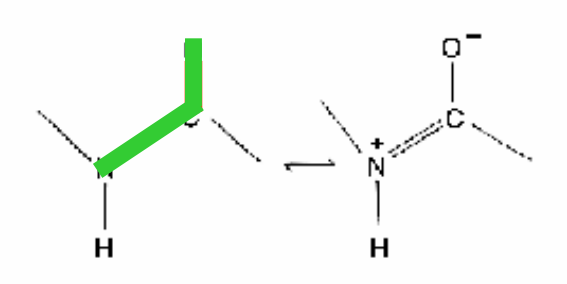
Concentration of acrylamide (%)	Protein size (kDa)
5	36-200
7.5	24-200
10	14-200
15	14-60

a method that separates protein by molecular weight over a range of about 10 to 300 kilodaltons (kDa). Samples are weighed and dissolved in sodium dodecyl sulfate (SDS). SDS is a negatively charged detergent that has both hydrophilic and hydrophobic regions. SDS likes to bind to proteins (1.4 g SDS/1 g protein) and to be in water. This SDS- protein-water interaction allows water insoluble proteins to dissolve in water, and to dissolve protein mixtures. Proteins are completely denatured. When an electric field is applied, the negative charge of the SDS causes the proteins to move through a clear acrylamide matrix toward the positive electrode. This matrix has holes in it that sieve out the proteins by molecular weight. Large proteins move more slowly through the matrix than the smaller proteins thereby separating proteins by molecular weight.

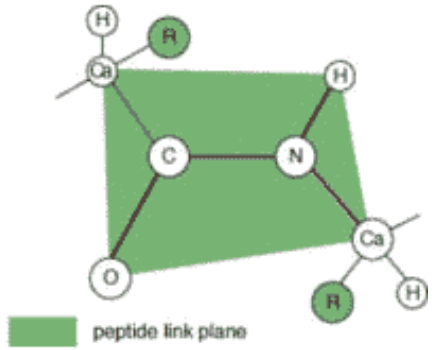
1 Da = 1 g/mol

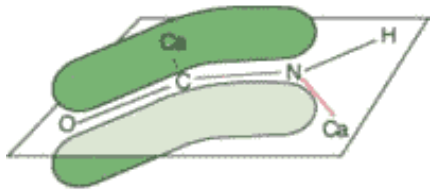
Average mol. weigth of 1 aminoacid: 110 Da

Planarity of the peptide bond



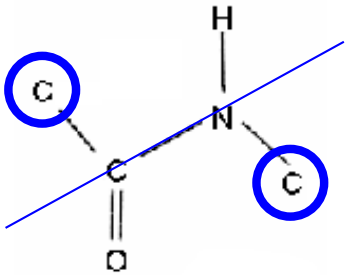
Partial double bond character of the N-C bond leads to restricted rotation
the region NH-CO is planar:



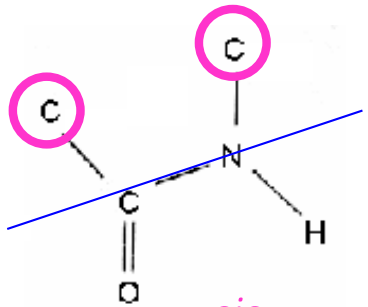


delocalisation of the π electrons over the entire peptide bond, rather than simply over the C=O bond

The peptide bond can assume a *trans* or a *cis* conformation: the *trans* form is favoured 1000:1.

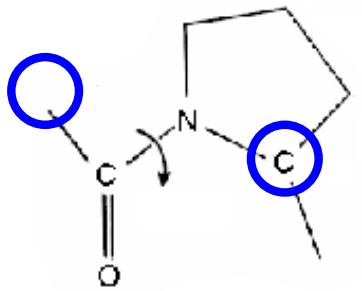


trans



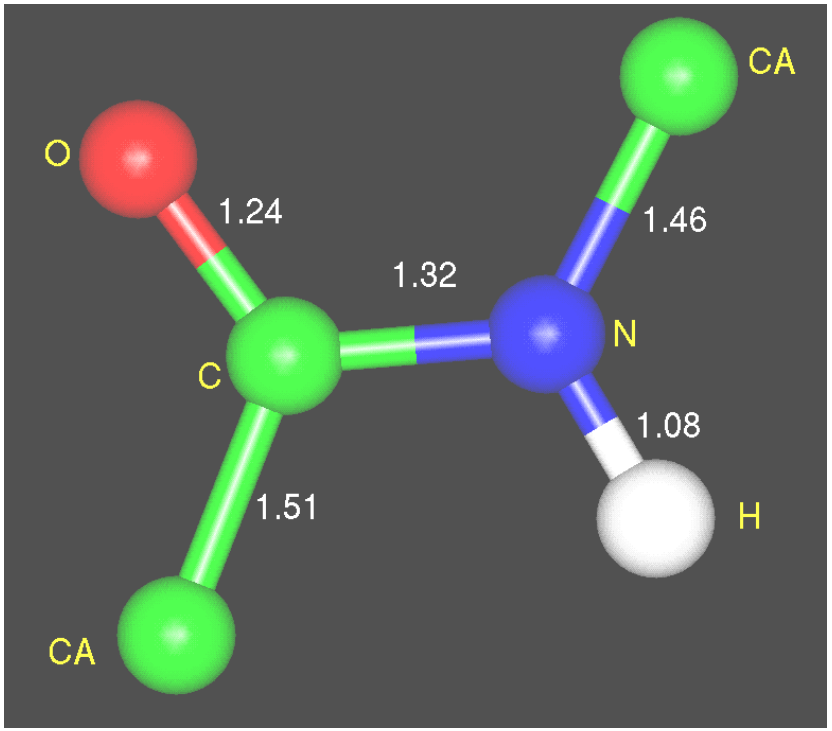
cis

In the case of prolines, the *trans* form is only favoured 15:1



trans
Proline

The ideal peptide



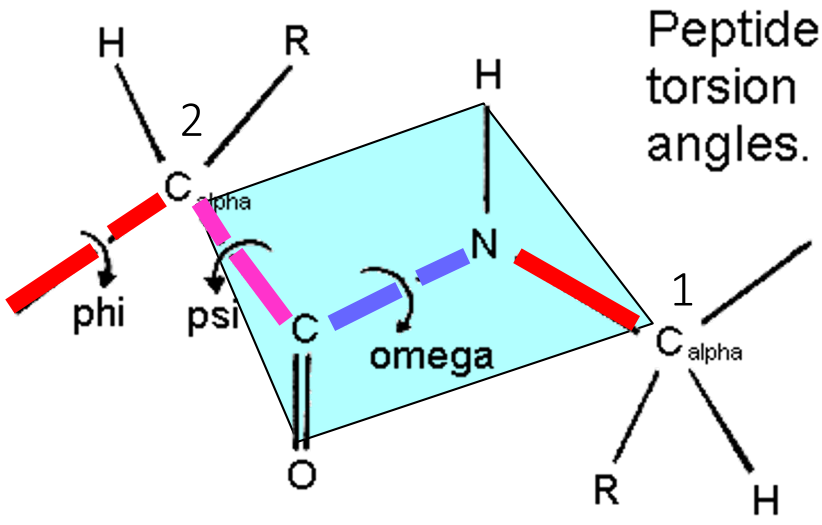
C-N single bond ~ 1.48 Å

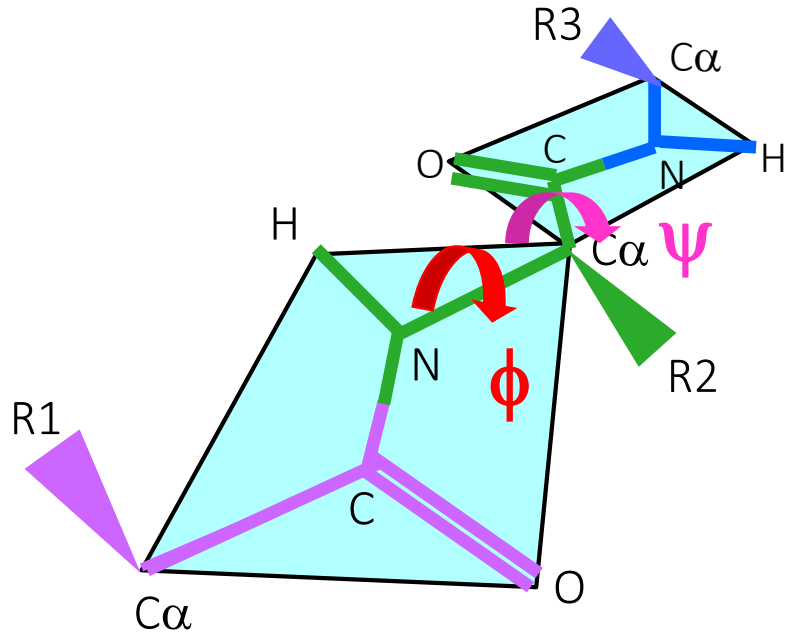
C=O double bond ~ 1.20 Å

peptide bond C-N = 1.32 Å (i.e. shorter than a single bond due to partial double bond character) while C=O bond is slightly longer

Peptide bond	Average length	Single Bond	Average length	Hydrogen Bond	Average (±0.3)
Cα – C	1.51 (Å)	C - C	1.54 (Å)	O-H --- O-H	2.8 (Å)
C - N	1.32 (Å)	C - N	1.48 (Å)	N-H --- O=C	2.9 (Å)
N - Cα	1.46 (Å)	C - O	1.43 (Å)	O-H --- O=C	2.8 (Å)

The torsion angles ψ and ϕ





omega (ω) = rotation around C-N bond
not allowed because of resonance,
therefore ω=180° (for trans)

planar region

phi (ϕ)= free rotation around Cα-N bond

psi (ψ)= free rotation around Cα-C bond

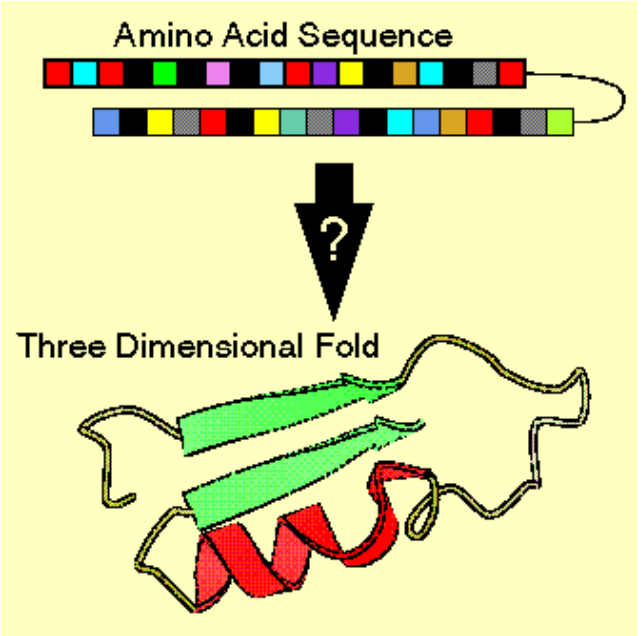
The main chain conformation is defined by the sequence of the (ψ,ϕ) angles: the list of the (ψ,ϕ) for each amino acid dictate the fold of the polypeptide chain, i.e. the 3D structure of the protein

How do proteins fold in the cell?

The amino-acid sequence specify the 3D structure, which is (probably?) the energy minimum for that particular sequence...

BUT how does a protein reach the correct three-dimensional fold?

by trying out all the possible conformations?



- consider the number of possible conformations of a chain of 100 amino acids
- assume each amino acid can have only 3 different conformations
- 3^{100} = 5×10^{47} possible different conformations
- if it took only 0.1 psec (10^{-13} sec) to try each possibility, it still would take 1.6×10^{27} years to find the minimum of energy!

➡ There must be a 'folding pathway'!!!

first forming local structures quickly, then packing them together

The “folding problem”

Experimental approach

Studying experimentally how folding of a particular protein occur in vitro by using techniques like NMR which can detect the presence of secondary structure elements in a partially unfolded protein (trying to determine the ‘folding pathway’)

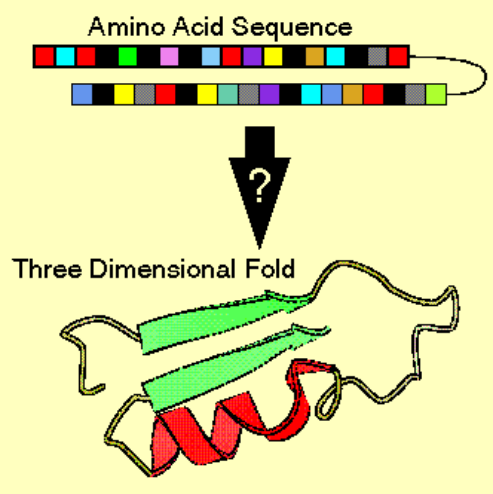
Studying experimentally how folding occur in the cell: some proteins fold by themselves, others require the help of other proteins called chaperones.

Theoretical approach

Using bio-informatics to predict the 3D structure from the amino-acid sequence. The sequence dictate the fold, but we are not very good at going from the sequence to the structure!

Problems?

- poor energy functions and parameters
- complexity
- treatment of solvent



Molecular evolution

Proteins evolve by changing little by little their amino-acid sequence

Changes are due to random mutations in the gene that code for that protein

- some mutations disrupt the structure and/or function of the protein and are eliminated by the selective pressure
- some mutations are 'neutral' and therefore allowed
- some (rare) mutations improve the functionality of the protein or change the function in a way that is advantageous for the cell

→ evolution will select the favourable mutations

A lot of small changes occurring in all protein sequences accumulate with time and are responsible for the variety of living forms we see.

By comparing amino-acid sequences of proteins we can build evolutionary trees:

- key residues (structurally or functionally) are usually conserved
- other residues are usually very similar in organisms that have diverged recently but more and more diverse in distantly related organisms

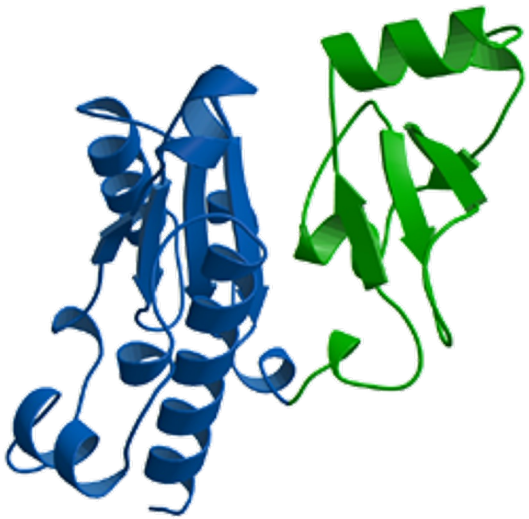
Overview of protein architecture

1) structure and chemistry of amino acids

2) how amino acids are linked together through peptide bonds to form a polypeptide chain

3) how the polypeptide chain folds in 3D:

- secondary structure elements (α-helix and β-sheet)
- how secondary structure elements pack together



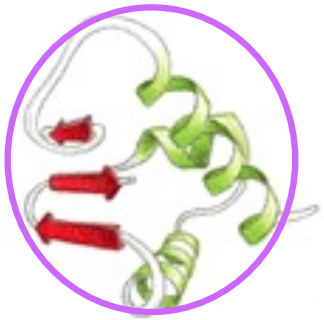
Protein architecture

Secondary structure

local organisation of the polypeptide chain

α-helix

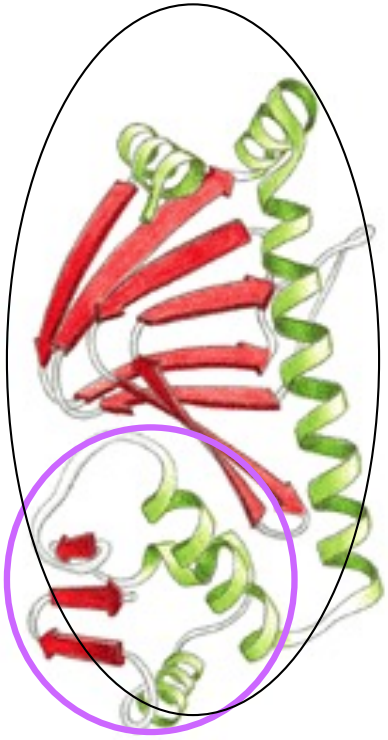
β-sheet



domain

Tertiary structure

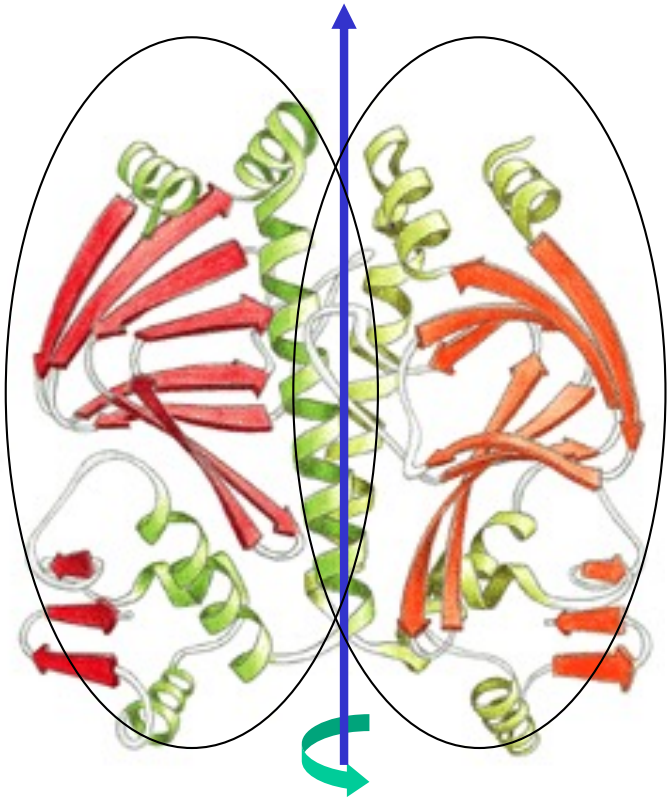
how the secondary structure elements pack together to give a 3D structure



monomer (or subunit)

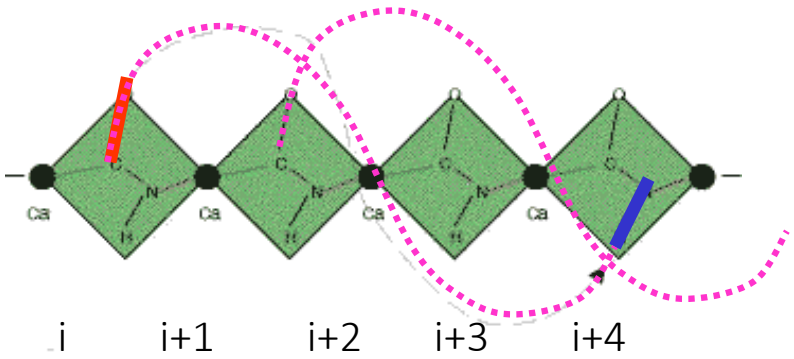
Quaternary structure

the number and relative position of the subunits in a multimeric protein



dimer (α2)

The α-helix

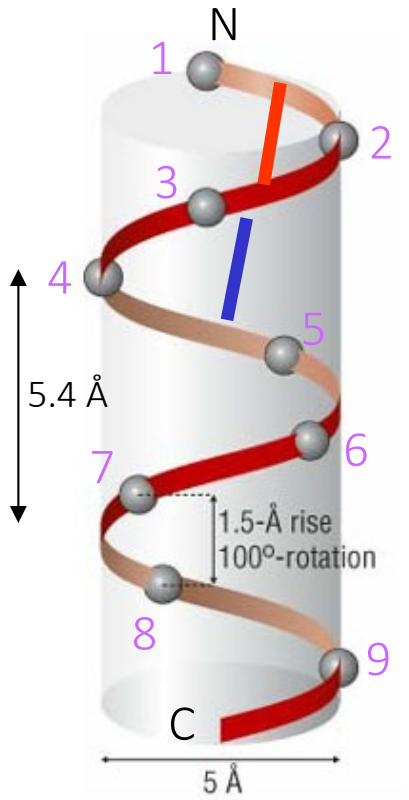


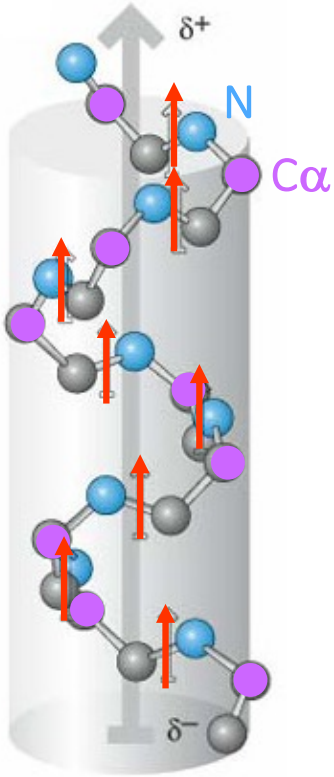
H-bonding pattern $CO_i \mapsto NH_{i+4}$
(local interactions)

all main-chain CO and NH are bonded

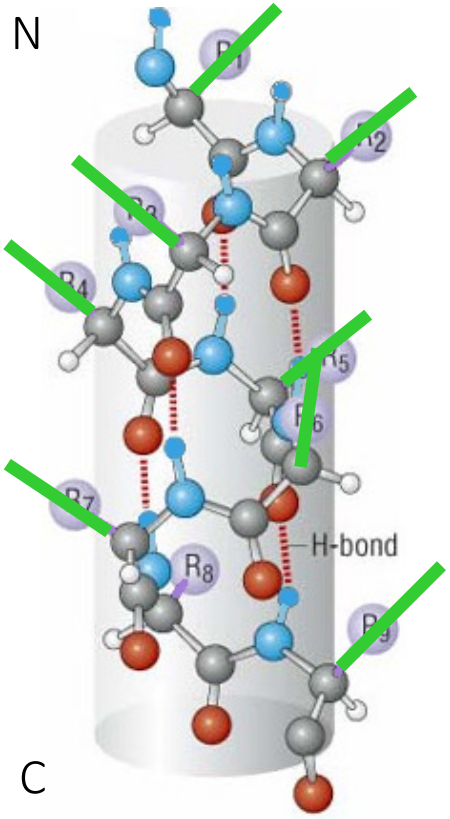
3.6 amino acids per turn;
1.5 Å rise per amino acid
$\mapsto$ 5.4 Å pitch

each peptide bond has a small dipole moment; in a helix all peptide bonds point in the same direction and generate a dipole pointing towards N





The α-helix

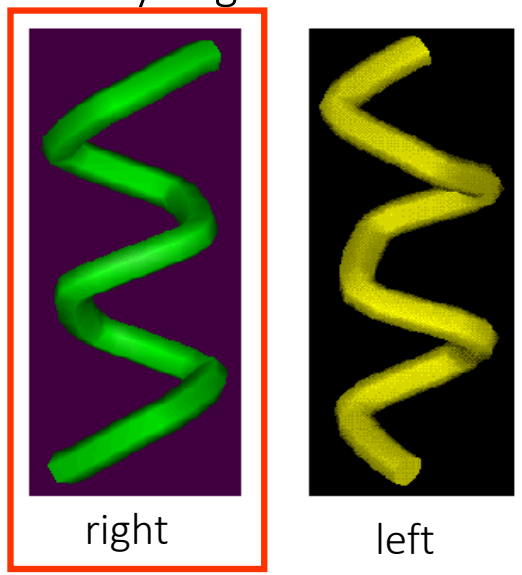


- nitrogen
- oxygen
- carbon
- R=side chain

rod-like structure with side chains extending outside

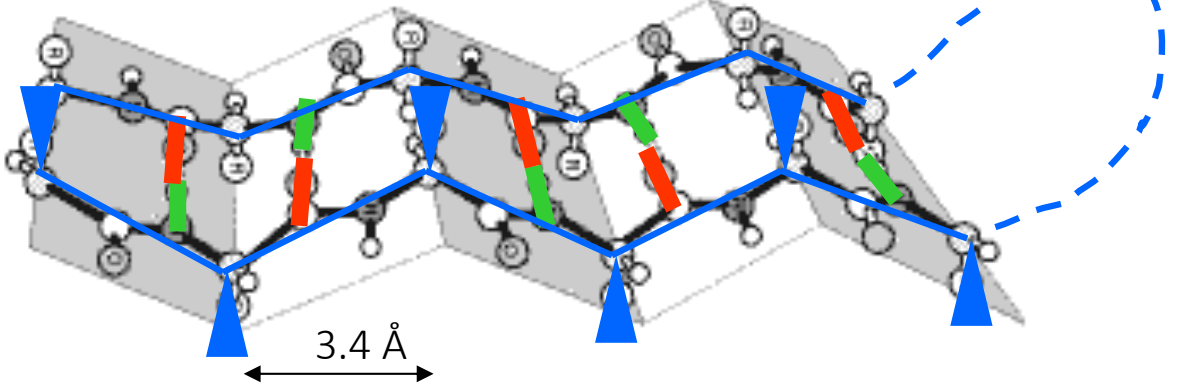
if the helix is oriented so that it goes from N (top) to C (bottom), the side chains point upwards

always right-handed



can accommodate all residues except proline

The β-sheet



anything between 2 and few hundreds amino acids

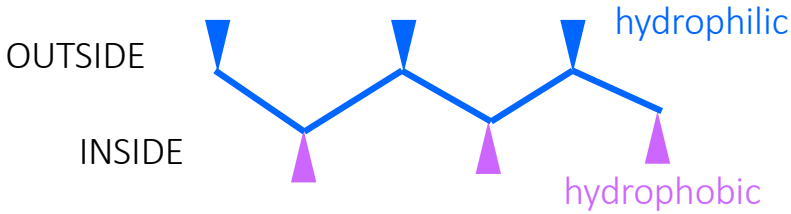
β-strand

β-strand
(zig-zag)

β-sheet

the polypeptide is almost fully extended (3.4 A per residue)

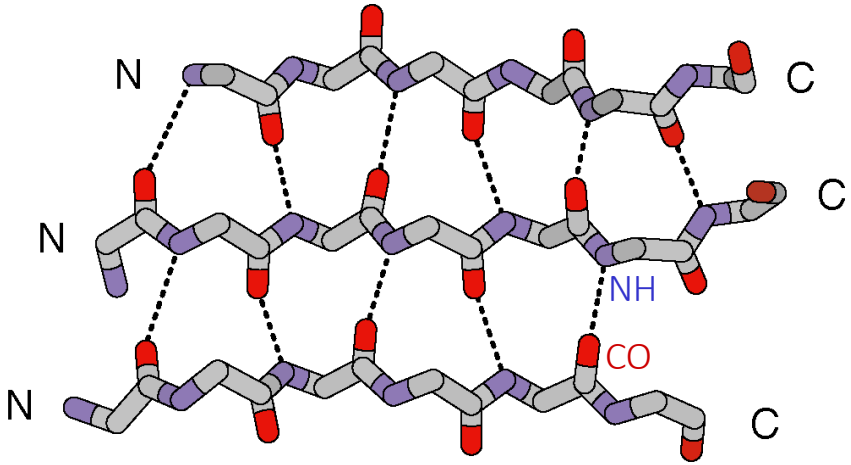
side chains points alternatively up and down

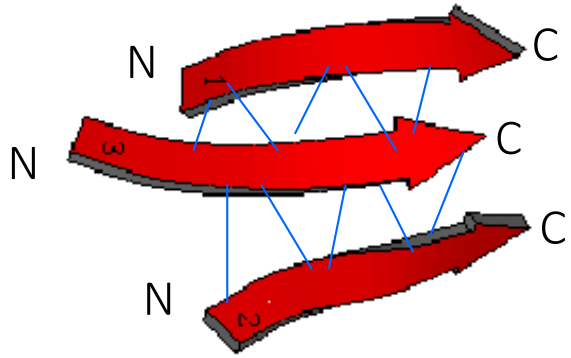


stabilised by main-chain:main-chain NH/CO hydrogen bonds between adjacent strands; contrary to the α-helix these are H bonds between NH/CO groups far apart in the amino-acid sequence

The β-sheet

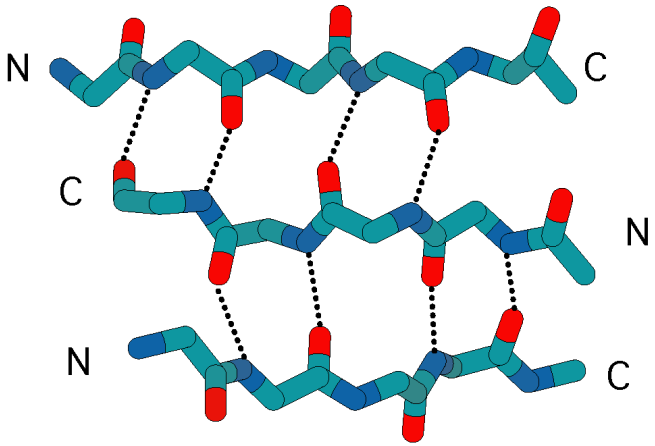
can be parallel (strands run in the same direction)

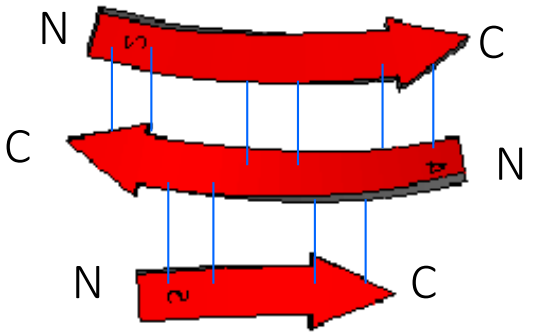




hydrogen bonding pattern

or antiparallel (strands run in opposite direction)





We often have mixed β-sheet, with some strands parallel and some antiparallel.

Tertiary structure:

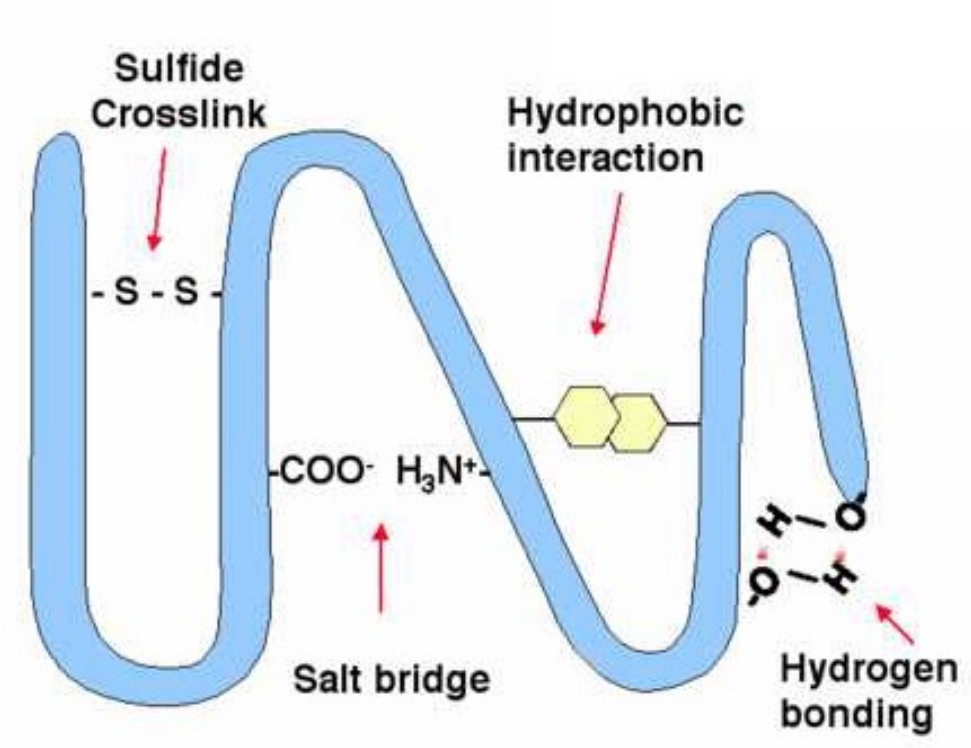
how the secondary structure elements pack together to give a 3D structure

3D structures are held together by “hydrophobic forces” and hydrogen bonds

hydrophobic side chains tend to cluster together in the interior of the protein

polar and charged amino acids interact with each other through hydrogen bonds and ionic interactions or gather on the outside of the protein where they can interact with water molecules

in some proteins S-S bonds and metal ions help to stabilise the 3D structure

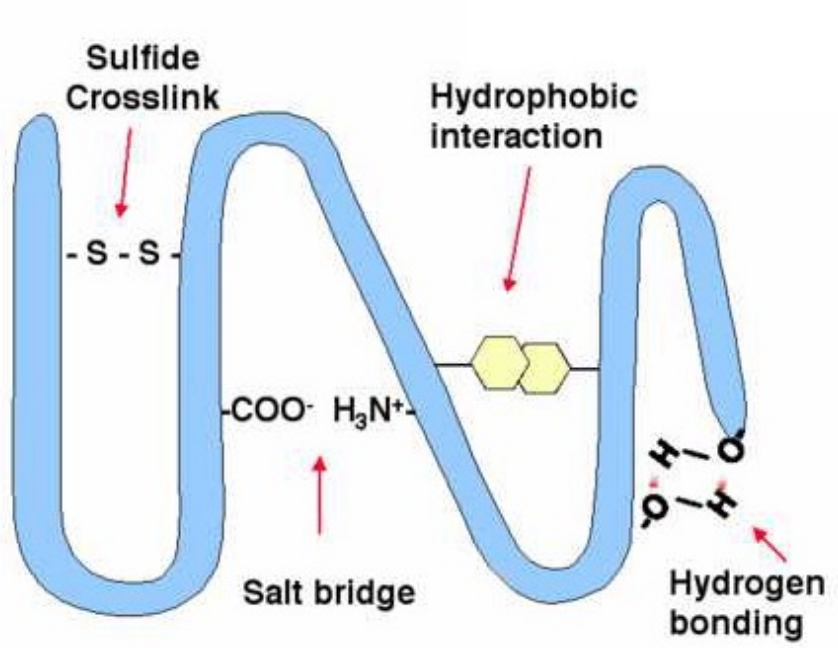


Tertiary structure:

All proteins have a well defined structure. A randomly arranged polypeptide has no biological activity

The function of a protein depends on the structure.

Proteins with similar sequences have similar structures (and similar functions), but not always the opposite is true: proteins with very different sequences can adopt similar conformations!

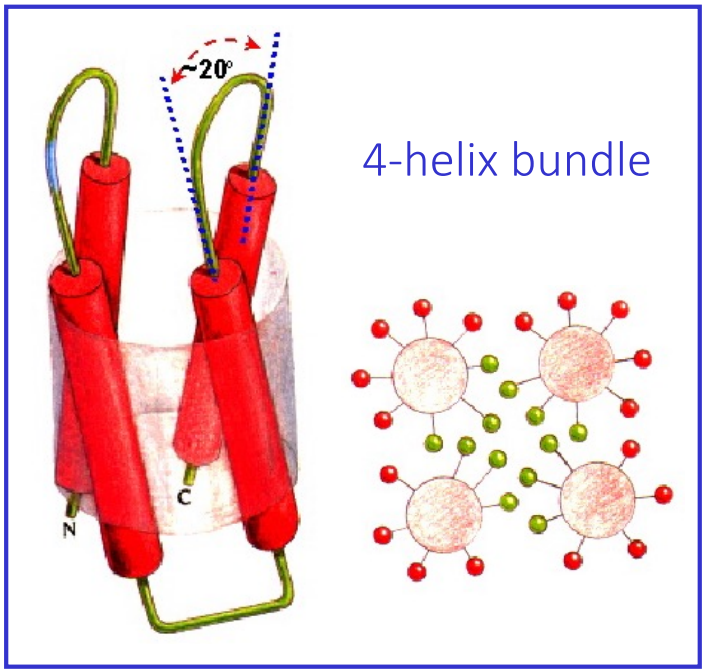


The structure is more conserved than the sequence.

Tertiary structure:
motifs in protein structures

Secondary structure elements are often connected to form structural motifs, i.e some specific geometric arrangements that occur often in protein structures; some of these motifs may be associated with certain functions, others have no specific biological function.

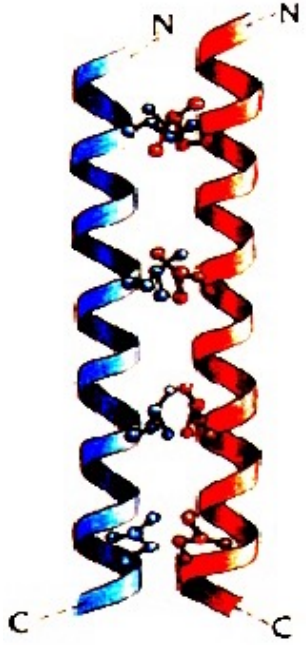
It is difficult to systematically list and classify all the motifs - here are examples of some of the common ones:



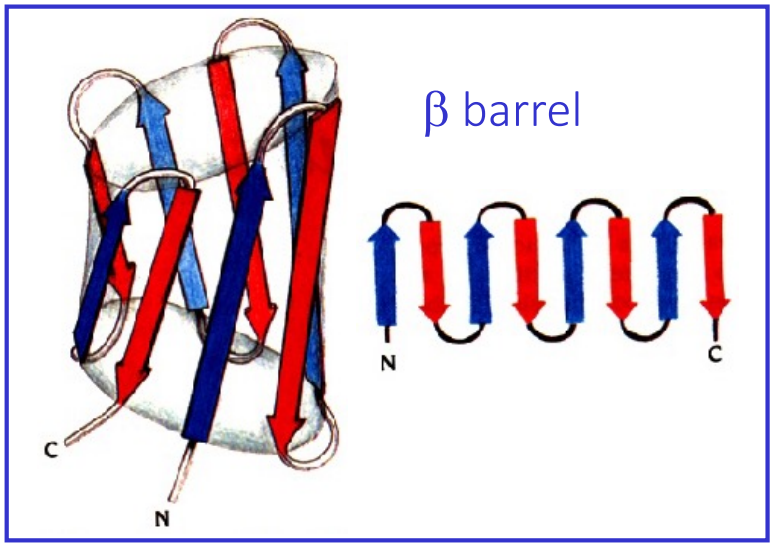
4-helix bundle

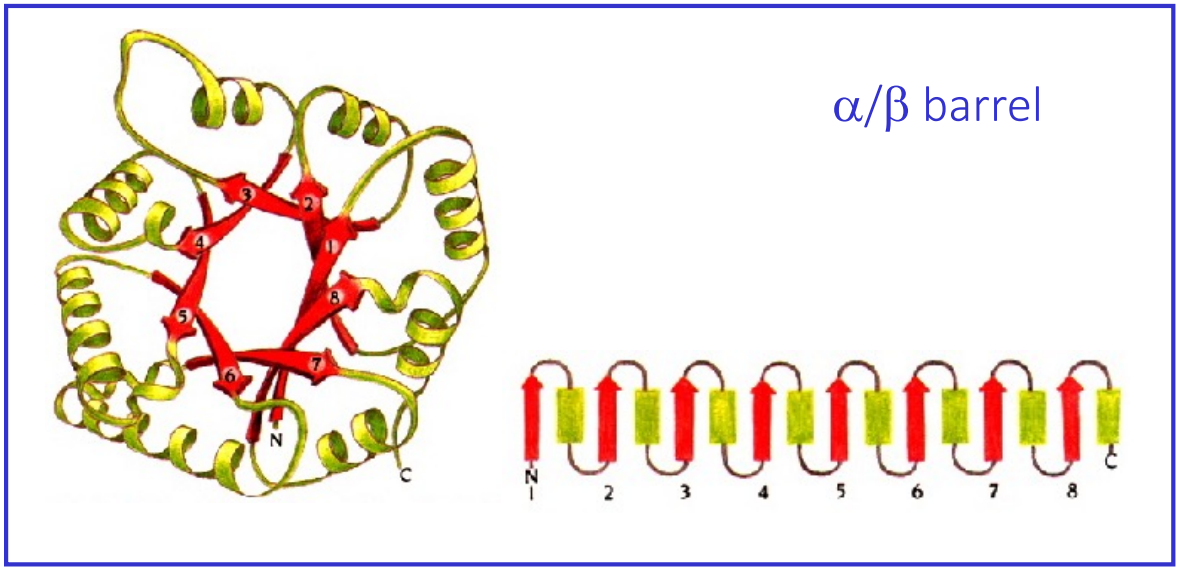
Leu zipper

a leucine residue repeated every 7 residues.



Tertiary structure:
motifs in protein structures



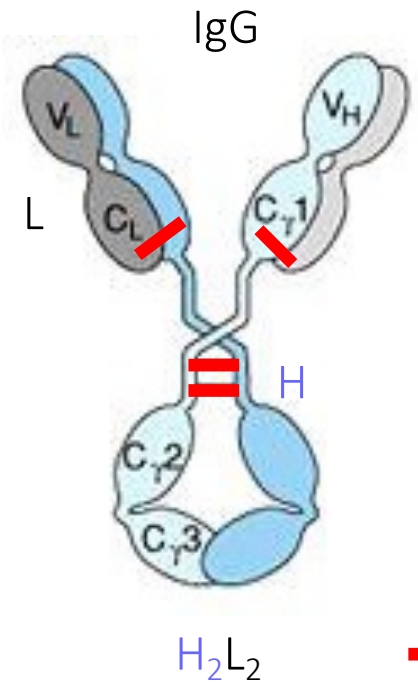


.... and many, many more!!!!

Quaternary structure:

how subunits aggregate to form multimeric proteins

Covalently-linked polypeptide chains

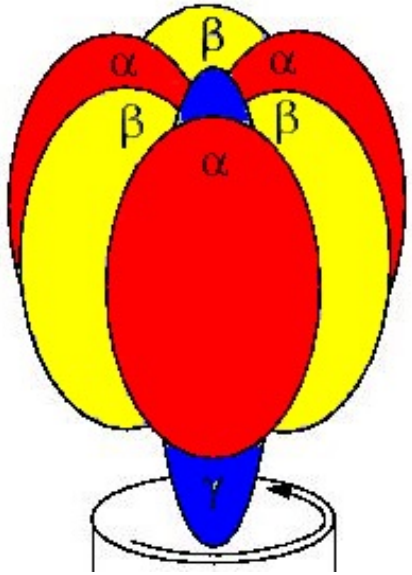


H_2L_2

For example an antibody is formed by two copies of a heavy chain H (in blue) and two copies of a light chain (in grey) connected by disulphide bridges

S-S bridges

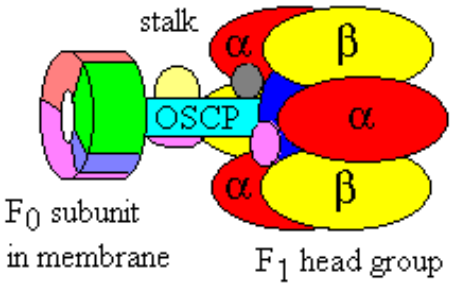
Hetero-multimers: **different** polypeptides aggregating together to form a unit.



$\alpha_3\beta_3\gamma\delta\varepsilon$

An example is the F1 head of the ATP synthase which is formed by 3 α subunits, 3 β subunit and one each of γ, ε, δ subunits.

The entire molecule is even more complex, with a transmembrane portion as well:

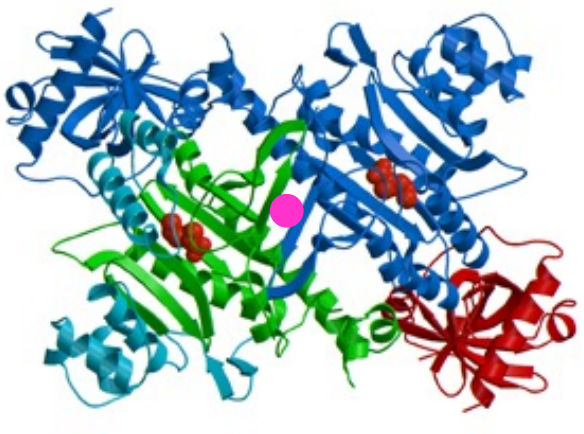


Quaternary structure:

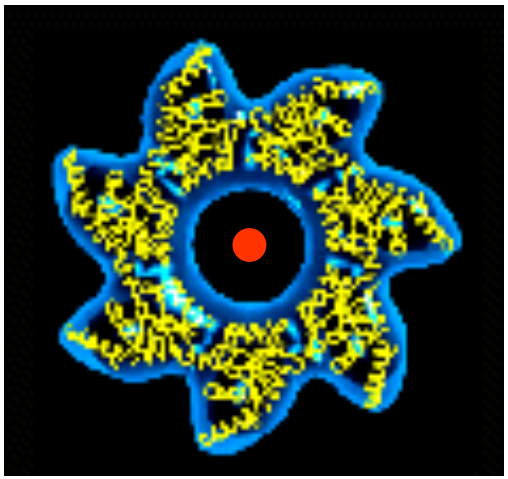
how subunits aggregate to form multimeric proteins

Homo-multimers: multiple copies of the **same** polypeptide associating non-covalently.

Such complexes usually exhibit rotational symmetry about one or more axes, forming dimers, trimers, tetramers, pentamers, hexamers, octamers, decamers, dodecamers, (or even tetradecamers in the case of the chaperonin GroEL).



Lysyl-tRNA synthetase:
● 2-fold axis



GROEL:
7-fold axis ●
72 symmetry

Quaternary structure:

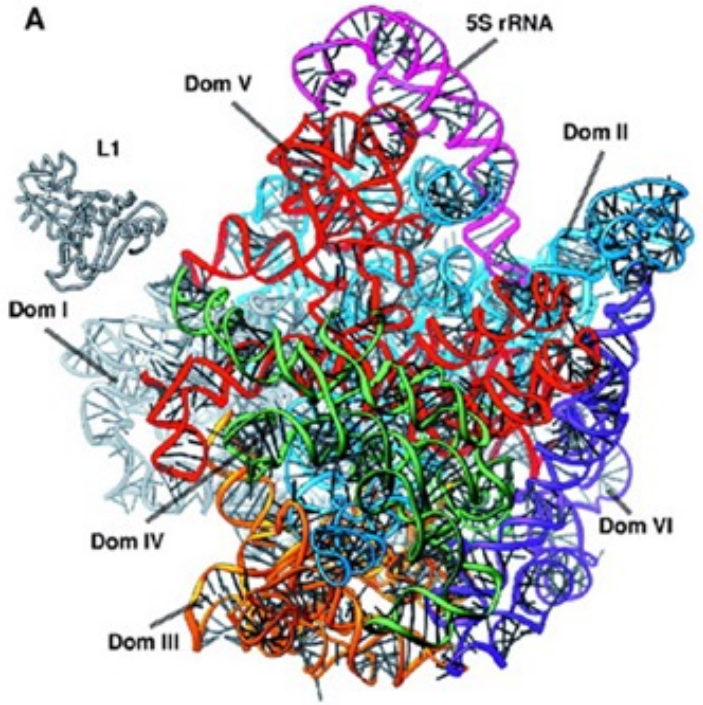
how subunits aggregate to form multimeric proteins

Larger Structures

The molecular machinery of the cell and indeed of assemblies of cells, rely on components made from multimeric assemblies of proteins, nucleic acids, and sugars. A few examples include :

- Viruses
- Microtubules
- Flagellae
- Ribosomes
- Histones

Here is the 3D structure of the large subunit of the ribosome



Noncovalent bonds and folding

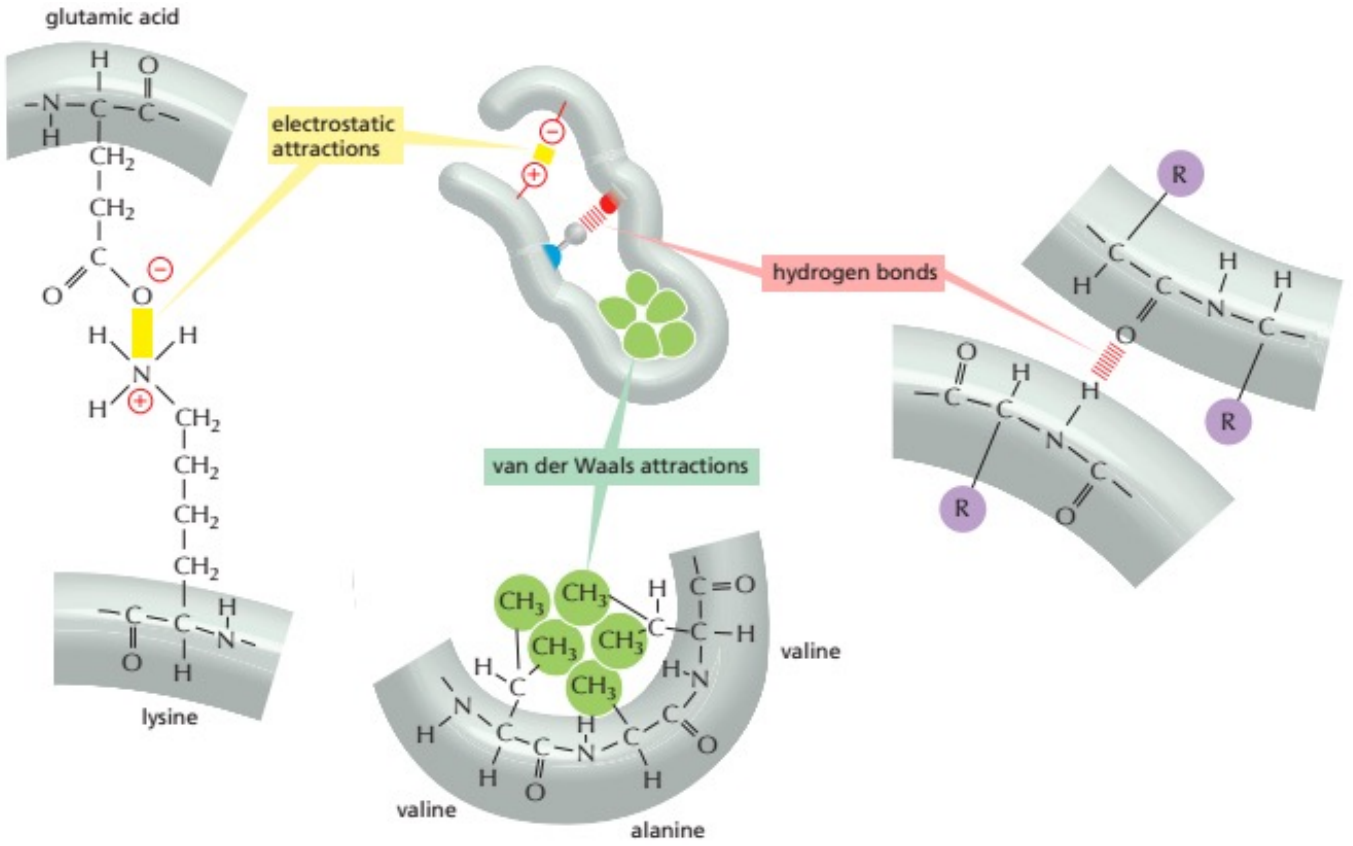


Figure 3–4 Three types of noncovalent bonds help proteins fold. Although a single one of these bonds is quite weak, many of them act together to create a strong bonding arrangement, as in the example shown. As in the previous figure, R is used as a general designation for an amino acid side chain.

Molecular forces

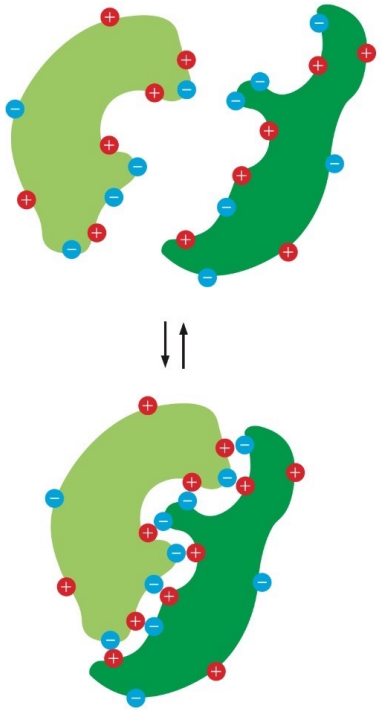
Govern how protein folds (DNA/RNA, lipid bilayer etc.) and which of its different conformations will predominate; drive ligand-macromolecules association

Covalent bonds:

- strength and direction

Non-covalent interactions:

- multipole interactions
  - ion-ion
  - ion-dipole
  - dipole-dipole ⇨ hydrogen bond
- induction interactions
- dispersion forces

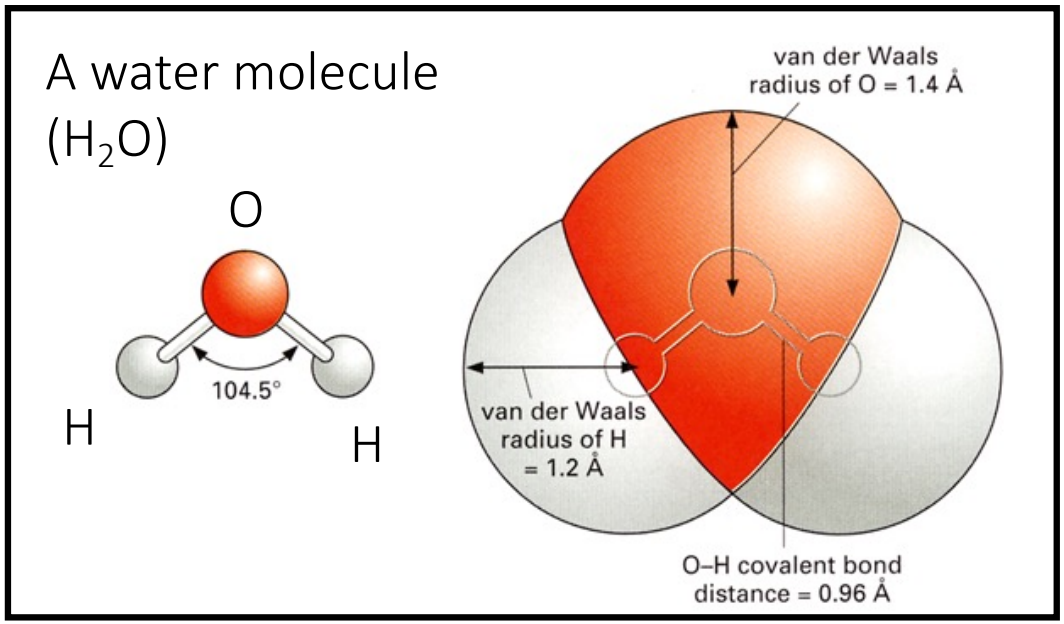


The final structure will be the result of the interplay of the different forces: complexity!

Covalent bonds

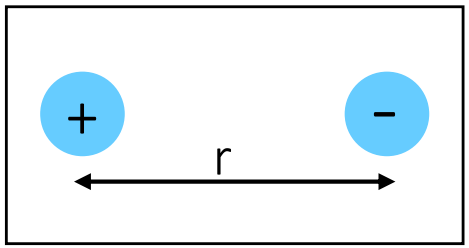
Covalent bonds are what hold “molecules” together

- strong (200-800 kJ/mol) ⇨ compare with RT~2.6 kJ/mol at 37º
- have well defined lengths
- have well defined directions



The Coulomb potential

ion-ion interactions



$$U = \frac{Q_1 Q_2}{4\pi\varepsilon_0\varepsilon_r \mathbf{r}}$$

50-350 kJ/mol

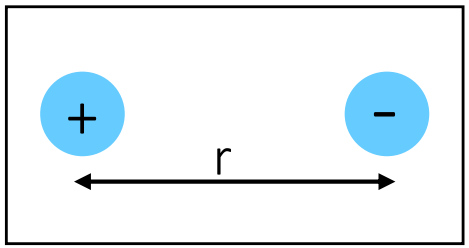
Characterizes the response of the surrounding medium to an electric field: depends on how easily the molecules are polarized

Water has a large value of ε_r (about 80). It counteracts the electric field (water mol. are highly polarizable, easily rotate)

In water ε_r is strongly T dependent decreasing by 0.46% per degree K near RT. At T= 300 K TS = -1.38 G, greater than the free energy G. Therefore, the Coulomb potential is a balance between ion-ion and ion-water molecule interaction. Ions make work on surrounding water forcing them to rotate and orient their dipoles

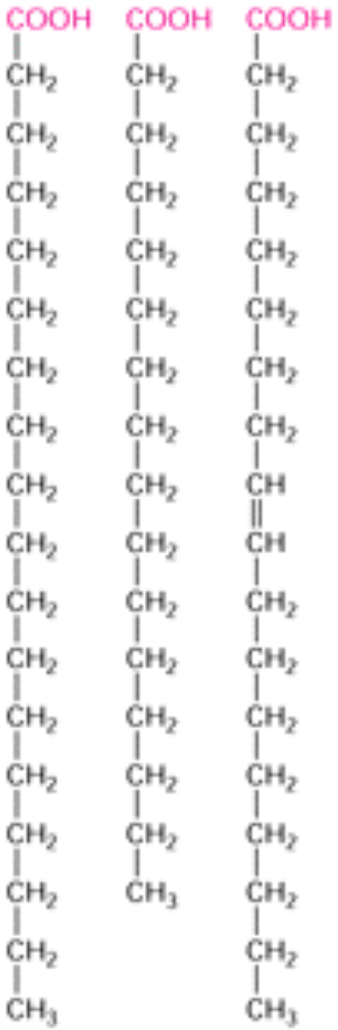
The Coulomb potential

ion-ion interactions



$$U = \frac{Q_1 Q_2}{4\pi\varepsilon_0 \varepsilon_r \mathbf{r}}$$

50-350 kJ/mol



characterizes the response of the surrounding medium to an electric field: depends on how easily the molecules are polarized

Hydrocarbons have ε_r of 2: the hydrophobic core of proteins and membranes experiences strong electrostatic interactions

1 kcal/mol = 4.2 kJ/mol = 0.043 eV

Electrostatic self-energy

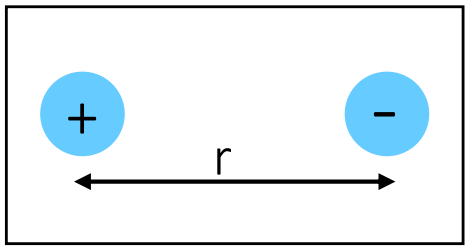
$$G = \frac{1}{\varepsilon_r r}\int_0^q q'dq' = q2/2\varepsilon_r r$$

Is the self-energy of a charge, or the energy of placing an ion in a dielectric medium (calculated from the work done to bring an increment dq' on the surface of a sphere with radius r and charge q')

For water, it is the **hydration energy**.
To transfer a Na+ ion with r = 0.95 Å from water to an hydrocarbon medium (ε goes from 80 to 2), the work necessary is of 85 kcal/mol.
In fact inorganic ions are generally insoluble in organic solvents.
It is difficult to move an ion inside a protein of a lipid bilayer!
Ions are always attracted towards the region with higher ε

Multipole interactions

ion-ion interactions

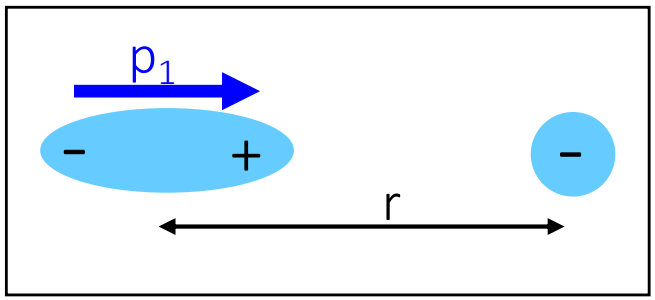


$$\frac{Q_1 Q_2}{4\pi\varepsilon_0\varepsilon_r \mathbf{r}}$$

50-350 kJ/mol

Even in neutral molecules, dipoles result from the unequal distribution of e^- due to differences in electronegativity between atoms.

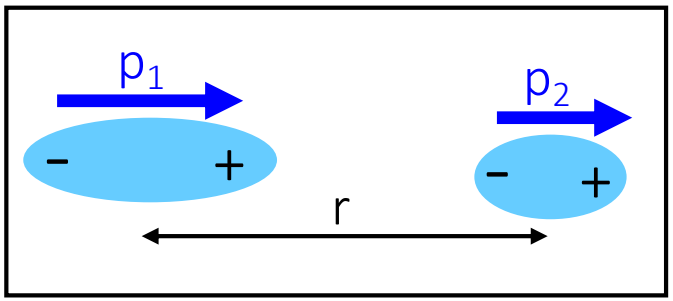
ion-dipole interactions



$$\frac{Q_1 p_1}{4\pi\varepsilon_0\varepsilon_r \mathbf{r}^2}$$

1-50 kJ/mol

dipole-dipole interactions



$$\frac{p_1 p_2}{4\pi\varepsilon_0\varepsilon_r \mathbf{r}^3}$$

0.1-10 kJ/mol

Induction forces

Ions and dipoles can polarise the electron cloud of an adjacent molecule. This causes an attractive force between the ion/permanent dipole and the induced dipole.

Interaction proportional to
- **r^{-4} for ion-induced dipole**
- **r^{-6} for permanent dipole/induced dipole interactions**

Dispersion forces

Random fluctuations of the electron clouds cause temporary dipoles even in uncharged molecules; these temporary dipoles will induce dipoles in the adjacent molecules causing a weak attractive force (He liquefies at 4K).

Van der Waals attractive forces!

0.4 kJ/mol—0.35 nm bond length
Does not change in water!!

Dispersion forces

Fluctuactions of transient dipole moments can be attractive or repulsive. The attractive configurations have a lower potential E than the repulsive ones, meaning have larger weights in Boltzmann average and therefore a net attraction.

The fluctuactions in the electronic structure responsible for the transient dipole moments are much faster than molecular rotation in liquids. Therefore such forces are not dependent on the specific medium.

Hydrophobic forces

Hydrophobic forces are very relevant in biology. They are primarily driven by an energy cost of creating hydrocarbon-water contact.
There is a reduction of entropy of water close of a hydrophobic surface: water becomes structured, even ice-like. It restricts the possible orientations close to the surface and decrease entropy.

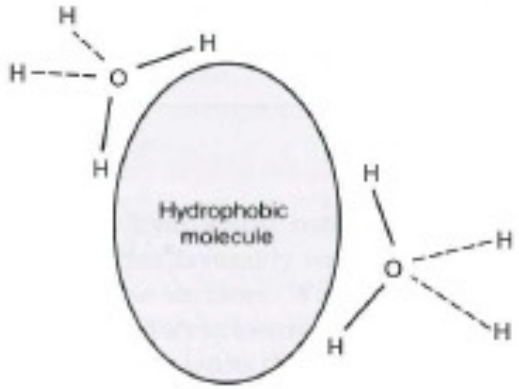
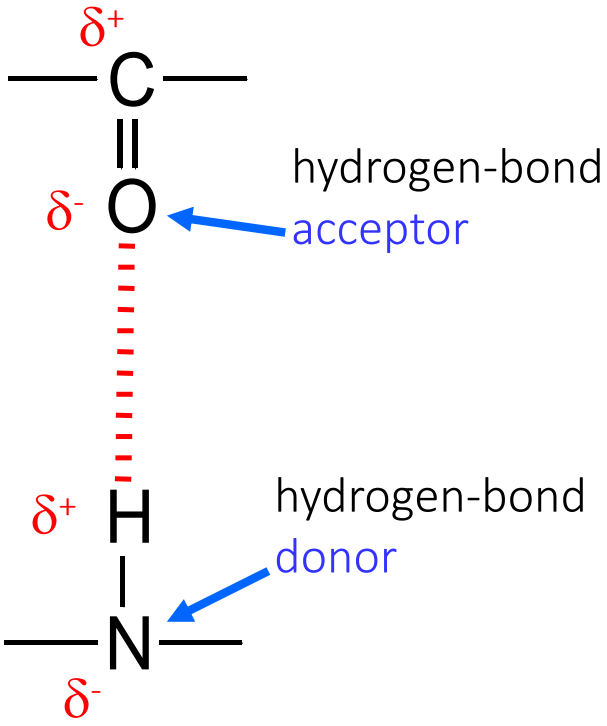


Fig. 2.7 Water molecules adjacent to a hydrophobic molecule suffer restrictions in orientation as they form hydrogen bonds with other water molecules.

Hydrogen bond

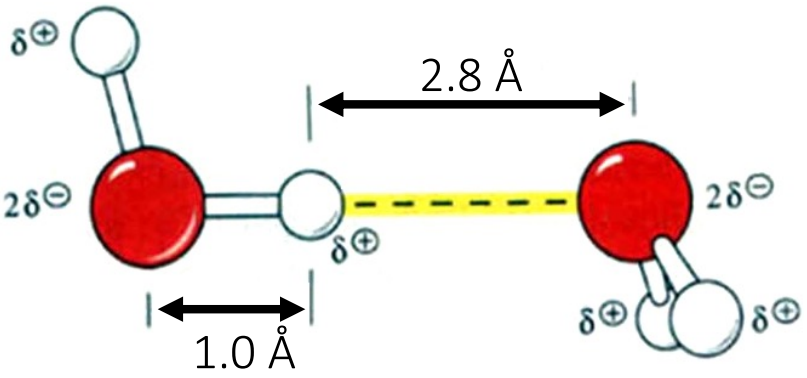
Hydrogen bonds are a particular case of a dipole-dipole interaction, unusually strong because the small size of the H atom allows the dipoles to come close to each other (~15-30 kJ/mol)

17 kJ/mol—**0.30 nm** bond length
Becomes 4.2 kJ/mol in water!!



Donors and acceptors must be electronegative atoms (O, N)

Hydrogen bonding in water:

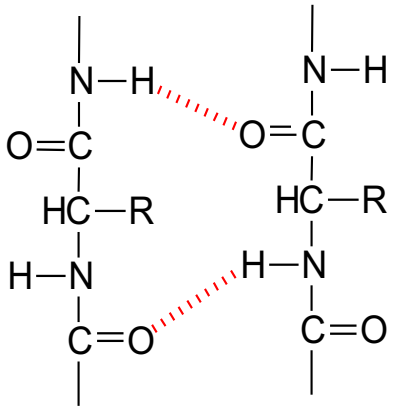


Hydrogen bonds have a defined lenght and orientation

Hydrogen bonds in biology

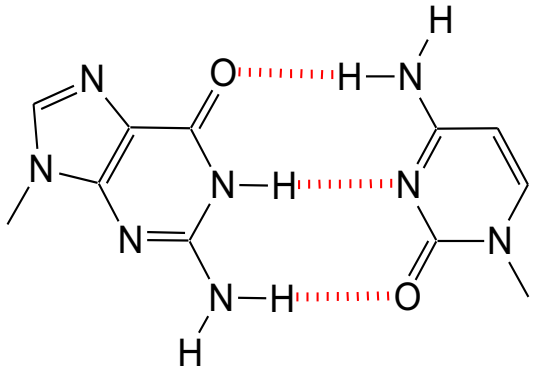
Hydrogen bonding interactions play a fundamental role in determining both the conformation of biological macromolecules and their interactions with other molecules.

The 3D structures of proteins are stabilized by hydrogen bonds between main-chain amide groups:



protein secondary structure: a β-sheet

The pairing of the bases in DNA is mediated by H-bonds:



Guanine-Cytosine base pair

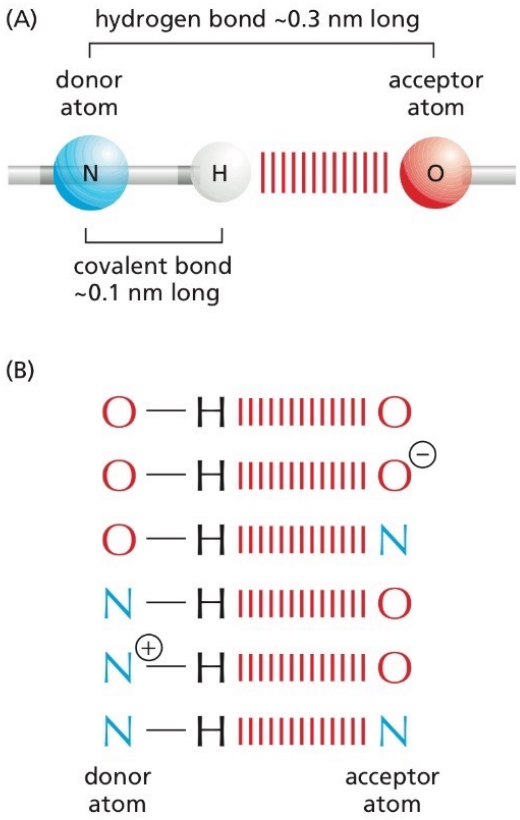


Figure 2–4 Hydrogen bonds. (A) Ball-and-stick model of a typical hydrogen bond. The distance between the hydrogen and the oxygen atom here is less than the sum of their van der Waals radii, indicating a partial sharing of electrons. (B) The most common hydrogen bonds in cells.

TABLE 2–1 Covalent and Noncovalent Chemical Bonds

Bond type		Length (nm)	Strength kJ/mole**	
			in vacuum	in water
Covalent		0.15	377 (90)	377 (90)
Noncovalent	ionic*	0.25	335 (80)	12.6 (3)
	hydrogen	0.30	16.7 (4)	4.2 (1)
	van der Waals attraction (per atom)	0.35	0.4 (0.1)	0.4 (0.1)

*An ionic bond is an electrostatic attraction between two fully charged atoms. **Values in parentheses are kcal/mole. 1 kJ = 0.239 kcal and 1 kcal = 4.18 kJ.

Fibrous proteins

Triple helix in collagen - next

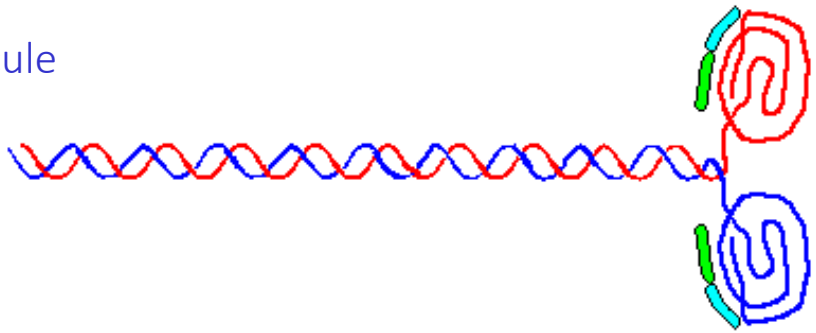
Coiled-coil α-helices present in keratin and myosin:

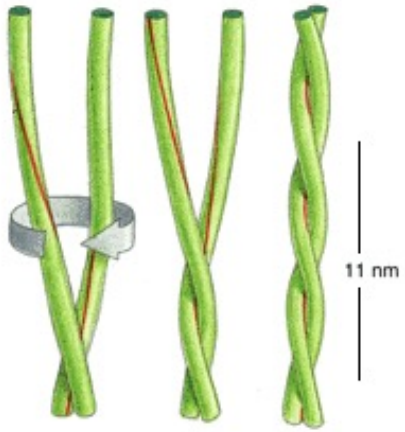
two a-helices twisted around each other to form a left handed coiled coil (7 residues repeat)

Example:a myosin molecule

heavy chains

light chains





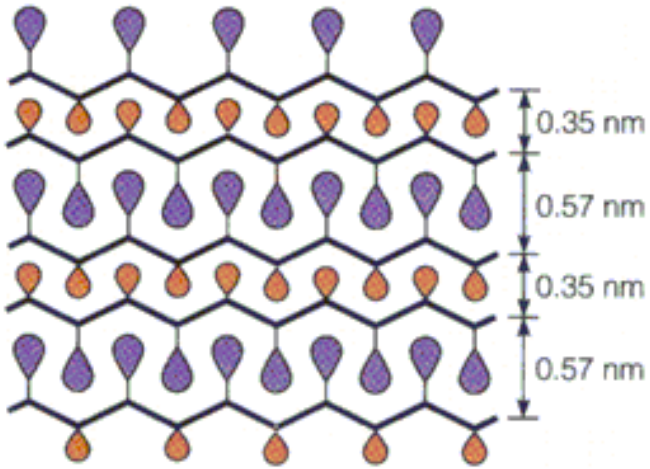
coiled-coil helices

β-sheets in amyloid fibres, spider webs and silk

antiparallel β-sheet whose chains extend parallel to the fibre axis

Ala or Ser

Gly

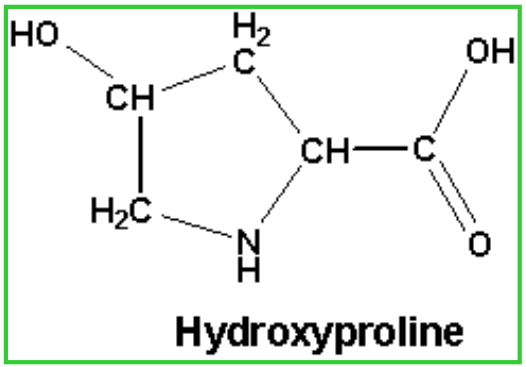


Fibrous proteins: the collagen helix

Collagens are family structural proteins forming the tendons and the extracellular matrix. Bones and teeth are made by adding mineral crystals to collagen.

Collagen is composed of three chains wound together in a triple helix.

Each chain is very long and consists of a repeating sequence of three amino acids: every 3rd amino acid is a glycine that fits in the interior of the triple helix; many of the remaining positions contain prolines and hydroxyprolines:



The enzyme that modifies a proline into hydroxyproline requires vitamin C; lack of vitamin C causes scurvy.

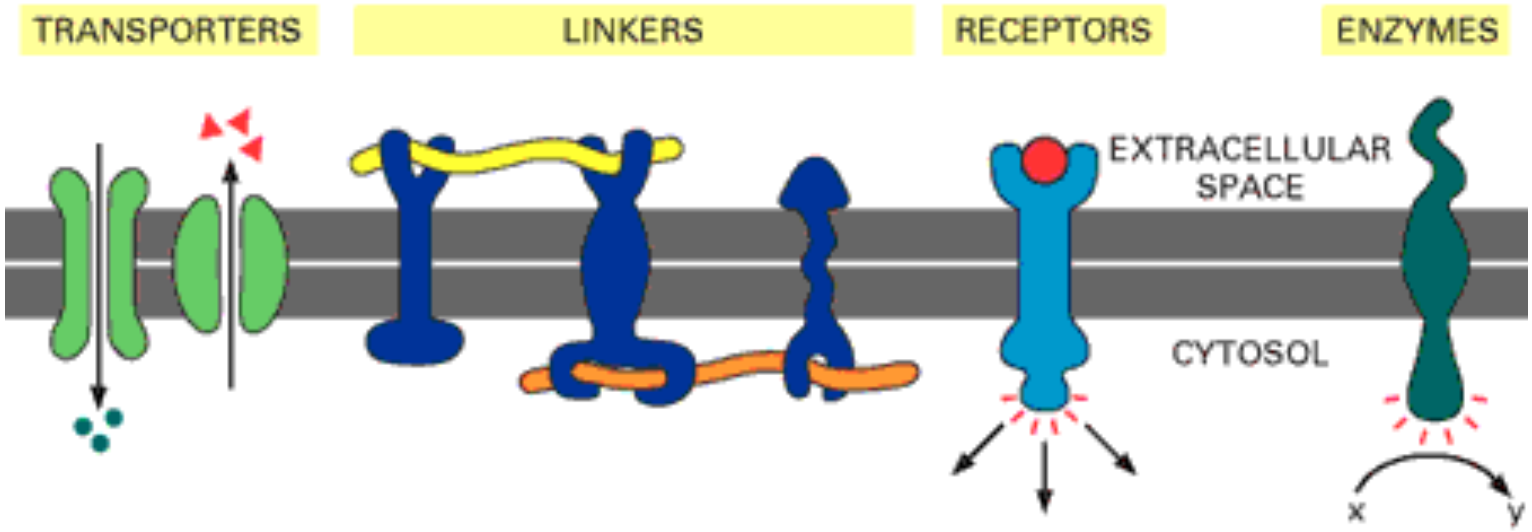
There are other non-standard aa (such as hydroxylysines) which are used to crosslink the chains.

Membrane proteins: biological roles

Membrane proteins are defined as proteins that sit in the lipid bilayer: they perform very different biological roles:

- pumps
- channels
- receptors
- cell-to-cell adhesion

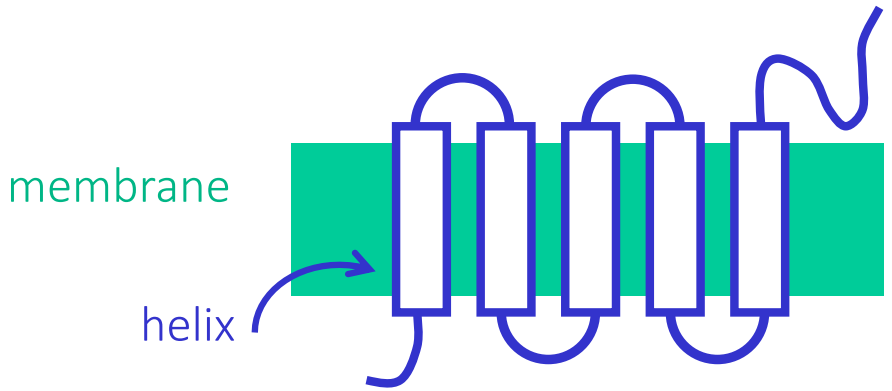
control the flow of chemicals and information between the inside and the outside of the cell and mediate communication between different cells.



Structures of membrane proteins

Less is known about the 3D structure of membrane proteins since in general they are much more difficult to crystallise than soluble proteins.

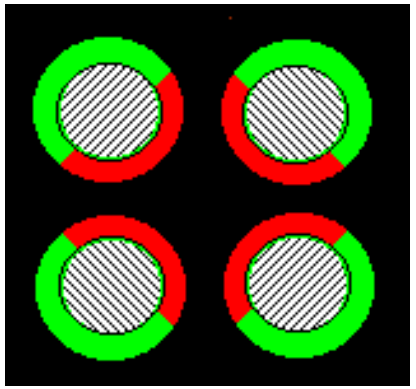
They are often built of α-helices spanning the membrane; but some are built of extended β-barrels (such as porins)



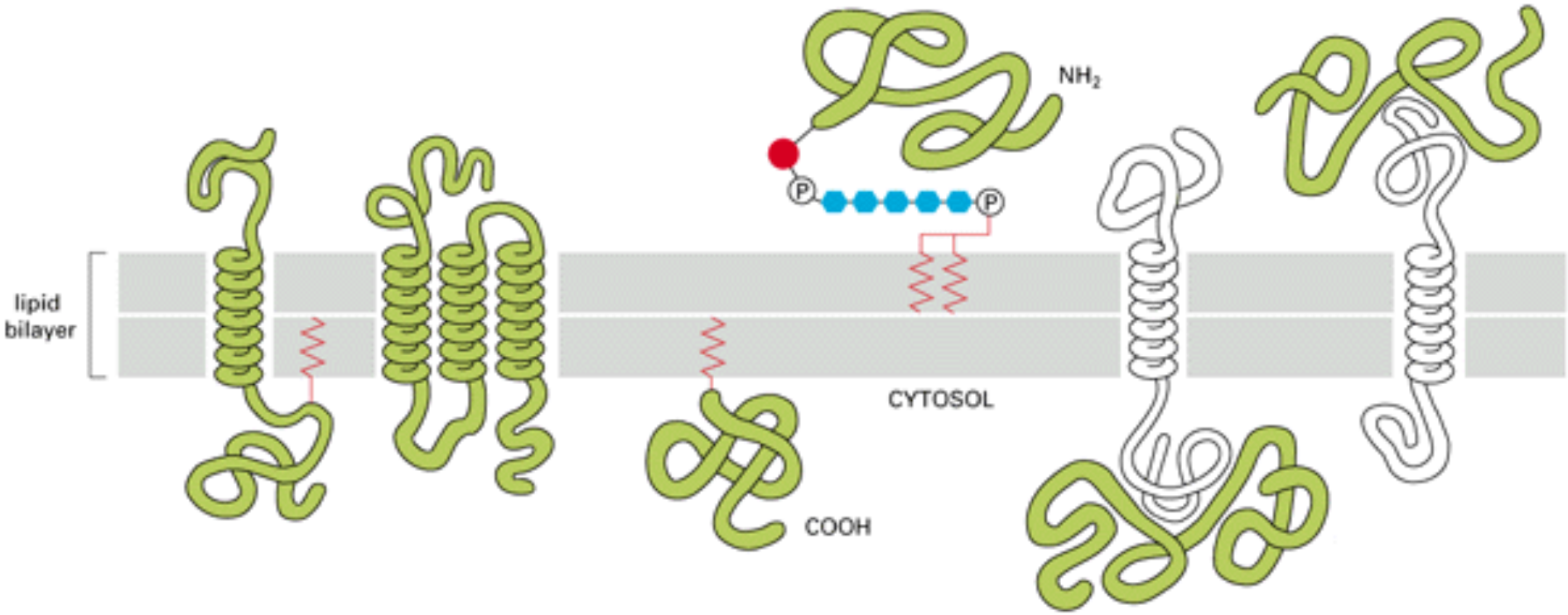
Contrary to soluble proteins, the hydrophobic residues will be on the outside, where they will interact with the chains of the lipids, while hydrophilic side chains will cluster inside

hydrophilic

hydrophobic



Membrane proteins associate with the lipid bilayer in various ways:

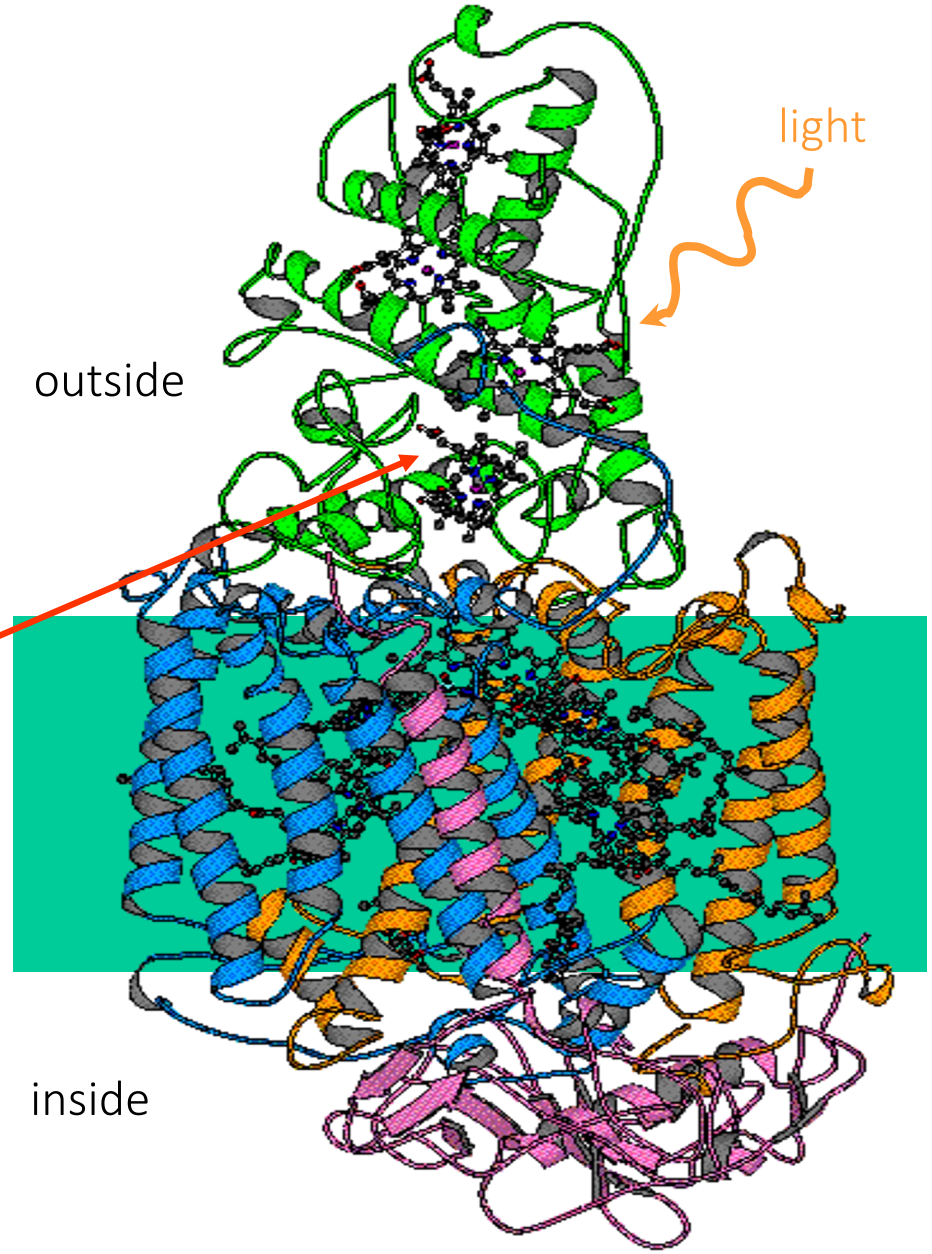


Examples of membrane proteins: the photosynthetic reaction centre

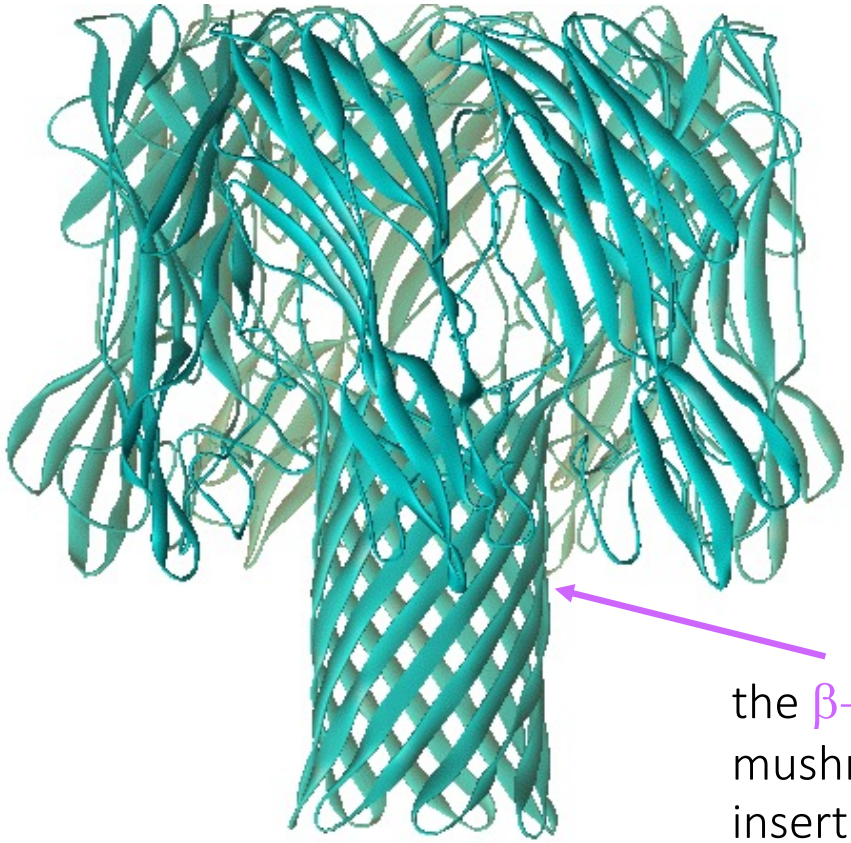
Structure determined by R. Huber, H. Mitchell & H. Deisenhofer (Nobel prize 1988)

Found in the membranes of chloroplasts and in photosynthetic bacteria; convert energy from the sun into electrical and chemical energy.

Contains a lot of pigments (such as chlorophylls, quinons, carotenoids, etc...) to capture photons

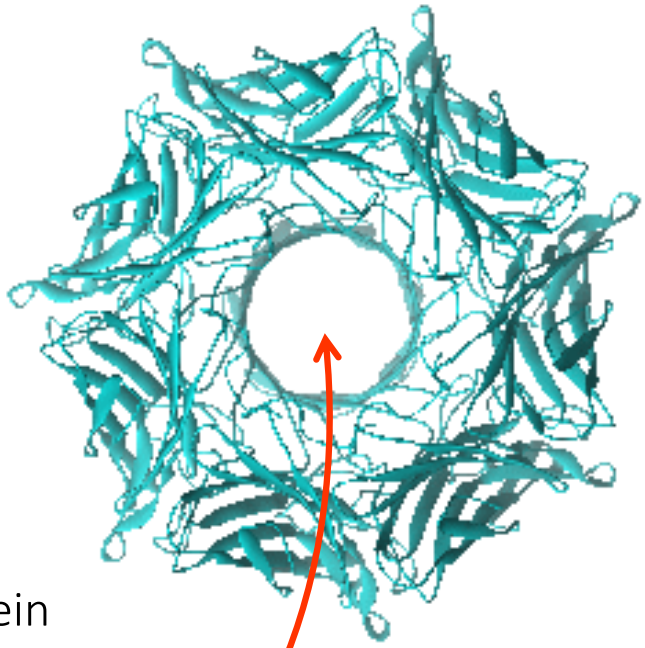


Examples of membrane proteins: the bacterial toxin a-hemolysin



seen from the side

seen from the top

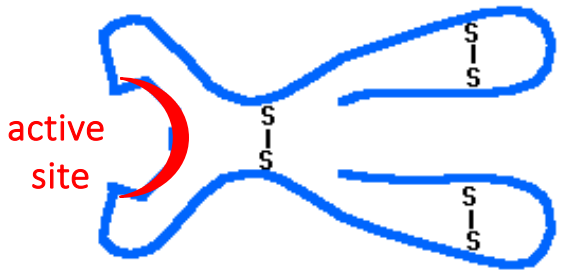


the β-barrel stalk of this mushroom shaped protein insert across the cell membrane and causes lethal permeability changes due to the central pore

Denaturation

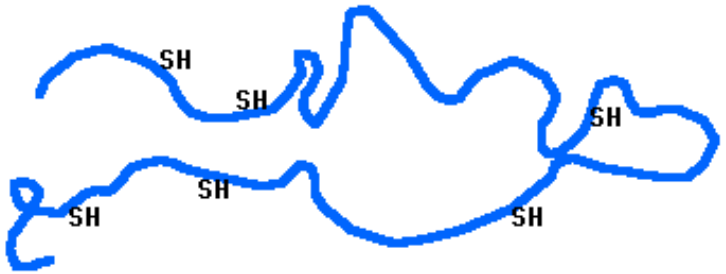
Many proteins can be unfolded and refolded:

FOLDED, active protein



urea, pH, high T

UNFOLDED, denatured protein



refold = regain biological activity

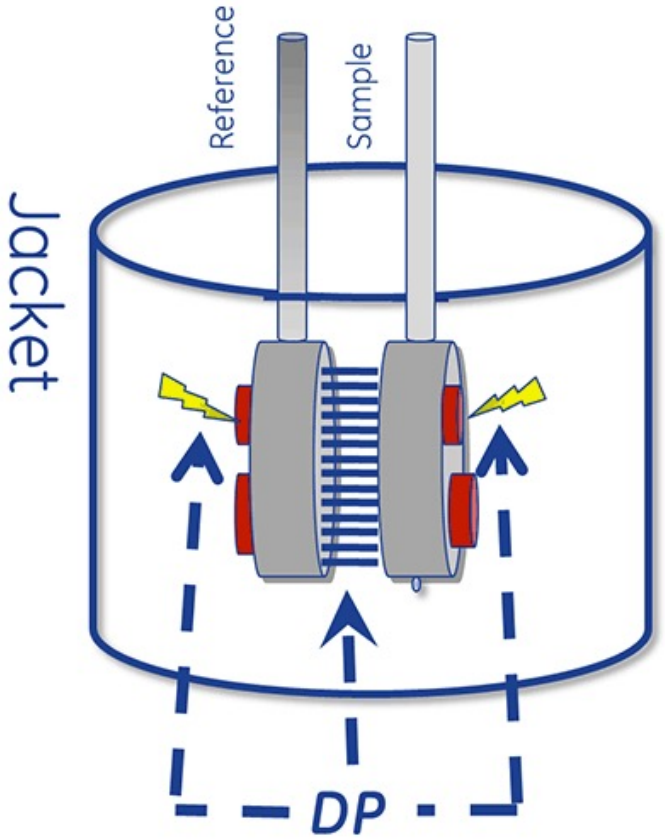
unfold = loose biological activity

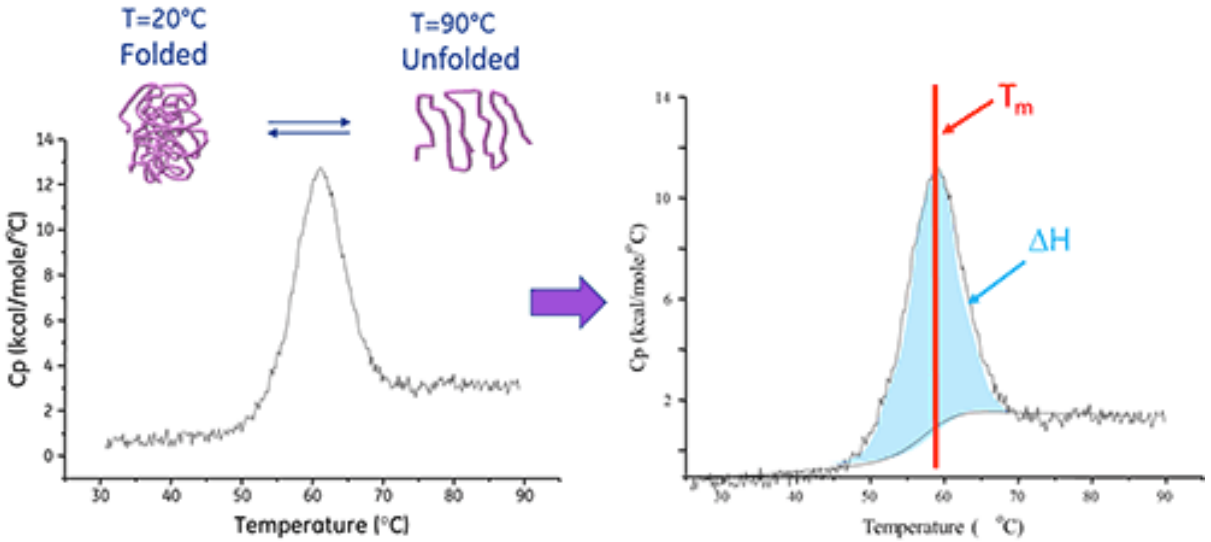
When a protein contains disulphide bridges (S-S)
we also need to reduce them to SH to unfold.

It does not work for all proteins - some proteins, once unfolded cannot be easily refolded again.

Denaturation

Many proteins can be unfolded and refolded: differential calorimetry





Summary

Amino acids: structure and properties

The peptide bond: polarity of the bond, concept of primary sequence, planarity and torsion angles.

Secondary structure: the α-helix and the β-sheet.

Tertiary structure and motifs.

Quaternary structure and multimeric proteins.

Fibrous proteins: coiled-coil, amyloid fibers, collagen triple helix.

Membrane proteins.

The folding problem.