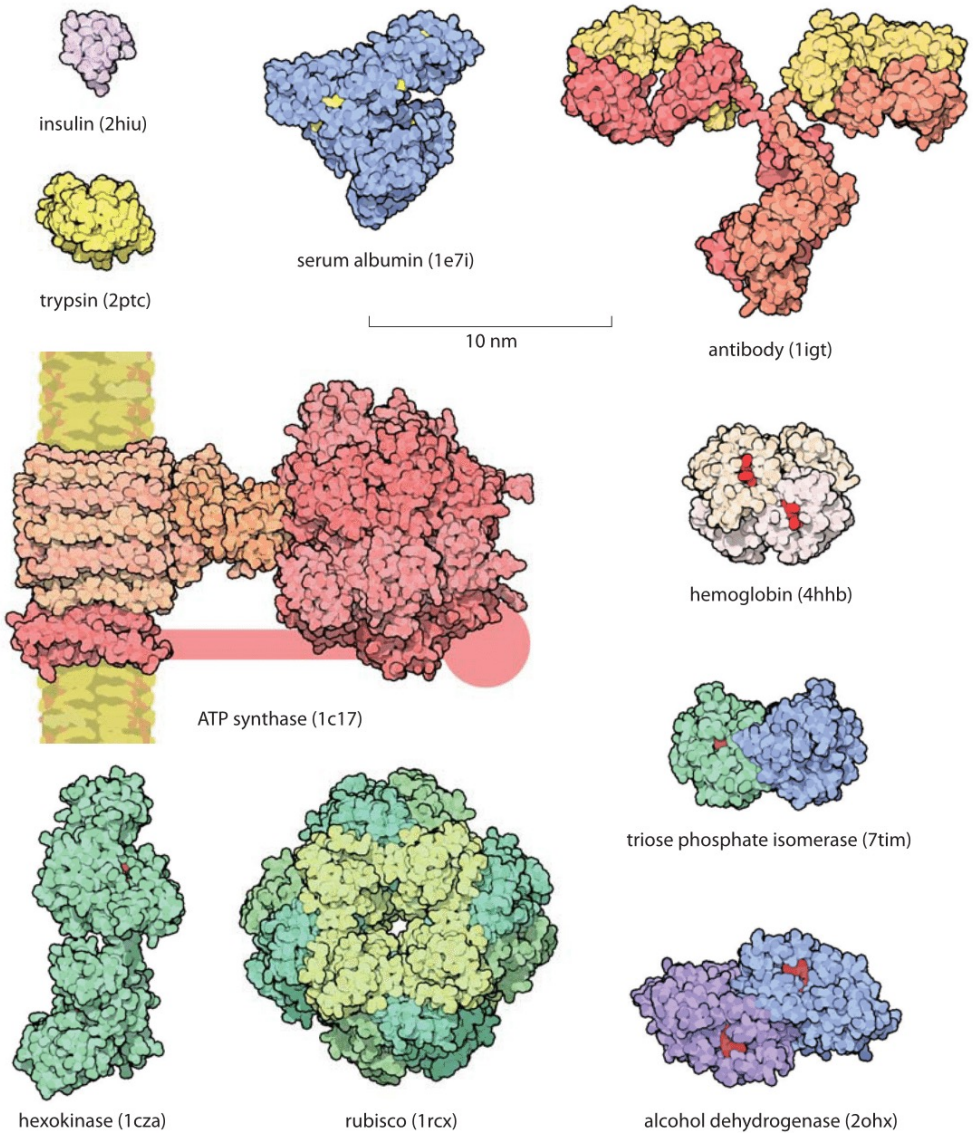
Amino acids and proteins

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

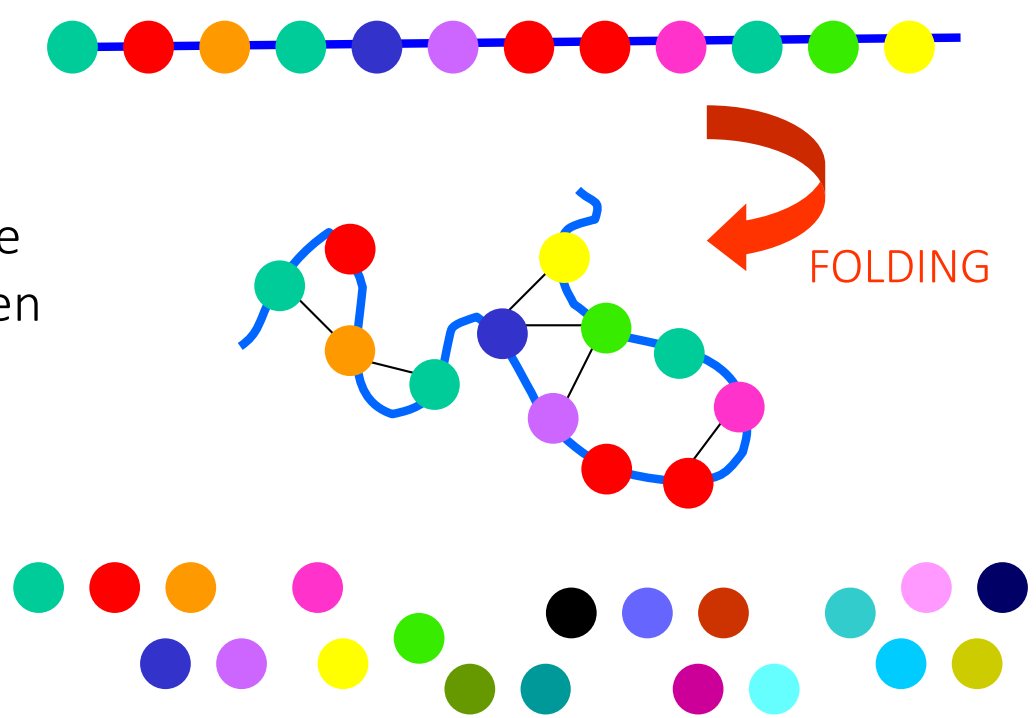
http://book.bionumbers.org/how-big-is-the-average-protein/

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

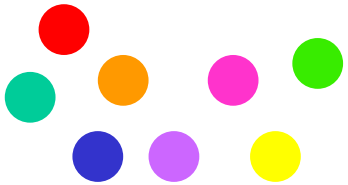
There are 20 different types of amino acid, each with different physico-chemical properties.



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

Overview of protein architecture

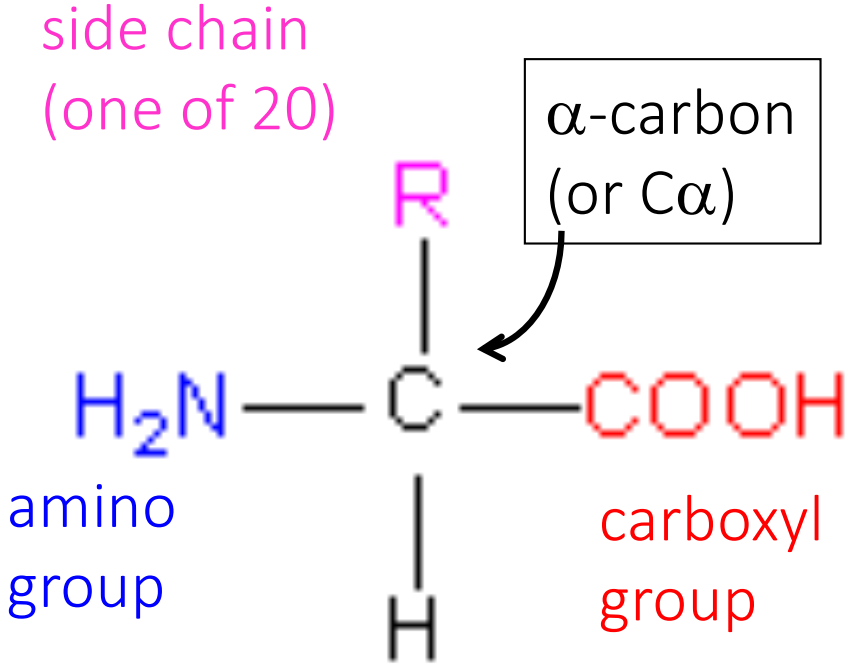
1) structure and chemistry of amino acids



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

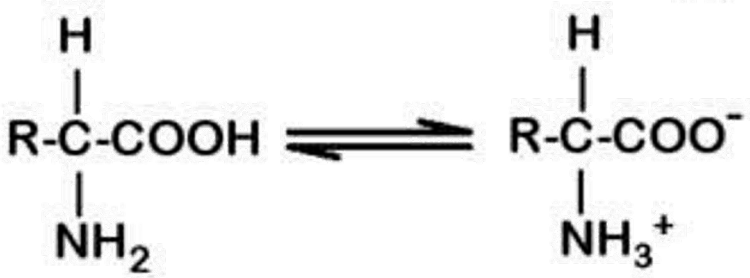
Structure of amino acids



At neutral pH:

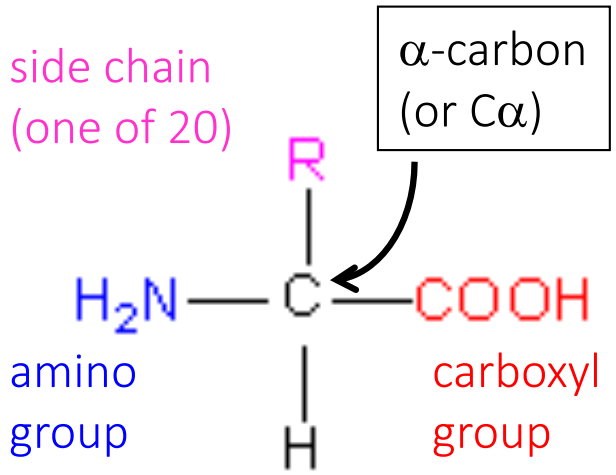
NH_2 is a base ⟶ NH_3^+

COOH is an acid ⟶ COO^-

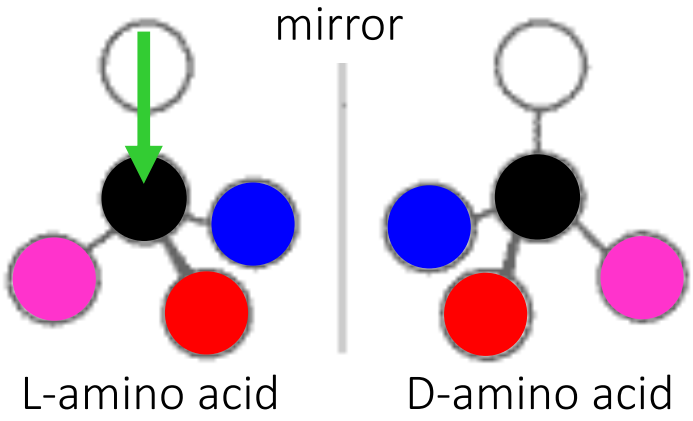


zwitterion
(dipolar form)

Structure of amino acids

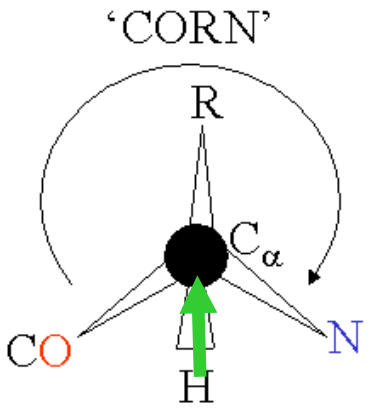


The Cα is an asymmetric carbon (bound to 4 different groups) and therefore is a chiral centre.
Two configurations (stereoisomers) are possible, which are one the mirror image of the other:

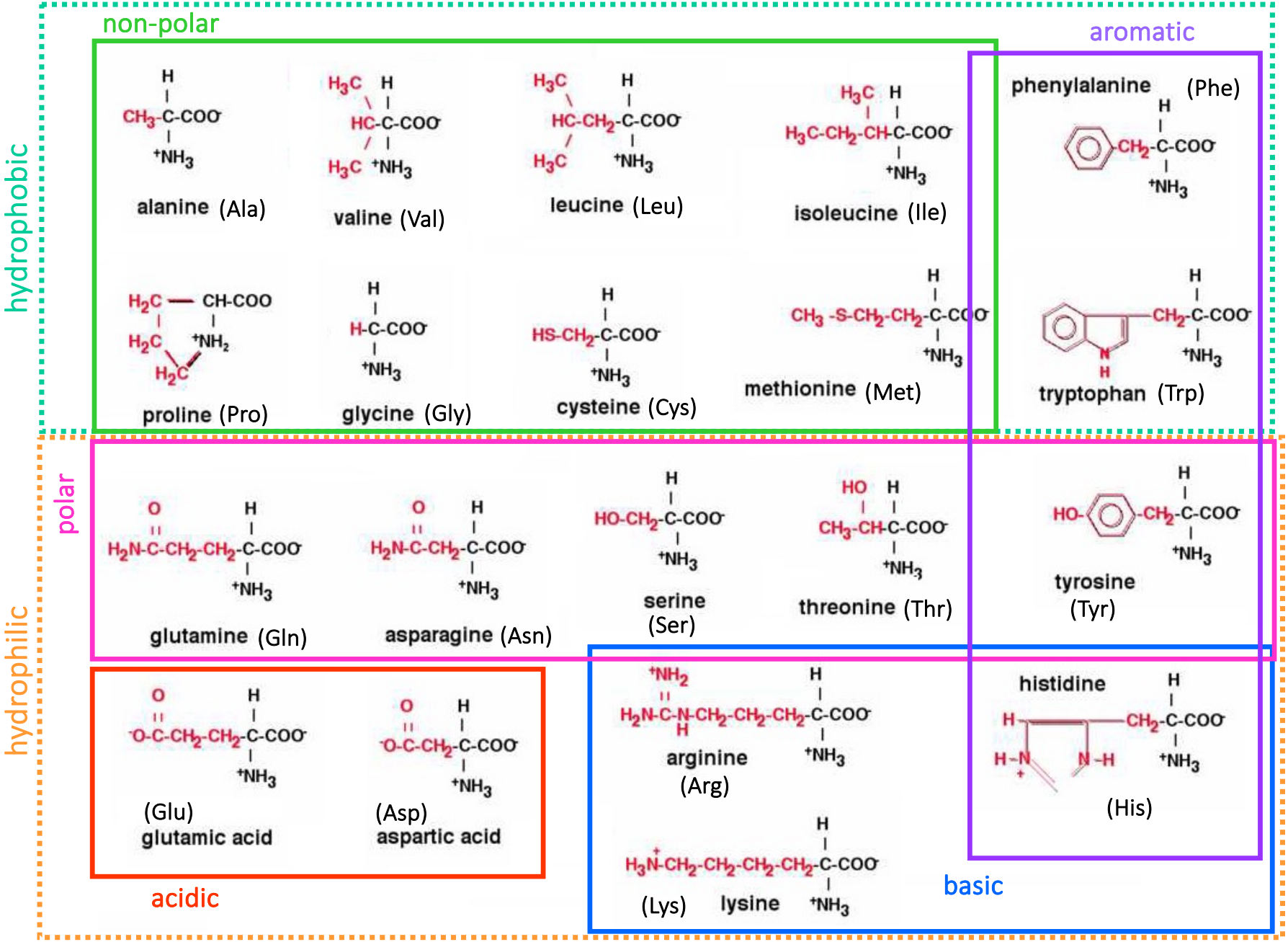


all amino acids in proteins are L!!

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

Charged side chains

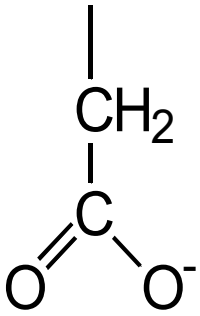
at neutral pH

negative charge

positive charge

Aspartic acid

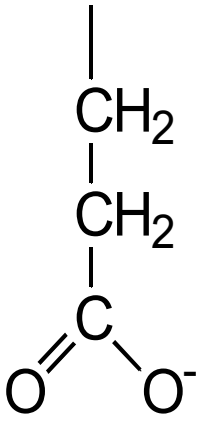
(Asp or D)



pK_a=3.9

Glutamic acid

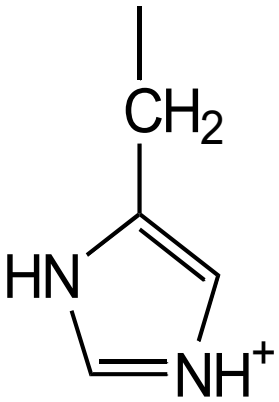
(Glu or E)



pK_a=4.2

Histidine

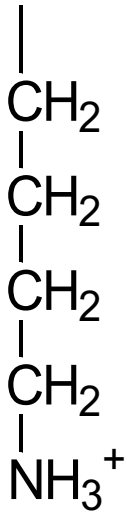
(His or H)



pK_a=6.1

Lysine

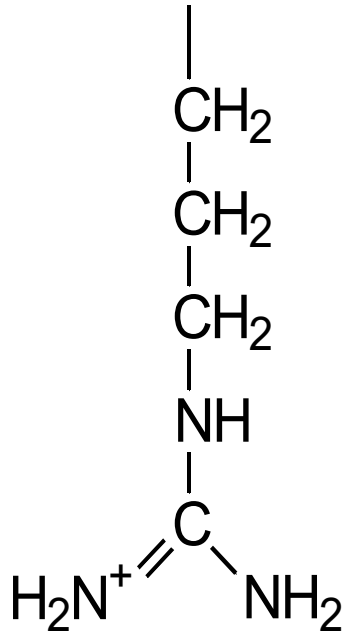
(Lys or K)



pK_a=10.8

Arginine

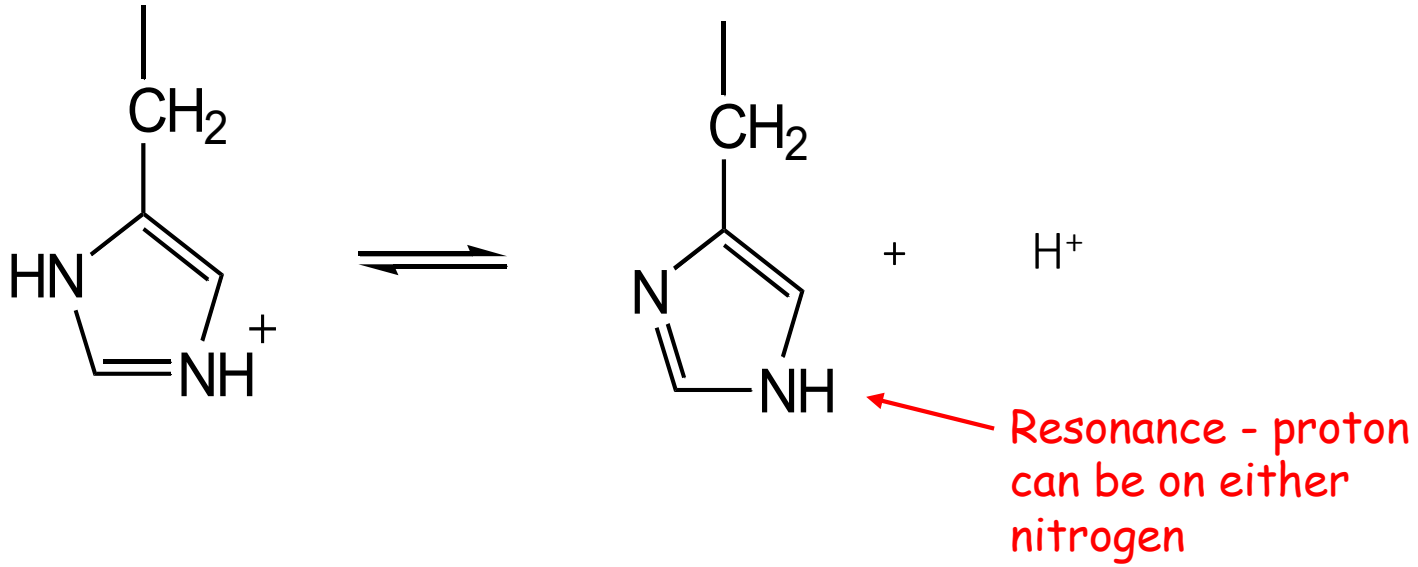
(Arg or R)



pK_a=12.5

Histidine

The side chain of a histidine residue is uncharged at high pH



$pK_a = 6.1$

Charge on His residue in a protein at neutral pH will depend on the local environment - small shifts of pH or local environment can change the charge of a His (important in enzyme mechanism)

pH

The acidity of a solution is measured on a "pH" scale where:

$$pH = -\log_{10}[H^+]$$

For pure water $[H^+] = 10^{-7}$ M and thus the pH = -log(10^{-7}) = 7

pH < 7.0 then [H+] > [OH-] → solution is acidic

pH > 7.0 then [H+] < [OH-] → solution is basic (or alkaline)

The interior of a cell is kept close to neutrality by the presence of buffers: weak acid and bases that can release or take up protons near pH 7, keeping the environment of the cell relatively constant under a variety of conditions.

pK_a

For an acid:

$$HA \rightleftharpoons H^+ + A^-$$

and the dissociation constant is

$$GH_2^+ \Leftrightarrow GH + H^+$$

$$GH \Leftrightarrow G^- + H^+$$

K is an equilibrium constant (in this case an acid dissociation constant)

where [X] indicates the molar concentration of X

We define:

$$pK_a = -\log_{10}(K_a)$$

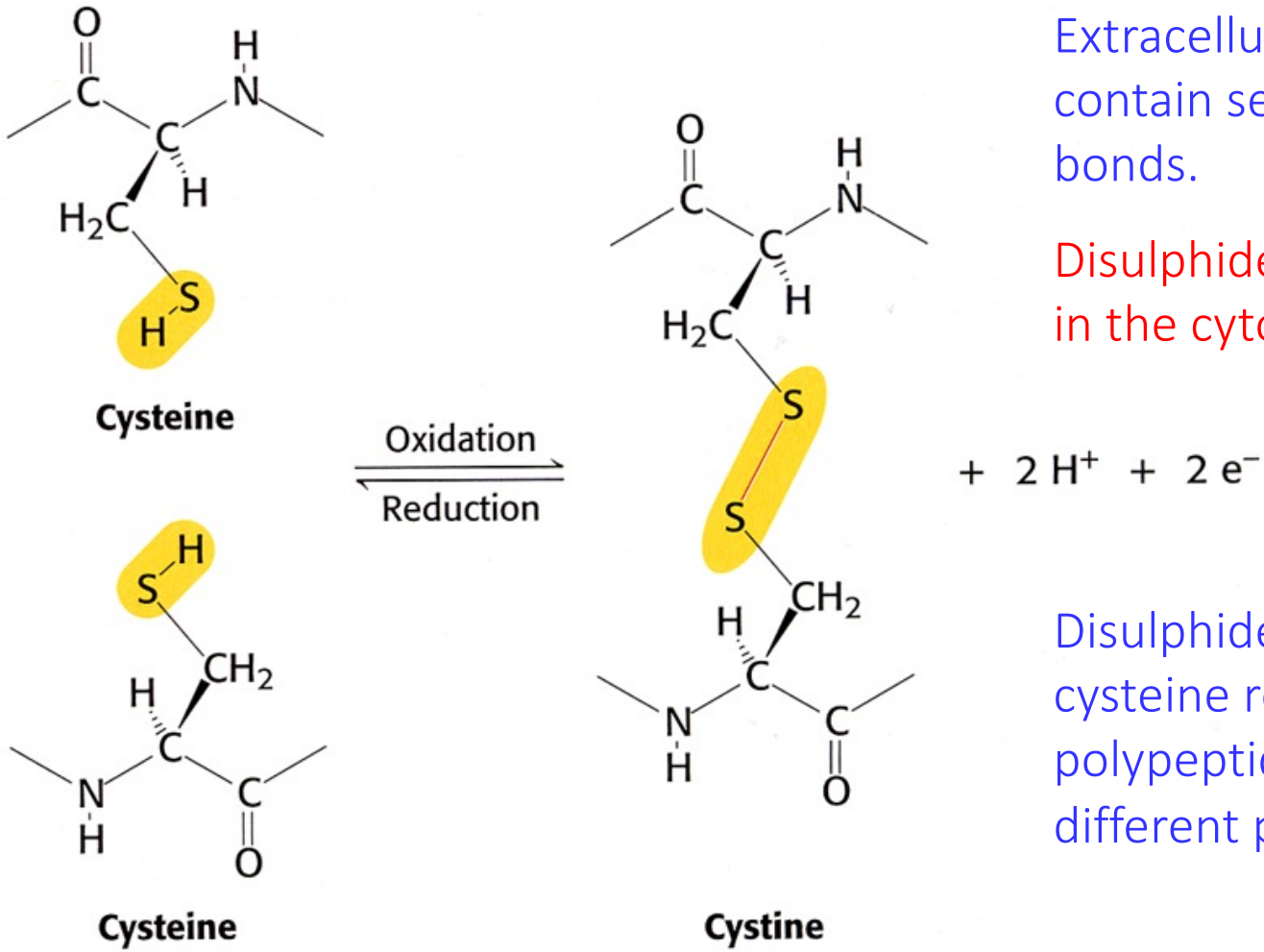
More dissociated → more equilibrium to right → larger K_a → smaller pK_a

Smaller pK_a → strong acid / weak conjugate base

Larger pK_a → weak acid / strong conjugate base

Disulphide bonds

A disulphide bond can form between two cysteine residues in proteins.



60 kcal/mol (251 kJ mol^{-1})

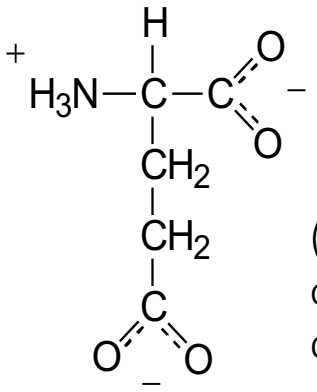
Extracellular proteins often contain several disulphide bonds.

Disulphide bonds do not form in the cytosol.

Disulphide bonds can link cysteine residues within a single polypeptide chain or on different polypeptide chains.

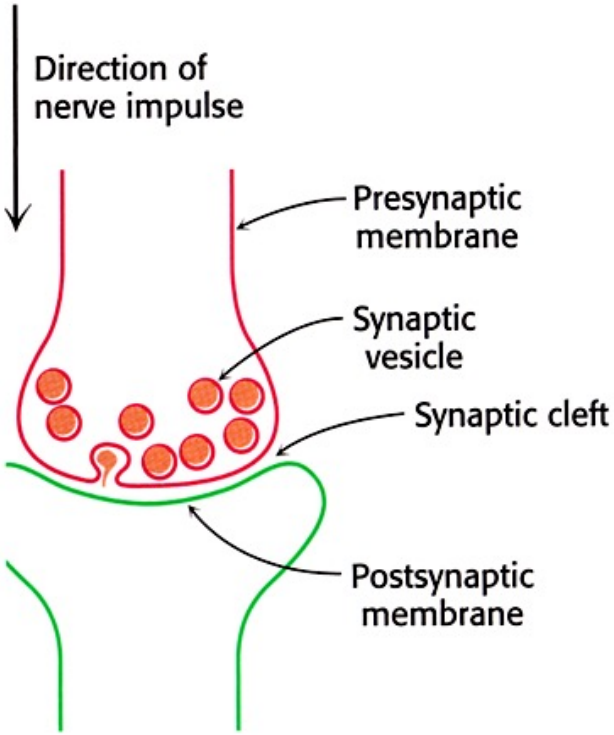
Amino acids as neurotransmitters

Glutamate is an excitatory neurotransmitters

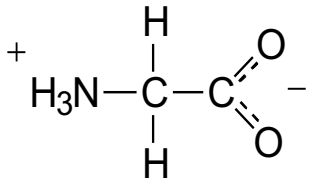


Glutamate

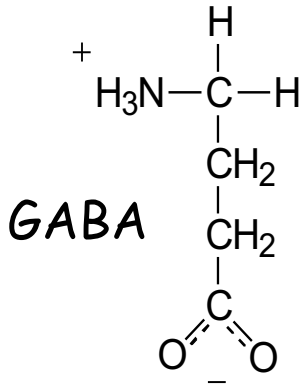
(release of glutamate leads to the opening of cation channels and the depolarisation of the membrane)



Glycine and γ-aminobutyric acid (GABA) are inhibitory neurotransmitters



Glycine



GABA

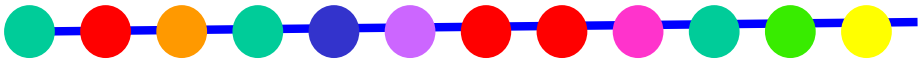
(release of glycine or GABA leads to the opening of Cl^- channels)

formed from glutamate by loss of carboxyl group

Overview of protein architecture

1) structure and chemistry of amino acids

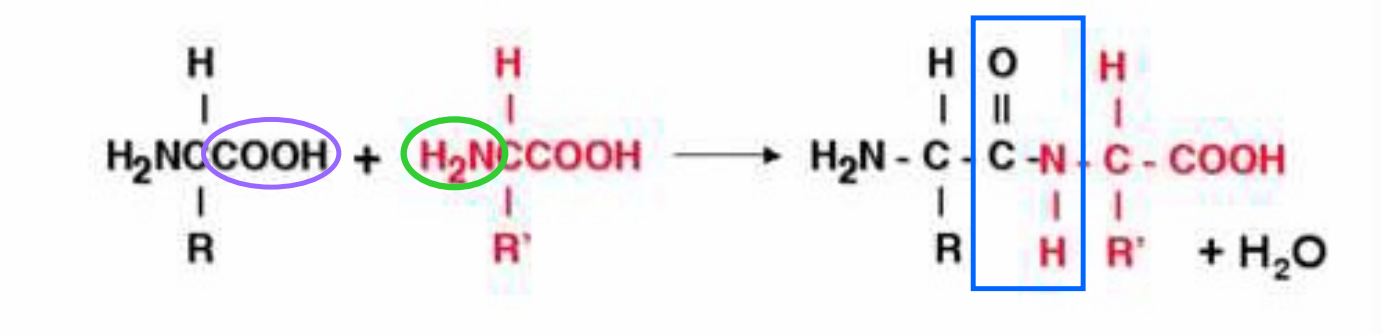
2) how amino acids are linked together through peptide bonds to form a polypeptide chain



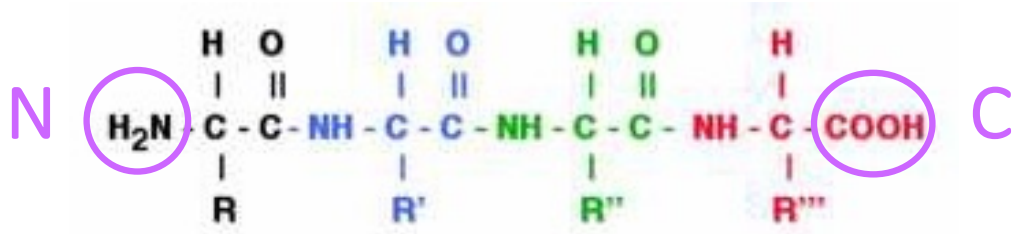
3) how the polypeptide chain folds in 3D
   - secondary structure elements (α-helix and β-sheet)
   - how secondary structure elements pack together

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

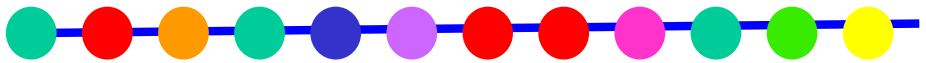
R A F G

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
- Average mass of 1 aminoacid: 100 Da
- 1 Dalton = 1 g/mol ≈ 1.7x10-24 g



- 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

- 1 Dalton = 1 g/mol ≈ 1.7×10^{-24} g (as derived in Figure 1)
- 1 nM is about 1 molecule per bacterial volume as derived in Figure 2, 10^{1}-10^{2} per yeast cell and 10^{3}-10^{4} molecules per characteristic mammalian (HeLa) cell volume. For 1 μM multiply by a thousand, for 1 mM multiply by a million
- 1 M is about one per 1 nm^3
- There are 2-4 million proteins per 1 μm^3 of cell volume
- Concentration of 1 ppm (part per million) of the cell proteome is ≈ 5 nM.
- 1 μg of DNA fragments 1 kb long is ≈1pmol or ≈10^{12} molecules
- Under standard conditions, particles at a concentration of 1M are ≈1 nm apart
- Mass of typical amino acid ≈100 Da
- Protein mass [Da] ≈100 x Number of amino acids
- Density of air ≈1 kg/m^3
- Water density ≈55 M ≈ x 1000 that of air ≈1000 kg/m^3
- 50 mM osmolites ≈1 Atm osmotic pressure (as shown in Figure 3)

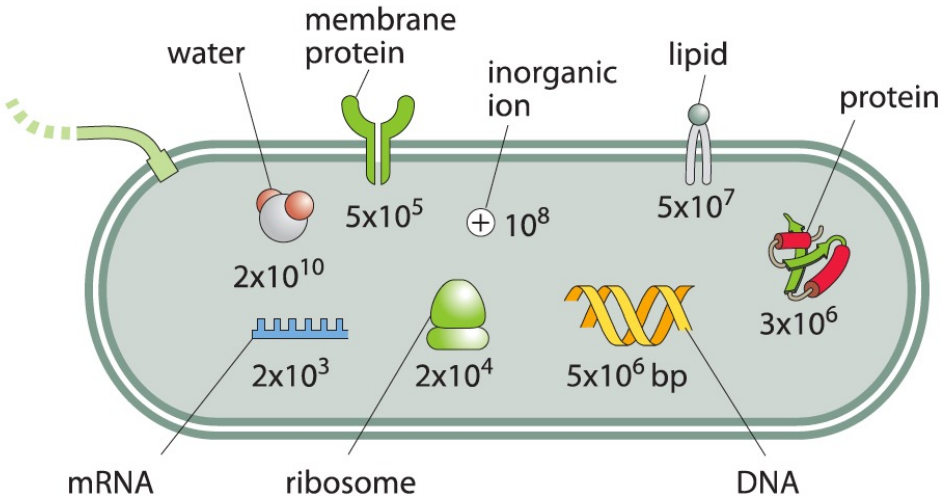
1 Dalton = 1 g/mol ≈ 1.7x10-24 g

- Water molecule volume ≈0.03 nm^3, $(\approx 0.3\ nm)^3$
- A base pair has a volume of ≈1 nm^3
- A base pair has a mass of ≈600 Da
- Lipid molecules have a mass of ≈500 -1000 Da
- 1 k_BT ≈ 2.5 kJ/mol ≈ 0.6 kcal/mol ≈ 25 meV ≈ 4 pN nm ≈ 4×10^{-21} J
- ≈6 kJ/mol sustains one order of magnitude concentration difference (=RT ln(10) ≈ 1.4 kcal/mol)
- Movement across the membrane is associated with 10-20 kJ/mol per one net charge due to membrane potential
- ATP hydrolysis under physiological conditions releases 20 k_BT ≈ 50 kJ/mol ≈ 12 kcal/mol ≈ 10^{-19} J
- One liter of oxygen releases ≈20 kJ during respiration
- A small metabolite diffuses 1 nm in ~1 ns
- 1 OD_{600} ≈0.5 g cell dry weight per liter
- ≈10^{10} carbon atoms in a 1 μm^3 cell volume

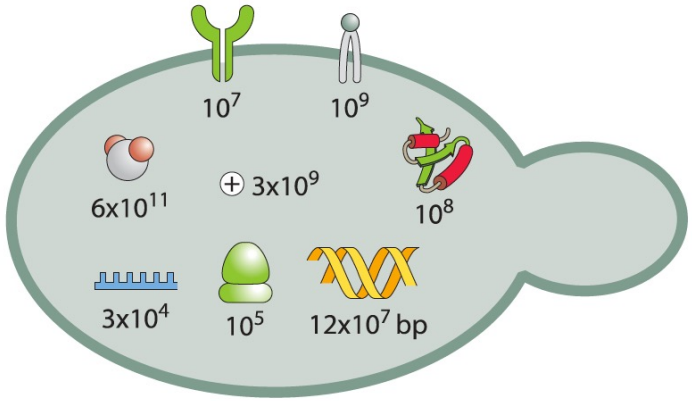
1 Dalton = 1 g/mol ≈ 1.7x10-24 g

The nucleus houses the billions of base pairs of the genome and is the site of the critical transcription processes taking place as genes are turned on and off in response to environmental stimuli and over the course of both the cell cycle and development.

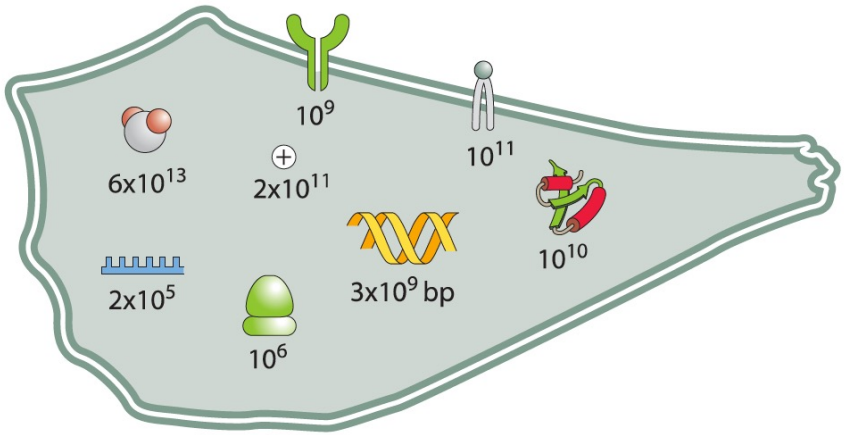
(A) bacterial cell (specifically, *E. coli:* V ≈ 1 μm^3; L ≈ 1 μm; τ ≈ 1 hour)



(B) yeast cell (specifically, *S. cerevisiae:* V ≈ 30 μm^3; L ≈ 5 μm; τ ≈ 3 hours)



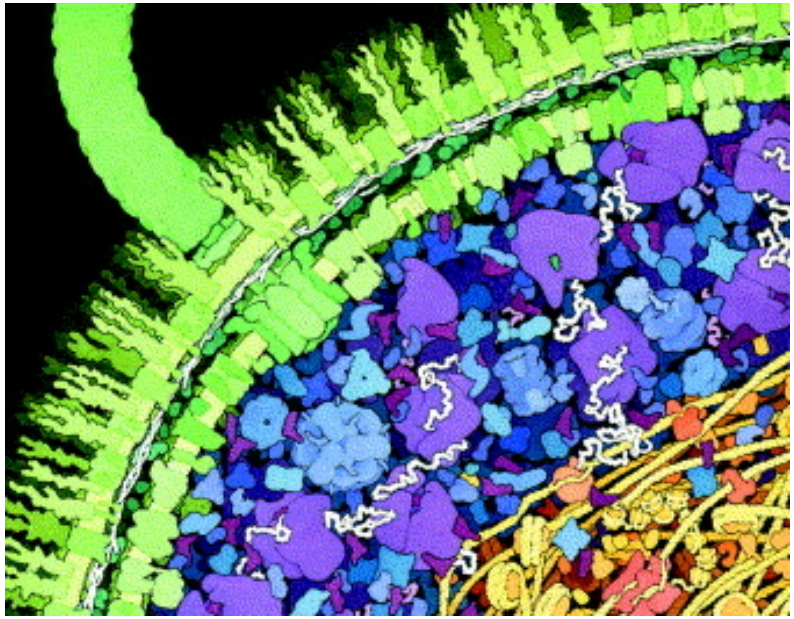
(C) mammalian cell (specifically, HeLa: V ≈ 3000 μm^3; L ≈ 20 μm; τ ≈ 1 day)



Given that there are several million proteins in a typical bacterium and these are the product of several thousand genes, **we can expect the "average" protein to have about 10^3 copies.**

property	*E. coli*	budding yeast	mammalian (HeLa line)
cell volume	0.3–3 μm^3	30–100 μm^3	1,000–10,000 μm^3
proteins per μm^3 cell volume		$2–4\times10^6$	
mRNA per cell	10^3-10^4	10^4–10^5	10^5–10^6
proteins per cell	~10^6	~10^8	~10^{10}
mean diameter of protein		4–5 nm	
genome size	4.6 Mbp	12 Mbp	3.2 Gbp
number protein coding genes	4300	6600	21,000
regulator binding site length		10–20 bp	
promoter length	~100 bp	~1000 bp	~10^4–10^5 bp
gene length	~1000 bp	~1000 bp	~10^4–10^6 bp (with introns)
concentration of one protein per cell	~1 nM	~10 pM	~0.1–1 pM
diffusion time of protein across cell ($D \approx 10\ \mu m^2/s$)	~0.01 s	~0.2 s	~1–10 s
diffusion time of small molecule across cell ($D \approx 100\ \mu m^2/s$)	~0.001 s	~0.03 s	~0.1–1 s
time to transcribe a gene	<1 min (80 nts/s)	~1 min	~30 min (incl. mRNA processing)
time to translate a protein	<1 min (20 aa/s)	~1 min	~30 min (incl. mRNA export)
typical mRNA lifetime	2–5 min	~10 min to over 1 h	5-100 min to over 10 h
typical protein lifetime	1 h	0.3–3 h	10–100 h
minimal doubling time	20 min	1 h	20 h
ribosomes/cell	~10^4	~10^5	~10^6
transitions between protein states (active/inactive)		1–100 μs	
timescale for equilibrium binding of small molecule to protein (diffusion limited)		1–1000 ms (1 μM–1 nM affinity)	
timescale of transcription factor binding to DNA site		~1 s	
mutation rate		10^{-8}–10^{-10}/bp/replication	

Cell crowding and diffusion constant



Brownian motion

Albert Einstein provided a theoretical explanation of Brownian motion in 1905, which helped confirm the atomic theory of matter. His work led to the development of the diffusion equation, linking the motion to the diffusion coefficient D, which measures how fast particles spread out over time:

$$\langle x^2 \rangle = 2Dt$$

where:

- $\langle x^2 \rangle$ is the mean squared displacement of the particle,
- D is the diffusion coefficient,
- t is time.

Einstein's work was later expanded upon by Jean Perrin, who experimentally verified the atomic nature of matter through his observations of Brownian motion.

D measures the rate of diffusion
is expressed as unit of area per unit of time!

Diffusion constant

The diffusion constant D is defined by **Fick's Law of Diffusion**, which describes the movement of particles from regions of higher concentration to regions of lower concentration. For **one-dimensional diffusion**, Fick's first law can be written as:

$$J = -D\frac{dC}{dx}$$

where:

- J is the diffusion flux (the amount of substance moving through a unit area per unit time),
- D is the diffusion coefficient (or constant),
- $\frac{dC}{dx}$ is the concentration gradient (change in concentration C over distance x).

This means the diffusion flux is proportional to the concentration gradient, and the proportionality constant is the diffusion constant D.

Diffusion constant

The value of the diffusion constant D depends on several factors:

1. **Temperature (T)**: As temperature increases, the kinetic energy of particles increases, leading to faster diffusion. Hence, D increases with temperature.
2. **Medium Viscosity (η)**: In more viscous mediums (like syrup vs. water), particles face more resistance when moving, leading to lower diffusion rates. Therefore, D decreases with increasing viscosity.
3. **Particle Size**: Larger particles have more inertia and move slower, resulting in a lower diffusion constant compared to smaller particles.
4. **Nature of the Medium**: The type of medium (gas, liquid, or solid) and the interaction between the particles and the medium also affect the diffusion constant.

Diffusion constant

Einstein Relation for Diffusion:

For a small particle undergoing Brownian motion in a fluid, the diffusion constant is related to the temperature, viscosity of the fluid, and the size of the particle. This relationship is given by the **Stokes-Einstein equation**:

$$D = \frac{k_B T}{6\pi\eta r}$$

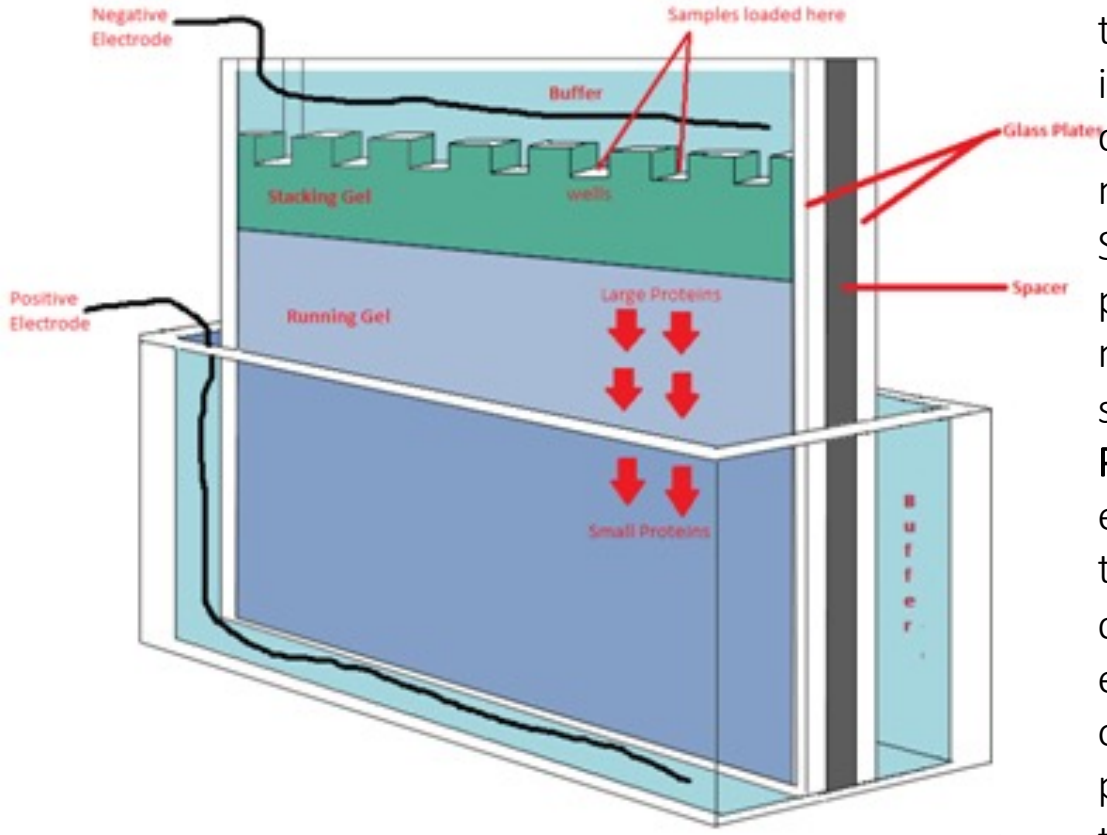
where:

- D is the diffusion constant,
- k_B is the Boltzmann constant ($1.38 \times 10^{-23}\ \mathrm{J/K}$),
- T is the absolute temperature,
- η is the dynamic viscosity of the fluid,
- r is the radius of the particle.

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.

SDS confers uniform negative charge to proteins, masking their native charges and making the separation dependent on molecular size, rather than charge or shape.

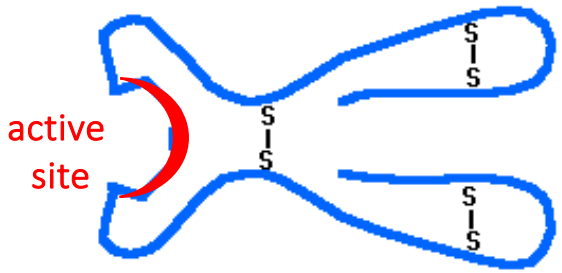
Proteins are completely denatured. When an electric field is applied, the negative charge of the SDS causes the proteins to move through a clear acrylamide matrix toward the positive electrode. This matrix has holes in it that sieve out the proteins by molecular weight. Large proteins move more slowly 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

Denaturation

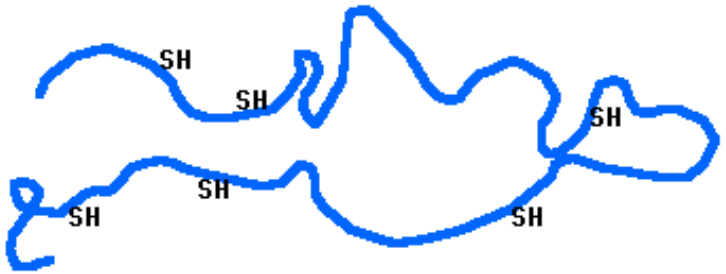
Many proteins can be unfolded and refolded:

FOLDED, active protein



urea, pH, high T

UNFOLDED, denatured protein



refold = regain biological activity

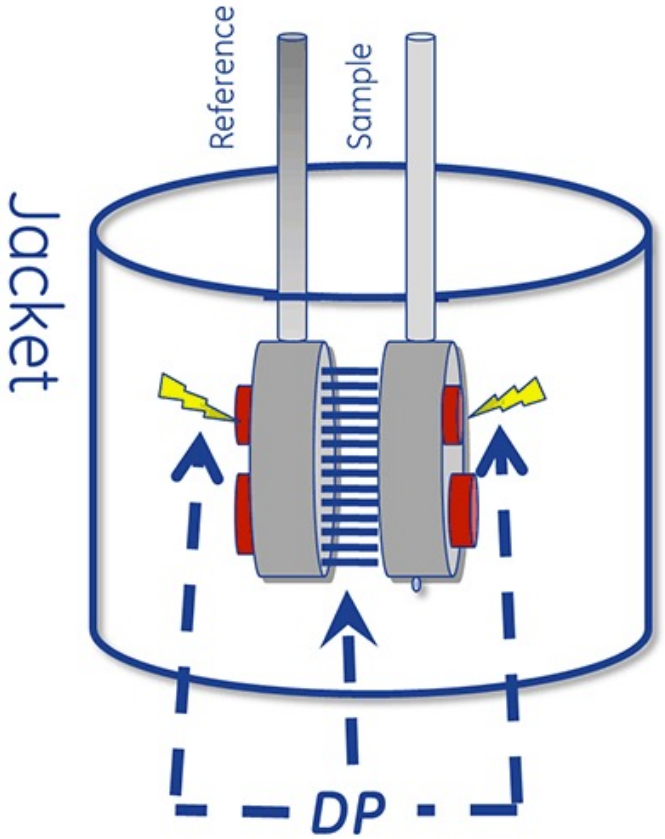
unfold = loose biological activity

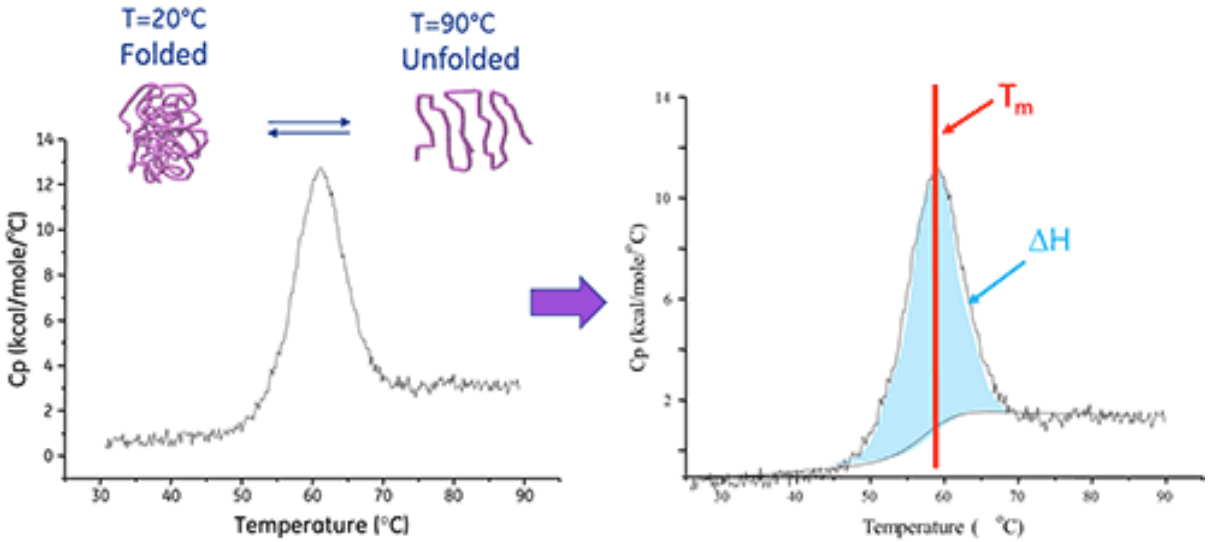
When a protein contains disulphide bridges (S-S)
we also need to reduce them to SH to unfold.

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

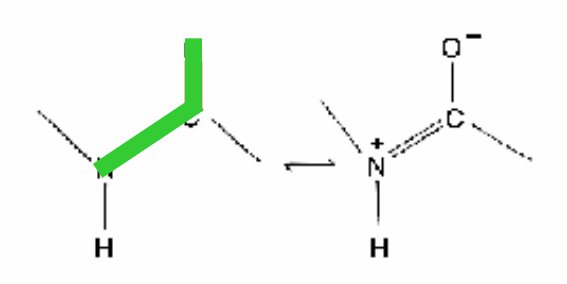
Denaturation

Many proteins can be unfolded and refolded: differential calorimetry

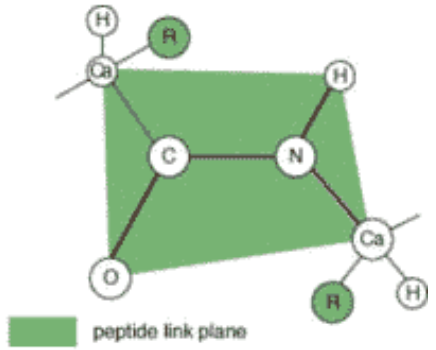


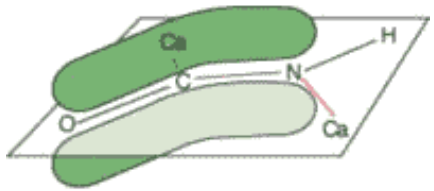


Planarity of the peptide bond



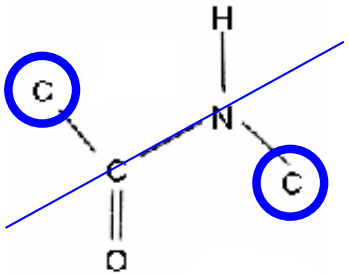
Partial double bond character of the N-C bond leads to restricted rotation
the region NH-CO is planar:



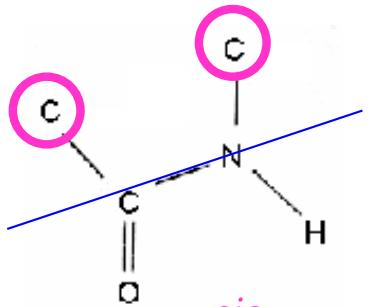


delocalisation of the π electrons over the entire peptide bond, rather than simply over the C=O bond

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

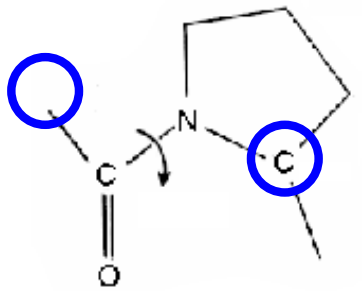


trans



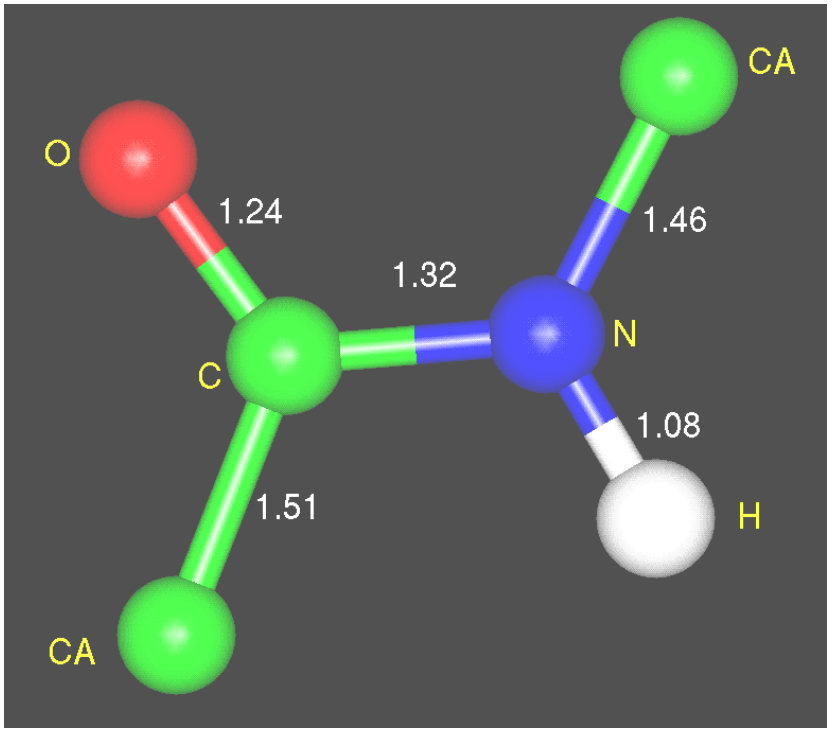
cis

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



trans
Proline

The ideal peptide



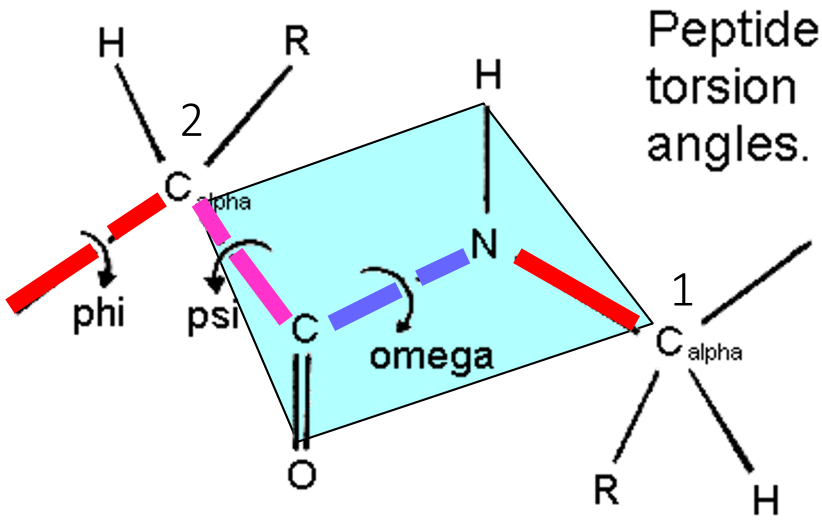
C-N single bond ~ 1.48 Å

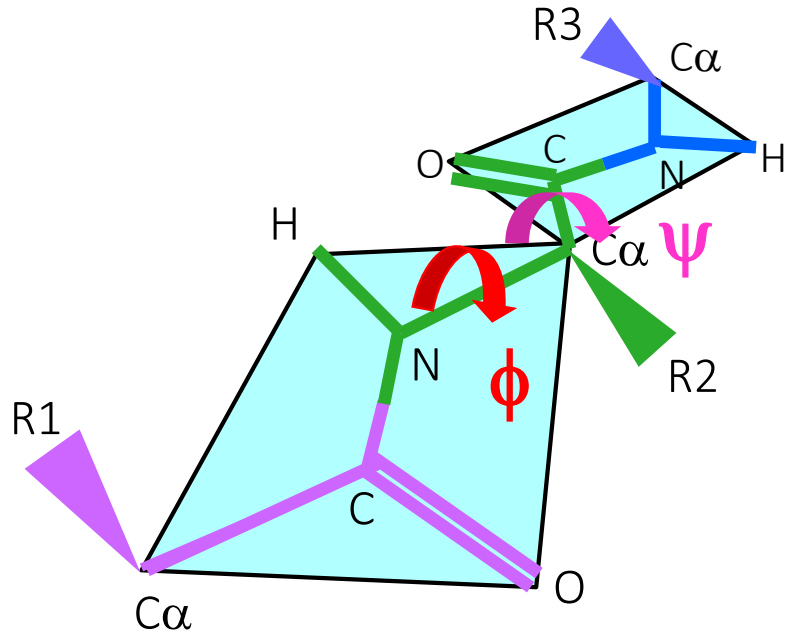
C=O double bond ~ 1.20 Å

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

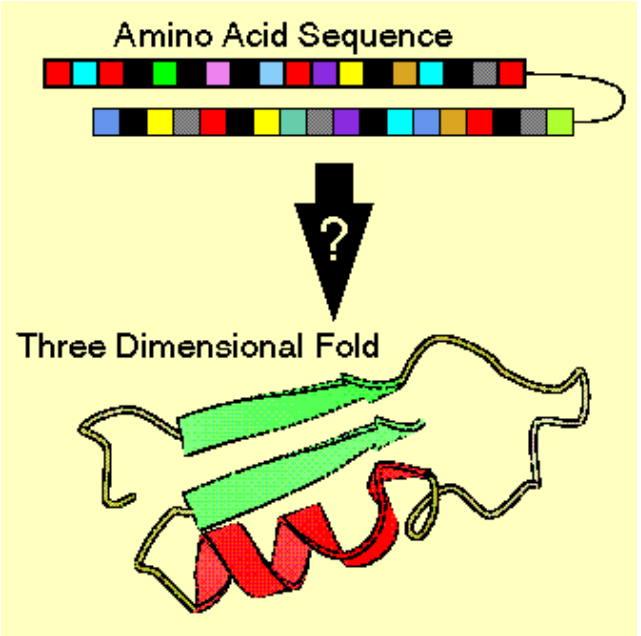
The main chain conformation is defined by the sequence of the (ψ,ϕ) angles: the list of the (ψ,ϕ) for each amino acid dictate the fold of the polypeptide chain, i.e. the 3D structure of the protein

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’)

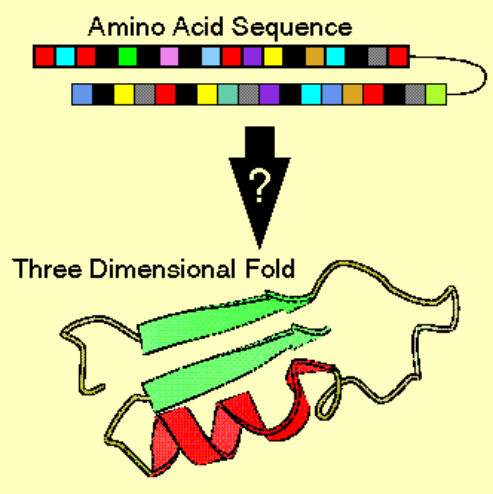
Studying experimentally how folding occur in the cell: some proteins fold by themselves, others require the help of other proteins called chaperones.

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

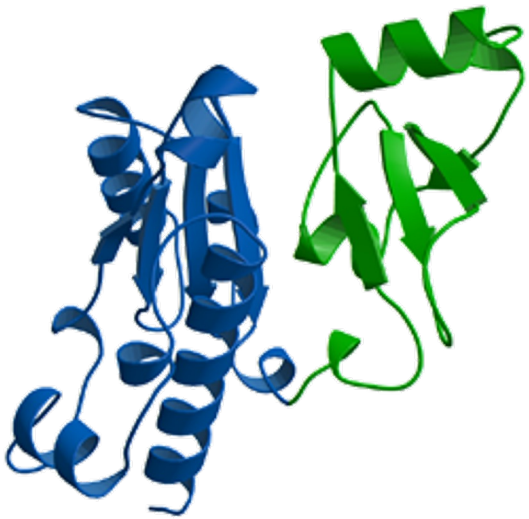
Overview of protein architecture

1) structure and chemistry of amino acids

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

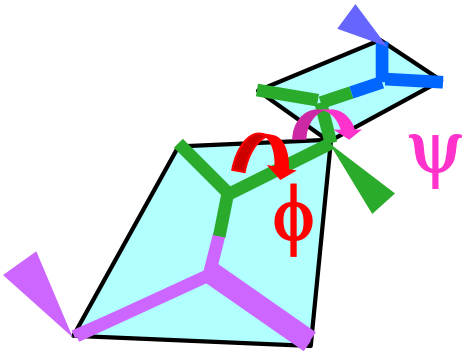
3) how the polypeptide chain folds in 3D:

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



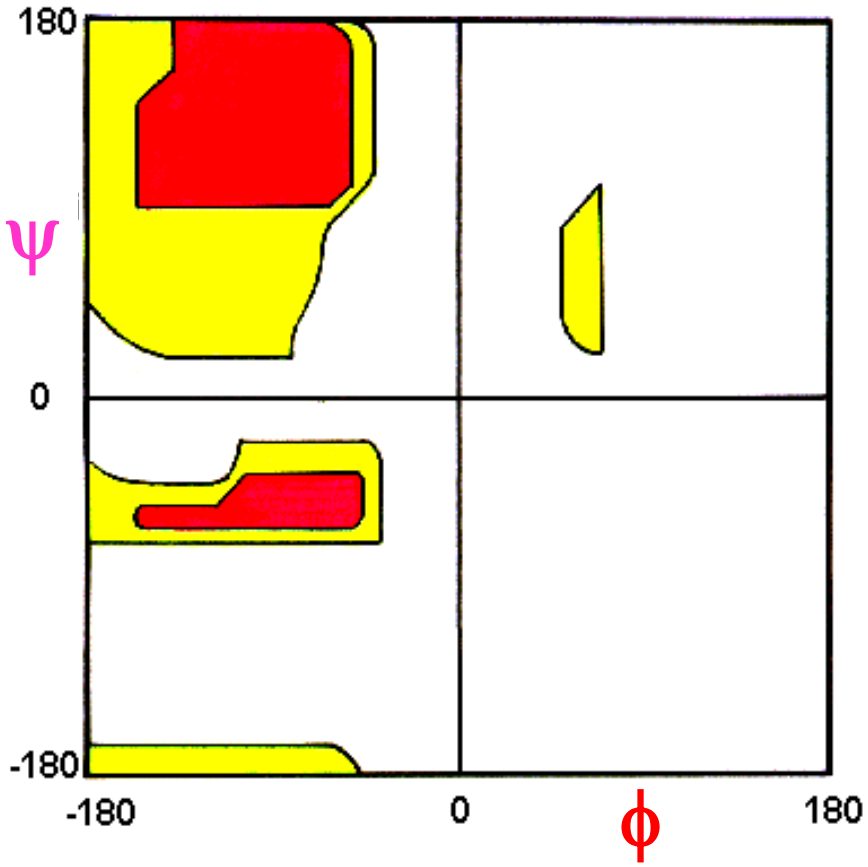
The Ramachandran plot

Because of steric clashes, only certain combinations of torsion angles are allowed: we can plot these allowed combinations in the (ψ,ϕ) plane - this is called the Ramachandran plot.

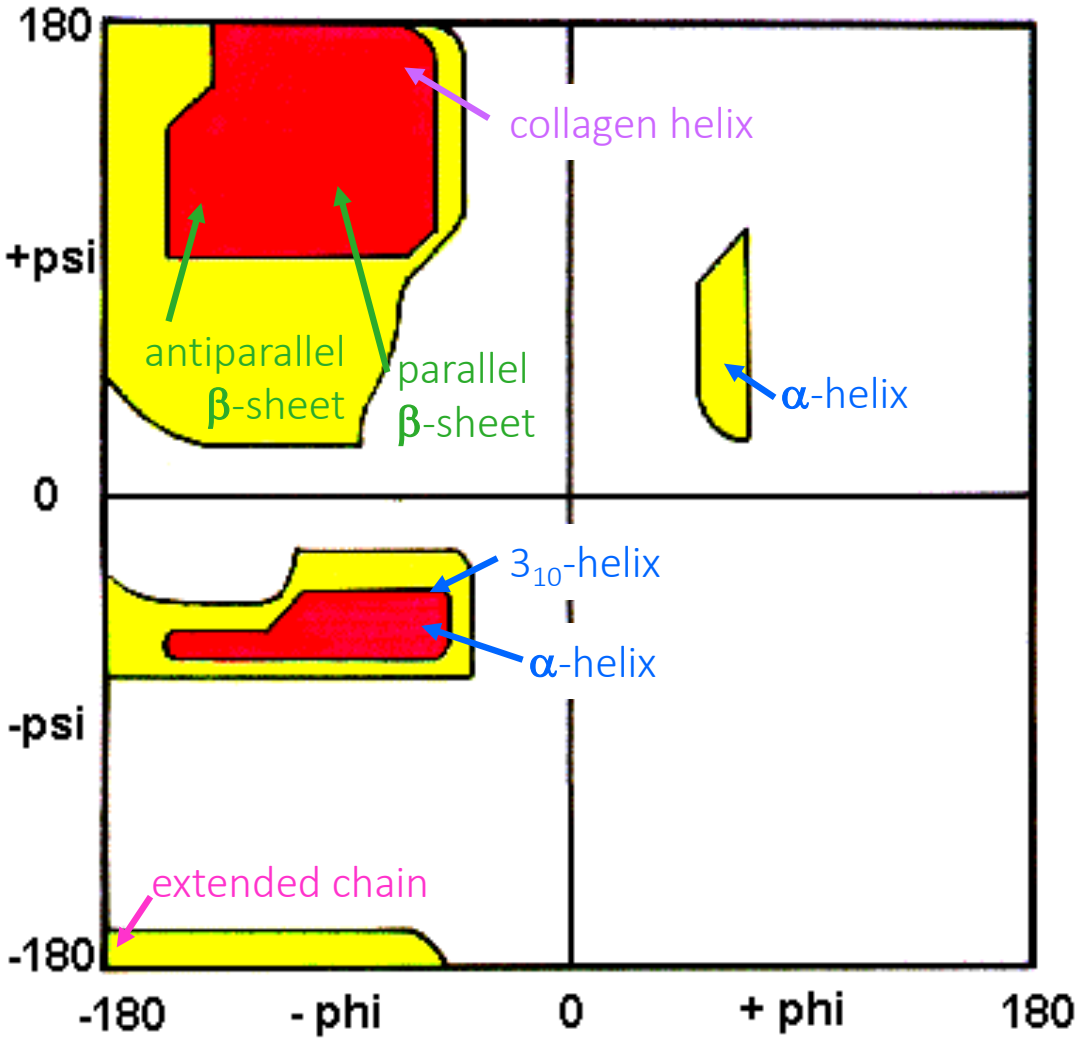


favorable regions for all aa

allowed regions for all aa



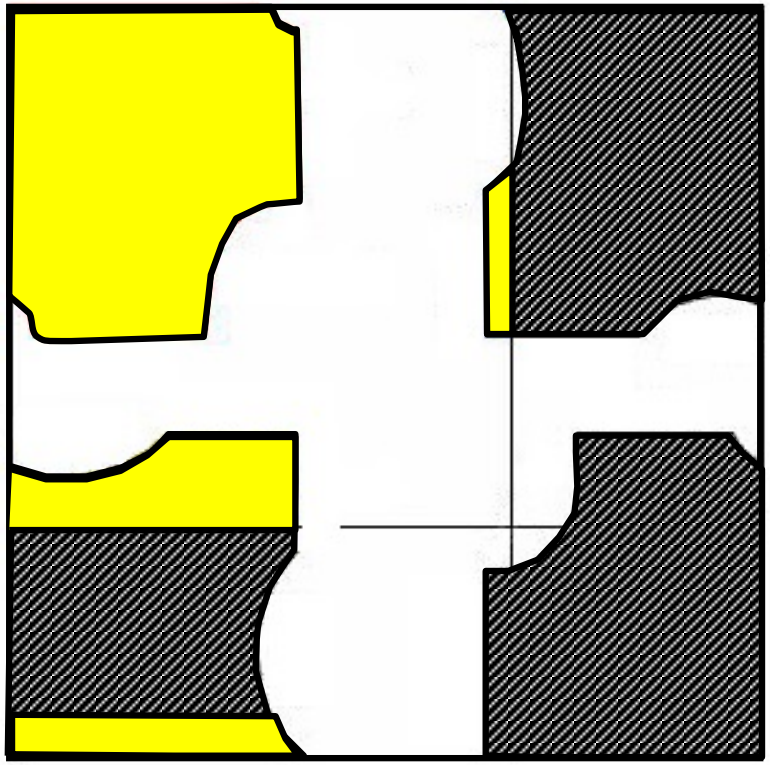
The Ramachandran plot: secondary structure elements



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 folding of the polypeptide chain, i.e. the 3D structure of the protein

Therefore secondary structure elements will be associated with specific average values of ψ, ϕ and therefore with specific regions of the Ramachandran plot.

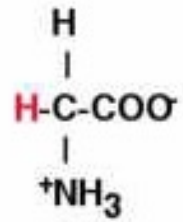
The Ramachandran plot: glycine residues



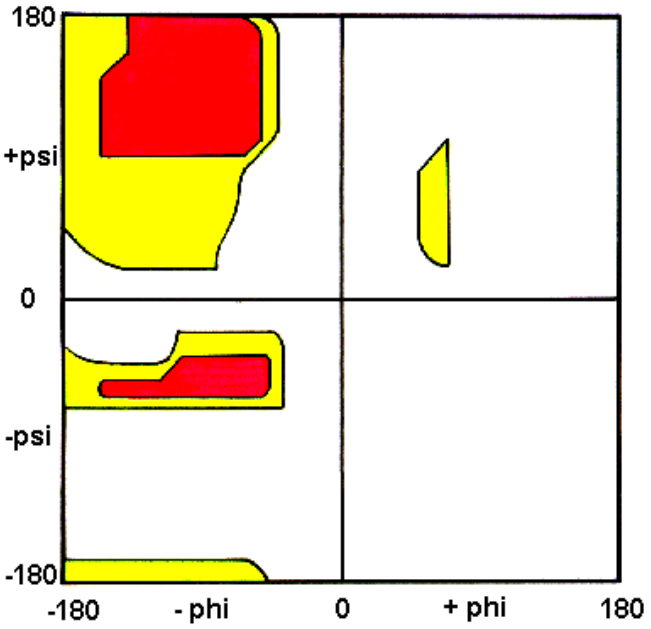
regions allowed only for glyine

regions allowed only for all aa

glycine



side chain=H (very small, hardly any steric hindrance)



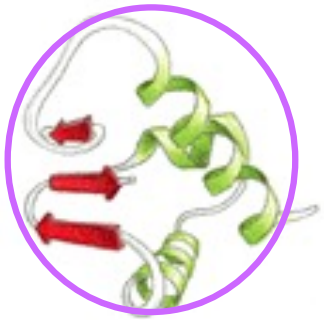
Protein architecture

Secondary structure

local organisation of the polypeptide chain

α-helix

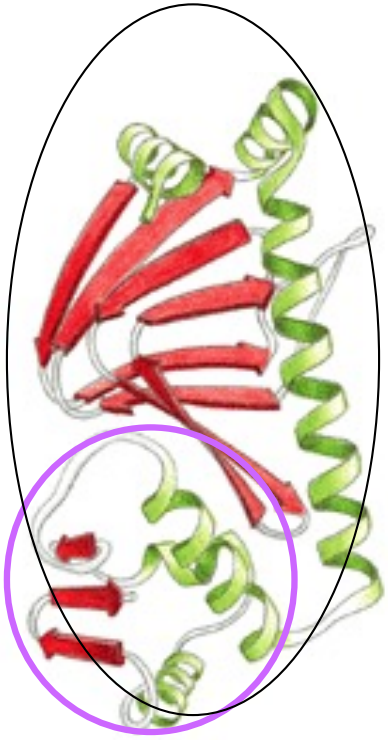
β-sheet



domain

Tertiary structure

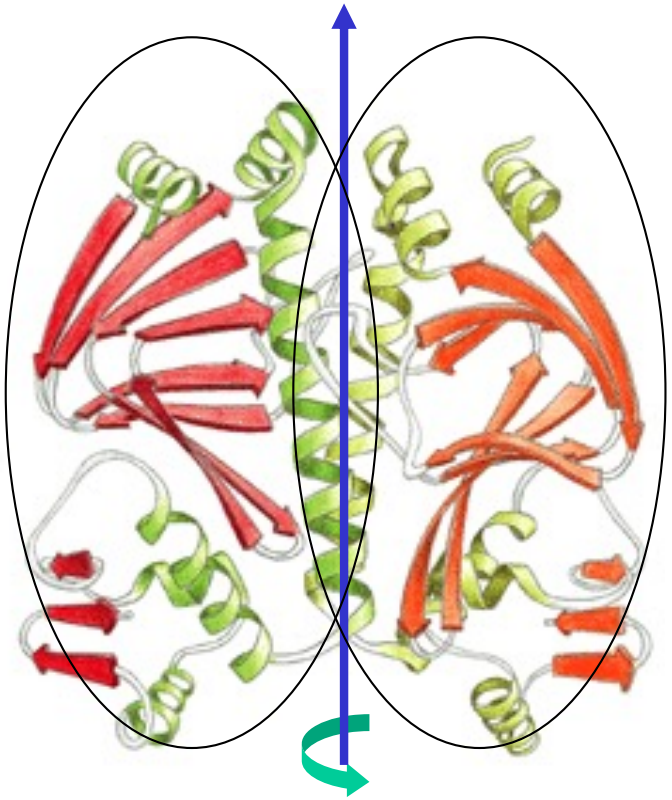
how the secondary structure elements pack together to give a 3D structure



monomer (or subunit)

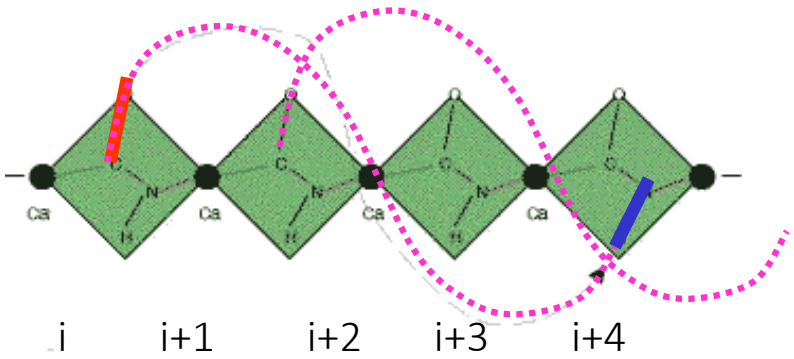
Quaternary structure

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



dimer (α2)

The α-helix

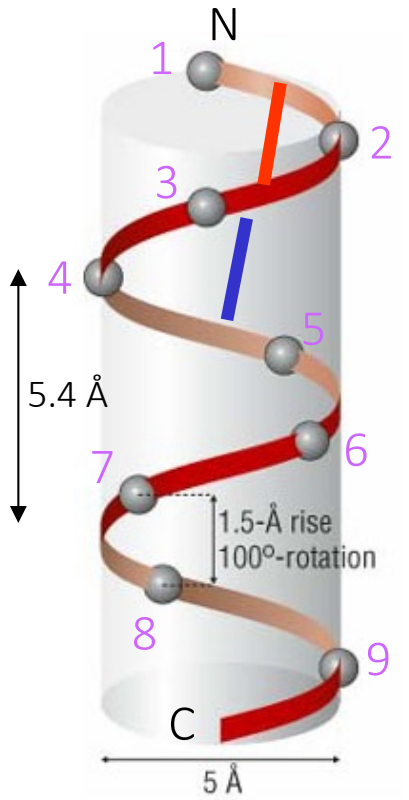


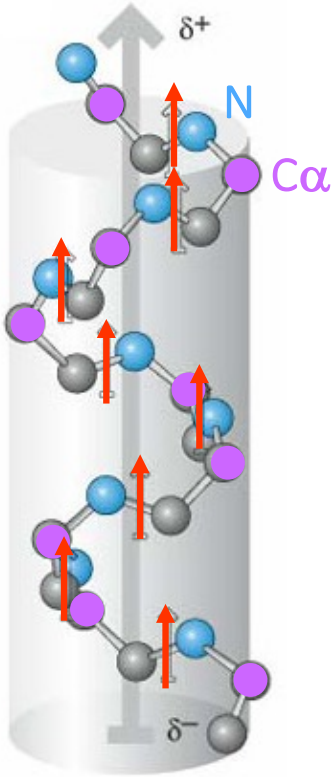
H-bonding pattern $CO_i \mapsto NH_{i+4}$
(local interactions)

all main-chain CO and NH are bonded

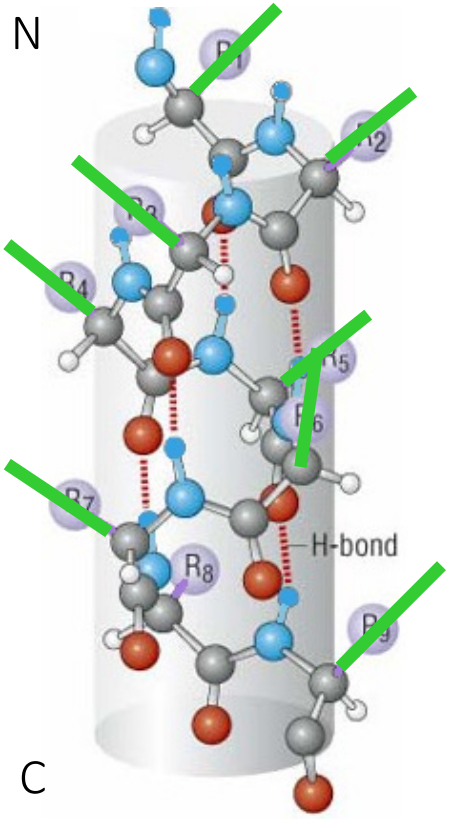
3.6 amino acids per turn;
1.5 Å rise per amino acid
$\mapsto$ 5.4 Å pitch

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





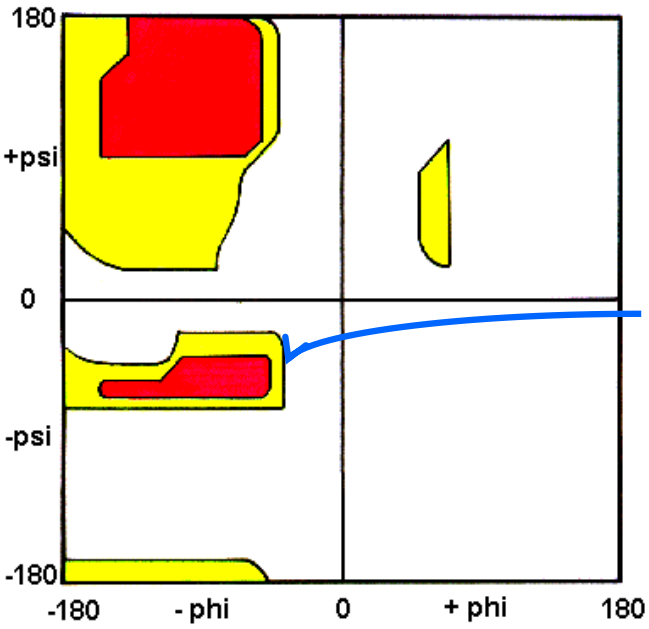
The α-helix



- nitrogen
- oxygen
- carbon
- R=side chain

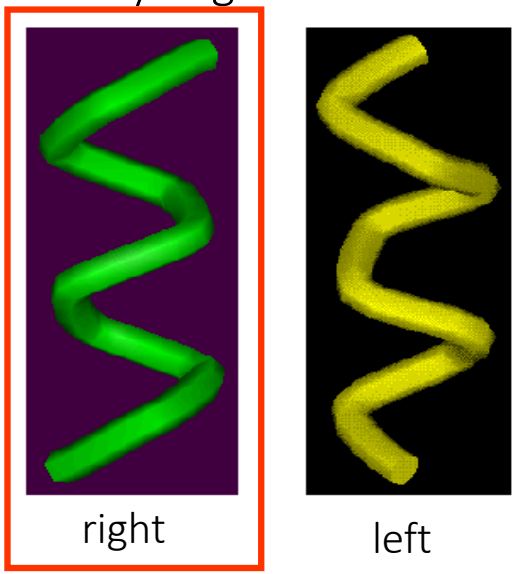
rod-like structure with side chains extending outside

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



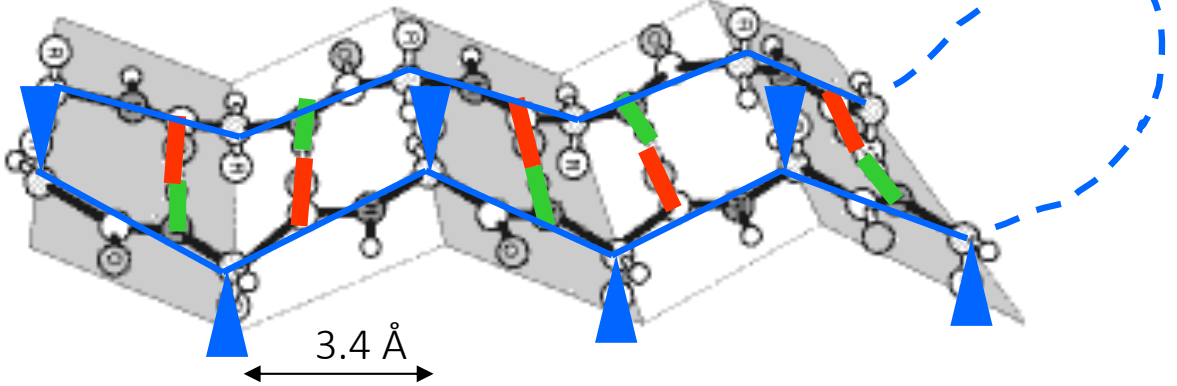
right-handed helix

always right-handed



can accommodate all residues except proline

The β-sheet



anything between 2 and few hundreds amino acids

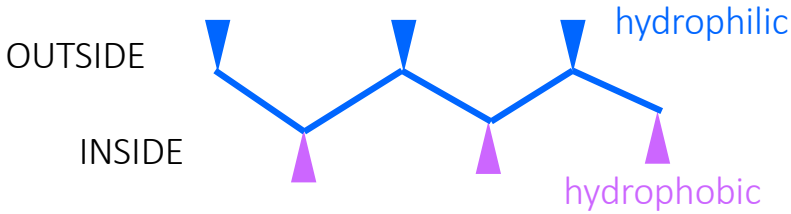
β-strand

β-strand
(zig-zag)

β-sheet

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

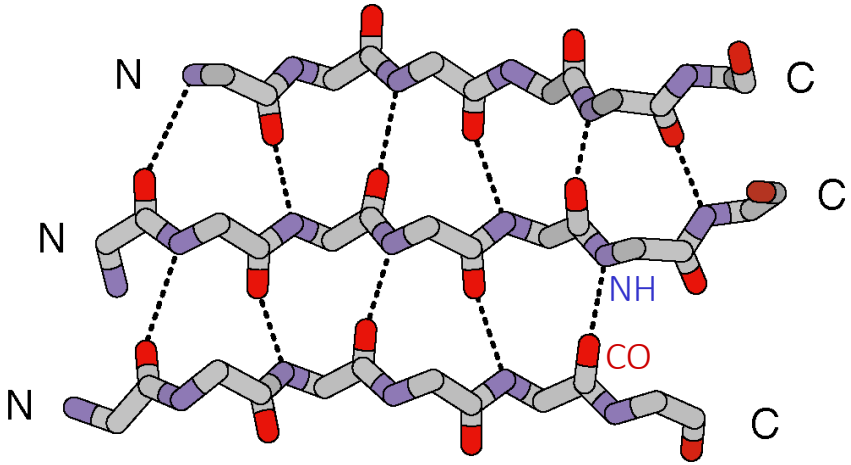
side chains points alternatively up and down

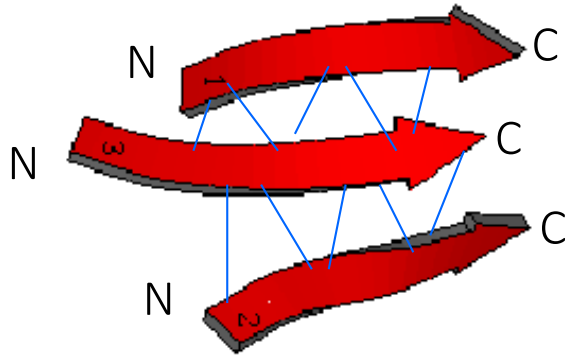


stabilised by main-chain:main-chain NH/CO hydrogen bonds between adjacent strands; contrary to the α-helix these are H bonds between NH/CO groups far apart in the amino-acid sequence

The β-sheet

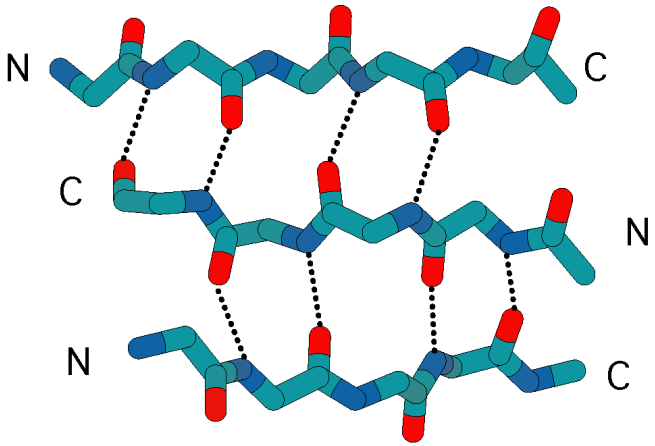
can be parallel (strands run in the same direction)

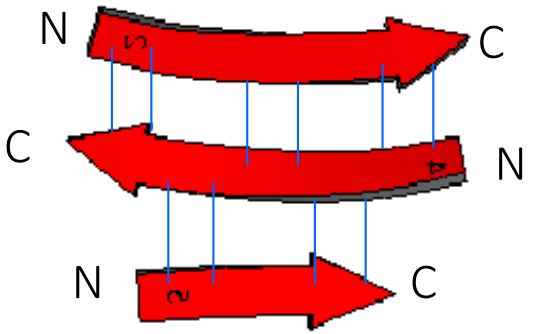




hydrogen bonding pattern

or antiparallel (strands run in opposite direction)





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

Tertiary structure:

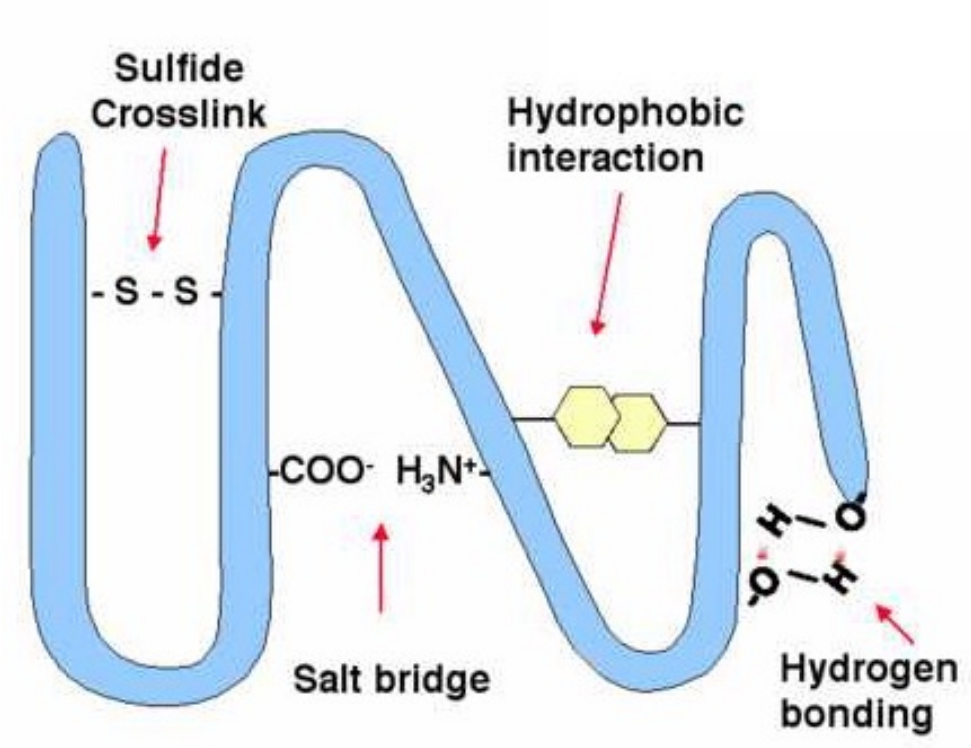
how the secondary structure elements pack together to give a 3D structure

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

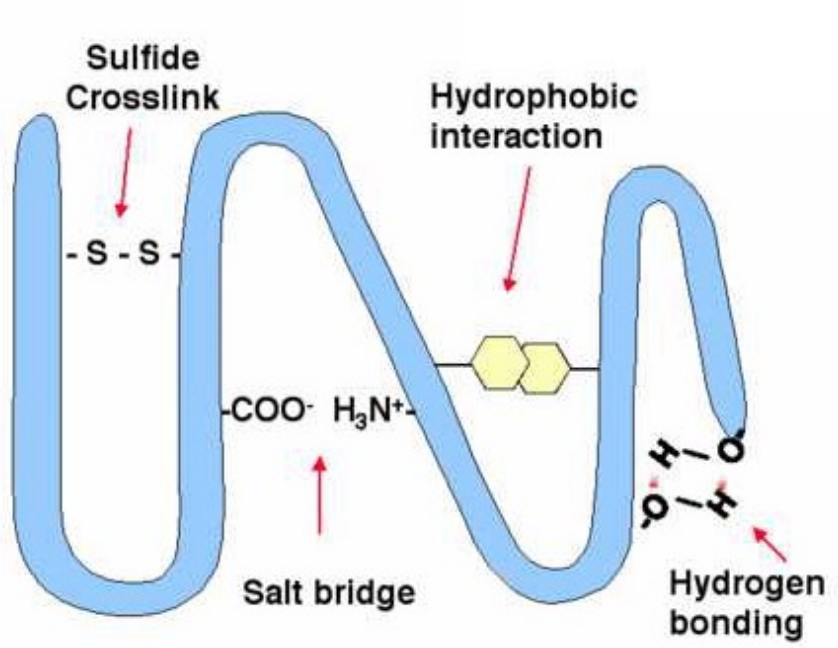


Tertiary structure:

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!

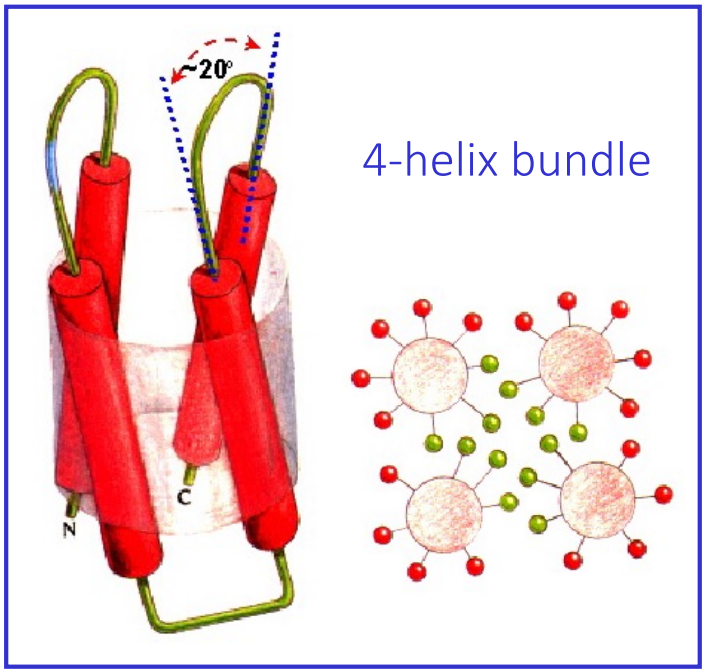


The structure is more conserved than the sequence.

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.

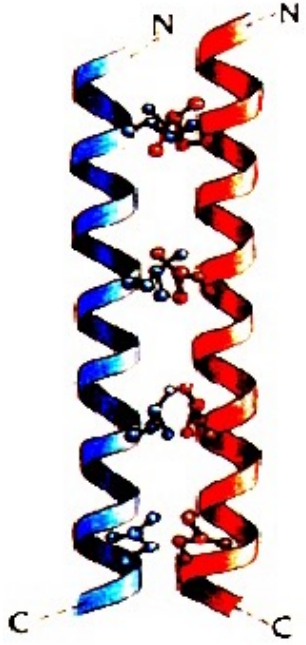
It is difficult to systematically list and classify all the motifs - here are examples of some of the common ones:



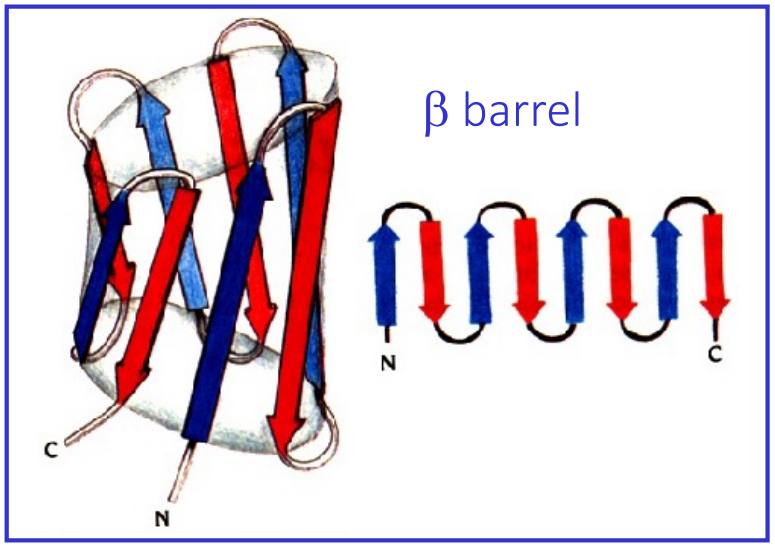
4-helix bundle

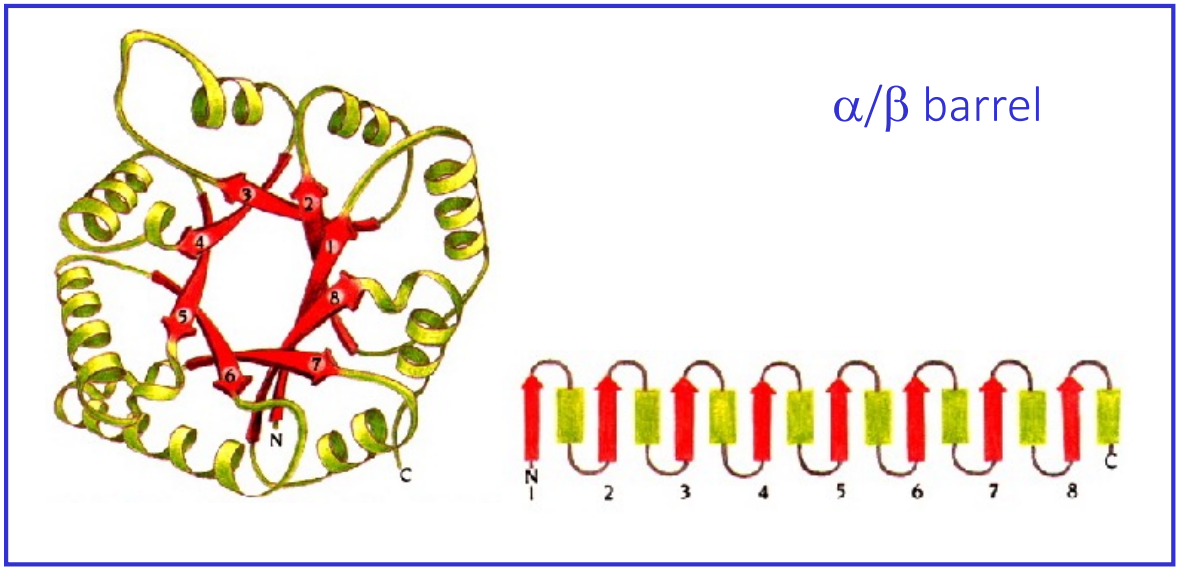
Leu zipper

a leucine residue repeated every 7 residues.



Tertiary structure:
motifs in protein structures



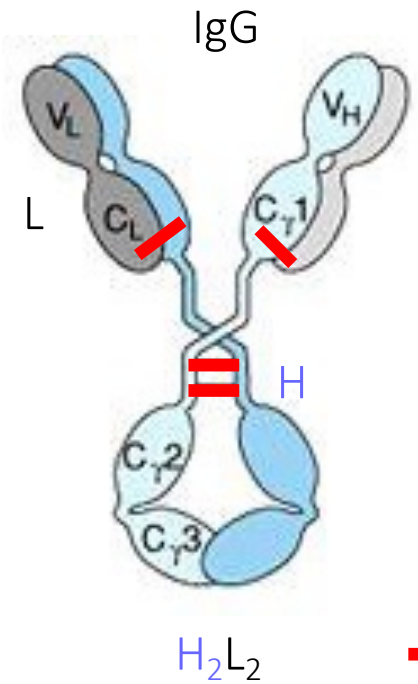


.... and many, many more!!!!

Quaternary structure:

how subunits aggregate to form multimeric proteins

Covalently-linked polypeptide chains

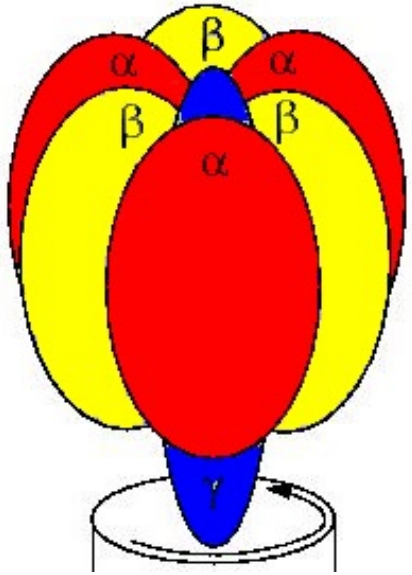


H_2L_2

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

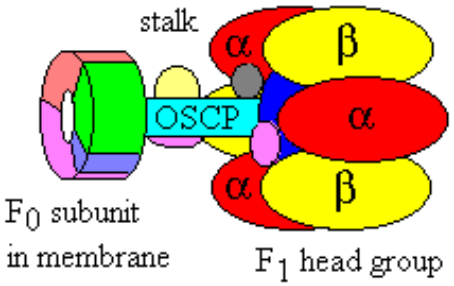
Hetero-multimers: **different** polypeptides aggregating together to form a unit.



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

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:

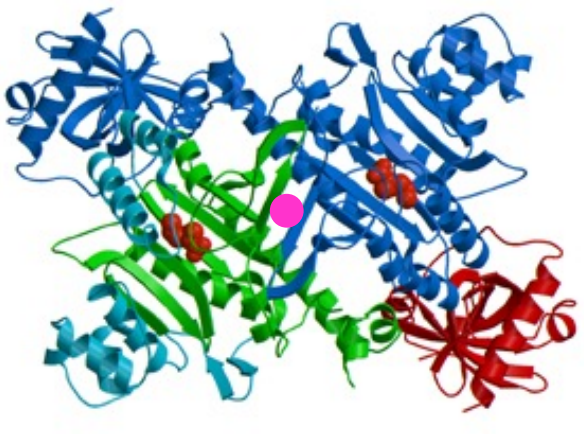


Quaternary structure:

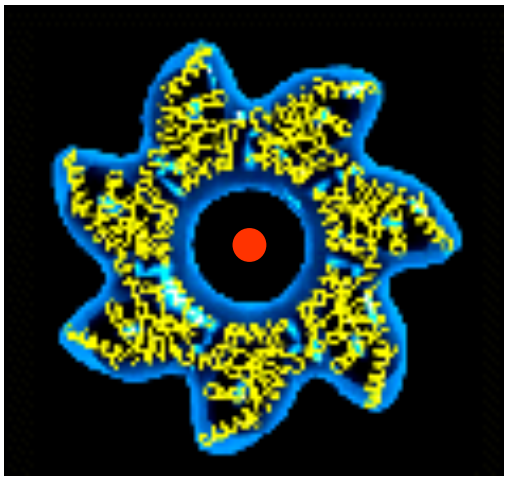
how subunits aggregate to form multimeric proteins

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:

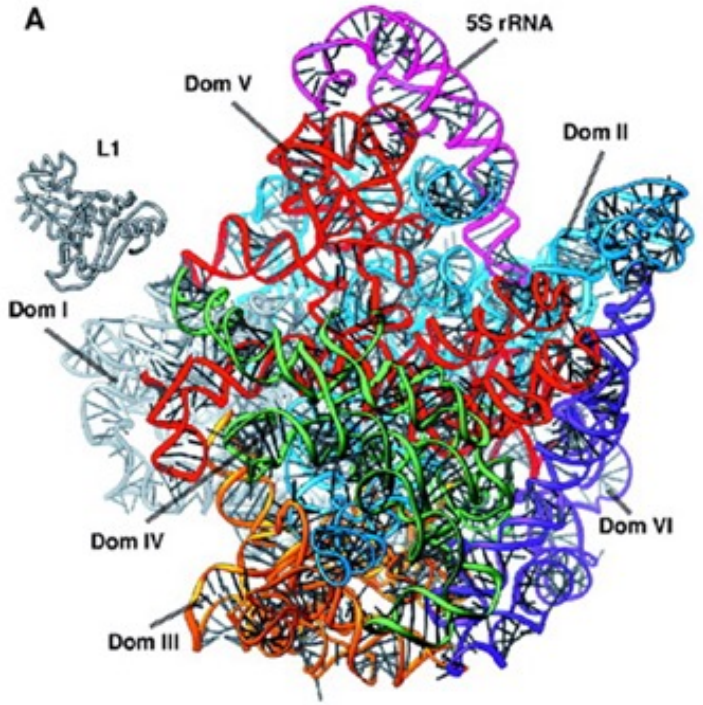
how subunits aggregate to form multimeric proteins

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

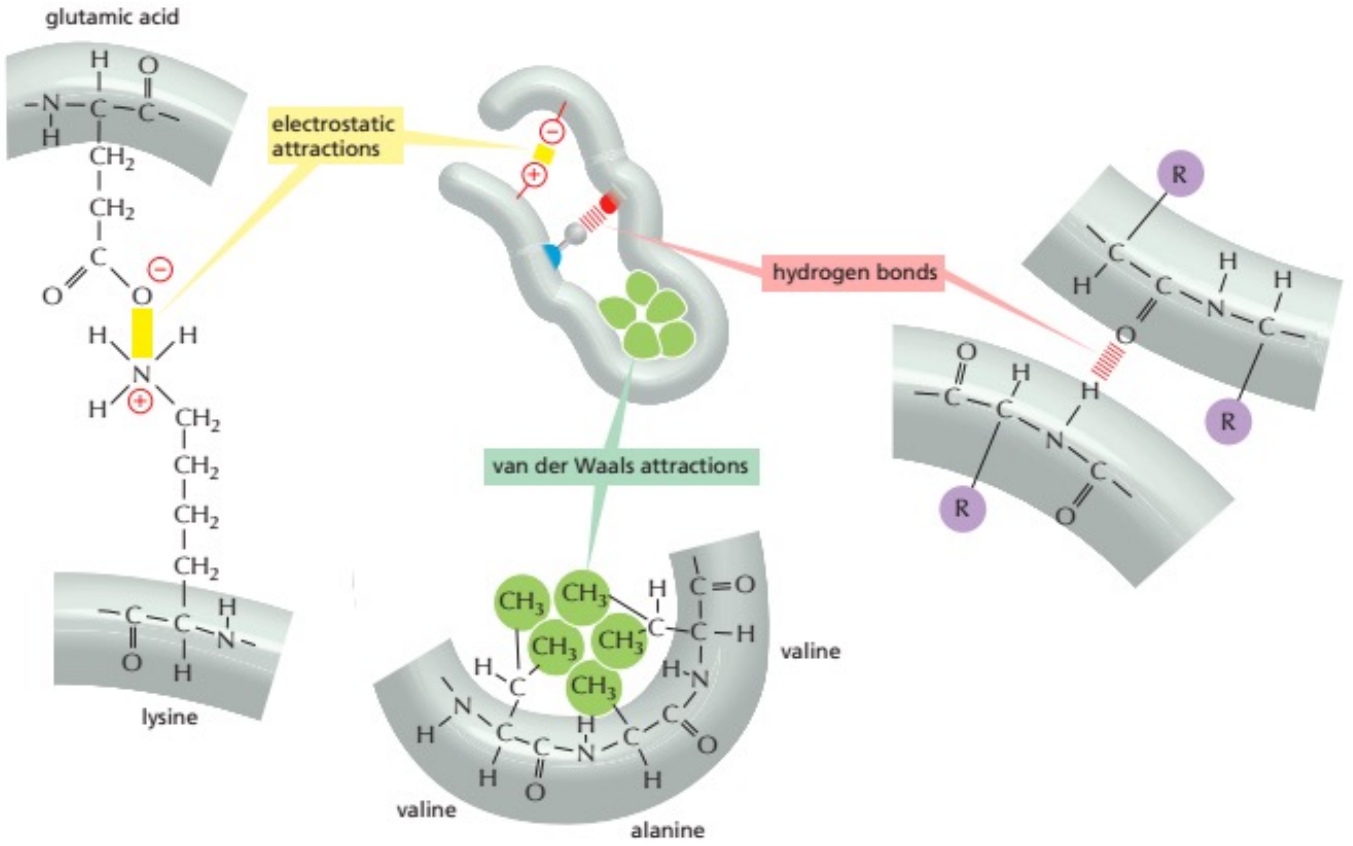


Figure 3–4 Three types of noncovalent bonds help proteins fold. Although a single one of these bonds is quite weak, many of them act together to create a strong bonding arrangement, as in the example shown. As in the previous figure, R is used as a general designation for an amino acid side chain.

Fibrous proteins

Triple helix in collagen - next

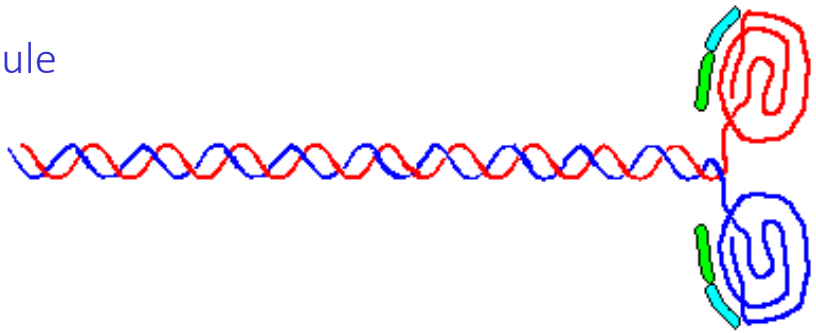
Coiled-coil α-helices present in keratin and myosin:

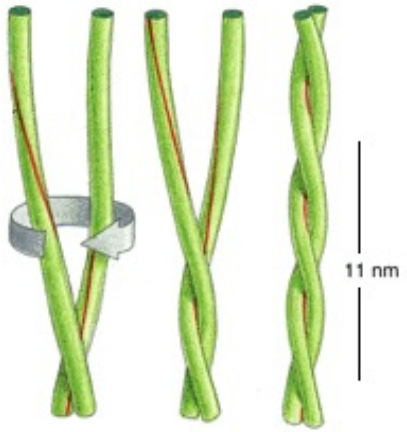
two a-helices twisted around each other to form a left handed coiled coil (7 residues repeat)

Example:a myosin molecule

heavy chains

light chains





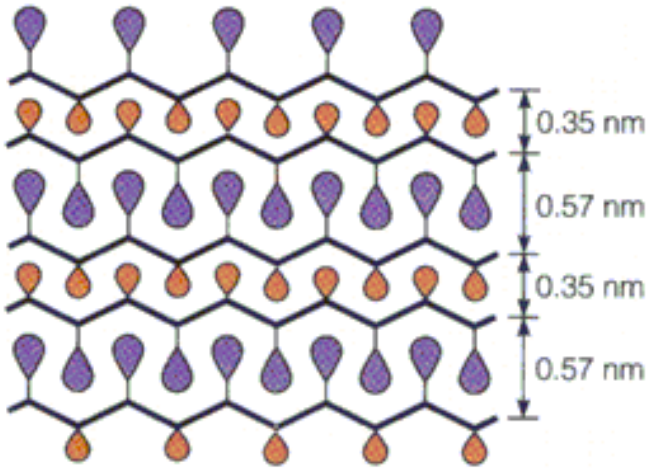
coiled-coil helices

β-sheets in amyloid fibres, spider webs and silk

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

Ala or Ser

Gly

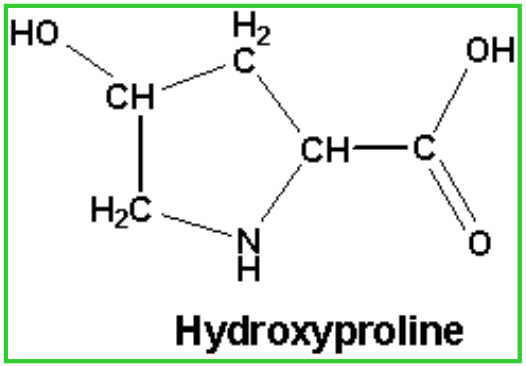


Fibrous proteins: the collagen helix

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.

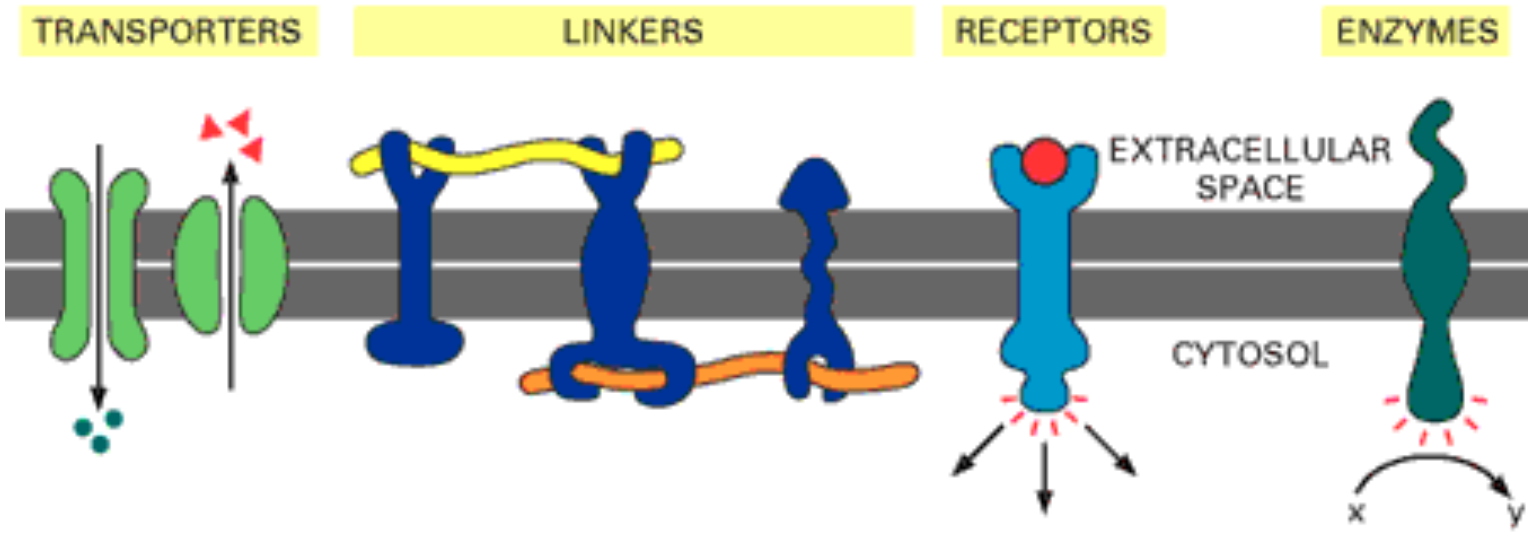
There are other non-standard aa (such as hydroxylysines) which are used to crosslink the chains.

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

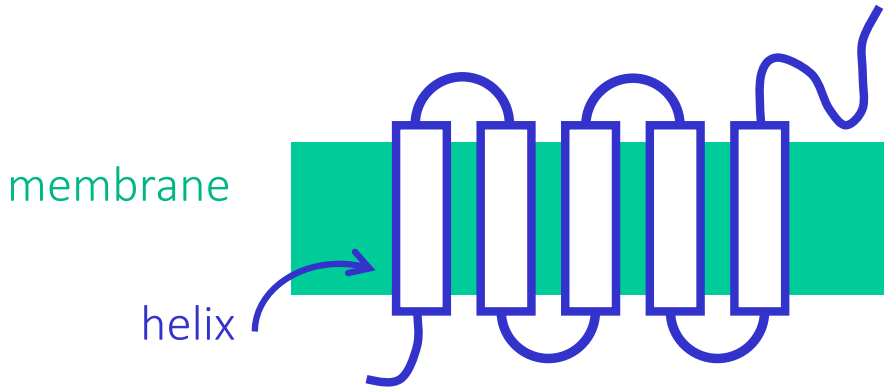
control the flow of chemicals and information between the inside and the outside of the cell and mediate communication between different cells.



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.

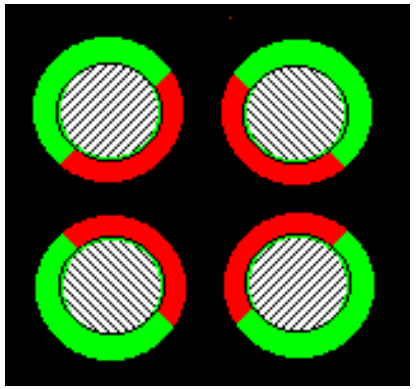
They are often built of α-helices spanning the membrane; but some are built of extended β-barrels (such as porins)



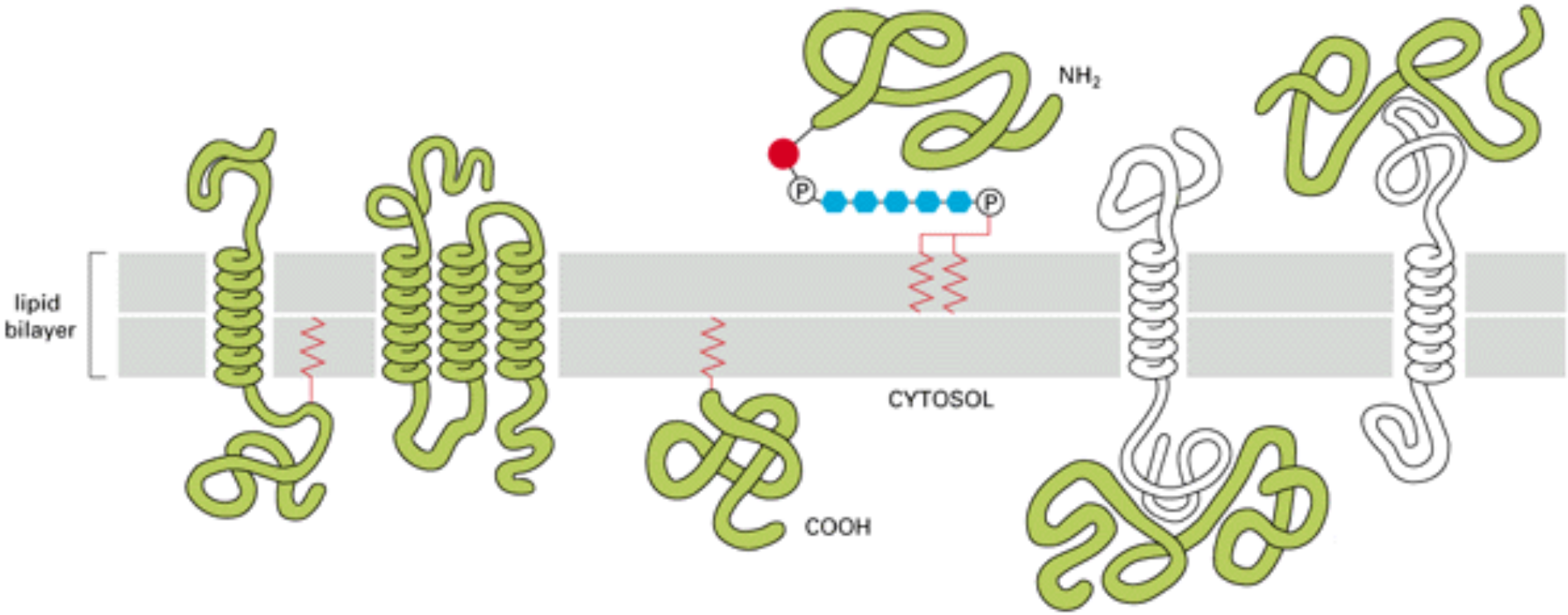
Contrary to soluble proteins, the hydrophobic residues will be on the outside, where they will interact with the chains of the lipids, while hydrophilic side chains will cluster inside

hydrophilic

hydrophobic



Membrane proteins associate with the lipid bilayer in various ways:

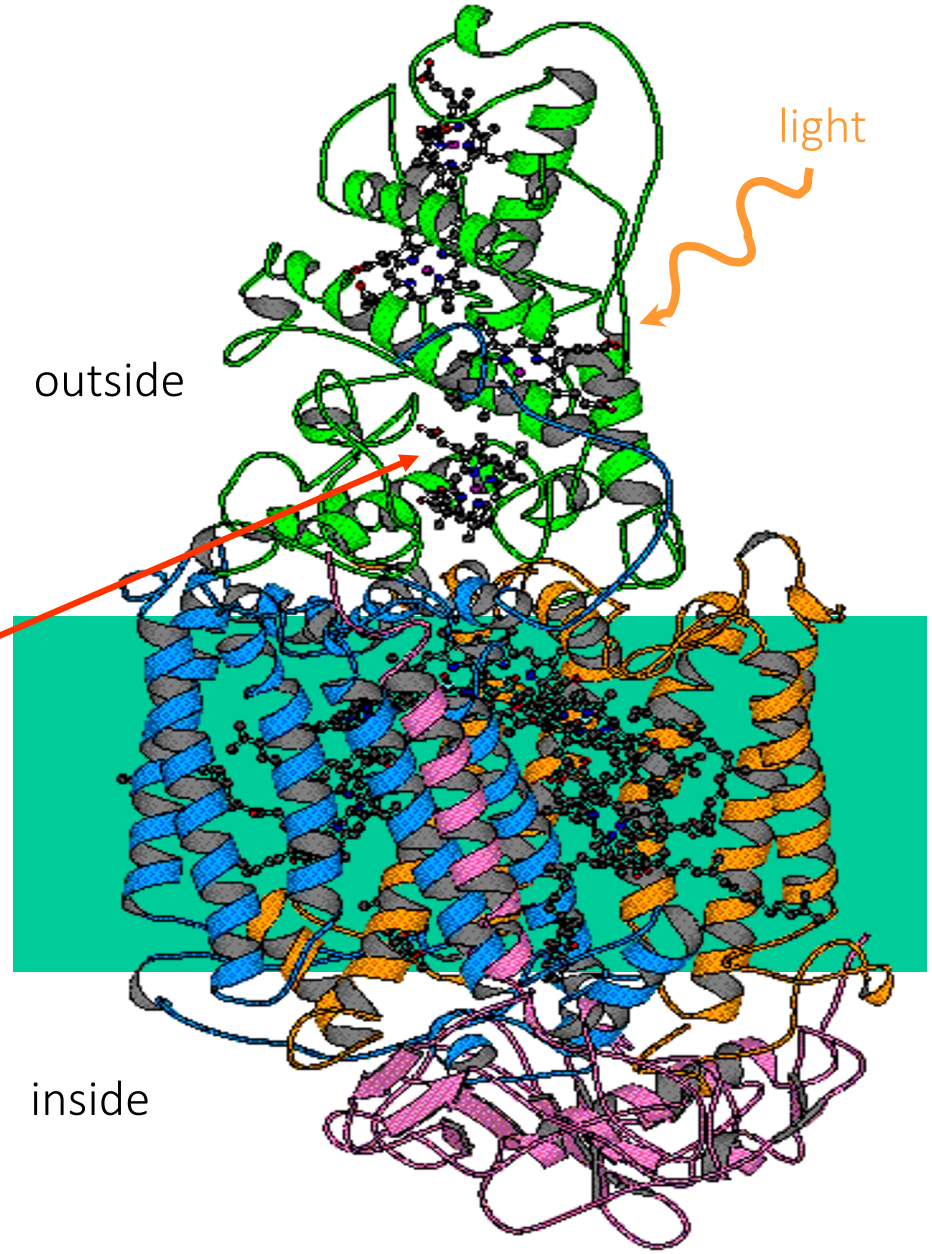


Examples of membrane proteins: the photosynthetic reaction centre

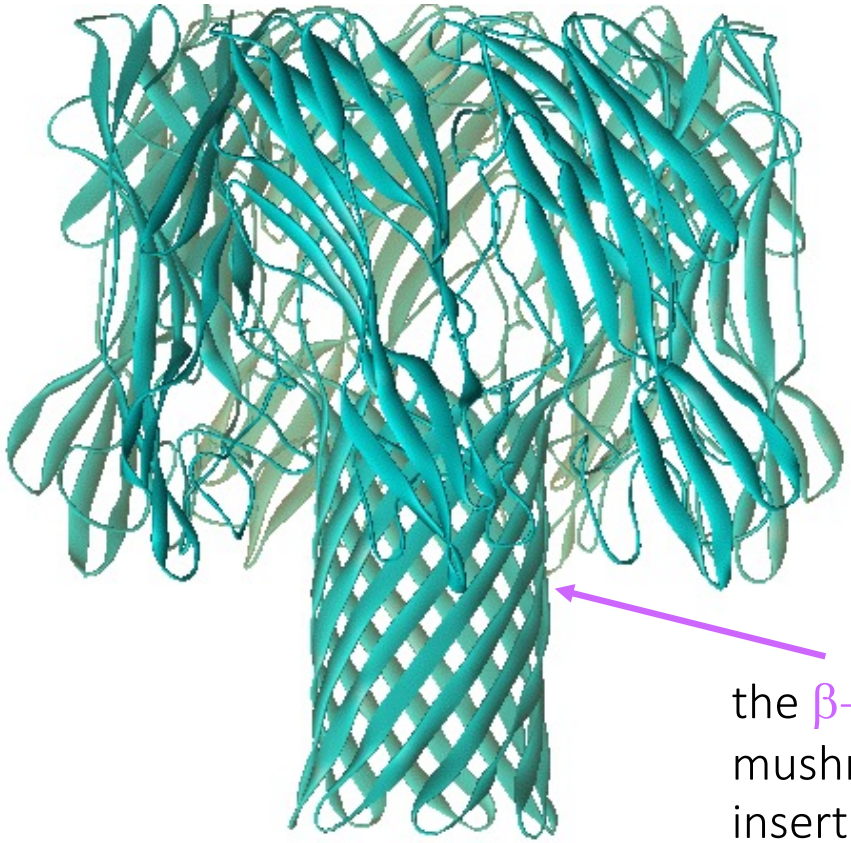
Structure determined by R. Huber, H. Mitchell & H. Deisenhofer (Nobel prize 1988)

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

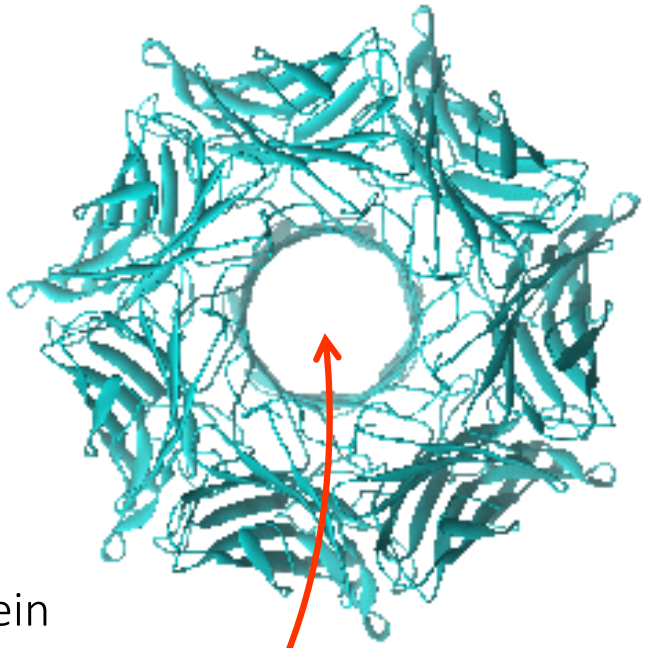


Examples of membrane proteins: the bacterial toxin a-hemolysin



seen from the side

seen from the top

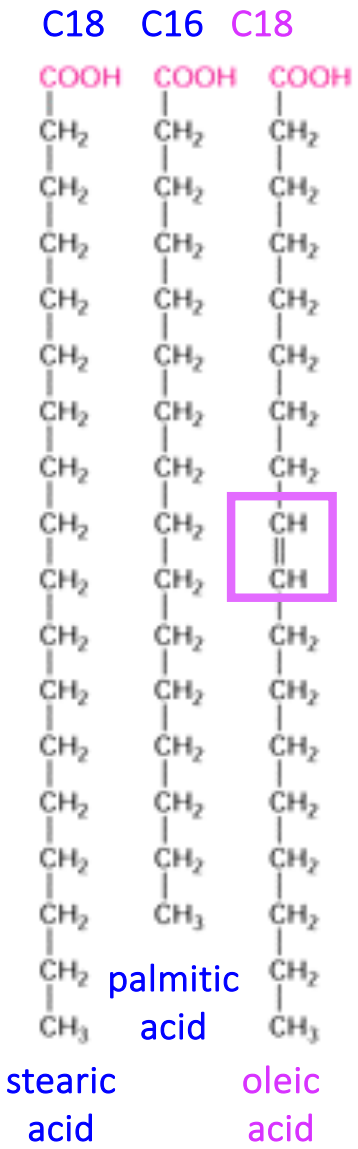


the β-barrel stalk of this mushroom shaped protein insert across the cell membrane and causes lethal permeability changes due to the central pore

Lipids

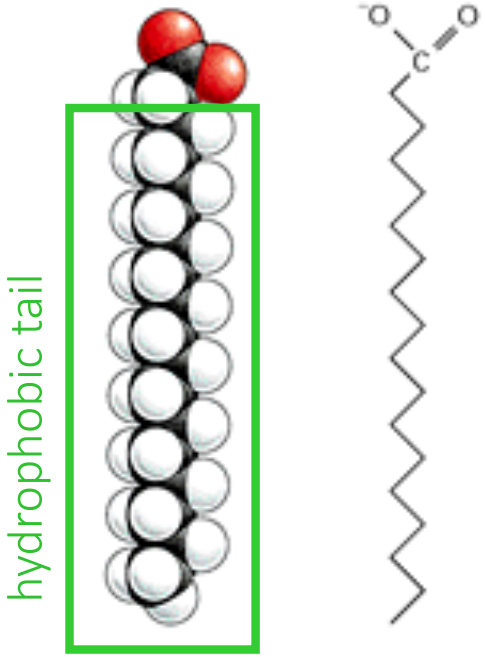
Fatty acids

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

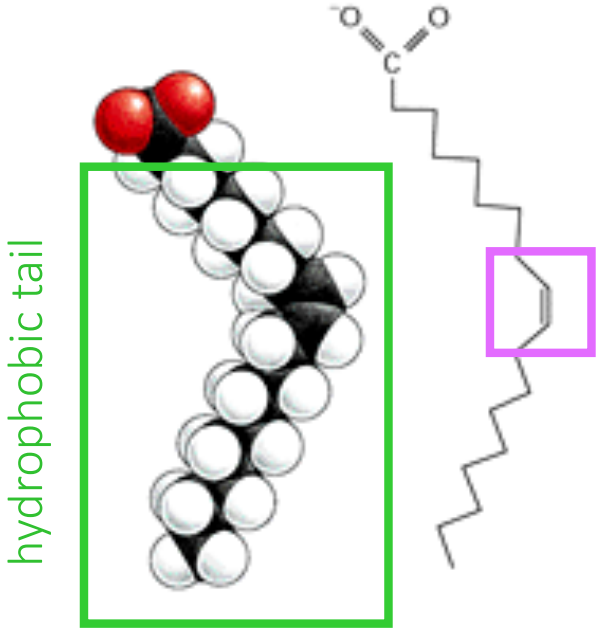


Some have one or more double bonds and are called unsaturated. The double bond is rigid and creates a kink in the chain; the rest of the chain is free to rotate

Stearic acid - saturated



Oleic acid - unsaturated



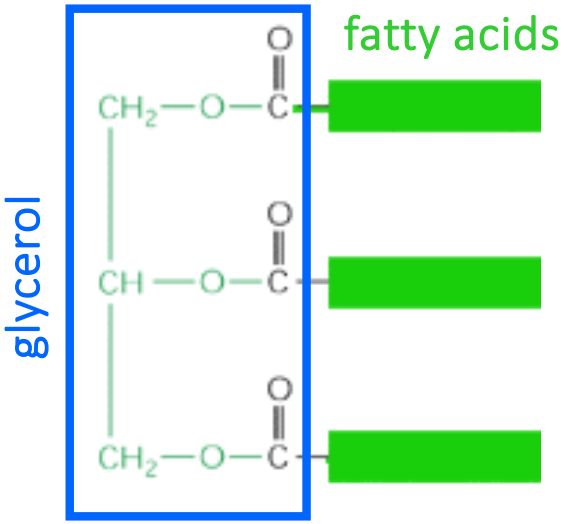
Fatty acids are used as E storage

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.

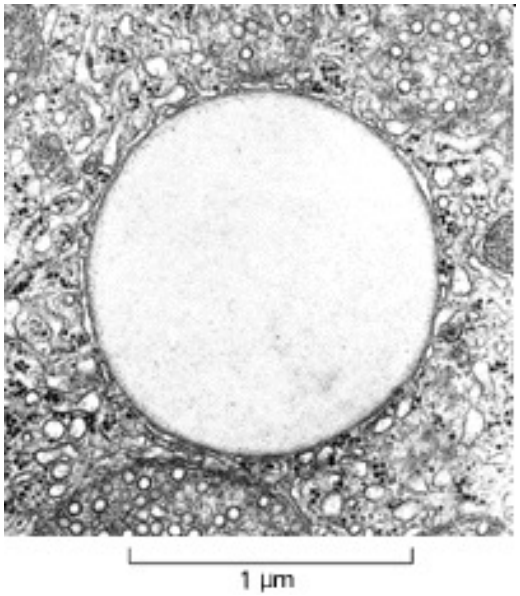
A fat molecule is composed of three molecules of fatty acid linked to glycerol: triacylglycerols (*triglycerides*).

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.

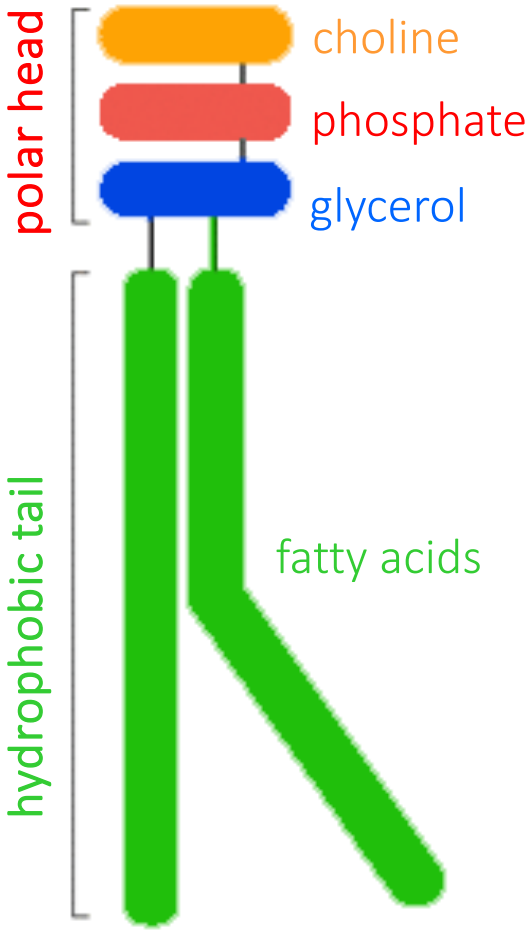


fat droplet

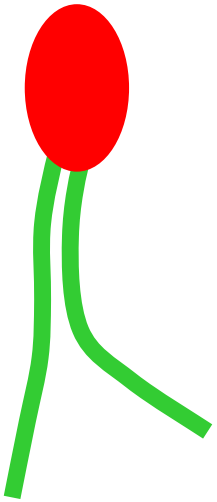


Phospholipids

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



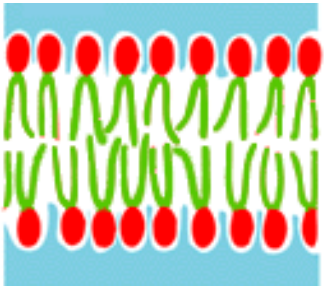
Phospholipids and membranes

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.

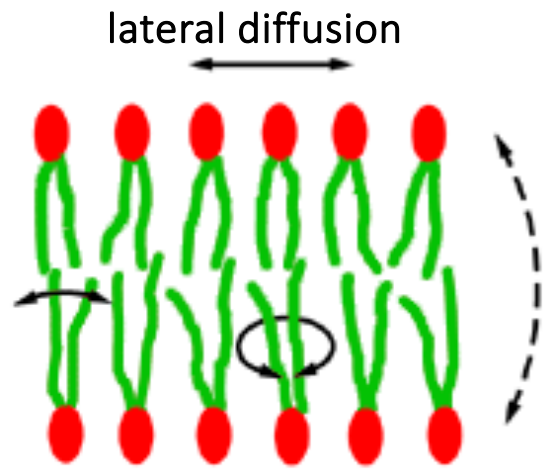
Cellular membranes are essentially made up by phospholipid bilayers.

micelle

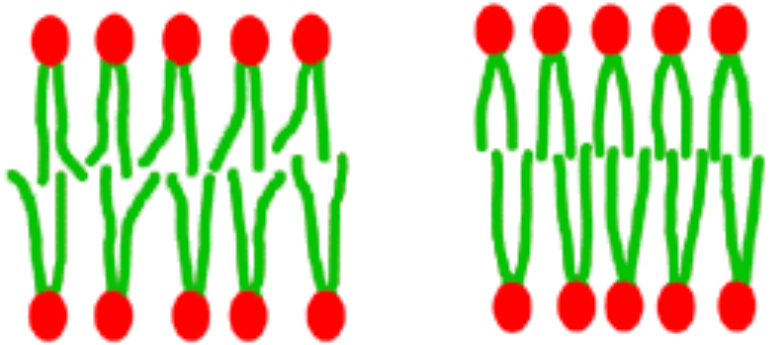


lipid bilayer

Phospholipid mobility



Saturated vs unsaturated lipids

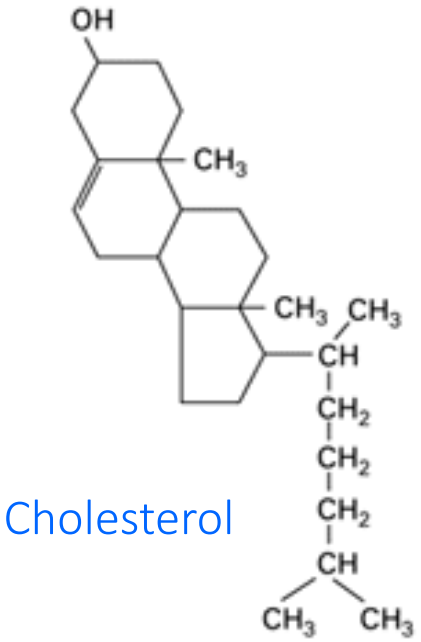


unsaturated, more difficult to pack, more fluid

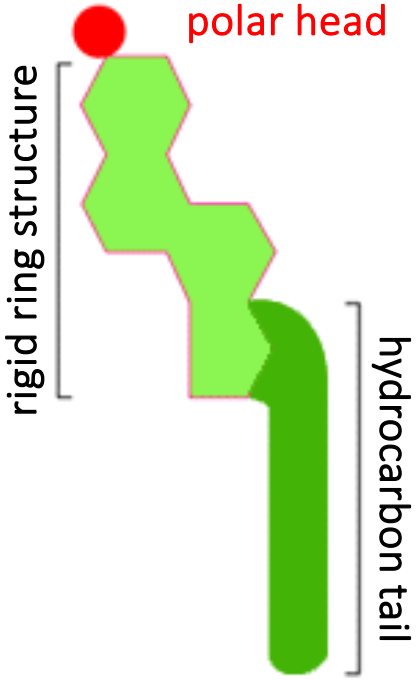
saturated, easier to pack, less fluid

Cholesterol and steroids

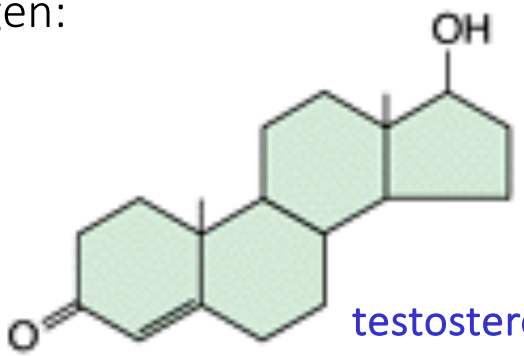
Steroids (such as cholesterol) have a rigid structure made up by 4 rings.



Cholesterol

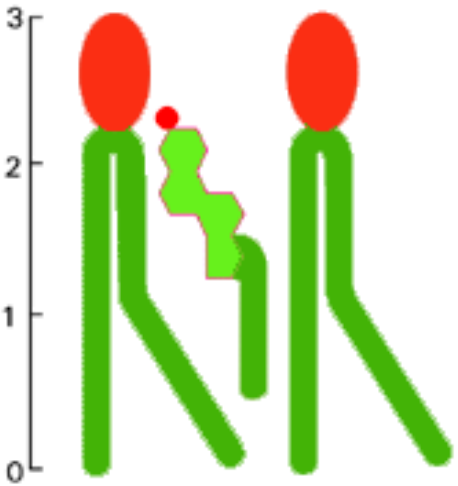


Other important steroid are the sex hormones, such as testosterone and estrogen:



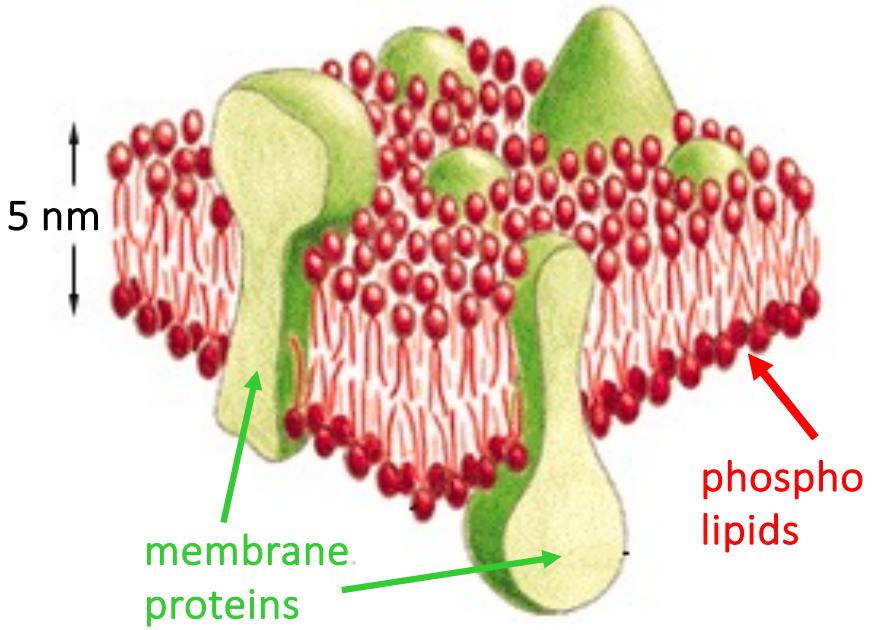
testosterone

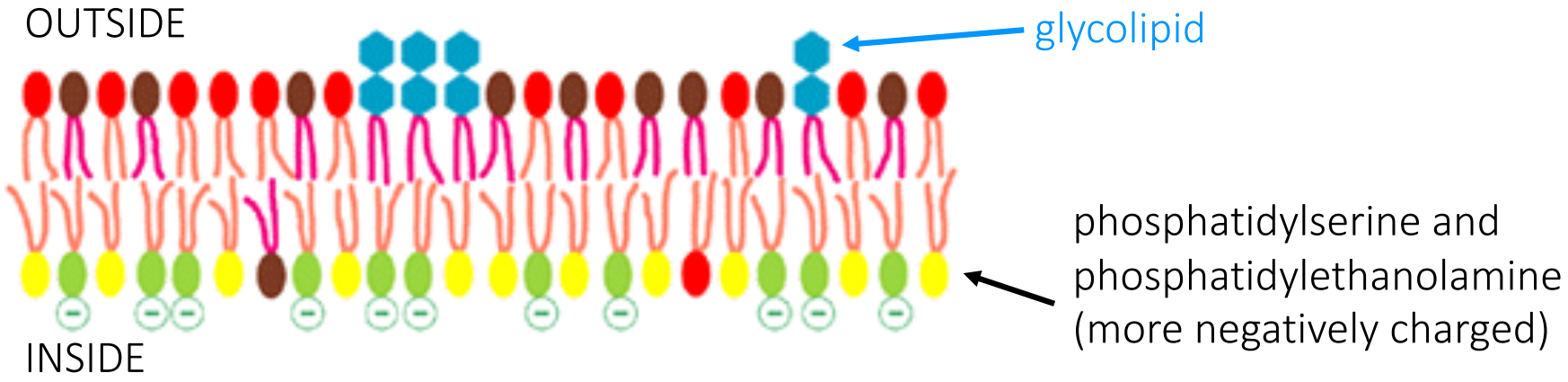
Cholesterol is an important component of the eukaryotic membranes and has a key role in controlling the membrane fluidity.



Cell membranes

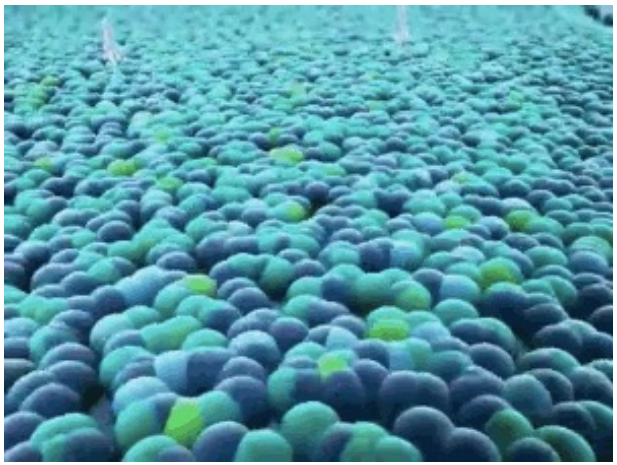
- biological membranes are fluid
- the fluidity is controlled by the % of saturated/unsaturated fatty acid and the % of cholesterol
- membranes are impermeable to ions and most polar molecules (H_2O is actively transported in)
- many proteins are embedded in the membrane
- the membrane is highly asymmetric

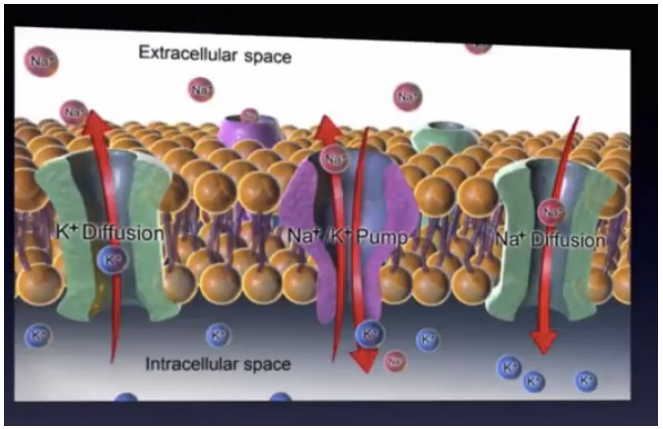


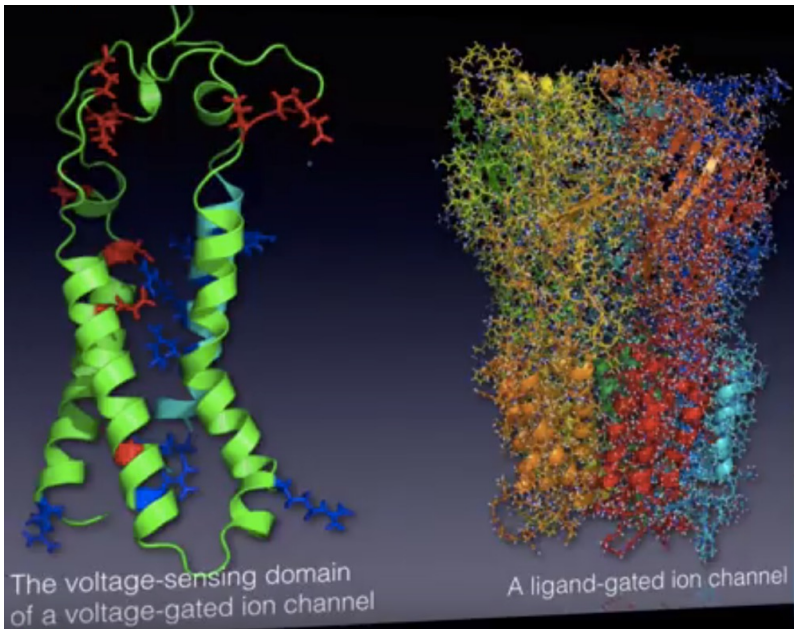


Cell membranes

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?



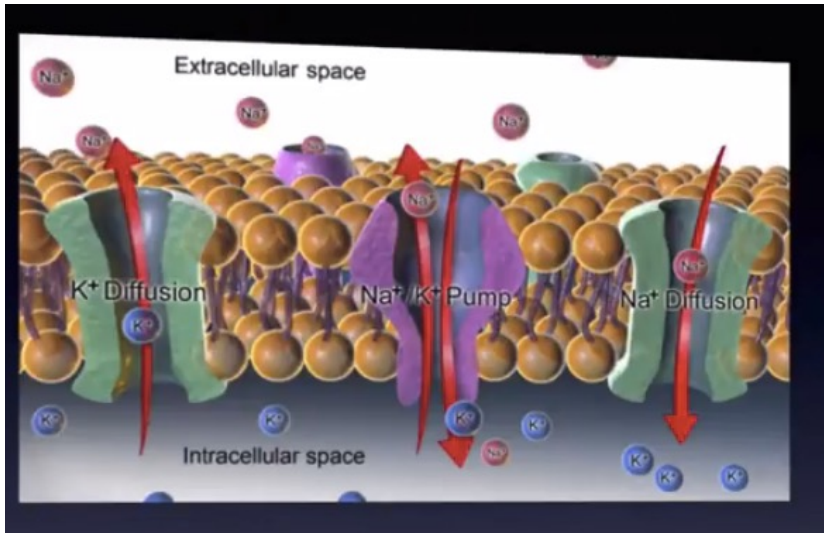


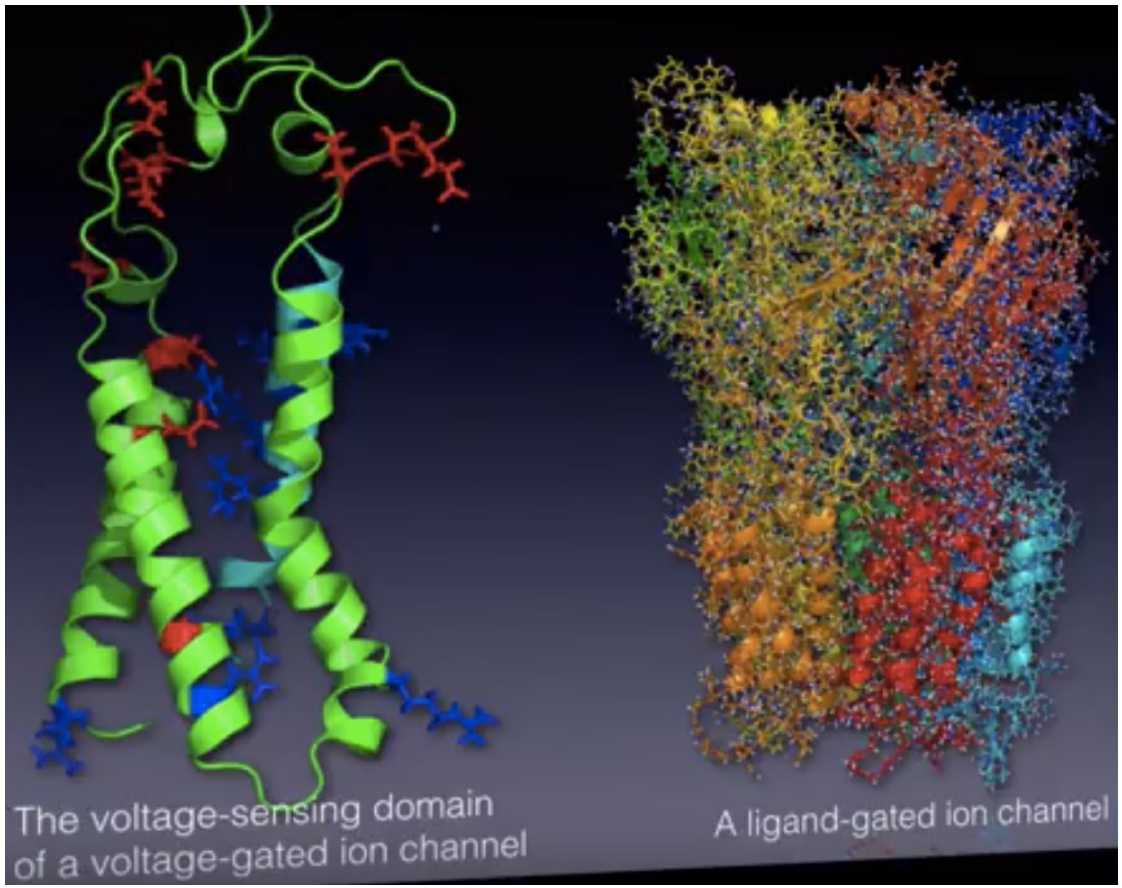


Cell is a highly organized and orderly structure: does it not obey the second law of thermodynamics?

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.

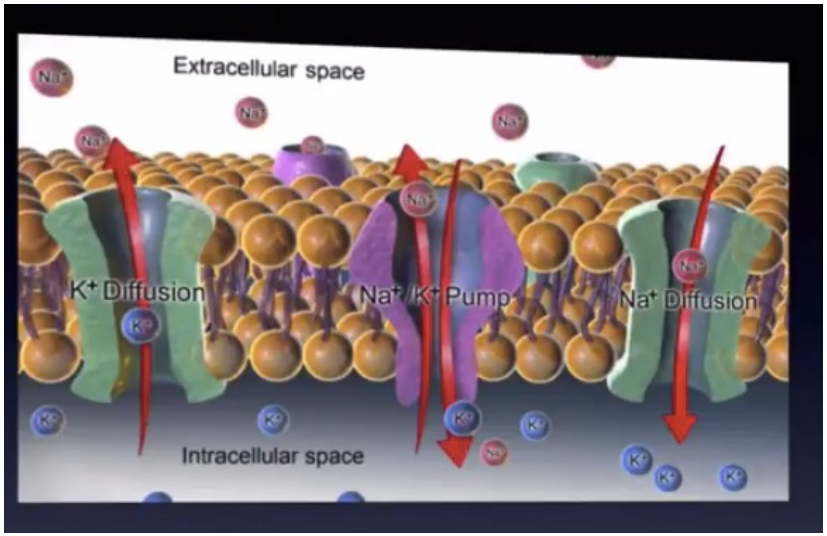


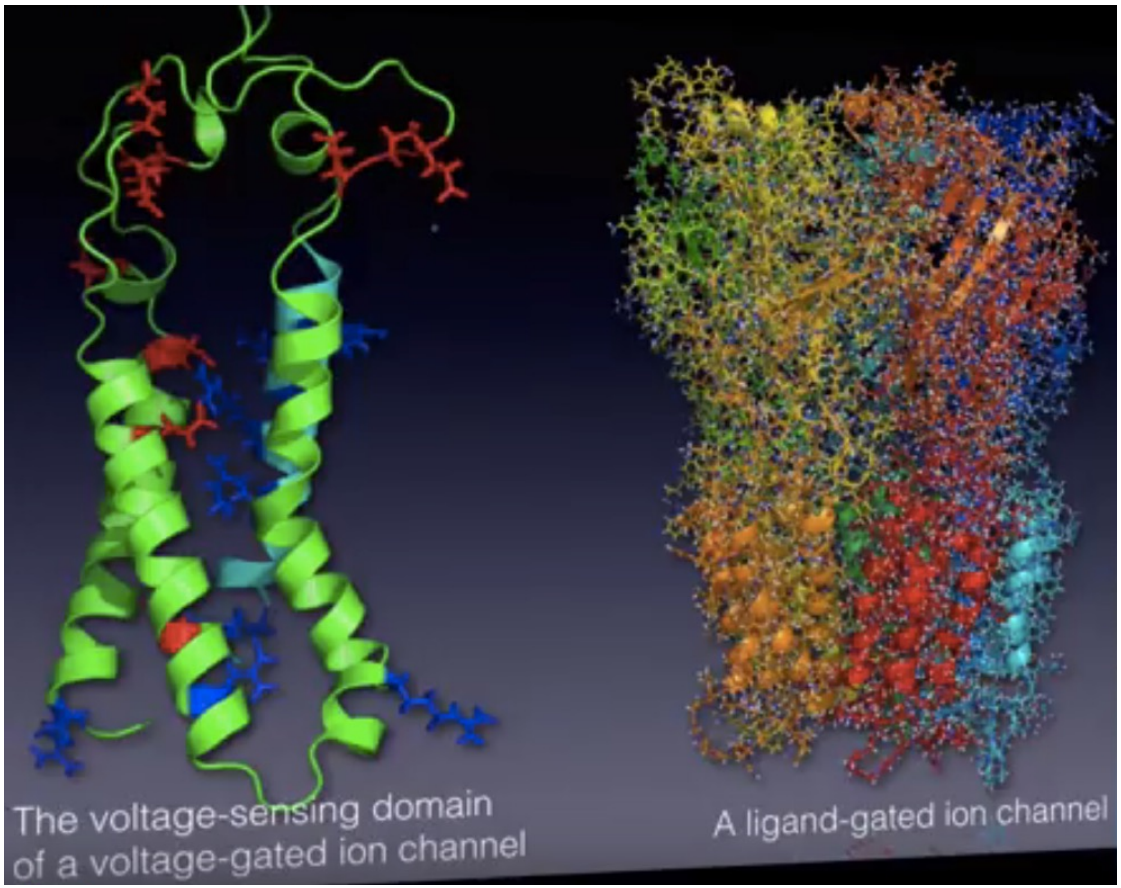


Biology is about distribute/generate/consume energy.

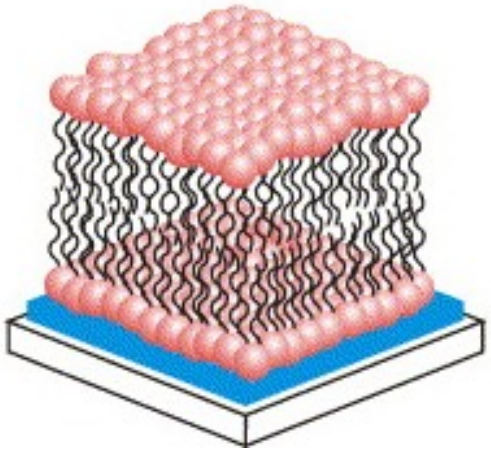
Ion channels are the voltage-sensing domains of a voltage-gated access (heart beats, nerve impulse), which make the cell working as a battery (selective opening of the channel).
Pumps are enzymes in the membrane which move ions in counter direction (ATPase).

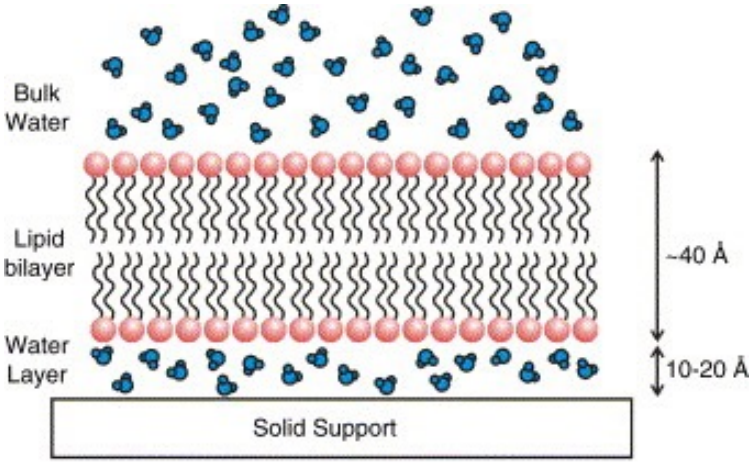
Ligand-gated channels are part of the nerve system (receiving side of the synapsis) and are amazingly specific! One mistake every billion (disease!)





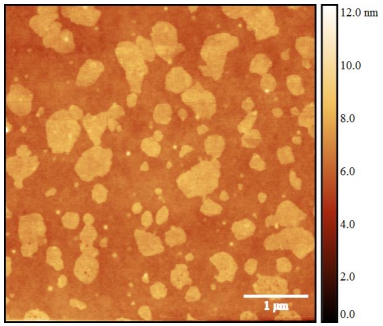
Artificial lipid bilayers



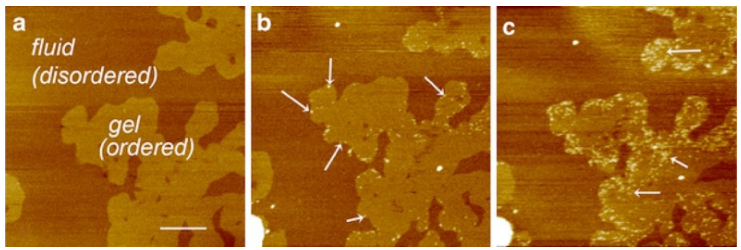


✓ Multicomponent model membranes for mimicking cellular membranes

Lipids behaviour



Protein-lipid interactions



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

Carbon nanotubes interactions