

Biophysics and structural bioinformatics I

D. Gilis & M. Rooman

Unité de bioinformatique génomique & structurale
UD3.203/UD3.204 (bâtiment U, campus du Solbosch)

Tél: 02/650.36.15 - 02/650.20.67

e-mail: dgilis@ulb.ac.be / mrooman@ulb.ac.be

Documents at: <http://uv.ulb.ac.be> and
<http://babylone.ulb.ac.be/~dgilis/MA1bioinfo.php>

Part 1

Introduction

1. Structure of biomolecules

- 1.1. Proteins
- 1.2. DNA
- 1.3. RNA

2. Experimental approaches to resolve the structure of biomolecules

2.1. Resolution of the 3D structure

- 2.1.1. X-ray crystallography
- 2.1.2. Nuclear magnetic resonance (NMR)

2.2. Information about the structure of biomolecules

- 2.2.1. Principle of emission and absorption spectroscopy
- 2.2.2. Circular dichroism (CD)
- 2.2.3. Infra-red spectroscopy (IR)

2.3. Thermodynamic stability of biomolecules

3. Intra- and inter-molecular interactions

- 3.1. Electrostatic interactions
- 3.2. Van der Waals interactions
- 3.3. Hydrogen bond
- 3.4. Amino- and cation- π
- 3.5. π - π stacking
- 3.6. Hydrophobic effect

Aims of this course:

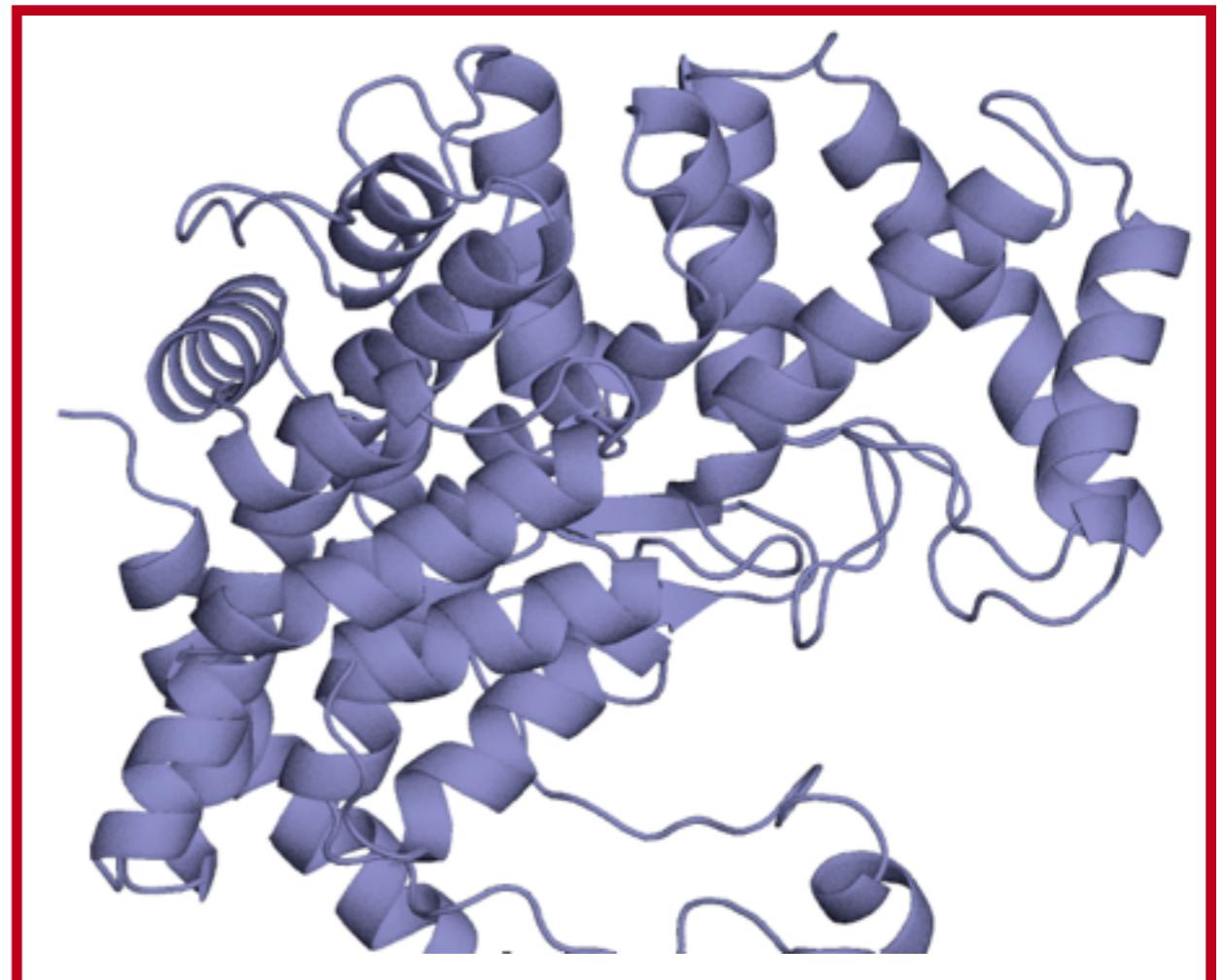
- To present biophysics approaches that are used to obtain information about the structure of biomolecules and their interactions.
- To present biophysics techniques that are used to analyze, to study, to predict the structure of biomolecules.
- To describe the methods used by these bioinformatics techniques: to understand the methods to choose those that are appropriate to solve a problem.
- To develop the capacity to be autonomous et to be critical to: (1) search for a bioinformatics program to solve a problem; (2) be aware of the possibilities and the limitations of the program.
- To learn to use bioinformatics program during practical lessons.

Introduction

We will consider these biomolecules: DNA, RNA, proteins. Why do we want to analyze their structure ?

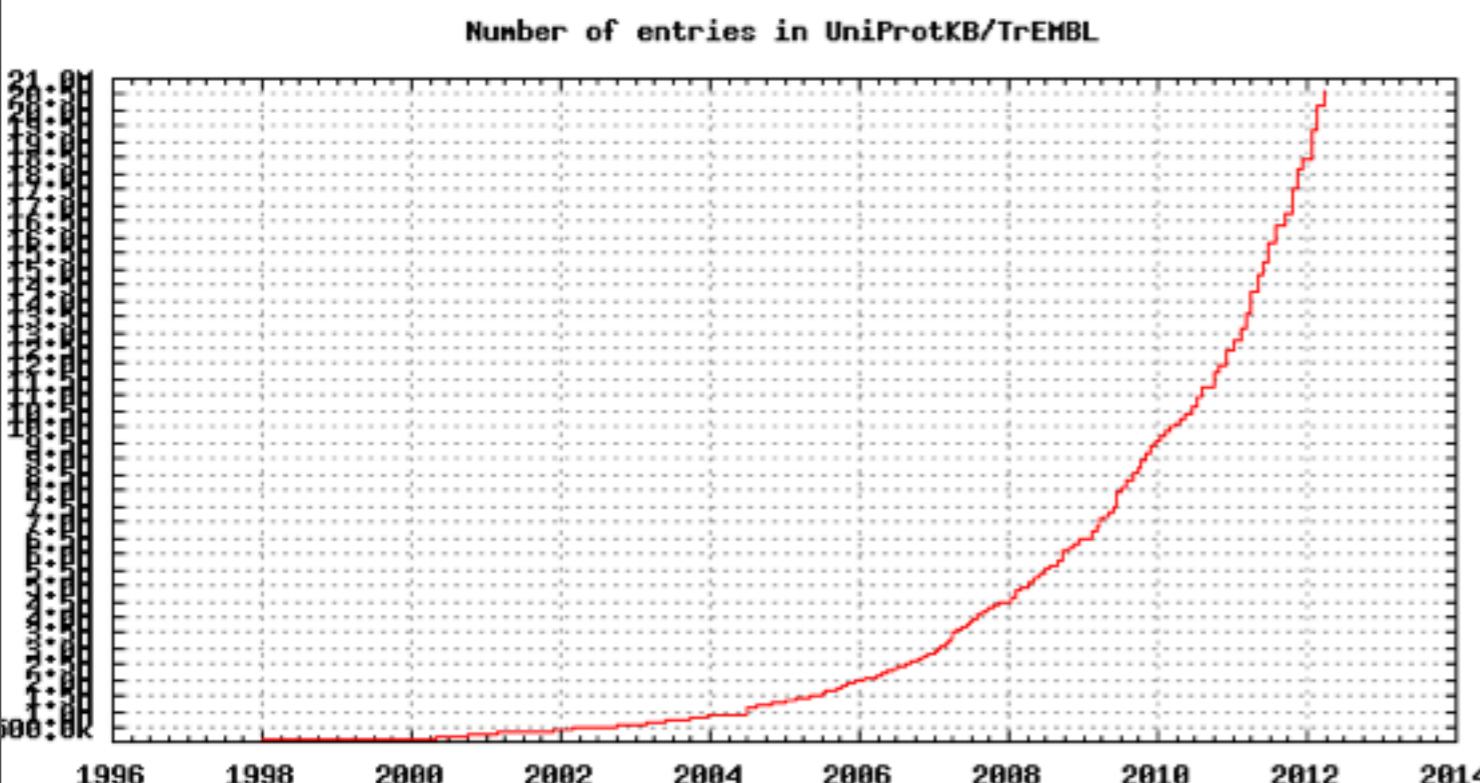
- Structural alignments → to analyze structural/functional patterns
- Methods to predict the structures

MSAKAISEQTGKEFLYKYICTSSAIQNRFKYARVT
PDTDWARLTQDHWPLLSERLVVKPDQLIKRRGKLG
LVGINLTLQVKVWLKQHLGQETTIANAKGILKNF
LIEPFVPHKQEEEFYVCIYAAREGDYVLFYHEGGV
DVGDVDAKAQKLLVAVDEKLNESDVKKHLLQHAPA
DKKDILASFICGLFNLYEDLYFTYLEINPLVVTKD
GVYILDLAAKIDATADYICKVKWGDVEFPPPFGRE
AYPEEAYIADLDAKSGASLKLTILNPKGRIWTMVA
GGGASVVYSDTICDLGGVNELANYGEYSGAPSEQQ
TYDYAKTILSLMTREKHPEGKILIIGGSIANFTNV

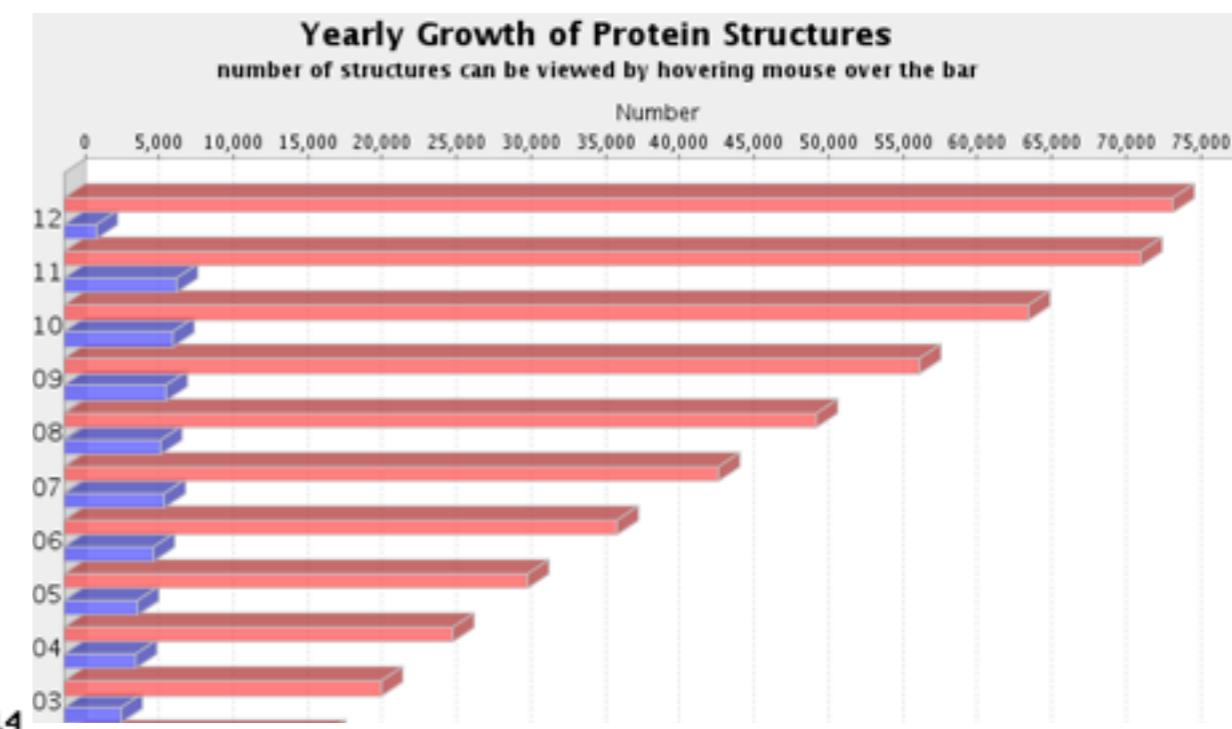


● Methods to predict the structures: why ?

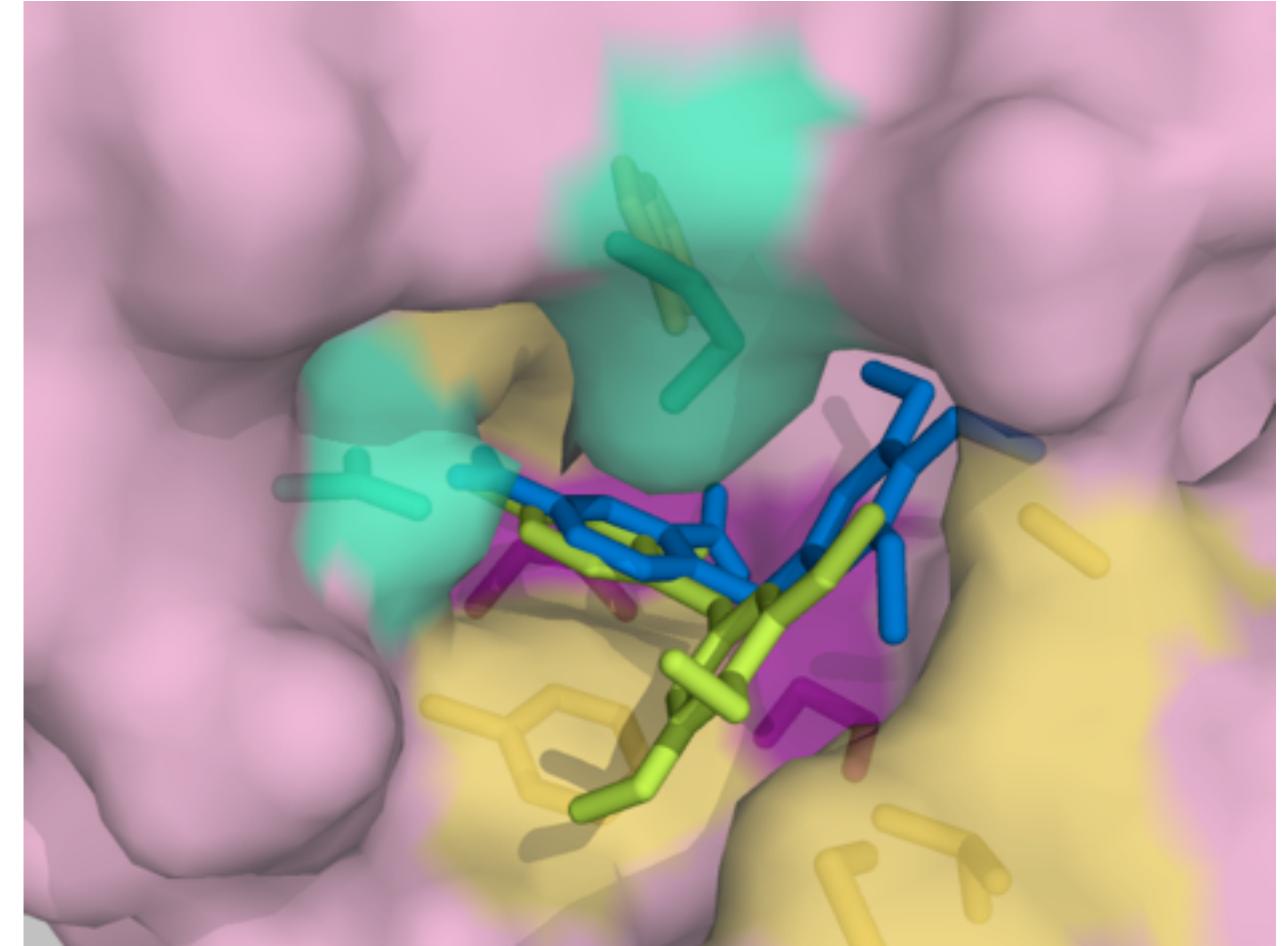
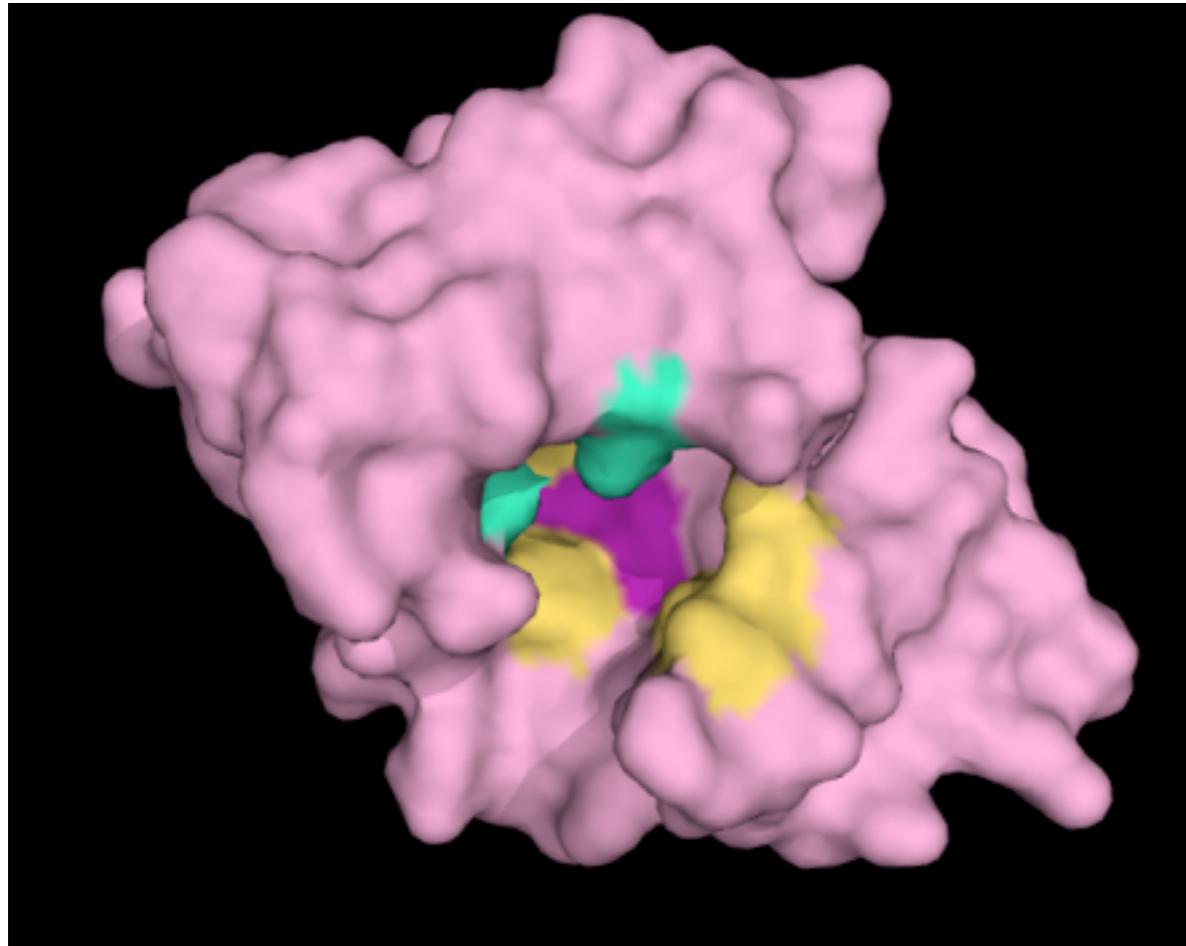
Number of sequences
April 2012: 20.639.311
July 2013: 41.451.118
(UniProtKB/TrEMBL)



Number of protein structures
April 2012: 74.612
September 2013: 86.646
(PDB)



Study of the function of proteins, design of ligands, modification of the physico-chemical properties,...



Biophysics part

Biophysics: application of physical techniques to understand et to study biological structures and functions.

1. Structure of biomolecules

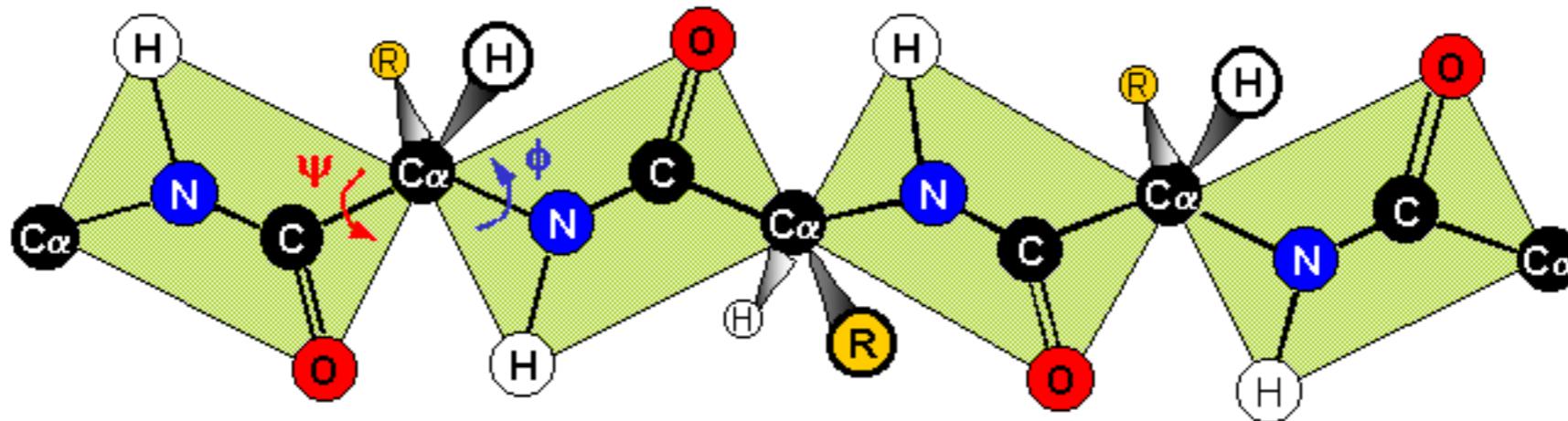
We will consider three types of biomolecules: proteins, DNA (deoxyribonucleic acid) and RNA (ribonucleic acid).

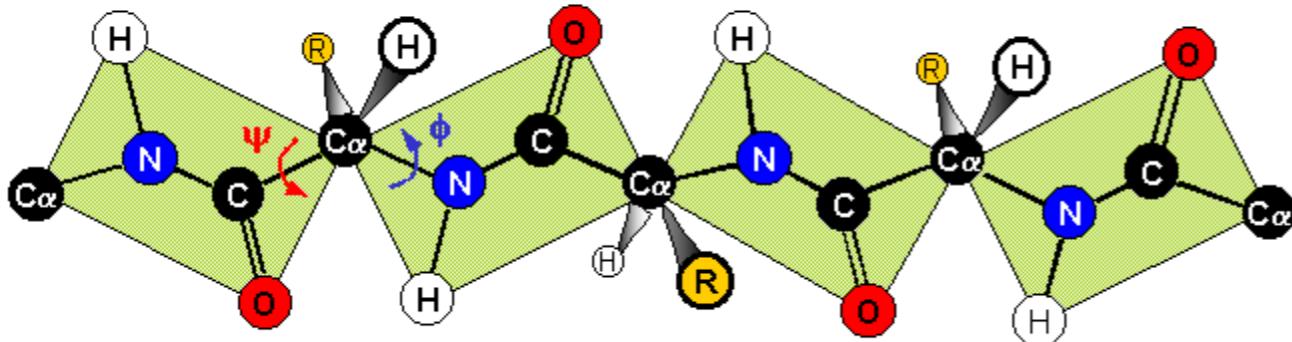
1.1. Proteins

Proteins are characterized by their amino acids sequence:

PYPFKLPDLGYPYEALEPHIDAKTMEIHHQKHHGAYVTNLNAALEKYPYLHGVEVEVLLRHLAALPQDIQTAVRNNGGGH
LNHSILFWRLLTPGGAKEPVGELKKAIDEQFGGFQALKEKLTOQAAMGRFGSGWAWLVKDPMGKLHVLSTPNQDNPMEGFT
PIVGIDVWEHAYYLKYQNRRADYLQAIWNVLNWDVAEEFFKKA

= primary structure

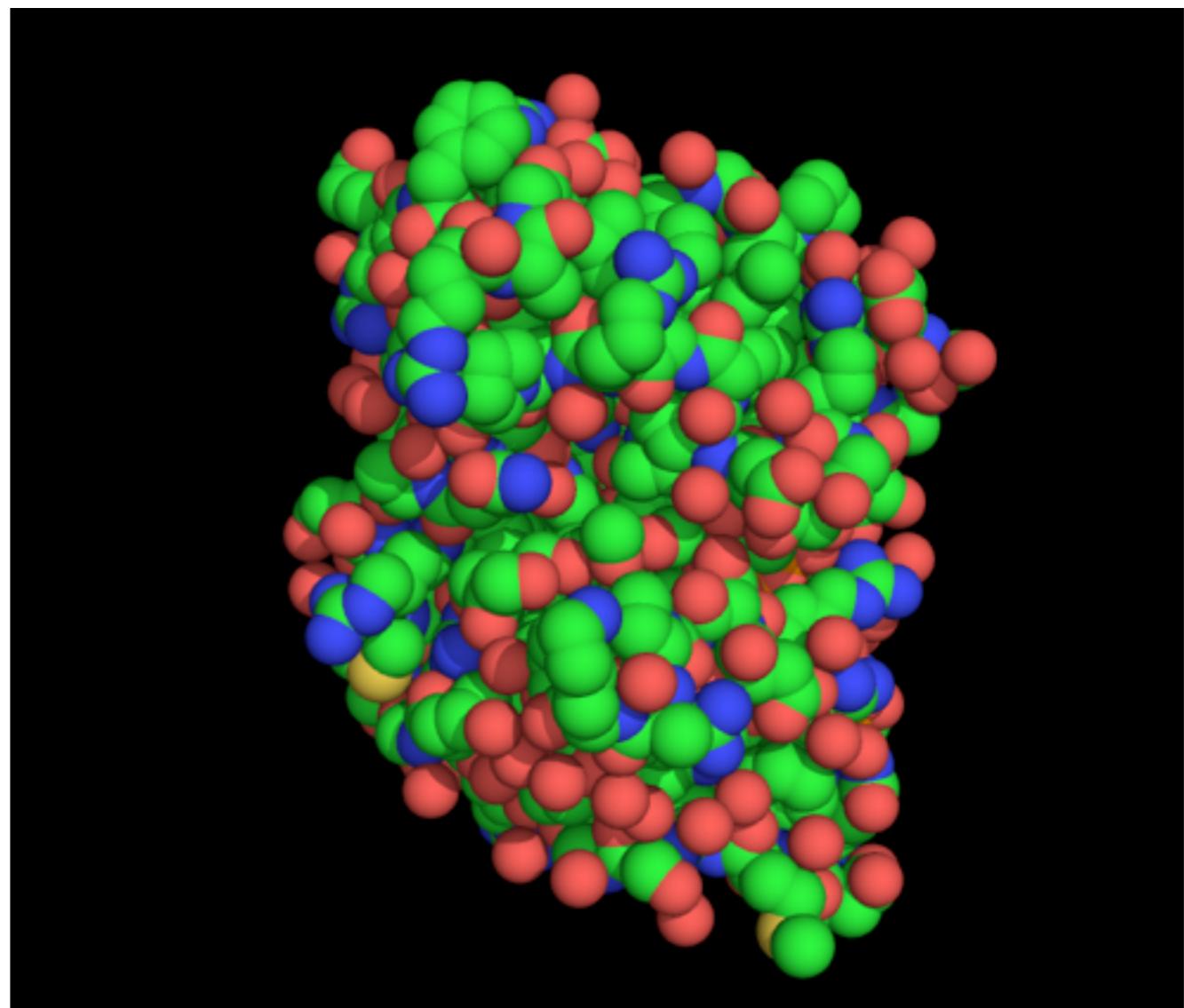


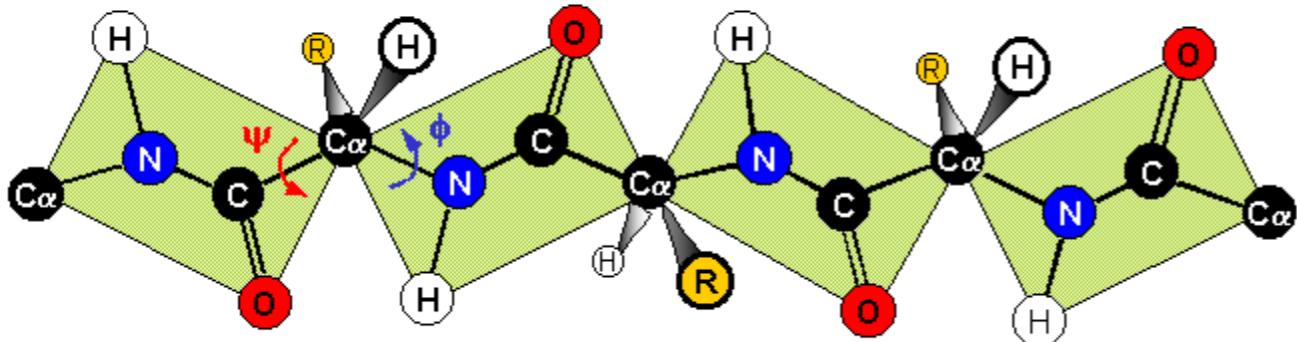


Amino acids are composed by several atoms. These atoms will interact (Van der Waals, electrostatic, hydrogen bonds, hydrophobic effect, ...).

The protein will fold and adopt a 3D structure.

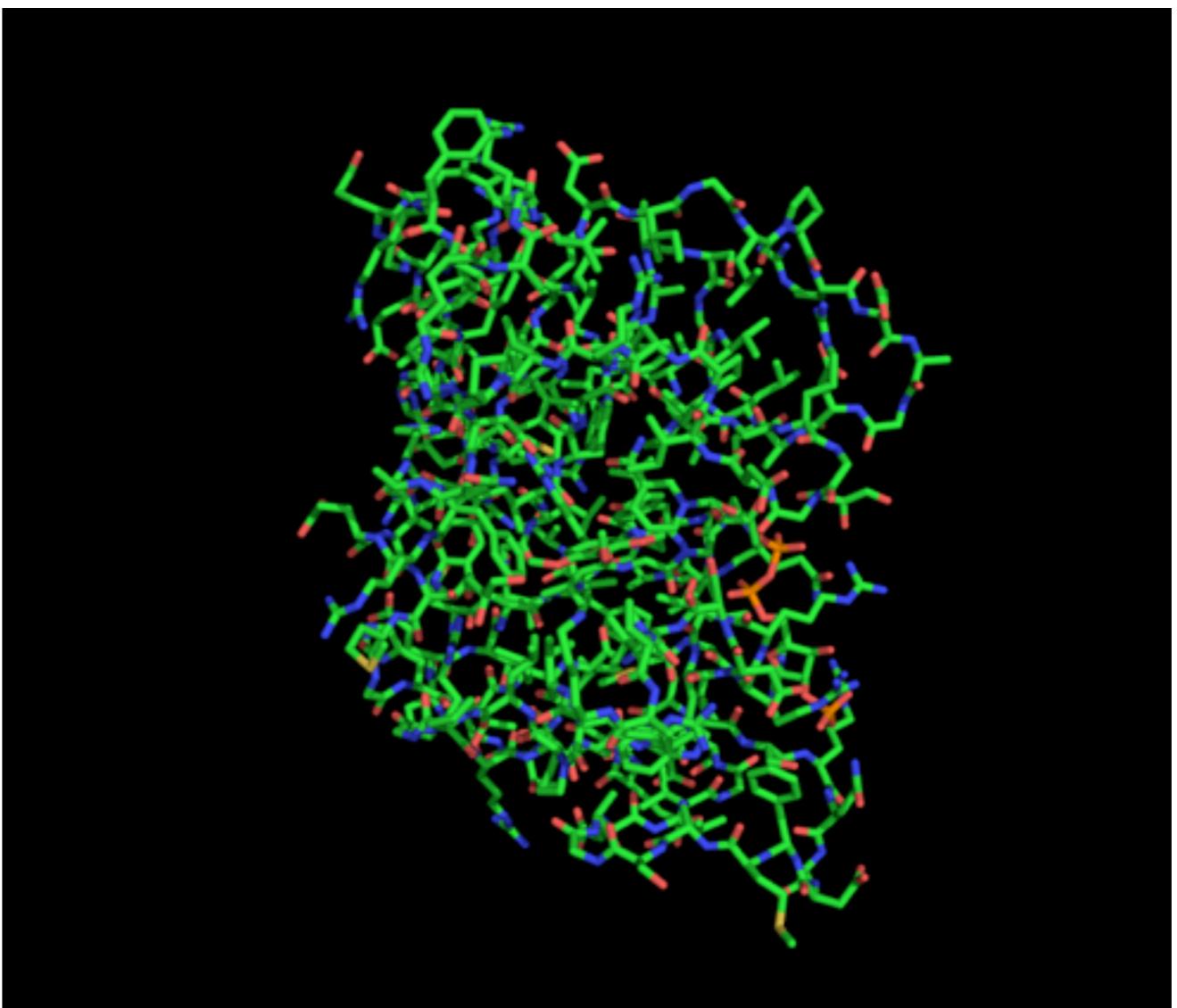
Volume of each atom

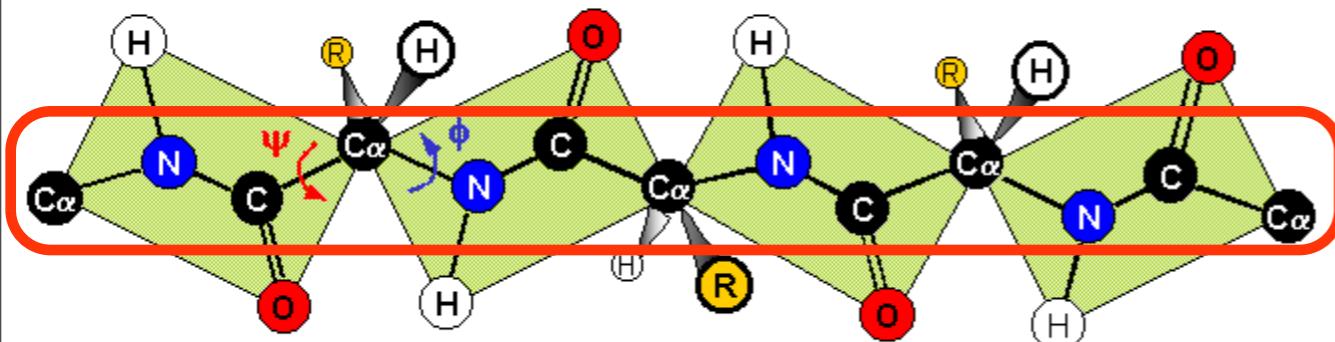




Amino acids are composed by several atoms. These atoms will interact (Van der Waals, electrostatic, hydrogen bonds, hydrophobic effect, ...).

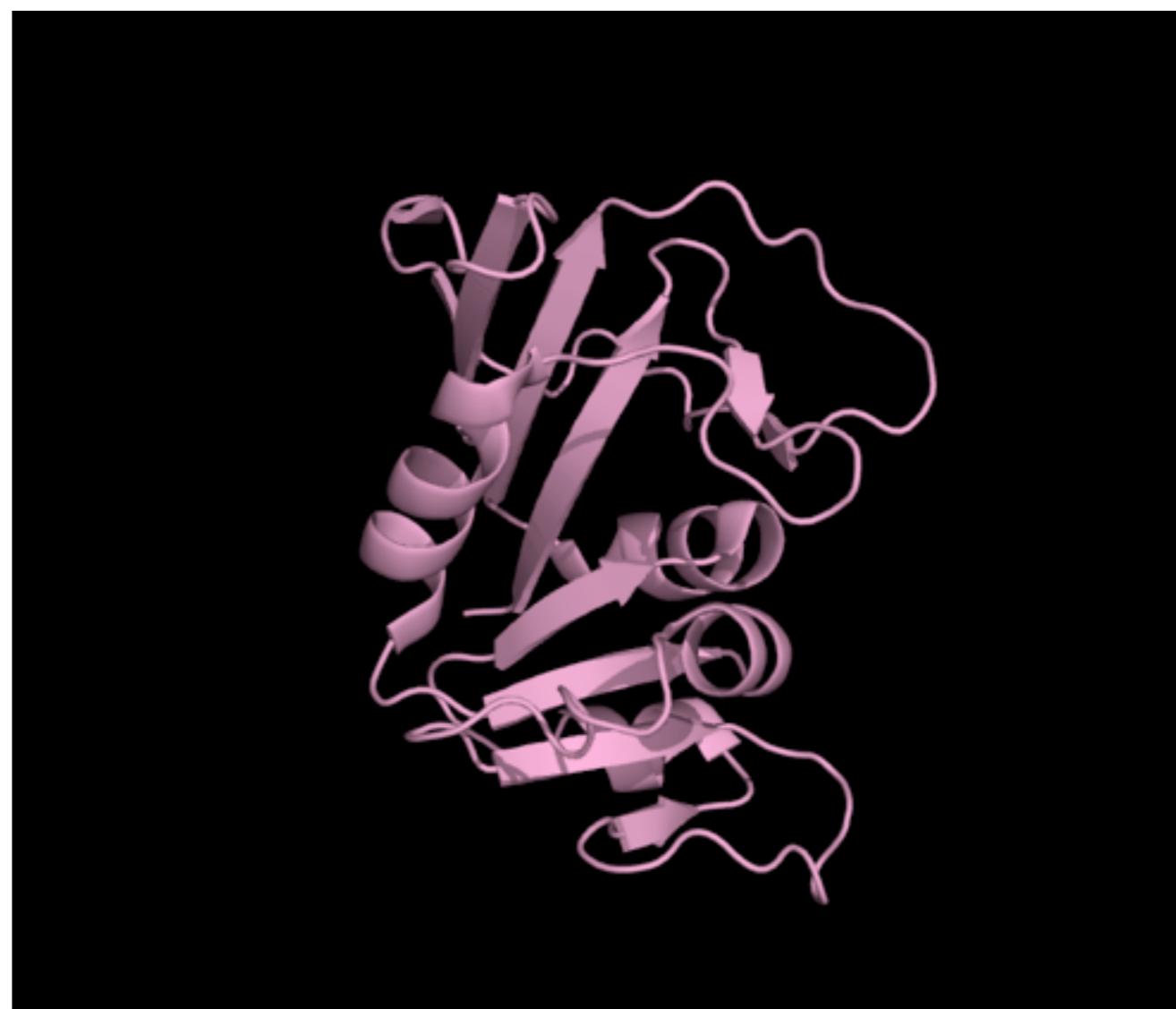
The protein will fold and adopt a 3D structure.



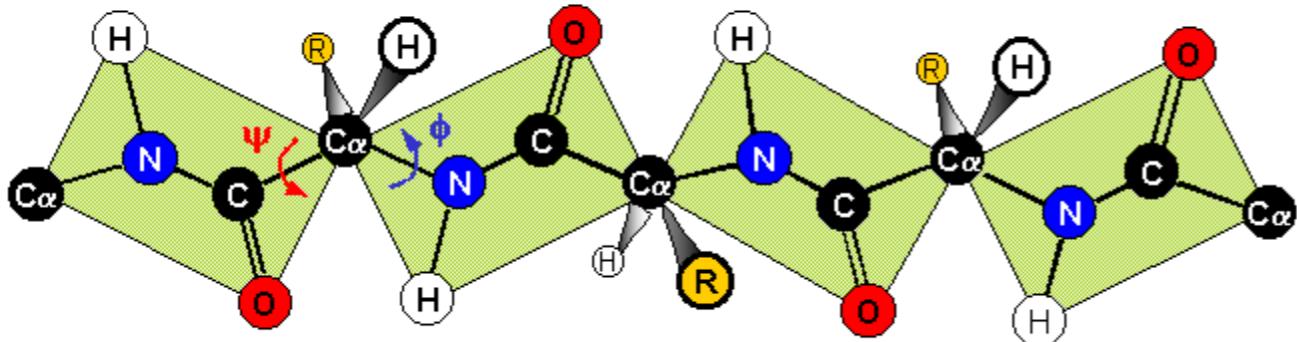


Amino acids are composed by several atoms. These atoms will interact (Van der Waals, electrostatic, hydrogen bonds, hydrophobic effect, ...).

The protein will fold and adopt a 3D structure.



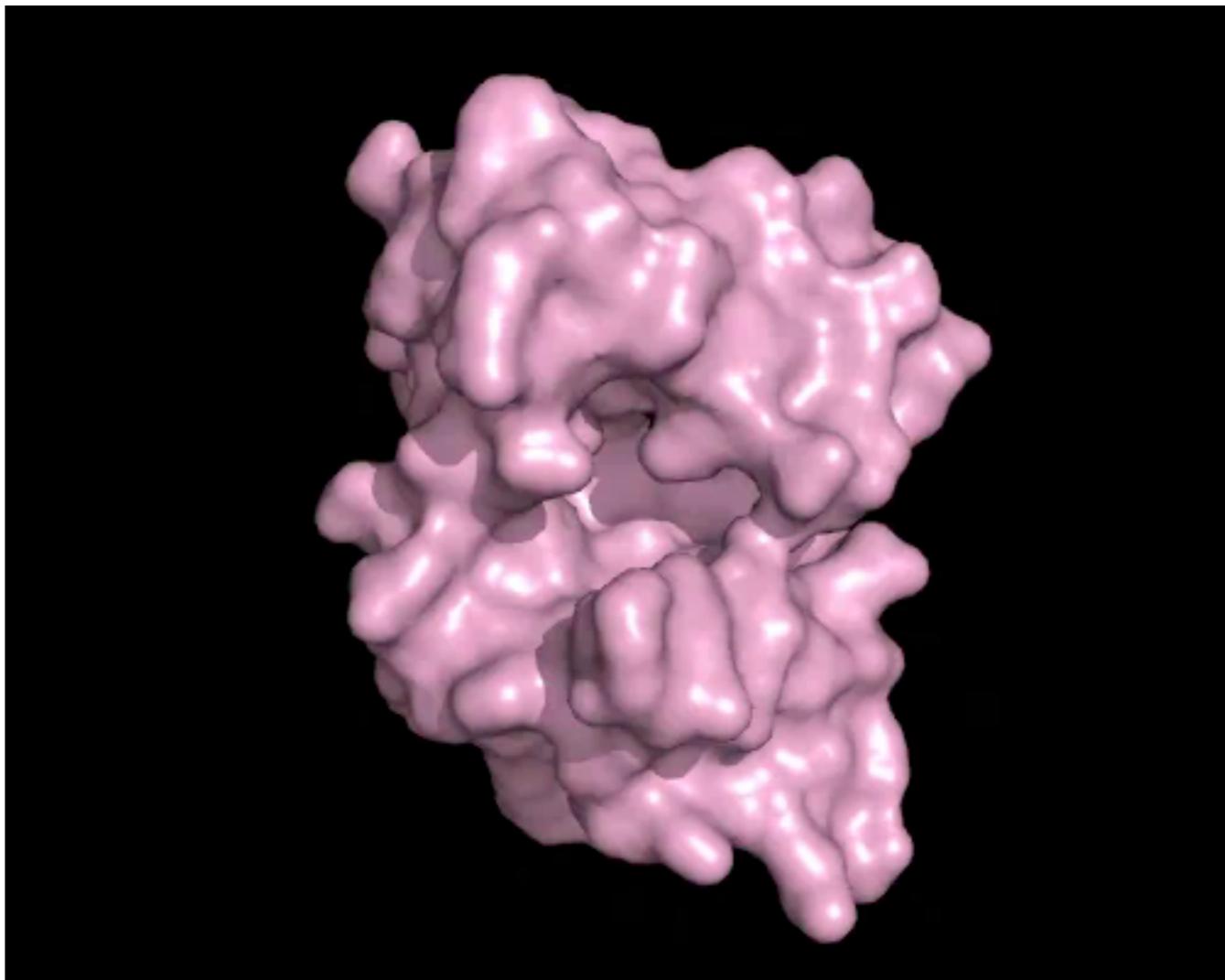
Main chain



Amino acids are composed by several atoms. These atoms will interact (Van der Waals, electrostatic, hydrogen bonds, hydrophobic effect, ...).

The protein will fold and adopt a 3D structure.

Surface



Primary structure: sequence of the protein

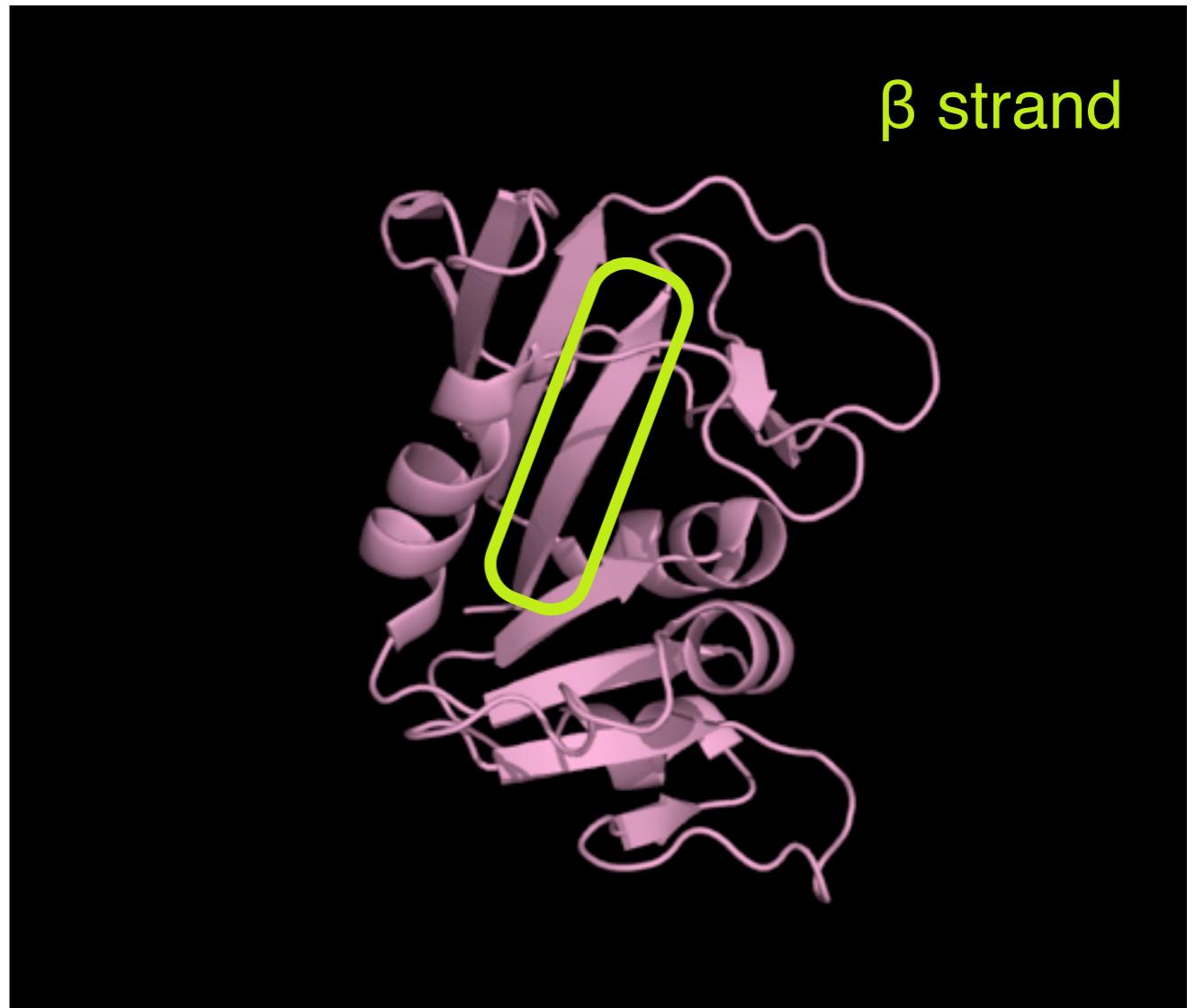
Secondary structure: local structuration of the protein.

a helix



Primary structure: sequence of the protein

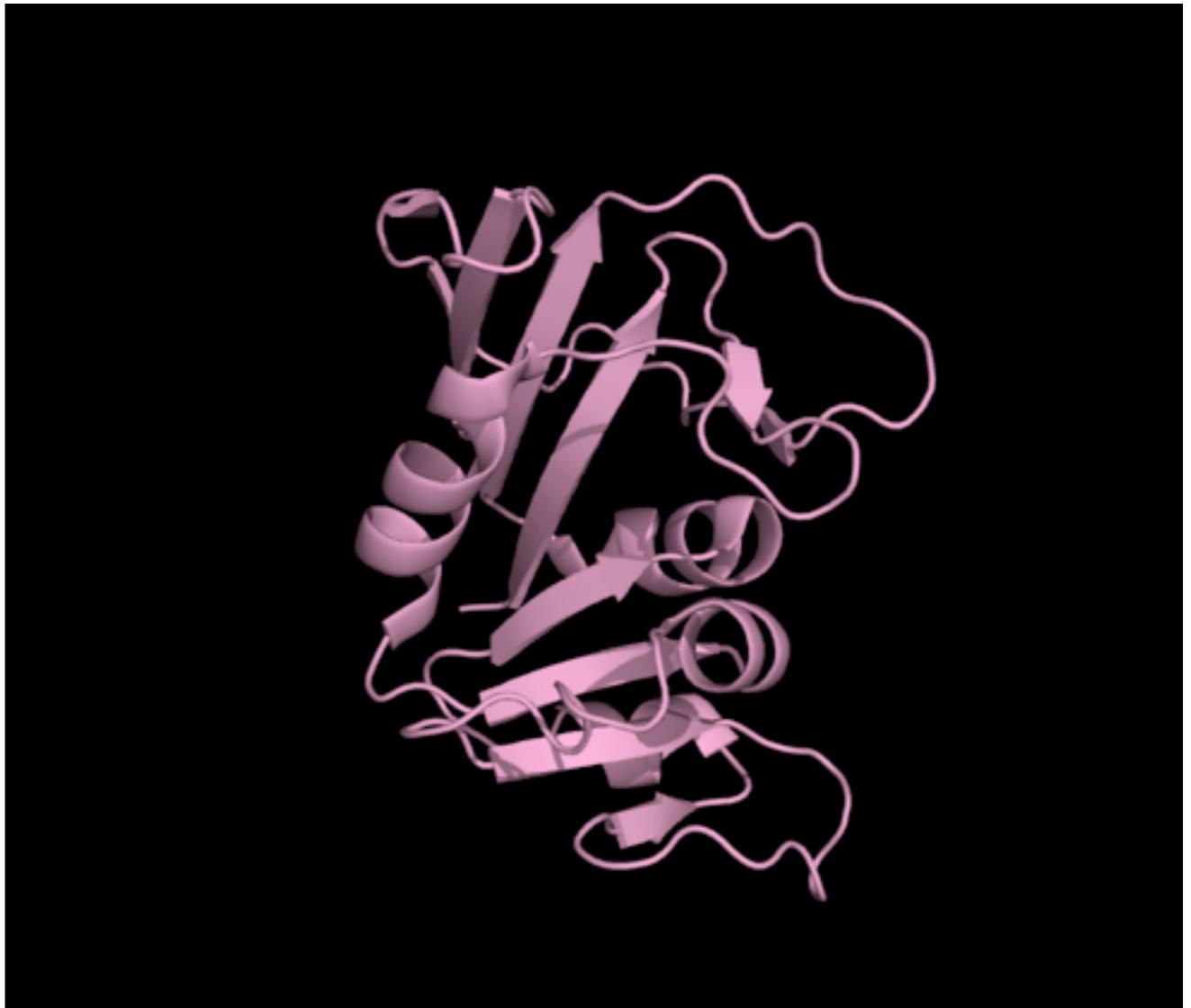
Secondary structure: local structuration of the protein.



Primary structure: sequence of the protein

Secondary structure: local structuration of the protein.

Tertiary structure: 3D structure of the protein

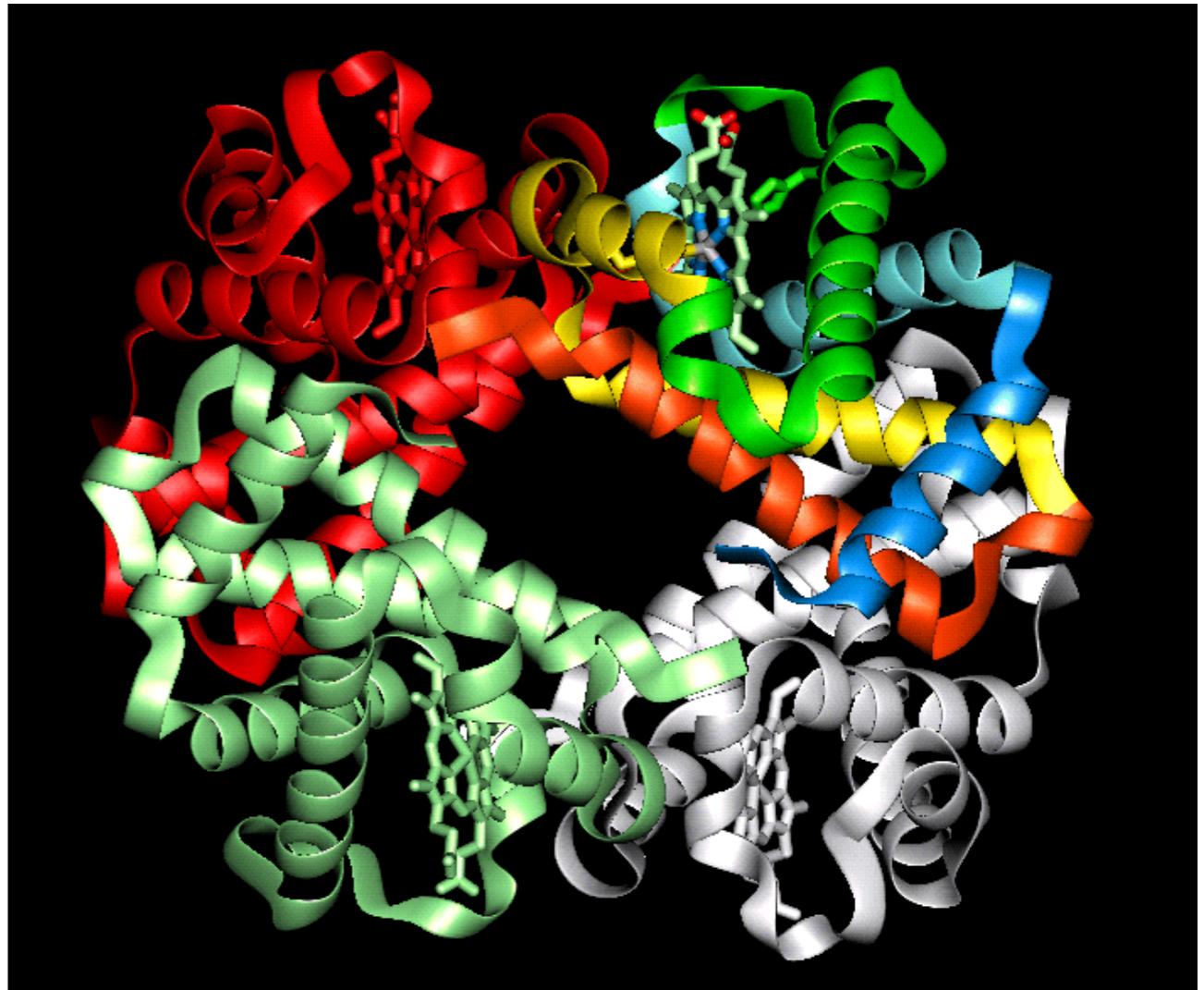


Primary structure: sequence of the protein

Secondary structure: local structuration of the protein.

Tertiary structure: 3D structure of the protein

Quaternary structure: structure of protein complexes



1.2. DNA

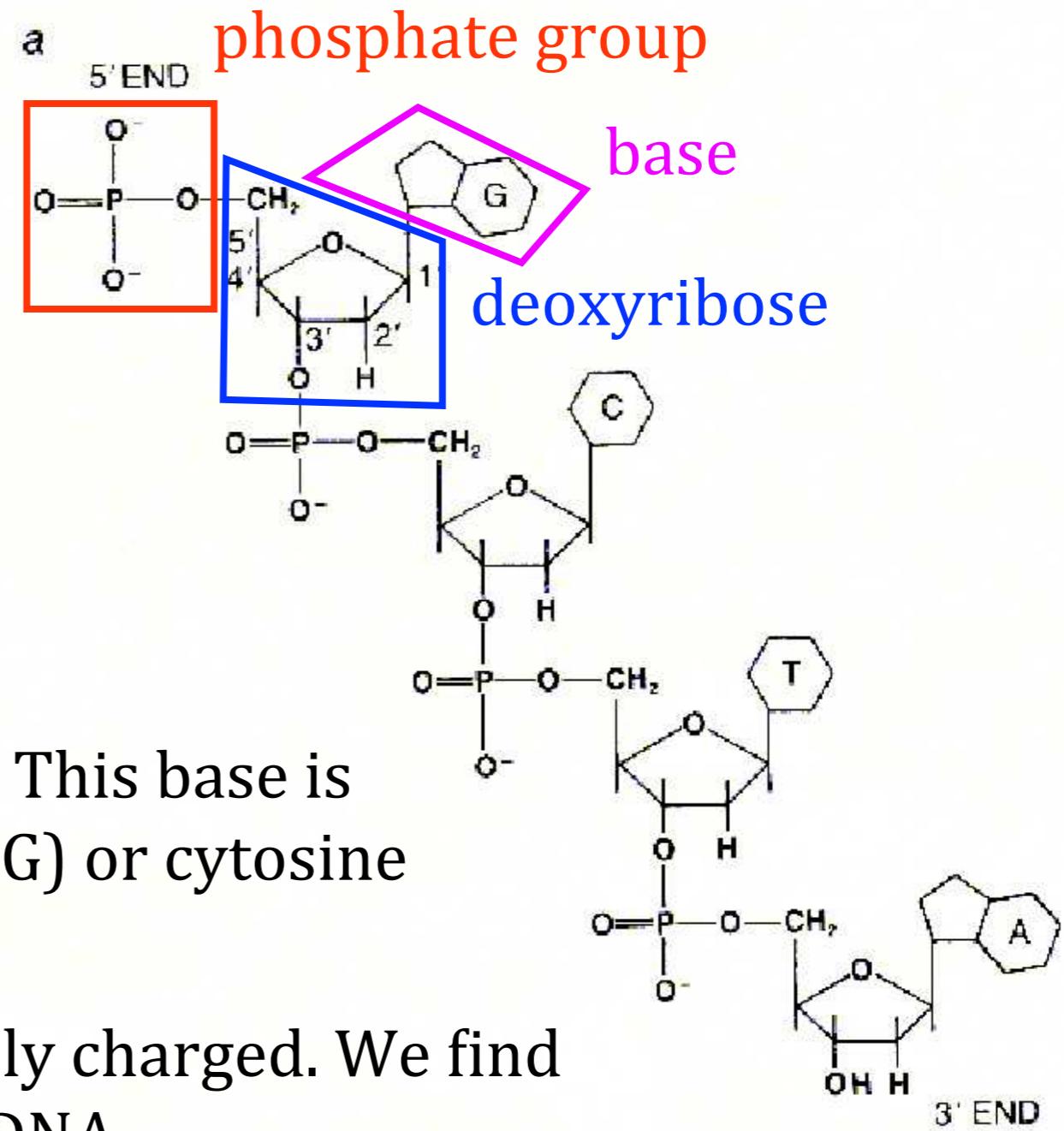
Deoxyribonucleic acid is composed by two strands, folded in a double helix (twisted ladder).

Each of the strand is composed by a main chain of a sugar (deoxyribose) and a phosphate group linked by a phosphodiester bond.

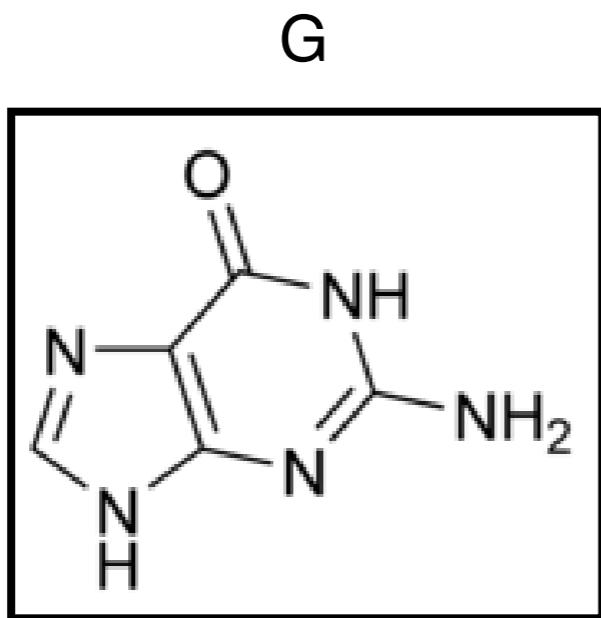
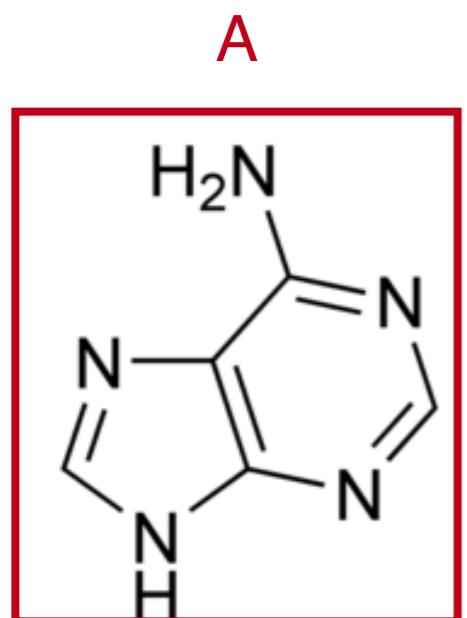
A base is linked to the deoxyribose. This base is adenine (A), thymine (T), guanine (G) or cytosine (C).

The phosphate groups are negatively charged. We find thus positive counter-ions around DNA.

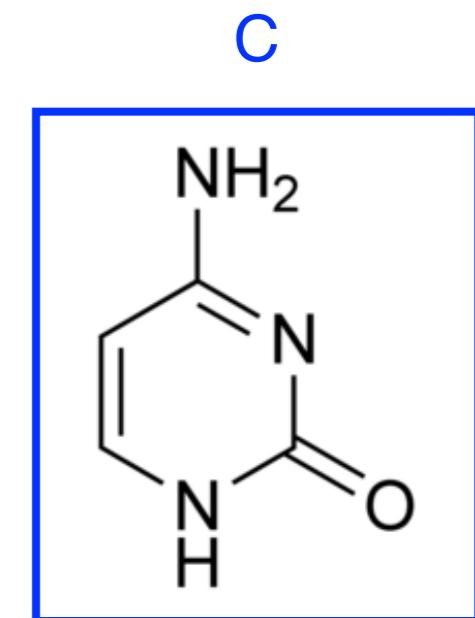
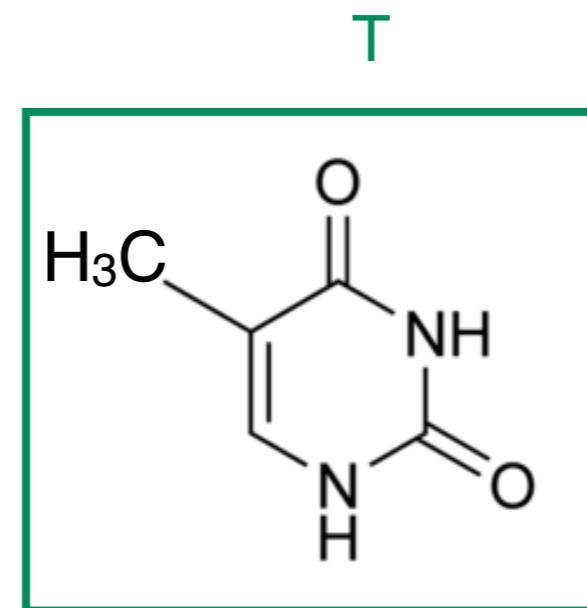
A DNA strand is oriented: the 5' part ends by a phosphate group and the 3' part by a sugar.



Purines

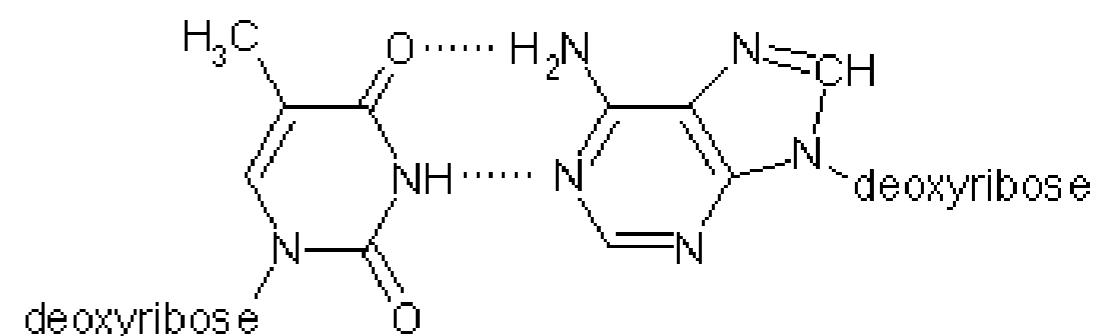


Pyrimidines

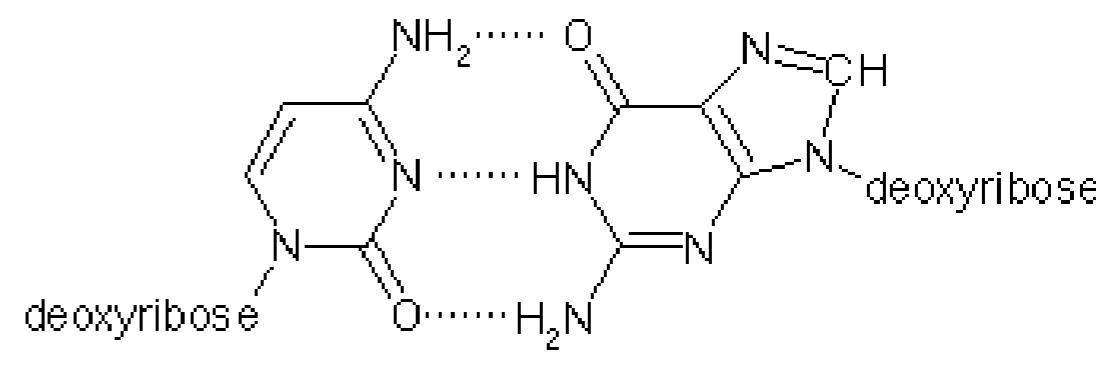


One of the associated strand is oriented 5'-3' and the second. For each adenine on one side of the chain, a thymine is opposite on the other strand; for each guanine on one side of the chain has a cytosine as its opposite.

Bases interact through hydrogen bonds: 3 for G-C pairs and 2 for A-T pairs.

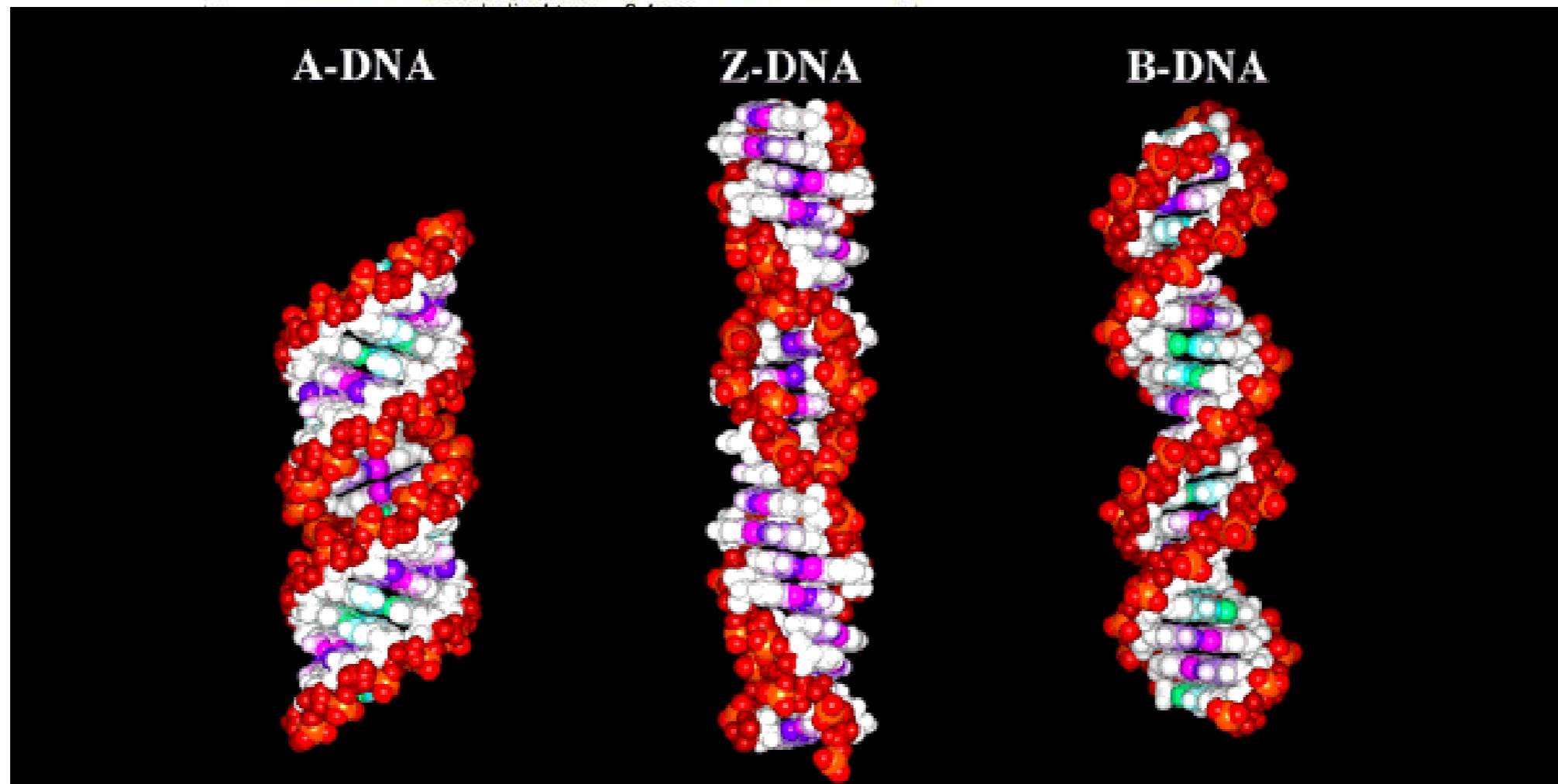
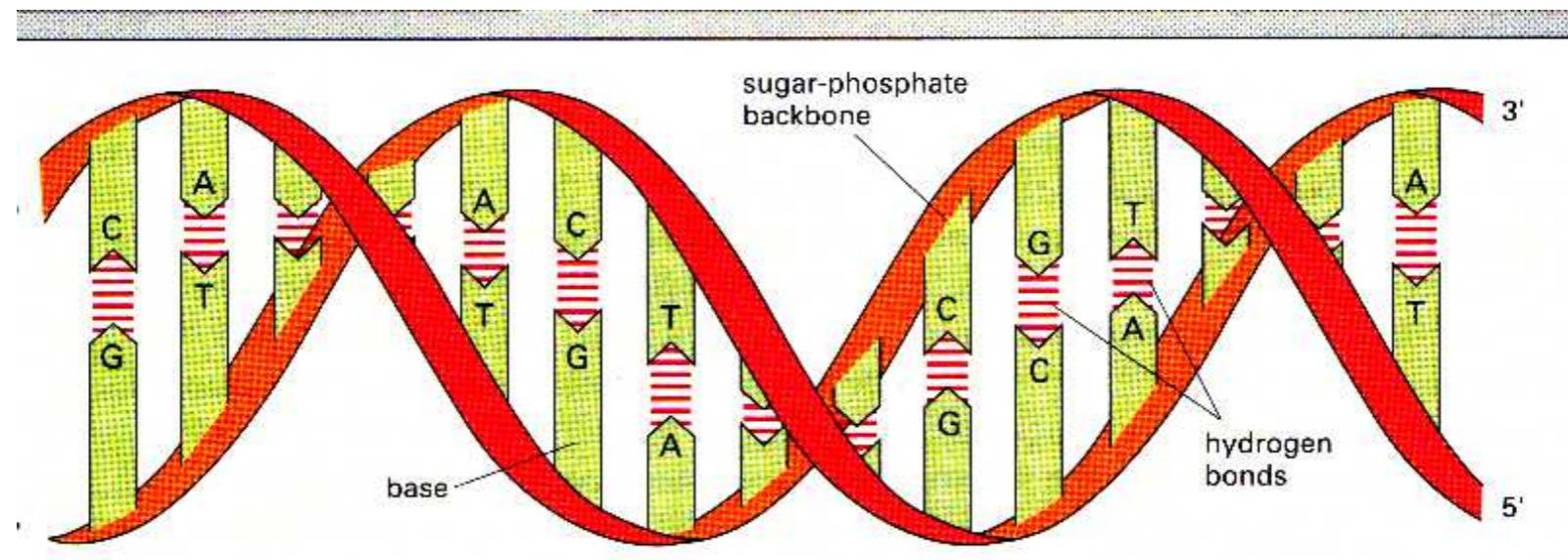


Thymine Adenine

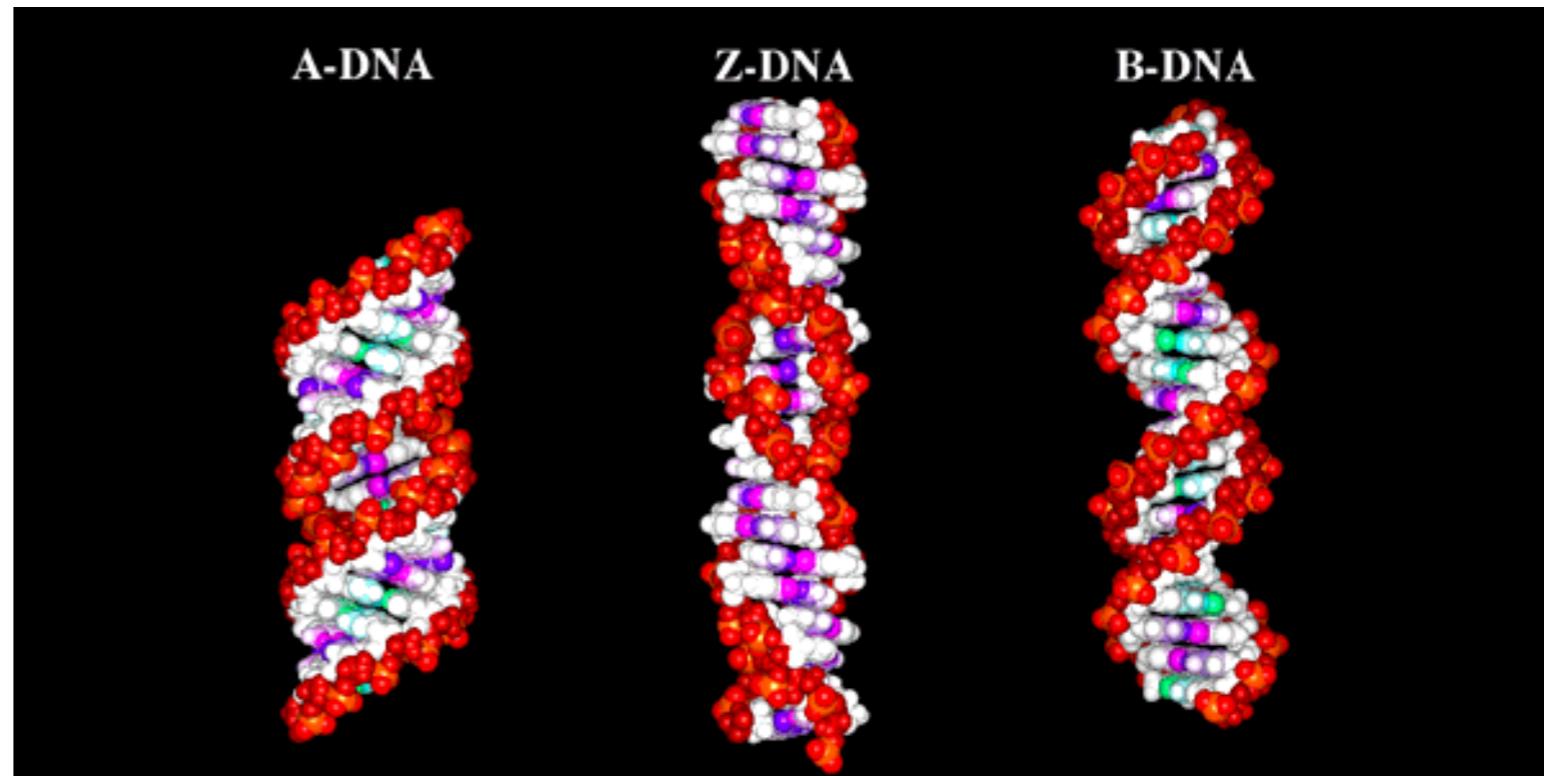


Cytosine Guanine

Structure of the double helix:



The most usual structure is B-DNA. But there exists also A, Z, ... structures.

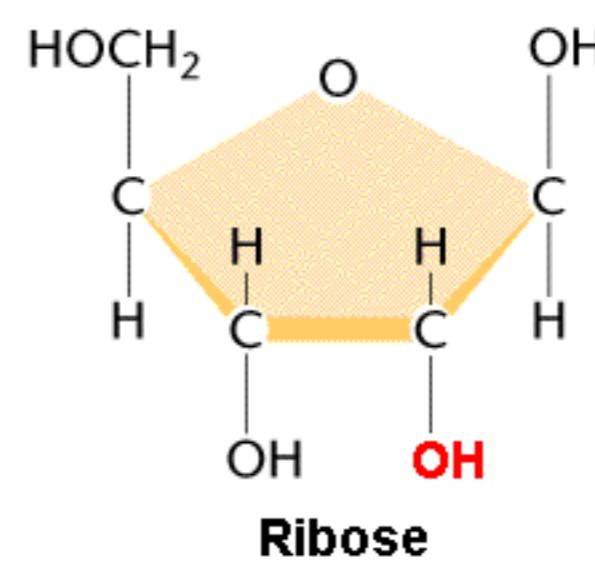
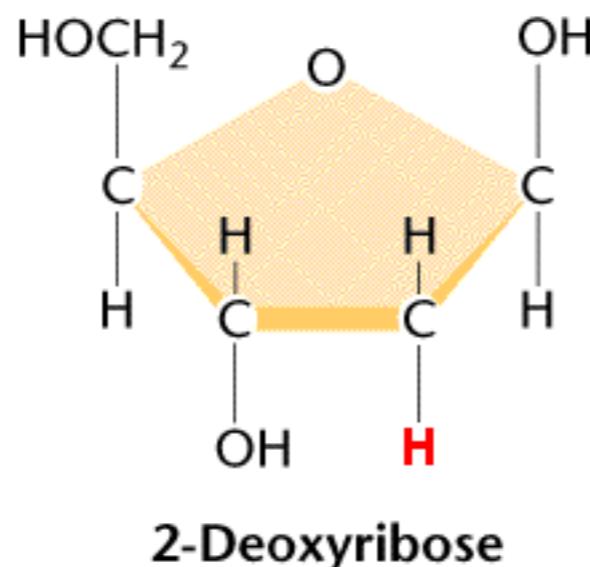


	A-DNA	B-DNA	Z-DNA
	Right-handed helix	Right-handed helix	Left-handed helix
Diameter of the helix	25,5 Å	23,7 Å	18,4 Å
Number of base pairs per turn	~11	~10	~12
Pitch per turn of helix	25 Å	34 Å	47 Å

1.3. RNA

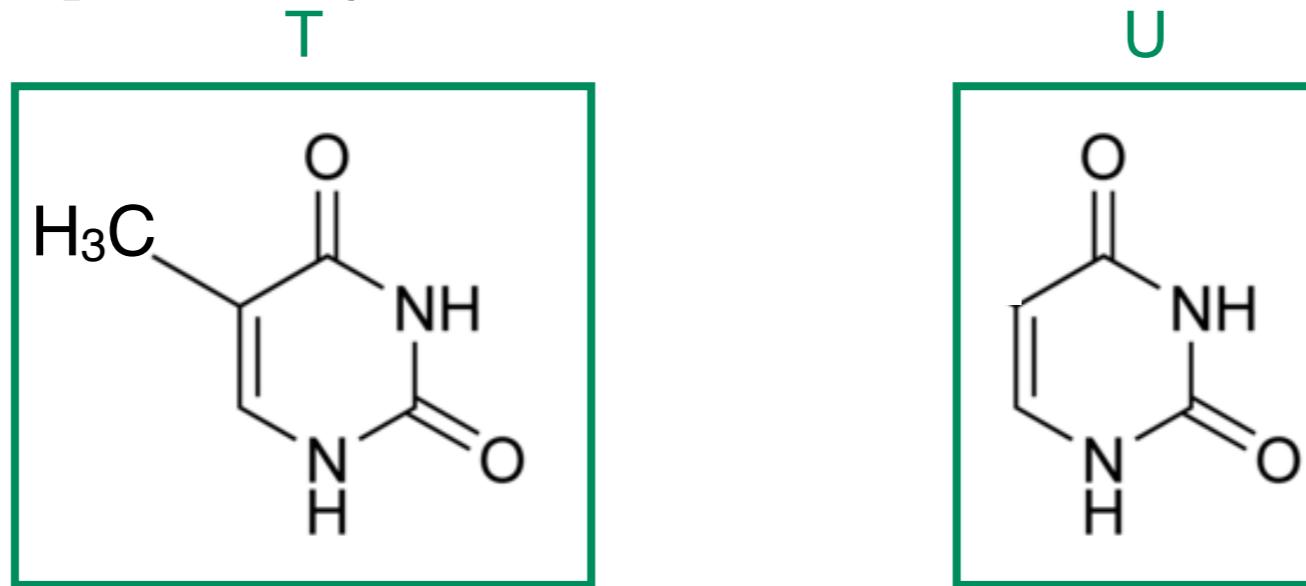
Ribonucleic acid (RNA) is very similar to DNA: one strand is also composed by phosphate groups and sugars, with a base attached to sugars. Differences with DNA are:

- RNA contains only one strand;
- the sugar is ribose:

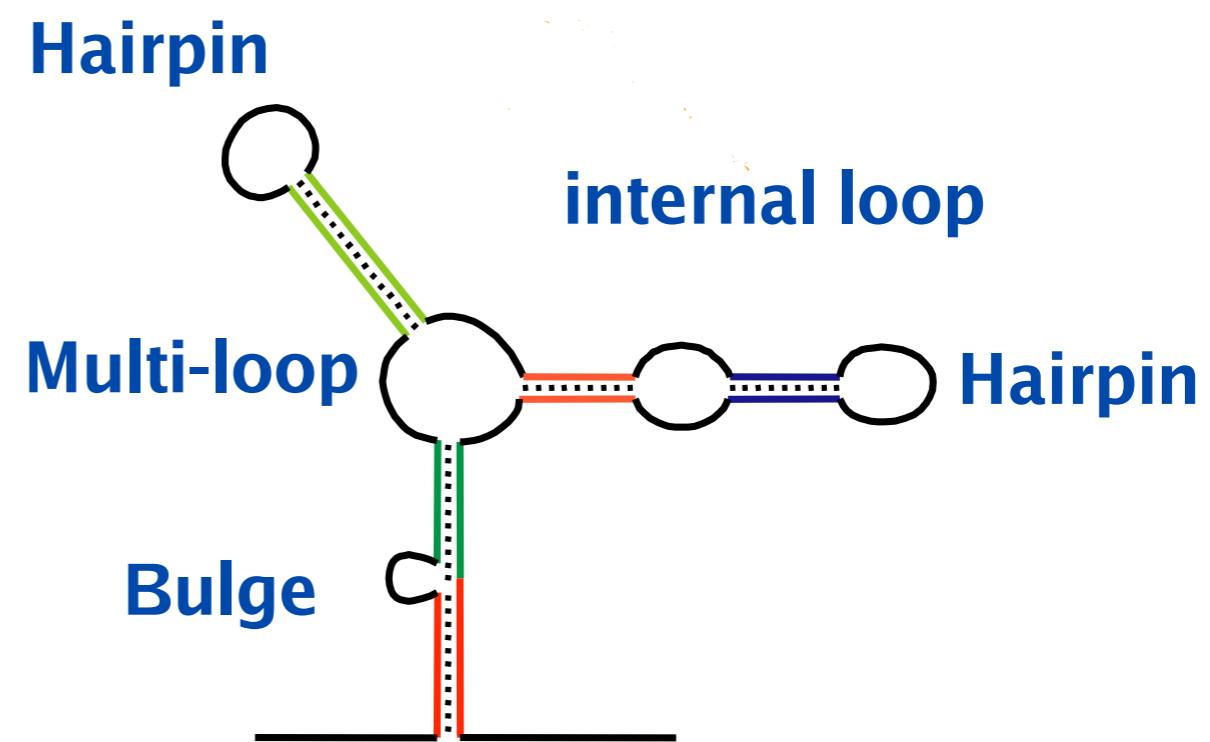
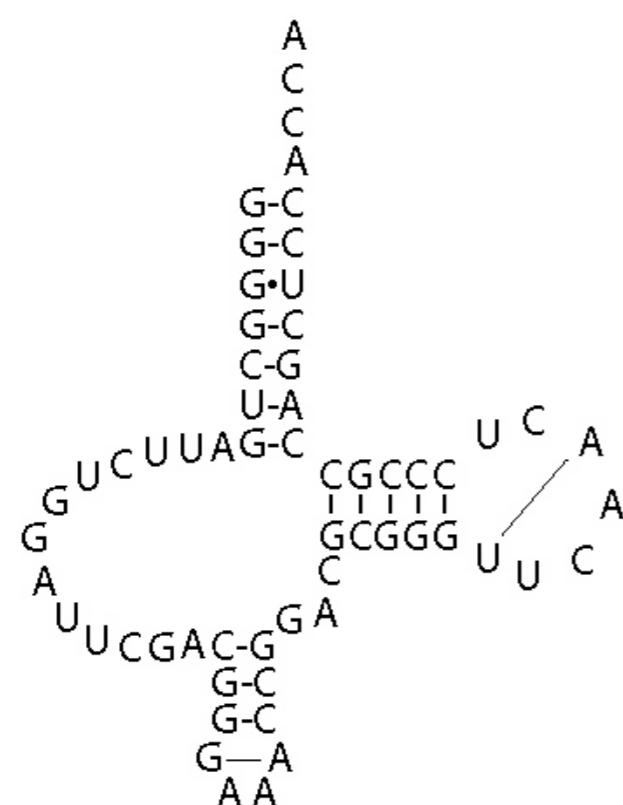


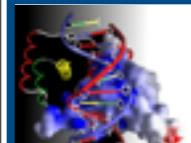
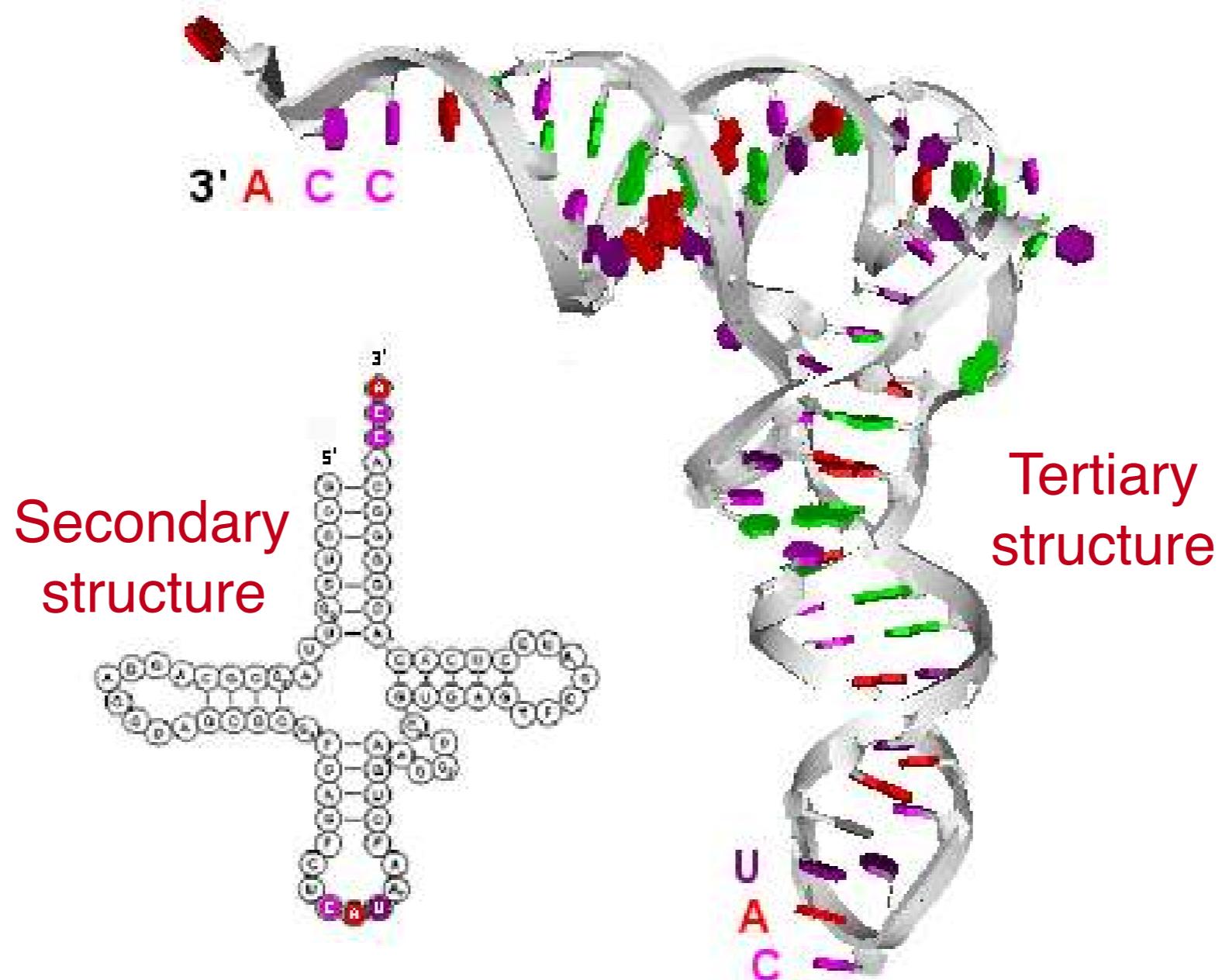
(Klug & Cummings 1997)

- thymine is replaced by uracile:

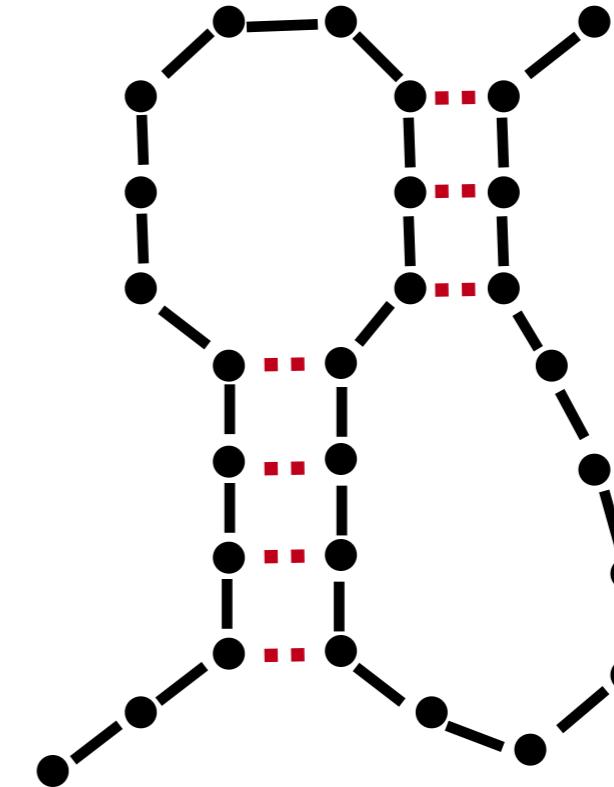
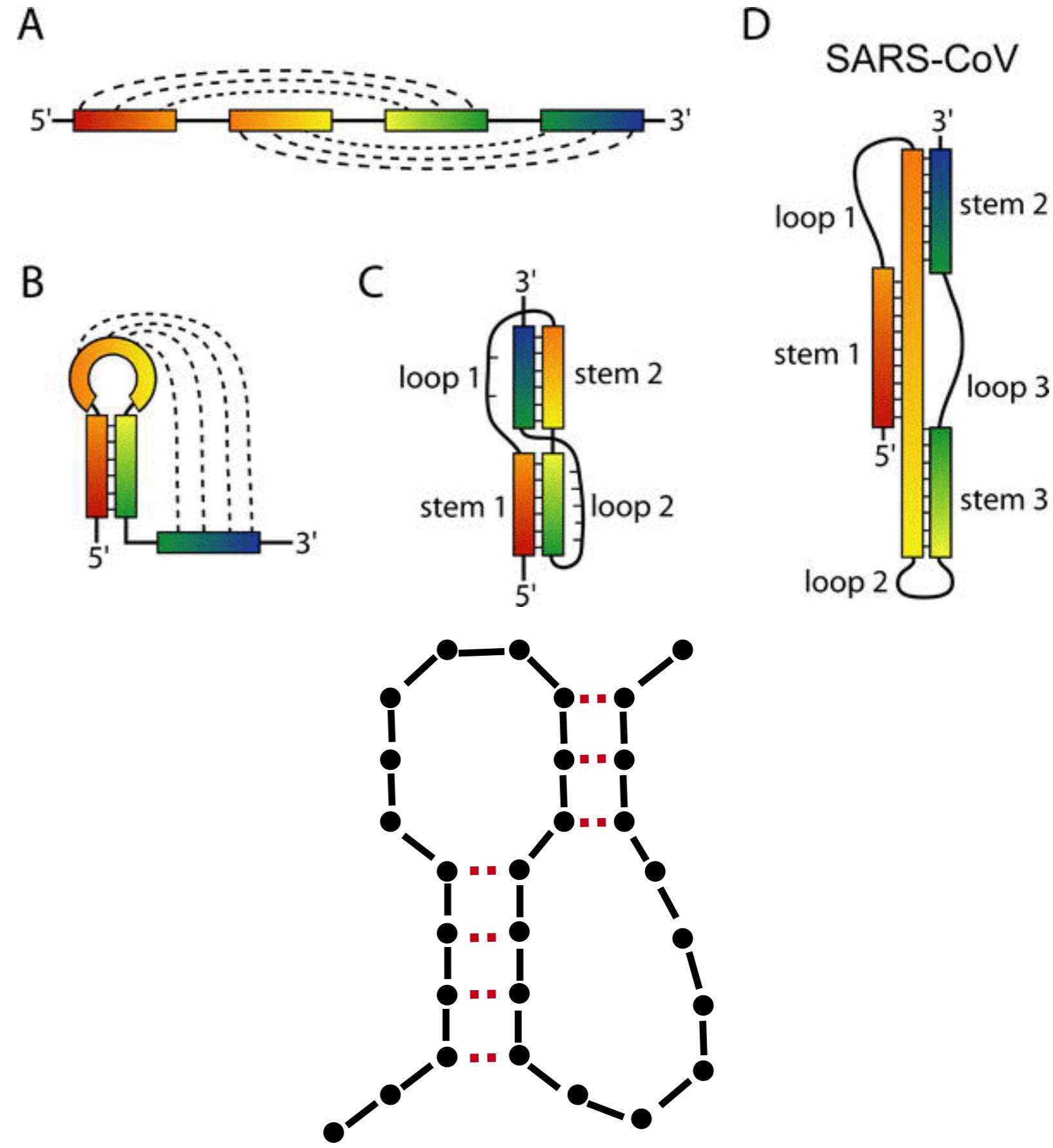
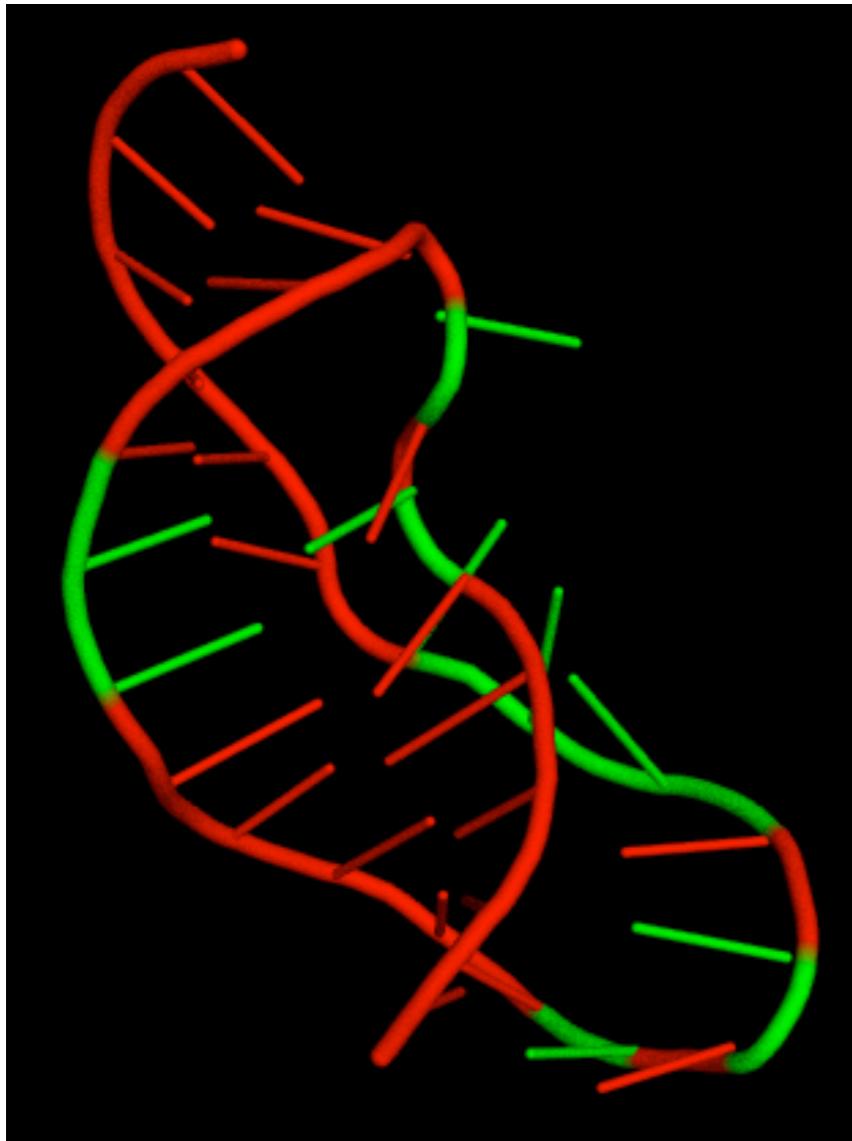


- RNA is not folded in helix as DNA, but can present a complex 3D structure by base pairing. Secondary and tertiary structures can be defined.





- pseudo-knots are also found in RNA:



2. Experimental approaches to resolve the structure of biomolecules

Two classes of methods are used to study/resolve experimentally the structure of biomolecules: those that provide information about the local structure (secondary structure of proteins), and those that resolve the 3D structure.

2.1. Resolution of the 3D structure

Two approaches are mainly used: X-ray crystallography (RX) and nuclear magnetic resonance (NMR).

The structures of proteins, DNA,... that have been experimentally resolved are stored in an online public database:

<http://www.pdb.org>

Experimentally resolved structures (September 2013, PDB)

Exp. Method	Proteins	Nucleic acids	Protein/NA complexes	Other	Total
X-Ray	77139	1481	4059	3	82682
NMR	8829	1044	193	7	10073
Electron microscopy	466	45	128	0	639
Hybrid	51	3	2	1	57
Other	150	4	6	13	173
Total	86635	2577	4388	24	93624

<http://www.pdb.org/pdb/statistics/holdings.do>

It is possible to download from the website a file (PDB file) containing the cartesian coordinates of each atom of a given biomolecule.

The screenshot shows the RCSB PDB website for structure 1J5N. The main page includes the PDB logo, a banner for PDB-101, and a statistics box. The central content area displays the title "Solution Structure of the Non-Sequence-Specific HMGB protein NHP6A in complex with SRY DNA". Below the title, the DOI and NDB ID are listed as 10.2210/pdb1j5n/pdb and 1J5N respectively. The "Primary Citation" section provides details from the journal article. To the right, there is a 3D ribbon model of the protein-DNA complex. A red arrow points to the "Download Files" menu, which offers various file formats including PDB File (Text), PDB File (gz), mmCIF File, and NMR Restraints (Text).

Structure of the PDB file:

header of the file: information about the biomolecule, experimental conditions, ...

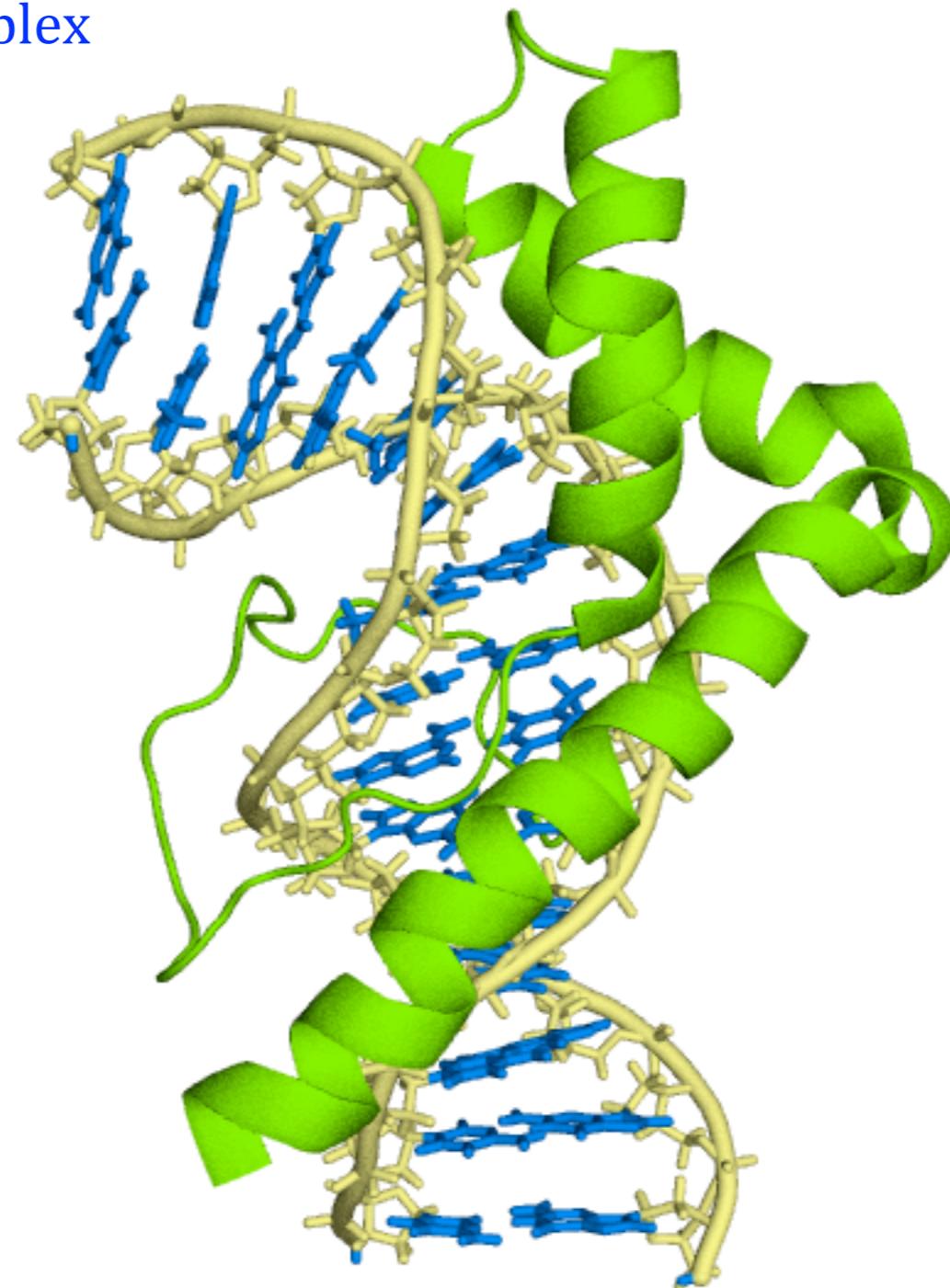
fields beginning by 'ATOM' or 'HETATM': cartesian coordinates of the atoms

Example: complex protein-DNA

atom number	atom	amino acid	chain name	amino acid number	x, y, z	
ATOM	1	N	MET	A	-11.818 -3.468 -20.506	1.00 0.00 N
ATOM	2	CA	MET	A	-10.466 -2.956 -20.851	1.00 0.00 C
ATOM	3	C	MET	A	-10.275 -1.535 -20.332	1.00 0.00 C
ATOM	4	O	MET	A	-9.572 -0.731 -20.943	1.00 0.00 O
ATOM	5	CB	MET	A	-10.260 -2.987 -22.367	1.00 0.00 C
ATOM	6	CG	MET	A	-10.620 -4.320 -23.002	1.00 0.00 C
ATOM	7	SD	MET	A	-9.806 -5.714 -22.201	1.00 0.00 S
ATOM	8	CE	MET	A	-8.919 -6.429 -23.584	1.00 0.00 C
ATOM	20	N	VAL	A	-10.906 -1.234 -19.202	1.00 0.00 N
ATOM	21	CA	VAL	A	-10.805 0.091 -18.601	1.00 0.00 C
ATOM	22	C	VAL	A	-9.427 0.310 -17.984	1.00 0.00 C
ATOM	23	O	VAL	A	-9.019 1.445 -17.740	1.00 0.00 O
ATOM	24	CB	VAL	A	-11.878 0.301 -17.518	1.00 0.00 C
ATOM	25	CG1	VAL	A	-13.210 0.674 -18.151	1.00 0.00 C
ATOM	26	CG2	VAL	A	-12.020 -0.944 -16.656	1.00 0.00 C

These PDB files can be opened with several visualization programs, allowing a detailed analysis of the structure:

Example: protein-DNA complex



2.1.1. X-ray crystallography

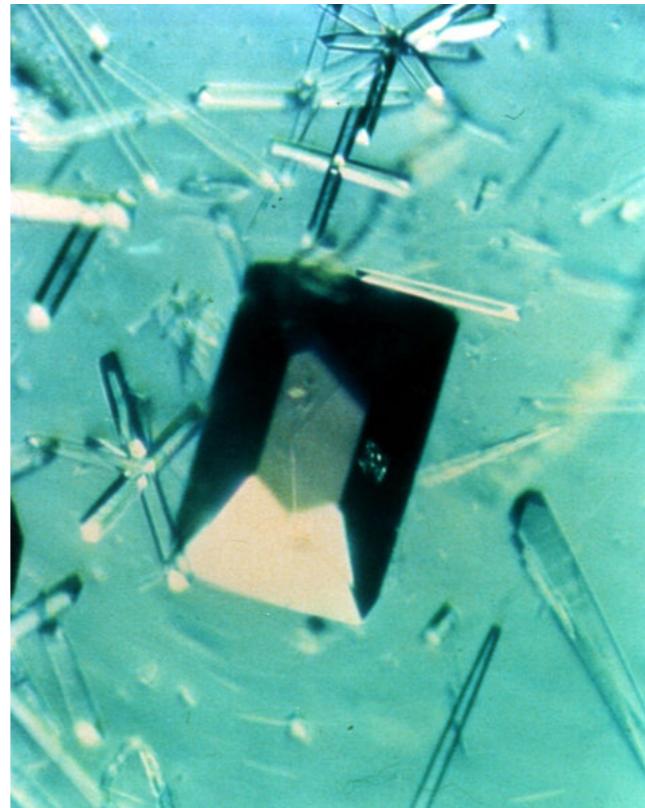
1958: Perutz & Kendrew resolve the first protein structure by X-ray crystallography (myoglobin); they received the Noble price in 1962.

The observation of an object requires a source of light compatible with the resolution that is needed. To distinguish atoms from each other, the radiation must have a wavelength comparable to the distance between atoms (about 1,5 Å). The radiations corresponding to this wavelength are electrons, neutrons or X photons.

In the case of X-rays, working with only one molecule will lead to a too small diffusion to get a sufficient signal. It is thus necessary to have a diffusion from a large number of molecules, so as to add the signals. To have a coherent diffusion, the molecules must be ordered.

=> necessary to work with a crystal, with a periodic arrangement of the molecules.

First: to obtain a crystal large enough and that will lead to a good diffraction.



It is necessary to have a pure and regular crystal.
A crystal is a periodic assembly of unit cells.

Obtaining a crystal can be very complex. It is very difficult to obtain crystals for membrane proteins.

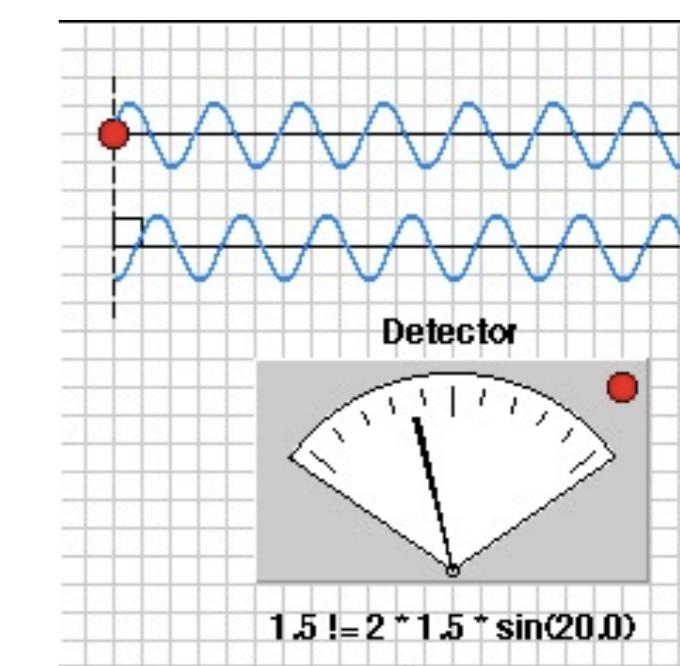
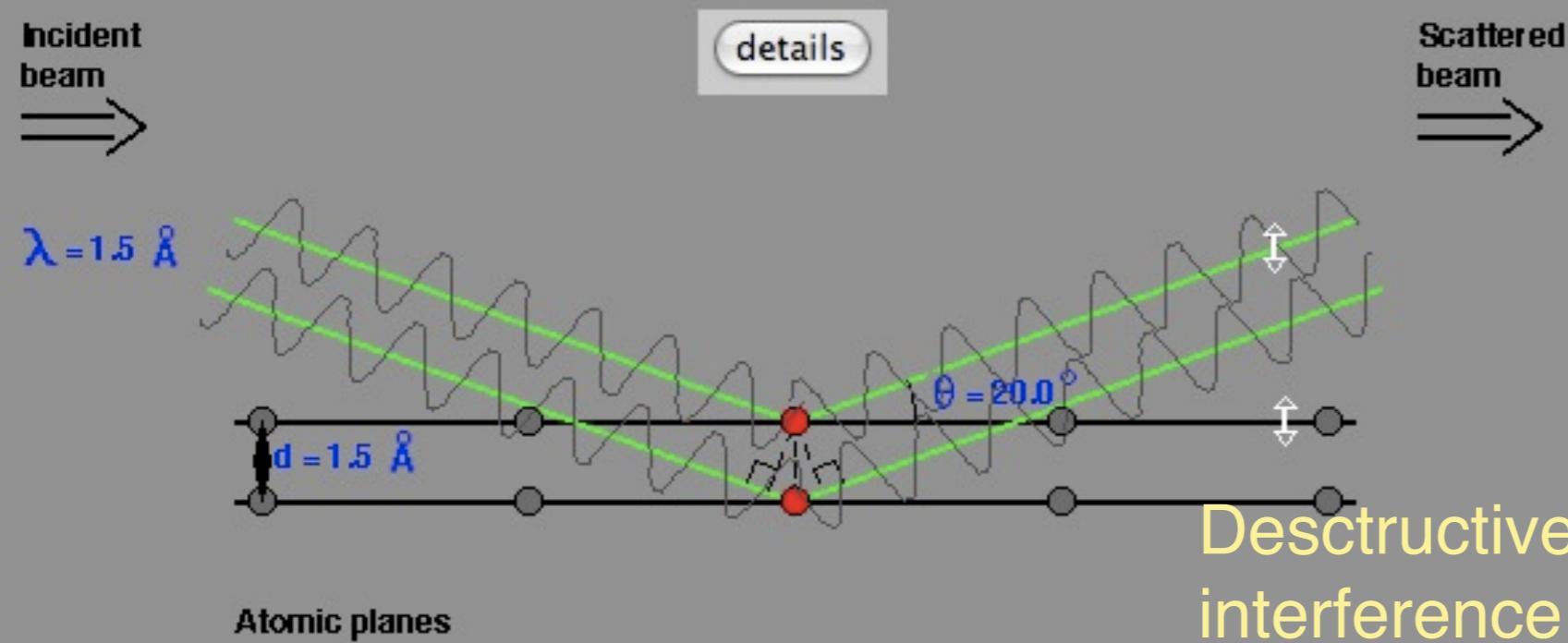
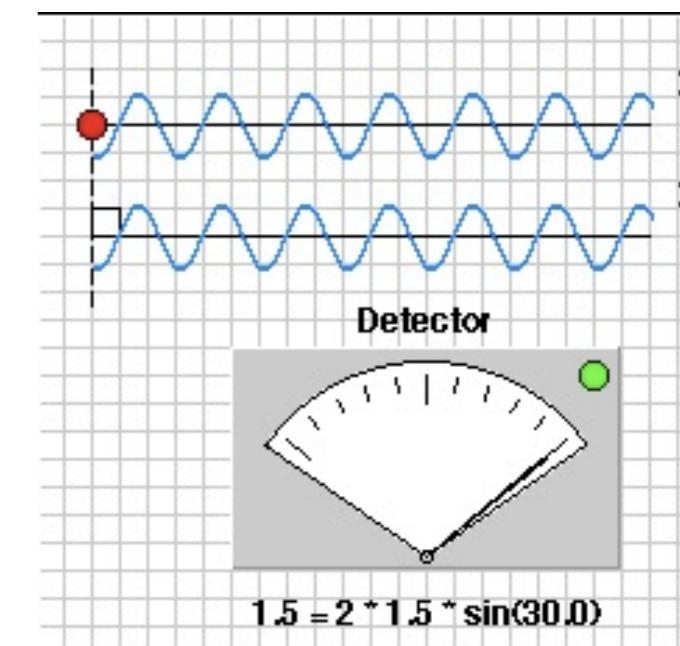
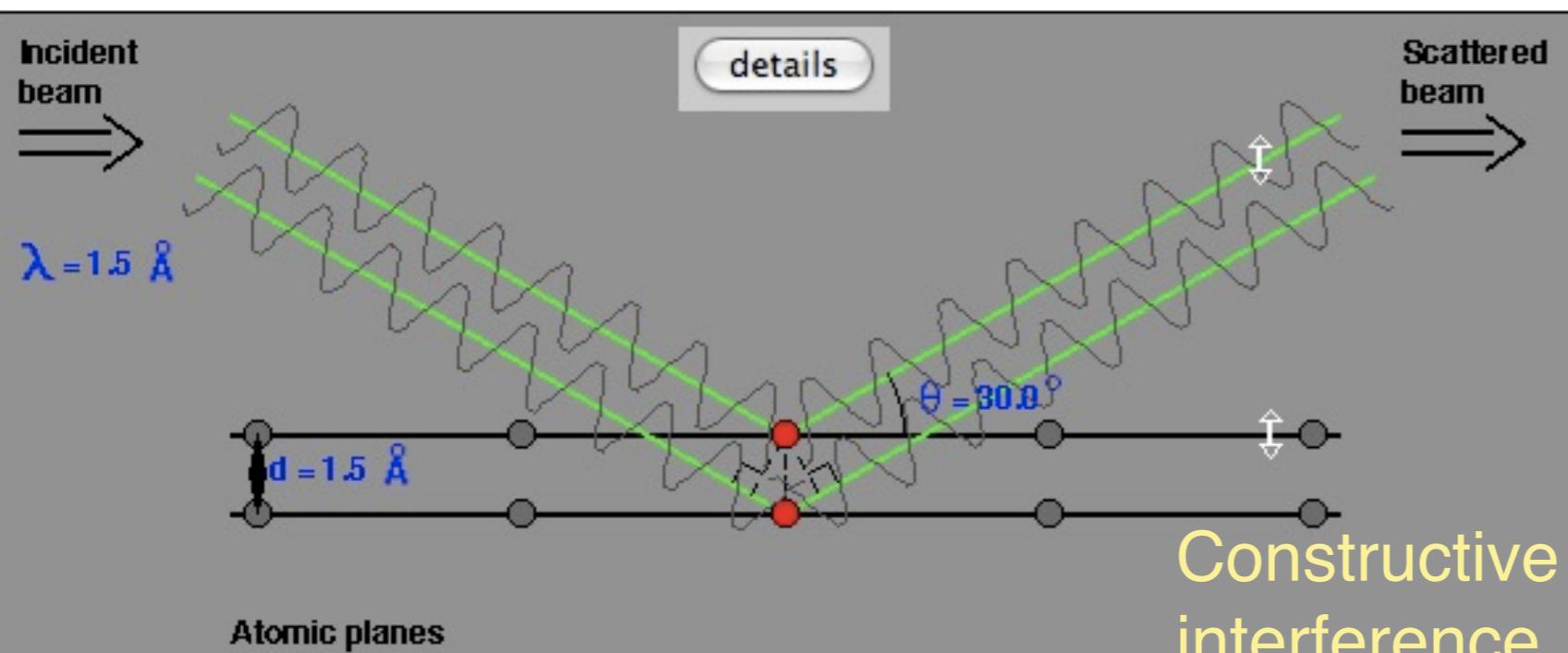
The molecules are in contact in the crystal only in some regions, and in these regions the molecules interact indirectly through solvent layers.

Is the structure of the protein the same in a crystal and in solution ?? Yes:

- the molecule in the crystal is also in an aqueous environment;
- crystals of enzymes are biologically active (and this activity depends on the structure);
- the structure of several proteins has been resolved both by NMR (in solution) and RX, and they are identical;

No lens in X-ray optics. It is thus necessary to study the diffraction pattern of these X-rays, which is a result of interference of the waves diffused by each atom.

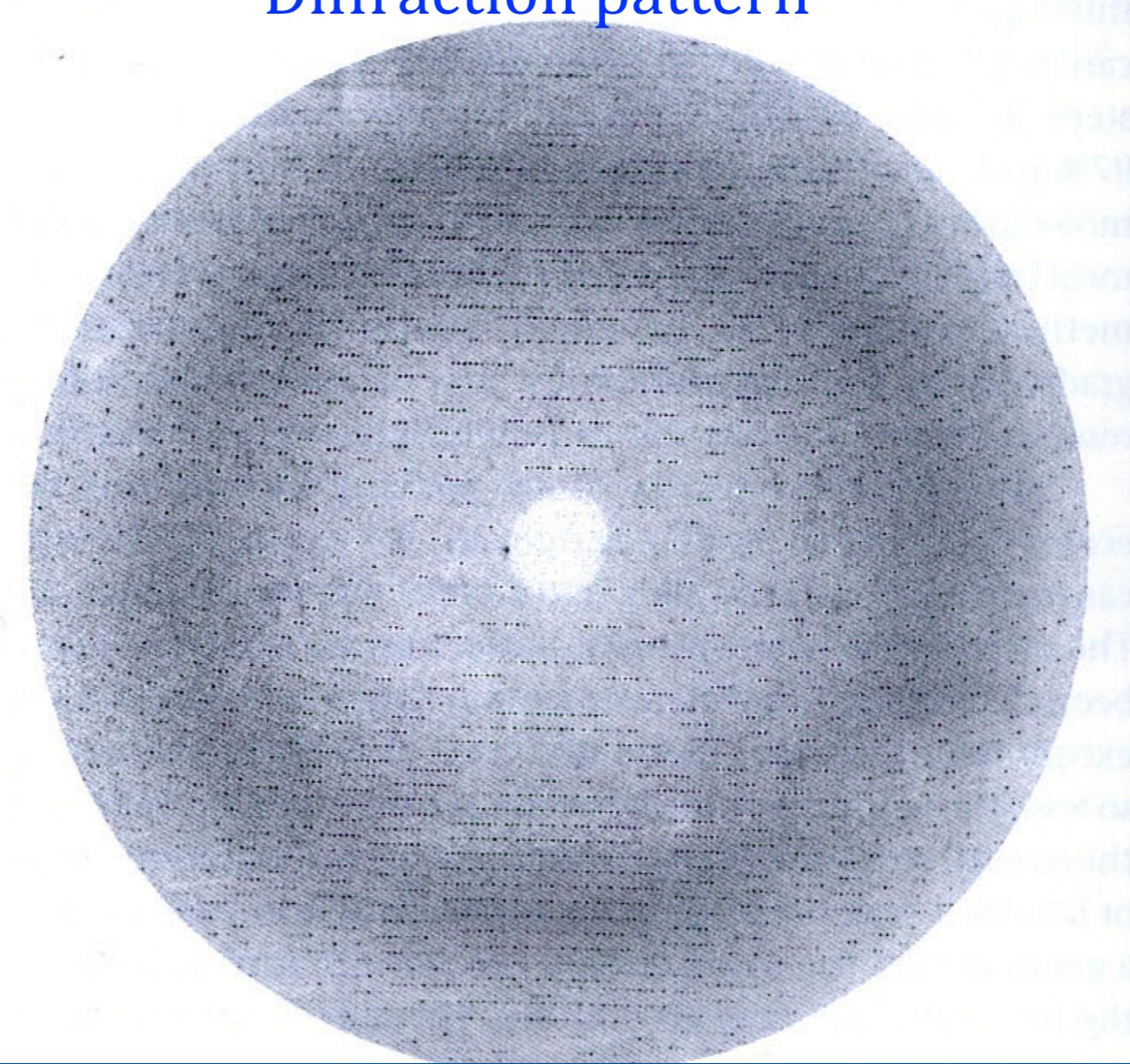
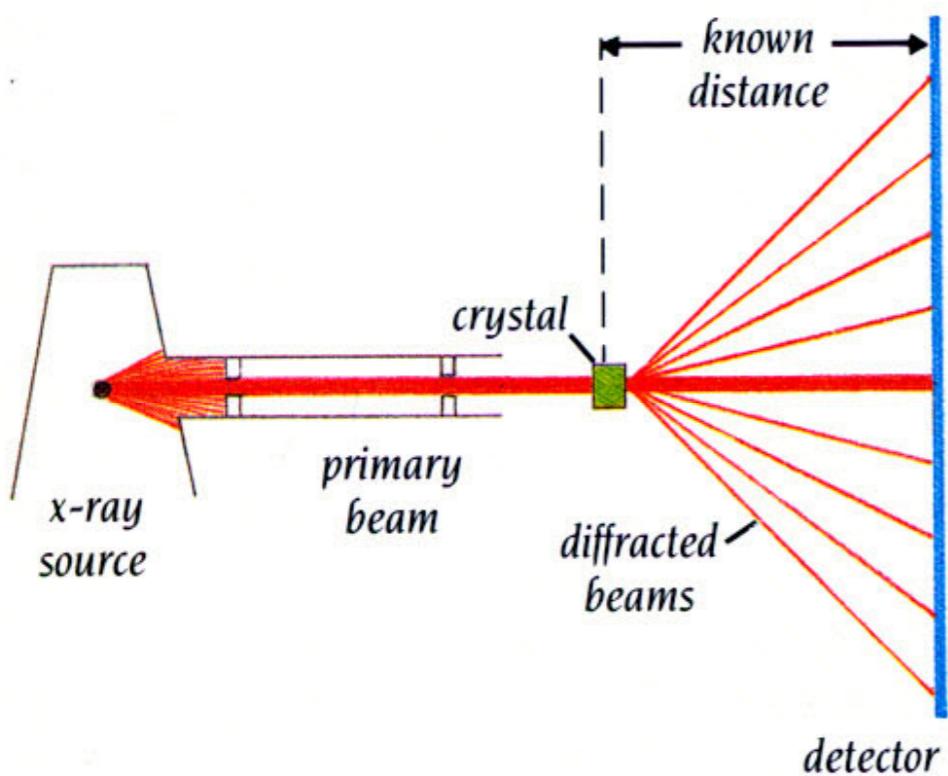
Constructive interference: when the rays have the same path length or when this path length is a multiple of the wavelength.



Necessary to have a good source of X-rays (rotating anode , synchrotron,...). A beam of monochromatic X-rays irradiate the crystal.

We will obtain a diffraction pattern (spots). The relative intensity of these spots will give an information that will be used to identify the molecular arrangement in the crystal. To reconstruct the whole crystal, the crystal is rotated 180°.

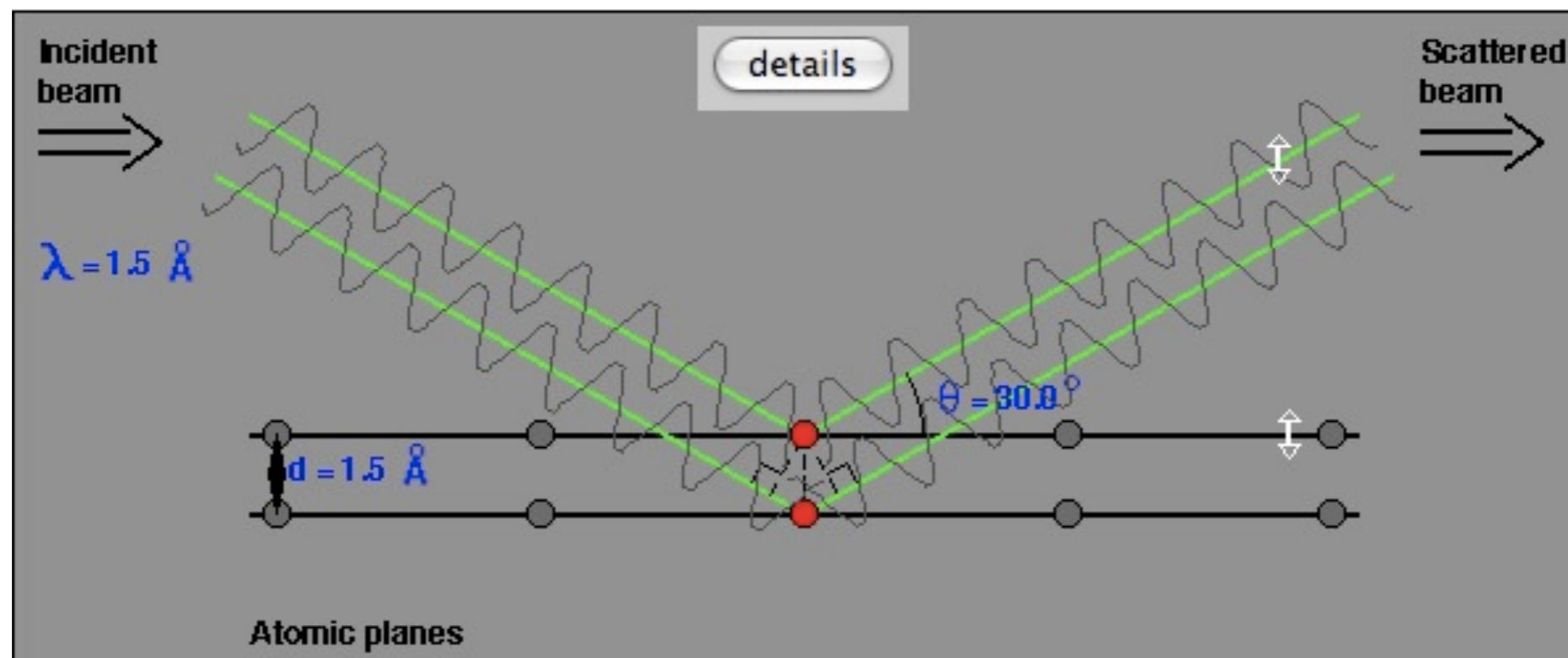
Diffraction pattern



The Bragg law provides the condition to have a constructive interference:

$$n \lambda = 2 d \sin \Theta$$
$$\Leftrightarrow d = \frac{n \lambda}{2 \sin \Theta}$$

where λ is the wavelength, Θ is the incident angle, n is any integer and d is the spacing between the diffracting plans.



$$d = \frac{n \lambda}{2 \sin \Theta}$$

There is an inverse proportionality between the diffracting plane spacing and the incident angle to have a constructive interference.

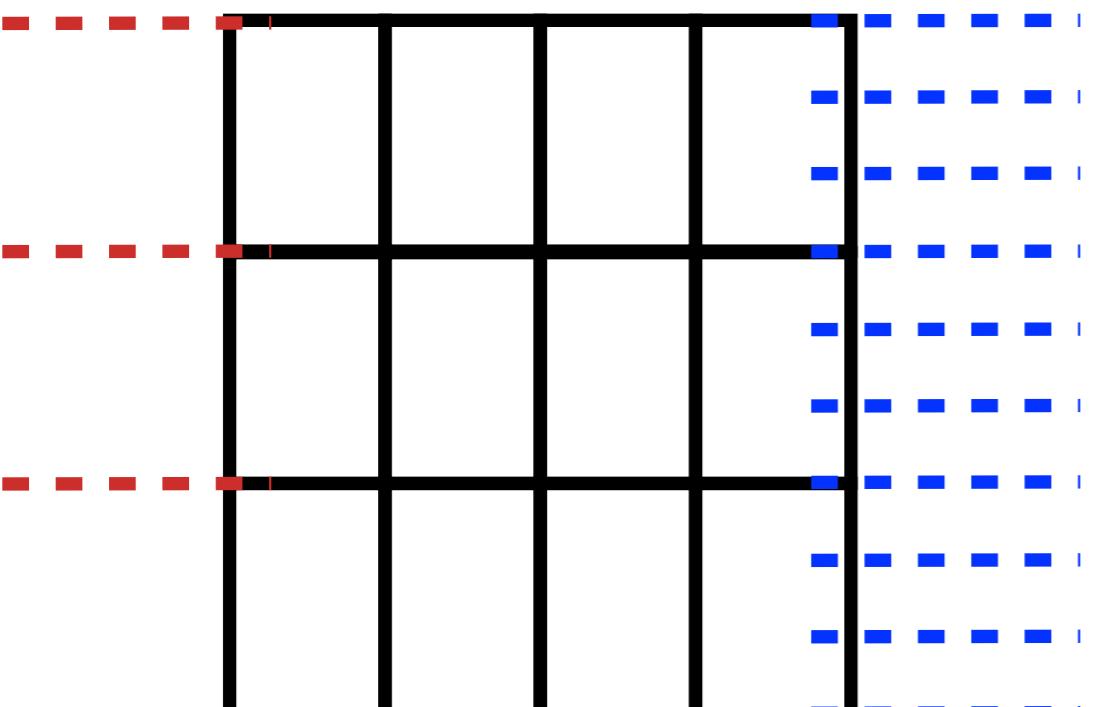
When Θ is equal to 0° : the distance is infinite, no diffusion, no information about the spatial arrangement of the atoms.

When Θ is equal to 90° : the wave is sent back to the source. The information is limited to a spacing between the planes that is equal to the half of the wavelength.

The diffraction spots (constructive interference between the diffused waves) will give information about the position of the electrons, and thus the atoms. The electronic density can be reconstructed from the diffraction pattern.

The cristal is obtained from the repetition of the unit cell (smallest group of molecules that is characteristic of a crystal lattice; the crystal can be obtained from the translation of the unit cell): 

- - - : separation between the planes of one unit cell along the vertical axe
- - - : separation between the planes of the third of the unit cell, along the vertical axe

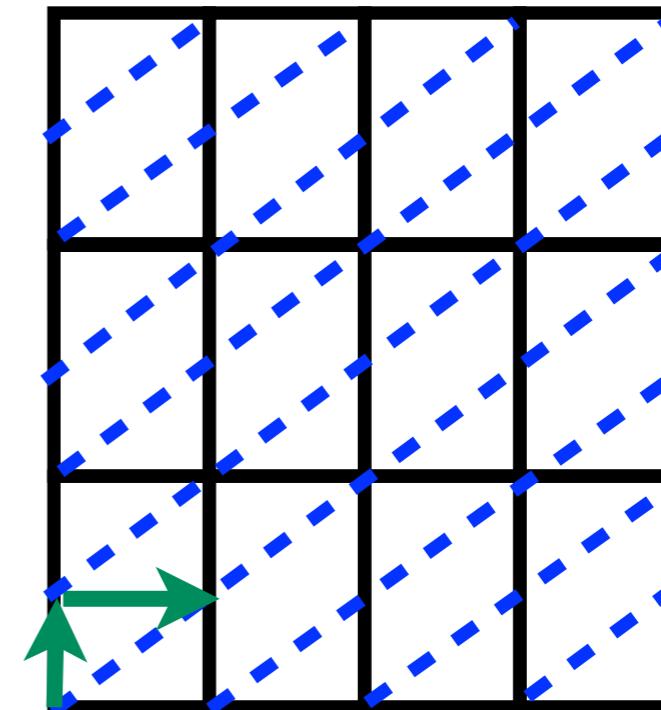
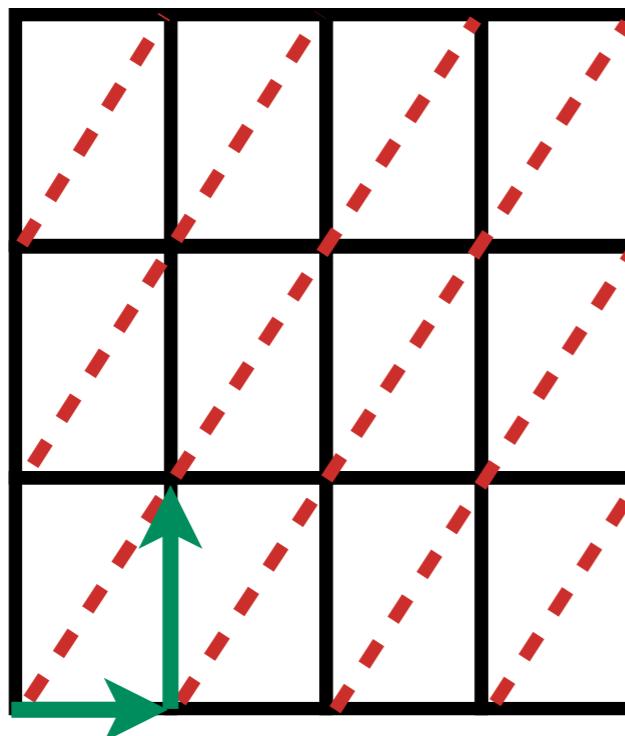


The axes are defined in the 3 directions of the space.

The planes separated by a red line are the planes $(1\ 0\ 0)$ because they divide une time the unit cell.

The planes separated by a blue line are the planes $(3\ 0\ 0)$ because they divide the unit cell in three.

- : planes (1 1 0), because the move correspond to one unit along x and 1 unit along y.
- : planes (2 1 0), because the move corresponds to an half unit along x et one unit along y.



These planes are referenced by the Miller indices ($h k l$).

Structure factor: $F(\mathbf{s})$, where \mathbf{s} is the diffraction vector.

$F(\mathbf{s})$ corresponds to the wave that results from the diffraction.

Structure factor for one electron: $F(\mathbf{s})=1e \exp(2\pi i \mathbf{s} \cdot \mathbf{r})$, where \mathbf{r} is the position vector of the electron and $1e$ is the amplitude. $F(\mathbf{s})$ is generally used as an adimensional number.

Structure factor of an object composed of a large number of electrons:

$$F(\mathbf{s})=\sum_j 1e_j \exp(2\pi i \mathbf{s} \cdot \mathbf{r}_j)$$

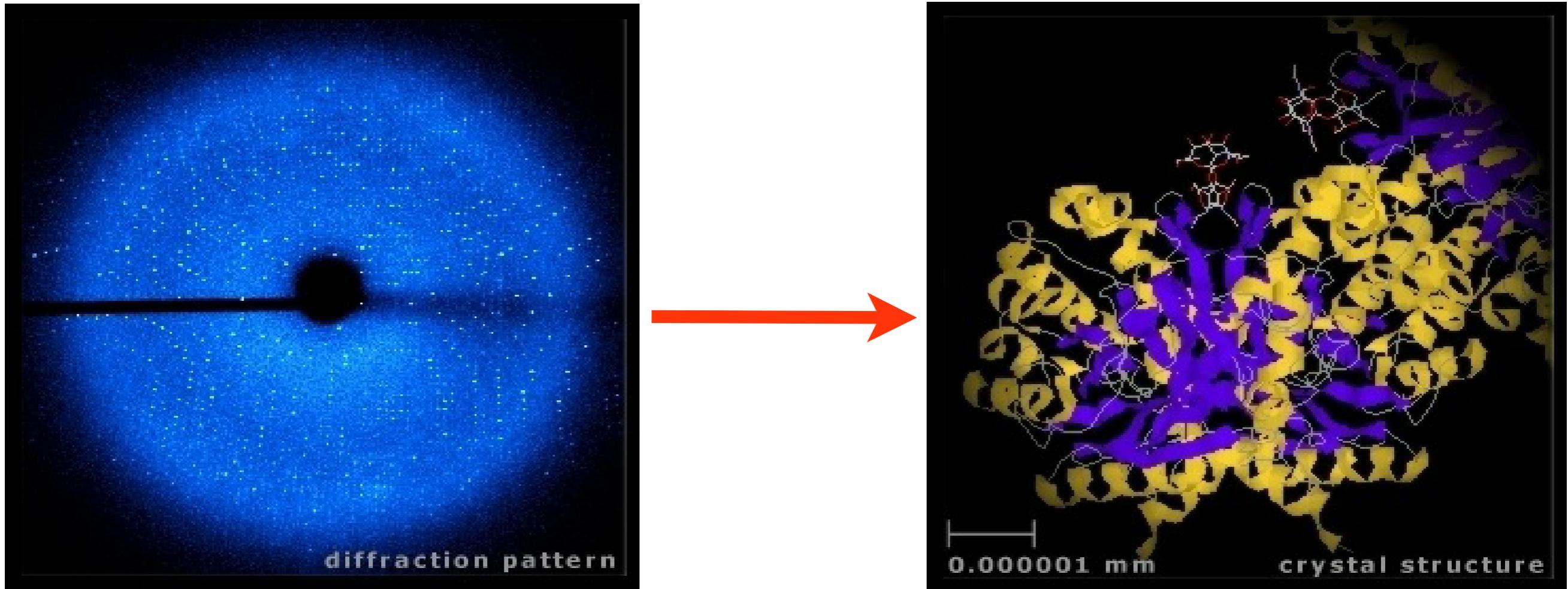
The diffraction at each point is the sum of the diffracted waves by each electron. The diffraction motif gives an information about the relative position of each electron.

Structure factor of a continuous distribution of electrons:

$$F(\mathbf{s})=\int_{\text{space}} \rho(\mathbf{r}) \exp(2\pi i \mathbf{s} \cdot \mathbf{r}) d\mathbf{r}; \mathbf{s} \text{ can be expressed as a function of the indices } h k l.$$

where $\rho(\mathbf{r})$ is the electronic density. The mathematical relationship between the electronic density and the structure factor is a Fourier transform.

The diffraction pattern corresponds to the Fourier transform of the electronic density. The goal is to reconstruct the electronic density from the diffraction pattern and to finally obtain a model of the structure:



The intensity of the spots, $I(hkl)$, can be measured on the diffraction figure. The amplitude of the structure factor is proportional to $|F_{hkl}|$.

The electronic density, $\rho(xyz)$, can be obtain from an inverse Fourier transform of the signal:

$$\rho(xyz) = 1/V \sum_h \sum_k \sum_l F(hkl) \exp[-2\pi i (hx+ky+lz)]$$

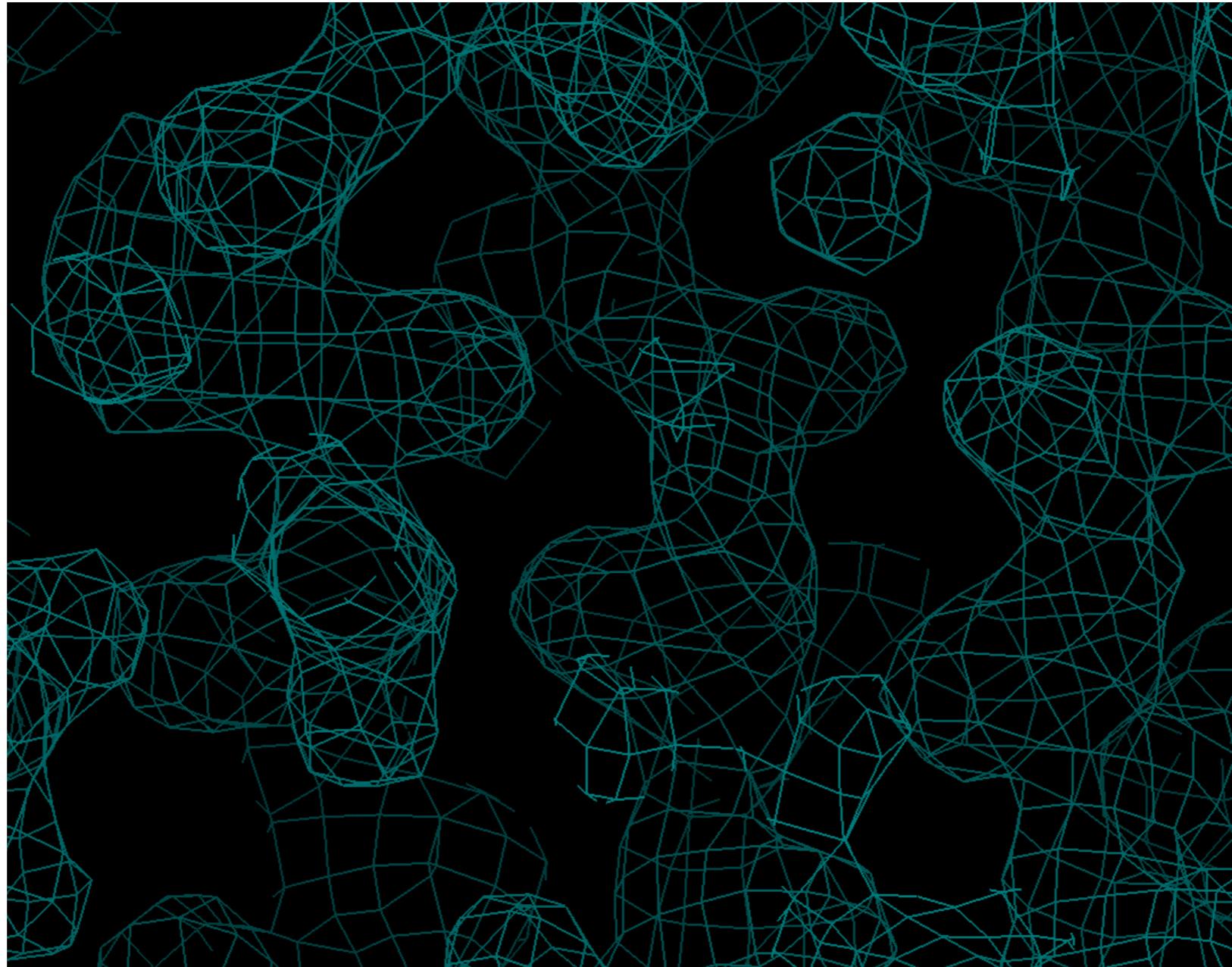
V is the volume of the unit cell. F is a complex number, and only its module can be obtained (see previous slide). Only the amplitude is known and not the phase.

Remind: a complex number $a + b \cdot i = r \cdot (\cos \Theta + i \cdot \sin \Theta) = r \cdot \exp[i \Theta]$, where r is the amplitude (module) and Θ the phase.

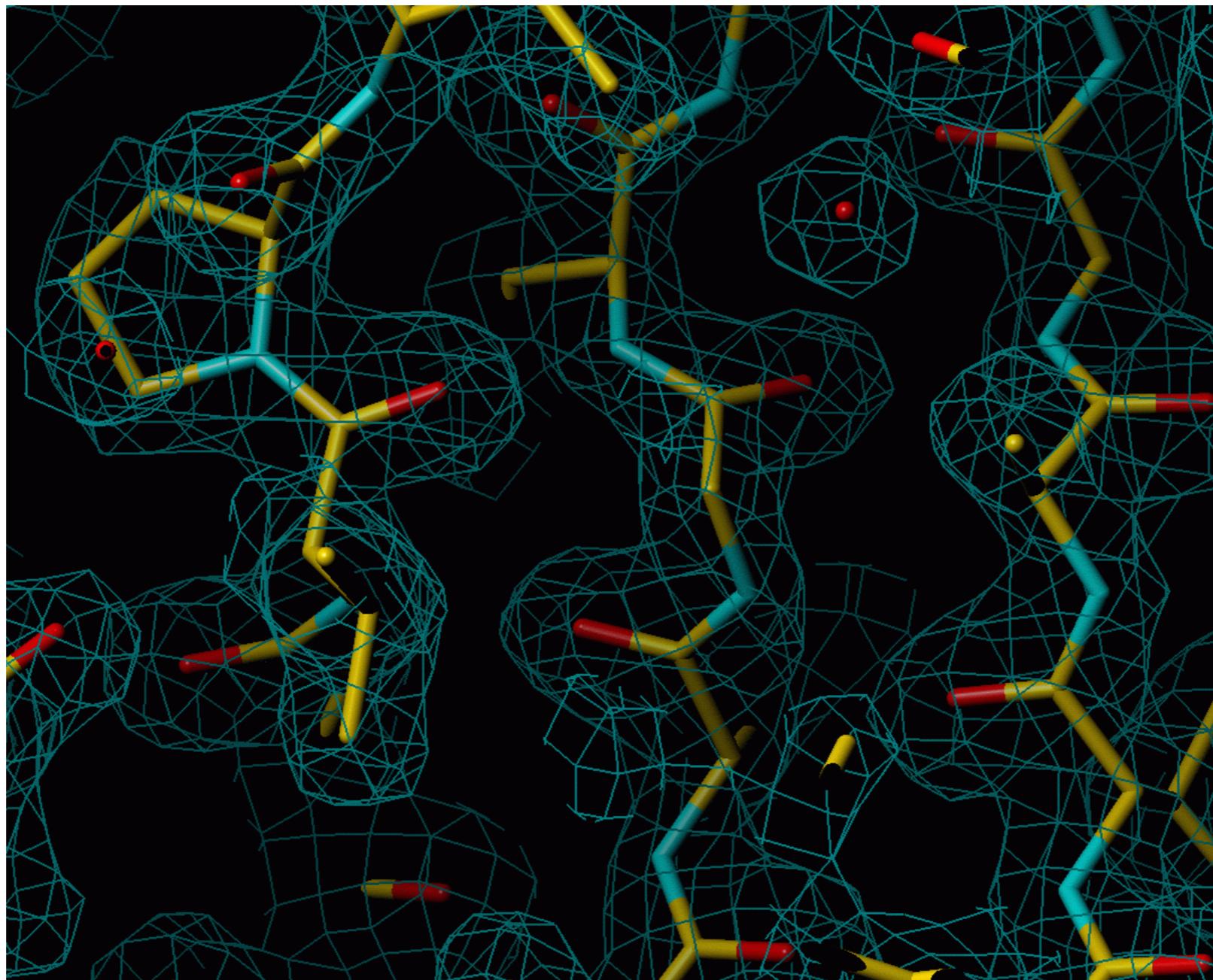
$$\rho(xyz) = 1/V \sum_h \sum_k \sum_l |F(hkl)| \exp[-2\pi i (hx+ky+lz-\alpha(hkl))]$$

where $\alpha(hkl)$ is the phase and $|F(hkl)|$ is the module of the structure factor. The phase being unknown, techniques must be used to obtain the phase and the electronic density. Several approaches have been developed (MIR: multiple isomorphous replacement, MAD: multiple anomalous dispersion, ...)

Finally, a 3D electronic density map is obtained:



Next step: to design a model for the structure; this model must be in agreement with the electronic density map:



The model is then refined to have a good agreement with the density map. Several approaches can be used for that purpose: constrained molecular dynamics, maximum likelihood refinement, manually, ...

Quality of a structure

The **R factor** : evaluates if the model agrees with the measured electronic density:

$$R = \frac{\sum_{h,k,l} |F_{\text{obs}} - F_{\text{calc}}|}{\sum_{h,k,l} |F_{\text{obs}}|}$$

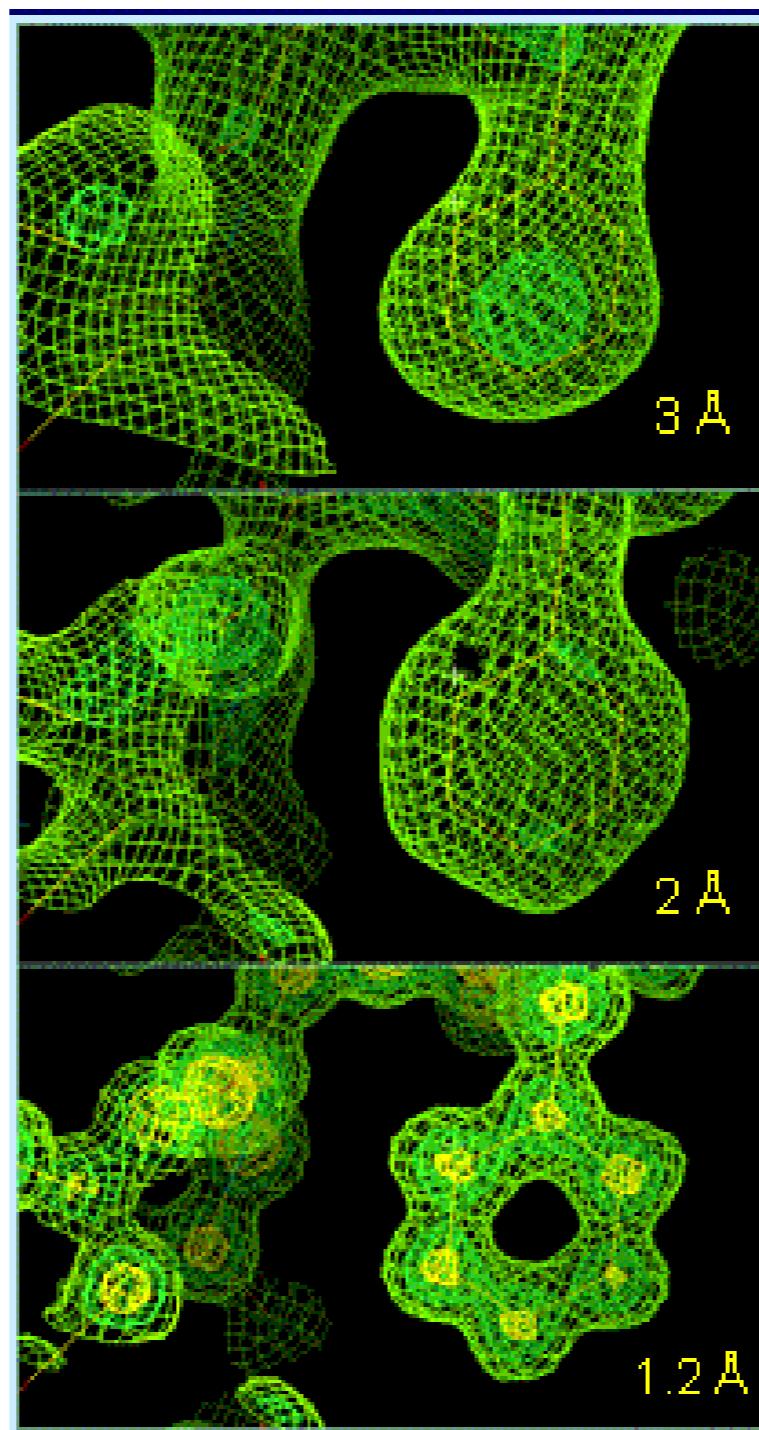
Another measure frequently given in crystallography: the **resolution**. The Bragg law provides a condition to have constructive interference :

$$n \lambda = 2 d \sin \Theta$$
$$\Leftrightarrow d = \frac{n \lambda}{2 \sin \Theta}$$

In theory (see previous slides), if Θ approaches 90° , one accesses to finer details, the limitation being the half of the wavelength. This is valid if the molecules are perfectly ordered and if the conformations are exactly the same in the crystal, with motionless atoms. The electronic density should present peaks at each atomic position.

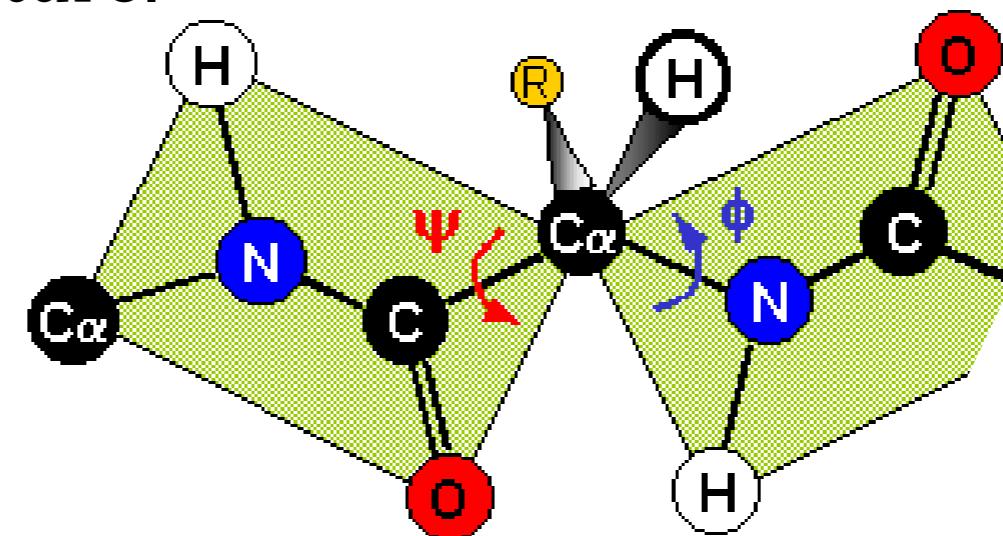
In reality, this is not the case, the diffraction motif disappears when Θ increases. The structures are limited to a detail level where some atoms are not resolved.

Resolution:

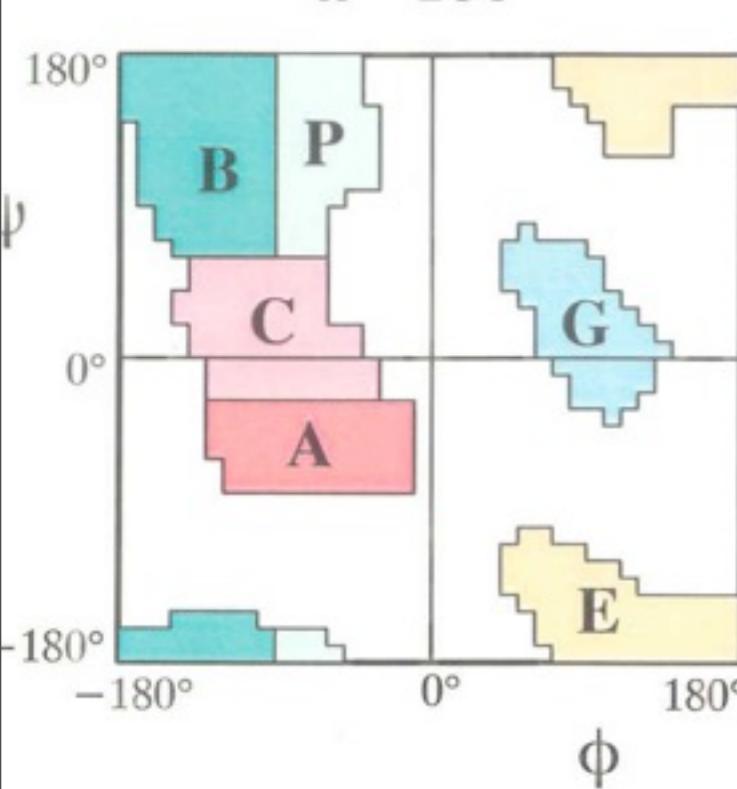


Structure validation

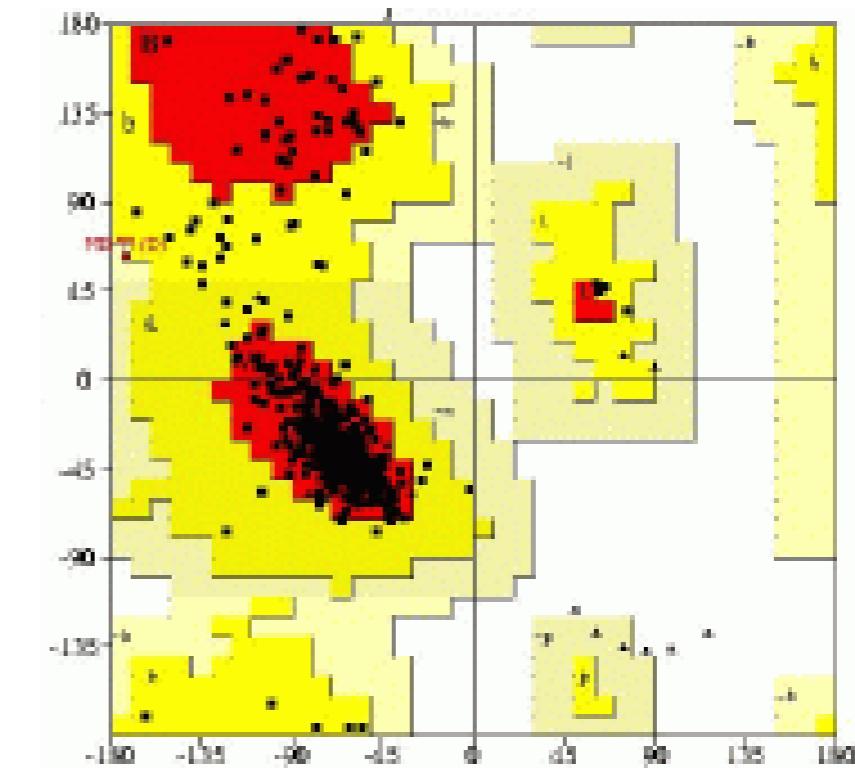
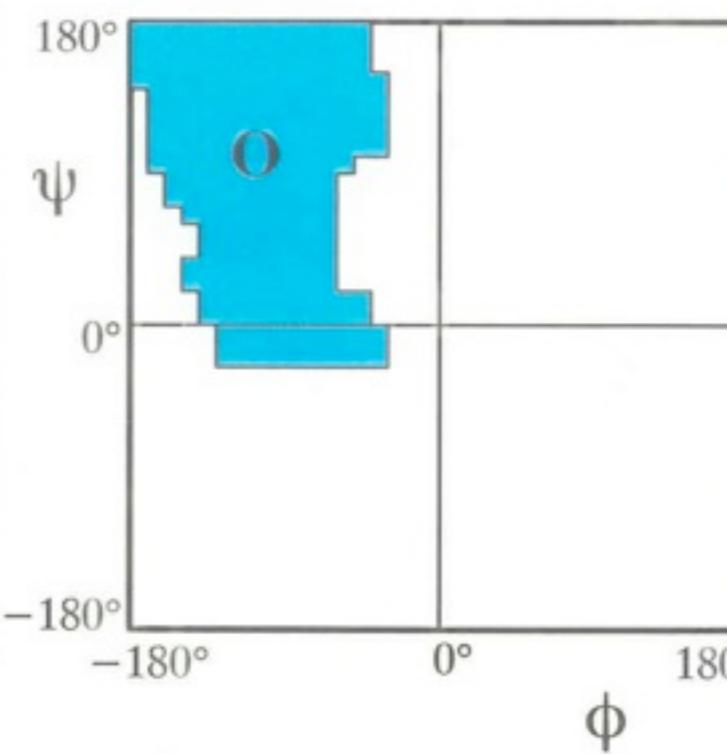
The plot of the value of the dihedral angles on the Ramachandran map can be used to validate the structure.



$\omega=180^\circ$



$\omega=0^\circ$



Procheck can be used to validate the structure. This program evaluates the quality of the stereochemistry of a structure. The program provides a Ramachandran plot, gives a graph with the χ_1 and χ_2 angles of the side chains, compares the structure to other structures presenting a comparable resolution (planarity of the peptidic bond, is the alpha carbon a tetrahedron, ...), bond lengths, ...

2.1.2. Nuclear magnetic resonance (NMR)

In NMR, the biomolecule is studied in solution; it is not necessary to obtain crystals. This technique can be used to study molecules that are partially folded (that are difficult to crystallise, and, if they are crystallised, disordered segments in solutions will be ordered in the crystal).

Proteins of “small” size (400 amino acids) can be studied in NMR. However new techniques allows the study of larger complexes.

NMR is a spectroscopic technique. But NMR is different from optical spectroscopic techniques such as infra-red spectroscopy (see next sections).

Optical spectroscopy: emission or absorption of an electromagnetic wave by a sample is studied. This emission or absorption corresponds to transitions between different energy levels. The matter-radiation interaction is through the electric component of the electromagnetic wave.

In NMR:

- it is necessary to induce a difference between the energy levels, by placing the sample in an magnetic field;
- the matter-radiation interaction is through the magnetic component of the electromagnetic wave, which involves smaller energies compared to the interaction through the electric component;
- the magnetization of the sample is studied, instead of the transition between energy levels.

Theory

Nuclear magnetic resonance requires that the components present a magnetic moment (μ) and an angular moment (J). All nuclei do not present these characteristics.

$\mu = \gamma J$ where γ is the gyromagnetic ratio. It depends on the nucleus and is related to its sensitivity.

$J = \frac{\hbar}{2\pi} I$ where \hbar is the Planck constant divided by 2π and I is the spin of the nucleus.

Nuclei such as ^{12}C , ^{16}O , ^2H present a spin equal to 0, whereas ^1H , ^{13}C , ^{15}N , ^{31}P present an half-integer spin ($I=1/2$) and can be used in NMR.

When a nucleus is placed in a magnetic field, B_0 , the energy is equal to:

$$E = -\mu \cdot B_0$$

If the magnetic field is oriented along the z-axis:

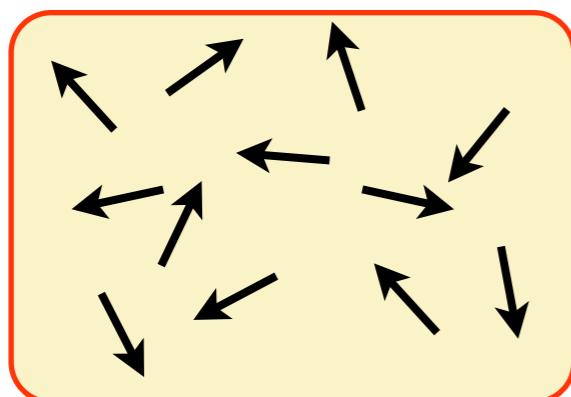
$$E = -\mu_z \cdot B_0 = -\gamma \hbar I_z B_0$$

In quantum mechanics, the allowed values for I_z , m_s , are quantified: $m_s = I$, $I-1$, $I-2$, ..., $-I$.

In this case, $I=1/2$, which means that there are two energetic levels:
 $E_1 = + 1/2 \gamma \hbar B_0$ et $E_2 = - 1/2 \gamma \hbar B_0$

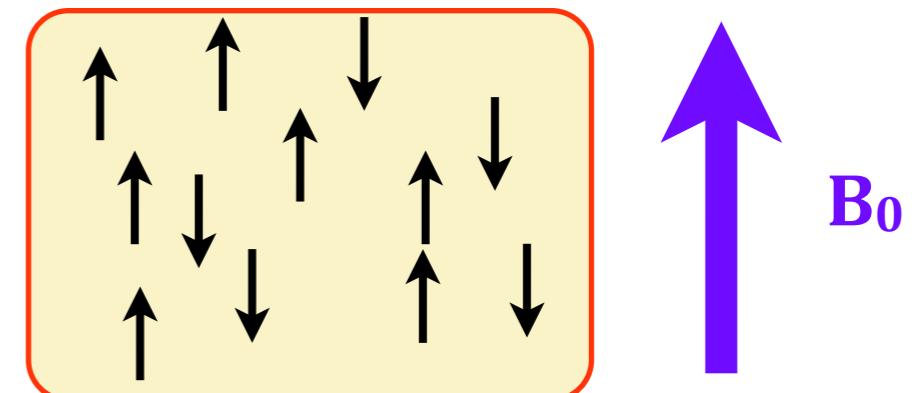
Without magnetic field, the nuclei
are unordered.

There is no energy difference
between the states.



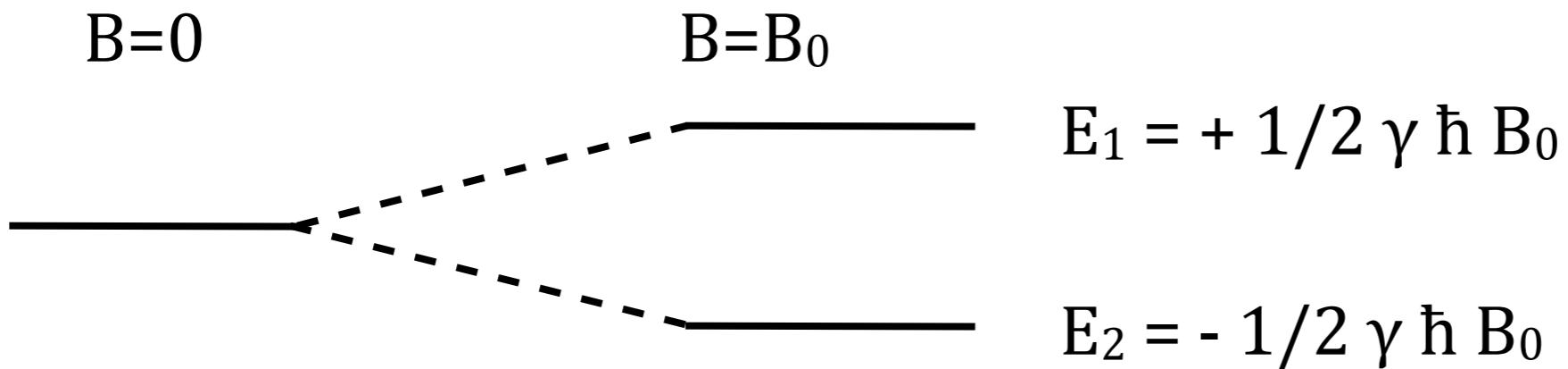
With a magnetic field, the nuclei with
a magnetic moment are oriented.

There is an energy difference between
the states.



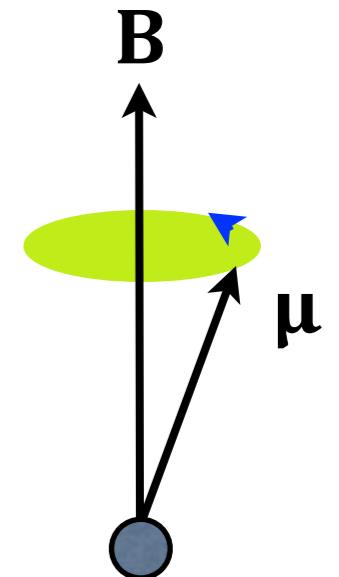
There is a small excess of nuclei that
are oriented along the field.

There is an energy difference between the levels:



$\Delta E = \gamma \hbar B_0 = h \omega_0$ where ω_0 is the Larmor frequency. It corresponds to the frequency of precession around the z-axis, or to the resonance frequency. In a magnetic field of 18,7 Tesla, the resonance frequency of a proton is equal to 800 MHz.

Precession



Note that the energy difference between two levels depends on the intensity of the magnetic field. The frequency of Larmor depends on B_0 and γ .

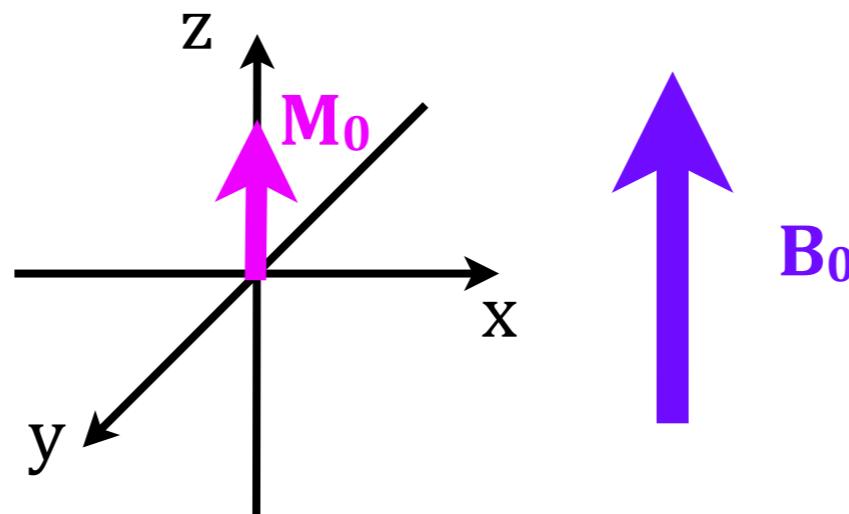
The ratio between the population of two levels is given by the Boltzmann law:

$$\frac{N_1}{N_2} = \exp [-\Delta E / kT]$$

For ^1H , $\gamma=2,6751987 \cdot 10^8$ Rad.T $^{-1}$.s $^{-1}$. Thus:

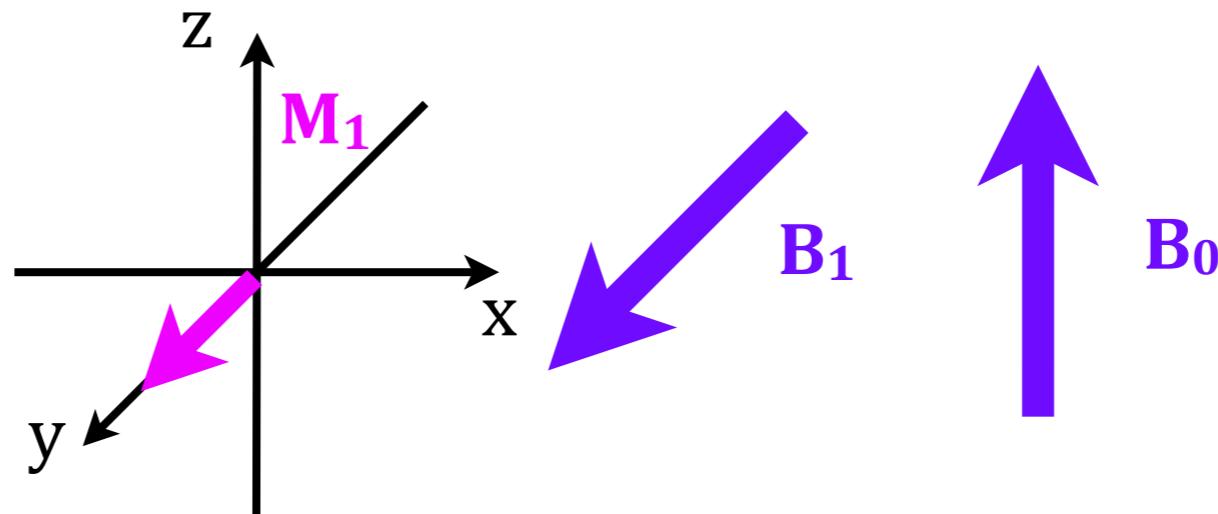
B ₀ (T)	ω ₀ (MHz)	ΔE (J/mol)	N ₂ /N ₁
4,7	200	2,20•10 ⁻³	1,000032
9,4	400	4,40•10 ⁻³	1,000064
14,1	600	6,61•10 ⁻³	1,000096
21,1	900	9,91•10 ⁻³	1,000144

The ratio between the population of the two levels is small, but increases with the intensity of the magnetic field. The magnetic moments of each nucleus add up, leading to a macroscopic magnetization: $\mathbf{M}_0 = \sum \mu$. This magnetization is oriented along the magnetic field (z-axis in our convention), and the x and y components are null.



There is only a precession around the z-axis.

A magnetic field \mathbf{B}_1 , perpendicular to \mathbf{B}_0 for instance, can be applied to perturb the system. If the frequency is close to the Larmor frequency, the magnetic moments will switch in the xy plane:



When the magnetic field \mathbf{B}_1 will be switched off, the magnetic moments will reorient along the z-axis. The magnetization will tend towards M_z , its z-axis component, whereas M_x et M_y will tend towards 0. In NMR, we will follow the evolution of the magnetization (relaxation).

There are two time constants:

T_1 : the spin-lattice relaxation, which measures the return to M_z

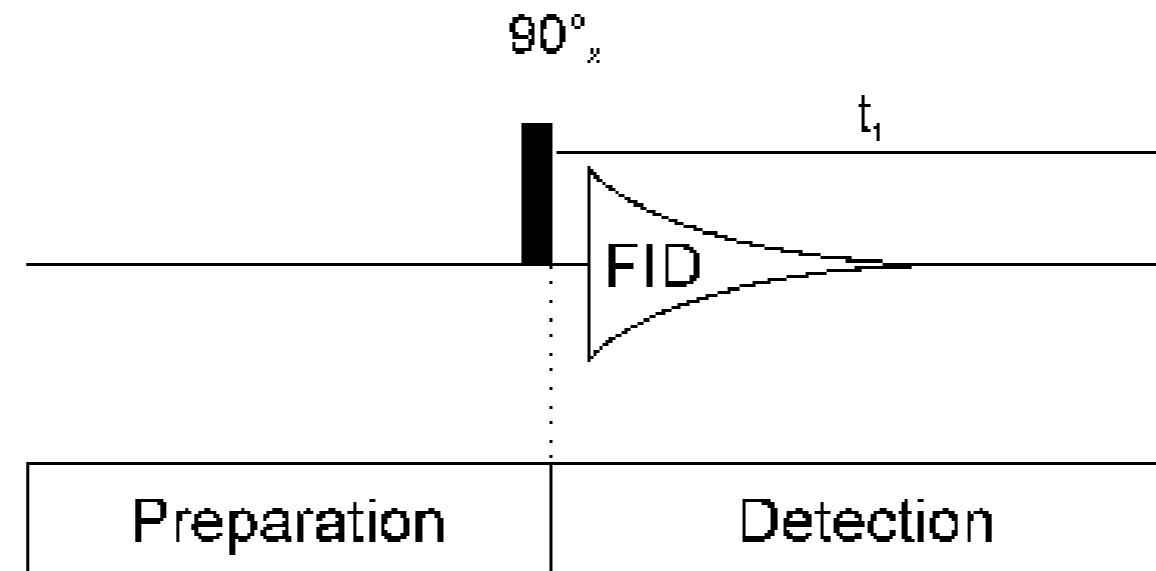
T_2 : the spin-spin relaxation, which measures the evolution towards 0 of M_x and M_y

Chemical shift:

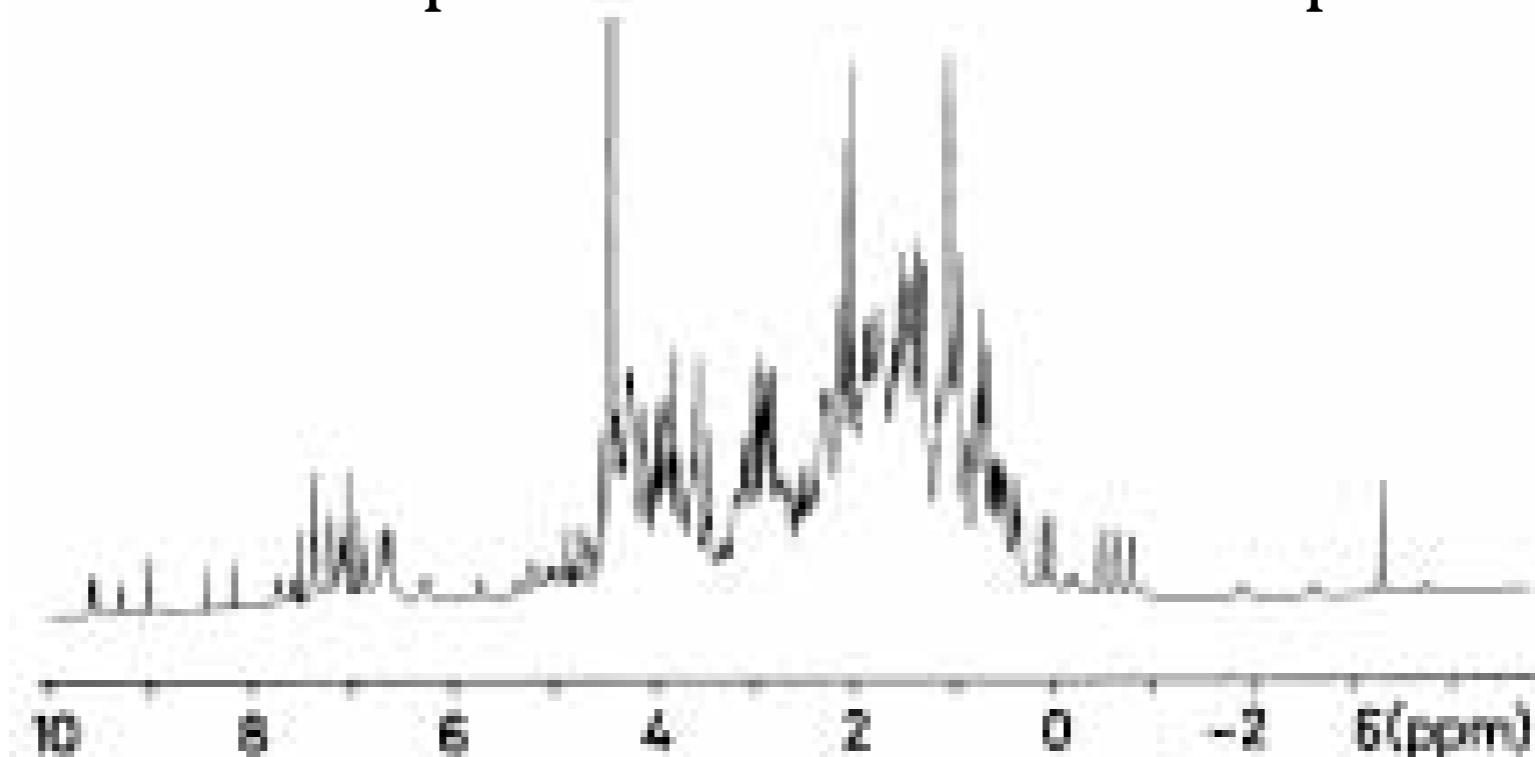
The frequency of resonance of different ^1H (for instance) depends on their environment. The other nuclei influence the magnetic field felted by the nucleus: $\mathbf{M}_0(1-\sigma)$, where σ is about 10^{-6} . It is thus important to have a very homogeneous magnetic field.

Practically: all the frequencies are not tested for \mathbf{B}_1 . Working with short time pulse permit to scan a broad spectrum of frequencies and to excite all the nuclei.

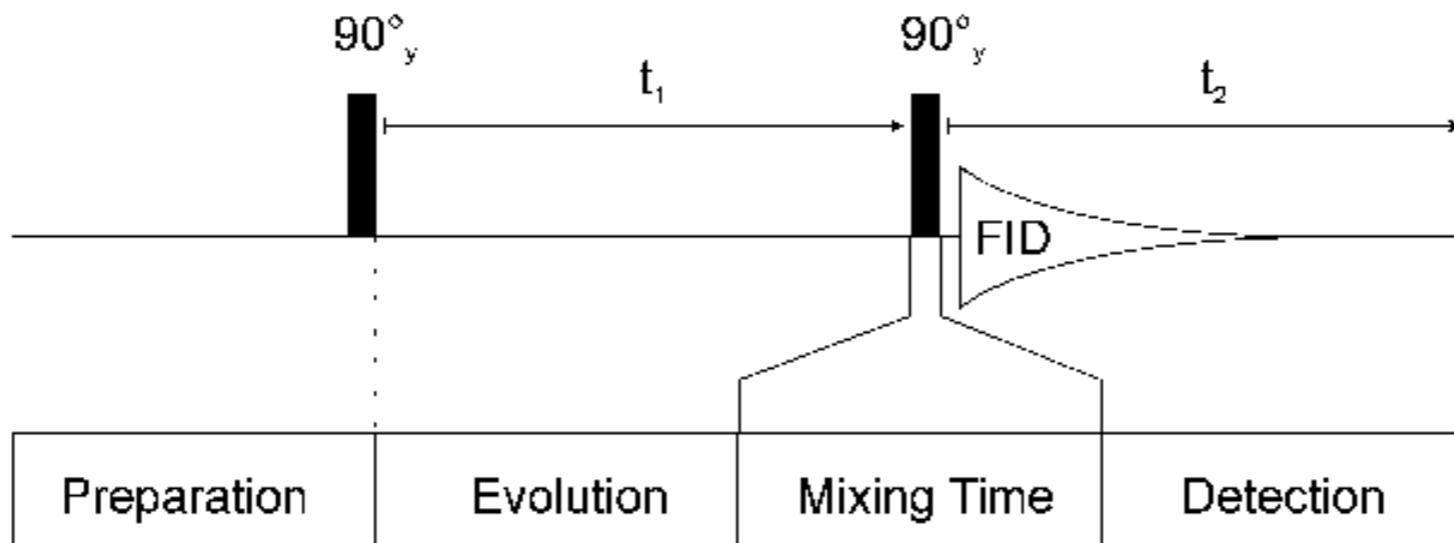
Different impulse sequences will lead to different behaviour and to different spectrum. In 1D NMR, there are two steps: En RMN 1D, nous avons deux phase: preparation (for instance an impulse of 90°) and detection.



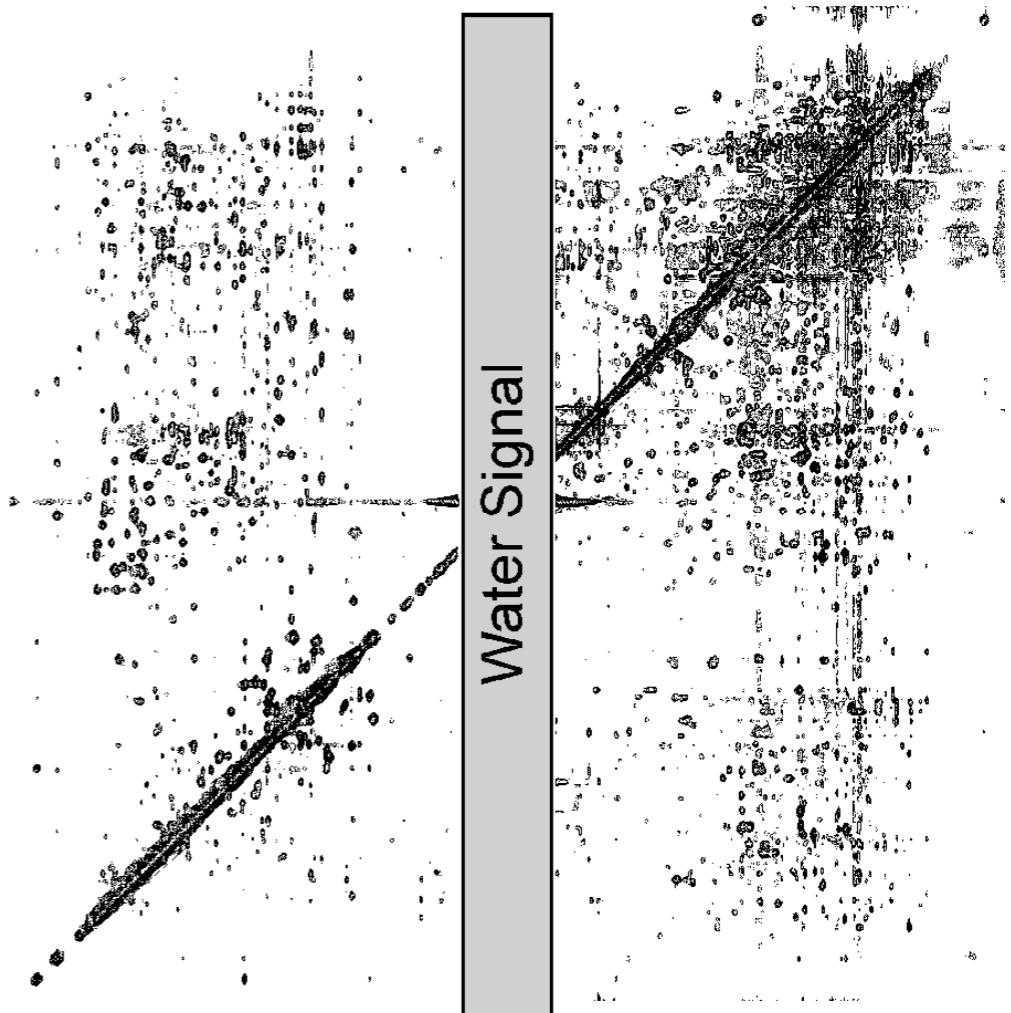
The spectrum obtained for a protein is difficult to interpret:



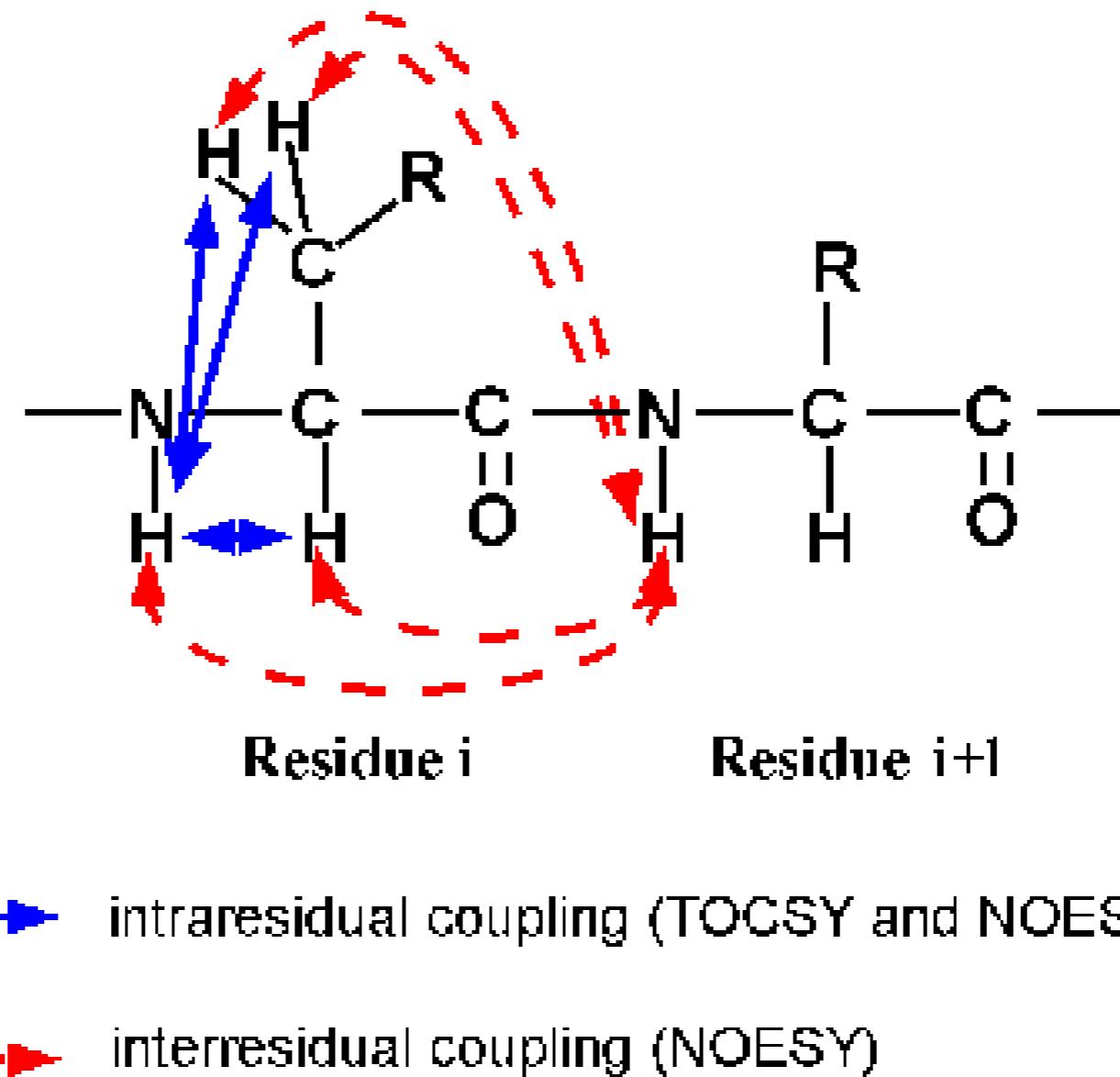
2D NMR is characterized by a different impulse sequence:



We obtain a spectrum on two frequency axis. It is easier to detect the ^1H that are in contact in the protein with this spectrum. Distance constraints are derived from the spectrum, and these constraints will be used to build a model of the 3D structure of the protein.



According to the experiment, contacts between specific ^1H can be highlighted:



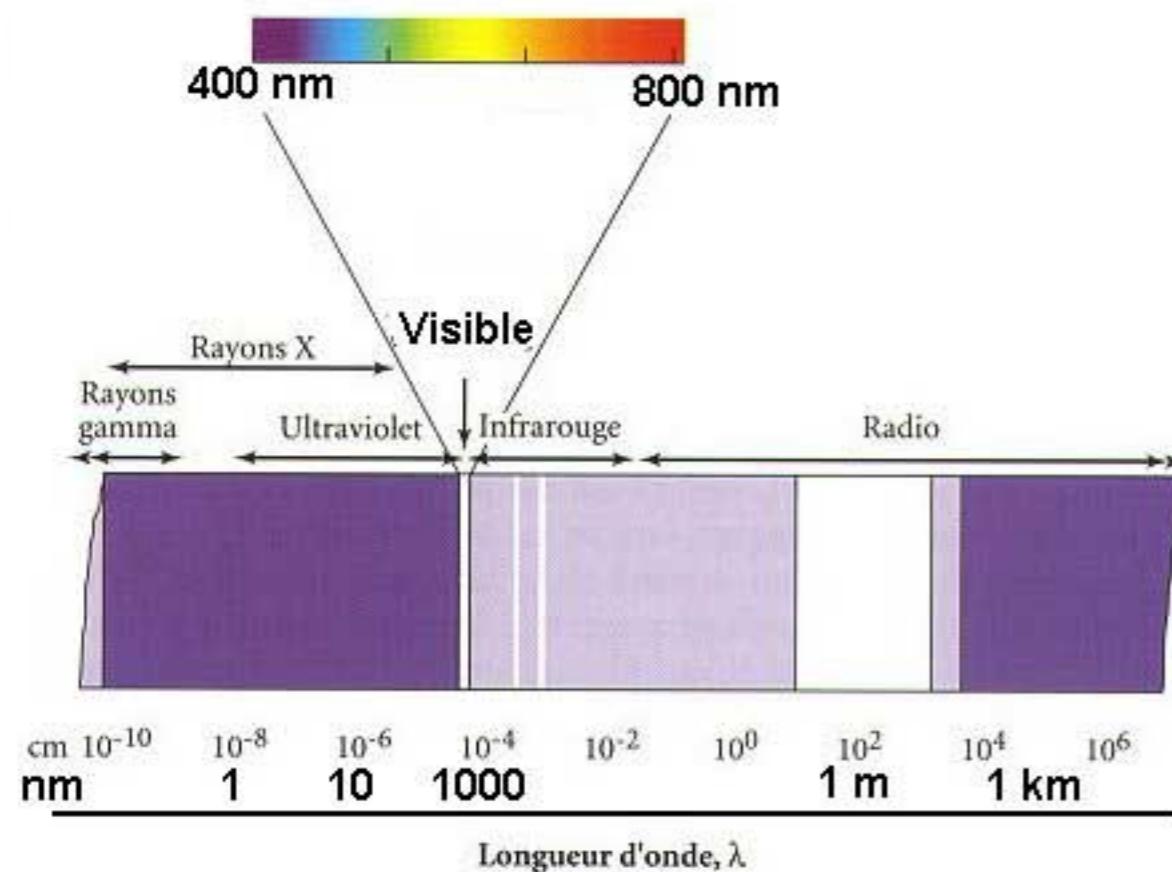
A model will be built according to the constraints obtained by NMR, and this model will be refined by *in silico* techniques.

2.2. Information about the structure of biomolecules

The methods that will be presented in this section are complementary to other approaches (NMR, Xray,).

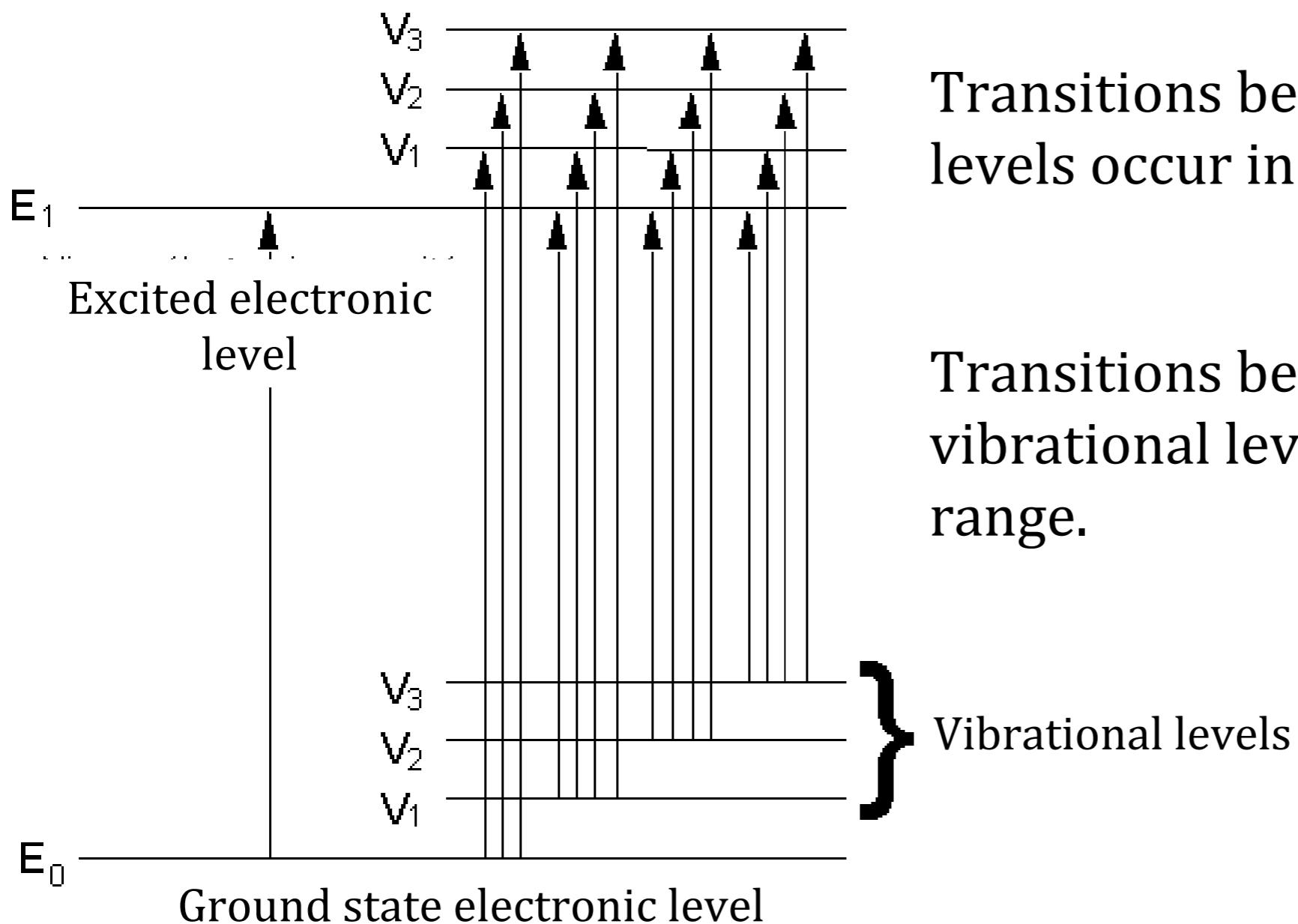
2.2.1. Principle of emission and absorption spectroscopy

In UV, visible, infra-red spectroscopy, the aim is to detect wavelengths at which a molecule absorb or emit an electromagnetic radiation. A molecule does not absorb/emit in a continuous spectrum; this absorption /emission is quantified. It corresponds to transition between energy levels.



The energy, E , associated to an electromagnetic wave of frequency f , and of wave length λ (c =speed of the light) is equal to:

$$E = h f = h \frac{c}{\lambda}$$



Transitions between molecular electronic levels occur in the UV-visible range.

Transitions between molecular vibrational levels occur in the infra-red range.

2.2.2. Circular dichroism (CD)

It is an absorption spectroscopy that can be used to:

- evaluate the percentage of secondary structure of a protein;
- follow a conformational change;
- study the effect of environment on the structure;
- study the folding and the denaturation of proteins;
- study the dynamics of biomolecules.

CD needs small quantities of biomolecules (about 100 micrograms, or less) and the measure is quite fast (about 1/2 hour).

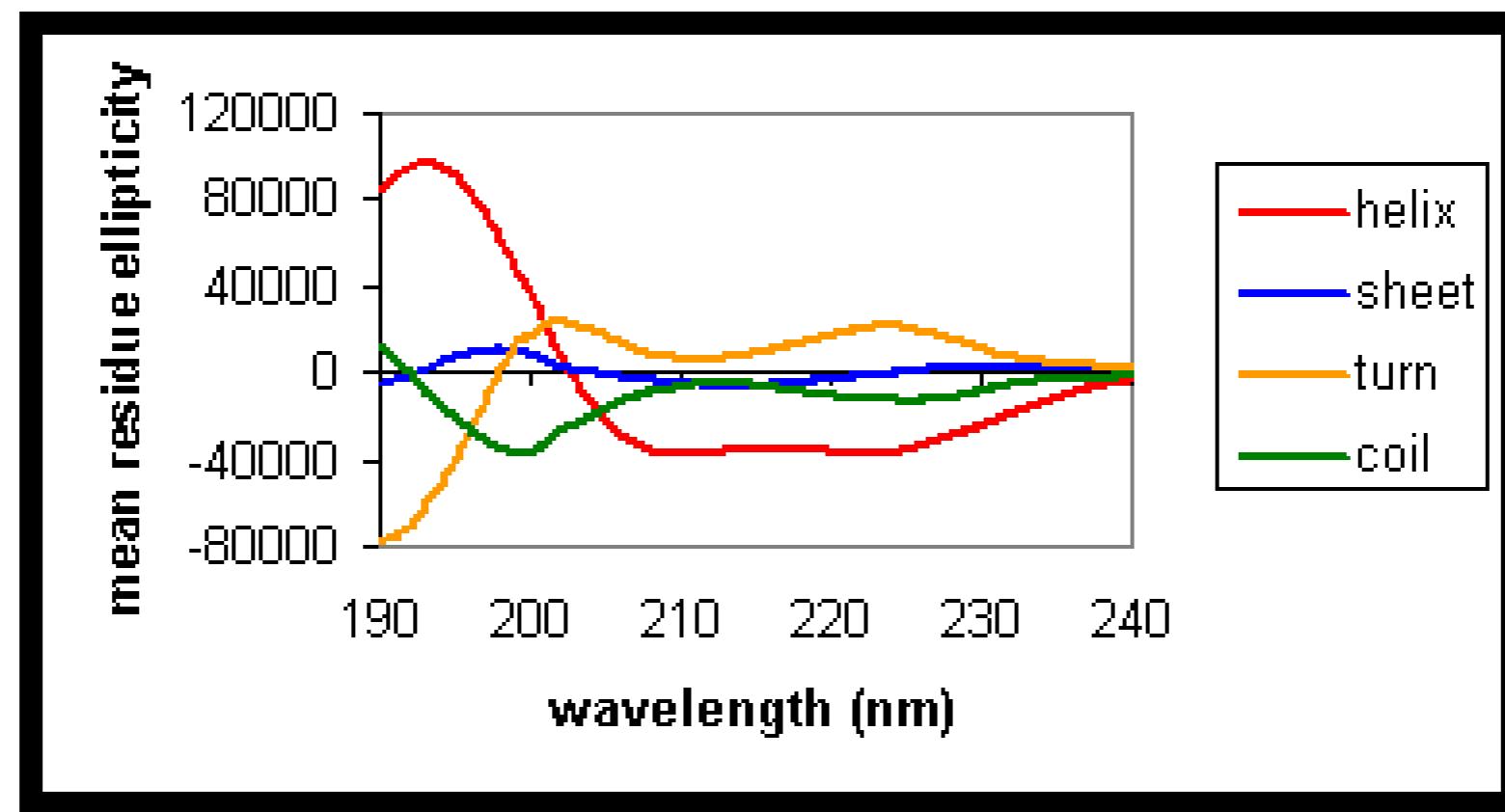
The differential absorption of left and right circularly polarized light is measured in CD. In the case of an electromagnetic wave circularly polarized, the norm of the electric field vector is constant and, at a fixed point in space, describes a circle as time progresses.

Chiral molecules present a differential absorption of right and left circularly polarized light, as well as some structures present in biomolecules (DNA, secondary structure of proteins). These molecules thus present a CD spectrum.

According to the Beer-Lambert law, at a given wave length, the difference between the absorbance of right and left polarized light is: $\Delta A = (\varepsilon_G - \varepsilon_D) C L$ where C is the concentration, L the length of the optical path and ε_G ε_D are the left and right molar extinction coefficients.

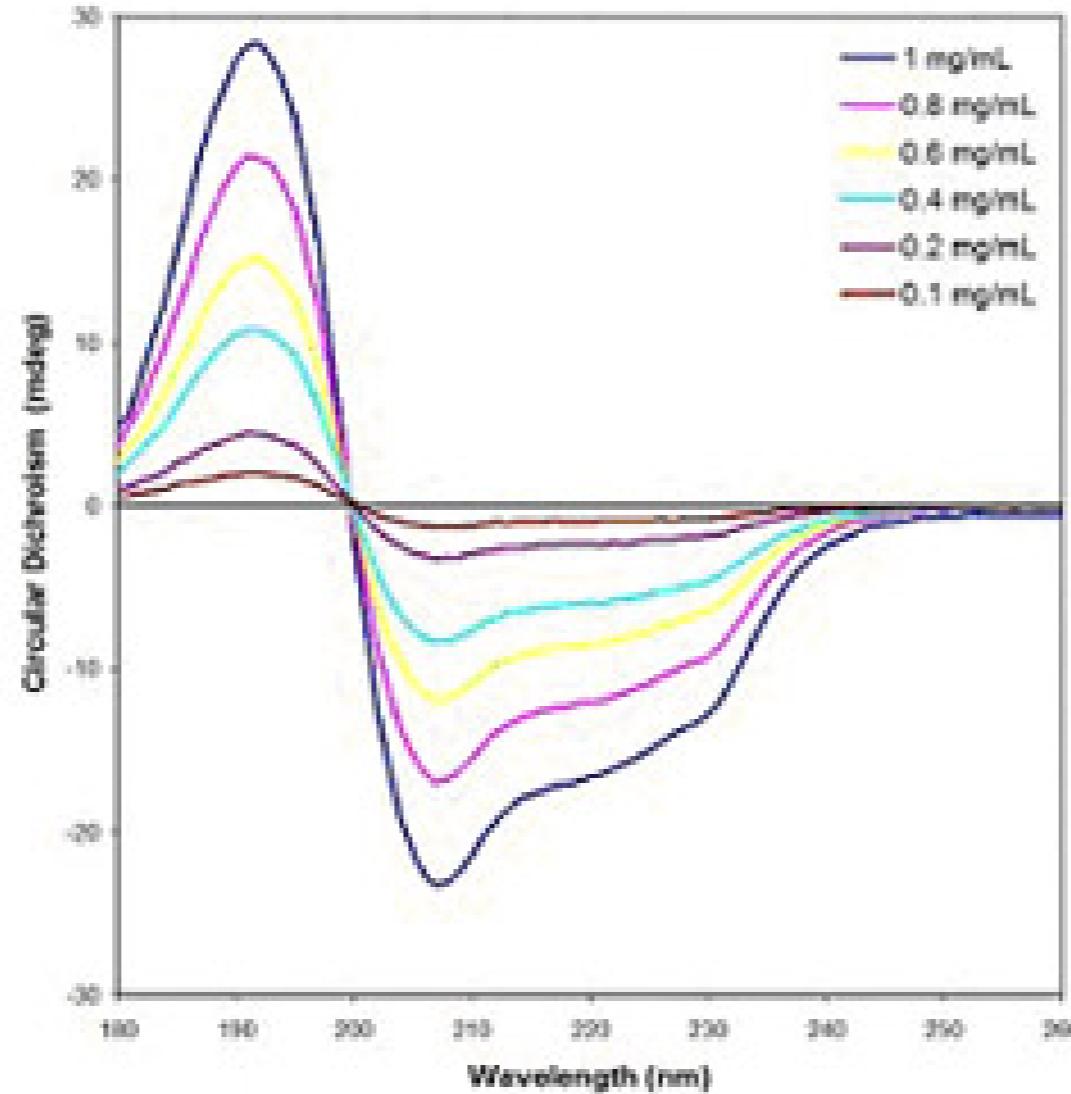
In CD, the molar ellipticity (θ) is measured; it depends on the molar extinction coefficients ϵ_G and ϵ_D ($\theta \div \Delta\epsilon$).

The ellipticity depends on the type of secondary structure in a protein:



The spectrum is in the wavelength range 190 - 300 nm (UV). This range corresponds to electronic transitions in main chains and side chains of peptides, and in purines and pyrimidines in DNA.

The spectrum evolves according to the experimental conditions:



The spectrum obtained for a protein is the weighted sum of the contribution of the different secondary structures that are present in the protein.

The spectrum deconvolution will lead to the percentage of each type of secondary structure.

2.2.3. Infra-red spectroscopy (IR)

The interaction light-matter can be used to obtain some information.

The UV range corresponds to transitions between molecular electronic levels. Infra-red radiation corresponds to transitions between molecular vibrational levels.

In the case of proteins, the absorption (or emission) of the NH and CO groups of the main chain will be studied.

Table 8-5

Characteristics of principal infrared absorption bands of the peptide group

Vibration	$\delta\mu/\delta R$	Hydrogen-bonded forms		Non-hydrogen-bonded	
		α Helix		β Sheet	
		Frequency (cm^{-1})	Dichroism	Frequency (cm^{-1})	Dichroism
N—H stretch	$\leftarrow\text{N}\text{—H}\rightarrow$	\leftrightarrow 3,290–3,300		3,280–3,300	⊥
Amide I (C=O stretch)	$\leftarrow\text{C}\text{=O}\rightarrow$	\leftrightarrow 1,650–1,660		1,630	⊥
Amide II	$\begin{array}{c} \uparrow \\ \text{H} \\ \downarrow \\ \leftarrow\text{C}\text{—N}\rightarrow \end{array}$	\leftrightarrow 1,540–1,550	⊥	1,520–1,525	
					<1,520?

SOURCE: Adapted from J. A. Schellman and C. Schellman, in *The Proteins*, 2d ed., vol. 2, ed. H. Neurath (New York: Academic Press, 1962), p. 1.

The amide I and II bands are used in IR spectroscopy of proteins. The wavelength at which the chemical groups absorb depends on the secondary structure in which they are included:

Principal Amide I Frequencies Characteristic of Protein Secondary Structures		
Conformation	H ₂ O	D ₂ O
α -helix	1650–1657	1647–1654
Antiparallel β -sheet	1612–1640; 1670–1690 (weak)	1628–1635
Parallel β -sheet	1626–1640	
Turn	1655–1675 1680–1696	
Unordered	1640–1651	1643

Since water also absorbs in the IR range, the solvent is often a H₂O-D₂O combination, or an hydrated film.

2.3. Thermodynamic stability of biomolecules

A few reminders of thermodynamics:

The internal energy, $U = q + w$, where q is the heat and w is the work; we will consider only mechanical work: $w = -p \Delta V$ (p =pressure, V =volume).

The enthalpy, $H = U+PV$

$$\Delta H = \Delta U + \Delta(PV) = \Delta U + V\Delta P + P\Delta V$$

The free energy change, $\Delta G = \Delta H - T\Delta S$

At constant volume: $q_V = \Delta U$

At constant pressure: $q_p = \Delta H$

The heat is also defined by:

$$q_V = n C_V \Delta T \text{ et } q_p = n C_p \Delta T$$

where C_p and C_V are the constant-pressure and the constant-volume heat capacity, respectively.

Some applications of microcalorimetry:

Protein-ligand
interactions

Enzyme-substrate binding

Nucleic acid-protein

Binding to a metal

Sugar-protein

Protein-protein interactions
Membrane receptors
Complexes between several
enzymes

Protein stability, wrong folding, aggregation

Thermodynamics of folding

Incorrectly folded proteins

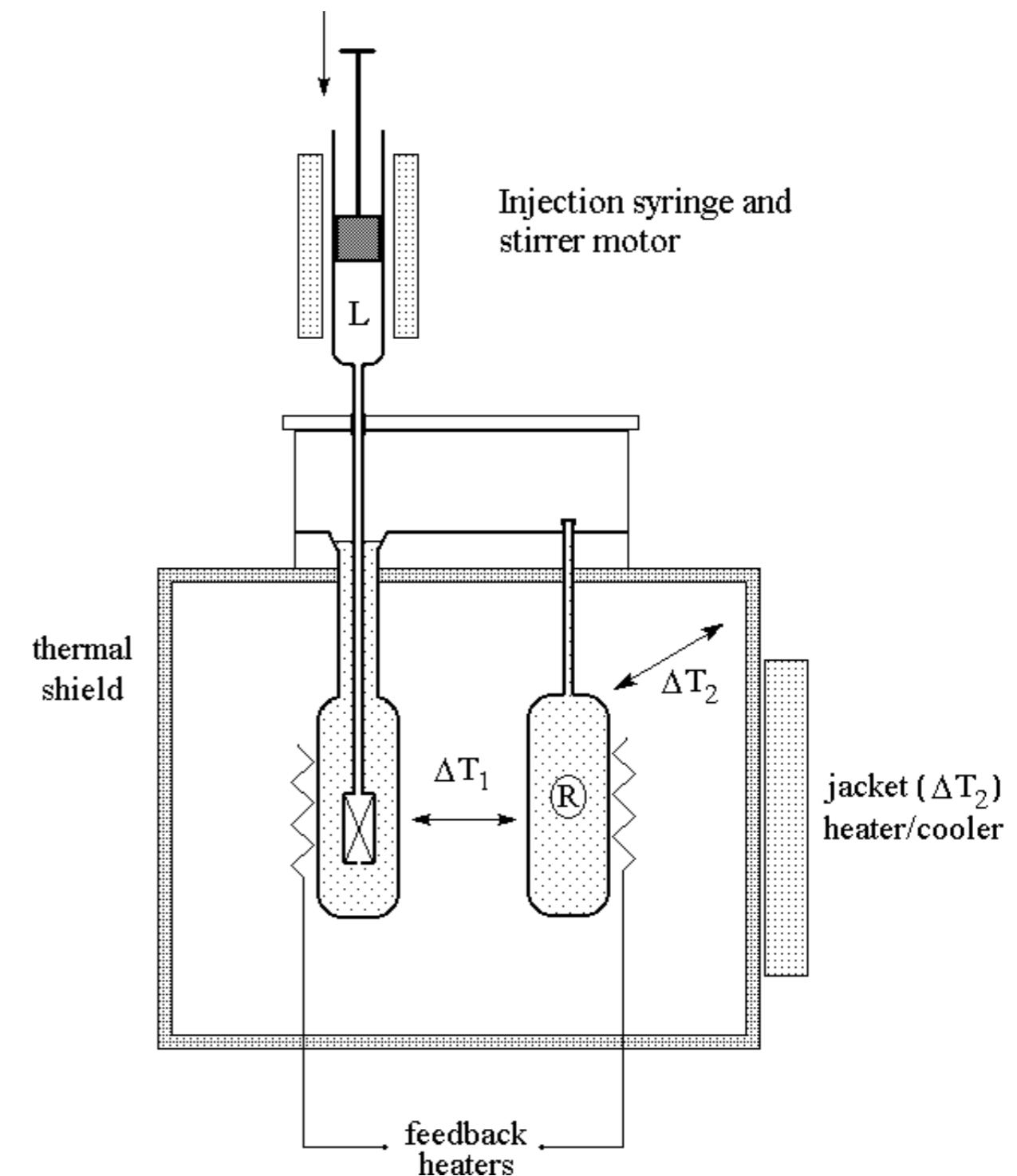
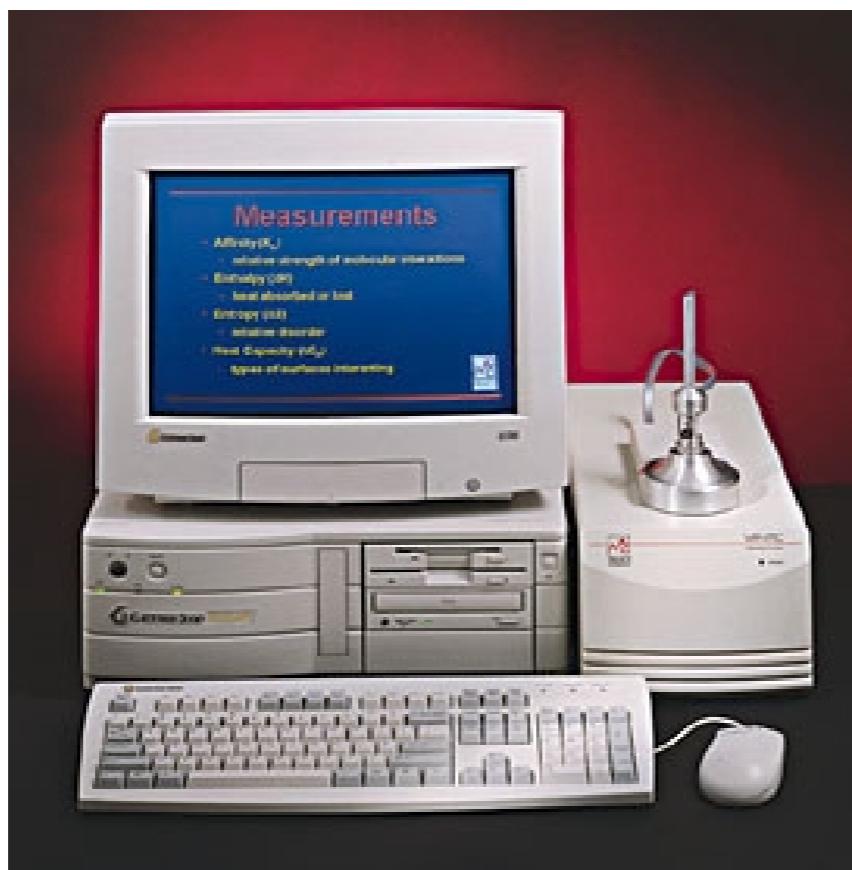
Drug-receptor interaction

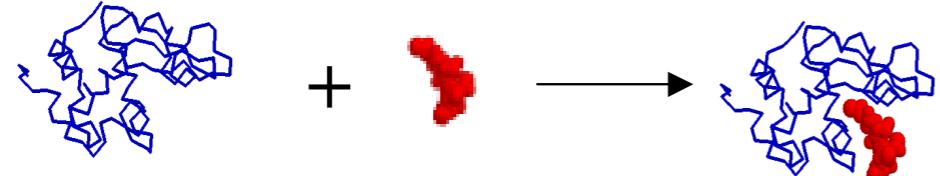
Thermodynamics of non-covalent interactions in solution

Approaches used:

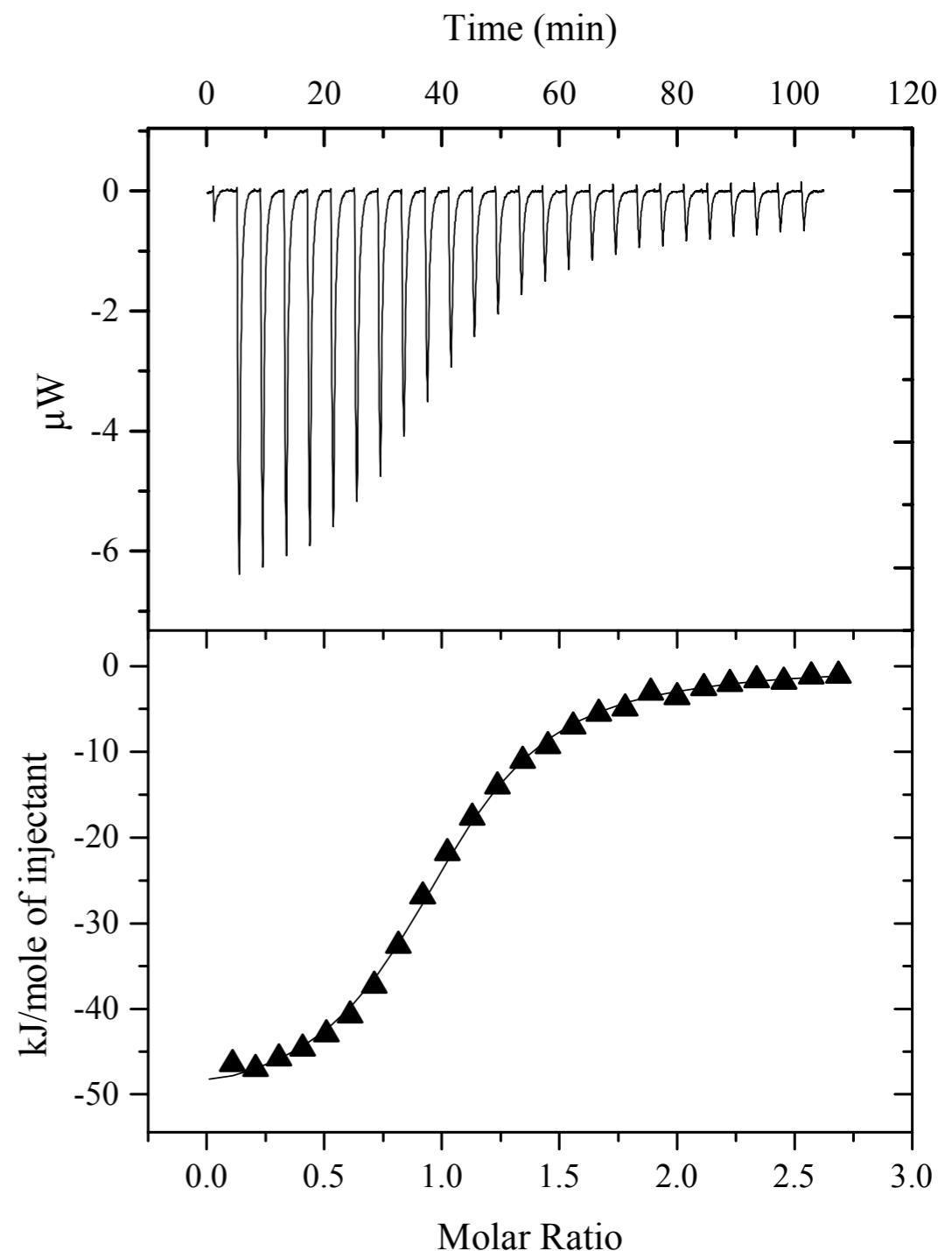
- ITC (isothermal titration calorimetry): it measures the heat emitted/absorbed during a reaction between molecules in a sample and molecules that are injected.
- DSC (differential scanning calorimetry): it measures the difference in heat required to increase the temperature of a sample compared to a reference as a function of temperature.

The calorimeter contains two identical cells: a reference cell and a cell to inject the sample. The power to keep an equal temperature between both cells is measured as a function of time.

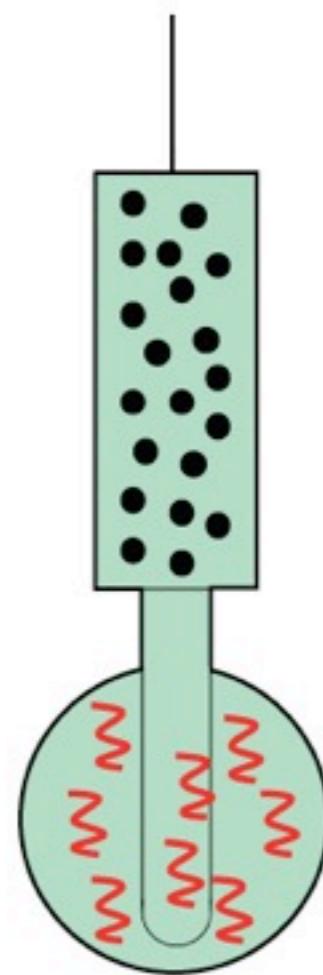




Each exothermic impulse corresponds to the injection of ligand in a solution of proteins. The lower graph is the integrated data; a curve has been fitted considering a one binding site model. The stoichiometry can be deduced, as well as the affinity and the enthalpy of the binding.



ITC - before titration



- Ligand - in syringe
- ~~~~ Protein (macromolecule) – in cell

From: <http://www.microcal.com/technology/itc-animation.asp>

The amount of heat emitted or absorbed after each injection depends on the amount of ligands injected, the concentration of the molecule in the cell, the number of binding sites, the ΔH of reaction.

- It requires a $\Delta H \geq \pm 3\text{-}5 \text{ kcal/mol}$ for an accurate determination.
- Concentration of the macromolecule about $1\mu\text{M}$ to 1mM in a volume about 1,700 ml.
- Needs a large amount of consumables.
- Less accurate with very strong or very weak binding ; the problem of strong binding can be bypassed by first creating a low affinity complex and then by creating a competition with a complex presenting a stronger binding.
- The method is slow (2-3 h).
- The ligand and the molecule must be in the same buffer to avoid peaks due to dilution.
- Take into account the protonation / unprotonation effect.

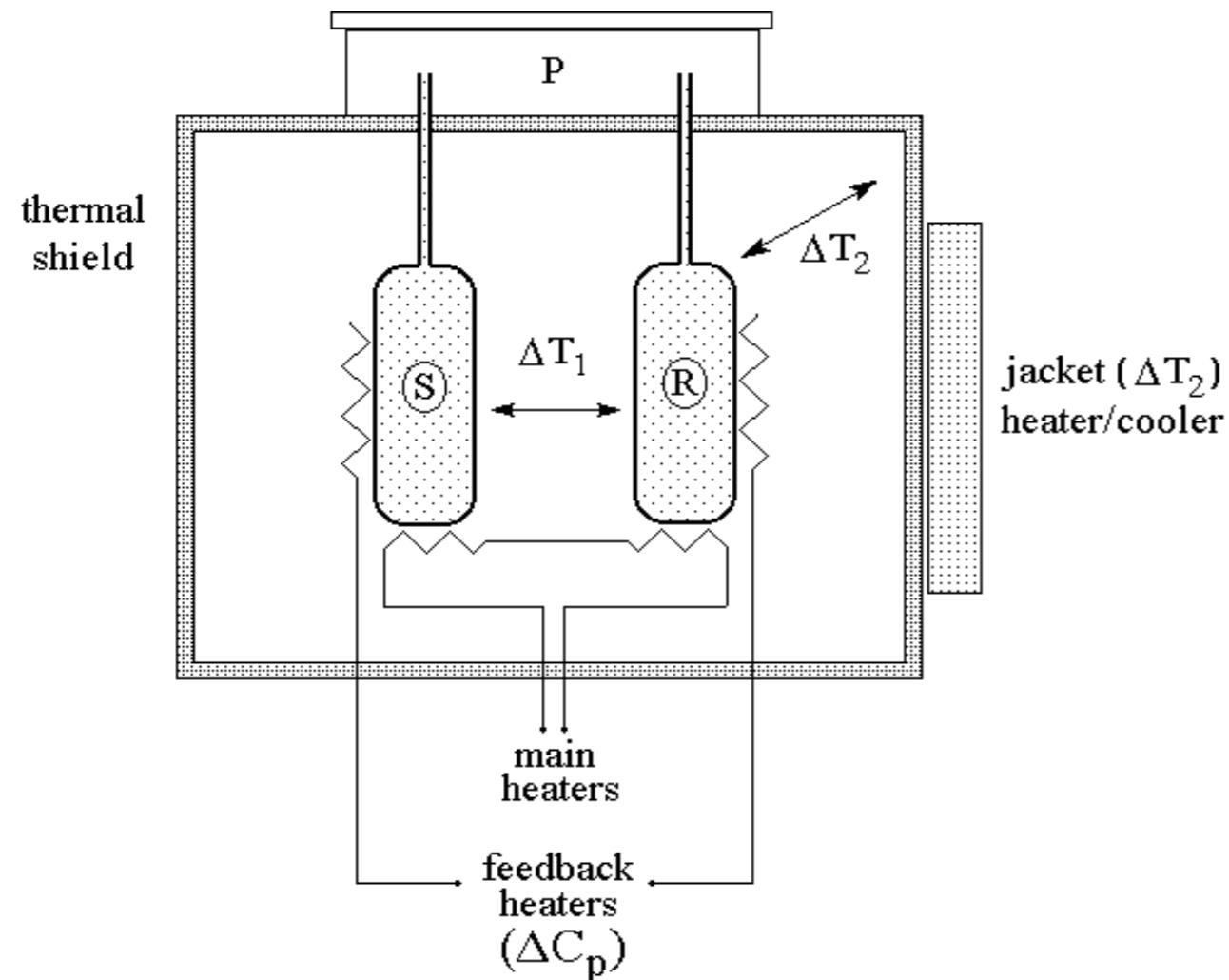
DSC

It measures the difference in heat required to increase the temperature of a sample compared to a reference as a function of temperature. It corresponds to the measure of the heat capacity of the dissolved sample. The dissociation of a complex, or the denaturation of a macromolecule will generate a peak at the transition temperature, Tm.

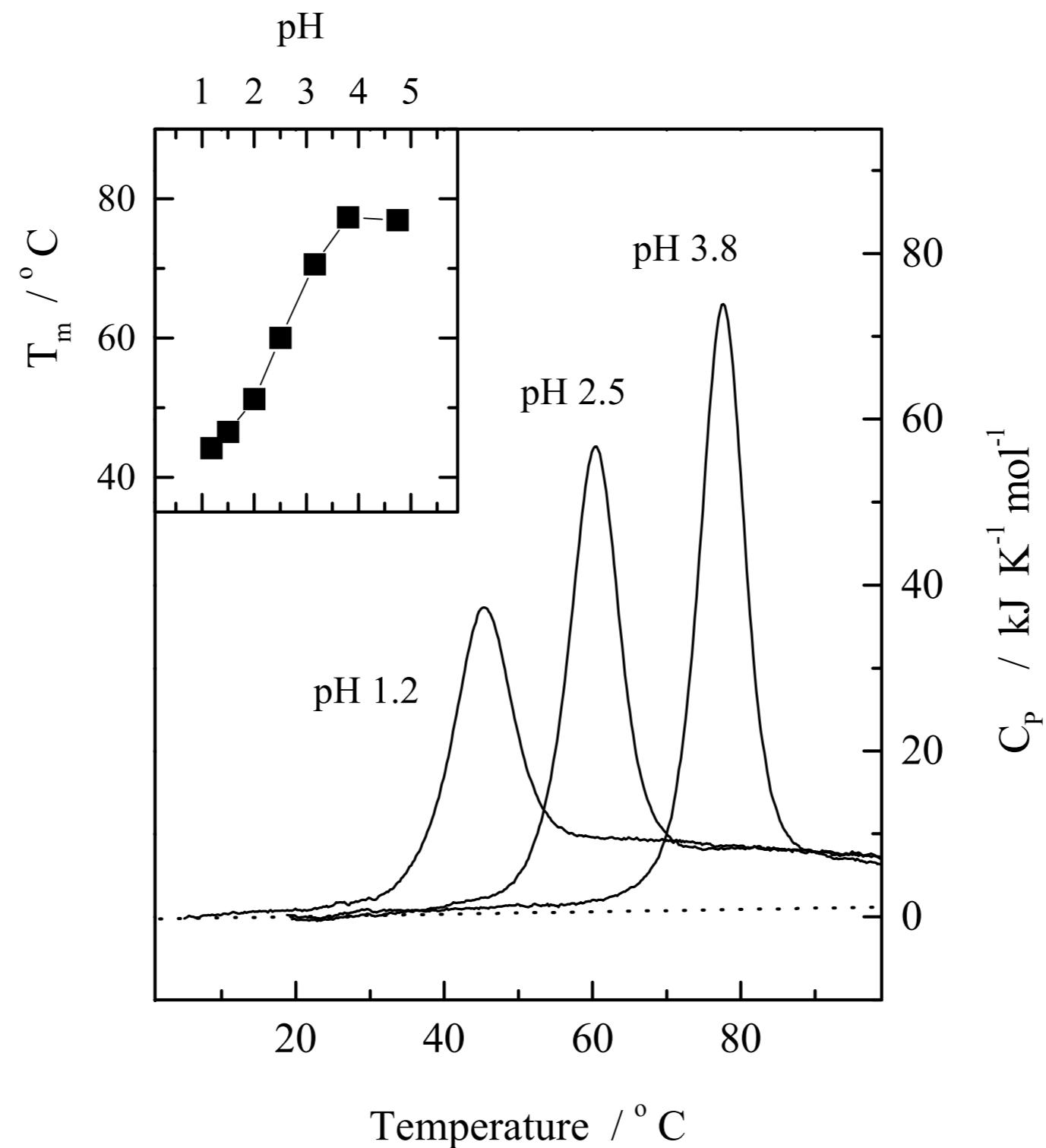
This technique allows:

- to measure the stability of macromolecules.
- to identify the number of domains in a protein.
- to determine the oligomeric state of a protein.

- The concentration of macromolecule is about $1\mu\text{M}$ - 1mM in a volume about 0,500 ml.



Typical DSC curve



3. Intra- and inter-molecular interactions

Biomolecules are thermodynamically stabilized by covalent and non-covalent interactions.

Differences between covalent and non-covalent interactions:

- Covalent bond: chemical bond characterized by the sharing of one or several electron pairs between atoms.
- The covalent bond is in general stronger than the other types of interactions, such as Van der Waals interaction. The strength of ionic bond is decreased by the presence of solvent ; in presence of water, the ionic interaction is weaker than a covalent bond.

Non-covalent interactions:

- hold the two DNA strands together;
- allow a protein chain to fold;
- allow the enzyme-substrate binding;
- allow protein-protein, protein-DNA, protein-nucleic acid assemblies (quaternary structure, binding of the transcription factors to DNA, assembly of large molecular complexes such as ribosomes, ...)

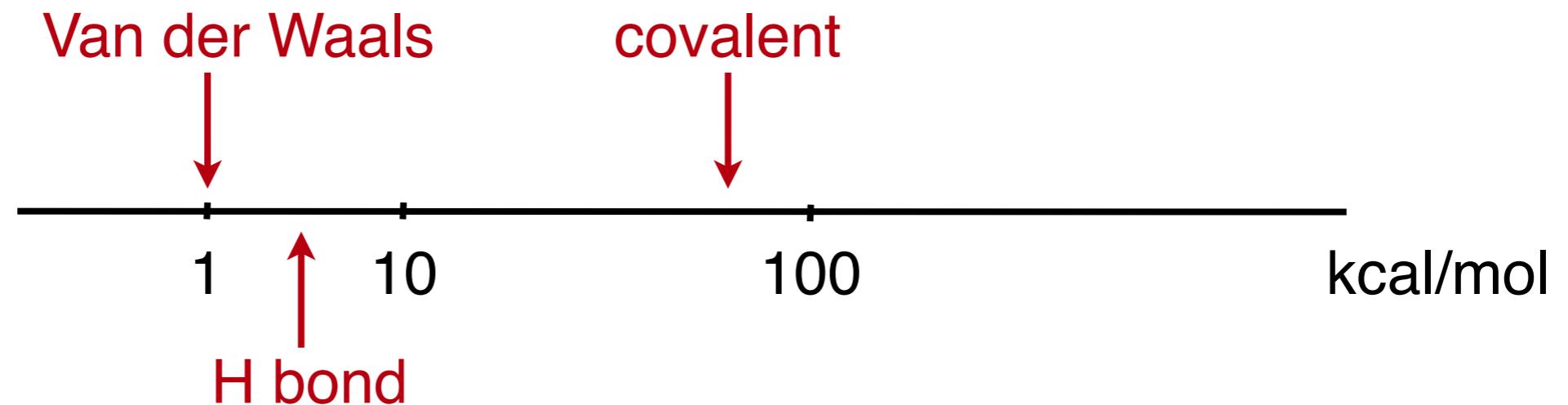
The energy of a covalent bond is about $100 - 150 \text{ kT}$. These bonds are thus not broken by thermal fluctuations.

The 3D conformation of biomolecules is due to the numerous non-covalent interactions. These interactions play an important role in the flexibility of these biomolecules.

The non-covalent interactions that will be described are:

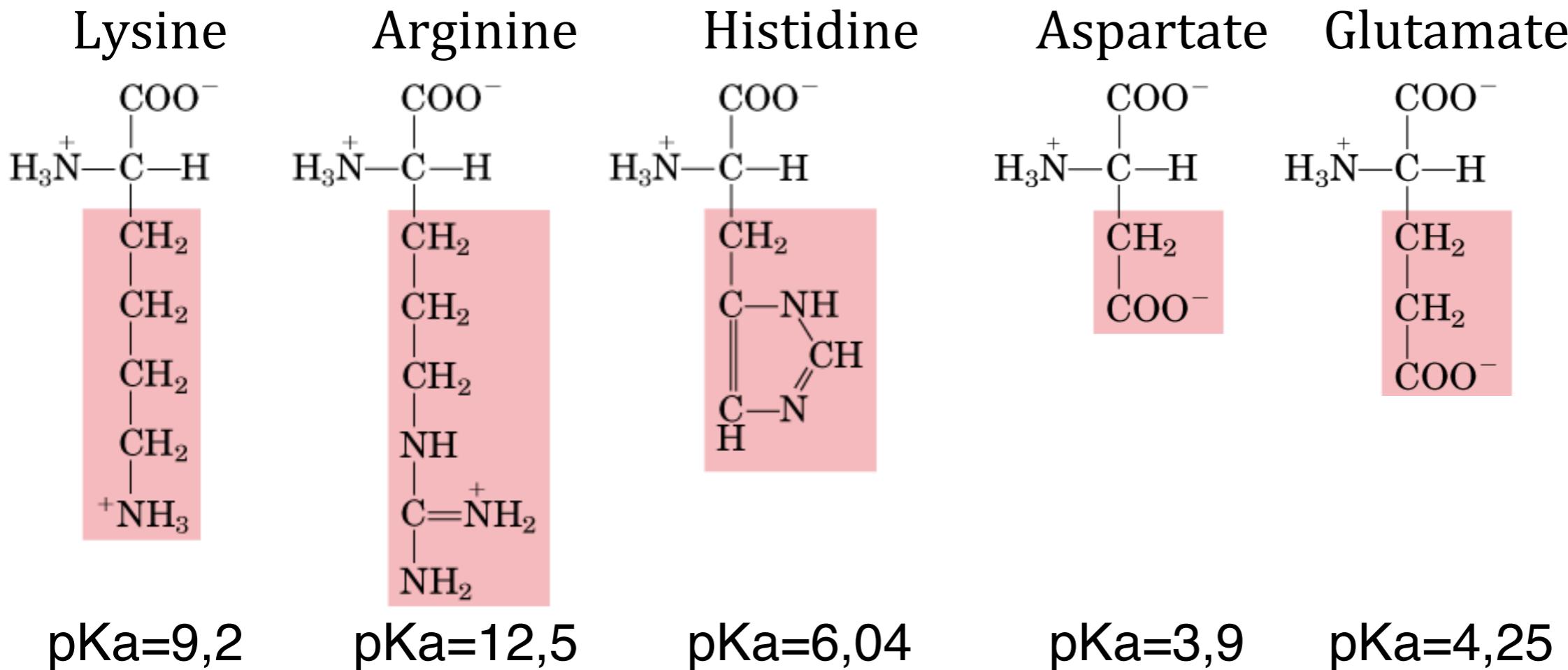
- electrostatic interactions;
- Van der Waals interactions;
- hydrogen bonds;
- cation- and amino- π ;
- π stacking
- hydrophobic effect.

The magnitude of non-covalent interactions is of the order of 5 kcal/mol [20 kJ/mol], whereas a covalent bond is about 90 to 100 kcal/mol [380-420 kJ/mol]. The stability of biomolecules is due to a large number of non-covalent interactions.



3.1. Electrostatic interactions

Among the 20 amino acids, some of them can present a negative or a positive charge, in function of the pH.



An electrostatic interaction can occur between these charged amino acids.

The electrostatic interaction can be large in magnitude, but it depends on the environment.

The electrostatic interaction can be large in magnitude, but it depends on the environment.

The force between two ions is equal to: $F = \frac{q_1 q_2}{\epsilon_r \epsilon_0 r^2}$

The interaction energy between two ions is equal to: $E = -\frac{q_1 q_2}{\epsilon_r \epsilon_0 r}$

$q_1 q_2$ are the charges, r is the distance between them, ϵ_r is the relative dielectric constant of the medium and ϵ_0 is the dielectric constant of vacuum.

solvant	ϵ
water	80
acetone	20,7
hexane	2,02

The dielectric constant decreases the intensity of the electrostatic interaction. This dielectric constant varies widely in a protein (it depends on the environment).

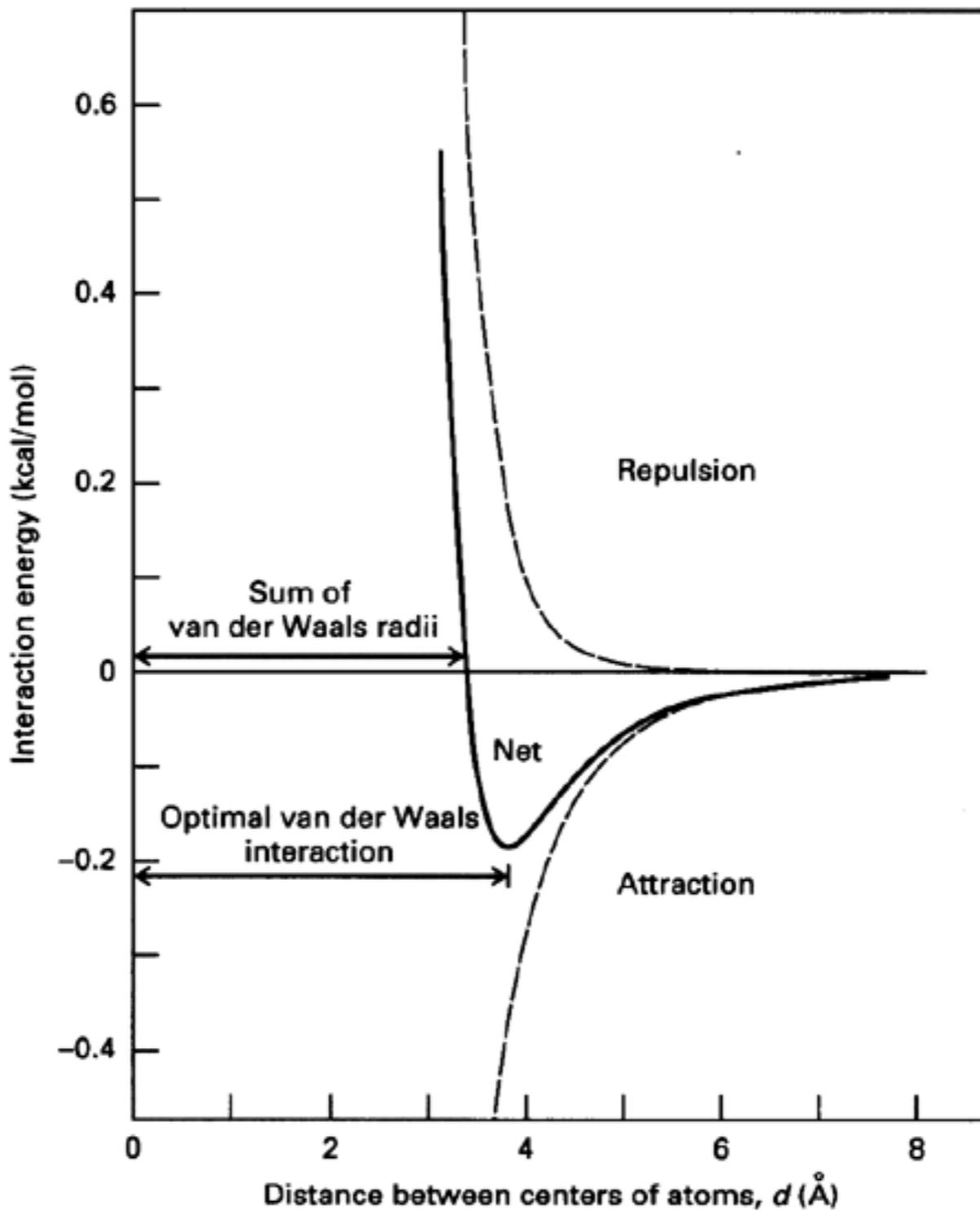
The salt concentration (ionic strength) will also affect the electrostatic interactions: a higher salt concentration will decrease the strength of the electrostatic interaction (the ions will compete with the charged amino acids).

3.2. Van der Waals interactions

There is an attraction due to induced polarization effects. These interactions are short-range and weak (they are inversely proportional to the sixth power of distance).

Van der Waals interactions present a short-range repulsive component and an induced dipole-induced dipole attractive component. The Lennard-Jones potential is often used as a model of the Van der Waals interaction as a function of distance.

Lennard-Jones potential

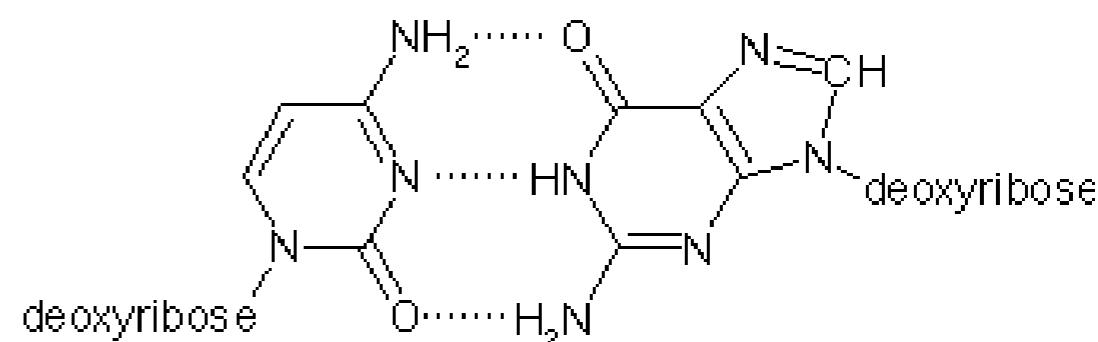
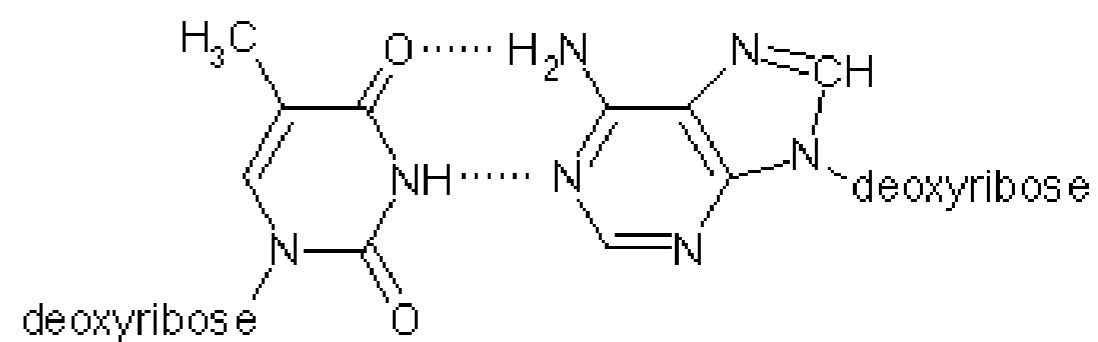


3.3. Hydrogen bond

An hydrogen bond corresponds to an interaction between an electronegative atom, such as oxygen or nitrogen, and an hydrogen atom binded to an electronegative atom.

Examples:

- H bond in water;
 - H bond between a C=O group and the N-H group of the main chain of a protein;
 - H bond between the C=O group and the O-H group the side chain of a serine or a threonine (proteins), of sugars;
 - H bond between nucleic bases of DNA.



There is a large number of H bonds in proteins, in their secondary structure (α -helices and β -sheets). They play an important role in the structure of proteins.

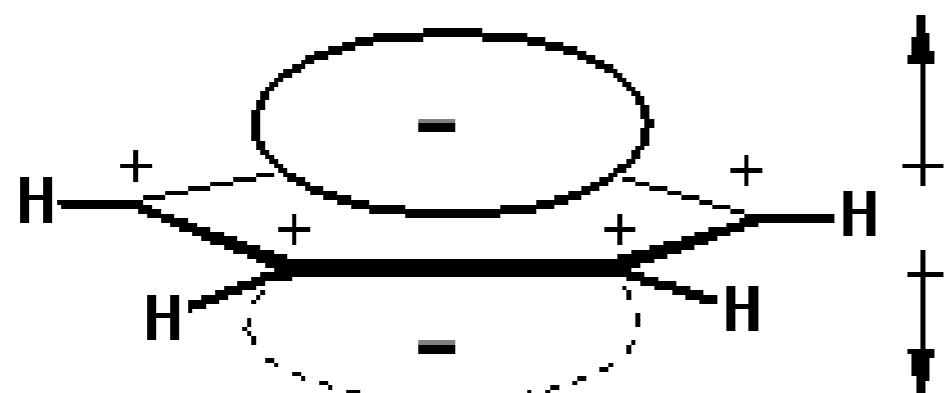
Is it a major factor in protein stability ?

Protein stability is defined as the free energy difference between the unfolded and the folded states. In the unfolded, the chemical groups that are able to make hydrogen bond do it with water. These H bonds are comparable to those that are formed within the folded protein. According to the enthalpy, the difference between both states is about zero. But the entropy change of the solvent must also be taken into account.

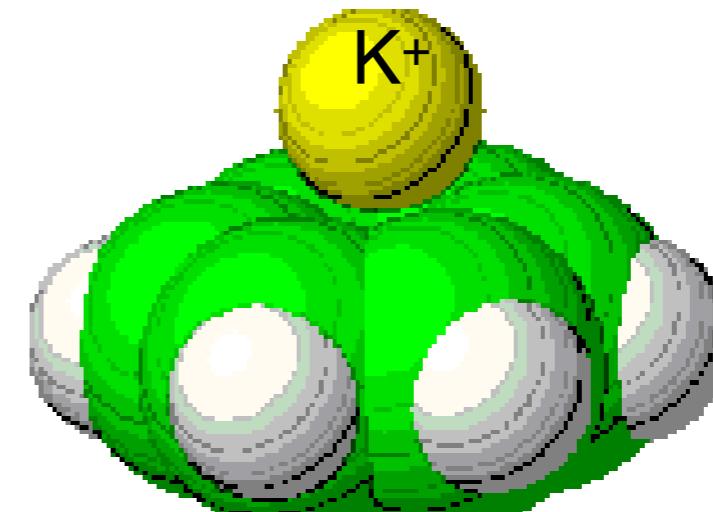
3.4. Amino- and cation- π

It is an interaction between a positive charge (or a positive partial charge) and delocalized π electrons that are present in an aromatic ring. The electrostatic force plays an important role, but induced dipole, polarisability, dispersion and charge transfer contributions are not negligible.

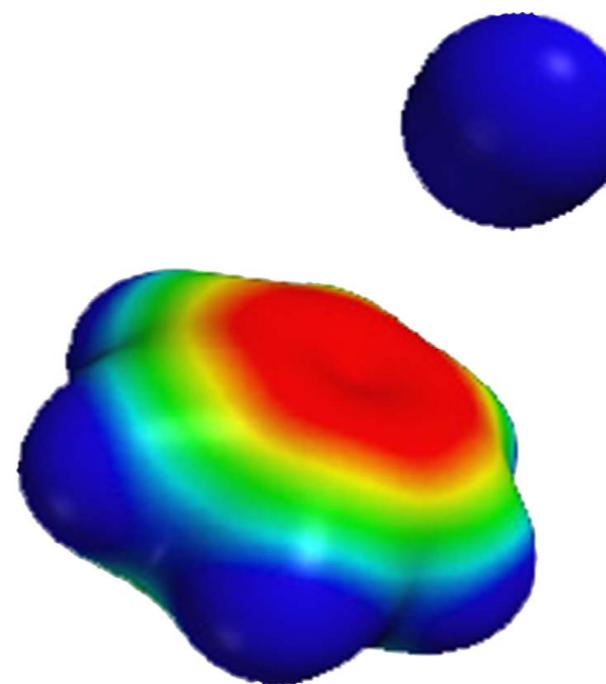
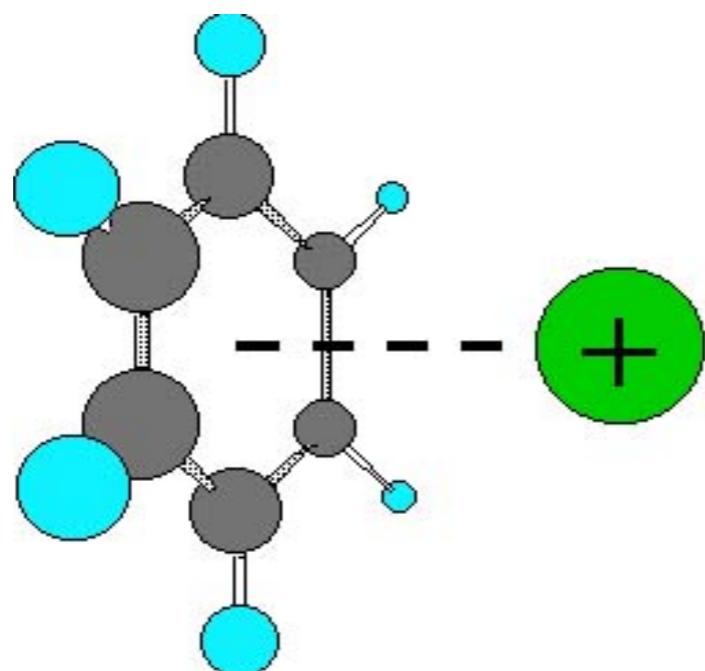
The charge dispersion in aromatic rings is shown below, as well as the quadrupole moment:



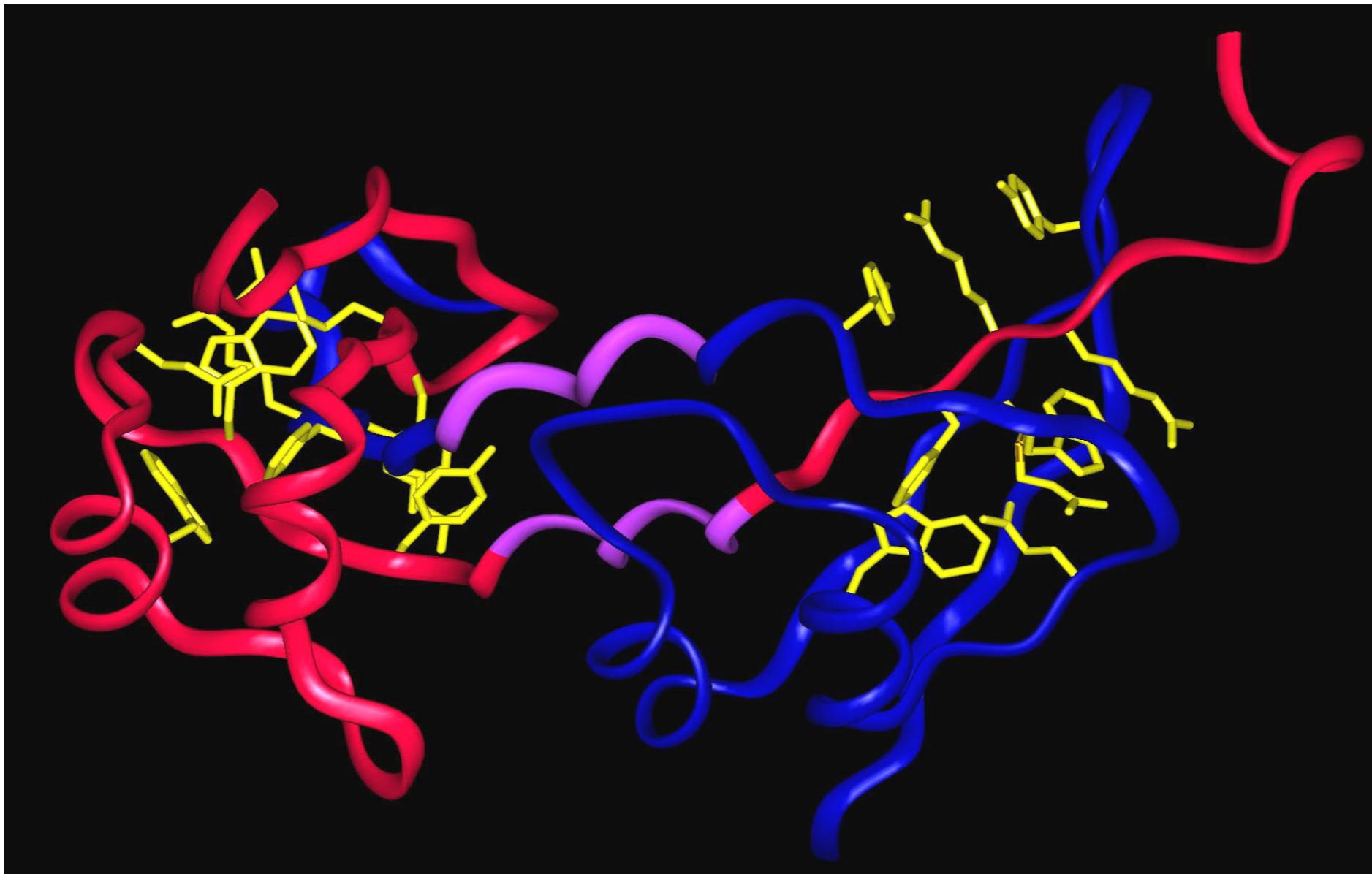
(a)



(b)



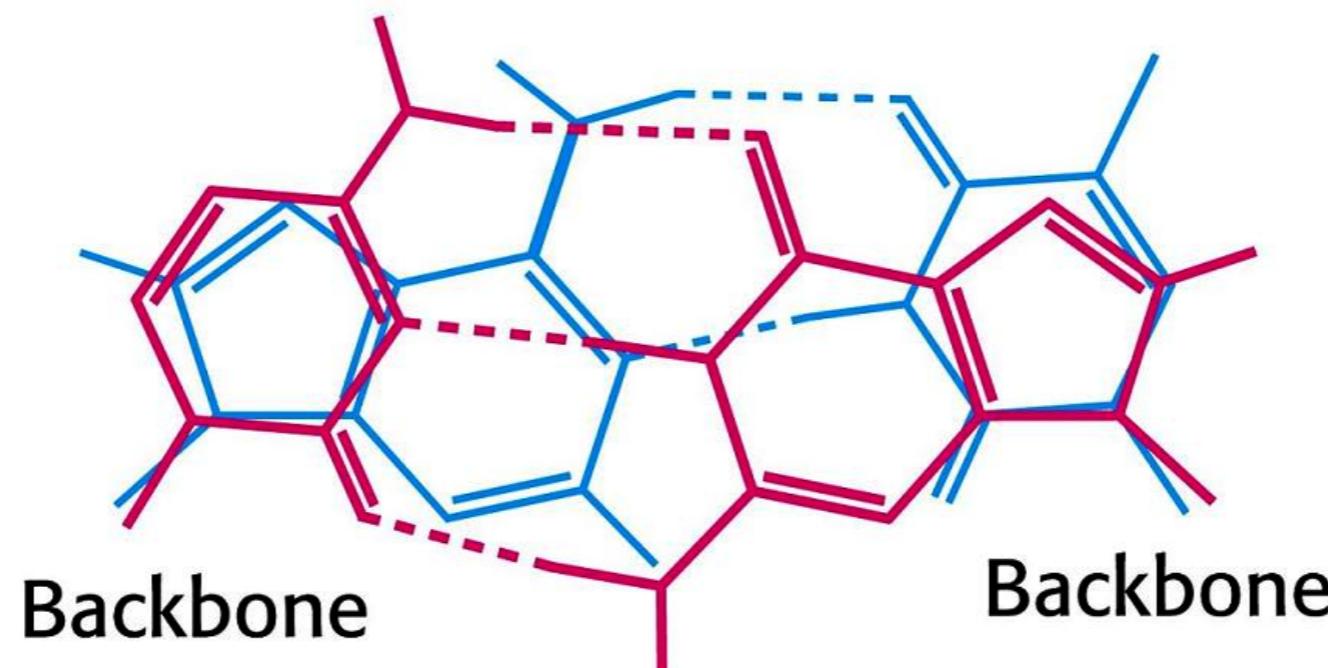
cations- π cluster in a protein (CksHs2)



3.5. π - π stacking

These interactions are found, among others, in DNA. They involve aromatic rings, where π electrons are delocalized.

A T geometry involves an electrostatic interaction. A parallel geometry involves an overlap between π orbitals of the aromatic rings.



3.6. Hydrophobic effect

This effect is important for protein folding. It results from the weak interaction between polar and non-polar molecules. Most of the non-polar residues tend to cluster in the core of the protein.

The thermodynamics of this effect is quite complex. This effect is an entropic consequence of the interactions between hydrophobic side chains and water.

Hydrophobic molecules are insoluble in water because they interfere with the structure of water (H bonds). A non-polar molecule creates a cavity that causes an increase in the order of water molecules around the cavity => decrease of the entropy of water (unfavorable).

