Silvia Onesti
silvia.onesti@elettra.eu

Elettra Sincrotrone Trieste

An introduction to Structural Biology
and its contribution to Biophysics

Biophysics: the beginning?

In 1943 Schrödinger gave a few lectures at Trinity College, Dublin, on "What is life: the physical aspects of the living cell"

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

What is life? what is special about living organisms?

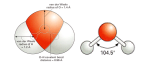
We know now that the distinction between life and non-life is not due to the molecules or the chemistry involved, but to the way these molecules are organised in a complex systems and are able to display a complex behaviour.

Biology is the coordinated interplay of large and complex macromolecules.

The molecules of life

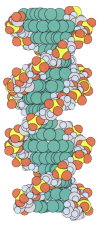
A living organism is made up by trillions and trillions of molecules.

The shape and spatial organization of the atoms within each molecule determine its properties and function.

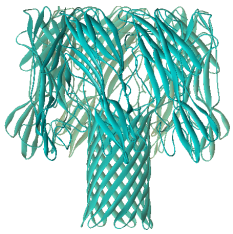


two different representation of a water molecules

phospholipid



DNA double helix



α-haemolysin, a bacterial toxin

Molecules in biology

From a functional point of view biological molecules can be divided into:

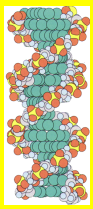
small molecules
- made and altered by individual steps of chemical reactions
- used as substrates for making macromolecules
- used to store and distribute energy for cell processes
- broken down to extract chemical energy
- used in signalling

macromolecules
- polymers made by a linear chain of building blocks
- made by linking a defined set of small molecules (monomers) through the repetitive use of a single chemical linkage
  - ↦ proteins: polymers made by a linear chain of amino acids
  - ↦ nucleic acids: polymers made by a linear chain of nucleotides

EVOLUTION of LIFE = EVOLUTION OF MACROMOLECULES

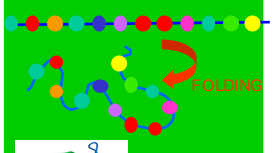
Macromolecules

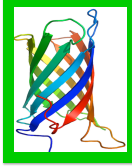
Nucleic acids



dsDNA

Proteins





Proteins:

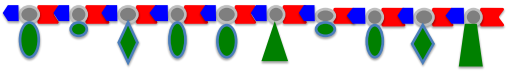
Proteins are *molecular machines* responsible for ALL the functions carried out within a cell:
- maintain cells' shape and function (structural role)
- control the synthesis and degradation of other molecules
- control their own synthesis and destruction
- allow cells to move and communicate
- control gene activity, in response to the environment
- direct the mechanisms that control inheritance and reproduction

Nucleic acids (DNA and RNA):

- contain a coded representation of all the proteins of the organisms
- contain a set of instructions that control when and in which quantity a protein must be produced.

Proteins and amino acids

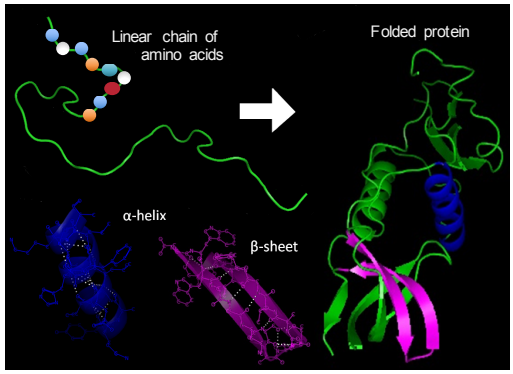
Amino acids are the building blocks that make up proteins: each amino acid has tow chemical groups (here red and blue) that bind to each other forming a chain, plus a ***variable lateral group***.



There are **20 different amino acids**, each with a different lateral group. Some are small, some are big, some have a positive or negative charge, some are hydrophilic, some are hydrophobic.

Each protein has a specific amino-acid sequence that dictate the protein shape and function.

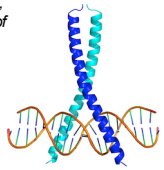
Protein folding



The shape of proteins

A protein is a linear chain of amino acids: the number and sequence of amino acids in each protein is different – there are small proteins (30 amino acids) and big proteins (30,000 amino acids!), globular proteins, elongated proteins, flexible proteins, rigid proteins…

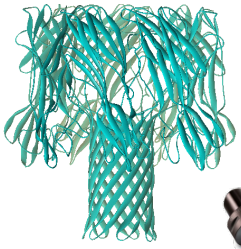
Collagen: makes tendons, bones, teeth and part of your skin



DNA binding protein that switches ON a gene

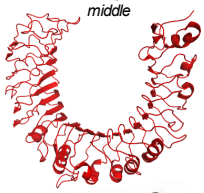
Protein involved in the control of cell proliferation – when too active it causes leukemia

The shape of proteins

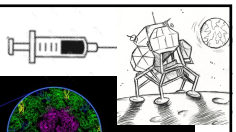


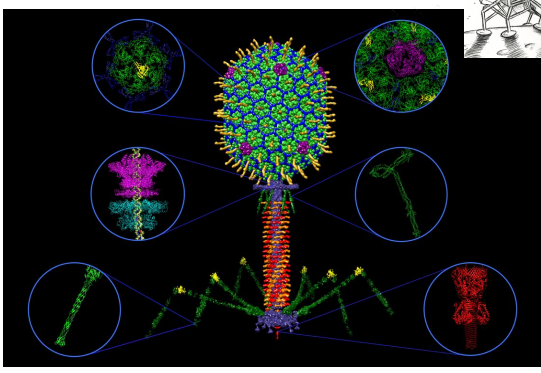
Bacterial toxin: it makes a hole in the cell membrane

Leucine-rich repeat: "grabs" other "things" in the middle



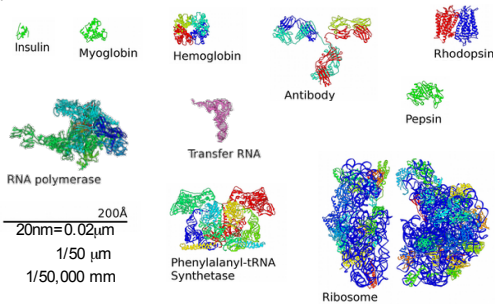
The shape of proteins





The size of macromolecules

There are small proteins (30 amino acids) and big proteins (30,000 amino acids!); some proteins assemble into complex machineries made up of many proteins and nucleic acids, such as the ribosome or a virus.



Why is structure important?

The structure of a protein dictates its function.

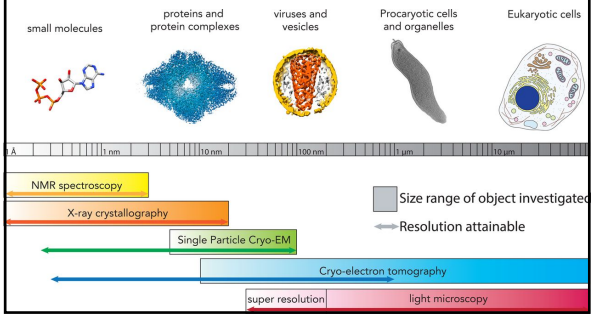
The function of a protein dictates its shape

Knowing the structure of a protein is thus essential to understand what the protein does.

- The *function* of a protein depends on its *structure*
- The *structure* depends on the specific *amino-acid sequence*
- The *amino-acid sequence* is controlled by the *DNA sequence of the gene* that codes for the protein.

The scale of living things

Here is a diagram of the sizes of various biological objects (cells and their components) drawn on a logarithmic scale, indicating the range of objects that can be visualised with different techniques.



An historical prospective

- 1934 first diffraction pattern from a protein
- 1953 double helical structure of DNA by fiber diffraction
- 1960s atomic structures of myoglobin, haemoglobin, lysozyme (an enzyme) by protein crystallography

Late 1980s

- 2-D NMR used to determine structures of small proteins
- structure determination of membrane proteins by crystallography
- structure determination of hundreds of proteins and nucleic acids

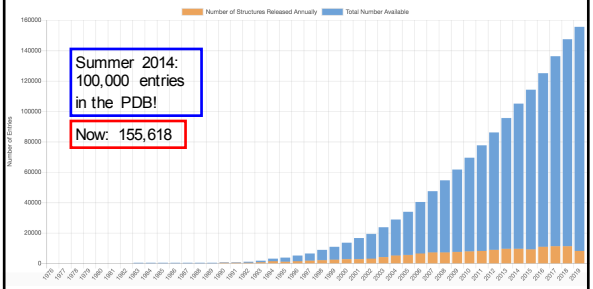
Late 1990s

- single particle cryo electron-microscopy - low resolution images of large complexes (eg. ribosome)
- structures of large assemblies at atomic resolution by PX (large viruses, the nucleosome, the ribosome, RNAPs, etc..)

2010s

- EM revolution – structures of single molecules to atomic resolution
- Development of cryo-

The Protein Data Bank (PDB)

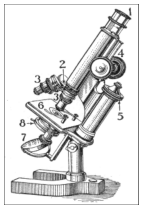


The number of atomic structures deposited in the PDB is growing exponentially.

Why can't we use a microscope?

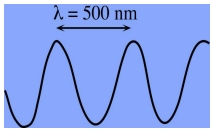
Normally, to look at small objects we use microscopes…

…but they can only provide images of things larger than the wavelength of light.



Protein (10 nm)

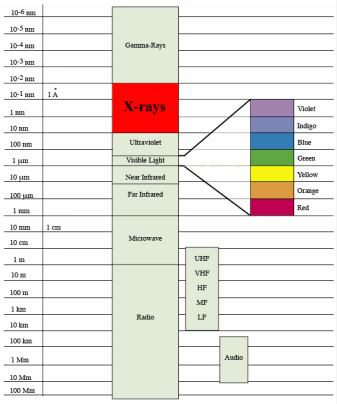
Atom (0.1 nm=1Å)



We could use **X-rays** (light of λ=0.1 nm, the **right size** for looking at atoms)...

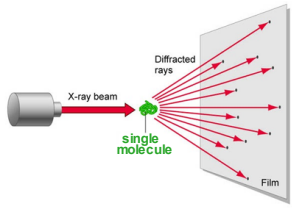
X-rays

X-rays:
E = 1-100 keV
λ = 10 nM-0.01 nm



…but we can't make X-ray microscopes as hard X-rays do not reflect or refract easily.

X-ray diffraction with single objects?

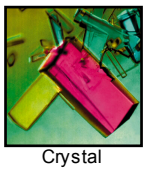


The intensity of the X-ray radiation diffracted by a single biological object (such as a protein or a cell) is very small.

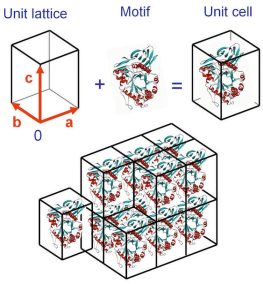
Even using very bright sources of X-rays (synchrotrons) it is impossible to measure .

New generation light sources called free electron lasers (FELs) produce even brighter light beams. FELs *may* be able to produce such intense X-ray beams that allow to measure the diffraction pattern from a single protein.

But we can use X-rays with crystals



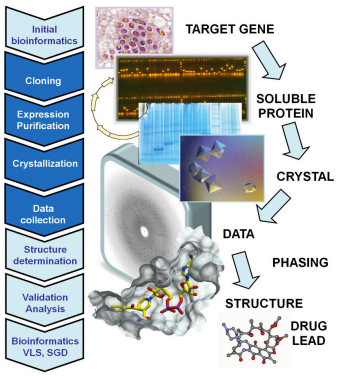
Crystal



Crystal: an ordered array of molecules

Constructive interference of the diffracted X-rays, so that all molecules scatter in phase and the signal adds up:
strong measurable signal.

The crystallographic pipeline



- no protein, no party!
  a lot of your time will be spent in cloning, protein expression and purification
- **no crystal, no party!**
  crystallization is the real bottleneck!
- once you get crystals, the fun begins:
  - data collection & processing
  - solving the phase problem
  - model building & refinement

MX: pros & cons

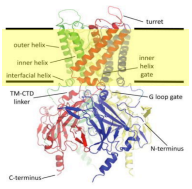
The application of X-ray diffraction to determine the 3D structure of biological molecules has been one of the most remarkable successes at the interface of physics and biology .

Since the early '60s, X-ray diffraction techniques have provided many thousands of crystal structures which have been instrumental in our understanding of biological processes at the atomic level, making crystallography the key to some of the major discoveries in modern biology

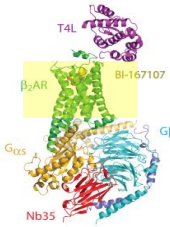
HOWEVER...

- MX relies on obtaining well ordered crystals
- to obtain crystals one needs large amounts of very pure protein, which can be difficult to achieve, especially with eukaryotic proteins subjected to post-translational modification or with membrane proteins

Membrane proteins

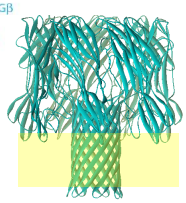


K+ channel



GPCR receptor

Hydrophobic residues on the outside, facing membrane lipids



Bacteriolysin

Part protruding from membrane can be a very short sequence (a few amino acids), a loop, or large, independently folding domains.

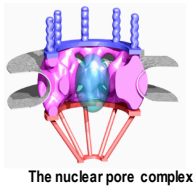
Beyond macromolecular crystallography?

However, as our knowledge of biology increases, so does the complexity of the problems we wish to tackle. Many important cellular processes are dominated by macromolecular complexes that are large, often polymorphic, with unstructured regions, and exist in a variety of functional states, making them a challenging target for crystallisation.

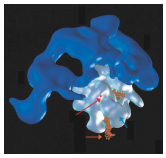
We now feel the need to go beyond macromolecular crystallography and to use an integrated approach by combining various structural biology techniques, such as NMR, MX, EM, SAXS – as well as many other biophysical and biochemical approaches.

Beyond macromolecular crystallography?

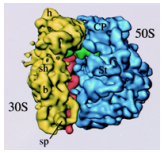
The textbook picture of the cell like a "soup", where protein and enzymes swims freely in the cytoplasm, has changed dramatically over the last 10-20 years. We now know that the interior of cells is very crowded and structurally organised into "territories", with many large macromolecular assemblies responsible for key biological processes.



The nuclear pore complex



RNAPII+Mediator



The 70S ribosome

Beyond macromolecular crystallography?

Many important cellular processes are dominated by macromolecular complexes that are large, often polymorphic, with unstructured regions, and exist in a variety of functional states, making them a challenging target for crystallisation.

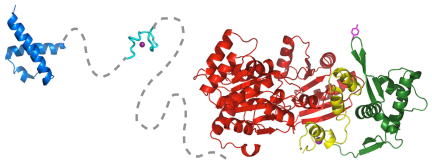
There is therefore the need for alternative methods that are able to cope with the size and complexity of large macromolecular systems, albeit at lower resolution:

- Single-particle (cryo) electron microscopy
- Small-angle X-ray scattering
- EM tomography

Data from these techniques can be combined with atomic structures from homologous proteins or fragments of the complex, to get a 3D picture of the complex architecture.

Beyond macromolecular crystallography?

Beside the size, we are more and more confronted with proteins that contain unstructured regions, exhibit polymorphism, have multi-domain architectures: these are also challenging target for crystallisation.



Here again we need alternative methods that are able to cope with flexibility and polymorphism

- Nuclear Magnetic resonance
- Small-angle X-ray scattering
- Single-particle cryo-electron microscopy

Classification based on type of radiation

Using X-rays

- Macromolecular crystallography (MX)
- Small angle X-ray scattering (SAXS)
- Fiber diffraction

Using neutrons

- Neutron crystallography
- Small angle neutron scattering (SANS)

Using electrons

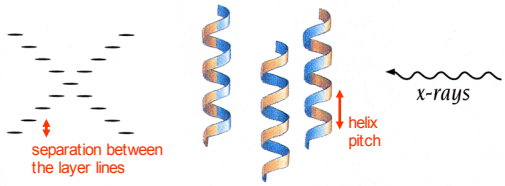
- Electron diffraction
- Scanning electron microscopy (SEM)
- Transmission electron microscopy (TEM) – single particle (cryo)EM

Nuclear magnetic resonance (NMR)

X-ray fiber diffraction

DNA fibers or fibrous proteins tend to pack side by side in an ordered manner, with the axis of their fibers lining up in a parallel fashion.

The polymers in fibers typically assume helical structures; this gives a characteristic cross-like variation in the diffraction pattern.



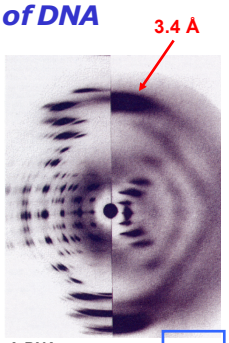
The separation between the layer lines is determined by the helix pitch: as the helix pitch increases the layer lines move closer together.

X-ray fiber diffraction: the structure of DNA

Diffraction pattern from two forms of DNA: A-DNA shows a higher degree of order in the way the fibers pack together (more 'spots').

The pitch of the A-form is 28 Å while for the B form is 34 Å (smaller spacing). Knowing the helix pitch, it is possible determine an approximate value for the radius of the helix from the inclination of the arms of the cross.

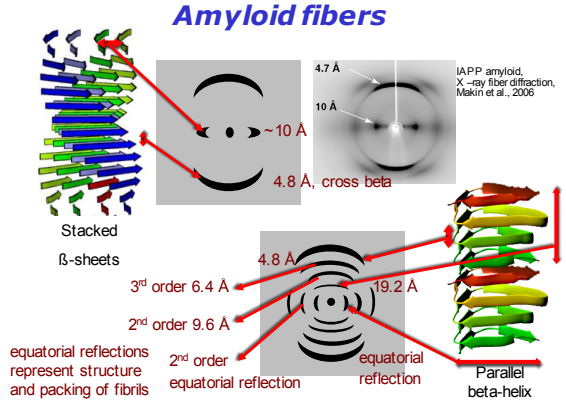
Arguments such as these, based on the X-ray fiber diffraction patterns obtained by M. Wilkins and R. Franklin, were used by J. Watson and F. Crick in the construction of the double-helical model for DNA



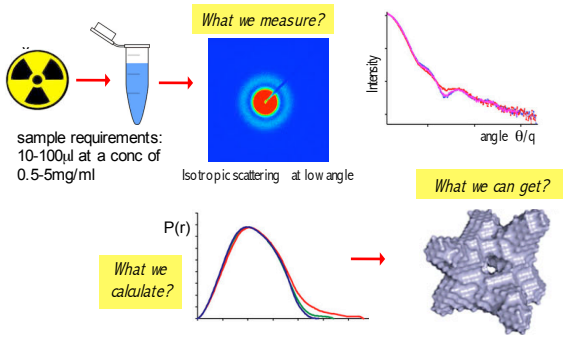
A-DNA

B-DNA

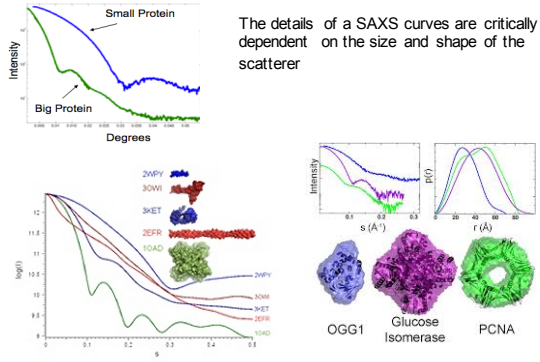
Amyloid fibers



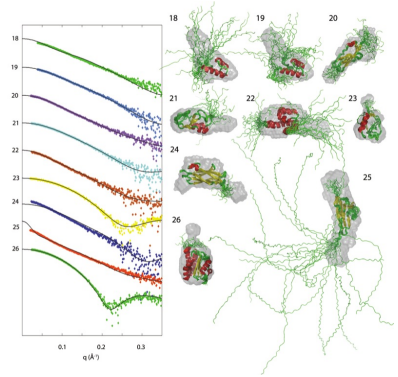
Small angle X-rays scattering (SAXS)



SAXS & molecular shapes



SAXS & disordered regions

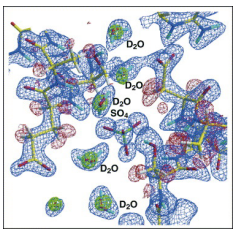


Neutron vs X-rays

- Neutrons interact with the *nucleus* of the atom, so the contribution to the diffracted intensity is different for each isotope (H and D)does not increase with the atomic number (light atoms matter).
- Neutrons interact with matter to a much lesser degree than X-rays: less damage but less signal

Neutron crystallography requires very big crystals, but provides enhanced visibility of hydrogen atoms: can help to determine of the protonation state of proteins, ordered water molecules and inhibitors.

Small angle neutron scattering (SANS) can help discriminate between protein/nucleic acids/lipids, etc.. (contrast variation/matching)



Electrons vs X-rays?

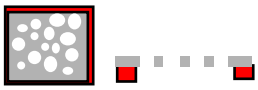
Most important difference between the two:

- the scattering cross-section is about 100x greater for electrons than for X-rays: significant scattering using electrons is obtained for specimens that are 1 to 10 nm thick whereas scattering or absorption of a similar fraction of an illuminating X-ray beam requires crystals that are 1-10 µm thick.
- electrons are much more easily focused than X-rays since they are charged particles that can be deflected by electric or magnetic fields: electron lenses are much superior to X-ray lenses; this allows the electron microscope to be switched back and forth instantly between imaging and diffraction modes.

Electrons-based techniques

- Electron crystallography (2D crystals)
- Fiber electron diffraction (helical fibers)
- Scanning electron microscopy (surfaces)
- Trasmission electron microscopy:
  - cells, organelles
  - single particles CryoEM
- Electron tomography (cells, subcellular structures, large macromolecular complexes)

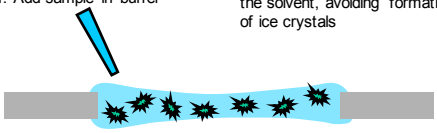
Single particle cryo-electron microscopy



Support for sample: "holey" carbon film

1. Add sample in buffer

2. Plunge in liquid ethane (-160 deg), so as to "vitrify" the solvent, avoiding formation of ice crystals

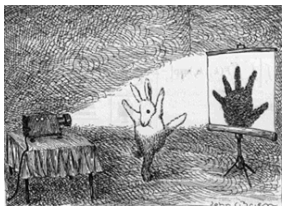


Single particle cryo-electron microscopy

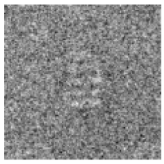
- image is 2D projection of original 3D object: 3D structure can be determined from a set of views at different orientations
- radiation damage is the ultimate limit on resolution – to avoid destroying the sample, one uses very low doses, obtaining very noisy images

We have two problems

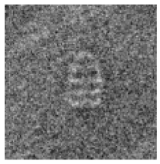
- getting a signal out of very noisy images
- getting a 3D structure out of 2D projections



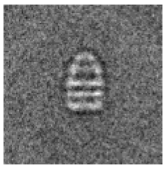
Single particle cryo-electron microscopy



Sum of 4 images



Sum of 8 images



Sum of 32 images

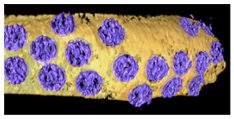
- ***Averaging*** large number of particles massively increases the signal/noise (BUT we have to make sure we are averaging views with the same orientation!)
- I can do the same ***for all the possible orientations*** of the molecule, and putting them together to get a 3D reconstruction.
- ***New detectors*** have allowed to reach atomic resolution from single molecules larger than 100-200 kDa

CryoET: the new frontier?

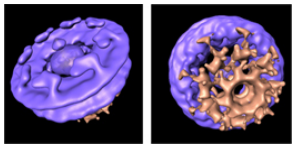
Standard single-particle Cryo-EM requires the purification of proteins and complexes, which is often difficult and/or detrimental to their structure or function.

Cryo-ET can provide three-dimensional insights into the unperturbed organization of tissues, cells and viruses in their native environment.

An example: the nuclear pore complex (NPC)

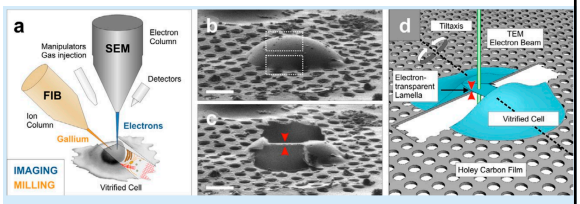


Surface rendered representation of a segment of nuclear envelope (NPCs in violet, membranes in yellow).



Structure of the NPC after averaging of subtomograms.

CryoET on cells/tissues

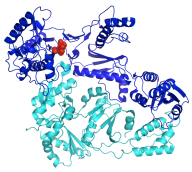


Focused ion beam (FIB) milling.
In this method, sample thinning is achieved by sputtering the specimen surface with gallium ions (panel a). Vitreous lamellae, cut at a shallow angle to the supporting grid and extending over tens of micrometers (panels c and d) were recently prepared from eukaryotic cells by cryo-FIB milling (Rigort et al., 2012a,c). Current development focuses on improving the reliability of the complete FIB-milling workflow and developing precise 3D targeting of low-copy number structures within large cellular volumes by correlative microscopy.

Structural Biology and Drug discovery

Most drugs target and inhibit proteins: once a **lead compound** (i.e. a ligand) is known, medicinal chemists modify it to increase the affinity, delivery, off-target effects, etc. --> Drug optimization.

Knowing the structure of a protein and especially the complex between the protein and a potential drug can speed up the process by suggesting the optimal modifications and reducing the number of attempts leading to a better drug.



In any case knowing the structure of a drug bound to its viral, bacterial or tumor target help us to understand the mechanism of action and fight resistance.

*This is a protein from the **HIV virus (HIV reverse transcriptase)** bound to an inhibitor (**Nevirapine**) that is a drugs against AIDS*

Structures can be VERY useful in biology!

However, remember that:

To FULLY UNDERSTAND the behaviour of a protein in the cell one needs to integrate data from cell biology, biochemistry, molecular biology, bioinformatics, various structural biology techniques, computational biology, biophysical chemistry, nanobiophysics, pharmacology, single molecule studies.

DNA replication: THE MOVIE!

Drew Barry – The Walter and Elizabet Institute of Medical Research, Melbourne