ICTP School on Synchrotron Light Sources and their Application

17th January 2025

Macromolecular crystallography & Structural Biology for medicine and drug design

Elettra Sincrotrone Trieste

Silvia Onesti
Elettra – Sincrotrone Trieste

1

"Seeing is believing"

Humans are ***highly visual creatures***. We use our eyes to make sense of the world.

We rely heavily on visual cues to find food, mates, shelter, requiring a *highly sophisticated processing of visual information*. Consequently, the number, size, and complexity of brain structures involved in visual processing has expanded dramatically and our brains have evolved to absorb, manipulate, and react to visual information in increasingly effective ways.

- 50 percent of our brain is involved in visual processing
- although we have five senses, the majority of the sensory feedbacks are visual. Our eyes are so good (and so much better than the rest of our senses) that we can grasp a visual scene in less than 1/10 of a second.
- we recall what 'we see' much better than what we read

We often think we "understand" something when we "see" it!

2

"Just look at the thing!"

Plenty of Room at the Bottom

Richard P. Feynman
(Dated: Dec. 1959)

What are the most central and fundamental problems of biology today? They are questions like: What is the sequence of bases in the DNA? What happens when you have a mutation? How is the base order in the DNA connected to the order of amino acids in the protein? What is the structure of the RNA; is it single-chain or double-chain, and how is it related in its order of bases to the DNA? What is the organization of the microsomes? How are proteins synthesized? Where does the RNA go? How does it sit? Where do the proteins sit? Where do the amino acids go in? In photosynthesis, where is the chlorophyll; how is it arranged; where are the carotenoids involved in this thing? What is the system of the conversion of light into chemical energy?

It is very easy to answer many of these fundamental biological questions; you just look at the thing! You will see the order of bases in the chain; you will see the structure

Transcript of a talk presented by Richard Feynman to the American Physical Society in Pasadena on Dec 1959

The central idea behind structural biology is that once we are able to "look" at "things" in great enough detail to discern the atomic structures, we will be able to understand the components of complex biological processes and their mechanism of action.

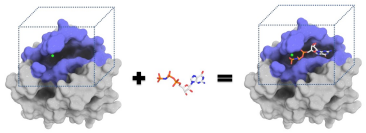
Indeed, structural biology has contributed substantially to major biological discoveries, over the last 70 years.

Moreover it has facilitated developments of therapeutic agents to cure diseases.

3

Why Structural Biology?

- Provide a framework for rational structure-based drug design:
  - virtual screening
  - optimisation of lead compounds
  - fragment-based drug design



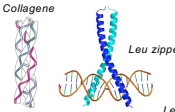
A protein kinase bound to Glivec, a drug against leucemia

4

Why Structural Biology?

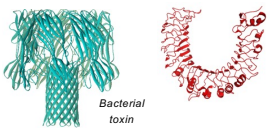
- Understanding basic biological processes at the molecular/chemical level

Collagene

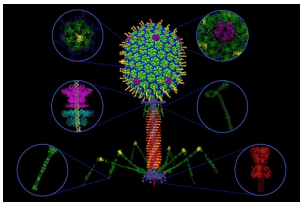


Leu zipper

Leucine-rich repeat



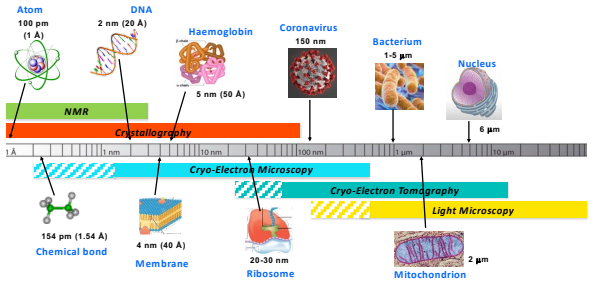
Bacterial toxin



BACTERIOPHAGE

5

The scale of life



6

An historical perspective

- 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 macromolecular crystallography (MX)

Then exponential growth of the number of protein structures...

late 1980s
- 2-D NMR used to determine structures of small proteins
- structure determination of membrane proteins & large complexes by MX

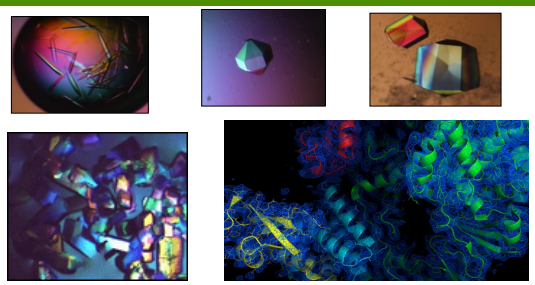
late 1990s
- single particle CryoEM used to determine low resolution images of large complexes (eg. ribosome)
- structures of large assemblies at atomic resolution by MX (the nucleosome, the ribosome, large viruses, RNAPs, etc..)

2010s
- EM revolution – structures of single molecules to atomic resolution

2020s
- AI-based predictions

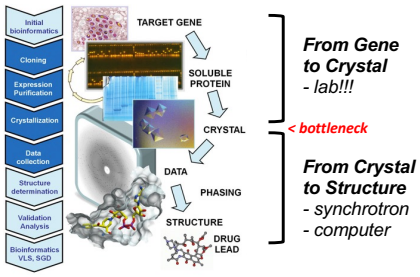
7

Macromolecular crystallography (MX)



8

The MX pipeline



From Gene to Crystal
- lab!!!

< bottleneck

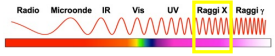
From Crystal to Structure
- synchrotron
- computer

9

How can we see proteins?

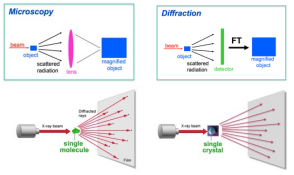
Normally, to look at small objects we use microscopes... but they can only provide images of things larger than the wavelength of light.

To get atomic resolution I need a light with a l close to atomic dimensions/bonds (0.1nm = 1 Å)



X-ray microscopy? **two problems**:

1. Microscopes use LENSES
   The quality of X-ray lenses is not suitable for high resolution images. I need to collect the scattered radiation on a detector and use an FT to get a magnified image (DIFFRACTION)

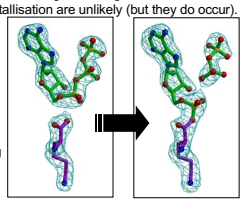


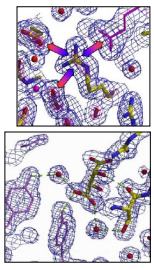
2. The signal I can obtain from a single protein is too small – I need to amplify it (CRYSTAL)

10

Does crystallisation affects the structure?

- The forces that hold molecules in a lattice are very much weaker than those that hold protein structures together, so gross conformational changes triggered by crystallisation are unlikely (but they do occur).
- Proteins crystallised in different crystal forms are often identical or almost identical.
- Some enzymes retain their biological activity within the crystal (a strong indication of a native-like structure).



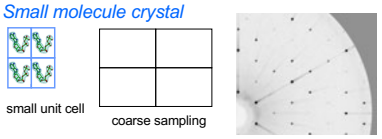


However the crystal lattice may favour a conformation that is not the dominant shape in solution.

11

MX vs small molecule crystallography

Small molecule crystal

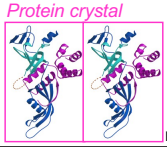


small unit cell

coarse sampling

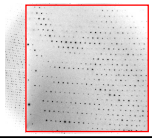
- a few, strong reflections
- diffract to high resolution

Protein crystal

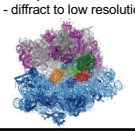


fine sampling

large unit cell

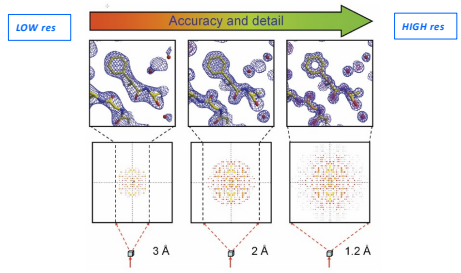


- many, weak reflections
- diffract to low resolution



12

Resolution in MX



13

The phase problem in MX

For small molecule I have few hundreds strong reflections to phase, and crystals are ordered and tend to diffract to very high resolution (close to the theoretical limit). Need to guess a small set of phases.

◆ Direct methods
1985 Chemistry Nobel prize: H. Hauptmann & J. Karle (+ Isabelle Karle?)

Apply physical constraints (such as "positivity", "atomicity", etc.) to the electron density to limit the phases to one or a few possible self-consistent sets (a sort of "good guess").

Direct methods are very powerful in solving small molecules, but cannot be used for proteins:
- huge number of reflections (a very large set of phases to guess)
- low resolution (atomicity does not hold)

14

The phase problem in MX

MIR (multiple isomorphous replacement)

Older method (Cambridge, 60') – relies on binding "heavy" atoms to the crystal and compare he diffraction pattern to the native. Trial and error search for good heavy atoms, it may take longer to get it right

MR (molecular replacement)

Older method (Cambridge, 70'-80') – relies on the expected similarity between the protein and another whose structure is known. Cannot solve de novo structures. Requires high homology (30% sequence identity?)

MAD (multiwavelength anomalous dispersion)

Relies on the absorption of specific wavelengths due to electronic transitions within the atom core. Similar to MIR but generally far quicker and more accurate. Requires high specification synchrotron radiation.

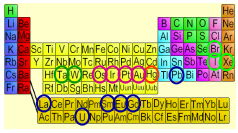
15

The phase problem in MX: MIR

Multiple Isomorphous Replacement method (MIR)

In 1953, Max Perutz (1914-2002) showed that the diffracted X-rays from protein crystals could be phased by comparing the patterns from crystals of the protein with and without ***heavy atoms*** attached.

In 1959, he employed this method to determine the molecular structure of hemoglobin. This work resulted in his sharing with John Kendrew the 1962 Nobel Prize for Chemistry.



Introduce "heavy" atoms at a relatively small number of sites on the protein molecules

Collect data from crystals of the protein alone, and crystals of the protein:heavy-atom complex: comparing the two data set helps to put constraints on the phases.

Finding heavy atoms that bind but do not disrupt the protein structure and the crystal (isomorphism) is done by trial and error

16

The phase problem in MX: MIR

The real breakthrough in using MIR for the determination of protein structures came when people learned how to deal with errors.

Blow D.M. & Crick F.H.C. (1959) "The treatment of errors in the isomorphous replacement method". *Acta Crystallogr.* **12**, 794-802

David Blow (1931-2004)

Francis Crick (1916-2004)

Error treatment is very complex – but it was an essential step in solving protein structures.

17

The phase problem in MX: MR

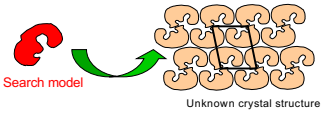
Molecular Replacement method (MR)

Tomato Bushy Stunt Virus

Human Rhinovirus 14

Michael Rossmann (1930–2019) used the structure of Tomato Bushy Stunt Virus, a plant virus, to determine the crystal structure of the Human Rhinovirus 14 (the common cold virus)
- in the early 80s.

In MR a known structure that is similar to the unknown structure as a search model. Initial phases are calculated from the correctly oriented and positioned model, as an approximation to the real phases.

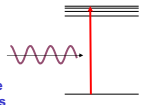


18

The phase problem in MX: MAD

Multiwavelength Anomalous Dispersion (MAD)

If the energy of photons in an incident X-ray beam is close to that of the binding energy of electrons, there will be small differences that will break the symmetry of the diffraction pattern.



Key point: Close to the absorption edge, small changes in λ give rise to significant changes in the phase of the scattered wave (anomalous dispersion)

Using MAD, we can collect multiple "derivatives" from the same crystal just by changing the X-ray wavelength very slightly.
>>> Need tunable X-rays sources with small bandwidth.

But at wavelengths of around 1 Å there are no electronic transitions for the "light" atoms (C, N, H, O, S, P) of biological molecules. Protein molecules can contain metals such as Zn, Fe and Cu which have accessible absorption edges – as an alternative I can replace Met residues by SeMet (Selenium has an absorption edge at 0.98 Å).

19

X-ray synchrotron data collectiom

Laboratory source

Single wavelength (Copper K_α λ = 1.5418Å)

Low intensity → long exposure times, radiation damage

Synchrotron sources

Variable wavelength → allow phasing using MAD/SAD

High intensity → short exposure times & less damage per photon

Most protein crystals are small and diffract poorly. The intense X-rays from a synchrotron has revolutionized structural biology and allowed to see structures of very large and complex proteins, including proteins of medical and pharmacological interest.

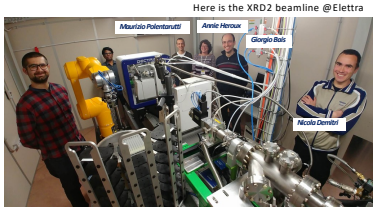
20

Synchrotron remote data collectiom

Most MX synchrotron beamlines are highly automated and offer remote data collection to both expert users and novices.

Well designed pipelines allow students, postdocs and PIs to send Dewars with frozen crystals and to collect data online, evaluating the quality of the diffraction, processing the data on-the-fly, and taking decisions as if they were at the synchrotron beamline.

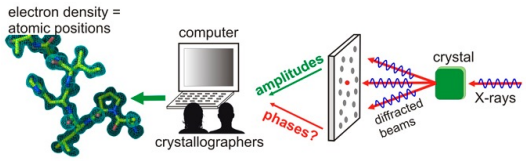
Here is the XRD2 beamline @Elettra



21

The MX pipeline: maps & models

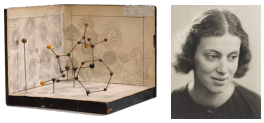
Once we have decent experimental estimates of the **phases** (from MIR, MR or MAD) we can calculate a Fourier Transform and obtain an electron density map, and from this get a 3D atomic model of our macromolecules.



22

The MX pipeline: maps & models

Model building in the early days…



Dorothy Crowfoot Hodgkin (1910-1994)

- structure of penicillin (1945)
- structure of vitamin B12 (1956)
- structure of insulin(1969)

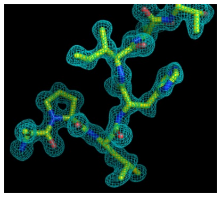
Chemistry Nobel prize 1964

John Kendrew (1917-1997) building the structure of myoglobin (Chemistry Nobel prize in 1962, with max Perutz)

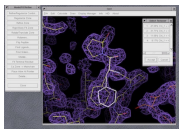
23

The MX pipeline: maps & models

We now use interactive computer graphics programs to build and analyse the atomic model of the protein.



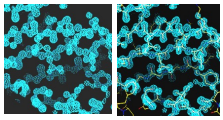
The programs allow display of maps and model in 3D, and permits full adjustment of the model (*e.g.* add/delete residues, tools for building from scratch, change residue type, change torsion angles, rotate side-chains, re-positioning of individual atoms or fragments, local stereochemical refinement,...)



24

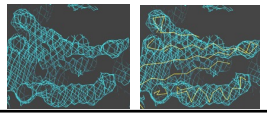
The MX pipeline: maps & models

The task of model building is to interpret the electron density maps in light of chemical knowledge, basic stereochemistry, chemical sequence, etc...
The level of interpretation depends on the resolution of the map:



Here is a 1 Å map

At very high resolution, individual atoms can be seen and fitted in the electron density blobs: the problem therefore is reduced to 'join-the-dots'

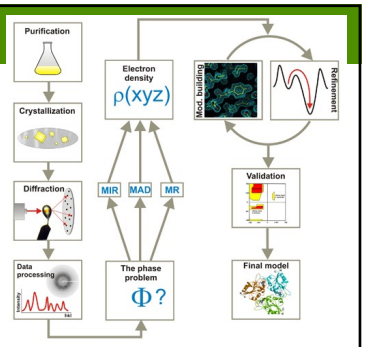


Here is a 6 Å map

At very low resolution only large features can be seen - for example helices look like rods and β-sheets can barely be detected.

25

The MX pipeline

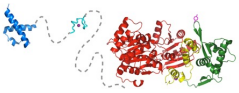


26

Beyond protein crystallography?

The problem

- Crystallographers have been so successful that they started to tackle larger and more complex structures
- Many structures have multiple domains connected by flexible links
- The textbook picture of the cell like a "soup", where protein and enzymes swims freely in the cytoplasm, has changed dramatically: cells are very crowded and macromolecules interact with each other in complex and dynamic ways



We need to integrate multiple approaches (MX, SAXS, CryoEM, NMR, Computational approaches, single molecule studies, etc...)

27

"Make the microscope 100x more powerful..."

Plenty of Room at the Bottom

Richard P. Feynman
(Dated: Dec. 1959)

We have friends in other fields—in biology, for instance. We physicists often look at them and say, "You know the reason you fellows are making so little progress?" (Actually I don't know any field where they are making more rapid progress than they are in biology today.) "You should use more mathematics, like we do." They could answer us—but they're polite, so I'll answer for them: "What you should do in order for us to make more rapid progress is to make the electron microscope 100 times better."

[...]

of the microsome. Unfortunately, the present microscope sees at a scale which is just a bit too crude. Make the microscope one hundred times more powerful, and many problems of biology would be made very much easier. I

The Nobel Prize in Chemistry 2017 Press Release

Cool microscope technology revolutionises biochemistry

We may soon have detailed images of life's complex machineries in atomic resolution. The Nobel Prize in Chemistry 2017 is awarded to **Jacques Dubochet, Joachim Frank** and **Richard Henderson** for the development of cryo-electron microscopy, which both simplifies and improves the imaging of biomolecules. This method has moved biochemistry into a new era.

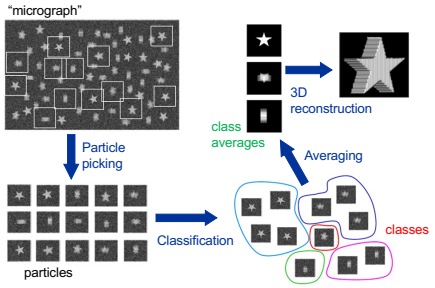
A picture is a key to understanding. Scientific breakthroughs often build upon the successful visualisation of objects invisible to the human eye. However, biochemical maps have long been filled with blank spaces because the available technology has had difficulty generating images of much of life's molecular machinery. Cryo-electron microscopy changes all of this. Researchers can now freeze biomolecules mid-movement and visualise processes they have never previously seen, which is decisive for both the basic understanding of life's chemistry and for the development of pharmaceuticals.

28

Single particle Electron Microscopy

TWO PROBLEMS

- low doses: getting a signal from very noisy images
- getting a 3D structure from 2D projections

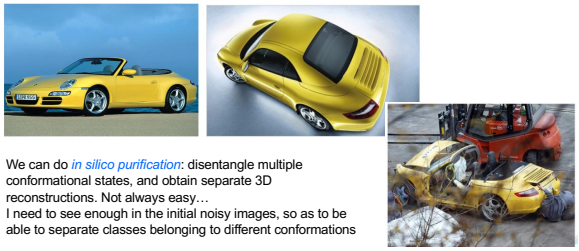


Slide from Helen Saibil

29

Polymorphic samples?

The sample may be present in multiple conformation and/or in partially unfolded forms.



We can do *in silico purification*: disentangle multiple conformational states, and obtain separate 3D reconstructions. Not always easy…
I need to see enough in the initial noisy images, so as to be able to separate classes belonging to different conformations

30

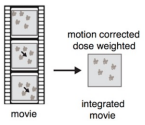
The EM revolution

HARDWARE

- a **new generation of electron detectors** of unprecedented **speed and sensitivity**
- more sensitive: lower doses, **less radiation damage, less noise**, higher signal-to-noise, **higher resolution**
- **fast readout** allow to compensate small movements that inevitably happen when the electron beam strikes the sample

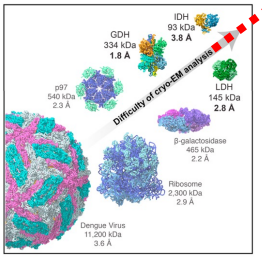
SOFTWARE

- **deblurring**: dozens of images are taken in rapid succession (movies), and beam-induced movements are detected and corrected
- powerful **image processing routines** became available
- **reliable and objective criteria for averaging** tens or hundreds of thousands of single-particle images



31

The EM revolution

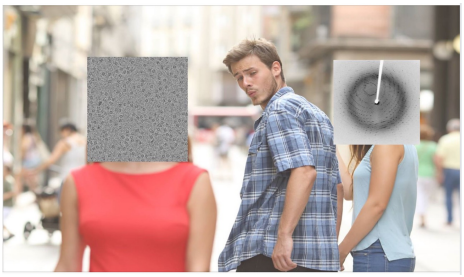


HOWEVER:

- one still needs to produce recombinant protein
- most of the CryoEM structures are at 3-3.5 Å or lower; getting to higher resolution is hard
- anything below 150 kDa is still very difficult
- collecting good data often requires a lot of work in optimizing conditions, avoid damage at the air:water interface, etc..
- drug design is possible at medium resolution, but not ideal

32

EM vs crystallography???

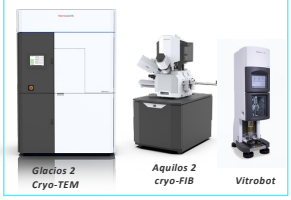


33

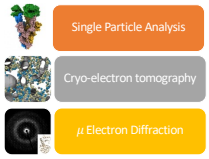
New CryoEM facility in Trieste

A new CryoEM facility is being established in Trieste and will be jointly managed by CNR-IOM and Elettra, providing further structural biology opportunities to the Elettra user community.

The facility will include:



The facility will therefore be able to cover the whole range of electron-based techniques



34

The role of AI in Structural Biology

AI-based predictions:

- Identify functional domains/domain boundaries for construct design where expression of the full-length protein is problematic.
- Provide models to obtain phases by molecular replacement in MX
- Provide models to fit medium/low resolution CryoEM/CryoET maps
- Provide models for computational biology/molecular dynamics/virtual screenings, etc..
- Design more stable mutants (i.e to increase the half-life of an antigen for vaccine development, to obtain stable protein/crystals)

35

The role of AI in Structural Biology

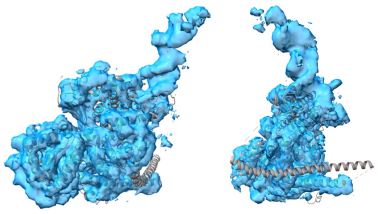
AI-based predictions: current limitations?

- AI works well with **single proteins or domains** – a lot of biology involves homomeric or multimeric complexes (AlphaFold Multimer)
- AI models do not include **ligands or co-factors** (often a "hole" where ligand/co-factor-metal should be) and **post-translational modifications**?
- AI does nor provide **proteins/nucleic acid complexes** (yet?)
- AI networks require **continuous training** on existing protein structures: difficult to predict the structures of proteins with folds that are not well represented in the PDB (**"the dark proteome"**)
- **Proteins are not a static or rigid assemblies**: many exist in multiple conformations and they change shape according to the ligands, substrates, partners they bind. AI gives me ONE conf.
- **Drug design requires better precision** than it can be now achieved with AI

36

The role of AI in Structural Biology

We still need experimental data!



37

Future perspectives?

Back to X-rays?

- Free Electron Lasers (FELs) produce ultrabright (10^9x synchrotron light and ultrashort (fsec) pulses of radiation, also in the X-rays range (XFELs).
- Diffract-before-destroy: data are recorded before any radiation damage occurs & the molecule/crystal explode – no need to freeze.
- Serial femtosecond crystallography (SFX): whole data sets can be derived from thousands of diffraction shots, each taken from a protein nano/micro-crystal before it is blown apart (*aka "The American Method: shoot first, ask questions later"*).
- Time resolved experiments? Often tricky to trigger the reaction in all the molecules inside a micro/nano-crystal.
- Moving towards 'pump-probe' experiments, leading towards *molecular movies*?
- More prosaically: more efficient and automatic inhibitor/drug/fragment screening pipelines

38

Future perspectives?

Beyond X-rays?

- CryoEM is still improving, allowing to look at smaller and smaller structures.
- Cryo-Electron Tomography is obviously the new frontier, allowing to visualise molecules in their natural environment – HOWEVER: medium resolution only with subtomogram averaging (and thus for macromolecules that are large, and numerous enough to be seen in the tomogram.
- In cell NMR? often signal is too small, probably due to crowdiness/multiple interactions?
- Further advances in AI-based predictions? It will certainly happen, but we will still need experimental validation. AI-tools are "tools" not competitors…

39