

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# X-Ray Diffraction in Biology: How Can We See DNA and Proteins in Three Dimensions?

---

Claudine Mayer

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/64999>

---

## Abstract

Knowing the three-dimensional structure of biological macromolecules, such as proteins and DNA, is crucial for understanding the functioning of life. Biological crystallography, the main method of structural biology, which is the branch of biology that studies the structure and spatial organization in biological macromolecules, is based on the study of X-ray diffraction by crystals of macromolecules. This article will present the principle, methodology and limitations of solving biological structures by crystallography.

**Keywords:** biological macromolecules, X-ray diffraction, monocristal, tridimensional structure

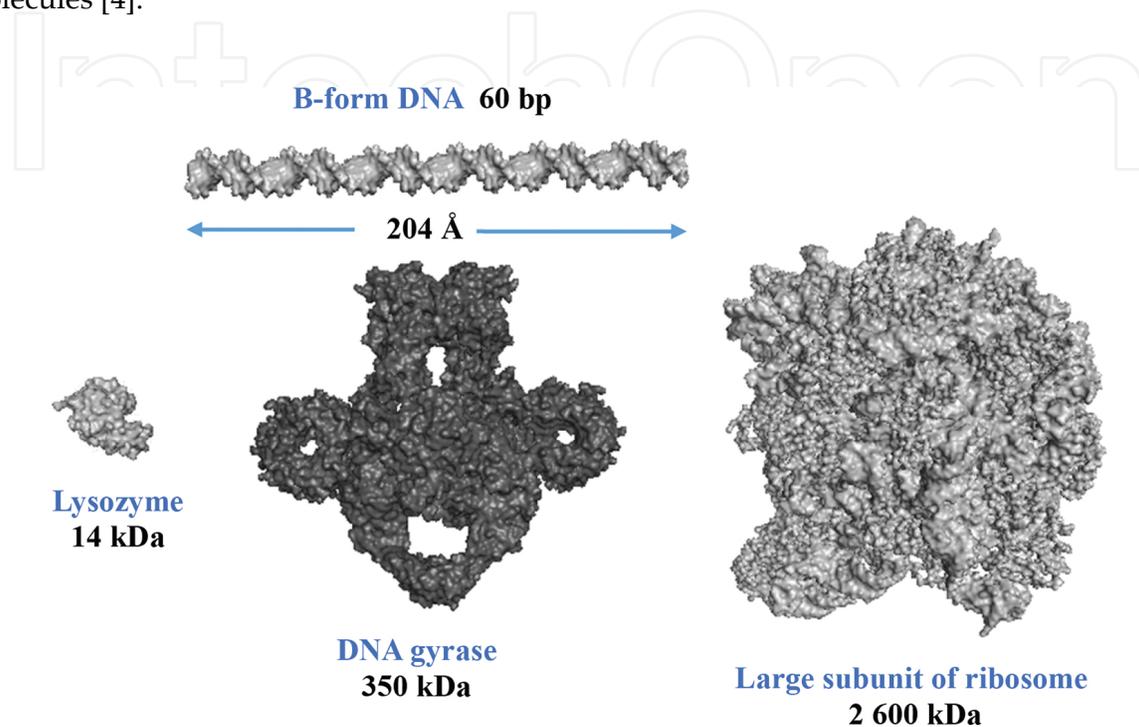
---

## 1. Introduction

In 1953, James Watson and Francis Crick revealed the double helical structure of DNA using the results of Rosalyn Franklin obtained by X-ray scattering on natural filaments formed by DNA molecules [1]. Proteins, the nanomachines essential to living organisms, have their “manufacturing plan” encoded in their DNA gene sequence [2]. During their synthesis, proteins adopt a specific three-dimensional structure that allows them to perform their functions within the cell. “Seeing” the structure of biological macromolecules, such as proteins or nucleic acids (RNA or DNA), allows researchers to elucidate the mechanisms of life in all organisms, and among many other applications, allows them to design new drugs [3].

“Seeing” proteins or nucleic acids in three dimensions, a dream or a reality? Could microscopy, a technic known since more than 350 years that allows to visualize biological cells, be the right

approach? Of course, the dimensions of these two objects, macromolecules and cells are very different: The cell size ranges generally from 10 to 100 microns ( $10^{-6}$  m), the dimensions of biological macromolecules, proteins or nucleic acids, are of the order of tens of angstroms ( $10^{-10}$  m) (**Figure 1**). To reach atomic details, the method of choice is crystallography, whose principle is based on the bombardment by X-ray of crystals composed of biological macromolecules [4].



**Figure 1.** Dimension of biological macromolecules represented at the same scale (picture provided by Dr Jérémie Piton). The length of a 60 base pairs DNA double helix is 204 Å.

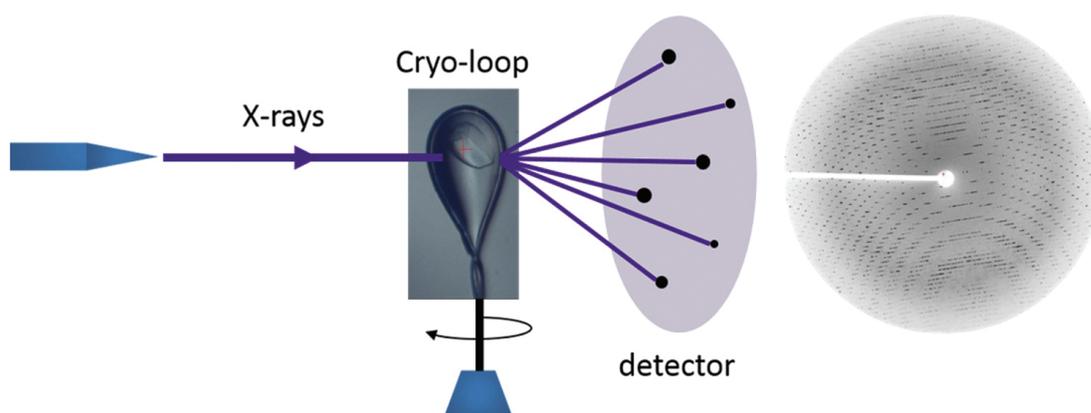
**Why using X-rays?** Their wavelength is of the order of the angström and thus corresponds to the distance between two bound atoms. **Why using a crystal?** To date, the conception of an X-ray microscope encounters two obstacles. First, the signal from a single macromolecule is too low, second, a device, such as lenses, generating a direct image of a macromolecules, does not exist for X-rays. Using a crystal, that contains about  $10^{15}$  identical macromolecules periodically arranged in the three directions of space, overcomes these obstacles.

In only 50 years, crystallography has become the technique of choice for the determination of structures of biological macromolecules at atomic scale, taking advantage of the major advances in the scientific fields as diverse as molecular biology, biochemistry, computer science, physics and more recently robotics. Today, crystallography is able to address the determination of three-dimensional structures of macromolecules more and more complex, more and more quickly. Currently, more than 25 crystal structures are deposited daily in the Protein Data Bank (<http://www.rcsb.org>)<sup>1</sup> [5].

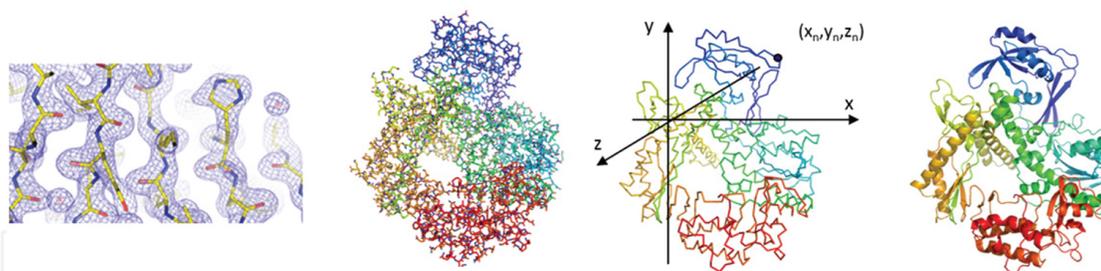
<sup>1</sup> The protein Data Bank (PDB) is a databank that contains 120,262 entries of macromolecules structures (protein, nucleic acids, complexes), 107,455 have been solved by X-ray crystallography (July 2016).

The physical principle of crystallography is based on X-ray diffraction by all the electrons constituting the atoms of all the macromolecules contained in the crystal (**Figure 2**). The analysis of these diffraction data then allows the crystallographer to calculate the electron density, which is the distribution of the electron cloud of the macromolecule in the crystal. This electron density provided it is sufficiently precise – this preciseness depends on the resolution of the diffraction data – allows the localization of each atom of the molecule, and thus the determination of its coordinates in the three-dimensional space [6].

A.



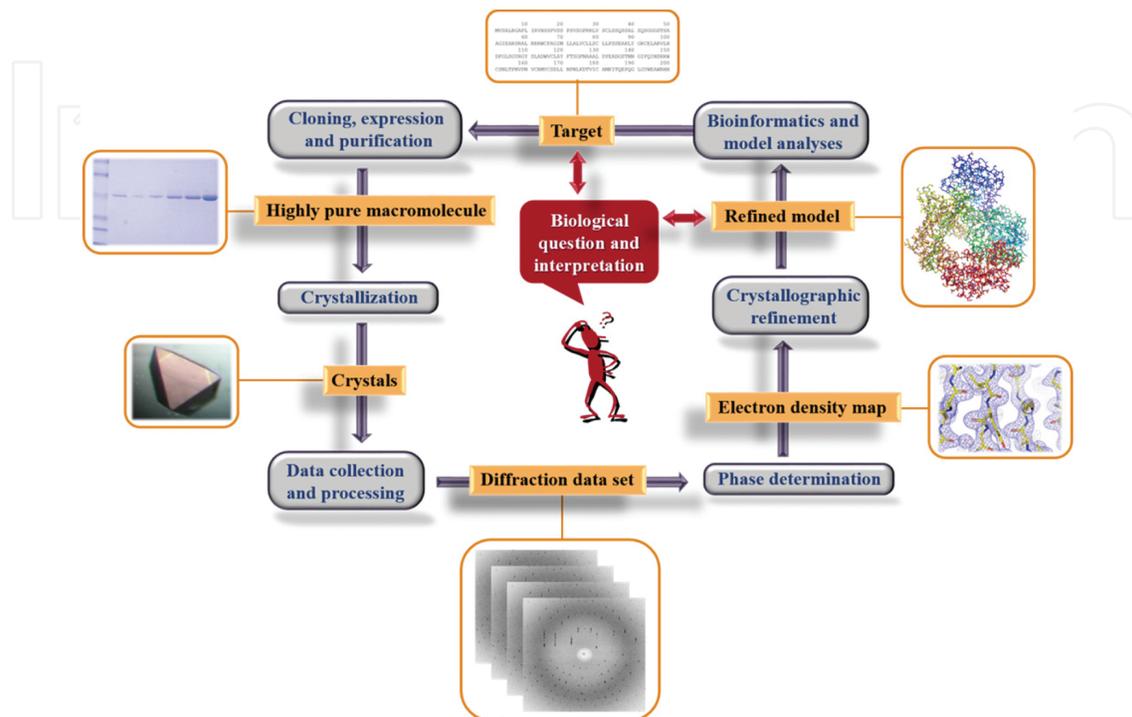
B.



**Figure 2.** The principle of crystallography. (A) A monochromatic X-ray beam bombards a crystal frozen in a cryo-loop that rotates on itself. The observed diffraction spots are the result of the impact on the detector of the wave diffracted by the electrons in the crystal. (B) Electron density map of a fragment of a macromolecule is represented (left). The three-dimensional structure of a macromolecule (here a protein) is represented in three ways: all-atoms, backbone and cartoon representation (see **Figure 9**).

To get this three-dimensional structure, several steps that falls within multiple disciplines are required (**Figure 3**). Each of these steps represents potential bottlenecks that need to be overcome. These are the production and the purification of the macromolecule, its crystallization, diffraction data collection and processing. Another crucial step is the determination of the phases of the measured signal, absolutely required to calculate the electron density. The last step is the refinement of the built structure, called the model, which will then be interpreted in the context of its biological function. The analysis of the model will thus raise new questions

leading to the resolution of other crystal structures, such as structure of a complex between the studied protein and its partners [7]. We will in the following sections describe each of these steps.



**Figure 3.** The main steps of the three-dimensional structure determination of biological macromolecules by crystallography.

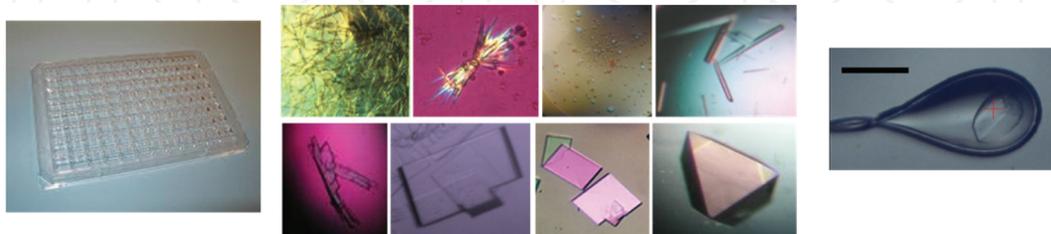
## 2. Steps upstream the structure determination

The first step, a step that falls within biology and includes molecular biology and biochemistry techniques, is the production of highly pure macromolecule in large quantity. Once the sequence of the macromolecule to be studied has been identified and characterized by bioinformatics analyses, the sequence corresponding to the gene of the macromolecule is cloned in an expression vector and produced classically in a bacterial organism (typically *Escherichia coli*). The macromolecule is then extracted from the bacterial cells and purified using chromatographic techniques. The prerequisite for the next step is to obtain a concentrated<sup>2</sup> (of the order of tens of grams per liter) and highly pure sample (greater than 98%) of the macromolecule.

The next bottleneck is based on physical chemistry, specifically crystallization which addresses concepts such as solubility of molecules and their transition from soluble state to a solid

<sup>2</sup> In a aqueous solution containing a buffer, salt and various additives.

crystalline ordered state [8]. This step, built on statistical screenings plays with the variation of parameters such as temperature, pH, concentrations of biological macromolecules, as well as nature and concentration of crystallizing agents and various additives [9]. Obtaining a single homogeneous crystal, that result to high quality diffraction data, represents a crucial step in the process of determining a macromolecular structure. In order to increase the success rate, crystallization robots are used today to screen more than several thousands of parameters. The size (from tens to hundreds of microns) and the morphology of the crystals are highly variable (**Figure 4**) and are not necessarily related to their diffracting power and quality.

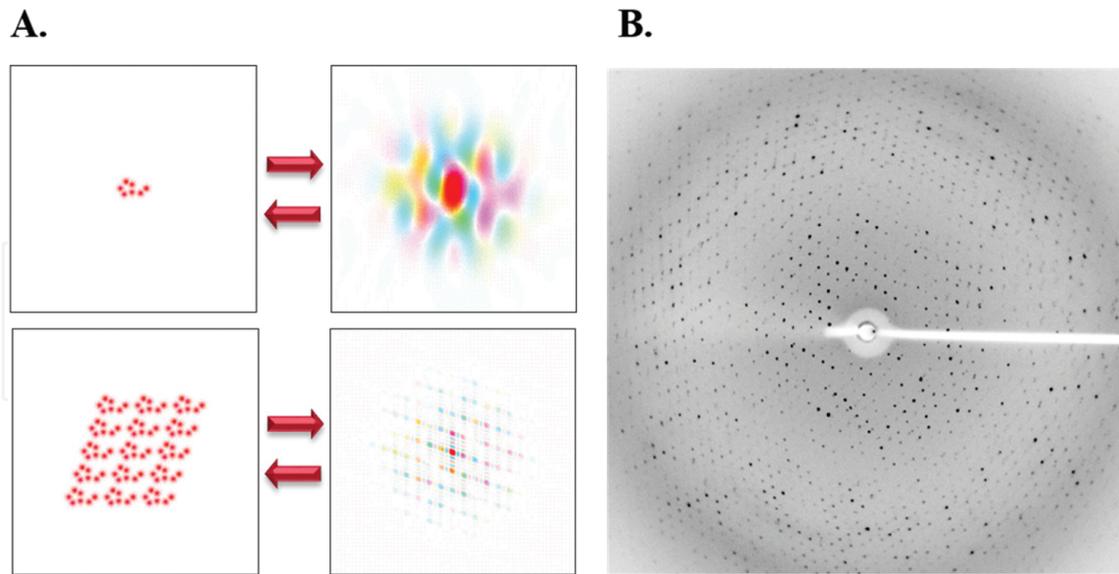


**Figure 4.** Crystals of biological macromolecules. Left, a typical crystallization plate used in crystallization robots that allows to screen 96 crystallization conditions. Middle, different crystals of macromolecule. Right, the crystal is shown in its cryo-loop (see Section 2.). The black bar is 100 microns.

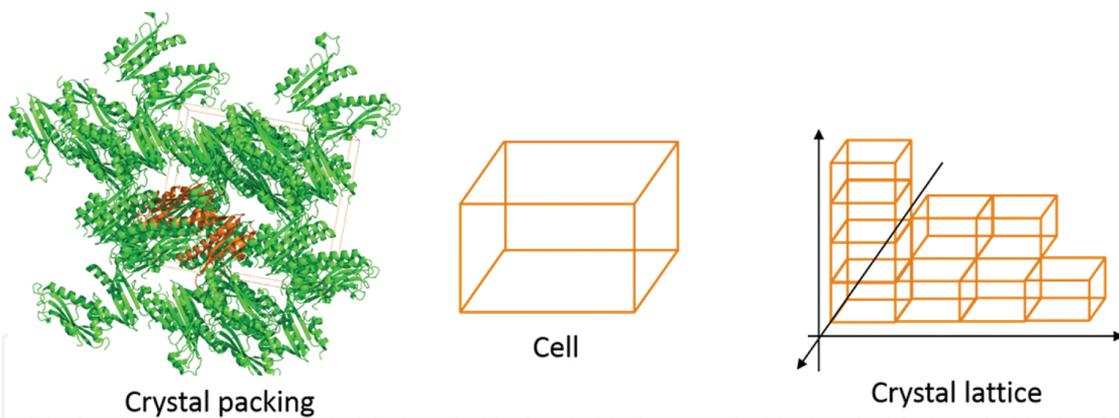
### 3. The diffraction data

The crystals obtained during the previous step are fished using a small loop (**Figure 4**), cryo-cooled to protect them from radiation damage [10], and then placed into a monochromatic X-ray beam produced by an appropriate source, either a rotating anode generator available in crystallography laboratories or a synchrotron radiation, the latter producing significantly more intense beams [11]. Under these conditions, the waves scattered by the electrons of the macromolecules that are three-dimensionally ordered in the crystal add up in given directions (the diffracted beam is characterized by a structure factor, **Figure 7**) and generate a diffraction spot on the screen of the detector (**Figure 5A**). All the spots, regularly spaced, form the diffraction pattern (**Figure 5A**). This diffraction pattern is reconstituted by using several hundreds of images, each corresponding to an orientation of the crystal that rotates on itself during the measurement of the diffraction data (**Figure 2** and **Figure 5B**). The information contained in each diffraction spot is characterized by the amplitude and the phase of the structure factor characterizing the corresponding scattered wave.

The three-dimensional distribution of the spots is directly related to the cell parameters, e.g. the three lengths of the parallelepiped that constitutes the volume element (the cell), which is regularly repeated in space (**Figure 6**) and allows to describe the crystal. The distribution of the spot intensities is directly related to the electron density distribution (the macromolecules) in the cell. Mathematically, this means that the diffraction pattern is the Fourier transform of the electron density (**Figure 7**).



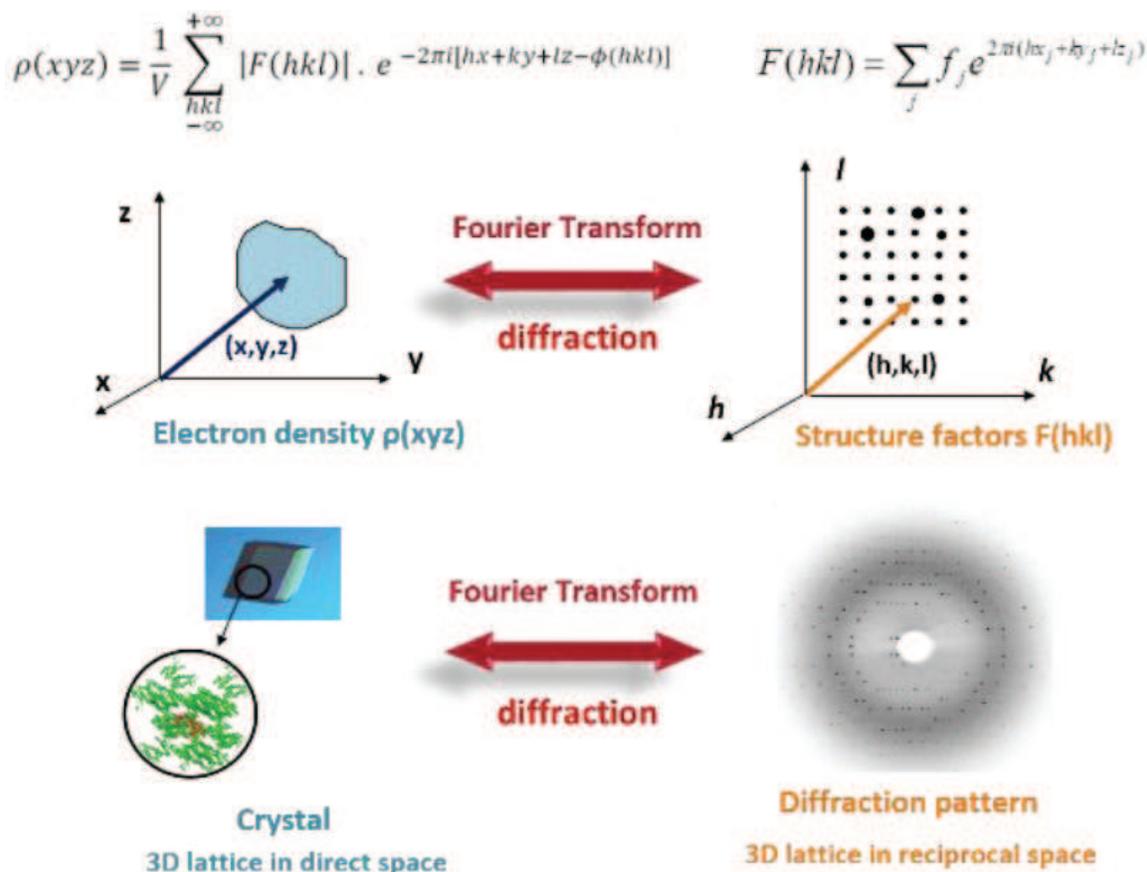
**Figure 5.** (A) The diffraction pattern (or Fourier transform) of a crystallized molecule generates a three-dimensional spot lattice (bottom), whose background image corresponds to the Fourier transform of a single molecule (top). The amplitude and phase of the diffracted beams are represented by the color brightness and the color hue, respectively (Kevin Cowtan's Picture Book of Fourier Transforms (<http://www.ytbl.york.ac.uk/~cowtan/fourier/fourier.html>)). (B) Example of detector image constituting the diffraction pattern. Hundreds of images are usually recorded. The spots at the image edge are high resolution spots, providing the most detailed information.



**Figure 6.** The macromolecules are ordered in the three directions of space and form the crystal packing (left). The smallest volume that is repeated by translation in all directions of space is the cell (middle and right). It forms a parallelepiped characterized by three vectors named  $a$ ,  $b$  and  $c$ .

The electron density contained in one cell can thus be calculated by inverse Fourier transform, a mathematical property of this transformation, provided the amplitude and the phase of all the diffracted beams are known (**Figure 7**). Whereas the amplitude is directly proportional to the intensity of the diffracted spots, the phase information is not experimentally measurable.

In summary, the crystal “realizes” a Fourier analysis producing diffraction data, and the crystallographer will calculate a Fourier synthesis to get the electron density contained in one cell (**Figure 7**).



**Figure 7.** Schematic summary of the relationship between the diffraction (structure factors and diffraction spots) and the electron density of the structure three-dimensionally packed in the crystal.

#### 4. From the diffraction data to the electron density

Three main methods exist for the estimation of the phases [12]. We have to remember here that the number of phases to be estimated is typically several tens to hundreds of thousands (the phase of each spot for which the intensity has been measured has to be estimated).

The first method is *molecular replacement*. It uses the known structure of a homologous protein. To date, approximately 60% of the structures found in the PDB were solved by this method [5]. It consists of constructing a virtual crystal by placing the homologous structure in the cell of the crystal studied using mathematical translation and rotation functions and comparing the diffraction pattern calculated from this virtual crystal and the measured diffraction data. Since the Fourier transforms of two homologous molecules placed in the same crystal are similar, the calculated phases are an excellent approximation of the phases of the measured signal [13–15].

The second method is *multiple isomorphous replacement*, which consists in diffusing heavy atoms (electron-rich) in the crystal [16]. In the first protein structure determination, the phase problem was solved using this method, those of the myoglobin and the hemoglobin [17, 18], by John

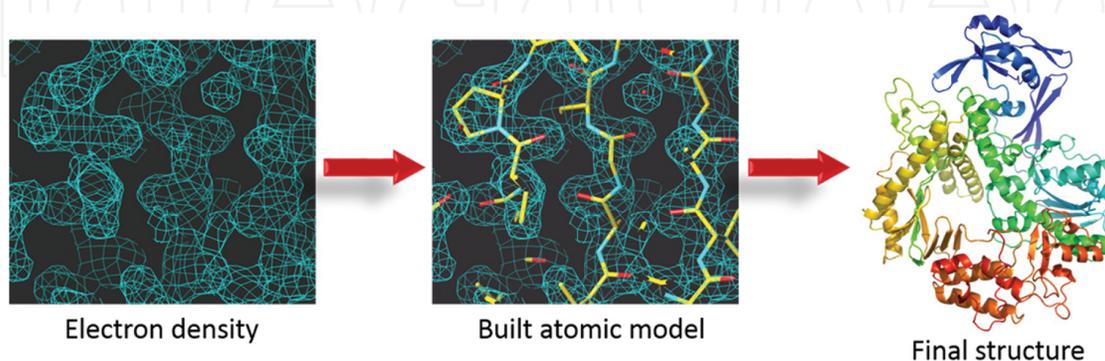
Kendrew and Max Perutz in 1960. The presence of the heavy element slightly modifies the diffraction intensities and the comparison of the diffraction pattern in the presence and absence of these heavy elements allows the estimation of the phases by triangulation, after having positioned the heavy atoms in the crystal lattice using methods known as Patterson functions [19].

The third method is *anomalous dispersion*, a specific property of the diffraction pattern when absorption of X-radiation is no longer negligible [20, 21]. This method consists in varying the incident beam wavelength around the absorption edge of one of the atom type contained in the molecule. Comparing the diffraction pattern at different wavelengths will allow the estimation of the phases using methods similar to that of the isomorphous replacement [22]. Selenium is often used because it has an absorption edge near to the wavelengths used (e.g. 1 Å). For proteins, selenomethionine, an amino acid for which the sulfur is replaced by selenium, is generally introduced biosynthetically [23]. In the case of nucleic acids, modified bases containing bromine are frequently used [24].

## 5. From the electron density to the structural model

Once a first set of phases is estimated, a first electron density map is calculated. If this map is sufficiently interpretable, the macromolecule can be built step by step in this map (**Figure 8**). A combination of automated algorithm and manual method available through interactive graphics softwares are used [25], leading to a final model composed of the three-dimensional coordinates of each atom of the cell content constituted by one or several macromolecules.

From that first built model, the diffraction intensities are calculated by Fourier transform and compared to the intensities experimentally measured. This comparison allows the step by step improvement of the model. This cyclical process is called the crystallographic refinement, alternating the search for global minimum of energy functions and manual reconstruction of the model [26].

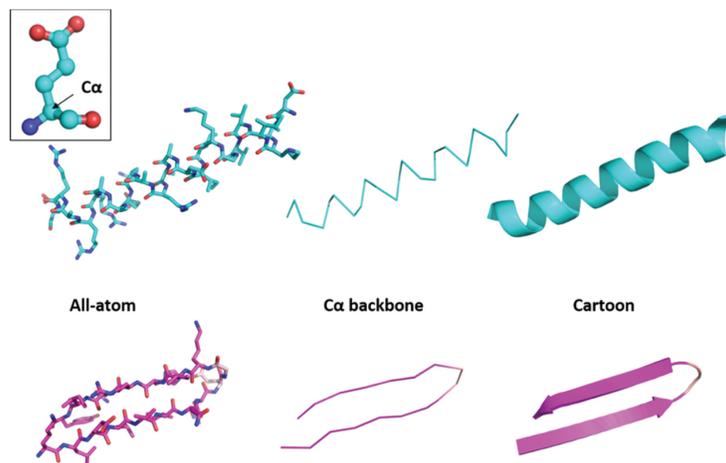


**Figure 8.** The calculation of the electron density map (left) allows the building of the atomic model step by step (middle) and leads to the three-dimensional model of the structure (right).

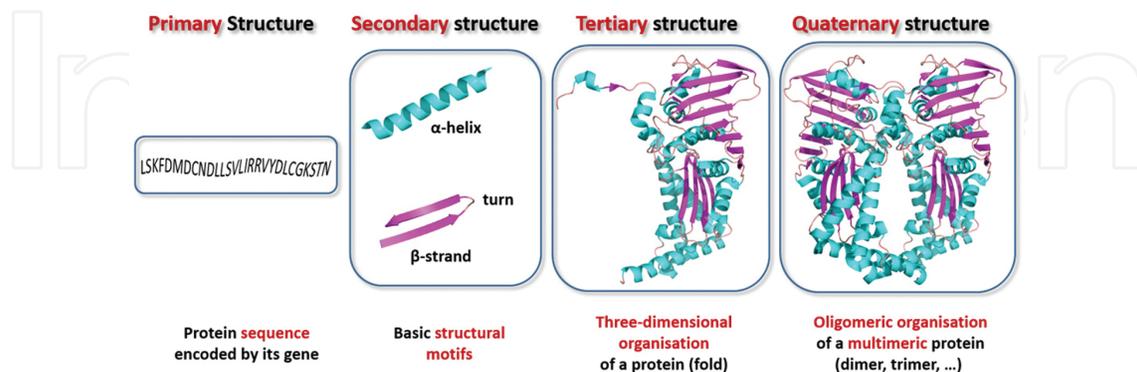
## 6. Steps downstream the structure determination

The final step, downstream the structure determination by X-ray diffraction, concerns the interpretation of the structure and its integration into the biological context [27–29]. It consists in the understanding of the structural result as a three-dimensional object and the appreciation of its function at the cellular or evolution level. The description of the interatomic interactions, the secondary structures (**Figure 9**), the domains and their arrangement that defines the fold or the tertiary structure (**Figure 9**), as well as the characterization of the shape, the electrostatic properties and the quaternary structure based on the content of the cell in the crystal packing, are often complemented by the study of the macromolecule in solution, to better characterize its oligomeric (**Figure 9**) and its dynamic behavior, alone or in the presence of interactors, if known. These studies use a variety of biophysical methods, such as mass spectrometry, analytical ultracentrifugation, light scattering, microcalorimetry or surface plasmon resonance

A.



B.



**Figure 9.** (A) The protein structures are represented by three modes of representation (see also **Figure 2**). The “all-atom” representation shows all the atoms in the protein, the representation “C $\alpha$  backbone” shows only one atom of each amino acid, the C $\alpha$  carbon atom, and cartoon representation shows the secondary structures in the shape of a helix for  $\alpha$ -helices and in the form of arrows for  $\beta$ -strands. (B) Protein structures are described in four levels, from primary to quaternary structure.

(Biacore® technology), etc ... [30]. In the case of enzymes, these studies will be coupled with enzymological approaches to determine the activity and the catalytic constants.

An analysis based on bioinformatics tools will allow to place the structure determined in the context of structural and evolutionary knowledge at a given time [31]. The lessons learned from these studies, often of primary importance, provide information including the classification of the structure and its sequence within a family counterparts, on the distribution and evolution of folding in the different domains of life (viruses, bacteria, archaea, eukaryotes), on the possible function when it is unknown, on the catalytic site and its spatial conservation and sequence, on the degree of oligomerization or on the existence of interaction with other partners, proteins, nucleic acids or ligands. A final type of study seeks to place the three-dimensional object into the context of the knowledge on the major biological mechanisms of life, such as knowledge on gene expression with transcriptomics, on complex formation with interactomics, etc ... This information will include the characterization of the partners of the studied macromolecule at the scale of the cell or the whole organism.

All these steps, from the structure determination to the biological interpretation, far from being the end of the story, are often the beginnings of new structural studies (**Figure 3**). These can be articulated around analyses of the relative importance of the components of the macromolecule, the aminoacids, by determining the structure of mutants, or the studies of the interactions with partners by determining the structure of macromolecular complexes.

## Author details

Claudine Mayer<sup>1,2,3\*</sup>

Address all correspondence to: mayer@pasteur.fr

1 Department of Structural Biology and Chemistry – Institut Pasteur, Paris, France

2 CNRS UMR 3528, Paris, France

3 Paris Diderot University, Sorbonne Paris Cité, Paris, France

## References

- [1] Watson J.D., Crick F.H.C., Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 1953, 171, p. 737-738.
- [2] Anfinsen C., The formation and stabilization of protein structure. *Biochem. J.*, 1972, 128, p. 737-749.

- [3] Scapin G., Structural biology and drug discovery. *Curr. Pharm. Des.*, 2006, 12, p. 2087–2097.
- [4] Blundell T.L., Johnson L.N., Protein Crystallography. Academic Press, New York, 1976.
- [5] Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E., The Protein Data Bank. *Nucl. Acids Res.*, 2000, 28, p. 235-242.
- [6] Wlodawer A., Minor W., Dauter Z., Jaskolski M., Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J.*, 2008, 275, p. 1-21. Review.
- [7] Wlodawer A., Minor W., Dauter Z., Jaskolski M., Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J.*, 2013, 280, p. 5705–5736. Review.
- [8] Chernov A.A. Protein crystals and their growth. *J. Struct. Biol.*, 2003, 142, p. 3–21.
- [9] Luft J.R., Collins R.J., Fehrman N.A., Lauricella A.M., Veatch C.K., DeTitta G.T. A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J. Struct. Biol.*, 2003, 142, p. 170-179.
- [10] Garman E.F., Schneider T.R. Macromolecular cryocrystallography. *J. Appl. Crystallogr.*, 1997, 30, p. 211-237.
- [11] Moffat K, Ren Z., Synchrotron radiation applications to macromolecular crystallography. *Curr. Opin. Struct. Biol.*, 1997, 7, p. 689–696.
- [12] Taylor G., The phase problem. *Acta crystallogr. D.*, 2003, 59, p. 1881-1890.
- [13] Rossmann M.G. The molecular replacement method. *Acta Crystallogr. A.*, 1990, 46, p. 73-82.
- [14] Tickle I.J., Driessen H.P., Molecular replacement using known structural information. *Methods Mol. Biol.*, 1996, 56, p. 173-203. Review.
- [15] Abergel C., Molecular replacement: tricks and treats. *Acta Crystallogr. D Biol. Crystallogr.*, 2013, 69, p. 2167-2173. Review.
- [16] Perutz, M.F., Isomorphous replacement and phase determination in non-centrosymmetric space groups. *Acta Cryst.* 1956, 9, p. 867–873.
- [17] Perutz M. F., Rossmann M. G., Cullis A. F., Muirhead H., Will G., North A. C., Structure of hemoglobin: a three-dimensional Fourier synthesis at 5.5 Å resolution, therefore obtained by X-ray analysis. *Nature*, 1960, 185, p. 416-422.
- [18] Kendrew J.C., Bodo G., Dintzis H.M., Parrish R.G., Wyckoff H., Phillipps D.C., A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, 1958, 181, p. 662-666.

- [19] Patterson A.L., A direct method for the determination of the components of interatomic distances in crystals. *Zeitschrift für Kristallographie*, 1935, 90, p. 517-542.
- [20] Read R.J., As MAD as can be. *Structure*, 1996, 4, p. 11-14. Review.
- [21] Blow D.M., How Bijvoet made the difference: the growing power of anomalous scattering. *Methods Enzymol.*, 2003, 374, p. 3–22.
- [22] Ealick S.E., Advances in multiple wavelength anomalous diffraction crystallography. *Curr. Opin. Chem. Biol.*, 2000, 4, p. 495–499.
- [23] Doublet S., Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.*, 1997, 276, p. 523–530.
- [24] Anderson A.C., O'Neil R.H., Filman D.J., Frederick C.A., Crystal structure of a brominated RNA helix with four mismatched base pairs: an investigation into RNA conformational variability. *Biochemistry*, 1999, 38, p. 12577-12585.
- [25] Emsley P., Debreczeni J.E., The use of molecular graphics in structure-based drug design. *Methods Mol. Biol.*, 2012, 841, p. 143-159.
- [26] Tronrud, D.E., Introduction to macromolecular refinement. *Methods in molecular biology series*, 2007, vol. 364: macromolecular crystallography protocols: vol. 2: 34 structure determination. p. 231–253, Humana Press Inc, Totowa, NJ.
- [27] Baker E.N., Seeing atoms: the rise and rise of crystallography in chemistry and biology, chemistry in New Zealand, January 2011, (New Zealand Institute of Chemistry), New Zealand.
- [28] Shi Y., A glimpse of structural biology through X-ray crystallography. *Cell*, 2014, 159, p. 995-1014.
- [29] Yonath A., X-ray crystallography at the heart of life science. *Curr. Opin. Struct. Biol.*, 2011, 21, p. 622–626.
- [30] Malik S.S., Shrivastava T., Protein characterization using modern biophysical techniques. *Advances in Protein Chemistry*, 2013, OMICS Group eBooks. (<http://www.esciencentral.org/ebooks/ebooks-about.php>)
- [31] Lecompte O., Thompson J.D., Plewniak F., Thierry J., Poch O., Multiple alignment of complete sequences (MACS) in the post-genomic era. *Gene*, 2001, 270, p. 17. Review.