

X-ray Crystallography of Biological Macromolecules

K Suguna

The three-dimensional structures of thousands of biological macromolecules have been determined by X-ray diffraction, since the first structures were reported about half a century ago. The structures revealed how critical the shapes and sizes of the molecules are to perform various biological processes in living creatures. In addition to providing atomic details of how the molecules function and interact with other molecules, the structures help in designing medicines.

Little was known when the mysterious X-rays were discovered by Wilhelm Röntgen in 1895, that they were to reveal the mysteries of life and play a major role in revolutionizing several disciplines of science in the following century. X-rays have a wavelength smaller than that of visible and ultraviolet rays and closer to the distances between atoms in molecules. The electrons of the atoms scatter X-rays and if identical molecules are arranged in a periodic manner in three dimensions as in a crystal, the scattered waves form an ordered diffraction image as the crystal acts as a natural diffraction grating. This image is a collection of a number of diffracted rays. The position and intensity of each of these rays depend on the arrangement of atoms – in other words, on the three-dimensional structures of molecules present in the crystal. By analyzing several diffraction images generated by changing the orientation of the crystal, the structures of the molecules can be derived mathematically.

A large number of molecules constantly at work keep the living organisms alive by performing numerous functions. The shapes of these molecules determine the type of functions they carry out. To determine the shapes or the structures of these biological molecules such as DNA and proteins, which are billions of times



K Suguna is at the Molecular Biophysics Unit, Indian Institute of Science, Bangalore. Her research involves the determination of the crystal structures of macromolecules by the X-ray diffraction method. She is interested in the structures and functions of biological macromolecules.

Keywords

X-ray diffraction, biological macromolecules, crystal structure, crystallography.



X-rays have the power to penetrate through molecules and reveal not only their shapes and sizes but also the intricate arrangement of the atoms that make up the molecules.

smaller than a mustard seed, one has to look beyond the light microscope. X-rays have the power to penetrate through molecules and reveal not only their shapes and sizes but also the intricate arrangement of the atoms that make up the molecules. The structures enable us to understand the precise nature and the exact mechanism of their functions and aid in designing drugs to diseases. The technique originally practised in physics laboratories entered the realm of biology. The experimental procedures and the subsequent calculations are complicated but the final structures are precise and provide a wealth of useful information on the physical and chemical nature of the molecules and explain the functions of biological molecules in atomic detail.

The Beginning

The first X-ray diffraction pictures were taken and the theory was worked out in the early 20th century. Max von Laue's first X-ray diffraction image of copper sulphate in 1912 followed by the elucidation of the Bragg's law, $2d\sin\theta = n\lambda$, by W H Bragg and W L Bragg in 1913 were mainly responsible for the establishment of X-ray crystallography. The theories and methods for the structure solution of small molecules were developed in the next 15 years and since then a large number structures have been determined with significant contributions not only to physics but also to chemistry and material science. Small molecules are made of two to a few hundred atoms in contrast to macromolecules consisting of thousands of atoms. The feasibility that the method can be applied to large molecules was established with the crystallization of urease in 1926 by J B Sumner and recording the diffraction pattern of pepsin in 1934 by J D Bernal and Dorothy Crowfoot (Hodgkin). W T Astbury recorded diffraction patterns from natural fibrous proteins as early as 1933. Macromolecular crystallography or the structure determination of large biomolecules emerged in the 1950's, the first breakthrough being the structure of the genetic material DNA. Guided by the diffraction patterns of fibrous DNA molecules recorded by Maurice Wilkins and Rosalind Franklin, with great ingenuity and intuition, the double helix structure of



DNA was proposed in 1953 by James Watson and Francis Crick. This finding, considered to be one of the greatest discoveries of the last century, not only revealed the secret of life, but is the origin of many fields in biology that we have today. The elucidation of the triple helical structure of collagen, the protein present in large amounts in the skin and the bones of vertebrates, proposed by G N Ramachandran¹ in 1954 based on fibre diffraction photographs is a great contribution to structural biology from India. One of his outstanding theoretical contributions is the much acclaimed Ramachandran map (1963) that explores the conformational space for proteins and has become a valuable and an indispensable asset in building protein structures.

Soon followed an equally important milestone, the structure determination of two proteins, haemoglobin and myoglobin by Max Perutz and John Kendrew in the late 1950's setting the stage for the determination of the three-dimensional structures of large molecules. Structures of lysozyme and other catalytic enzymes, and the protein hormone insulin were elucidated in the following decade. These structures enabled the visualization of the α -helix and β -sheets known as the secondary structural elements of proteins ingeniously proposed by Linus Pauling in 1951. *Figure 1* shows the crystal structures of DNA, collagen and haemoglobin.

¹ Details of G N Ramachandran's contributions to crystallography and protein structure can be found in *Resonance*, Vol. 6, No. 10, 2001.

Protein data bank accession
<http://www.rcsb.org>

Programs used to generate the figures of the molecules

PyMOL

(<http://www.pymol.org>)

Weblab Viewer

(<http://www.msi.org>)

RIBBONS version 2.0

by M Carson (1992)

MolScript by P Kraulis, *Journal of Applied Crystallography*, Vol.24, pp.946-950, 1996.

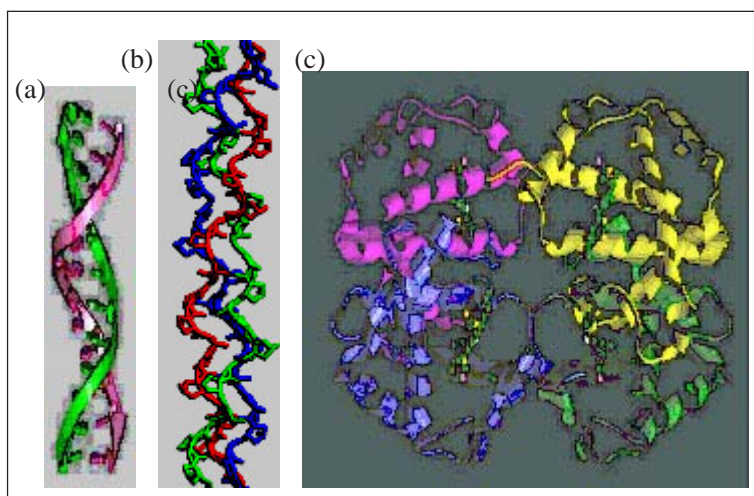


Figure 1. Structures of (a) the B-form of DNA (b) collagen and (c) haemoglobin. The pictures are drawn using the coordinates of the molecules taken from the Protein Data Bank.



Even though thousands of proteins have been crystallized until now, there is no defined method to produce crystals.

Methods of Structure Determination

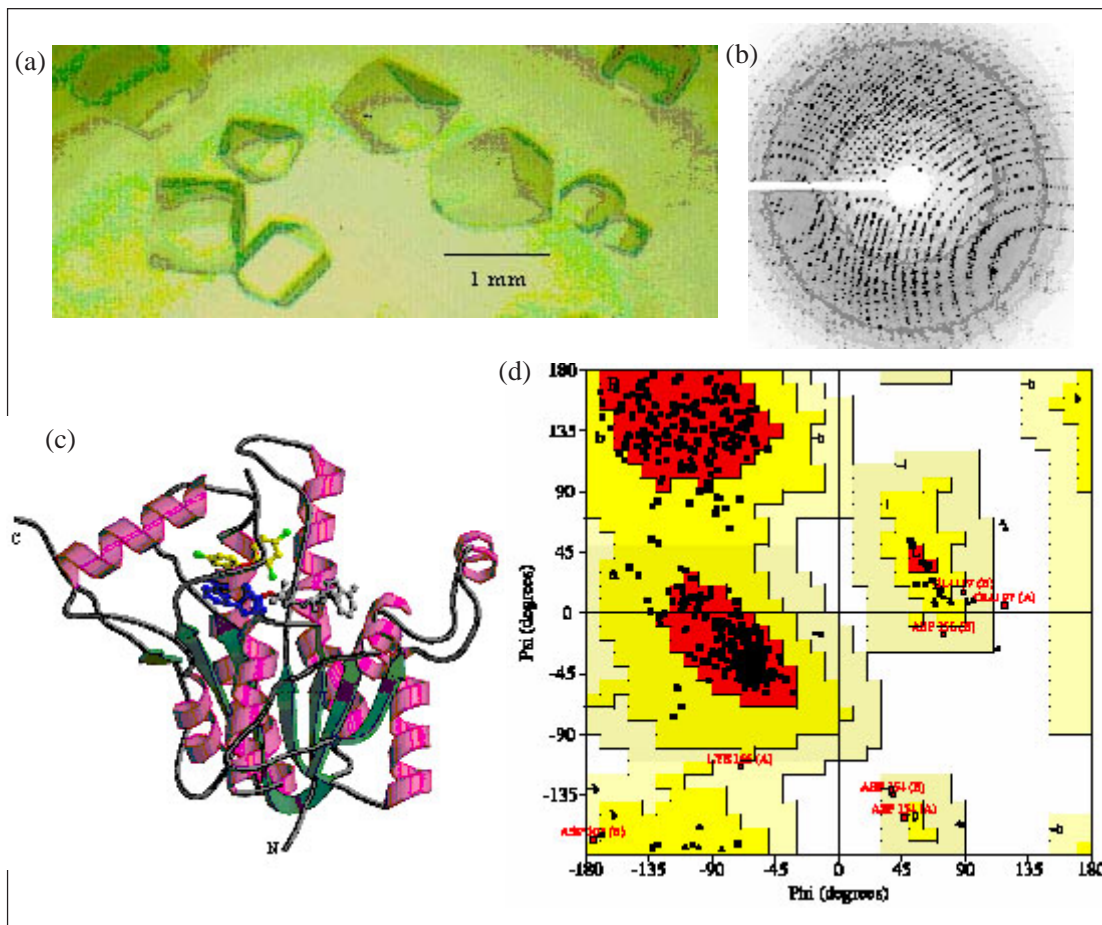
Molecules like the DNA and collagen can be drawn as fibres and exposed to X-rays. Proteins have to be made in the form of single crystals which can then produce hundreds of distinct diffracted rays. In any structure elucidation process, crystallization is the first and the most crucial step but the most uncertain one. Even though thousands of proteins have been crystallized until now, there is no defined method to produce crystals and crystallization of proteins has remained as a trial and error method. For each new protein, hundreds of conditions have to be screened with a large number of precipitants like salts and organic solvents, varying several parameters like the pH, concentration of the protein and the precipitant, temperature etc. Also highly pure and homogeneous proteins are required for crystallization.

Fibre diffraction provides only limited data whereas a large number of diffraction data are available from the crystals of macromolecules. From the intensities of these diffracted rays, we could calculate the amplitude of the structure factors (which is given by $|F| = \sqrt{I}$) but the phases of the diffracted rays can not be obtained directly from the intensities which is possible in the case of small molecules. To calculate the electron density values in the crystal which give the positions of the atoms, both amplitude and phase of each reflection are required. Various methods are employed to solve this phase problem for macromolecules: (i) Multiple Isomorphous Replacement method (ii) Molecular Replacement method and (iii) Anomalous dispersion method. *Figure 2* shows the crystals, diffraction pattern and the structure of a protein from the malarial parasite, *Plasmodium falciparum* and the corresponding Ramachandran map. The structure of this enzyme has been determined in our laboratory by the molecular replacement method.

Available Structures and their Implications

X-ray crystallography gives us the precise position of each atom present in these large molecules made up of thousands of atoms:

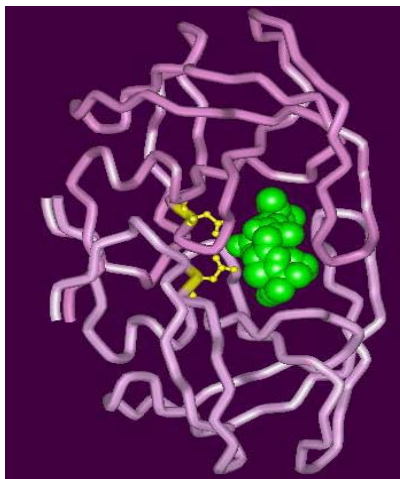




mainly carbons, oxygens, nitrogens, hydrogens and a small percentage of other atoms. These details enable us to understand various processes involved in biological systems in detail. We now know how oxygen from the lungs is carried to the tissues by a protein called haemoglobin which makes the blood look red, and how carbon dioxide is carried back from the tissues to the lungs and how oxygen is stored by the protein, myoglobin. The mechanism by which food is digested by the scissors-like action of certain enzyme molecules in saliva, stomach and the small intestine has been worked out. How the body's defence molecules capture and fight with any foreign unwanted invaders of the body like bacteria or viruses; how certain proteins bind to DNA and switch ON or OFF some reactions; how the signals to

Figure 2. (a) crystals of a protein from the malarial parasite, *Plasmodium falciparum*, (b) one of the several diffraction images from the crystal, (c) the structure of the protein, (d) the Ramachandran map of the amino acid residues of the protein. Proline residues are shown as triangles and the other residues as squares.

Figure 3. The structure of a protein from the AIDS virus. This protein is an enzyme called HIV protease which has two equal and symmetric molecules shown in pink. Small molecules such as the one shown in green are designed to fit perfectly in the cavity of the protein to block access to its atoms (shown in yellow) which are essential for its function. Such small molecules are being used in the treatment of AIDS to inhibit the function of the protein thereby controlling the spread of the virus in the body. The picture is drawn using the coordinates of the molecules taken from the Protein Data Bank.



the brain are controlled; how proteins are synthesized by huge molecules called the ribosomes; how signals from outside the cell are transmitted into the cell triggering a series of reactions, are understood at the atomic level once the structure of the molecule involved has been determined.

Structural work on the virus that causes common cold and on the proteins from the virus causing flu offered ways to find medicines for these ailments. Drugs to combat AIDS are the best example of structure-based design (*Figure 3*). Designing a drug is like finding the key to fit perfectly into the lock. Structures provide sites on the surfaces of molecules of a disease-causing pathogen that can be blocked to inactivate the proteins to stop further spread/growth or the survival of the pathogens thereby containing the diseases. The major challenge in drug design is to find compounds that specifically inhibit pathogen functions but do not interfere with the host machinery. An alternative approach could be to block the sites on the host receptor such that the entry of the pathogen or the pathogen-host interaction can be prevented. A few small molecules designed based on the structures of proteins from HIV, the virus causing AIDS are being used in the treatment of AIDS now. It is amazing that a small molecule made of less than 100 atoms is able to fight the gigantic virus. From the structure of a major protein, the fine

The major challenge in drug design is to find compounds that specifically inhibit pathogen functions but do not interfere with the host machinery.



details of photosynthesis have been worked out. This is the most important chemical reaction on the earth by which food is synthesized by making use of water and carbon dioxide in the presence of sun light. Proteins or DNA can not achieve much in isolation but have to interact with other molecules, small or big, to maintain life. Crystal structures of a number of these complexes between molecules have been determined which have completely changed our perception of the meaning of life. Structural biology makes us think in terms of molecules. We may be asleep at night but molecules are at work, tirelessly carrying out reactions with uncanny precision recognizing and interacting with the right molecules at the right moment.

Until 1970 the structures of about 10 proteins were known. Now, every day more than ten structures are being determined. New and constant technological developments in various fields contributed to this enormous growth. Due to recombinant DNA techniques it is possible to obtain large amounts of pure proteins required for structure analysis and to prepare desired mutations to study their effects on structure and function. Enormous increase in the performance and speed of computers tremendously reduced the time taken for various calculations involved and helped in 3-d visualization of molecules. Synchrotron sources not only provided highly brilliant beam of X-rays but also made it possible to tune their wavelengths. One can measure the intensities of the reflections with devices like the imaging plates and charge coupled devices which replace the conventional photographic methods. With the new devices the time taken for data collection and processing has enormously reduced and the accuracy in the measurements has considerably increased. The structures of more than 30,00 proteins (which include mutants, complexes and different forms) have been deposited in the protein data bank now. Now the boundaries between various conventional fields of science are diffusing. With the overlapping scientific interests, new areas like biophysics, molecular biology, biotechnology, genetic engineering, structural biology, structural genomics etc. are emerging.

Until 1970 the structures of about 10 proteins were known. Now, every day more than ten structures are being determined.



Suggested Reading

- [1] M G Rossmann and E Arnold, *Crystallography of Biological molecules, International tables for crystallography*, Kluwer Academic Publishers, Vol. F, 2001.
- [2] G N Ramachandran and G Kartha, Structure of collagen, *Nature*, Vol. 174, pp. 269-270, 1954.
- [3] G N Ramachandran, C Ramakrishnan and V Sasisekharan, Stereochemistry of Polypeptide chain configuration, *Journal of Molecular Biology*, Vol. 7, pp. 95-99, 1963.
- [4] M Vijayan, N Yathindra and A S Kolaskar, (Eds), *Perspectives in Structural Biology*, Universities Press (India) Limited, 2001.
- [5] M Vijayan, Structural genomics of microbial pathogens - An Indian programme, *Current Science*, Vol. 85, pp. 878-885, 2003.

Today the crystal structures of thousands of proteins that perform diverse functions in humans and other species are known. In addition to a large number of groups all over the world, there are about a dozen groups in our country involved in Macromolecular crystallography, a major centre being at the Indian Institute of Science. We have a national X-ray Facility for Structural Biology to collect diffraction data from the crystals of macromolecules which can be used by crystallographers in the country. Since the 1990's structures of dozens of biomolecules have been studied by these groups, which include proteins from humans, plants, the bacterium causing TB, the parasite that is transmitted by mosquitoes and causes malaria, the virus that causes diarrhea, viruses from plants and DNA. More resources and access to sophisticated facilities like the synchrotron radiation will enhance the research activities in this field in India.

Conclusion

The structures of biological molecules provide enormous information at the level of the individual atoms and tell us how exactly the molecules function and interact with other molecules. They are used to design small molecules such as drugs to fight diseases. Many more structures have to be determined to understand the vast variety of biological processes and macromolecular crystallography continue to have an impact on modern biology. The structures are available in public domain data bases. Anyone with a PC and an internet access can view these amazing molecules and their various features in colour. Recent advances in computer graphics allow us to peep into any molecule whose structure is available and study it in detail. X-rays have allowed us access into this fascinating micro-, or in today's terminology, nano-world.

Address for Correspondence

K Suguna

Associate Professor

Molecular Biophysics Unit

Indian Institute of Science

Bangalore-560 012

Email:

suguna@mbu.iisc.ernet.in

