

## Status and prospects of macromolecular crystallography

M R N MURTHY

*Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India.  
e-mail: mrn@mbu.iisc.ernet.in*

Evolution of a field in the days to come is an extrapolation of its yester years. This extrapolation is at best applicable to the immediate years to come. The very nature of science is such that any long term extrapolation is doomed to fail.

X-rays were discovered by Roentgen in the year 1895. The nature of this radiation was far from clear at the time of its discovery. Its properties, however, suggested that it is likely to be electromagnetic radiation of very short wave length. The experimental verification awaited the insightful observation of Max von Laue, who suggested that such short electromagnetic radiation should be diffracted from crystalline materials due to the periodic arrangement of molecules in crystals. Laue worked out the equations relating the direction of the incident X-rays, crystal unit cell parameters and orientation and the directions in which X-rays are scattered by constructive interference. This was verified almost immediately by Friedrich and Knipping. A couple of years later, W L Bragg suggested a simpler way of studying the interaction of X-rays with crystalline array of molecules. The father and son William Bragg and Lawrence Bragg also realized that it should be possible to determine the structure of simple inorganic crystals using X-ray diffraction. Before the second decade of the twentieth century, it was established that X-rays will provide a powerful means to probe the structure of the atomic world. However, in the beginning, determination of even simple structures was a formidable task. The initial structures determined were those of inorganic crystals, where the structures could almost

be 'imagined' from the atomic compositions and X-rays provided a means of verification. During the next twenty years, X-ray diffraction was used to determine the structures of small organic molecules. These were awfully complex exercises and demanded crystallographic expertise and extensive manual calculations.

The main difficulty in the structure determination by X-ray diffraction was (and continues to be) the so called 'phase problem'. The structure of the molecules that constitute crystals is contained in the X-ray diffraction data. It is a straightforward calculation, although numerically very intensive, to derive the structure if the amplitudes and phases of X-ray reflections are known. While experimental methods allow accurate determination of diffraction amplitudes, their phases cannot be directly recorded. In the initial phase of X-ray structure analysis, it was extremely difficult to determine the phases and hence structure determination was an arduous task. There were other difficulties as well. X-ray machines could provide abysmally low intensities compared to today's standards, measurement of accurate intensities of X-ray reflections obtained from weak X-ray sources was difficult and no computers were available for the numerical work needed.

In the mid thirties, Bernal and his student Crowfoot recorded good diffraction data from crystals of the protein pepsin demonstrating for the first time the possibility of determining the structures of biological macromolecules by X-ray diffraction. However, no method was available for the estimation of phases of reflections from protein

**Keywords.** X-rays; crystallography; biological macromolecules; three-dimensional structure; structural biology; perspectives; trends; prospects.

crystals. In the fifties, Max Perutz and his colleagues developed the 'isomorphous replacement method' for the determination of phases of protein reflections. The method was based on comparative analysis of the X-ray diffraction intensities from a protein crystal and several of its isomorphous derivatives, where a heavy atom is substituted or added at specific sites of the protein without disturbing the crystal structure. Using this method, the structures of myoglobin and haemoglobin were determined in the early sixties by Kendrew and Perutz, respectively. These structures were nothing like structures of inorganic and simple organic molecules. They were bewilderingly complex. Only over decades, the architectural details of proteins began to be appreciated. With the work of these pioneers, X-ray crystal structure analysis of biological macromolecules had become a reality.

In parallel, several mathematically oriented crystallographers, among them notably, Herbert Hauptmann and Jerome Karle, had realized that the atomicity of molecular structures and the fact that electron density is always positive could provide very powerful conditions for the calculation of phases directly from the intensity data of a crystal provided the crystal diffracted to atomic resolution (significant intensities at large angles to the incident X-ray beam, typically  $\sim 42^\circ$  for  $\text{MoK}\alpha$  radiation of wavelength 0.71 Å). The application of their ideas and equations culminated in the current powerful 'direct methods', making structure determination of small organic molecules a routine technique that could be completely automated in most cases. These methods are not yet applicable to proteins as the quality of diffraction from protein crystals is severely limited due to the inherent disorder in these crystals. However, direct methods have indeed been successfully used to determine the location of heavy atoms in isomorphous derivatives, a necessary key step of phase determination in this method.

The classical technologies for the collection of X-ray diffraction data, the Weissenberg and precession methods based on photographic plates as well as detector methods based on counters were inadequate to record the thousands of reflections that are simultaneously scattered by crystals of macromolecules. 'Screenless' methods that circumvent the low efficiency of the classical methods were developed in the seventies. The 'screenless' photographs were extremely complex and processing photographs to derive X-ray intensities was a tedious manual task. It was also necessary to 'know' the crystals and mount the crystals in a predetermined orientation for X-ray data collection. Determination of a protein structure needed skill, patience and several years of hard labour.

Synchrotrons became available as extremely powerful sources of X-rays for protein crystallography in the late eighties. These mammoth machines provide X-rays with fluxes millions of times more intense than laboratory sources. On these expensive machines, it was impractical to waste time to orient the crystals for data collection. Hence crystals were exposed to X-rays in unknown orientations. There was a desperate need for methods for the automatic determination of crystal orientation using the reflection spot positions on the X-ray diffraction photographs and for automation in the derivation of accurate X-ray intensity data. Extremely clever and powerful methods for this 'autoindexing' were developed in the late eighties and early nineties. In parallel, there were impressive theoretical developments in every aspect of crystallographic theory. Synchrotron also allowed use of multiple wavelengths for data collection that could be exploited for phase determination without resorting to several heavy atom derivatives for structure determination. Also, computers with extremely large storage capacity and high computing speeds became available at very low cost. These developments made macromolecular crystallography much like organic chemical crystallography, a task that could be almost completely automated.

Initial studies on proteins were mostly on those that are available in vast quantities, proteins that could be easily purified and stable. With the development of molecular biology, it became possible to clone and express almost all proteins and obtain in quantities sufficient for structure determination. Therefore, structures of all families of proteins were attacked. From a few humble protein structures in the seventies, the number of structures determined so far has soared to over fifty thousand. The major challenge in macromolecular crystallography today is obtaining good crystals that diffract to high resolution. This has still eluded systematic development.

There are three schools of macromolecular crystallographers today. The first group is the 'structural genomics' group, which aims at determining the structures of as many proteins as possible from a selected organism, with the hope that the results obtained will greatly contribute to human health and welfare and understanding of biochemical processes. These 'structural genomics' initiatives are also responsible for introducing a high degree of automation in macromolecular crystallography. Cloning, crystallization, screening for usable crystals, crystal mounting, data collection and processing, and to a large extent structure determination steps have been automated. The second group tackles very large macromolecular complexes. In living cells, often, it

is not individual molecules, but a collection of a large number of molecules that are non-covalently although strongly held, that are functionally significant. Determination of these huge structures is still a daunting task and has to be carried out by expert crystallographers. The third group consists of those who have continued the earlier tradition of carrying out intensive studies on individual proteins or groups of related proteins with the aim of elucidating structure-function relationships. The contributions of all three groups are important and they are mutually complementary. It is now well established that protein function is a result of its precise three dimensional structure, the juxtaposition of chemical groups in space. The standard practice, at least in the West, is to carry out initial characterization of crystals in the home source and make a trip to synchrotron for the rest of data collection.

With this background of development and current status of a magnificent field, we could try to speculate on the future perspectives. The genomic initiatives that have been started will lead to large collections of protein structures that will deeply influence our understanding of cellular function. The new structures that will become available due to these structural initiatives will provide opportunities for a generation of structural biologists to work out the significance of the structural data. Crystallographers will play an increasingly prominent role in providing structural information at more complex levels of biological organization. Structures of several 'molecular machines' that are likely to be available in the near future will provide new and deeper insights into cellular function. In this regard, the combination of X-ray crystallography and electron microscopy will also be very significant. For several years to come, it appears that the work of 'small lab' crystallographers working intensively on the structure and function of individual proteins will also continue to be important. Such results are essential for understanding protein

structure, function, stability, evolution and use of proteins as tools for industrial and pharmaceutical applications.

Although at present massive synchrotrons are extremely important for macromolecular crystallography, it might change in days to come. It might be possible to develop home sources that provide sufficient flux that will make trips to synchrotron unnecessary. There are also efforts to develop 'table top synchrotrons' that could be installed in individual institutions at a fraction of the cost of synchrotrons. Finally, there is the possibility of free electron lasers becoming available as ideal X-ray sources.

In India, although crystallographic research has always been a prominent activity, and parallels the activity elsewhere, we could discern some significant features. The Indian contribution to crystallographic theory by an earlier generation of crystallographers was very significant. This breed of crystallographers, mostly from physics background, is sadly no more. Present day Indian crystallographers are mostly chemists and biochemists. Therefore, unless a very committed mathematician or physicist appears on the horizon, the Indian contribution to crystallographic theory does not appear to be bright. Biological crystallography was started only in the late eighties in India. Despite this late start, significant structural work has been accomplished in the country. This heavy productivity will for sure continue for several years. Unfortunately, structural studies on large macromolecular systems have not been started at all. This phase could begin if a synchrotron becomes available and research in biological cryo-electron microscopy (which again is sadly lacking) begins in earnest. It would be hazardous to predict if such work will soon begin in the country. One could, however, be confident that in India and elsewhere, structural work on biological macromolecules will continue to be an important component of biological research in the decades to come.