Microbial pathogens – an Indian platform for structure-based inhibitor design

M. Vijayan

Although efforts to initiate macromolecular crystallographic studies in India started in the mid-1970s at the Indian Institute of Science (IISc), Bengaluru and in the late 1970s at the Bhabha Atomic Research Centre, Mumbai, definitive preliminary results in the area at the two centres began to emerge only in the early 1980s. The efforts received a major impetus in 1983 when the Department of Science and Technology, New Delhi decided to handsomely support the Bengaluru centre under its Thrust Area Programme. The Bengaluru centre also came to be recognized as a national nucleus for the development of the area. By the 1990s, work in the area began to spread to different institutions in the country with the support of other granting agencies as well, such as the Department of Biotechnology and the Council of Scientific and Industrial Research, New Delhi. Macromolecular crystallography in India came of age by the turn of the century, although the expansion of the area in the country continued unabated even afterwards. Some of the programmes pursued then and continued to be pursued now such as those on lectins and plant viruses at Bengaluru and on mammalian secretions at the All India Institute of Medical Sciences, New Delhi, have a distinctly Indian flavour. However, it was felt that it was time to address problems which are still more directly relevant to India.

Molecular structural biology of microbial pathogens

Infectious diseases constitute a major problem confronting poor countries, including India. Structural studies on proteins from the microbial pathogens which cause such diseases, are of considerable relevance to the country. That the sequences of the genomes of some of these pathogens, particularly that of *Mycobacterium tuberculosis* which is the causative agent of tuberculosis (TB), became available in the nineties added impetus to such studies. The world over, structural genomics (proteomics) studies were emerging in a highly organized manner. In this scenario, this writer, along with others, orchestrated the need for a national effort on the structural genomics of microbial pathogens[1]. The most important component of this effort turned out to be structural studies on mycobacterial, particularly TB, proteins.

India has a long tradition of mycobacterial research. During the last decades of the twentieth century, molecular biology approaches also began to be applied extensively in this research. The first structural biology work in the area involved the homology modelling of *M. tuberculosis* RecA (*Mt*RecA), carried out as part of a larger effort at IISc[2]. This was soon followed by the brilliant annotation of an important *M. tuberculosis* gene using bioinformatics approaches, carried out at the Institute of Microbial Technology, Chandigarh[3]. The first crystal structure of a mycobacterial protein to be solved in India was that of *Mt*RecA at IISc[4] in 2000. This was then one of the handful of TB proteins of known three-dimensional structure. Around this time, a TB structural genomics consortium, based in the US and supported by National Institutes of Health (NIH), but with worldwide participation, was established. The IISc group was one of its early members. Subsequently, other research groups from India also joined the consortium which in its early years was useful particularly for networking among TB structural biologists.

In the meantime, structural studies on TB proteins were taken up by many other laboratories in India. Currently, X-ray crystallographic studies on mycobacterial proteins, their complexes and mutants are being carried out in about a dozen institutions in the country. The proteins studied include those involved in DNA replication, recombination, repair and modification; transcription and translation; amino acid synthesis, degradation and modification; fatty acid, mycolic acid and peptidoglycan synthesis; biosynthesis of cofactors, prosthetic groups and carriers, and signalling. Work is underway on chaperones/heat shock proteins and toxin–antitoxin systems as well. It turns out that more than 10% of the TB proteins whose three-dimensional structures have been analysed worldwide, have been determined in Indian laboratories[5]. They include many essential and important proteins. Indeed, Indian contributions form an important component of the global effort on the structural biology of TB proteins.

Though not as extensively as in the case of mycobacterial proteins, structural studies on proteins from some other pathogens are also underway in the country. For instance, crystallographic and related investigations on proteins from *Salmonella typhimurium* are in progress in at least three laboratories in India. Focused structural studies on proteins from the malarial parasite are being pursued in a couple of laboratories. Similar efforts on proteins from *Leishmania donovani*, *Entamoeba hystolitica* and a couple of viruses have also gathered momentum. Early efforts in macromolecular crystallography in India have been concerned mostly with proteins from plant and animal sources. Partly through exclusively individual initiatives and to an extent through loosely coordinated efforts, structural work on proteins from microbial pathogens has reached a reasonable level of maturity in the country during the past decade.

Know your enemy

The long-term primary objective of structural studies, as indeed of biochemical and molecular biology investigations, on microbial pathogens is to understand the basic biology of the organisms. With the advent of antibiotics, it was hoped that infectious diseases could be brought under control. However, that was not to be. The organisms rapidly developed resistance to existing drugs and there is need for developing new drugs, against which again resistance is likely to develop. Pathogens like *M. tuberculosis* have been with humanity for millennia and they are unlikely to go away in a hurry. Therefore, we need to wage a long-term battle with them. For that, we

need to understand the organisms as well as we can.

Work on pathogens can also be used to specifically address important biological issues in consonance with the approach described as directed basic research by Chidambaram[6]. The idea is to encourage the scientist to use, say, *M. tuberculosis* as a model system instead of *Escherichia coli* for explaining the phenomenon of interest to her/him. That would then help not only in elucidating the phenomenon of interest, but also for advancing the understanding of the concerned pathogen. The work at the Centre for Genetic Engineering (CGE) established at IISc in the mid-1980s and subsequently merged with the Department of Microbiology and Cell Biology nearly a decade later, provides a good example of this approach. The members of the Centre were free to choose their own research problems, but they were encouraged to use mycobacteria as models. This resulted in a great revival of modern mycobacterial research at the Institute. Those who originally constituted the CGE are now among the leaders of Indian biology. Their contributions to the understanding of the basic biology of mycobacteria have also been enormous. Their role in the development of the structural biology of mycobacterial proteins at the Institute and elsewhere has also been substantial.

In addition to serving the long-term purpose of understanding the basic biology of pathogens, fundamental research, including that involving structural biology, can form a basis for applications such as drug development. It is often good fundamental research and a prepared mind that lead to applications. As Ramakrishnan, a pioneer in modern TB research in the country, along with Chandrasekhar, mentioned in a paper in 1999, 'To sum up, the need to develop new drugs against *M. tuberculosis* remains an important one; and basic research on the biochemistry and molecular biology of the organism is essential if we are to have any hope of doing so.'[7] In the present context, molecular biology encompasses molecular structural biology as well.

Structure-based inhibitor design

The availability of the three-dimensional structures of a number of important proteins from a pathogen provides a platform for the structure-based design of inhibitors as a first step in drug development. Perhaps the best examples of drugs developed through structure-based design are those for AIDS. Most of them are inhibitors of HIV protease or HIV reverse transcriptase. HIV integrase is also recognized as a drug target. Likewise, the known drugs of influenza are inhibitors of influenza virus nuraminidase. Design of inhibitors based on structural information with a view to using them as possible drug candidates, is currently an active area of research. Inhibition of an essential or important biological macromolecule (usually a protein molecule) is only a necessary condition. Many other factors such as bioavailability, toxicity, membrane permeability, etc. need to be considered before an inhibitor is accepted as a possible drug. Furthermore, there are other lengthy and complicated processes involving trials, commercial evaluation, approval by component authorities, etc. In the total scheme of things, the design of inhibitors, though intellectually challenging, is the least expensive component. That also does not call for elaborate organizational structures and is in the nature of normal laboratory research effort.

Need for new or modified paradigms

Rational drug design, including that based on structural information, generally involves identification of a validated target and then discovering through screening or designing a small molecule that interferes with its function. This general approach has yielded rich dividends, but has probably entered the phase of diminishing returns. Perhaps, a more holistic approach is now called for. The practice of combination therapy for TB, for instance, is a step in that direction. In this instance, a few targets are being simultaneously targeted. Going one step further, it is desirable to adopt a holistic approach at the early stages of drug design itself.

It is often difficult to a priori predict which of the proteins in an organism are possible drug targets. Small viruses usually have around a dozen genes in their genomes. In the case of HIV, at least three gene products are treated as drug targets. In the case of *M. tuberculosis*, with around 4000 open reading frames in its genome, the number of possible drug targets could be several hundreds, including some unexpected ones. For example, RecA is a critical protein involved in DNA-recombination, the function of which in eubacteria is primarily DNA-repair. For long it was not considered as a drug target. However, it was subsequently realized that RecA is involved in the development of drug resistance. With hindsight, this role of RecA appears almost obvious. Now RecA is thought of as a possible co-target or adjunct target. Thus, it is not advisable to rule out any important protein as a drug target. The effect of simultaneously interfering with the functions of several proteins is difficult to predict, although systems biology approaches might provide some clues. It has to be experimentally investigated, for which we need inhibitors for a large number of proteins.

The above considerations lead to the suggestion of a plausible approach involving structure-based design of inhibitors. The approach involves the design of inhibitors for a large number of important proteins from an organism, without being too concerned about the essentiality of each individual protein. The choice of proteins could be left to the concerned investigators, without imposing restrictions on the basis of the currently prevailing paradigm. Large-scale design of inhibitors is now technologically feasible and is not far too expensive. Inhibitor design involves biochemistry/molecular biology, structural biology, bioinformatics and organic synthesis. Efforts in each of these areas are becoming easier by the day. In the 1980s, I recall that the X-ray analysis of a variant of a known crystal from of lysozyme was enough to secure a Ph D from IISc and to publish a paper in the *Journal of Biological Chemistry*. Nowadays, many crystallographic theses report several structures each. Likewise, I recall that homology modelling of a protein was a non-trivial operation even in the mid-1990s. It is now almost only a click away. Comparable progress has taken place in other relevant areas as well. Therefore, large-scale structure-based inhibitor design is a feasible proposition. It is also not likely to cost more than what is normally provided by granting agencies for good fundamental science projects. The crux of the approach is to produce baskets of different inhibitors to choose from. Once a few are chosen for further efforts at drug

development involving simultaneous targeting of several proteins, it then becomes a different ball game requiring large organization and funds. That phase does not come under the purview of the present discussion. In any case, drug development is not the only use of inhibitors. They are indispensable tools in biological research. Therefore, designing of inhibitors is an intrinsically worthwhile exercise, quite apart from its utility in drug development.

Conclusion

As indicated earlier, detailed structural information on a large number of proteins from different pathogens is now available in Indian laboratories. This is particularly true in relation to *M. tuberculosis*. The time is now propitious to initiate concerted efforts in the area of structure-based inhibitor design. Efforts with emphasis on validated targets should certainly be encouraged. In addition, it is also important to support the more holistic approaches of the type outlined above. Apart from other things, there is a crying need to develop drugs for infectious diseases, including TB. India now has the competence to contribute substantially to addressing this need. In this note, I have focused on structure-based efforts, as I am particularly familiar with them. Our efforts in this area should involve proven paradigms as well as modified or new paradigms. In the present context, the adage 'let a hundred flowers bloom and let a hundred ideas contend', should guide us.

1. Vijayan, M., *Curr. Sci.*, 2003, **85**, 878–885.
2. Kumar, R. A., Vaze, M. B., Chandra, N. R., Vijayan, M. and Muniyappa, K., *Biochemistry*, 1996, **35**, 1793–1802.
3. Bachhawat, N. and Mande, S. C., *J. Mol. Biol*., 1999, **291**, 531–536.
4. Datta, S. *et al*., *Nucleic Acids Res*., 2000, **28**, 4964–4973.
5. Arora, A. *et al*., *Tuberculosis*, 2011, **91**, 456–468.
6. Chidambaram, R., *Curr. Sci*., 2007, **92**, 1229–1233.
7. Ramakrishnan, T. and Chandrasekhar, P., *J. Biosci*., 1999, **24**, 143–152.

M. Vijayan is in the Molecular Biophysics Unit, Indian Institute of Science, Bengaluru 560 012, India.
e-mail: mv@mbu.iisc.ernet.in

Is 'compiler construction' a dead subject?

Pinaki Chakraborty

It will not be an exaggeration if I say that today we live in the age of computers. We are surrounded by computers and other programable devices. Almost all software programs that run on these devices are written in high-level programing languages like C, C++ and Java. These programing languages allow software developers to specify what the program is supposed to do in a human intelligible form. This property of the high-level programing languages makes it convenient for software developers to write and debug programs. Unfortunately, computers understand none of these high-level programing languages. A computer can only run a program written in its machine language. A machine language is a machine-specific low-level language, and is difficult to understand and use by software developers. So, a special type of software program called compilers is used to bridge the gap between the high-level programing languages and the machine languages. A compiler translates a program written by a software developer in a high-level programing language into machine language.

The first realistic compiler was developed by a team led by John W. Backus in 1957. That compiler translated programs written in the FORTRAN (FORmula TRANslation) programing language into the machine language of the then latest IBM 704 computers. When the developers were commissioned to develop that compiler, they hardly had any idea of the difficulty of the project which they expected to complete within six months. However, they ended up consuming two and half years of time and 18 man-years of effort to complete the project. Their experience taught two important lessons to the computer science community. First, compilers are complex programs and a subject called 'compiler construction' should be formally established. Second, and more importantly, compilers are useful software programs that can revolutionize the art of computer programing. Serious research and repeated development activities over the years have by now standardized the structure and the internal working of compilers. However, both high-level programing languages and computer architectures have been evolving continuously since 1957. Consequently, compilers have been forced to evolve too.

A few years ago while studying at Jawaharlal Nehru University, I once heard a senior professor from another premier university in India, who was delivering an invited talk in a conference in our university, make a passing remark that nobody works on compilers anymore. So, is compiler construction a dead subject? Courses on compiler construction are taught in both undergraduate and post-graduate-level computer engineering programs in most universities in India. However, these courses are often taught in a dry and uninteresting manner with either little or absolutely no laboratory support. Moreover, hardly anybody does research on compilers in India. However, things are quite different abroad, especially in the top universities. There are active research groups working on compilers. Courses on compiler construction are taught based on a programing exercise. This programing exercise is often the largest and the most sophisticated program that computer engineers write in their student life.

There are quite a few reasons for studying compiler construction and researching compilers. A decent knowledge of compilers helps software developers to write programs with desired characteristics like small size when translated into machine language, less running time, better fault tolerance and low power consumption. For programs that will be used many times by multiple