Structural genomics of microbial pathogens – An Indian programme*

M. Vijayan

Structural genomics, simply stated, seeks to determine the structures of all proteins coded by genomes of known sequence, using X-ray crystallography, NMR and bioinformatics. The known principles of protein architecture and the available information on the structural and functional classification of proteins, make this an approachable objective. The early excessive preoccupation with folds has now been substantially overcome. The emphasis is now on the determination of a collection of related proteins coded by a given genome or a set of similar proteins coded by different genomes. Taking advantage of the existing strengths in the country, a national programme on the structural genomics of microbial pathogens is being pursued. A major component of the programme is concerned with proteins from Mycobacterium tuberculosis. Comparative structural genomics of viruses forms another important component. Although not part of the concerted effort, structural studies on proteins from parasites are also gathering momentum. India has reasonably well-equipped laboratories for carrying out the programme. A major lacuna in the effort is caused by the absence of an Indian synchrotron X-ray facility. The results obtained so far in the programme have been encouraging. Particular attention is being paid to marry the requirements of quantity and quality. The overall objective of the programme is to advance our detailed understanding of selected microbial pathogens at the molecular level and to promote applications that flow from it where possible, under the overall umbrella provided by the genomics effort in the country.

DURING the past decade more than a hundred genomes, including the human genome have been sequenced. A majority of them are those of simple prokaryotes, while many are those of complex, multicellular eukaryotes. The sequencing of several hundred more genomes is currently in progress. There are millions of different organisms in the biosphere and each one is endowed with its own unique genome. The number of genes or open reading frames varies widely. Among the genomes with known sequences, that of *Mycoplasma genitalium* is the smallest. It contains about 480 genes. The human genome is believed to contain about 30,000 genes. The main function of most of the genes is to carry information regarding the sequences of amino acids in the proteins they code for. It is not often easy to identify the genes in a given genome sequence, particularly in the case of complex organisms. Elucidation of the biological roles of the proteins coded for by the genes is still harder. An important, and often indispensable, step in the process is an understanding of three-dimensional structures of the concerned proteins. It is said that the determination of the three-dimensional structures of all the proteins encoded by it constitutes the ultimate annotation of a genome. In any case, once proteins are synthesized using the information encoded in the genes, much of the metabolic functions are carried out by proteins. They also constitute a substantial part of the structural components of any organism. Obviously, the biological function or role of any protein is dependent on its three-dimensional structure. Therefore, in order to understand an organism in all its molecular detail, it is important to know the three-dimensional structures of the proteins it contains.

Structural genomics

The genome of each organism typically codes for thousands of proteins and there are millions of different organisms. Thus, several million proteins, each different from any other to different extents, exist in nature. Simply stated, structural genomics aims at determining the three-dimensional structures of all these proteins using X-ray crystallography, NMR spectroscopy and bioinformatics. Stated in this manner, the problem might appear impossible. The problem is formidable, but not as formidable as it might

*Dedicated to Prof. S. Ramaseshan on his 80th birthday.

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

first appear for a variety of reasons. We know, to borrow an expression from Jane Richardson[1], a great deal about the anatomy and taxonomy of proteins. Furthermore, we have a reasonable understanding of the evolutionary relationships among proteins and protein families. This enables us to approach the problem with a reasonable degree of confidence.

It is well-known that polypeptide chains locally fold into helices (a), sheets (b) and loops which have no regular structure. On the basis of these secondary structural elements, protein structures can be divided into three broad classes. Those made up primarily of helices (a-proteins), those containing helices and sheets (a/b-proteins) and those made up primarily of sheets (b-proteins). Proteins in each class can be further classified into different architectural types. For example, the barrel, the sandwich and the roll are different architectural types in the a/b class. In each architectural type, there are usually several folds or domains defined substantially by topological connectivities. These folds are the basic modular units of protein architecture. There are proteins which contain only one domain, but most proteins are made up of two or more domains. A globular unit containing one or more domains can, and often does, form part of a multi-subunit protein. The subunits in a multimeric protein could all be the same or different.

The classification outlined above[2,3] is only approximate. Protein structures obviously did not evolve to enable us to classify them! The border lines between different classes, types and folds are often hazy. Yet, this classification has been found to be useful in understanding protein structure. Structural relationships among proteins in terms of folds often shed light on functional relationships as well. The classification also helps in establishing evolutionary relationships. In terms of folds and evolutionary relationships, proteins have also been grouped into families and superfamilies, with some degree of success.

X-ray crystallography has been, and continues to be, the method of choice for determining the three-dimensional structure of proteins. In recent years, NMR has also been used extensively for this purpose. It is now obligatory to deposit the atomic coordinates, which define the structure, in the Protein Data Bank (PDB)[4], before the results of a protein structure determination are published. In fact, PDB has been among the first databases to be set up[5]. Currently, there are more than 20,000 entries in the PDB. They contain not only independent proteins, but also several ligand complexes and mutants. Even a liberal criterion leads only to about 3500 unique proteins in the PDB. They represent about 700 different unique folds. To what extent has the 'fold space' already been mapped? How many more folds exist in nature? These are questions difficult to answer. The estimates of the total number of folds in nature vary widely.

Much of the early discussion on structural genomics centred around folds. A major goal, if not the main one, was perceived to be identifying all the existing folds. Many, including crystallographers like the present writer, were somewhat uncomfortable with the overemphasis on folds. Folds merely provide a framework. They are important, but there is much more to a protein than the fold. Each protein molecule has to be examined in all its complexity and subtlety. Happily, over the years the emphasis has to a great extent shifted from folds to the structure and function of proteins. What distinguishes the genomics approach from the normal efforts at structure determination is mainly the scale and strategy of the operation. The emphasis in structural genomics is on a collection of related proteins from a given genome or a set of similar proteins from different genomes, than on individual proteins. Furthermore, there is an added emphasis on automation to deal simultaneously with a large number of proteins. The most important first step in the work is the cloning of the desired genes and their expression in large quantities, and the purification of the resulting proteins. In the subsequent X-ray investigations, the initial, and often the rate-limiting step is crystallization. Once suitable crystals are obtained, the available techniques are adequate to obtain structure solution in most situations. However, efforts are underway to speed up X-ray intensity data collection and the crystallographic analysis. New, powerful NMR techniques are also being developed. Structural genomics also relies on bioinformatics approaches at different stages. Such approaches are often used for identifying the genes to be expressed. This is often referred to as target identification. Proteins belonging to the same family have sequences similar to different extents. If the structure of one or more proteins of the family is known, then the approximate structure of the remaining proteins can be constructed using the known structure(s) as the template, employing homology modelling. *Ab initio* prediction of the three-dimensional structures of proteins from sequence information alone, has been a dream of computational biologists. But that still appears to be a distant dream.

The Indian effort

Structural genomics of mycobacteria

Exploratory efforts in structural genomics began in the late 1990s. The first major structural genomics programmes in North America, Europe and Japan were launched in 2000 (refs 6 and 7). During the same period, the first steps were taken towards an Indian initiative in the area. These steps were centred around research on *Mycobacterium tuberculosis*, the causative agent for tuberculosis (TB) at the Indian Institute of Science (IISc), Bangalore. Through discussions among the concerned scientists at IISc, Central Drug Research Institute (CDRI), Lucknow; National Institute of Immunology (NII); New Delhi, and the Institute of Microbial Technology (IMTech), Chandigarh, the outline of an Indian TB structural genomics ini-

tiative emerged. In the meantime, a TB Structural Genomics Consortium, with considerable international participation, was launched in the US. Along with scientists from Europe, New Zealand and other countries, those involved in the Indian TB structural genomics effort also joined the consortium. (In the meantime, the IMTech group moved to the Centre for DNA Fingerprinting and Diagnostics (CDFD, Hyderabad.) This is perhaps the first, and so far the only major international structural genomics initiative in which Indian scientists are participants. The structure of *M. tuberculosis* RecA, solved at Bangalore in a collaborative effort involving the group of Muniyappa, was among the first few structures of TB proteins to be solved worldwide. To date the structures of over 30 TB proteins have been determined using X-ray crystallography. Among these, five are from India.

As indicated earlier, *M. tuberculosis* RecA (MtRecA) was the first TB protein to be X-ray analysed in India[8]. The structure of *E. coli* RecA (EcRecA) and the partial structure of its complex with ADP were the only structural information available on RecA proteins. The difficulty in the crystallization of the RecA protein, presumably on account of its aggregation properties, appears to be the reason for the paucity of structural information on it. In addition to solving the structure of native MtRecA, the Bangalore group determined the structures of several of its nucleotide complexes as well[8,9]. RecA is a multifunctional protein involved in DNA-binding, ATP hydrolysis, LexA-binding and SOS response. RecA molecules aggregate into filaments. In *E. coli*, the filaments are known to further aggregate into bundles. It is known that DNA binds into the central groove of the helical filaments. The regions of the molecule involved in different functions could be clearly identified (Figure 1). It turns out that the filaments seen in the crystals, generated by 6_1 screw axes, are similar to those found in physiological conditions in electron microscopic studies. Therefore, not only the structure, but also the aggregation observed in the crystals is biologically relevant. It is immediately obvious from the crystal structures that the surface of MtRecA filaments contains negative patches, while that of EcRecA filaments is substantially neutral (Figure 2). Furthermore, in the crystal structures there are several inter-filament hydrogen bonds in EcRecA, while there is only one in MtRecA. This provides a rationale for the absence of readily detectable bundle formation of MtRecA in solution.

The nucleotide-binding region in bacterial RecA proteins is substantially conserved in sequence. Yet, nucleotide-binding and ATPase activity are weaker in MtRecA than in EcRecA[10]. The crystal structure provides a ready explanation for this observation in terms of the size of the nucleotide-binding site. The site, particularly the P-loop involved in binding phosphate, expands in MtRecA with respect to that in EcRecA (Figure 3), leading to looser binding of nucleotides by the former. This also points to the possibility of the design of selective inhibitors of RecA.

The fast-growing, non-pathogenic *M. smegmatis* is often used as a model for the slow-growing, pathogenic *M. tuberculosis*. In order to gain further insight into the structure and function of mycobacterial RecAs, the structures of *M. smegmatis* RecA (MsRecA) and several of its complexes were analysed[11], again in collaboration with Muniyappa. MtRecA and MsRecA have more than 90% sequence identity. The aggregation properties and distribution of charges of the surface of the molecular filament are similar in the two proteins. The nucleotide-binding

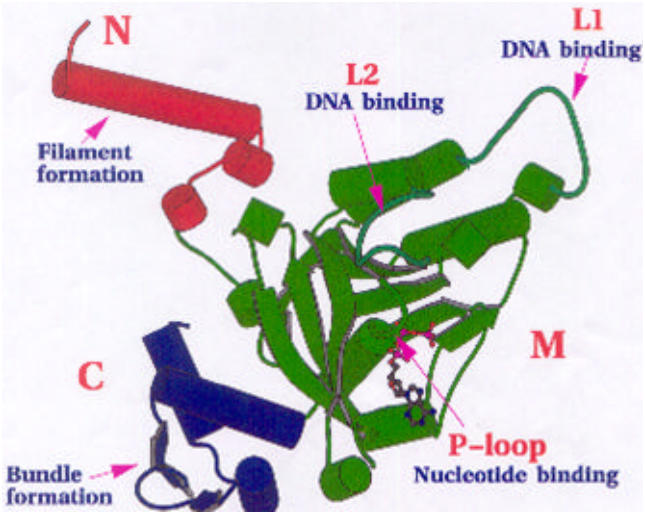

Figure 1. Structure of MtRecA (PDB codes: 1M04, 1M05). The three domains (N, M, C) are shown in different colours. Regions associated with different functions are indicated. All figures, except Figures 2 and 6 *a*, have been generated using MOLSCRIPT[25].

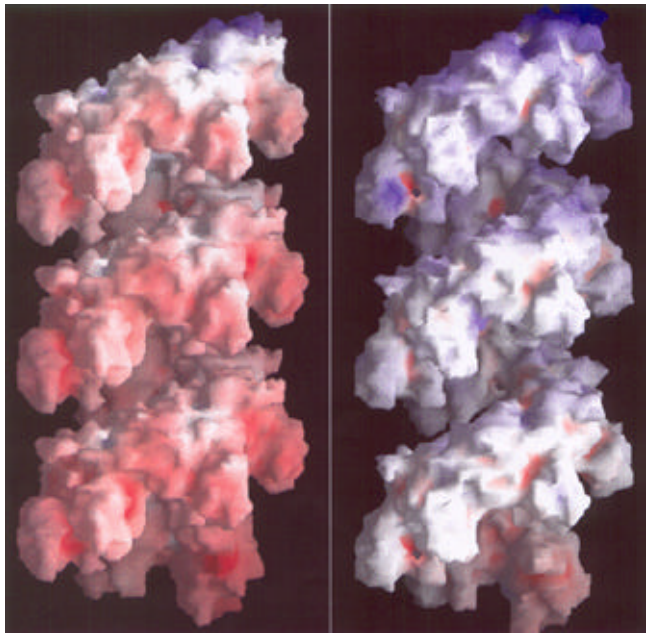

Figure 2. Surface representation of (left) MtRecA and (right) EcRecA filaments (1G19, 2REB). Red and blue indicate negative and positive charges respectively. The figure was prepared using GRASP[26].

site of MsRecA also expands with respect to that of EcRecA. However, differences exist between MtRecA and MsRecA, especially in the case of the dATP complexes, in the disposition of the nucleotides in the binding site. These differences provide a framework for examining the observed differences in the nucleotide-binding properties of the two RecAs[12]. The complexes of both MtRecA and MsRecA show that nucleotide-binding invariably involves the movement of Gln 196 (MsRecA numbering; Figure 4), which is the first residue in one of the two DNA-binding loops. This observation leads to an understanding of the mechanism for the transmission of the effect of nucleotide-binding to the DNA-binding region.

Single-stranded DNA-binding protein is also important in DNA repair and recombination. This protein protects single-stranded DNA, formed transiently during replication, repair, recombination, etc., from nuclease and chemical attacks as well as prevents it from forming aberrant secondary structures. The structure of the protein from *M. tuberculosis* (MtSSB) has been solved at Bangalore[13,14] in collaboration with Umesh Varshney and his colleagues. A subunit of MtSSB, like that of other SSBs, consists of a DNA-binding domain with OB-fold (Figure 5), and a disordered C-terminal domain believed to be involved in interactions with proteins. The DNA-binding domain of MtSSB, however, has an additional strand at the C-terminus. The most noteworthy difference between MtSSB and SSBs of known structure, lies in quaternary association. An SSB tetramer can be described as a dimer of dimers. Unlike in other SSBs, the first dimerization in MtSSB is facilitated by a clamp mechanism involving the C-terminal strand in the DNA-binding domain, referred to earlier (Figure 6 *a*). This and other factors endow MtSSB with a unique quaternary structure which is substantially different from that of other SSBs (Figure 6 *b*). This appears to shorten the length of DNA required to wind around the tetrameric SSB molecule.

The Bangalore group has targetted many more genes concerned with recombination and repair, protein synthesis, cell division and stringent response. Work on several of them is in progress. Expression, crystallization and structure analysis of TB proteins are underway at CDFD, NII and CDRI as well. The genes targetted at these centres pertain to LAM biosynthesis, drug resistance-related proteins, and heat-shock proteins and chaperones (CDFD);

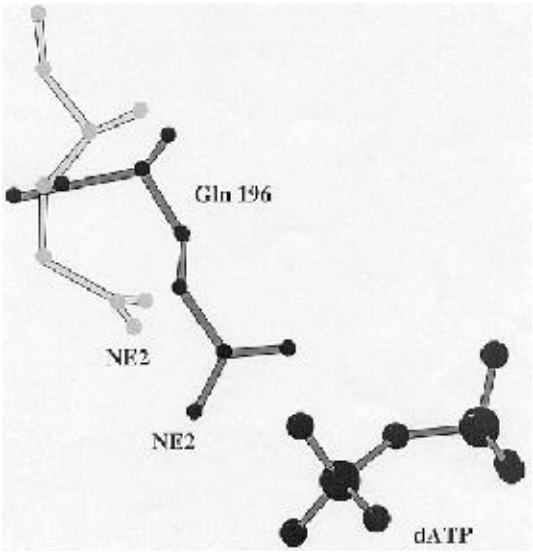

Figure 4. Movement Gln 196 in MsRecA in response to dATP-binding (1UBC, 1UBG). Position of the residue in the native structure is shown in grey, while that in the dATP complex is shown in darker shade. Only the terminal phosphates in the dATP molecules are illustrated.

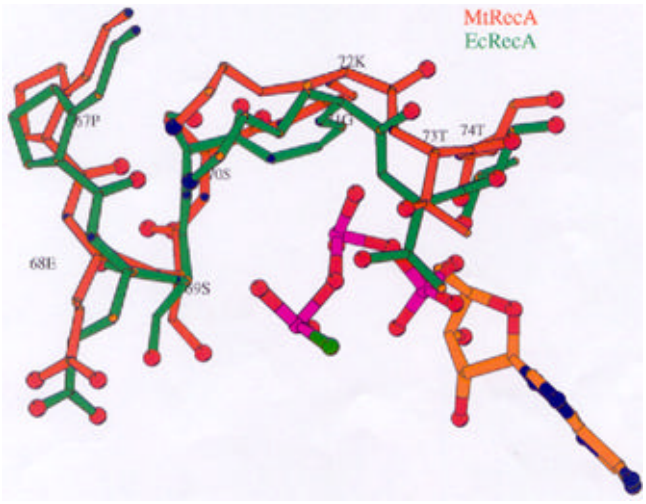

Figure 3. Expansion of P-loop in MtRecA with respect to that in EcRecA (1M05, 2REB). Thicker bonds are used in the representation of bound ATP-*g*-S.

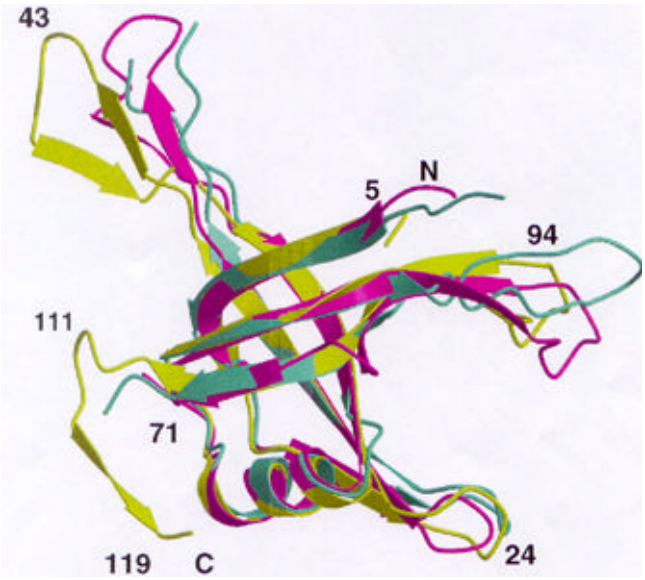

Figure 5. Superposition of DNA-binding domain in MtSSB (yellow; 1UE6), *E. coli* SSB (magenta; 1EYG) and human mitochondrial SSB (cyan; 3ULL). Note the C-terminal strand in MtSSB.

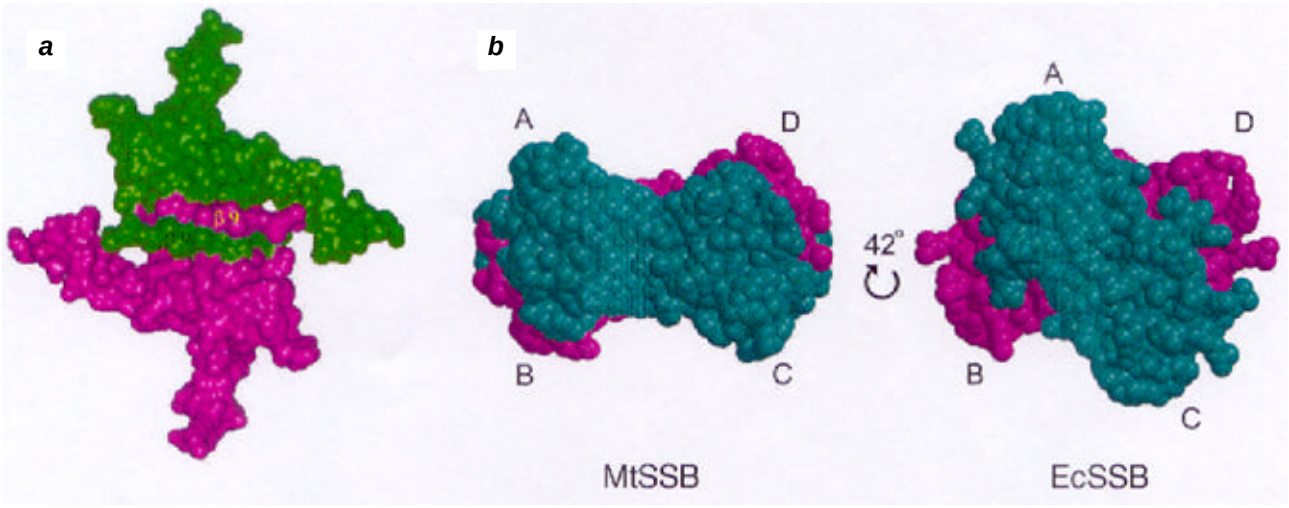

Figure 6. ***a***, Clamp mechanism in MtSSB dimer (1UE6), and ***b***, MtSSB and EcSSB tetramers (1EYG) illustrated as space-filling models. The four subunits are designated as A, B, C and D. ***a*** was generated using INSIGHT II.

polyketide biosynthesis, glycolysis, kinases, proteases and nucleases (NII); and DNA gyrase and lyase, mycolic-acid synthesis and AdoMet-Mtases (CDRI).

The work of Mande and colleagues at IMTech/CDFD has been particularly noteworthy. Their approach has been comprehensive, involving molecular biology, computational and crystallographic techniques. From a structural point of view, the structure solution of chaperonin-10 from *M. tuberculosis*[15] has been an important step in this effort (Figure 7). Yet another important structure to emerge from India was that of the gene product of Rv2118c, solved by the group of Subramanya at CDRI (Figure 8)[16]. The protein was initially of unknown function, but the structure and the subsequent biochemical work showed it to be an AdoMet tRNA methyl transferase.

Comparative structural genomics of viruses

Viruses constitute an important class of pathogens. They are also the largest objects that can be crystallized and X-ray-analysed. The sheer magnitude of the endeavour has meant that crystallographers of only a few countries have ventured to undertake virus crystallography. India is one such country, the others being USA, UK, Sweden and Japan. At Bangalore, Murthy and his colleagues, in association with Savithri's group, have determined the structures of two plant viruses, namely Sesbania Mosaic Virus (SMV) and Physallis Mottle Virus (PhMV; Figure 9)[17,18]. They have also carried out extensive biochemical and crystallographic studies on the structure, assembly and function of these viruses[19,20]. In another crystallographic effort concerned with viruses, Hosur, Kannan and others at the Bhabha Atomic Research Centre, Mumbai have constructed and determined the structure of a tethered dimer of the

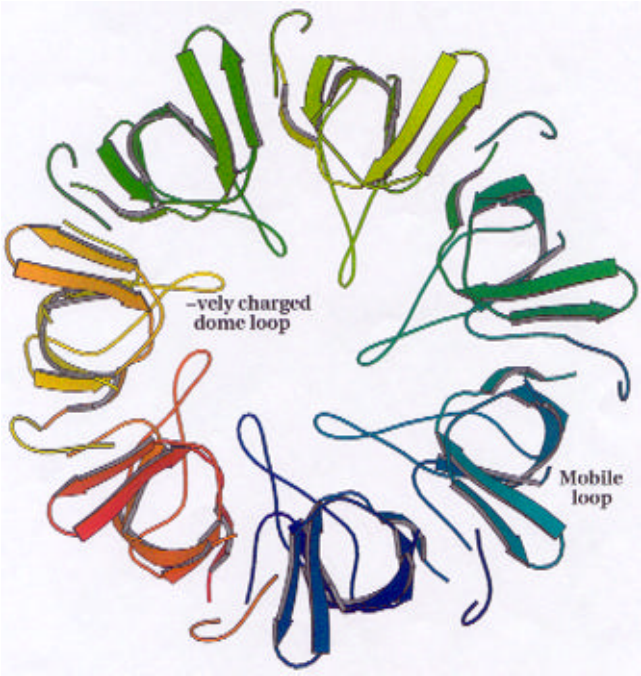

Figure 7. Structure of *M. tuberculosis* chaperonin-10 (1HX5).

HIV protease (Figure 10)[21]. Extensive studies of different kinds using this tethered dimer are currently in progress. Several other studies on viruses are also in progress in different laboratories in India.

Taking advantage of the above strengths, work on viruses also has been brought under the national effort in the structural genomics of microbial pathogens. Viruses have small genomes and they have not received much attention in structural genomics programmes in the West. The Indian programme, conceptualized primarily at Bangalore, involves

the structure analysis of proteins from a set of related viruses or strains, and hence has been described as comparative structural genomics. Comparative structural genomics studies at Bangalore, being pursued by Murthy, Savithri and their colleagues, have registered considerable progress. Other workers are currently attempting similar studies on rotaviruses and morbilli viruses.

Structural studies on parasite proteins

Although not as part of an organized effort, crystal structures of proteins from *Plasmodium falciparam*, a causative agent for malaria, have begun to be determined in India. The first such structure to be solved in the country was that of triose phosphate isomerase from the parasite by Murthy at Bangalore in association with Balaram and Hema Balaram[22]. More recently, Amit Sharma and his associates at the International Centre for Genetic Engineering and Biotehnology, New Delhi have determined the structure of a gametocyte protein essential for the sexual development of the parasite[23]. *Leshmania donovani* is another parasite of particular relevance to India. Rahul Banerjee from the Saha Institute of Nuclear Physics, Kolkata, in association with Alok Datta from Indian Institute of Chemical Biology, Kolkata has recently solved the structure of a cyclophilin from *L. donavani*, the first protein from the organism to be X-ray-analysed[24]. Structure determination of proteins from the two parasites is expected to gather momentum in the years to come.

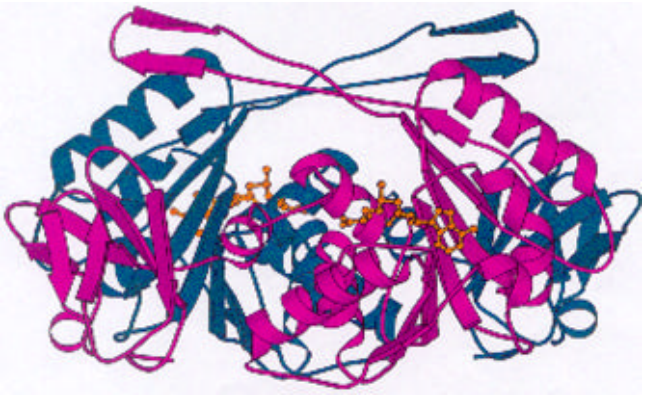

Figure 8. Structure of AdoMet-dependent methyl transferase from *M. tuberculosis* (1I9G).

The approach

A programme on the structural genomics of microbial pathogens is appropriate for a country like India, where diseases caused by them are endemic. It, like any other effort on these pathogens, should undoubtedly contribute, whenever possible, to the development of diagnostics, vaccines and drugs. However, it is important to adopt a longer-term perspective. Microbial pathogens are the only predators of human beings. With the development of antibodies and such other agents, it was thought that we had conquered many of the infectious diseases. But many of the pathogens developed drug-resistance and re-appeared to threaten humanity. In order to fight them on a long-term basis, we need to understand them in every molecular detail. Contributing towards this understanding should be the long-term objective of the programme. More often than not, only mature science leads to useful applications.

Target selection often lays emphasis on proteins which are likely to have novel folds and those whose function cannot be surmised through sequence comparisons with proteins of known structure. This is as it should be, but exclusivity in this regard can be counter-productive. We

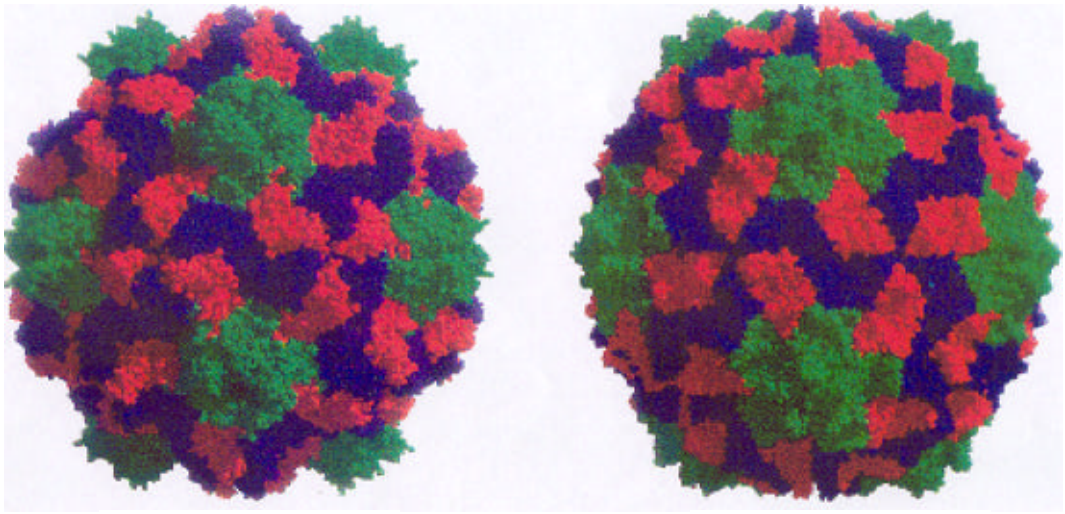

Figure 9. Space-filling representations of Physallis Mottle Virus (1QJZ) and Sesbania Mosaic Virus (1SMV).

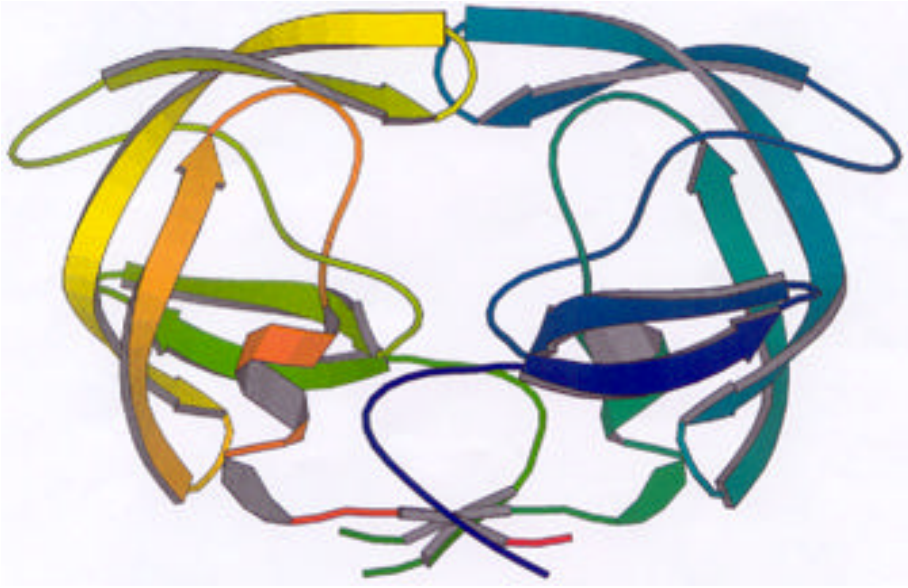

Figure 10. Structure of a tethered mutant of HIV protease (1G6L).

need not blindly follow paradigms developed elsewhere in different contexts, especially those with an immediate commercial objective. It is often small differences in homologous proteins that give rise to differences between organisms. The structures already solved in India under the programme appear to bear this out[9,11,14,15]. In a statistical sense, the structure of a non-homologous protein is likely to yield more biological information than that of one with a homologous partner. But this is not always true. The cumulative effect of a large number of small differences often leads to crucial differences in biological properties. In fact, our current understanding of the structure, function and biological role of most proteins is too incomplete to permit us to be far too choosy. Therefore, the structural genomics programme should aim at dealing with as many appropriately chosen sets of related proteins as is possible.

The molecular structure of a protein is not an end in itself. Structure determination is an essential step in understanding a protein. But in a wholesome effort, it should be preceded and followed by biochemical and other investigations. Successful efforts are often characterized by the synergy between structural and functional studies. The scope of structural genomics should ideally encompass comprehensive studies on sets of related proteins, with structure as the central anchor. Therefore, it is important to select sets in which collaborating biochemists and molecular biologists are deeply interested.

Absence of synchrotron X-ray source – the Achilles heel

Most of the necessary facilities, except a synchrotron source, are in place at different centres. The absence of an Indian synchrotron X-ray source is a major lacuna. Among the recent protein-structure determinations, about 90% have used synchrotron radiation. Despite their proven competence and reputation, the Indian crystallographers are among the deprived 10%. This has begun to seriously affect the competitiveness of Indian macromolecular crystallography. Given the high expenses involved in international travel and logistical problems involved in getting in and out of the country carrying fragile and perishable protein crystals, use of facilities abroad is no real alternative to the availability of an Indian synchrotron source. Discussions on the construction of an Indian synchrotron facility for X-ray diffraction studies, have been on for more than two decades. But progress in this direction has been painfully slow. This is beginning to compromise the effectiveness of the Indian structural genomics effort.

Concluding remarks

The national structural genomics effort has had a reasonable start. The participants in the effort are well-networked, with emphasis on cooperation rather than competition. Given the quality, dedication and expertise of the participating scientists, Indian effort in the area is expected to have a significant impact on the national and the international scene. Despite the existence of several structural genomics programmes in different parts of the world, the content and the objectives of structural genomics are still widely debated. One major issue is concerned with the requirement of determining a large number of protein structures and the thoroughness with which each protein is studied. The endeavour in India has been to marry the requirements of quantity and quality, to produce good

science. The overall objective of the Indian structural genomics programme under discussion, has been to advance our detailed understanding of selected microbial pathogens at the molecular level and to promote applications where possible, under the overall umbrella provided by the genomics effort in the country.

1. Richardson, J. S., The anatomy and taxonomy of protein structure. *Adv. Protein Chem.*, 1981, **34**, 168–339.
2. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. and Thronton, J. M., CATH—a hierarchic classification of protein domain structures. *Structure*, 1997, **5**, 1093–1108.
3. Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C., SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Bol.*, 1995, **247**, 536–540.
4. Berman, H. M. *et al*., The Protein Data Bank. *Nucleic Acids Res.*, 2000, **28**, 235–242.
5. Bernstein, F. C. *et al*., The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 1977, **112**, 535–542.
6. *Nature Struct. Biol.* (*Suppl.*), 2000, **7**, 927–994.
7. *Prog. Biophys. Mol. Biol.*, 2000, **73**, 289–376.
8. Datta, S. *et al*., Crystal structures of *Mycobacterium tuberculosis* RecA and its complex with ADP-AlF(4): Implications for decreased ATPase activity and molecular aggregation. *Nucleic Acid Res.*, 2000, **28**, 4964–4973.
9. Datta, S., Ganesh, N., Chandra, N. R., Muniyappa, K. and Vijayan, M., Structural studies on MtRecA–nucleotide complexes: Insights into DNA and nucleotide binding and the structural signature of NTP recognition. *Proteins: Struct. Funct. Genet.*, 2003, **50**, 474–485.
10. Kumar, R. A., Vaze, M. B., Chandra, N. R., Vijayan, M. and Muniyappa, K., Functional characterization of the precursor and spliced forms of RecA protein of *Mycobacterium tuberculosis*. *Biochemistry*, 1996, **35**, 1793–1802.
11. Datta, S., Krishna, R., Ganesh, N., Chandra, N. R., Muniyappa, K. and Vijayan, M., Crystal structures of *Mycobacterium smegmatis* RecA and its nucleotide complexes. *J. Bacteriol.*, 2003, **185**, 4280–4284.
12. Ganesh, N. and Muniyappa, K., *Mycobacterium smegmatis* RecA protein is structurally similar to but functionally distinct from *Mycobacterium tuberculosis* RecA. *Proteins: Struct. Funct. Genet.*, 2003 (in press).
13. Saikrishnan, K. *et al*., Crystallization and preliminary X-ray studies of the single-stranded DNA-binding protein from *Mycobacterium tuberculosis*. *Acta Crystallogr.*, 2002, **D58**, 327–329.
14. Saikrishnan, K. *et al*., Structure of *Mycobacterium tuberculosis* single-stranded DNA-binding protein. Variability in quaternary structure and its implications. *J. Mol. Biol.*, 2003, **331**, 385–393.
15. Taneja, B. and Mande, S. C., Structure of *Mycobacterium tuberculosis* chaperonin-10 at 3.5 Å resolution. *Acta Crystallogr.*, 2002, **D58**, 260–266.
16. Gupta, A., Kumar, P. H., Dineshkumar, T. K., Varshney, U. and Subramanya, H. S., Crystal structure of Rv2118c: An AdoMet-dependent methyltransferase from *Mycobacterium tuberculosis* H37Rv. *J. Mol. Biol.*, 2001, **312**, 381–391.
17. Bhuvaneshwari, M., Subramanya, H. S., Gopinath, K., Nayudu, M., Savithri, H. S. and Murthy, M. R. N., Structure of sesbania mosaic virus at 3 Å resolution. *Structure*, 1995, **3**, 1021–1030.
18. Sri Krishna, S., Munshi, S. K., Hiremath, C. N., Sastri, M., Savithri, H. S. and Murthy, M. R. N., Three-dimensional structure of physalis mottle virus: Implications for the viral assembly. *J. Mol. Biol.*, 1999, **289**, 919–934.
19. Sangita, V. *et al*., Determination of the structure of the recombinant $T = 1$ capsid of sesbania mosaic virus. *Curr. Sci.*, 2002, **82**, 1123–1131.
20. Uma Shankar, M., Murthy, M. R. N. and Savithri, H. S., Mutation of interfacial residues disrupts subunit folding and particle assembly of physalis mottle tymovirus. *J. Biol. Chem.*, 2003, **278**, 6145–6152.
21. Pillai, B., Kannan, K. K. and Hosur, M. V., 1.9 Å X-ray study shows closed flap conformation in crystals of tethered HIV-1 PR. *Proteins: Struct. Funct. Genet.*, 2001, **43**, 57–64.
22. Velankar, S. S. *et al*., Triosephosphate isomerase from *Plasmodium falciparum*: The crystal structure provides insights into antimalarial drug design. *Structure*, 1997, **5**, 751–761.
23. Sharma, A., Sharma, I., Kogkasuriyachai, D. and Kumar, N., Structure of a gametocyte protein essential for sexual development in *Plasmodium falciparum*. *Nature Struct. Biol.*, 2003, **10**, 197–203.
24. Banerjee, R., Datta, M., Sen, M. and Datta, A. K., Crystallization and preliminary X-ray analysis of cyclophilin from *Leishmania donovani*. *Acta Crystallogr.*, 2002, **D58**, 1846–1847.
25. Kraulis, P. J., MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, 1991, **24**, 946–950.
26. Nicholls, A., Sharp, K. A. and Honnig, B., Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.*, 1991, **11**, 281–296.

ACKNOWLEDGEMENTS. The work reviewed in this article has been carried out by many colleagues mentioned in the text, in different parts in the country. My own efforts in the area have been in collaboration with K. Muniyappa and Umesh Varshney and with the support of Nagasuma Chandra and K. Sekar. Students and postdoctoral fellows who have contributed to the mycobacterial structural work at Bangalore include Sunando Datta, K. Saikrishnan, N. Ganesh, R. Krishna, M. M. Prabu, M. B. Vaze, J. Jeyakanthan, K. Purnapatre, J. Venkatesh and N. Acharya. I thank Sunando Datta, K. Saikrishnan and Pankaja for help in the preparation of the manuscript. The national programme on microbial pathogens has been supported generously by the Department of Biotechnology, Government of India.

Received 28 August 2003