MEETING REPORT

Advances in structural biology and structure prediction*

With the advent of structural proteomics projects – both commercial and academic – and the consequent technological advances of the last decade, many structural biologists feel that their research area has entered a 'golden age'. Structural proteomics consortia and projects around the world have driven technology to allow more rapid and cost-effective structure determination of relatively simple biological macromolecules with the aim of comprehensively covering protein fold space, providing a vast array of near-atomic resolution molecular structures on which to build models and form testable hypotheses for similar proteins of (as yet) unknown structure. In addition, computational and high-throughput screening tools are now routinely used to extract clues to the function of proteins for which only the structure is available. Efficient structure determination has also enhanced the battery of techniques for exploring druggable targets from pathogens. By now, there is a conceptual shift driving a new direction for structural proteomics – the need to understand networks of macromolecules and the functioning of macromolecular machines, to endeavour to close the gap between the classically reductionist structural biology and the emergent field of systems biology.

The Eleventh Annual Symposium of the Association for the Promotion of DNA Fingerprinting and other DNA Technologies (ADNAT11) was convened at the Centre for Cellular and Molecular Biology (CCMB), Hyderabad. Entitled 'Advances in Structural Biology and Structure Prediction', with 21 invited speakers, 68 posters and attended by some 400 delegates from around India and overseas, the symposium explored some successes of high-throughput structural proteomics, specifically as applied to pathogenic drug targets, in addition to more classical hypothesis-driven structure/function analysis of macromolecules of special scientific interest and significance. Further, there were discussions of some of the issues in the new direction of structural proteomics and how traditionally different themes in structural biology research – both experimental and theoretical – must collaborate to fully realize the potential of structural biology as it gradually steps closer to systems biology.

The conference opened with a keynote address from M. Vijayan (Indian Institute of Science (IISc), Bangalore), in which he charted nearly 50 years of structural biology research in India, starting from the earliest work conducted by G. N. Ramachandran on the structural analysis of the collagen triple helix structure and development of the famous and highly relevant tool, commonly called as the Ramachandran plot. Vijayan tracked the history of Indian structural biology from there, including his own personal account of the years of work that went into the eventual establishment of the first Indian macromolecular crystallography facilities at IISc (supported by the Department of Science and Technology, New Delhi) and Bhabha Atomic Research Centre, Mumbai. He recapitulated the rapid growth of Indian structural biology since that time, pointing to a number of key research achievements (both empirical and theoretical) from the thirty or so macromolecular crystallography groups at work in India today. This historical perspective served both as a retrospective for many senior scientists in the audience and as an inspirational account for the younger members.

The conference then switched gear with the first scientific session examining the use of X-ray crystallography in tackling diseases. Results were presented in this session by Tom Blundell (University of Cambridge, UK), Amit Sharma (International Centre for Genetic Engineering and Biotechnology, New Delhi) and Matthias Wilmanns (European Molecular Biology Laboratory, Hamburg, Germany). Blundell mentioned his earlier dilemma of a 'dream or a vision? From gene to protein to drug'. He showcased computational tools and experimental approaches developed over the years in his laboratory and applied by Astex Technology (Cambridge) for using high-throughput crystallography and a library of small molecular fragments to explore 'chemical space' with the aim of producing lead compounds for drug development. He discussed how the close association of fundamental academic science and venture capital-driven industrial science can be highly beneficial for drug discovery and tackling diseases. He elaborated on a major challenge in drug discovery – how to disrupt the extensive surfaces of multiprotein complexes involved in a disease, at present best tackled in a low-throughput mode, and cited the example of the complex of human recombinase Rad51 and breast cancer-associated protein BRCA2 as an exciting target for chemotherapy[1].

Sharma followed with discussions of the threat posed by malaria in the developing world (claiming nearly 3 million lives every year) and showed recent research highlights from his laboratory. Special emphasis was given to cytoadherence and erythrocyte invasion and his studies of Duffy-binding-like domains (DBLs), of which more than 200 are found in the *Plasmodium falciparum* proteome, for example, in the *P. falciparum* erythrocyte membrane protein (PfEMP1). Human-infecting plasmodium strains invade erythrocytes via Duffy antigen receptor for chemokines (DARC), apparently through a strong acid–base interaction, as explored in a seminal study from Sharma's laboratory[2].

Next, Wilmanns described the X-MTB pipeline (www.xmtb.org), a project coordinated by himself and funded by the German Government's Ministry for Science and Education. The mission of X-MTB is 'to help combat tuberculosis by employing a structure-based approach for identifying novel lead compounds directed against specific *Mycobacterium tuberculosis* targets'. Tuberculosis is also major cause of death in the developing world, especially in the case of HIV co-infection and now that strains displaying multi-drug resistance are emerging. X-MTB has to date solved crystal structures of 34 proteins from 242 targets selected, for example, for up- or down-regulation during the *M. tuberculosis* infection cycle. The project has led so far to four chemical inhibitors, currently under intensive *in vitro* and *in vivo* studies. Integral to

*A report on the 11th ADNAT Convention held at the Centre for Cellular and Molecular Biology, Hyderabad during 23–25 February 2007.

the success of the X-MTB pipeline has been the development of the non-pathogenic *M. smegmatis* as an inducible expression host for producing recombinant *M. tuberculosis* proteins in high cell-density fermentation.

The next, rather eclectic, session started with N. Srinivasan (IISc) describing an approach for detecting remote homology recently devised by his group. The approach uses a cascade of computationally generated protein-like sequences to act as 'intermediates' in finding homology between proteins of highly divergent (yet not unrelated) sequences and proves to be extremely effective in detecting similarities between test-set proteins of similar structure, but sequences that initially appear unrelated.

Dinakar M. Salunke (National Institute of Immunology, New Delhi) described his work on antibody–antigen recognition, based on examination of numerous X-ray crystal structures. He has recently explained how a single antibody can promiscuously recognize different antigens by conformational changes in the hypervariable region, thus expanding the repertoire of the primary antibody response[3]. Geoff Barton (University of Dundee, UK) discussed computational analysis of protein–protein interactions. He described a tool developed in his group, the structures, interfaces and alignments for protein–protein interactions database (SNAPPI-DB)[4] to extract more information on protein–protein interactions than is available simply from the crystallographic asymmetric unit and incorporate descriptions of domain structure and organization, and biologically relevant quaternary structures. A second project currently under way in his group uses large-scale experimental analyses (e.g. co-expression and co-localization data) with predictions of domain structure, post-translational modification, disorder, etc. to postulate some 37,000 protein–protein interactions.

The second day of the conference commenced with two lectures on nucleic acid polymerization. Patrick Cramer (University of Munich, Germany) introduced his group's work on elucidating the function of RNA polymerase II in mRNA transcription. The approach is to reconstitute RNA polymerase II with various inhibitors and substrates and take snap-shots of this 2 MDa molecular machine in 'functional complexes' using crystallography. Recent efforts in his laboratory include examining RNA polymerase II activity in transcription-coupled DNA lesion repair (TCR), for example, in the case of UV-induced cyclobutane pyrimidine dimers (CPDs)[5]. They found that CPD-directed misincorporation of uridine into the mRNA stalled transcription. This stalling leaves the DNA lesion inaccesible within the polymerase and a model has been proposed in which the polymerase, the transcript and the lesion bearing-DNA fragment are all removed from the template DNA and the gapped DNA is exposed to repair factors.

Luca Pellegrini (University of Cambridge) discussed DNA repair, recombination and replication. He emphasized biochemical and crystallographic work from his laboratory on the higher eukaryotic nucleoprotein filament forming RAD51 recombinase and its control by tumour suppressor BRAC2 and the orthologous prokaryotic system in which RecX acts as a negative regulator of RecA[1,6].

Next was an interdisciplinary session, with a general theme of translation. Frédéric Allain (ETH Zürich, Switzerland) delivered the EMBO Young Investigator Lecture on control of post-transcriptional gene regulation (through splicing and at the translational level) by mRNA binding proteins, studied in solution by nuclear magnetic resonance (NMR) spectroscopy. His laboratory has identified mechanisms of single-stranded and stem-looped RNA recognition in splicing regulation. Further, structures of free sterile alpha motif (SAM) of yeast Vst1p (a translational regulator) and in complex with the stem–loop RNA Smaug Response Element (SRE) reveal that the RNA is recognized by a shape-specific – as opposed to sequence-specific – mechanism[7].

Osamu Nureki (Tokyo Institute of Technology, Japan) followed with a presentation on the high-specificity chemistry underlying accurate RNA to amino acid translation, as examined by X-ray crystallographic freeze-frames taken at key points in the process. He focused on two main themes: the complete crystallographic analysis of the dynamics of CCA-addition[8], in which six crystal structures have revealed the basis for tRNA 3′CCA terminus maturation, mediated by CCA-adding polymerase, which is able to discriminate between CTP and ATP throughout their sequential addition; and a series of three crystal structures of MnmA thiouridylase–tRNA complexes, exploring RNA sulphuration of anticodon uridine 34 in glutamate, glutamine and lysine tRNAs – a critical modification limiting codon–anticodon wobble to minimize translational errors[9].

Single particle 3D cryoelectron microscopy (cryoEM) was then put in the spotlight by Rajendra Kumar Agrawal (State University of New York, Albany, USA). His laboratory uses cryoEM to study mammalian mitochondrial ribosomes, the machines tasked with producing 13 proteins crucial for ATP energy generation. Comparison of reconstructed 3D images of bovine mitochondrial ribosomes with crystal structures of bacterial ribosomes reveals that the overall protein–RNA quaternary structures are markedly different[10]. Important functional sites, such as the mRNA entry and polypeptide exit sites, appear to be topographically different in the mitoribosome, implying mechanistic differences in protein translation mediated by organellar and cytosolic ribosomes.

In the final session of the second day, R. Sowdhamini (National Centre for Biological Sciences, Bangalore) presented work from her group on identifying and comparing the serine protease compliment of two plant genomes – *Arabidopsis thaliana* and *Oryza sativa*[11]. This systematic *in silico* cross-genome comparison suggests functional groupings of previously uncharacterized serine proteases, which may be of importance when mapped onto other crop plants of commercial and social importance.

Chandra S. Verma (Bioinformatics Institute, Singapore) outlined some recent, as yet unpublished, molecular dynamic simulations on the phosphorylation of ubiquitin ligase MDM2 and how this inhibits binding to, and down regulation of, the oncogenic transcription factor p53. Moving on from purely *in silico* studies to 'visual proteomics'[12], Wolfgang Baumeister (Max Planck Institute of Biochemistry, Martinsried, Germany) described the use of cryoelectron tomography in producing a 'macromolecular atlas of the cell'. He explained that while 'molecular structures are mostly deterministic and therefore repetitive, higher order structures are subject to stochastic variations and, therefore, have unique topologies'. Thus, bridging the gap between well-defined molecular structures and examining the distribution and interaction networks of major higher order machines (e.g. ribosomes, proteasomes, nuclear pores) within a cellular context, relies on

accurate nanometre scale, optimal signal-to-noise, minimized image distortion, cryoelectron tomographic measurements. Following data acquisition, subtle computational interpretation of the tomograms must be carried out. Interpretation of the tomograms requires placing of library 'templates' – models of known macromolecular assemblies derived, for example, from cryoEM reconstructions. This task is complicated by the substantial molecular crowing within cells and organelles. Amongst numerous tomographic examples, Baumeister showed state-of-the-art visual proteomics studies on *Thermoplasma acidophilum*, for which a macromolecular atlas can be constructed using a library of 70 templates and counting.

The third day of the conference started with a session on receptor–ligand interactions. Michael W. Parker (St Vincent's Institute of Medical Research, Victoria, Australia) talked about structure-based insights into the function of granulocyte macrophage colony-stimulating factor (GM-CSF) cytokine receptors and presented a co-crystal structure of a soluble form of the GMRα receptor in complex with GM-CSF, yielding insights into the functional properties of cytokine receptor systems.

Will A. Stanley (CCMB, Hyderabad) presented the synergistic use of a variety of biophysical and structural analyses as applied to understanding how Pex5p, the major receptor for protein import into the peroxisomal matrix, recognizes the type 1 peroxisomal targeting signal, PTS1[13]. Pex5p explores a set of loose, 'snail-like' conformations, while PTS1 is highly flexible in solution. When the two meet in the cytosol, Pex5p locks-down into a 'ring-like' conformation, burying and rigidifying PTS1 in a thermodynamically passive intermolecular recognition. This facilitates subsequent steps in PTS1 translocation from the cytosol to the peroxisome lumen.

The next session addressed the aspects of bacterial pathogen–host interactions. K. Swaminathan (Institute of Molecular and Cell Biology, Singapore) explored typhoid virulence, showing new data on a type IV pilin, PilS, from *Salmonella typhi*, which has recently been crystallized in his lab[14] and described how it specifically targets the cystic fibrosis transmembrane conductance regulator. Debasish Chattopadhyay (University of Alabama, Birmingham, USA) talked about how Pneumococcal surface protein A (PspA) of the Gram-positive pathogen, *Streptococcus pneumoniae* thwarts the antimicrobial action of host lactoferrin by direct electrostatic interaction, as demonstrated by a (yet to be published) co-crystal structure. S. Ramaswamy (University of Iowa, USA) showed data on the activity of the Nod1 receptor, and the structure of its caspase activation and recruitment domain[15]. Nod1 is a cytosolic receptor able to detect bacterial enteroinvasion and consequently activate pro-inflammatory and pro-apoptotic pathways.

To wrap up the conference, Derek T. Logan (Lund University, Sweden) described the allosteric substrate specificity regulation found in ribonucleotide reductase (RNR)[16]. Crystallographic studies of class II RNR from *Thermotoga maritima* in complex with cognate effector/substrate pairs (e.g. dTTP and GDP) clearly show that a single loop ('loop2') mediates communication between the effector and substrate. Effector binding causes a conformational change in loop 2, promoting recognition of the appropriate substrate and allowing some degree of cooperative effector/substrate binding.

In summary, ADNAT11 gave a broad view of current structural biology, both empirical and theoretical. Many speakers discussed the advantages and advances that structural biology can bring to medical and pharmaceutical sciences, while others talked about the use of structural techniques and approaches to understand fundamental aspects of life sciences, to examine the fine mechanistic details of biological processes that help shape the conceptual understanding of such processes from a wider standpoint. Much was made, both explicitly and implicitly, of the tools now routinely used for high-throughput and high-volume approaches to structural proteomics and integrating near-atomic resolution structural data into functional proteomic networks – and even atlases of the subcellular environment. Aside from the purely scientific aspects of the meeting, a general feeling of energy and enthusiasm from the participants, especially younger students, through discussions at lectures, poster presentations and breaks between sessions was strongly felt and appreciated, as were the cultural events and local cuisine.

1. Pellegrini, L., Yu, D. S., Lo, T., Anand, S., Lee, M., Blundell, T. L. and Venkitaraman, A. R., *Nature*, 2002, **420**, 287–293.
2. Singh, S. K., Hora, R., Belrhali, H., Chitnis, C. E. and Sharma, A., *Nature*, 2006, **439**, 741–744.
3. Sethi, D. K., Agarwal, A., Manivel, V., Rao, K. V. and Salunke, D. M., *Immunity*, 2006, **24**, 429–438.
4. Jefferson, E. R., Walsh, T. P., Roberts, T. J. and Barton, G. J., *Nucl. Acids Res.*, 2007, **35**, D580–D589.
5. Brueckner, F., Hennecke, U., Carell, T. and Cramer, P., *Science*, 2007, **315**, 859–862.
6. Pellegrini, L. and Venkitaraman, A., *Trends Biochem. Sci.*, 2004, **29**, 310–316.
7. Oberstrass, F. C., Lee, A., Stefl, R., Janis, M., Chanfreau, G. and Allain, F. H., *Nature Struct. Mol. Biol.*, 2006, **13**, 160–167.
8. Tomita, K., Ishitani, R., Fukai, S. and Nureki, O., *Nature*, 2006, **443**, 956–960.
9. Numata, T., Ikeuchi, Y., Fukai, S., Suzuki, T. and Nureki, O., *Nature*, 2006, **442**, 419–424.
10. Sharma, M. R., Koc, E. C., Datta, P. P., Booth, T. M., Spremulli, L. L. and Agrawal, R. K., *Cell*, 2003, **115**, 97–108.
11. Tripathi, L. P. and Sowdhamini, R., *BMC Genomics*, 2006, **7**, 200.
12. Nickell, S., Kofler, C., Leis, A. P. and Baumeister, W., *Nature Rev. Mol. Cell Biol.*, 2006, **7**, 225–230.
13. Stanley, W. *et al.*, *Mol. Cell*, 2006, **24**, 653–663.
14. Balakrishna, A. M., Tan, Y. Y., Mok, H. Y., Saxena, A. M. and Swaminathan, K., *Acta Crystallogr. Sect. F*, 2006, **62**, 1024–1026.
15. Coussens, N. P., Mowera, J. C., McDonald, C., Nunez, G. and Ramaswamy, S., *Biochem. Biophys. Res. Commun.*, 2007, **353**, 1–5.
16. Larsson, K. M., Jordan, A., Eliasson, R., Reichard, P., Logan, D. T. and Nordlund, P., *Nature Struct. Mol. Biol.*, 2004, **11**, 1142–1149.

ACKNOWLEDGEMENTS. I thank CSIR, New Delhi for sponsoring a Visiting Scientist Fellowship to work in the laboratory of Dr R. Sankaranarayanan, CCMB, Hyderabad, India. I also the R. Sankaranarayanan (Convener) and K. Guruprasad (Co-convener) of ADNAT11 for organizing the meeting and the invitation to write this report.

Will A. Stanley, ARC COE Plant Energy Biology, MCS Building M316, University of Western Australia, 35 Stirling Highway, Crawley 6009 WA, Australia.
e-mail: will.stanley@uwa.edu.au