BARC's JOURNEY IN STRUCTURAL BIOLOGY

Vishal Prashar*, Mukesh Kumar*, M. V. Hosur and K. K. Kannan
Protein Crystallography Section, Bio-Science Group
Bhabha Atomic research Centre
Mumbai - 400085, India

*Email: vishalp@barc.gov.in; mukeshk@barc.gov.in

Abstract
This article provides a detailed account of BARC's pioneering journey in structural biology research. Beginning in the 1960s with neutron diffraction studies, the focus expanded in the 1970s to include X-ray crystallography of macromolecules. Over the years, BARC has established state-of-the-art facilities to investigate key biological processes at the molecular level through X-ray crystallography. Looking forward, the group is poised to target human disease and environmental challenges through structural biology research.

1. Pioneering Structural Biology Research at BARC

The twentieth century witnessed transformative changes in biological research. Advancements in technology enabled researchers from other disciplines such as physics and chemistry to study life systems, leading to the emergence of several new disciplines. One such discipline was 'Molecular Biology', where different cellular processes are explained in terms of involvement of specific molecules. These molecules, both large and small, chemically interact to drive biological functions. The next level of advancement was the emergence of the discipline 'Structural Biology' which enables biological processes to be understood at the atomic level, thereby rationalizing 'specificity' which is a hallmark of living systems. Unraveling how biological molecules such as proteins, DNA, cofactors, drugs, etc. bind to each other in discharging their functions is key to advancing rational molecular design for use in medicine, agriculture, and bioengineering. In keeping with the tradition of being at the forefront of research, BARC initiated

programs in Neutron Physics Division (NtPD) to characterize structures and interactions of biological molecules at the atomic-level by using physical techniques of diffraction. The radiation used in these diffraction experiments was neutrons and X-rays respectively for structures of small molecules such as amino acids, nucleotides, etc. and large molecules such as proteins, nucleic acids, etc. BARC with its nuclear reactor as a source of neutrons, was in a unique position in the country to undertake the task of neutron diffraction. The neutron crystallography was initiated at Trombay in the early sixties, while the X-ray crystallography of large molecules began later, in the late seventies. Incidentally, BARC was only the second institution in the country to venture into the challenging multi-disciplinary task of determining three-dimensional structure of a macromolecule such as a protein molecule. Some of the key developments in BARC around these activities over the years are described briefly below.

Neutron diffraction techniques at BARC, like those at other institutions world-wide, were in their infancy during the 1960s, with a strong emphasis on building instruments and developing crystallography software. Over the years, the Trombay group has maintained a position at the forefront and employed neutron diffraction to address a wide range of critical scientific questions. Among their notable contributions were the determination of crystal structures of several amino acids using neutron diffraction and the precise characterization of hydrogen bonds, which are crucial for intermolecular recognition and biological specificity.

The infrastructure setup for X-ray diffraction studies of biological macromolecules was made under the leadership of Dr. R. Chidambaram & Dr. K. K. Kannan in mid-1970s. It included: 1) a rotating anode X-ray generator procured from M/S Elliot, UK, 2) an optical scanner for digitizing X-ray diffraction patterns, 3) a precession camera, 4) an oscillation camera (Arndt-Wonacott), 5) a cold cabinet for use in protein extraction and purification, 6) fraction collector and 7) equipment for gel-electrophoresis, etc. The computer software required for macromolecular crystallography was modified to run on the PDP computers. A driver software was developed to drive the computer-controlled microdensitometer used to digitize oscillation films. Later, the software were installed on the Norsk-Data (ND-500)-system of computers that became available at that time in BARC. The graphics software used in the visualization of protein models (FRODO) and also in the interpretation of experimentally obtained electron density maps into atomic coordinates for the protein molecule, were installed on a PDP8-controlled Vector General graphics work station. This was a major advancement over the use of a stack of Perspex sheets for visualization of electron density maps (**Fig. 1**). A computer program called FRDICT was developed to create molecular dictionary for any molecule required by FRODO to enable visualization and real-time manipulation of that molecule in the graphics system. The Vector General computer graphics setup was employed to analyze the electron density of the C-subunit of an insect virus, the Black Beetle Virus. Further, computer codes were written to include stereochemical restraints in the least-squares refinement protocols and the crystal structure of triclinic hen egg-white lysozyme was refined to a low R-factor.

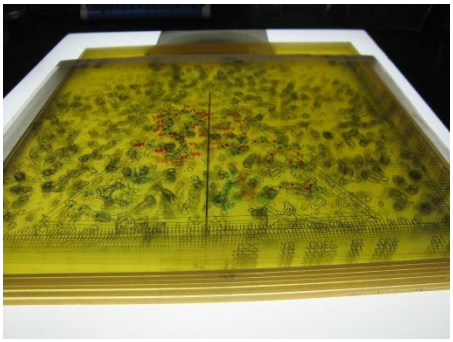

Fig. 1: A typical stack of perspex sheets used to visualize contoured electron density maps in 3-Dspace before the advent of computational graphics (until the 1980s)

2. Expanding Research and Building a National Facility

The first protein pursued at BARC for three-dimensional structural studies was Carbonic Anhydrase (**Fig. 2**), an enzyme that catalyses the reversible reaction between carbon dioxide (CO_2) and water (H_2O) to produce bicarbonate (HCO_3^-) and protons (H^+), which is involved in many physiological processes. Binding of sulfonamide drugs to the human enzyme were explored crystallographically, as a step in the design of better inhibitors of carbonic anhydrase. The crystal structure of carbonic anhydrase complexed with the product bicarbonate was determined, and this work identified active-site residues and also gave insight into the molecular mechanism of this enzyme. This structure identified presence of a hydrogen bond network involving a histidine residue and water molecules as essential for the activity of this enzyme.

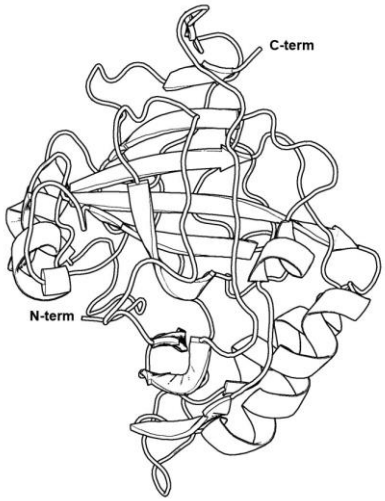

Fig. 2: A cartoon representation the 3D structure of Human Carbonic Anhydrase showing beta-sheet and helical regions

There were several other notable achievements of 1980s and 1990s. The presynaptic neurotoxin-phospholipase Notechis II-5, isolated from the venom of the Australian tiger snake, was the first protein to be crystallized and characterized within India. Further, carbonic anhydrase II was purified from erythrocytes of buffalo blood procured from the slaughter house and crystallized. This was the first enzyme structure determined from India using only the in-house facilities. Since X-ray crystallography of macromolecules is a challenging task because of its multi-disciplinary nature and was being pursued only in very few developed countries at that time, two institutions in India, Indian Institute of Science (IISc), Bangalore and Bhabha Atomic Research Centre, Trombay, were chosen for extra support to pursue this line of research: Therefore, the National Facility for Macromolecular Crystallography was setup at BARC and several experimental facilities needed for this activity were created with partial funding from the Department of Biotechnology, Govt of India. While IISc adopted the stand of creating infrastructure exclusively for diffraction data collection, BARC adopted a different strategy. We preferred to set up a self-sufficient wholesome facility that would cater to all aspects of this work, including the gene cloning, protein expression and purification, crystallization, diffraction data collection and computation. Such an approach allowed us to independently pursue any research project of our interest without seeking extra-institutional collaboration. The equipments procured and installed as a part of the Facility included DNA synthesizer (**Fig. 3a**), High-density fermenter for bacterial cell culture, Controlled temperature incubator, FPLC system with different types of separation columns (**Fig. 3b**), Peptide synthesizer, RAXIS-II imaging plate system for diffraction data collection (**Fig. 3c**) and Silicon graphics work stations for molecular modeling and electron density interpretation. Human carbonic anhydrase, an enzyme naturally found in red blood cells, was recombinantly over-expressed in E. coli bacteria by graduate student Arun Mohanty to produce large quantities for crystallographic studies. Notably, this was the first gene to be cloned in India for crystallographic work, and it was the result of a collaboration between the molecular Biology Division (MBD) and NtPD, BARC. In collaboration with MBD, crystallographic studies of a multi-enzyme complex containing plant RuBisCO was undertaken. Single crystals of the complex, isolated from spinach leaves and harboring RuBisCO bound to RuBP, were successfully grown and characterized. X-ray diffraction data were collected to a resolution of 2.5 Å at the Photon Factory synchrotron in Japan. Structural analysis confirmed the canonical L_8S_8 subunit arrangement of RuBisCO within the complex and revealed well-defined electron density for the other component enzymes and the bound RuBP substrate.

3. Advancing Structural Biology: From HIV to COVID-19 and Beyond

The facilities made available under the National Facility for Macromolecular Crystallography enabled to undertake several other projects including the plant-based toxins gelonin and saporin, HIV-1 protease, drug-resistant mutants of HIV-1 protease, HIV-1 protease substrate interactions, etc. Ribosome inactivating proteins, Gelonin and Saporin, have a potential to be conjugated with suitable antibody to form immunotoxins

for targeted killing of cancer cells. Gelonin was crystallized and diffraction data were collected on the rotating anode X-ray generator. The structure showed two cysteine residues, Cys44 and Cys50, within the protein sequence form a disulfide bond, rendering them unavailable for conjugation with antibodies. Based on structural analysis, a region of the molecule involved in intra-dimer interactions was proposed as a suitable site for introducing a cysteine residue to enable antibody conjugation and the subsequent production of immunotoxins. Saporin, which has a potential to be used as a more effective immunotoxin, was also purified and crystallized.

(a)

(b)

(c)

Fig. 3: Photographs of (a) DNA Synthesizer (1995) (b) FPLC System(1995) (c) RAXIS-II imaging plate system mounted on rotating anode X-ray generator (1996)

HIV-1 protease, a crucial enzyme for the virus's survival, has long been a target for drug development against AIDS. However, capturing the enzyme in action with its natural substrates proved challenging. At BARC, a breakthrough could be made by discovering that the protease "flaps" adopt a closed conformation even without a bound ligand (**Fig. 4**). This finding, combined with substrate soaking and X-ray diffraction on numerous crystals, allowed to obtain the first-ever crystal structures of an active HIV-1 protease complexed with true substrates, not just analogs. This groundbreaking research, published in PNAS, revealed a critical finding: a low-barrier hydrogen bond between the catalytic aspartate residues. Additionally, these studies provided evidence for the tetrahedral intermediate, a key step in the proposed catalytic mechanism. Six reaction intermediates were successfully trapped in protein crystals, enabling the proposal of a detailed mechanism of action for the enzyme. Further, neutron diffraction studies helped elucidate the protonation state of the trapped tetrahedral intermediate. These studies also delved into mutations in HIV-1 protease that arise under clinical conditions, conferring resistance to three major FDA-approved drugs: saquinavir, nelfinavir, and ritonavir. By determining the crystal structures of these drug-resistant mutants, both liganded and unliganded, it became possible to elucidate the molecular mechanisms of resistance. This work provided valuable insights for overcoming these resistance pathways.

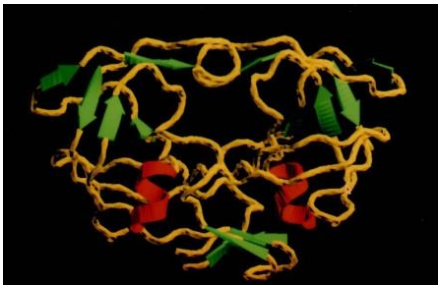

Fig. 4: Ribbon diagram of unliganded HIV-1 protease in 'closed' flap conformation

Subsequently, structural studies were conducted on a variety of other proteins to elucidate their structural and functional details. These investigations have encompassed proteins with diverse biological roles. PSP94, a semen protein associated with prostate cancer, was structurally characterized, revealing a unique dimeric architecture. In the realm of microbial enzymes, structural studies on phosphatases, including PhoN and AphA from Salmonella and SPAP from Sphingomonas, provided insights into catalytic mechanisms and facilitated protein engineering for bioremediation applications. The structure of Translin was determined to explore its potential involvement in RNA binding. To identify potential radioprotective agents, the Kelch domain of Keap1 was structurally characterized. Additionally, the structural elucidation of KatB, a manganese catalase essential for Anabaena survival, was undertaken. The mechanism of action of drFrnE, a protein from Deinococcus radiodurans, was investigated through structural analysis. In the field of photosynthesis, the structures of phycoerythrin, phycocyanin, and allophycocyanin, components of the phycobilisome, were determined to understand light-harvesting processes. Within the context of mosquito control, the structure of Cqm1, a receptor for mosquito larvicidal toxins, was elucidated. The human UVSSA protein, involved in DNA repair, was characterized structurally to map its nucleic acid binding activity. Nanobodies targeting human thyroglobulin were developed and structurally characterized, demonstrating their potential for therapeutic applications.

With the outbreak of COVID-19 pandemic in 2020, it was decided to extend our expertise on HIV protease to the proteases of SARS-CoV-2, like the Papain-like protease (PLro) and the Main protease (Mpro). These two proteases are essential for viral replication and are considered as attractive drug targets. High-throughput screening of a compound library was employed to identify PLpro inhibitors. Notably, compound ATA emerged as a potent candidate, exhibiting low micromolar inhibition of PLpro. Detailed characterization using enzymatic assays, isothermal titration calorimetry (ITC), and other biophysical methods confirmed its inhibitory activity. Furthermore, in vitro studies demonstrated significant antiviral potential against SARS-CoV-2. In vivo evaluation in a Syrian hamster model of SARS-CoV-2 infection revealed that oral administration of ATA significantly reduced viral loads in throat swabs. Independent structural analysis of Mpro crystallized with the peptide substrate revealed covalent attachment of the product peptide in the active site. In a separate line of investigation, Ceftazidime and Sennoside

A were identified as potential inhibitors of RNA binding by the SARS-CoV-2 Nucleocapsid (N) protein. Multi-dimensional NMR spectroscopy confirmed their inhibitory activity and delineated the binding sites. NMR titration experiments demonstrated concentration-dependent chemical shift perturbations, signifying ligand interaction.

The structural biology facilities have been further augmented through the acquisition and installation of advanced instrumentation. The facility's capabilities were initially expanded with the commissioning of a microfocus sealed tube X-ray diffraction system, followed by the installation of a high-flux liquid gallium metaljet X-ray source coupled with a Pilatus pixel detector. To enhance efficiency in the crystallization process, high-throughput screening and imaging systems were integrated. Complementing these structural biology tools, a comprehensive suite of biochemical and biophysical instrumentation was added, including but not limited to isothermal titration calorimeter, differential scanning calorimeter, differential scanning fluorimeter, circular dichroism spectropolarimeter, Fourier-transform infrared spectrometer, surface plasmon resonance system, dynamic light scattering system, and spectrofluorimeter.

Concurrent with the development of in-house capabilities, a dedicated protein crystallography beamline was established at the INDUS-II synchrotron, RRCAT, Indore. This beamline enables high-resolution diffraction studies on macromolecular crystals, including proteins, DNA, and their complexes. The beamline offers precise energy tunability in the range of 5 - 20 keV, facilitating single and multi-wavelength anomalous diffraction experiments. With data acquisition times ranging from 5 to 30 seconds per frame, the beamline serves as a valuable resource for researchers across the nation. Infrastructure for remote usage of synchrotrons abroad for diffraction data collection was set-up at HBNI, Anushaktinagar.

4. Way forward

Building upon the foundation in structural biology, ongoing exploration of intricate molecular mechanisms underlying critical diseases and environmental issues will be continued. The structures and functions of key proteins will be elucidated to develop innovative therapies and sustainable solutions. Studies on SARS-CoV-2 enzymes and Plasmoredoxin are aimed at developing new antiviral and antimalarial drugs, while research on cancer targets MTHFD2 and KRAS focuses on uncovering novel cancer therapies. Additionally, the complex interplay between the microbiome and cancer drug metabolism is being investigated to optimize treatment modalities. Research into the structural properties of enzymes, such as terephthalases, aims to engineer solutions for plastic degradation. These efforts are intended to translate findings into tangible benefits for human health and the environment.