Crystallography beyond Crystals: PX and SPCryoEM

Ramanathan Natesh

Protein Crystallography (PX) or Macromolecular Crystallography has been unarguably one of the great intellectual achievements of 20th century. In fact, it has developed to be a strong base for 'hybrid methods' for macromolecular structure determination by a combination of PX and Single Particle Cryo Electron Microscopy (SPCryoEM). This article describes how PX and SPCryoEM are independent methods, how they contribute as hybrid methods, their complementarity and supplementarity and their current status. This article though not a complete review of PX or SPCryoEM of macromolecular protein complexes, serves as a guideline in understanding the underlying principles.

1. Introduction

Despite the knowledge and popular use of PX techniques in the scientific community, it is less known to the general public. The International Year of Crystallography (IYCr) 2014 is being organized jointly by the International Union of Crystallography (IUCr) and UNESCO, which has taken many measures to popularize crystallography. A UNESCO brochure on *Crystallography Matters* can be found here http://www.iycr2014.org/about/promotional-materials.

One question you might ask is, 'why is knowing the protein structure important?' All living beings are made of cells and there are thousands of proteins that function within and outside these cells. Suppose, there is a protein in your body, which is responsible for maintaining proper blood pressure and is not functioning properly, so your blood pressure rises. By knowing the three-dimensional structure of this protein, one can design inhibitors[1] that will prevent protein from over working and the blood pressure will reduce. This approach forms the basis for Structure-Based Drug Design (SBDD), extensively used by pharmaceutical

Ramanathan Natesh is a Ramalingaswami Fellow – DBT and Assistant Professor at the School of Biology, Indian Institute of Science Education and Research, Thiruvananthapuram. His research interests involve Molecular Structural Biology studies of proteins and its complexes in human health and diseases using two principle techniques viz., Protein Crystallography and Single Particle Cryo Electron Microscopy along with range of other biophysical and biochemical techniques.

[1] Inhibitors are also called as drug molecules that stop the function of the protein by binding to them specifically.

Keywords
Protein crystallography, single particle cryo electron microscopy, hybrid methods, protein structure, macromolecular complexes, image processing.

companies to assist in drug design. There are several method available to elucidate the structure of proteins. These include primarily Protein Crystallography (PX), Nuclear Magnetic Resonance (NMR), Atomic Force Microscopy (AFM), and more recently Single Particle Cryo Electron Microscopy (SPCryoEM). Other cognate techniques like neutron diffraction, mass spectroscopy, small angle X-ray scattering are employed to obtain special structural details.

2. Protein Crystallography

Proteins are made of atoms that join together to form amino acids, which in turn join together to form a long amino acid chain which folds to form a functional protein. PX involves four main steps which are graphically represented in *Figure* 1 and discussed below. Each step is equally important by itself and there is no one step that is more important than the other. An informative movie involving these steps can be found at URL: http://youtu.be/-go6XzLg8wc.

2.1 *Sample Preparation*

Typically, a protein sample preparation will include purification of the protein directly from the source (e.g., fungus, extremophilic bacteria, etc.) or by recombinant protein production which is the

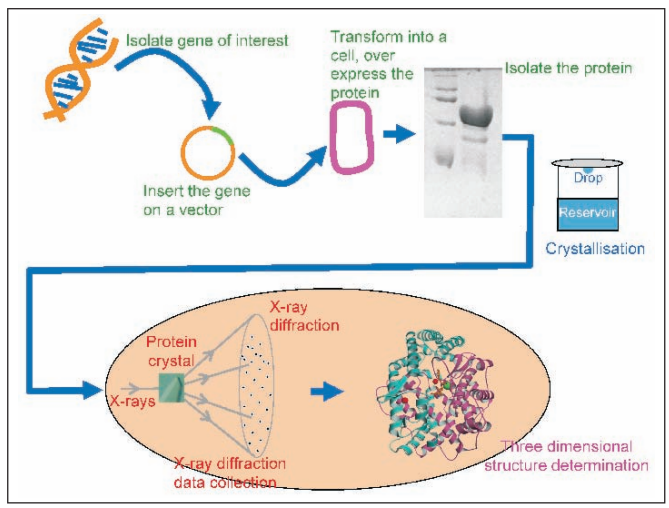

Figure 1. Four major steps in protein crystallography (PX): a) Sample preparation (green); b) Crystallisation (blue); c) X-ray analysis/X-ray diffraction data collection (red); d) 3D structure determination (magenta).

over expression of protein in *E. coli*[1]. The recombinant method, in particular, is more popular nowadays, amongst the many recombinant methods which have been developed in the past few decades. There are some limitations, which are however, beyond the scope of this review.

The protein purity should be better than 95 % as a rule of thumb. 'Isolate the protein' step in the figure shows the purity of the protein expressed on an SDS-PAGE[2] gel.

2.2 *Crystallisation*

Once a pure protein is obtained, it will be extensively screened for crystallisation conditions. To do this, an equal volume (typically in the range of 0.2—2 μL) mixture of protein and reservoir (containing precipitating agent and additives) is placed on a siliconised coverslip and inverted over a reservoir (typically 500 μL) and sealed for vapour exchange to take place (*Figure* 1, blue text). This method, called the 'hanging drop method', is the most famous method among crystallographers; there are other methods like crystallisation under oil (Batch Method), sitting drop method, dialysis method, free interface methods, etc. [6]. Crystallisation processes have gained a boost with the introduction of robotic crystallisation which can set up 96 wells of very small volumes of 0.2 μL in less than a few minutes. This has enabled a large number of trials in a very short time and lesser storage space. In PX, growing a useful protein crystal is a bottleneck and is the most serious problem. Robotic crystallisation has eased this problem a little by high throughput technology to maximize the protein crystallisation targets and to identify the problematic protein for crystallisation.

2.3 *X-ray Diffraction and Data Collection*

Atoms in proteins are separated by distances of the order of 1 Å. On an electromagnetic spectrum, X-rays provide radiation of wavelength around 1 Å. Hence, to see protein molecules at atomic resolution, X-rays and not visible light (used in light microscope[3]) are useful. See *Box* 1 for a short history of the development of

[1] We isolate a gene of interest (for which the protein is to be expressed) from the source by molecular biology techniques, insert it in a vector and transform (by transformation we get the gene of interest riding on a vector into *E. coli* cells) it into the *E. coli* cells to over express the protein of interest and purify the protein by affinity methods (e.g., His-Tag affinity), and by other methods where the separation is based on size.

[2] SDS-PAGE is a biochemical method to separate biological macromolecules, usually proteins or nucleic acids, by separation based on their length and mass-to-charge ratio.

[3] Why cannot one see the protein molecules in 3D using a light microscope? There is a limit to which one can use light microscope. One cannot see objects that are much smaller than the wavelength of visible light (that are used in light microscope). The resolving power of the *microscope* is ultimately limited by the *wavelength* of light (4000–6000 Å for *visible light;* 1 Å = 10^{-10} *meter*).

Box 1. Development of X-ray Sources

Development of in-house X-ray sources has its history and is a topic by itself for physicists. The best available type of an in-house X-ray generator is the 'rotating anode generator' as compared to the earlier 'sealed tube X-ray generator'. Development of synchrotrons (a type of massive machine that can generate extremely high intense light and X-rays) has revolutionised crystallography. Large research facilities housing synchrotrons are used by crystallographers working in areas such as biology, chemistry, materials science, physics, and geology (http://www.iycr2014.org/about/promotional-materials).

Synchrotrons were originally used by particle physicists and X-rays emitted by them were an annoyance for them and were treated as a waste by-product, but soon it was realized that they are of much higher brilliance than obtained by conventional in-house sources. In 1970's when it was shown that they could be used for biological samples and they showed a dramatic enhancement of resolution in diffraction as compared to the best home source. Synchrotrons have evolved since then and in 1980's to 2nd and 3rd generation synchrotron sources with tuneable X-ray wavelength and very high brilliance [3]. Since it is obvious that higher brilliance allows the examination of smaller crystals, the development of Free Electron Lasers (FEL's) which produce billion time brighter X-rays than the existing synchrotron sources in the 21st century heralds a new future [3].

X-ray sources. Protein crystals grown are subjected to X-ray diffraction (*Figure* 1, in red) which gives rise to diffraction patterns (*Figure* 2).

But then, why do we need crystals? Can't we see proteins in solutions using the eyes of X-rays? The answer is that X-ray scattering by a single molecule of protein will be extremely weak and noisy and will be difficult to detect and analyse. On the other hand, in crystals, a huge number of molecules are in the same orientation and the signal gets added up and raised tremendously. The crystal acts as an amplifier. There are, of course, constructive and destructive interferences giving rise to spots that we see in a diffraction pattern. *Figure* 2 shows an actual diffraction from a protein crystal.

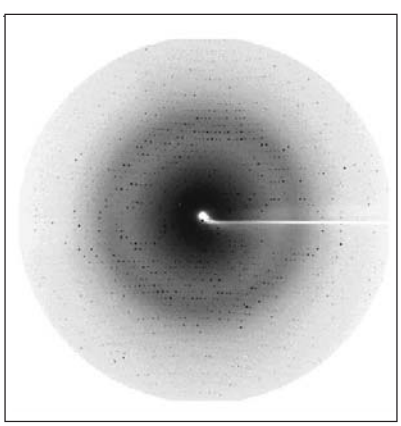

Figure 2. An example of diffraction image taken from an ACE protein crystal at a synchrotron source [1].

A diffracted wave consists of both, *amplitude* (related to intensity) and *phase*. During diffraction data collection, intensities of diffracted spots are detected by sophisticated detectors. However, the phase of each spot cannot be measured and hence, is not known. But this *phase* is so

important, that without it the structure of complicated proteins with thousands of atoms cannot be solved. This is the famous 'Phase Problem' in crystallography. Phases cannot be measured directly, they have to be deduced indirectly.

Step four involves methods for solving the phase problem and structure determination. After data collection, the intensities of spots are indexed and put together in one file with the index (called *h*, *k* and *l*) and intensity, with no phase term associated with it. The final step involves ways to identify the phase of each diffraction intensity.

2.4 *Three-Dimensional Structure Determination*

Protein structure determination starts with an aim to solve the phase problem by using either a model which gives us *a priori phase* information (rough starting values) or using one of the *ab initio* methods without any *a priori* information.

In the first case, which is termed as Molecular Replacement (MR), the model is a protein whose structure is already known and is similar to the protein of interest, based on prior information of sequence similarity. Computational scientists can build rough models in some cases where there is very little sequence similarity. MR can be carried out using programs like EPMR, AMoRe, Phaser, Molrep, MrBump, etc. In all these programs, the basic concept is to place a known model in the crystal (using the data collected for the protein of interest) and compute guesses for phases.

In *ab initio* methods, protein crystallographers use a more complicated method of adding a few heavy atoms to the crystal. These heavy atoms are interpretable from the diffraction data and one can identify their positions in the crystal by direct methods. The contribution of these heavy atoms to the diffraction data is strong and using the positions of heavy atoms in the crystals, one can get the positions of the rest of the diffracting atoms which are from the proteins in the crystal. These methods, which come in various recipes, viz., Multiple Isomorphous Replacement (MIR), Single

Isomorphous Replacement (SIR), Multiwavelength Anomalous Dispersion (MAD), Single Wavelength Anomalous Dispersion (SAD), a combination of MIR and MAD, etc., are used to get the electron density.

Once the phase problem is solved, protein crystallographers get a rough map called the 'electron density map' through a process known as Fourier transformation. A rough model is already available in the case of MR. In the case of *ab initio* methods, the starting model is built from scratch. Model building is carried out using graphics packages like 'O', 'Coot', etc. Initial models are never good and have a lot of errors. From then on, it is a matter of improving the initial model. To do this, the model is fitted properly into the electron density map, and improved using a process called refinement, in which the atomic model is adjusted to improve the agreement with the measured diffraction data. This is calculated as the R-factor which is the average fractional error in calculated amplitude compared to the observed amplitude. A rule of thumb is that a good structure will have an R-factor of $< 25\%$.

An example of a final electron density map at ultra high resolution is shown in *Figure* 3. In the figure, the cyan mesh represents the electron density which gives the position of atoms in the crystal. Atoms are shown as small spheres in green, yellow, blue and red for sulphur, carbon, nitrogen and oxygen, respectively. The bonds between the atoms are coloured according to the atom colours.

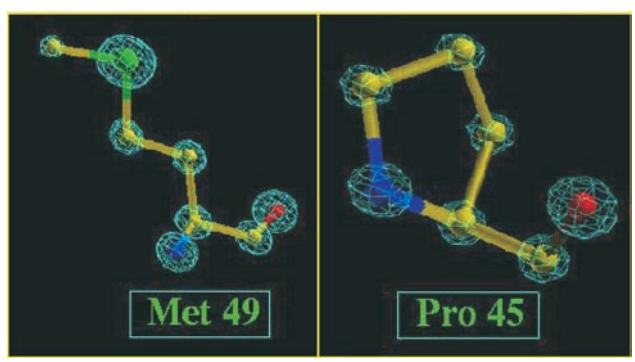

Figure 3. An example of ultra-high resolution electron density map (cyan) for methionine (Met) and proline (Pro) residues from the crystal structure of xylanase [2].

The process of refinement is usually validated during each cycle and after final refinement. Validation is done through programs like PROCHECK, WHATCHECK, MolProbity, etc. The Ramachandran plot, which is used in all the validation programs, is a pioneering contribution by G N Ramachandran[4] [7] and is a good indicator of the quality of a structure. Crystallographers use this plot to verify the refinement and to validate the structure before they finalise it. If all is well after validation, the final coordinates obtained and the data are submitted to the Protein Data Bank (www.rcsb.org). The final solved structures are usually represented as shown in *Figure* 4.

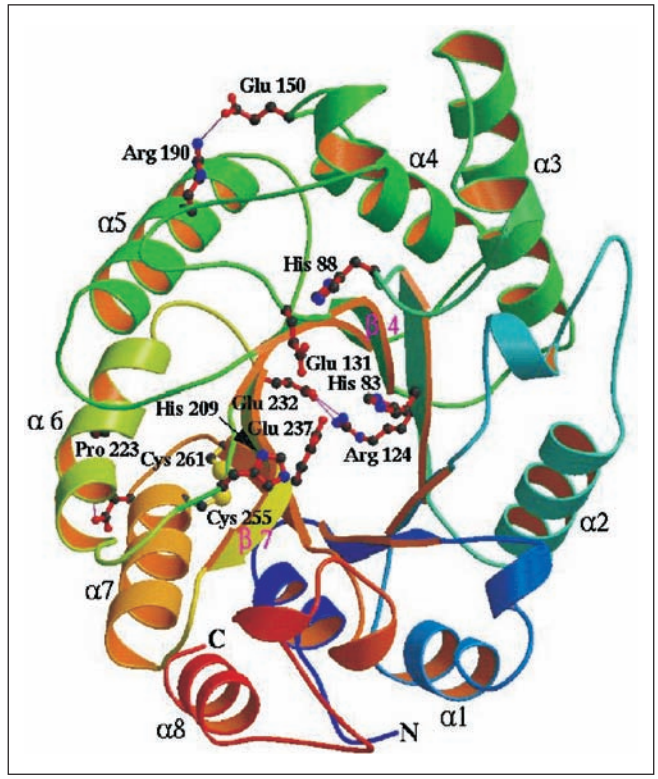

Figure 4. View along the barrel axis of the crystal structure of *Thermoascus aurantiacus* Xylanase. Figure reproduced from [4].

From a small beginning in the 1970's with a handful of structures in the Protein Data Bank (PDB), the number of protein structures deposited with the PDB has crossed 100,000 on 13th May 2014. This is a new milestone for the PDB which happened in this International Year of Crystallography. PX efforts paved the way for the first protein and DNA structures in 1950's. The first of the protein structures to be solved was that of a protein called 'myoglobin' by John Kendrew and Max Perutz, who published the very first high-resolution protein structure in 1958. The work for which had started nearly 20 years before 1958. Kendrew and Perutz received the Nobel prize in Chemistry in 1962 for this work.

PX efforts in India started in the early 1980's and the very first crystal structures were obtained 1994 [9]. Currently, more than 30 institutions in India are carrying out PX studies. A state-of-art synchrotron source in India is yet to come. Indus-2, a synchrotron radiation source built at RRCAT, Indore, is of nominal electron energy of 2.5 GeV and a critical wavelength of about 4 Å. Home-

[4] G N R has made original contributions to the methods of structure determination by X-ray crystallography [8]. See G N R issue, *Resonance*, 2001.

Single Particle Cryo EM in combination with Protein Crystallography is emerging as a powerful tool in visualising macromolecular assemblies at pseudo atomic resolution.

built crystallography beamline BM21 is operational at Indus-2 for PX. Although it is a technology demonstrator for India, it is not quite as powerful as other synchrotron sources around the world [9].

3. Single Particle Cryo Electron Microscopy (SPCryoEM)

Advances in electron microscopy (EM) hardware, low-temperature methods and image-processing software have made CryoEM an important complement to X-ray crystallography and NMR for macromolecular structure determination [10]. Although X-ray crystallography is unarguably still the most popular and successful method in 3D structure determinations, crystallisation of macromolecules is a bottleneck. Homogenous sample preparation is more crucial in protein crystallisation. Also macromolecular structure determination by NMR becomes very difficult for protein larger than 30 to 50 kDa. Developments in the 80's and 90's in SPCryoEM played an important role in bringing CryoEM as a new tool for macromolecular structure determination, apart from the 'classical' structure techniques viz., PX and Protein NMR which can deal with only a subset of larger protein complexes [11]. However, a combination of macromolecular crystallography and SPCryoEM creates a powerful tool that provides informational values greater than the individual techniques. For example, by fitting atomic resolution structure of constituent fragments into CryoEM reconstruction of full macromolecular complexes, scientists can achieve far more accuracy in positioning the atomic structures that exceeds the resolution of the CryoEM reconstruction. SPCryoEM, by itself, also considered a powerful tool with sub-nanometre resolution of CryoEM 3D reconstructions becoming possible for even asymmetric particles nowadays.

CryoEM needs molecules larger than a few hundred kDa to give sufficient signals to locate it in a noisy low-contrast image and then to determine its orientation with sufficient accuracy. For PX structure determination < 150 to 200 kDa proteins or protein complexes are better, whereas, for SPCryoEM, proteins >200

kDa are more suitable. The larger the better it is for SPCryoEM. The small the better it is for PX. Hence, these two methods together as 'Hybrid Methods', can make a very good complementary method to study the whole range of macromolecular sizes right from tens of KDa's to hundreds of MDa's range.

SPCryoEM method involves four major steps as explained below.

3.1 *Sample Preparation*

Sample preparation is similar to sample preparation in PX. Many a time, large macromolecular complexes are expressed and purified as individual constituent fragments and reconstituted (combined in stoichiometric ratio) to form full macromolecular complexes. When the sample is ready, the specimen has to be prepared for data collection as described.

3.2 *Specimen Preparation*

There are two variations of Single Particle Electron Microscopy specimen preparations for macromolecular complexes structure determination. One is the more routine method for negative stain EM and the other one is 'frozen hydrated specimen' or 'vitrified specimen' for CryoEM.

The negative stain EM is simple, quick and provides high contrast, but only the overall shape is revealed by negative contrast and the native structure may be distorted by dehydration [10]. In this method, the complex is placed in a pool of heavy-metal stain on a carbon-support film over a specimen grid (usually copper grid). Then, the stain is slowly blotted with a filter paper and the grid is air-dried to leave a fine coat of metal cast on the complex. *Figure* 5a shows a schematic diagram of a negatively stained

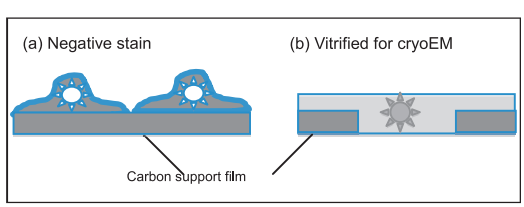

Figure 5. Macromolecular protein complexes specimen (a) negatively stained complex with high contrast lying on continuous carbon film (b) complex lying in a vitrified ice on a holey carbon grid with low contrast. The complex is suspended in a vitrified ice (hole) thus preserving the solution state.

sample. The particle is outlined with good contrast by the heavy metal stain. However, there might be artefacts due the nature of the heavy metal used and also the protein structure may be distorted by dehydration as shown by somewhat flattened particle in *Figure* 5a.

In the case of the frozen hydrated specimen, the specimen preparation is carried out in a holey grid (grid with holes) supported by carbon film as shown in *Figure* 5b. The native structure is maintained and the internal structure is recorded, at the cost of very low contrast. The process of vitrification, which was first standardised by Jacques Dubochet and coworkers, was published for the first time in his classic 1981 paper ([12] and references there in). It is an interesting story of how they wished to make water a glassy material (vitreous) instead of ice crystals for CryoEM data collection.

Figure 6 schematically shows the vitrification method. In this, a drop of solution 3.5 μL typically is applied on the grid and blotted with filter paper to leave a very thin layer (< 2000 Å) just before plunging in a coolant with high heat conductivity, usually liquid ethane [10]. Since the vitrification step is very rapid, it is possible to capture very short-lived structural states by spraying the ligand on to the grid before it is plunged into the liquid ethane. The vitrified specimen is now ready which can then be moved to a storage box in liquid nitrogen for later use in the transmission

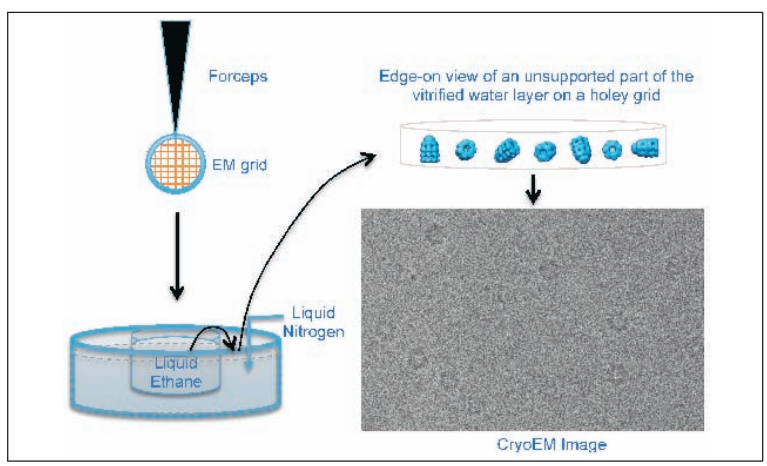

Figure 6. Vitrification of cryoEM specimens. A cryoEM grid with a thin film of solution (< 2000 Å of thickness) is plunged into liquid ethane for vitrification. The frozen specimen is transferred into liquid nitrogen before it is imaged at liquid nitrogen temperatures on a TEM. To the right top is a schematic edge-on view of a part of frozen water layer, with macromolecular complexes trapped in different orientations. Bottom right, part of a cryoEM image showing weak and noisy views of the complexes.

electron microscope (TEM) for data collection. There are other recipes of SPCryoEM sample preparation, e.g., Cryo –ve staining, etc. These are, however, not dealt with in this review.

To achieve sub nanometer resolution in SPCryoEM maps, a 200k kV or a 300 kV FEG TEM is essential.

3.3 *Data Collection*

A standard electron source in a TEM is a heated tungsten filament. Electrons are accelerated by a very high voltage (typically 100 to 300 kV) under vacuum conditions. LaB_6 (lanthanum hexaboride) filament is another better source that produces higher coherence and current density than a tungsten filament. To achieve sub-nanometre resolution, data collection for an SPCryoEM is carried out on a high coherent, bright electron microscope and this is achieved by a Field Emission Gun (FEG) TEM, which is more coherent and >500 times brighter than LaB_6 source, with a very small spread of energies [13].

When required, just before data collection, the grid is transferred from the storage box to a special cryo-transfer holder (at –195°C) placed on a cryo workstation. Then the cryo transfer holder is carefully transferred into the TEM while keeping the cryo shutter (on the cryo transfer holder) closed so as to keep the grid shielded from atmosphere (*Figure* 7). The cryo shutter on the specimen minimizes ice contamination during transfer of grid into the TEM.

Once inside the TEM, the cryo shutter is opened under the stable high vacuum condition of TEM column. Data collection is carried out by highly sensitive CCD cameras or in a more classical way on a photographic film, which has to be digitized. The advantages of a photographic film are its extremely fine pixels and large detection area. Whereas, with CCD, we directly get a digitized image and there is no need to develop photographic film. A high sampling of 7–14 micron/pixel is important to achieve better resolution. One of the biggest problems in CryoEM data collection is the low signal to noise ratio. High efficiency detectors with low noise would be ideal for exploiting the full potential of SPCryoEM. These are addressed in the new genera-

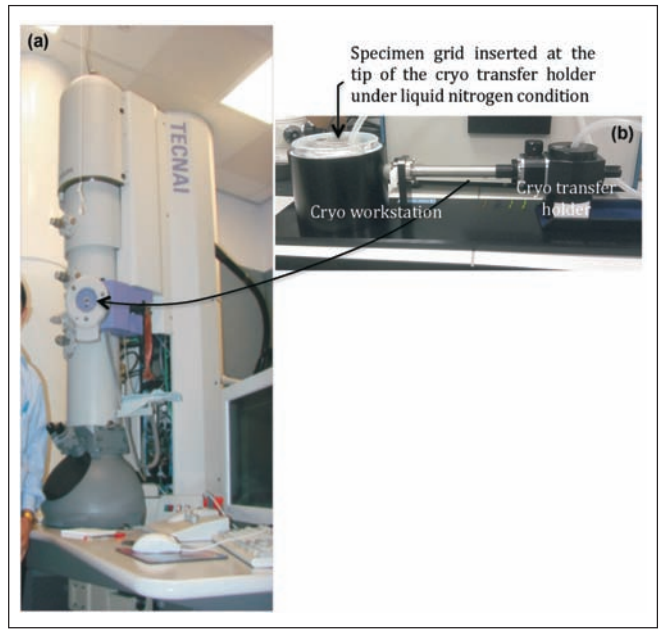

Figure 7. (a) 200 kV Transmission Electron Microscope equipped with Field Emission Gun (FEG). (b) Cryo-holder docked onto cryo workstation. After inserting the specimen on to the cryo-holder (not in scale to microscope), it is carefully transferred to the microscope as shown by the arrow mark.

tion of devices that detect electrons directly (e.g., new direct CMOS detectors). Design of digital cameras in EM continues to improve and the latest cameras are able to register electrons over a broad energy range and cover large areas with smaller pixels so that the detector area becomes comparable to or bigger than EM films [16].

3.4 *Image Processing and 3D Reconstruction*

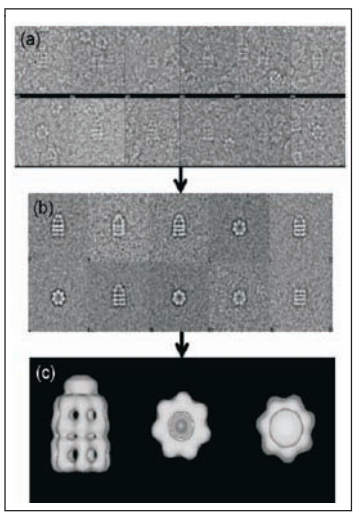

Figure 8. (a) CryoEM image from Figure 6 are cut into boxes and aligned to bring them to the same centre. (b) Class averages of images shows significant improvement in signal-to-noise ratio. (c) 3D model obtained.

A limitation of SPCryoEM in image processing is that, for practical purposes, the macromolecular complex should be > 400 kDa. Having said that, it should be at least greater than a few hundred kDa [14] in order to give a sufficient signal to locate it in a noisy low-contrast image and then to determine its orientation with sufficient accuracy. Early ideas for CryoEM were developed using negatively stained samples. The initial alignment of particles is carried out with a reference image(s) (either from crystal structure coordinates brought down to EM blob resolution or by cross correlation of the images with their rotationally symmetrised average). Alignment with the reference brings them all to the same centre (*Figure* 8a). With a

large enough (few thousand images) data set to give good statistics, 'multivariate statistical analysis' (MSA, [10]) can be used to sort similar images into subsets. Addition of these similar images gives class averages, which provide a significant improvement in the signal-to-noise ratio as shown in *Figure* 8b. Class averages can reveal information about orientations and the heterogeneity present in the data set.

3D electron microscopy image analysis of macromolecular complexes is popular among computational biologists.

A starting 3D model is obtained through two approaches: conical tilting or the use of common-line projections. In the first approach, a rough model is generated through conical tilt, which is an *ab initio* method used to generate reference models from EM images taken at a pair of known angles. The second approach is to find the common lines between pairs of images, which are different projections of the same 3D structure. Class averages described above can be used effectively to generate and compare all possible line projections to obtain their relative orientations ([10] and references there in). This procedure of 'angular reconstitution' is used to obtain a set of initial angle assignments that generate the first 3D model [10]. The 3D model is re-projected into all angular directions to create reference images. Each raw image is then compared with the reference image. The raw images are aligned and given the angles of reference images, which give the highest correlation. At this stage more than one 3D reference can be introduced in order to separate images in the data set into groups corresponding to different components of a mixture. The aligned new set of images are again grouped and averaged to give a new set of improved 3D map. The procedure is iterated until the maps converge to get a final 3D reconstruction(s). An example of such a 3D reconstruction of GroEL-ES-ATP7 is shown in *Figure* 9. Re-projections of the maps are compared with the preceding set of class averages to check for consistency.

The obtained CryoEM 3D reconstruction map of macromolecular protein complexes are usually rigid body-fitted with individual constituent fragments to get a picture of the full complex in the reconstruction. The number of reconstructions solved to resolution maps better than 4Å resolutions have contributed to *ab initio*

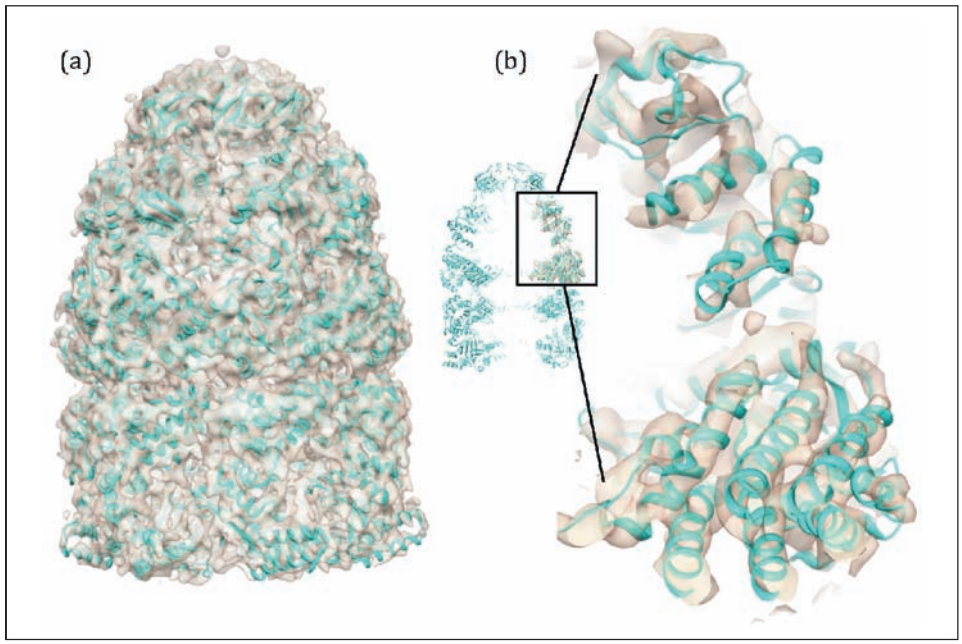

Figure 9. A 3D reconstruction of GroEL-ES in presence of ATP shown in sand colour, fitted with secondary structures shown in cyan. (a) Complete reconstruction of GroEL tetradecamer (798 kDa) in complex with GroES heptamer (70 kDa). (b) A GroES monomer zoom-in of a part of the tetradecamer shown on slice of the tetradecamer (inset) (Figures are generated with Chimera [5] using EMDB entry EMD-1180.

chain tracing from CryoEM map alone ([11] and references therein). In principle, CryoEM is not only restricted to SPCryoEM; there are 2D crystals, helical assemblies, tomography, etc., which are not dealt with in this review.

2D crystallography and helical assemblies are somewhat similar to SPCryoEM and can give better than 4 Å resolution due to symmetry in molecules imaged. However, Cryo Electron Tomography (CryoET) methods are mainly used for cellular and organelle level imaging and 3D reconstruction in near-native states to a resolution of approximately 4–5 nm and have proven to provide amazing ultrastructure pictures of cells. This method is being developed for visualizing the molecular organization within organelles or cells [15]. The present challenge for CryoET is to beat the current limitation of 4–5 nm resolution and reach the 2–3 nm resolution in order to visualize macromolecular complexes *in situ* within the cells ([15], [16]).

From the time of the first –ve stain 3D reconstruction of helical T4 tail structure (DeRosier and Klug, 1968) to the reconstruction from multiple projections of asymmetric systems (Hoppe *et al*,

1974) and reconstruction using unstained biological material (Hart *et al* 1968) to the development of first cryogenic preservation of macromolecular complexes 35 year ago ([12] and references there in) to its present state of better than 4 Å resolution in solution state preserved in vitrious water, SPCryoEM has come a long way. This was possible because of better, high brilliant, coherent FEG gun electron source machines, better detectors like direct electron detectors, improved image processing software and advancement in capabilities like large storage, speed, parallel processing computational hardware. Sub-nanometre resolutions (better than 10 Å) are becoming a more affordable and realisable reality for >400 kDa macromolecular assemblies. A resolution better than 5 Å for macromolecular assemblies with high symmetry viruses is a present-day reality. This drastic improvement in resolution over the last few decades has enabled many scientists to tackle bigger problems in structural biology.

Box 2 briefly describes the Indian efforts and provides the current state of the art which promises a bright future in this methodology. An encouraging fact is that a low resolution (22 Å) CryoEM 3D reconstruction of Vibrio cholerae haemolysin oligomer, the

Box 2. Some Recent Intiatives

Jayati Sengupta who joined IICB, Kolkata in June 2008 is establishing the 300 kV FEG TEM facility currently and is carrying out the test data collection and standardisation at the time this article is being written.

Manidipa Banerjee who moved to IIT Delhi in April 2010 has recently (in Feb 2014) setup a 200 kV FEG TEM running lab.

Ramanathan Natesh from Helen Saibil's lab. Birkbeck college, London, UK (where he was trained in SPCryoEM and image processing on a Wellcome Trust Fellowship), joined IISER Thiruvananthapuram in Aug 2010. Currently he is setting up a PX facility and aiming to set up SPCryoEM and tomography imaging facility at IISER Thiruvananthapuram Vithura Campus (which is under construction).

Partha Pratim Datta (who moved to IISER-Kolkata in Dec 2009) and Kalyan Mitra (who returned to CDRI Lucknow in July 2009) are among the five people who got formal training in SPCryoEM and are setting up their labs in India. Partha and Jayati got trained in Wadsworth centre in USA under R K Agrawal and J Frank respectively, Kalyan from Purdue, in Michael Rossman's lab and Manidipa from Jack Johnson's lab at The Scripps Research Institute, USA.

first CryoEM study from India, has been published [17]. Sub-nanometre resolution (better than 10 Å) 3D reconstruction of SPCryoEM data collected within India is yet to come.

4. PX and SPCryoEM: The Hybrid Method, Differences, Complementarity and Supplementarity

Going hybrid can help to visualize large macromolecular complexes. To illustrate this in simple words, let us imagine that the SPCryoEM 3D reconstruction of GroEL tetradecamer in complex with GroES heptamer shown in *Figure* 9a does not have a crystal structure of the full complex (868 kDa), and the crystal structure of constituent fragment (57 kDa GroEL and 10 kDa GroES) is known. By domain fitting of the crystal structures of constituent fragment 57 kDa GroEL (shown in the black box of *Figure* 9b along with its zoom-in) and 10 kDa GroES (capping as a lid on the top) and applying seven-fold symmetry one can get a pseudo-atomic model of the complete GroEL tetradecamer in complex with the GroES heptamer.

Conformational changes of heptamer GroEL in the top ring and the same sequence GroEL polypeptide chain (shown in red in *Figure* 10) in a different conformation in the bottom ring can be distinctly seen. Such drastic conformational changes within a 3D reconstruction would give many functional details beyond the resolution of the EM maps from the pseudo-atomic model fitted into the SPCryoEM 3D reconstruction. In fact, it can go beyond the static pictures of molecular structures so far achievable. Some examples are given below.

A recent study of dynamics of GroEL describes a flexible fitting of crystal structure of individual components (one constituent fragment of GroEL (*Figure* 9b) into the CryoEM reconstructions. Applying computational analysis to a large SPCryoEM dataset of a GroEL ATPase mutant, Clare *et al* [18] were able to determine six distinct 3D reconstructions representing different GroEL-ATP states. Hence, by going hybrid one can see large macromolecular in action [11].

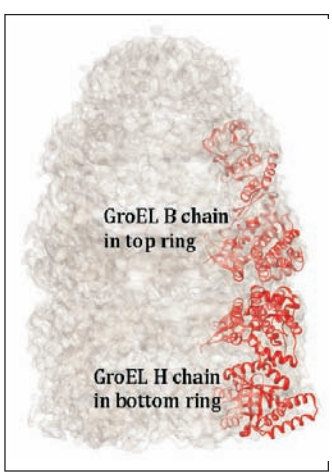

Figure 10. A complete pseudo-atomic model from the 3D reconstruction of GroEL tetradecamer (798 kDa) in complex with GroES heptamer (70 kDa) can be obtained by fitting GroES monomer (shown in red). A marked difference in conformation between GroEL constituent fragment B chain (57 kDa) in top ring and the bottom ring H chain can be seen distinctly (Figure generated with Chimera[5] using EMDB entry EMD-1180.

EM is not free from crystallography. The use of 2D crystallography has been a very useful method to achieve high-resolution structures of membrane proteins, which usually crystallize as 2D crystals. In 1964, Aaron Klug, (Nobel Prize, 1982), showed that the principles of structure determination by X-ray diffraction could be used to develop crystallographic electron microscopy, which can solve quite complex structures, including those of intact viruses. Klug and DeRosier gave to the world the first 3D reconstruction from 2D crystals using CryoEM. Electron crystallography using EM is different from X-ray crystallography because it uses 'images' as primary data, rather than the diffraction patterns. Translated into Fourier lingo, the availability of images means that the 'Phase Problem' known in X-ray crystallography (described in sub-sections 2.3 and 2.4) does not exist in EM. The electron microscope, in Hoppe's words, [19] is a 'Phase Measuring Diffractometer'. Another important factor that 2D electron crystallography is possible with electron microscope rather than X-rays because electrons interact with matter more strongly. Hence, even 2D crystals no larger than few micrometres in diameter provide sufficient contrast to produce, upon averaging, interpretable images. On the other hand, X-ray diffraction requires 3D crystals of macroscopic dimensions for producing statistically acceptable measurements [14].

SPCryoEM also forms a complementary method for those proteins that are less amenable for crystallisation or NMR methods. The output of EM can be used as input to PX and output of PX can be used as input to SPCryoEM, thus providing information feed to other methods supplementing the information to bridge the gap. We have seen how atomic resolution crystal structure of constituent fragments supply information as input to a SPCryoEM map. In a reverse manner, EM can also provide supplementary information to crystallography. EM images can be used as a model for MR. For example, low-resolution phase information from SPCryoEM maps can be used to phase low-resolution crystallography maps and phase extension can be carried out in steps to obtain phases for high-resolution crystallographic native

	PX	SPCryoEM
Specimen	Highly ordered crystals (Protein conformation exist in crystalline lattice)	No need for crystal (The protein complexes are frozen in time and space in solution state in vitrified water/buffer)
Size	No size restriction. Generally the smaller the protein size, the better is resolution	> 200 kDa, preferably > 400 kDa.
Quantity	10 mg/ml	~100 to 1000 fold less protein sample required than X-ray crystallography
Sample	Homogenous sample. Heterogeneous sample impossible	Homogenous Preferred for high resolution. Heterogeneous sample acceptable. The only structural biology method that can provide 3D structures from heterogenous populations.

Table 2. PX and SPCryoEM techniques – a comparison.

data set [20]. In fact, this is not very different from 2D crystallography where you take high resolution diffraction (e.g., 2 Å) images in EM and use low resolution (e.g., 10 Å) images phases to do phase extension to high resolution [21].

At present, crystallography is still the most used method because it gives unprecedented information about the atomic positions. It is an extremely powerful and well-matured method, to the extent that, for crystals that diffract very well to ultrahigh resolution (better than 0.95 Å) it can provide information including hydrogen positions. But crystallography needs highly ordered crystals in order to give a high resolution structure. Some of the statistics of PX and SPCryoEM are briefed in *Table* 2.

5. The Future of PX and CryoEM

Science has no barriers and interdisciplinary will be the mantra for future science. Though this review deals with PX and CryoEM, however, without the combined interdisciplinary effort of structural biologists, chemists, computational biologists, mathemati-

cians, physicists, engineers and architects, none of the progress would have been possible or will be possible in future.

The future of PX lies in the use of this methodology to understand the basic problems in biology. Using PX, new methods like fragment-based drug design are useful to find new ways to develop new medicines in treatment of diseases: cancer, cardiovascular disease, inflammation, infectious disease, neurological disorders, etc. More complex structures and more complicated problems like protein network interactions can be addressed. This is related to map the interaction of proteins or the structure genomics, which aims to solve the structure of all the proteins coded by the genome in all the organisms.

Introduction of free-electron laser is already creating excitement in the crystallography community with the possibility of solving the structure from smallest of crystal ever possible before. Hence it can be foreseen that PX will continue to thrive and provide key structures useful for betterment of mankind.

CryoEM, being a new method, has a lot of potential for the future in terms of its technological development and robustness. Better detectors, 3D reconstruction of small asymmetric particles, more rigorous method for heterogeneous population segregation, quality control and validation along with the dream of sub-nanometre tomographic images of cells will be the ultimate.

Acknowledgements

The author would like to thank Manidipa Banerjee and Jayati Sen Gupta for a critical reading of the manuscript and helpful suggestions. The authors would also like to thank Professor Helen Ruth Saibil for valuable discussions and comments on the manuscript. The author will like to thank Wellcome Trust (UK) for Academic Fellowship, Department of Biotechnology (DBT, India) for Ramalingawai fellowship. The author will like to thank IISER Thiruvananthapuram and its past and present Director, Deans and his colleagues for their unstinted support and stimulating environment.

Suggested Reading

[1] **R Natesh, S L Schwager, E D Sturrock and K R Acharya, Crystal structure of the human angiotensin-converting enzyme-lisinopril complex, *Nature*, Vol.421, pp.551–4, 2003.**

[2] **R Natesh, K Manikandan, P Bhanumoorthy, M A Viswamitra, and S Ramakumar, Thermostable xylanase from Thermoascus aurantiacus at ultrahigh resolution (0.89 A) at 100 K and atomic resolution (1.11 A) at 293 K refined anisotropically to small-molecule accuracy, *Acta crystallographica Section D, Biological crystallography*, Vol.59, pp.105–17, 2003.**

[3] **E M H Duke and L N Johnson, Macromolecular crystallography at synchrotron radiation sources: current status and future developments, *P Roy Soc a-Math Phy.*, Vol.466, pp.3421–52, 2010.**

[4] **R Natesh, P Bhanumoorthy, P J Vithayathil, K Sekar, S Ramakumar and M A Viswamitra, Crystal structure at 1.8 A resolution and proposed amino acid sequence of a thermostable xylanase from Thermoascus aurantiacus, *Journal of Molecular Biology*, Vol.288, pp.999–1012, 1999.**

[5] **E F Pettersen, *et al*, UCSF Chimera—a visualization system for exploratory research and analysis, *Journal of Computational Chemistry*, Vol.25, pp.1605–12, 2004.**

[6] A McPherson, *Crystallisation of Biological Macromolecules: Cold* Spring Harbor Laboratory Press, New York, 1999.

[7] G N Ramachandran, C Ramakrishnan and V Sasisekharan, Stereochemistry of polypeptide chain configurations, *Journal of Molecular Biology*, Vol.7, pp.95–9, 1963.

[8] K Venkatesan, Professor G N Ramachandran's Contributions to X-ray Crystallography, *Resonance*, Vol.6, pp.8–15, 2001.

[9] M Vijayan, Macromolecular Crystallography in India in the Global Context, *Journal of the Indian Institute of Science*, Vol.94, pp.103–8, 2014.

[10] H R Saibil, Macromolecular structure determination by cryo-electron microscopy, *Acta crystallographica Section D, Biological crystallography*, Vol.56, pp.1215–22, 2000.

[11] G C Lander, H R Saibil and Nogales, E. Go hybrid: EM, crystallography, and beyond, *Current Opinion in Structural Biology*, Vol.22, pp.627–35, 2012.

[12] J Dubochet, Cryo-EM—the first thirty years, *Journal of Microscopy*, Vol.245,pp.221–4, 2012.

[13] E V Orlova and H R Saibil, Structural Analysis of Macromolecular Assemblies by Electron Microscopy, *Chem. Rev.*, Vol.111, pp.7710–48, 2011.

[14] J Frank, *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*, 2nd ed. Oxford; New York: Oxford University Press; 2006.

[15] W Baumeister, Electron tomography: towards visualizing the molecular organization of the cytoplasm, *Curr. Opin. Struct. Biol.*, Vol.12, pp.679–84, 2002.

[16] V Lucic, A Rigort and W Baumeister, Cryo-electron tomography: the challenge of doing structural biology in situ, *The Journal of Cell Biology*, Vol.202, pp.407–19, 2013.

[17] S Dutta, B Mazumdar, K K Banerjee and A N Ghosh , Three-dimensional structure of different functional forms of the Vibrio cholerae hemolysin oligomer: a cryo-electron microscopic study, *J. Bacteriology*, Vol.192, pp.169–178, 2010.

[18] D K Clare, D Vasishtan, S Stagg, J Quispe, G W Farr and M Topf, *et al*, ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin, *Cell*, Vol.149, pp.113–23, 2012.

[19] W Hoppe, Electron-Diffraction with the Transmission Electron-Microscope as a Phase-Determining Diffractometer - from Spatial-Frequency Filtering to the 3-Dimensional Structure-Analysis of Ribosomes, *Angew. Chem. Int. Edit.*, Vol.22, pp.456–85, 1983.

[20] E J Dodson, Using electron-microscopy images as a model for molecular replacement, *Acta crystallographica Section D, Biological Crystallography*, Vol.57,pp.1405–9, 2001.

[21] G Wisedchaisri and T Gonen, Fragment-based phase extension for three-dimensional structure determination of membrane proteins by electron crystallography, *Structure*, Vol.19, pp.976–87, 2011.

Address for Correspondence
Ramanathan Natesh
School of Biology
Indian Institute of Science Education and Research (Transit campus)
Computer Science Building
(1st floor), CET Campus
Trivandrum 695016, India
Email: natesh@iisertvm.ac.in, natesh72@hotmail.com