

CryoEM as a complement to current techniques in protein structural analysis

Jack Elands & Werner Hax

FEI Company, Eindhoven, the Netherlands



CryoEM allows scientists to elucidate protein structures quickly and will play a crucial role in structure-based drug discovery.

Protein structure is fundamental to the operation of all biological systems. The completion of the Human Genome Project offers a tremendous opportunity to elucidate structure at the most fundamental level. Although the amino acid sequence of a protein ultimately determines its structure, we know the structures of only a fraction of the proteins encoded by the genome. The determination of these structures, and the elucidation of their functions, is the great challenge of our generation.

Protein crystallization remains a major hurdle for protein structure determination. Nuclear magnetic resonance (NMR) avoids the requirement for crystallization, but is difficult to use with large molecules and complexes. Cryoelectron microscopy (CryoEM) offers an alternative technique that also avoids the requirement for crystallization and, complementary to NMR, is best suited for analysis of secondary, tertiary, and quaternary structure - the levels at which structure-activity-relationships operate. It allows investigation of proteins in their natural state and is sensitive to the dynamics of molecules. Perhaps most importantly, with recent advances in automation, it offers the prospect of much faster structural determination than either X-ray diffraction (XRD) or NMR. It is certainly a powerful, new addition to the structural biologist's toolbox and may significantly shorten the pharmaceutical development cycle through contributions to structure-based drug discovery.

Conventional determination

The primary tools for determining protein structure have been XRD and NMR spectroscopy. Both offer atomic-level spatial resolution, but require relatively large amounts of purified protein. XRD further requires that the proteins be crystallized. Although the determination of structure, once XRD data have been acquired, has become a rapid and routine operation, determination of crystallization conditions remains difficult and time consuming. Many proteins, including some of the most interesting, such as membrane-bound proteins (receptors and ion channels) and those that function to transport materials across the cell or internal organelle membranes, have proven stubbornly resistant to crystallization. Crystallization conditions are typically not physiological and may well introduce changes in structure. NMR has the advantage of working with proteins in solution, but it is difficult to use on large molecules. Neither technique provides easily interpretable, intuitive results.

Cryoelectron microscopy

CryoEM uses a transmission electron microscope (TEM) to examine frozen hydrated samples. In a TEM, a coherent beam of high-energy electrons 'illuminates' the sample. Lenses beyond the sample focus the transmitted electrons into an image of the sample that is projected onto a viewing screen or detector array. The

process is similar to the projection of an image from film to screen using a slide projector. The contrast in the image involves both amplitude- and phase-contrast components. In unstained, frozen, hydrated samples, phase contrast prevails.

Although a TEM is capable of sub-Ångstrom resolution, sample preparation and possible beam damage, caused by the interaction of the electron beam with the frozen hydrated sample, limit resolution in most biological materials. Generally exposure must be restricted to less than 10-20 electrons per square Ångstrom to avoid beam damage. At these signal levels, the signal-to-noise ratio of the image is rather low and visual examination would reveal little, if any, information. However, the information is there and can be extracted by averaging a large number of images of identical particles. This may be considered analogous to the superposition of X-rays diffracted by the many perfectly aligned molecules of a crystallized protein in XRD. Given a sufficiently large number of images (~300-100,000), investigators have reported resolutions as good as 0.7 nm - good enough to discern α -helices within the protein structure at the secondary and higher levels.

Sample preparation

Obviously, a TEM brings with it its own set of sample preparation requirements. The sample must be thin enough to transmit



Keep discovering

World Drug Alerts Plus
Derwent Drug File

Get all the information you need to discover new drugs and bring them to market with *World Drug Alerts Plus* and *Derwent Drug File*. These two premier biomedical databases contain more than two million vital pharmaceutical records drawn from patents, journals, and conferences. You can access significant data about new and established drugs, helping you at every stage of the discovery and development process.

You can search *World Drug Alerts Plus* and *Derwent Drug File* via the user-friendly *Derwent Discovery™* interface. You can also cross search the databases with *Investigational Drugs database (IDdb)*, the leading competitive intelligence and R&D monitoring service.

www.thomsonscientific.com/discovery

electrons, the thinner the better, with typical thicknesses of less than 100 nm. Furthermore, the sample must tolerate the vacuum environment of the TEM, required to prevent the scattering of electrons by gas molecules. Traditionally, chemically fixing and drying the sample to remove all water and other volatile components have met this requirement. Such harsh treatment unavoidably introduces artifacts and, in the case of proteins, would destroy the structure that is the object of the observation. Freezing, however, offers a perfect solution to immobilizing volatile sample components. In CryoEM the sample is frozen extremely quickly (freezing velocity > 10,000 K/sec) in a thin layer of vitrified water. Because the vitrification process proceeds so rapidly, the water molecules have 'no time' to crystallize and form an amorphous 'liquid' that preserves sample structure in its pristine natural state. This is one of the important advantages of CryoEM over techniques that impose non-physiological conditions upon the sample. It also offers the opportunity to freeze and observe the sample at various stages in a dynamic process.

Single particle analysis

There are currently two approaches to deriving structural information from proteins: single particle analysis and cryo-electron tomography. SPA is a somewhat misleading name since it in fact refers to the summing, *in silico*, of images of a large number of identical particles to improve signal-to-noise ratio and, consequently, resolution. The name refers to the fact that each of the summed images is of a single, identical, randomly oriented particle. The

problem then becomes one of aligning the images such that the random noise events average out while the signal accumulates (Figure 1). Because each sample is exposed only once to the electron beam, the number of images is limited only by the number of available protein molecules. Resolutions in the 0.7-1.0 nm range can be derived from sets of 300-100,000 images, depending upon particle symmetry. Improving the resolution by a factor of two is theoretically possible, but may require an order of magnitude increase in the number of images. This approach is limited to molecules that are relatively rigid and can easily be statistically classified using multivariate statistical analysis. Averaging molecules that are slightly dissimilar, or not correctly classified, introduces a blurring effect.

Clearly, data acquisition and, in particular, processing time are significant factors controlling the practical limits of SPA. The prospect of acquiring and analyzing hundreds of thousands of images is daunting; however, recent developments in automation offer the potential for improvement.

Vitrified samples are usually prepared on a holey carbon film supported by a metal grid. An aqueous sample solution is deposited on the grid and then subjected to plunge freezing in a liquid coolant. After vitrification the sample must be isolated during transport to the TEM in order to prevent the condensation and freezing of water from the ambient environment. We recently introduced a system called Vitrobot that automates the entire vitrification process and provides complete isolation during sample transport.

The sample introduced into the TEM consists of a thin vitrified ice film covering the sample grid that will be exposed in hundreds of locations by the holes in the carbon film substrate. Researchers at Scripps Research Institute have developed a program (Leginon) that automates the entire analysis. By acquiring and manipulating images at progressively higher magnifications, it builds a library of images, each one a projection of an identical particle at some random orientation. The images are then sorted by similarity and summed to improve signal to noise ratio and resolution.

Analysis of the data set to determine protein structure is complex and the detailed procedures depend upon the type of protein or protein complex being analyzed. For symmetric assemblies such as the helical filaments of the tobacco mosaic virus, or the icosahedral structure of the cow pea mosaic virus, a single image of the assembly provides multiple copies of the component molecule, and knowledge of their symmetric relative orientations reduces the complexity of the analysis. Less symmetric particles such as hemocyanin from the keyhole limpet pose much a greater challenge for both data acquisition and analysis time (Figure 2).

CryoET

CryoET of proteins fixed in a thin vitrified ice layer also examines a single particle, but rather than looking at a large number of projections, each from a different (though identical), randomly oriented particle, it acquires a limited series of projections from one particle as it is incrementally tilted through a range of precisely controlled angles (Figure 3). The resolution of the technique is a function of the specimen thickness, the overall range of tilt, the number of images, and the increment of tilt between images. Since the same particle is exposed repeatedly to the electron beam, the ultimate limit to resolution is the number of exposures it can tolerate before its structure is significantly degraded by the beam. Currently, tomographic analysis of vitrified suspensions offers resolution in the range of 2-4 nm.

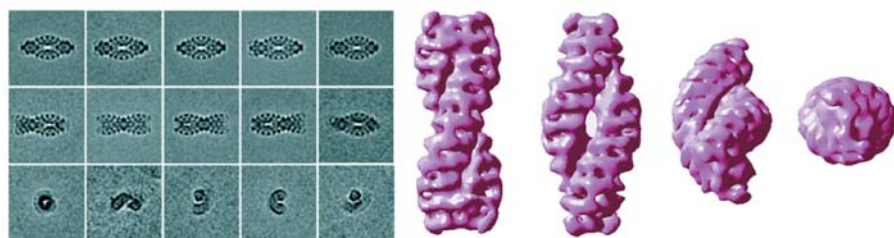


Figure 1. (a) High-resolution TEM (FEI Tecnai) images of *Drosophila melanogaster* tripeptidyl peptidase II acquired as part of a CryoEM single particle analysis. (b) The resulting 3D reconstruction with 3.3-nm resolution. Courtesy of Sali et al *Nature* 422: 216-225; ©2003 Nature Publishing Group.

The automation of CryoET faces a set of challenges quite different from those encountered in SPA. The number of images, limited by the total beam exposure the sample will tolerate, is much smaller for CryoET than for SPA. While this might seem to reduce the computational complexity of the analysis, it also places a premium on the information extracted from each image. CryoET does not enjoy the luxury of the nearly unlimited set of data that is available to the sorting and averaging techniques used in SPA. Fortunately, CryoET does have the advantage of knowing more about each image in its data set, specifically, the precise change in orientation from image to image. Sidec Technologies of Stockholm, Sweden has developed an approach to CryoET analysis (Sidec Electron Tomography or SET) that extracts structural information using a proprietary algorithm known as Constrained Maximum Entropy Tomography (Sidec COMET). Starting with an initial reconstruction derived from back-projections of the image set, the COMET algorithm refines the model by deriving the smoothest 3D conformation that could have produced the observed data (Figure 4).

Although CryoET, applied to vitrified suspensions of proteins and/or protein complexes, is currently limited to about 2-nm resolution it has the advantage of being truly single-particle: the structural information it delivers is derived from a single instance of the particle. This offers important benefits in dynamic applications such as the study of flexible proteins and conformational changes that occur during binding interactions between proteins. One recent study by Bongini and colleagues reported such observations of immunoglobulin (Figure 5). By acquiring a gallery of structures from a number of individual particles, the investigators were able to determine the distribution of various conformations and derive a potential energy model to describe the angles among the two fragment-antigen binding (Fab) arms and the effector domain (Fc) stem.

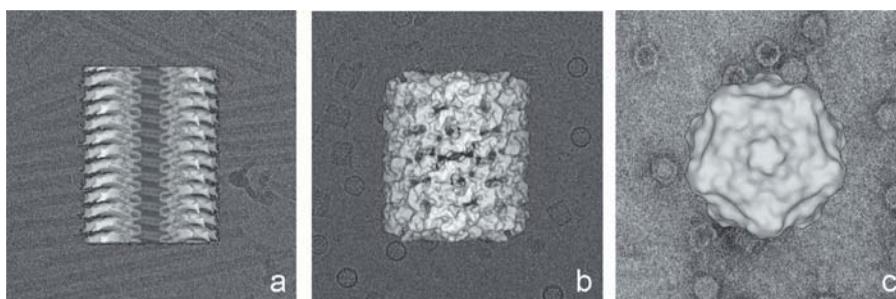


Figure 2. SPA reconstructions of (left to right) tobacco mosaic virus, keyhole limpet hemocyanin, and cow pea mosaic virus superimposed on representative TEM (FEI Tecnai) images of the particles. Courtesy of Carragher et al *Journal of Synchrotron Radiation* 11: 83-85; ©2004 IUCr.

Analysis time

It is not unusual for the complete determination of a protein structure, including expression, purification, crystallization and analysis, to take several months using XRD. A number of automated systems have been developed for expression and purification and substantial progress has been made in automating the search for crystallization conditions. In most cases, it is this search that consumes most of the time. NMR, though it eliminates the requirement for crystallization, is computationally more complex. NMR structural analysis times are also typically measured in months. In contrast, CryoEM can provide structural information within a single day. CryoET using Sidec's proprietary COMET algorithm promises structural analysis with 2-4 nm resolution in a day or less. SPA operates on the same time scale. As proof of concept, the Scripps

group recently reported complete analysis of TMV coat protein with 10-Å resolution within 20 hours of first inserting the grid into the TEM. This included the acquisition and analysis of images of some 50,000 individual particles.

CryoEM versus NMR/XRD

Protein crystallization remains the greatest barrier to structural analysis with XRD. Much of the low hanging fruit has already been picked and it is not unreasonable to expect that the remaining proteins will prove harder to crystallize. Membrane proteins, designed by nature to exist in a non-aqueous environment, may never be crystallized. Globular proteins and flexible proteins also pose significant crystallization challenges. Although NMR avoids the crystallization requirement it is limited in its ability to deal with large molecules and molecular complexes. CryoEM is a

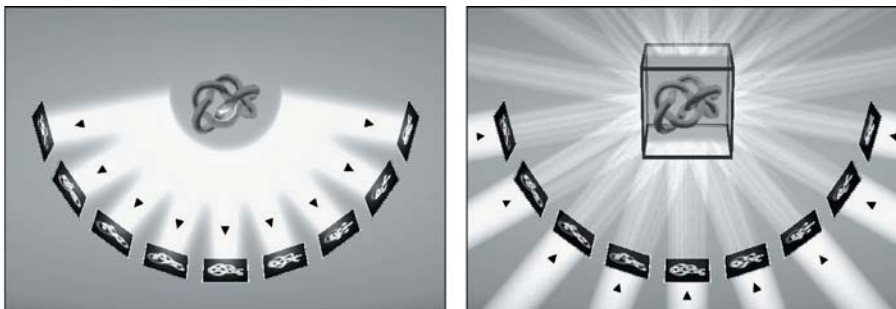


Figure 3. Electron tomography acquires multiple 2D projections of the molecule as it is tilted through a sequences of precise angular increments. The projections are then 'back-projected' to calculate the 3D model. Sidec COMET analysis acts as a mathematical filter to improve the signal to noise ratio and resolution of the model. Courtesy of Sali et al *Nature* 422:216-225; ©2003 Nature Publishing Group.



Figure 4. Membrane receptor expressed in the cell membrane neurotrophin receptor p75 (light red) visualized in situ using Sided Electron Tomography. The transmembrane receptor domains are marked in dark red. The membrane itself is not visible at this magnification. To locate the receptor among other cell proteins (grey), gold-labeled antibodies (yellow and blue) were used. Courtesy of Sided Technologies, Stockholm, Sweden.

complementary tool, making up for its lack of atomic-level resolution with its ability to analyze large, uncrystallized structures. One of the most promising approaches to structural analysis is hybrid analysis that fits atomic resolution structures determined by XRD and NMR into the EM density maps (Figure 6).

Though CryoEM currently lacks the resolution to determine atomic structure, it is certainly capable of elucidating the higher-level structure that plays an important role in determining the structure-activity-relationships between multiple proteins and between drugs and proteins that are considered drug targets. Nature itself greatly values higher-level structure, often conserving shape and function over molecular sequence: mutations that change the sequence without impacting shape and function cannot be selected for or against. Moreover, many, if not most, proteins act in complexes with other proteins, another example of the importance of higher-level structural information.

While the determination of the 3D structure down to the atomic level will remain the province of XRD and NMR for the foreseeable future, pharmaceutical researchers are placing growing emphasis on the rapid determination of higher-level structure. An informal survey conducted while researching this article uncovered numerous examples:

- One group was able to determine that the target protein expressed by a cell line they used for screening was very different from the one that was expressed by the native cell line. They stopped and redeveloped the screening line.
- Another group was able to narrow a selection of ligands when they compared the conformational changes induced in the target by the binding of the candidates with the changes induced by the binding of the natural analog.
- Yet another group is studying the complexes of growth hormone and its receptors in situ - still embedded in the cell membrane.
- A group studying a specific membrane protein was able to determine that it appears predominately as a tetramer.

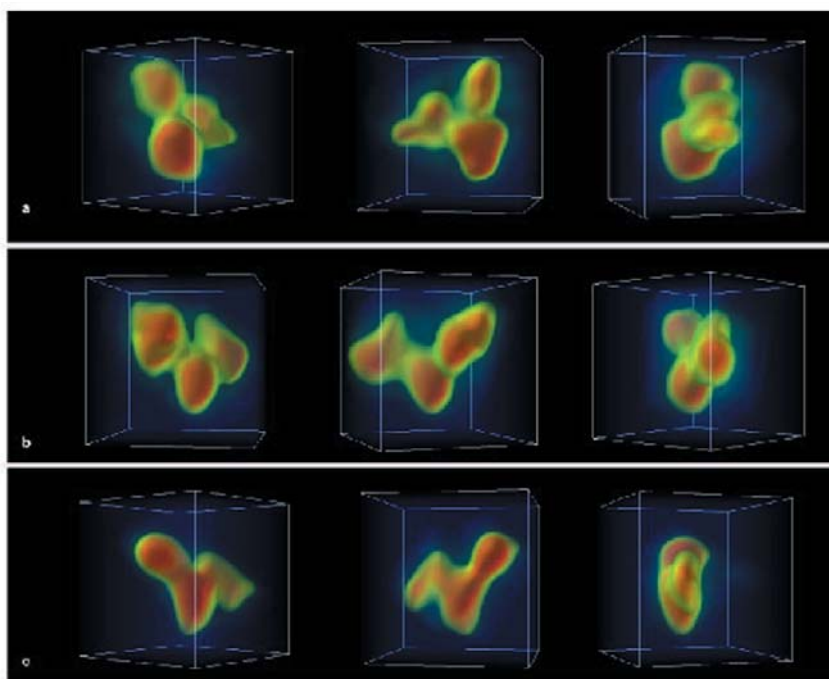


Figure 5. Three different IgG molecules (A, B, C) modeled with CryoET. Each model is shown in three different orientations. From a collection of such models it was possible to derive a potential energy model to describe the distribution of angles observed among the Fab arms and Fc stem. Courtesy of Bongini et al *Proceedings of the National Academy of Sciences* 101:6466-6471; ©2004 PNAS.

CryoEM offers an unprecedented opportunity to examine proteins in their natural state. There has long been concern that some of the more extreme conditions applied to promote crystallization may introduce structural artifacts. The plunge freezing used with CryoEM avoids this concern: if the process occurs rapidly enough to prevent the rearrangement of water molecules into ice crystals, it is reasonable to assume that it is also fast enough to prevent rearrangements within the target protein. Its rapidity also offers the prospect capturing 'snapshots' of dynamic processes such as the previously mentioned work on immunoglobulin.

Structure-based drug discovery

Structure-based drug discovery has great promise in the coming years as the genome is converted into a structural proteome and structure-

activity-relationships are determined. CryoEM can play an important role in that process. Although the human genome encodes for tens of thousands of proteins, many of these can be grouped into a much smaller number of categories whose members are related genetically, structurally, and functionally. CryoEM may allow researchers to classify proteins 'at a glance', streamlining both target prioritization and lead discovery. This could be particularly powerful when applied to candidates initially identified through partially shared genetic sequences. The ability to rapidly determine structure and terminate unlikely projects early in the discovery process holds tremendous value for pharmaceutical manufacturers.

CryoEM opportunity

CryoEM is a powerful new tool for protein structural analysis. It is complementary to existing techniques, offering less resolution, but avoiding the requirement for crystallization and working well with larger molecules and molecular complexes. Most importantly, it can determine higher-level structure in a fraction of the time required by conventional techniques and promises

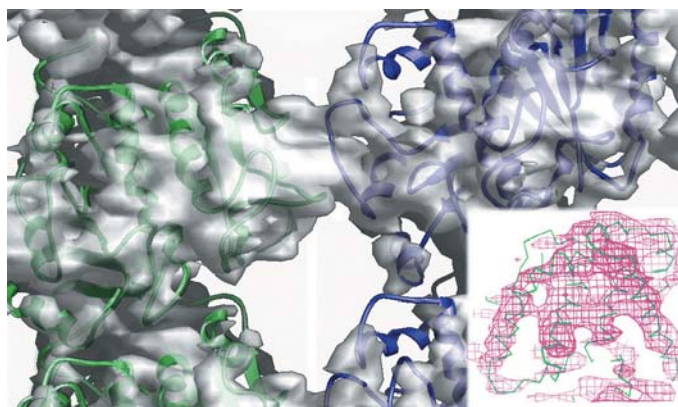


Figure 6. Hybrid analysis showing the atomic model (ribbon) of the protein tubulin 'docked' into the CryoEM density map of the assembled microtubule. Courtesy of Sali et al *Nature* 422:216-225; © 2003 Nature Publishing Group.

to play a critical role in structure-based drug discovery as researchers sift through the huge volumes of data created by current attempts to characterize the entire human proteome.

Jack Elands

Director of Business Development, Life Sciences

Werner Hax

Strategic Marketing Manager, Life Sciences

FEI Company
PO Box 80066
5600 KA Eindhoven
Netherlands

Email:
jelands@feico.com
whax@feico.com

www.feicompany.com

FURTHER READING

Bongini L, Fanelli D, Piazza F *et al* (2004) **Freezing immunoglobulins to see them move.** *Proceedings of the National Academy of Sciences* **101**:6466-6471.

Debe DA, Hambly K (2004) **Supporting your pipeline with structural knowledge.** *Current Drug Discovery* March:15-18.

Carragher B, Fellmann D, Guerra F *et al* (2004) **Rapid, routine structure determination of macromolecular assemblies using electron microscopy: Current progress and further challenges.** *Journal of Synchrotron Radiation* **11**:83-85.

Potter CS, Fellmann D, Milligan RA *et al* (2002) **Automated cryo-electron microscopy.** *IEEE International Symposium for Biomedical Imaging* 2002 IEEE 7-10 July, Washington, DC.

Sali A, Glaeser R, Earnest T *et al* (2003) **From words to literature in structural proteomics.** *Nature* **422**:216-225.

Medalia O *et al* (2002) **Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography.** *Science* **298**:1209-1213.

Burke J (2004) **Protein crystals, the next hurdle.** *Current Drug Discovery* March:19-22.

Abbott A (2002) **The society of proteins.** *Nature* **417**:894-896.