WORKSHOP ON THE VALIDATION AND MODELING OF ELECTRON CRYO-MICROSCOPY STRUCTURES OF BIOLOGICAL NANOMACHINES

STEVEN J. LUDTKE

Verna & Marrs McLean Dept. of Biochem. & Mol. Biology, Baylor College of Medicine,
1 Baylor Plaza , Houston, TX 77030, USA
sludtke@bcm.edu

CATHERINE L. LAWSON

Rutgers, The State University of New Jersey, Department of Chemistry & Chemical Biology and Research
Collaboratory for Structural Bioinformatics, 610 Taylor Road Piscataway, NJ 08854, USA
cathy.lawson@rutgers.edu

GERARD J. KLEYWEGT

Protein Data Bank in Europe, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK
gerard@ebi.ac.uk

HELEN M. BERMAN

Rutgers, The State University of New Jersey, Department of Chemistry & Chemical Biology and Research
Collaboratory for Structural Bioinformatics, 610 Taylor Road Piscataway, NJ 08854, USA
berman@rcsb.rutgers.edu

WAH CHIU

Verna & Marrs McLean Dept. of Biochem. & Mol. Biology, Baylor College of Medicine
1 Baylor Plaza, Houston, TX 77030, USA
wah@bcm.edu

Electron cryo-microscopy (cryoEM) is a rapidly maturing methodology in structural biology, which now enables the determination of 3D structures of molecules, macromolecular complexes and cellular components at resolutions as high as 3.5Å, bridging the gap between light microscopy and X-ray crystallography/NMR. In recent years structures of many complex molecular machines have been visualized using this method. Single particle reconstruction, the most widely used technique in cryoEM, has recently demonstrated the capability of producing structures at resolutions approaching those of X-ray crystallography, with over a dozen structures at better than 5 Å resolution published to date . This method represents a significant new source of experimental data for molecular modeling and simulation studies. CryoEM derived maps and models are archived through EMDataBank.org joint deposition services to the EM Data Bank (EMDB) and Protein Data Bank (PDB), respectively. CryoEM maps are now being routinely produced over the 3 - 30 Å resolution range, and a number of computational groups are developing software for building coordinate models based on this data and developing validation techniques to better assess map and model accuracy. In this workshop we will present the results of the first cryoEM modeling challenge, in which computational groups were asked to apply their tools to a selected set of published cryoEM structures. We will also compare the results of the various applied methods, and discuss the current state of the art and how we can most productively move forward.

1. Electron Cryo-microscopy

Electron Cryo-microscopy is a versatile experimental technique with several sub-specialties, each of which has its own unique strengths and weaknesses, which must be taken into account when modeling molecular structures or validating results. We briefly introduce each technique, highlighting the most important aspects of each from a modeling perspective.

1.1. *Single Particle Reconstruction*

Single particle reconstruction is the most widely used of the cryoEM methodologies for macromolecular structure determination, responsible for over 80% of the entries in the EMDB (http://EMDatabank.org). In this technique, purified macromolecules in aqueous buffer are vitrified and imaged, yielding images of individual particles in largely random orientations in a layer of vitreous ice. These images are extremely noisy due to the need to avoid radiation damage. They represent a snapshot of the solution conformation at the time of vitrification, thus the particle population includes any structural variability present in solution. Images of tens of thousands to millions of particles are selected and processed using a complex series of algorithms which determines the 3D orientation of each particle, corrects for microscope artifacts, and in certain cases separates the particles into multiple classes based on conformation, ligand binding or other attributes. These particles are then used to produce one or more 3D reconstructions at resolutions as high as 3.5 - 4.5 Å, for example[1-3].

1.2. *Electron Cryotomography*

Historically this technique has been used to for lower resolution studies of cellular architecture and subcellular structures, which, while highly interesting, is not particularly relevant to molecular modeling. Recently, however, a hybrid approach between tomography and single particle reconstruction has gained popularity. In this technique, a tomographic reconstruction is performed on a specimen similar to that used in single particle reconstruction. This provides extremely noisy 3D reconstructions of individual macromolecules, which can then be aligned and averaged in 3D[4]. The advantage over traditional single particle reconstruction is the presence of 3D information for each particle, rather than having only a single 2D projection of each. This 3D information can be used to resolve ambiguities between particle orientation and changes in particle conformation. While this technique is very powerful for studying difficult specimens or specimens displaying structural variability, at present its resolution is limited to ~20-30 Å even in the best cases. Thus, from a modeling perspective it is suitable only for docking large X-ray structure fragments.

1.3. *2D Crystallography*

Electron crystallography was used to elucidate some of the earliest membrane protein structures [5], and still remains a powerful technique for systems that are resistant to 3-D crystallization, but may naturally form 2D arrays. While this technique can produce exceptional resolution in the plane of the crystal, exceeding 2 Å in one case[6], the resolution in the orthogonal direction is necessarily much worse due to the experimental geometry. Nonetheless, this remains a powerful technique that can produce maps amenable to standard X-ray structure model building methods.

1.4. *Helical Reconstruction*

This technique determines the structure of macromolecules arranged in helical arrays. These arrays may be either naturally occurring or 2D crystals that have been formed on the surface of lipid tubes. The advantage of the helical experimental geometry is that a single filament provides images of the target protein and affords full 360-degree tomographic coverage. Until recently, this technique was capable of resolutions well beyond those achieved with single particle

reconstruction[7]. However, thanks to recent strides in single particle reconstruction, the gap has narrowed considerably. The best structures using this method have sufficient resolution for traditional X-ray model building methods.

2. Challenges of CryoEM Map Interpretation

Each of the above techniques are powerful, but each also has limitations. Here we will focus on interpretation of maps determined using single particle analysis, since more than 80% of deposited structures have been determined using this method. The fundamental challenge in any single particle analysis project is the high noise level present in the data owing to the need to avoid radiation damage. As the resolution improves, this problem becomes worse, as radiation damage tends to destroy high resolution features first. Single particle reconstruction intrinsically relies on averaging together large numbers of particles. This raises the question of how to assess the interpretability of reconstructed maps. Resolution in this field is a measure of the noise-levels present in the final reconstruction, and is quite distinct from resolvability, which can be adjusted without impacting measured resolution.

The standard resolution metric requires one to split the data into even and odd halves, generate two 'independent' reconstructions, then compare them by Fourier shell correlation (FSC). The resolution is then the point at which the FSC value falls below a threshold value. Unfortunately the FSC is susceptible to overestimation due to noise/model bias[8] and a number of other possible artifacts. While there are rarely any issues with the overall accuracy of a single particle reconstruction, there is some uncertainty over what level of detail in any given structure can be safely interpreted. For example, it is possible to filter a 5 Å resolution map so apparent sidechain densities are visible, but it is almost certain that such densities are simply noise.

Because cryoEM is now able to achieve resolutions that enable molecular interpretations at the near-atomic level, there is a critical need for model data validation tools as well as improved methods for map interpretation.

3. The CryoEM Modeling Challenge 2010

The idea to host a cryoEM modeling challenge (ncmi.bcm.edu/challenge) was developed in order to provide the modeling community with a standard set of maps to test their methods against, enabling comparison of results, and to improve awareness within the cryoEM community of the range of available tools. Unlike a true blind test of the various computational methods as provided by CASP (www.predictioncenter.org/casp9), the modeling challenge utilized known structures and challenged any interested groups to apply their methods to one or more of the structures, with the goal of improving existing map interpretations or developing new tools for map/model validation. The provided maps cover a range of different symmetries, particle sizes, resolutions and experimental methods. The challenge will conclude at the beginning of December, 2010, and results will be presented and discussed in this PSB workshop. After the conclusion of the workshop, all submitted results will be made permanently accessible to the public .

4. Available Modeling Techniques

There are many different computational techniques that can be used to interpret cryoEM maps, depending on the resolution range the map falls in. Each of the following sections describes a category of possible submissions in the cryoEM Modeling Challenge.

4.1. *Volume Interpretation*

Volume interpretation represents a class of techniques that can be applied to structures where the resolution is insufficient for molecular modeling approaches. We have divided these techniques into three broad categories. The first category is map segmentation; separating a map into meaningful sub-regions. Segmentation may be accomplished in a variety of ways, depending on the available information. For example, if crystal structures of domains or components of the map are available, they can be docked into the cryoEM reconstruction. De novo segmentation methods may attempt to perform automated segmentation based on, for example, the location of low-density regions combined with the symmetry of the structure. Validation of results remains a major issue for this technique.

The second technique is secondary structure element annotation. At sub-nanometer resolution, α-helices become resolvable, and as the resolution improves further, β-sheets become discernible, eventually showing strand separation. In this intermediate (~5-10 Å) resolution range, tools for automatic identification and localization of secondary structure elements become quite valuable, but again, in marginal cases there are validation issues. In addition, in this resolution range, it becomes possible to dock crystal structures with much higher levels of confidence.

The final technique in this class is Cα protein backbone tracing. In the 3.5-5 Å resolution range, it is often possible to perform unambiguous tracing of the protein backbone directly from the density map. Some methods for achieving this rely on additional information, such as sequence-based secondary structure prediction or the existence of a crystal structure of a homologue, to help resolve ambiguities.

4.2. *Modeling*

These methods yield true atomistic models derived from cryoEM density maps. The first of the three methods in this class is related to rigid-body docking described above. The implementation of this method may take many forms, and some methods are resolution-dependent. In many cases where flexible modeling is considered impractical, larger models will be broken into domains, for example at hinge points, to attempt to elucidate more information about differences between the cryoEM structure and the model. Variations of this method have been used in cryoEM for decades, even on structures at very low resolutions. Once again, the major difficulty lies in establishing the reliability of the final results.

The second class of modeling techniques comprises flexible docking methods. Rather than simply finding the best 3D position and orientation for an atomistic model within a cryoEM map, in this method the atomic positions are locally adjusted to better match the experimental data. This can be used to model structures in various conformational states, or can make corrections to

homology models. However, again there are serious questions related to the level of detail at which such flexible docking can be trusted. For example, at ~8 Å resolution, α-helices are clearly resolved, but if the flexible fitting were to try to use the density map to modify sidechain orientations, the results would obviously be invalid. Groups developing these techniques are working to establish how to balance molecular modeling energy functions against the need to match the information content of the experimental data.

The final technique is true ab initio modeling based on cryoEM maps. This includes established methods for model building in X-ray crystallography. Since the typical resolution of cryoEM experiments is still below the levels typical for crystallographic studies, new techniques are being developed that hopefully allow for accurate model building at lower resolution. An important point, however, is that the two techniques are not entirely the same. While cryoEM and X-ray crystallography both produce density maps, the specific artifacts (e.g. image distortion and image alignment errors in the case of cryoEM and model bias in crystallography) present in each are not necessarily the same, and the definitions of resolution used in the two communities are not entirely compatible. Accurate atomistic modeling has been performed on a number of cryoEM maps at ~4 Å resolution, a resolution that is generally regarded as marginal in X-ray work.

5. Conclusions

As of September 2010, there were over 50 registered participants in the modeling challenge. Many of the major modeling groups using physics and statistical based simulation and cryoEM density restraints are actively participating and applying their methods to the six cryoEM targets selected for the challenge. Many other groups are applying their tools for specific aspects of cryoEM map analysis for segmentation, secondary structure element identification and de novo modeling. Representatives from several groups have been invited to present their work at the workshop, and there will also be a panel discussion of the results.

6. Acknowledgements

References

1. X. Zhang, L. Jin, *et al.*, *Cell* **141**, 472-82 (2010).
2. Y. Cong, M. L. Baker, *et al.*, *Proc Natl Acad Sci U S A* **107**, 4967-72 (2010).
3. J. Z. Chen, E. C. Settembre, *et al.*, *Proc Natl Acad Sci U S A* **106**, 10644-8 (2009).
4. J. Walz, D. Typke, *et al.*, *J Struct Biol* **120**, 387-95 (1997).
5. R. Henderson, P. N. Unwin, *Nature* **257**, 28-32 (1975).
6. T. Gonen, Y. Cheng, *et al.*, *Nature* **438**, 633-8 (2005).
7. C. Sachse, J. Z. Chen, *et al.*, *Journal of molecular biology* **371**, 812-35 (2007).
8. A. Stewart, N. Grigorieff, *Ultramicroscopy* **102**, 67-84 (2004).
9. C. Lawson, M. Baker, et al., *Nucleic Acids Research,* in press (Jan. 2011)