EMDB Advisory Panel Report 2013

Meeting details: Held on March 8th at Baylor College of Medicine, Houston.

Panel members: Paul Adams (Lawrence Berkeley Lab, Chair), Richard Henderson (MRC-LMB, Cambridge), Bram Koster (Leiden University), Maryann Martone (UCSD), and Andrej Sali (UCSF).

Summary

It was the unanimous opinion of the Advisory Panel that the EMDB team has made significant advances in infrastructure and methods over the last five years, and that these developments have accelerated over the last 12 months. This achievement has been demonstrated in no small part by several important publications arising from community meetings led by EMDB. The team shows excellent internal integration, and has also become a focal point for the growing EM community. In particular the efforts of Catherine Lawson and Ardan Patwardhan are of note; their dedication to the EMDB mission is critical to current and future success. The team has gathered significant momentum, making continuity of funding essential to maximize the return on NIH investment. The panel also noted that the NIH funding for the EMDB benefits greatly from the leveraging of complementary programs at the participating institutions. This synergy has enabled rapid progress, and places the team in a unique position to develop the standards, deposition tools, and validation algorithms for the future of EM-based structural methods. The future plans in these areas were considered very strong, well thought out and likely to have high impact on the community. Of particular note are the efforts to rationalize EM model and data deposition across the wwPDB, development of analysis and reporting services for deposited structures, and the multiple approaches to map and model validation.

The EMDB future plans are ambitious, but commensurate with the needs of the community. The panel had several recommendations that may help the team focus their efforts in the coming year. There is a need for a balance between the development of new tools and methods, and playing a leadership role in the EM community. Therefore, it will be important to leverage the activities of others to promote community discussion and interactions. A specific suggestion of the panel is to provide a link and means of signing up to the 3dem mailing list from the EMDataBank web site. This could be combined with a modernization of the 3dem list to enable web access. In addition, as the techniques and tools evolve in the cryoEM community, there will be more overlap and opportunities for cross-fertilization with other communities and community resources, e.g. the American Society for Cell Biology's Cell Image Library. It is therefore important to keep an eye towards interoperability rather than duplication of resources.

With regard to validation methods, there are many exciting developments in progress. The team is encouraged to adopt, and promote, the use of a map versus model Fourier shell correlation (FSC) calculation to assess model quality. Much of the initial validation developments are focused on data internal to the team, which is a realistic path during these early stages. However, over time, the team is encouraged to collaborate more broadly in testing validation approaches as this will help disseminate their new methods and educate the community. The team is also commended on their efforts to engage

researchers from other fields where model validation is quite well developed, e.g. X-ray crystallography, and these efforts should continue. The panel recognized that many of the analyses and deposition tools would be of great use outside of the web-based deposition systems. Although it is currently impractical for the team to distribute these tools, it could be considered as part of a long-term strategy for engaging the EM community. Examples of successful community tool repositories include the Neuroimaging Tools and Resource Clearinghouse (NITRC; http://nitrc.org).

The number of EM depositions to the databank is rapidly increasing, helping establish a unique position for the EMDB in shaping the growth of the EM community. To best target their efforts, they are encouraged to research how EM models and data downloads are being used by the broad scientific community. Finally, to ensure that EM researchers are aware of planned EMDB developments, and best able to provide timely input, the team should provide a concise, high-level timeline of plans via the EMdatabank web site.

Detailed Feedback

EMDataBank Services

Catherine Lawson and Ardan Patwardardhan presented an overview and plans related to the EMdatabank Services. The current 3DEM pipeline was discussed from deposition, annotation, integration, through to dissemination. The yearly data deposition shows an exponential growth with a three-fold increase between 2009 and 2011. The upload data model system description is greatly improved by separating the required information on methods from structure information. The EMDataBank defined an XML file to combine method metadata with the data sets. A procedure to automate the harvesting from the electron microscopes used is in a test phase with Baylor and FEI. The status and future plans to integrate EM data into the wwPDB deposit and annotation system was outlined. The EMDB is now consolidated with the PDB in a single ftp archive. EMDataBank's recognized expertise will be used to disseminate 3DEM validation standards, which were also published by the EMDataBank in the past period. Powerful EM search tools, statistics plots, map visualization pages and visualization tools have been developed. The committee regarded the graphs and figures produced as part of validation reports as highly authoritative tools that deserve publication by themselves. The EMDataBank team is highly active within the 3DEM field at meetings and workshops gathering feedback from the community. From the community, procedures to upload electron tomography and sub-tomogram averaging data depositions are required. Close collaborations are in place with industrial partners and scientific groups to carry out pilot electron tomography data repositories, among others making use of the Aspera system for data transfer and Omero for data presentation.

3DEM Tools & Validation

The panel were impressed by a new tool developed by the EMDB during the last few months consisting of a "simple" EM validation report together with a map/model analysis that involved production of displays that overlay the map and model and plot the fraction of main chain and side chain atoms that are included in the map at different contour levels. The committee felt that, once this useful program had been developed further in-

house and possibly included as part of an EMDB publication, it would be welcomed by the cryoEM community as a tool for viewing map/model combinations rapidly and conveniently.

Map validation: following recommendations from the EM validation task force (EM VTF) that met in 2010, Steve Ludtke described progress towards the introduction of a number of procedures for map validation. It was planned that these validation tools might be requested or suggested as part of the map deposition procedure. Extensive progress towards acceptance of tilt pair validation had been made. A number of software packages, including EMAN2, had now implemented it in similar ways and it had proved to be helpful. Another idea was to use different software packages to see whether the same or a similar structure emerged by processing the same data using the different protocols used in the different packages. This also provided some reassurance and allowed a semi-quantitative consistency check to be made by calculating the FSC between the 3D maps determined using different packages. Finally, a 3D variance map could provide guidance concerning which parts of a structure were more and which were less variable.

Model validation: an informative procedure for measuring how well a model fits into a 3D density map, particularly after procedures (of several different kinds) that involve flexible fitting of the model, is to calculate the FSC between the experimental map and a 3D map calculated directly from the model. The advisory committee encouraged this calculation to become one of the options proposed by EMDB. Great progress was presented in fitting models into medium resolution maps, and building models into higher resolution maps. It would be useful to present ensembles of fits where possible, e.g. if there is a stochastic component to the model building or multiple different algorithms that can be applied. This will make it easier to understand where there is uncertainty in the fit model. It would be interesting to see how these model uncertainties correlate to the 3D variance maps. Finally, the team has reached out to the crystallographic and macromolecular validation communities to allow them to make use of existing tools for model optimization and validation. Progress in adopting these methods has been rapid and bodes well for future developments. The team is encouraged to strive for EM atomic model quality that is the equal of the current gold standards in crystallography, and better than older models created using outdated computational methods. The impressive 3D reconstructions presented at the meeting provide high hope that this will be achieved.

EM Challenge: the EM Challenge web site and workshop were seen by the SAB as a most positive contribution to the EM community. Thus, the Challenge should continue to evolve, taking into account existing comments noted in the slide deck, including (i) focus on well-defined map-based modeling challenges to allow for concrete validation and outcomes; and (ii) select a meeting location that is more accessible than Hawaii. In addition, there are four more suggestions. First, the assessors should consider distinguishing between automated and non-automated methods. Second, quality criteria for atomic EM-based models should probably include all those already used by PDB for structures determined by X-ray crystallography and NMR spectroscopy as well as some that are EM-specific. Third, a possible venue for the meeting, maximizing attendance, could be just before the next Keystone Hybrid Methods meeting, before the ISMB

conference (similarly to the satellite 3DSig conference), or before some other relevant meeting; the Challenge meeting could then include more dialogue, demonstrations of software, etc. Fourth, while the broad categories assessed by the Challenge seem appropriate, perhaps it is worth assessing also primarily EM-based models that are supported by additional data (such as chemical cross-links), although such a variation may be premature at this stage and might be better explored independently in a possible Hybrid Methods Challenge.

Looking to the future of model deposition, perhaps the EMDB should perform an experiment and allow depositors of EM-based models to list in free format what data and considerations other than fitting into an EM map they used to compute their models. If a sufficient number of depositors take this opportunity, EMDB may be in a better position to decide how to document the data and protocols used to compute EM-based models.