EMDB Advisory Panel Report 2014

Meeting details: Held on May 10th at Rutgers University, New Jersey.

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

Summary

The Advisory Panel strongly felt that the EMDB team has made significant advances in the last year, and has chosen to focus efforts in areas of high impact. Of particular note are the developments in map validation and structure deposition. It was also felt that the progress in model validation was excellent and timely, as the cryo-EM field moves towards atomic resolution models. The availability of integrated deposition of models and maps is key as it provides users with feedback about the atomic models. The analytical tools for model and map presentation as well as their analysis are considered very strong and are clearly already having an impact. Extending the validation approaches to include RNA molecules is a welcome development, with important consequences for the community. One question that EMDB will have to consider is when to approach journals about validation requirements for publication of cryo-EM based structures.

The EMDB future plans are ambitious, but in line with the needs of the community. The continued development of model-based validation methods in the next year is excellent. It is important that these efforts be coordinated with the proposed modelling challenge and the EM-VTF activities. The panel had several additional recommendations that may help the team focus their efforts in the coming year. Firstly, the EMDB should not shoulder the full burden of developing new methods, standards or tools if others are able to contribute. In particular, there are opportunities to engage the small angle X-ray/neutron scattering communities and the RNA structure community to help with some aspects of hybrid method definitions and validation metrics. One of the biggest potential challenges the EMDB will face is a rapid explosion in the use of cryo-EM as a result of new detector hardware. In particular, this may result in a significant increase in the number of atomic-resolution cryo-EM structures. The EMDB will need to be prepared for this growth by providing map/model validation tools to help depositors verify the correctness of structures and minimize EMDB manual effort at deposition. The panel felt that it would be important to allow for deposition of map variance and local map resolution, for example, to assess model uncertainty, and to inform subsequent molecular modeling. Finally, to ensure that validation metrics occupy the same importance in cryo-EM that they now do in macromolecular crystallography, the EMDB should strive to play a central role in developing community agreement about their use and meaning. In this regard, interactions with scientific journals may be important in establishing community standards.

The panel also encourages the EMDB to consider their long-term impact and role in the community. As computing resources become more and more available, it will eventually be practical to store the experimental data for cryo-EM maps and models. This is a challenging topic and one fraught with potential problems if not executed well. The EMDB should consider how to enable such depositions in the future and under what circumstances. The panel was delighted to see that a number of data sets are already

being held by the EMDB in a pilot scheme, which should provide some useful metrics for a longer-term plan. The EMDB is also very well placed to play a central role in defining the representation of complex biological systems – which will remain one of the key strengths of the cryo-EM method. They are urged to continue their efforts in defining the dictionaries to define such systems, with an eye to supporting integrative modeling in the future. Finally, the EMDB should think about what their focus should be three to five years from now. One area in which the panel felt the EMDB would be crucial and key to the success of the cryo-EM field is providing comprehensive machinery for identifying potential pitfalls in the analysis of experimental data and developing the appropriate validation metrics to avoid them. Without these, there is real danger that the field is undermined by a few high-profile mistakes. Below, the panel has detailed feedback in two specific topic areas.

Data sets for validation and development

The panel recommends that the EMDB focus on obtaining and curating 10 to 12 full datasets, in the form of micrographs or aligned movie frames. This was felt to be sufficient to provide enough diversity for testing purposes but not too onerous in terms of data management. Along with the raw information, it will be important to keep a reasonable set of intermediate results, for example class averages. This information will be important in reproducing results and also assessing where methods have failed, or improved. Clearly, the final map from a benchmark reconstruction also is required. One challenge in supporting development and new validation metrics is capturing information about the computational protocols used. Currently there is no accepted standard for recording the computational steps and parameter values. The EMDB is encouraged to keep in contact with developers who are actively defining infrastructures for protocol capture, such as the EMX group.

The modeling challenge

The panel was very supportive of the EMDB hosting a “challenge” to help assess the state-of-the-art in interpretation of cryo-EM maps, as this is likely to spur even more developments in this area. It was felt that this would work best if the modeling had targets across the 5 to 3 Å resolution range, with a total of approximately 10 structures. It is important that there be enough targets for the results to have statistical meaning. One of the important outcomes of the challenge could be an assessment of uncertainty for the different parts of the model and the maps, and the correspondence between the two. The participants should aim for proposing models whose precision reflects the precision of the map.

One powerful approach in establishing the challenge would be the creation of tests with a “known” answer. This could be achieved by using the same sample with a better detector to obtain a benchmark structure, and then use a worse detector to generate the test data for the challenge. Practically, it will be necessary for the EMDB to be involved in the organization and judging of the challenge. They are encouraged to look to the CASP competitions for potential models for operating the challenge and judging the results. One approach to help minimize manual effort and provide objectivity is to use automated model analysis tools where possible. It would also be beneficial for the cryo-EM field if judging involved others outside the field. Finally, it would be best to open up

the modeling challenge to researchers and developers outside the cryo-EM community, with an eye to bringing in new approaches.

The panel felt that the team's recent developments in the area of RNA modelling were particularly timely, and necessary, given the continued growth of cryo-EM for studying complex systems such as the ribosome. They are therefore encouraged to include RNA maps at high, medium, and low resolution if possible, in the modeling challenge.