2011 EM Databank (EMDB) Advisory Committee Meeting
EBI, Hinxton, UK, October 3

The EM Databank (EMDB) Advisory Committee met at the European Bioinformatics Institute (EBI) in Hinxton, Cambridge, UK from 9:30 am – 3:30 pm, on Monday, October 3, 2011. Members in attendance were Joachim Frank, Chair (Columbia University), Achilleas Frangakis (Goethe University, Frankfurt), Richard Henderson (MRC, Cambridge), Maryanne Martone (UC San Diego), and Michael Rossmann (Purdue University). Attendant observer was Paula Flicker (NIGMS).

After a welcome address by **Dr. Gerard Kleywegt** (EBI), **Dr. Wah Chiu** (PI, Baylor) gave an overview over the aims and organizational framework of the EMDB, which is a collaborative project between RCSB PDB, PDBe, and Baylor-NCMI funded by NIGMS, BBSRC, and EMBL. Dr. Chiu started by describing how EMDB has addressed 2010 AC recommendations. Among these were (1) arrangements to obtain benchmark data sets from the community, for dissemination to allow testing of new tools, (2) remediation of missing-species and strain information, and (3) elimination of the 2-year hold. Regarding tools for segmentation, labelling, and visualization, EMDB has agreed to confine its role to facilitation of collaborations within the community, rather than developing separate sets of tools. As the current funding period will end in May 2012, a renewal application has been submitted in July 2011, with a focus on the development of validation methods and standards.

Dr. Helen Berman (Rutgers) gave an overview over the current status of the EMDB. Since the establishment of EMDB, in 2002, there has been an exponential growth of EM map depositions, with the latest (Sep 28, 2011) status being 1150 map entries and 407 model entries. Six out of 12 major journals now have a "strong" policy on EM data deposition, meaning depositions are mandated as part of the publication requirements. A world map showed distribution of >9000 visits to the website by 90 countries, with 1/3 originating in the USA. Within the last year, three outreach activities were organized: (1) the EM Validation Task Force, coinciding with the last AC meeting in September 2010, (2) the Modelling Challenge Workshop in Hawaii, in January 2011, and (3) the upcoming Workshop on Data Management challenges, in Dec 2011. Dr. Berman listed seven publications in 2011 by PIs and staff associated with the EMDB.

Drs. Cathy Lawson and **Ardan Patwardhan** (Rutgers) gave an update on the new features of the deposition, archiving, and retrieval of maps from the website. The important news is that the EM data will be part of the redesigned PDB Deposition and Annotation System (2012). On-line visualization of EM maps has been improved, and data quality has been improved by remediation involving individual contacts to depositors. Search still has very basic functionality, but plans have been made to allow metadata search, formulate complex queries, and add query and search results widgets.

Dr. Gerard Kleywegt presented an outline of the efforts by EMDB directed at validation. In the current thinking, approaches to validation are not imposed by the deposition site but will emerge from community consensus, to be facilitated by meetings of a task force with wide community representation. Such validation task forces have been convened in one-year intervals for X-ray, EM, and NMR, with

another one following for SAXS in 2012. The EM task force, convened in September 2010, made a number of detailed recommendations which are summarized in an article to be published in Structure. Among the recommendations, scientific journals should be encouraged to stipulate map deposition before an article is published.

Dr. Steve Ludtke gave a presentation about the Cryo-EM Modelling Challenge Workshop in Hawaii in January 2011 with 58 participants representing ten research groups, among these the major modelling groups (Levitt, Baker, Sali, Schulten). In all, 136 maps were submitted and 13 software packages were considered. A total of 13 maps of 6 biological targets in resolution ranges from 2.5 to 24Å were analyzed. Dr. Ludtke showed results obtained with the ribosome, GroEL/ES and Aquaporin as examples.

Dr. Matt Baker and Greg Pintilie (Baylor) described the efforts that have gone into building, integrating and validating tools for cryo-EM maps. Progress over the years, from manual model building to automated tools such as Pathwalker and Gorgon were summarized, and possible scoring of fitted structures into given density maps were discussed.

Overall Comments

The EMDB has clearly solidified its position and leadership in the EM community, and has worked toward establishment of a versatile user-friendly platform for the deposition of three-dimensional EM maps. Transparency of the origin of deposition (i.e., from which side of the Atlantic) and coordination of the two servers has now been fully achieved. Particularly commendable are EMDB's efforts to facilitate community consensus on validation and similar outreach activities in the areas of modelling and data management. Given the central role EMDB has attained in the past years, consideration could be given to an expansion of its website to include community networking. This committee wholeheartedly endorses the plans of EMDB to work on developing validation standards and tools, as formulated in some of the specific aims of the NIH grant renewal, but expresses some concerns about the appropriateness of its role in the development of specific tools such as 3D segmentation.

Specific Comments on Past Year Activities

Although the Cryo-EM Modelling Challenge Workshop was deemed partially successful, there was a feeling, articulated by some workshop participants and alluded to by Dr. Steve Ludtke in his presentation, that the Workshop's goals were not defined well enough. As reported by Dr. Ludtke, there were six different biological targets and 13 density maps with resolutions varying from 2.4 to 24 Å. As a result, there were sometimes only 2 or 3 modellers for a particular combination of modelling software and biological target, limiting the ability to compare results. For future meetings, it would be important to have definable goals for the outcome of such a workshop. For example, how should modelling success be measured? - perhaps by a Fourier Shell Correlation between map and model? How should the protein stereochemistry be refined? How should the model parameters such as B-factors on particular residues be treated? It would be most useful to have a list of goals before the participants arrive, with an obligation to summarize the conclusions before departure.

In addition, the choice of venue should be more carefully considered in the planning of future meetings, given the economic reality and the cost of airfares outside the continental US.

Comments on Plans for the Coming Year and the Specific Aims of the Renewal Grant

There are four Specific Aims. The first Aim relates to the testing of EM data using currently available validation software. The 2nd, 3rd and 4th Aims relate to the need and procedures for archiving all EM maps and structure coordinates referred to in published articles. There is no question on the necessity and need for the development of archiving procedures. The only major concern is for defining what data should be archived in view of the large quantity of data that is now being produced by tomographic studies. The advice of this Committee is to archive only those tomograms that were actually mentioned in any ensuing publication. As an optional resource for authors, deposition of all data relating to the study -- namely tilt-series, tomograms and sub-tomogram arrays -- should be made possible. The Committee would recommend considering a "cloud"-type storage system for these tomographic data sets, which would make them available across various platforms. Viewing tools for the tomograms should be provided, as achieved through an interface such as the O.M.E.

Another question touched upon in the Committee's deliberations was whether and to what extent deposition of raw data (micrographs, single-particle images) used in published single-particle reconstructions should be facilitated and encouraged. Because of the need for test data on which to hone new community-developed algorithms, such as for classification, the EMDB is urged to provide the means for such optional depositions.

More contentious is the work (Specific Aim #1) on gathering and developing programs for validating published, or about to be published, EM results. At this time about half a dozen programs have been collected and tested for rigid-body fitting into cryoEM maps, and about an equal number of programs that are designed for flexible fitting. At the same time work has started to use the accumulated information for writing more perfect programs that have all the advantage of the gained wisdom. This procedure could constitute some conflict of interest unless great care is taken to fully express the origin of ideas and procedures. In addition, the knowledge gained should be made quickly and freely available to allow all those interested to implement improved procedures at the earliest stages. A website acting as community platform (see below) would be an excellent vehicle to achieve this.

The limited amount of information available suggests that many of the currently available validation programs fail to take care of the most obvious problems such as clashes between different fitted subunits and lack of preservation of good stereo-chemical geometry.

Suggestions for a Community Platform

The EMDB has made great strides in establishing a new community site and tools for making deposition of structures easier within the EMDB/PDB. As the EMDB has the opportunity to become the hub of the structural biology community, it is worth exploring the use of more social networking tools to allow the community to interact on-line through community forums and wikis, rather than exclusively through e-mail. In that way, the collective experience and knowledge of the community can be made available through search engines and the portal. These tools are also effective

means of keeping content up to date, as any tutorials or manuals produced quickly become out of date.

Constitution of Advisory Committee

Currently the AC is reconstituted by the PI each year in an ad-hoc way, and a chair is assigned with little advance notice. The AC of this meeting feels that the ensuing lack of continuity poses a problem in advising the PI and co-PI in a consistent way, and the PI is therefore urged to consider appointing a standing committee, given the importance and wide impact of the issues the EMDB seeks to address in the coming years.