

**2010 EM Databank (EMDB) Advisory Committee Meeting
Rutgers University, September 30**

The EM Databank (EMDB) Advisory Committee met at Rutgers University from 8:45 am – 3:30 pm, Thursday, September 30, 2010. Members in attendance were Joachim Frank, Chair (Columbia University), Tao Ju (Washington University), Maryanne Martone (UC San Diego) and Andrej Sali (UCSF). Attendant observers were Michael Rossmann (Purdue University) and Paula Flicker (NIGMS).

Dr. **Wah Chiu** (PI, Baylor) gave an introduction into the organizational structure and personnel of the EMDB and its current funding on both sides of the Atlantic, NIH grant (1R01 GM079429) which supports teams in the US at the National Center for Molecular Imaging at Baylor and PDB at Rutgers, and BBSRC grant BB/G0225771/1 which supports the PDBe team at the European Bioinformatics Institute and Professor JR Swedlow at the University of Dundee. After the introduction the meeting comprised a series of presentations on progress and future directions of EMDB. Dr. **Helen Berman**, co-PI, gave an overview of the activities of EMDB over the past year. She explained the way suggestions and concerns contained in the AC report from 2009 were addressed, including comments on EM formats, harvesting of meta-data, ease of deposition process, and outreach activities. In fact, two activities, the meeting of the EM Validation Task Force in the preceding two days and the EM challenge workshop planned for 2011, were partially in response to suggestions by the AC. EMDB depositions are soon to reach the symbolic mark of 1000, and associated PDB depositions have been made for 1/3 of these maps. The deposition and retrieval system has been streamlined and, very importantly, full coordination of the US and European sites with synchronous updating has been implemented. The EMDB had a prominent presence at 8 meetings in the past year, and lists 7 publications during that time frame. Dr. **Batsal Devkota** went into technical details of the EMDB site design, contents most frequently accessed, logistics of pdb-EMDB linkage, map remediation for file header consistency, and the weekly update cycle coordinated with PDBe. Dr. **Cathy Lawson** gave an outline of priorities for improvements and additions, which include integration of EMDB with the world wide PDB deposition and annotation system, and incorporation of EM density map validation based on the recommendation of the EM Validation Task Force.

Drs. **Matt Baker** (Baylor) presented his efforts on development and application of EM tools, including EMAN, Gorgon, and Coot. **Gerard Kleywegt** (EBI) presented technical aspects of validation of atomic structures determined by X-ray crystallography, as an example of what is needed by the EM community.

Dr. **Greg Pintilie** (Baylor) and **Powei Feng** (Rice University) gave an overview of the visualization tools available from EMDB and discussed recent developments in off-line segmentation of EM maps and tomograms. EMDB currently supports two online tools for visualization, (1) **EM Viewer**, a light-weight Java applet developed by Powei Feng under the guidance of Dr. Joe Warren (Rice) that supports display of raw and labelled 3D EM maps, and (2) **OpenAstex Viewer**, an open source visualization tool that is capable of displaying both EM maps and PDB structures. Dr. Pintilie presented and demonstrated **Segger**, a semi-automatic segmentation tool for EM maps implemented as a Chimera plug-in based on water-sheds and scale-space filtering. Powei Feng presented and demonstrated an interactive segmentation program for particles and filaments in tomograms of platelets, which allows “snapping” of user-drawn curves onto shapes of interest and automatic tracing through multiple slices.

Overall, the Advisory Committee is impressed by the progress EMDB has made since last year, accomplishing most of the declared goals. They have made improvements and changes in response to the suggestions in the Committee's 2009 report. In our view, the members of the EMDB team have a unique opportunity to form the nucleus of the community interested in determining macromolecular structures by EM, particularly if they start to establish more interactive on-line forums through Web 2.0 tools. They acknowledge that there are many additional tasks to be accomplished and that priorities must be set because of limitations in resources. For this reason, we believe it is particularly important that the EMDB define its scope carefully.

We saw some examples of tools that were designed for segmentation of cellular level tomography. However, these types of tools are in use and being developed by many other groups that are engaged in tomography of subcellular structures and by those working on serial electron microscopy datasets. While the tools that were shown looked promising, we are not sure they fell within the scope of the EMDB, especially in light of the need to prioritize. A similar reservation applies to specialized segmentation, labeling, and visualization tools for macromolecular complexes. Because it is so difficult to create and maintain these types of resources, the EMDB should be reaching out to join forces with other communities and establish collaborations to help build and promote these types of resources. At the least, EMDB should consult and learn from their experiences before embarking on extensive development in these areas.

The EMDB efforts on bringing together the electron microscopy and molecular modeling communities have been very productive thus far. For example, preceding the EMDB advisory meeting, there was the inaugural 2-day meeting of an EM Validation Task Force, during which approximately 25 participants discussed issues in the interpretation of EM maps of importance to the PDB. As a result of the meeting, a paper will be published in a structural biology journal, including recommendations that will eventually lead to community-wide standards for interpreting, archiving, and distributing EM maps and molecular models based on these maps. EMDB has played a key role in this effort thus far; in fact, the effort was initiated by EMDB.

EMDB should continue to facilitate this effort, as it is most uniquely positioned to do so. In addition to creating a benchmark of EM datasets for systems of known structure (mentioned elsewhere in this report), EMDB should also consider continuing to lead an effort to finalize community-wide standards for molecular modeling based on EM maps. This effort could include mobilization and recruitment of research groups interested in contributing at a substantial level, organizing meetings of these contributors, and facilitating the development of a software suite for certain interpretation and assessment tasks. To support such an effort on an appropriate scale (*cf.* potential involvement of many other research groups currently not contributing to EMDB), EMDB could consider asking for an administrative supplement to increase its budget.

A minor suggestion is for EMDB to do its best to identify the composition and stoichiometry of all samples corresponding to its already archived EM maps; as new structural information becomes available for many individual proteins and other molecules, knowing the identities of

the molecules in the sample for which an EM map has been obtained is essential for any future molecular modeling efforts based on this map.

There was a strong sentiment expressed in the Committee advocating the elimination of holding time for EM maps, which is currently two years after publication. Arguments for continuation of the present policy apparently repeat arguments that were made in the X-ray crystallography community at the early stage of pdb depositions, but they were eventually overruled by considerations taking account of the interests of the scientific community as a whole.

A special comment is in order concerning a suggestion that was debated during the EM Validation Task Force meeting. We support the motion to create benchmark data in the pipeline of EM data creation and analysis and make them freely-accessible on EMDB. There are at least two reasons to do so. First, the benchmark data will allow researchers from a broader community of computational science to contribute to the methods and tools related to EM. Their participation will be important to ensure that state-of-art techniques are used in the analysis and visualization of EM data, and will also foster diversity in the software programs. Second, the benchmark data will be crucial for validation of the various methods and tools on a common basis. It is only with this validation that the users in the EM community will understand the strength and weakness of each tool, and use them appropriately. If provisions for deposition of benchmark data are difficult to reconcile with the present scope of the grant, the PIs of EMDB may wish to consider putting in a supplement application to the NIH grant.