RCSB Protein Data Bank Advisory Committee
Report of September 14th 2012 Annual Meeting
Rutgers University, New Brunswick, New Jersey

Chair: Cynthia Wolberger

Membership: Stephen K. Burley, R. Andrew Byrd, Jack Chirikjian, Wah Chiu, Paul Craig, Roland Dunbrack, Andrzej Joachimiak, Ann C. Palmenberg (absent), Sue Rhee (absent), Andrej Sali (absent), David B. Searls, Cynthia Wolberger, and Cathy Wu

US Government Representatives: Peter McCartney (NSF representative, present for telephonic discussion), Ward Smith (NIH-NIGMS representative), DOE representative (absent), and NIH-NLM representative (absent)

RCSB Leadership: Helen Berman, Phil Bourne, Martha Quesada

RCSB PDB AC E-mail Addresses:
cwolberg@jhmi.edu, stephenkburleymd@gmail.com, rabyrd@ncifcrf.gov, jgc@georgetown.edu, wah@bcm.edu, paul.craig@rit.edu, roland.dunbrack@fccc.edu, acpalmen@facstaff.wisc.edu, andrzejj@anl.gov, rhee@acoma.stanford.edu, david.b.searls@gmail.com, wuc@udel.edu, sali@salilab.org

US Government Agency Representative E-mail Addresses:
pmccartney@nsf.org, roland.hirsch@science.doe.gov, ward.smith@nih.gov

RCSB Leadership E-mail Addresses:
berman@rcsb.rutgers.edu, bourne@sdsc.edu, mquesada@rcsb.rutgers.edu

Executive Summary

The Advisory Committee to the Research Collaboratory for Structural Bioinformatics (RCSB) met in New Brunswick, New Jersey on 14th September 2012 to consider management and enhancement of the Protein Data Bank (PDB).

Agenda items included
(1) Responses to 2011 RCSB PDB AC Recommendations;
(2) State of the PDB;
(3) Data In: Deposition, Annotation, and Remediation;
(4) Data In: Common Deposition and Annotation Tool Development;
(5) Data Out: Website, Mobile Apps and Hardware;
(6) Education and Outreach;
(7) Future plans; and
(8) Plans for non-competing grant renewal.

The meeting was held in the Rutgers University Center for Integrative Proteomics and opened by Professor Ken Breslauer, Dean of Life Sciences and Vice-President for Health Science Partnerships.

The Overview was presented by Berman, beginning with responses to the 2011 RCSB PDB AC Recommendations. A summary of recent activities was subsequently provided by Young, Quesada, Bourne, Rose, Zardecki, and Dutta.

The Committee remains of the opinion that the evolving PDB represents the preeminent source of experimentally determined macromolecular structure information for research and teaching in biology, biological chemistry, and medicine. The Committee again commends the RCSB and US Government agencies for their ongoing commitment to ensuring facile, open access to a secure, singular experimental data archive for macromolecular structural biology that will be maintained in perpetuity for the public good.

The RCSB PDB web site and query software continue to undergo substantial re-engineering to keep pace with the demands of the ever-expanding global user community. PDB-user interactions are facilitated by the Electronic Help Desk, Electronic News, Molecule of the Month (in collaboration with David Goodsell at the Scripps Research Institute), Conference Presentations/Participation, Workshops, and Publications. Outreach efforts include various posters, multiple task force and local advisory meetings, informal one-on-one discussions, formal interviews of PDB users, college courses, an RCSB PDB *Mobile* application for smart phones and tablet devices, and presentations/workshops for educators (kindergarten through graduate/professional). The RCSB also continues to strengthen its involvement in collaborative and consultative relationships. Most important among these is the wwPDB, a global partnership involving RCSB PDB, BMRB, PDBe, and PDBj.

The Committee reiterates its view that Berman's leadership of the NIH-funded Protein Structure Initiative Structural Biology Knowledge Base is an important adjunct to her RCSB PDB responsibilities.

Responses to 2011 RCSB PDB AC Recommendations

• Data In: Continue ongoing efforts to improve representation of ligands
RCSB PDB Response: New tools and dictionaries have been developed for complex ligands.

• Common Deposition and Annotation Tool: Estimate impact of tool on annotation speed and efficiency to help project load balancing requirements.
RCSB PDB Response: Load balancing outline drafted: impact testing to follow.

• Outreach: Improve mobile device access to outreach and education resources
RCSB PDB Response: Improved version of RCSB PDB *Mobile* for Apple iOS devices (iPad/iPhone/iPod) with broader functionality; Android version in development.

PDB Metrics

In aggregate, 9250 (9768*) depositions were processed between January 1st and December 31st 2011 with a two-week average turnaround (* 2012 projection).

Breakdown of depositions by discipline in calendar 2011 was as follows:

X-ray:	8,548	(92% of total, up from 8,186 in 2010)
NMR:	582	(~6%, from 604 in 2010)

EM:	90	(<1%, up from 73 in 2010)
Other:	27	(<1%)

Breakdown of depositions by wwPDB processing site in calendar 2011 was as follows:

RCSB PDB:	5938	(64%)
PDBj:	1816	(20%)
PDBe-EBI:	1496	(16%)

Breakdown of depositors by location in calendar 2011 was as follows:

North America	46%
Europe	27%
Asia	17%
Industry	7%
South America	1%
Australasia	2%
Africa	<1%

Monitoring of www.rcsb.org continues to show significant year-on-year increases in website traffic for April 2012 *versus* April 2011:

Website Visits:	+10% (617,748, up from 557,911)
Unique Visitors:	+11% (258,130, up from 230,825)

2011 RCSB PDB AC Discussion

Data In: Deposition, Annotation, and Quality Assessment

A review of current structure deposition/annotation systems and ongoing efforts regarding data remediation was presented by Dr. Jasmine Young, Lead Annotator. In 2011, RCSB PDB Annotators processed ~64% of all new PDB entries (significantly exceeding the proportion of entries coming from North America, ~46%), which represents a 3 percentage point increase over 2010. As before, the complexity of PDB depositions continues to increase, particularly in the proportion containing non-covalently bound ligands, including inhibitors. The ever-increasing workload will be handled in the near future by the common deposition and annotation system (described in more detail in the next section). A focus of the presentation and discussion was on efforts and challenges in remediating existing PDB entries. The most recently completed remediation, released in July 2011, addressed annotation of inhibitors and antibiotics that facilitates searches of small molecules and peptides against the new Biologically Interesting molecule Reference Dictionary (BIRD). Ongoing remediation efforts that will be implemented over the next two years include carbohydrates, post-translational modifications, metal-containing ligands, translation of non-standard crystal frame and dissociated assemblies, and recalculation of B factors. There was an extended discussion of several issues related to remediation. One issue centered on how depositors would be informed about changes to their depositions and whether they would be given the opportunity to remediate the coordinates themselves. A second issue that was discussed was how to deal with multiple versions of the same coordinates that will result from remediation. While the current policy is to assign a new ID

and obsolete the old coordinates, there was discussion of assigning different version numbers instead. A third issue was how the PDB communicates with authors about the deposition of biological assemblies, as there are cases where the published biological assembly does not match that identified by the author. While many of these can be fixed by remediation, the communication issue will need to be addressed. The PDB will devise a plan to deal with these issues with input from the wwPDB Advisory Committee, with a goal of finalizing this by the end of this calendar year.

wwPDB Common Deposition and Annotation Tool

The Advisory Committee was updated by Dr. Martha Quesada, Deputy Director, on progress in developing a common tool for all wwPDB sites that will enable automated deposition and annotation. The Committee was impressed by the progress over the past year and was pleased to learn that the wwPDB is on schedule to roll out the first version at the end of 2012. This tool will be critical to the ability of all PDB sites to handle the continually increasing number of depositions, which are estimated to increase by up to 70% over the next five years. Improvements to processing modules, new standardization on PDBx and the decreased need for email exchanges should greatly streamline the deposition process once the automated tool has been fully implemented. A remaining challenge that is recognized by the PDB leadership is how to manage the transition period, in which the tool is adopted by all three wwPDB sites and annotators must process entries using both the old and new systems. The Committee urges the PDB to consider the final testing and implementation phases carefully to ensure a smooth transition. Adequate training of users and ongoing communication will be important in this effort.

Data Out: Website and Impact

Dr. Philip Bourne, Associate Director, and Dr. Peter Rose described the newest additions to the searching and browsing functionality of the PDB web site. An impressive array of new features were presented, some already implemented and others to be rolled out this fall, that further add to the functionality of the web site and facilitate connecting macromolecular structure with biology. The Committee was pleased to see that previous suggestions to identify and serve different user communities were acted upon in the design of new search features. The new feature providing pre-calculated structural alignments, as well as a tool for generating customized alignments, was viewed as a highly desirable addition. Expanded search options including Pfam annotation and access to LINK records further increase database functionality. Another attractive feature that will be implemented this fall is the option to view search results in different formats, including abbreviated lists and structure icons. The new Timeline feature, which makes it possible to display the chronology of structure determinations for an individual investigator or type of macromolecule, was viewed as a useful pedagogical tool that will particularly appeal to younger users.

The new RCSB PDB *Mobile* app, designed to work on the iPhone/iPad platforms and made available recently, is an impressive new addition that further expands the ways in which the PDB can be accessed by users. An Android version is under development and should be available soon. There was some discussion of the relative priority that should be given to mobile app development and support in the future, in view of a possible flat or decreased budget in the future. In light of data indicating that mobile device users are expected to exceed desktop users by 2014, it was noted that these devices are likely to be of increasing importance in the very

near future. The Committee encourages the PDB to continue to monitor the relative use by mobile devices so that future efforts can be allocated accordingly.

The Committee was enthusiastic about the development of a Structural View of Biology resource, which will connect protein families, metabolic and signaling pathways, and sets of drug targets to structural information in the PDB. This can only broaden the utility of the PDB to non-experts in the scientific community in addition to further serving education efforts.

As in previous years, worldwide website utilization continues to grow, with an 8.5% increase in overall utilization and a 14% increase in unique visitors. The US remains the most active user of www.rcsb.org by country, with India and Japan now second and third, respectively. The main RCSB PDB web page remains the most-used entry point by far (>5M visits), but direct access of specific pages through Google and other search engines is significant (~2M). The Committee views favorably the plans to balance the processing workload on individual wwPDB sites.

Education and Outreach

A review of outreach and educational activities was presented by Dr. Shuchismita Dutta and Christine Zardecki. The committee was highly impressed by the large array of education and outreach activities and the clear impact they are having on the different target communities. The PDB-101 education section, with its Molecule of the Month, has been a remarkable success as made evident by the tremendous increase of 18% in site visits over the past year – more than double that for the RCSB PDB web site as a whole. The Committee views this as a testament to the high quality and outstanding success of the RCSB PDB education and outreach efforts.

An outstanding array of education and outreach offerings were presented, some that have already been implemented and others that are in the planning stages. In addition to programs targeting K-12 and undergraduate students, the RCSB PDB has now targeted K-12 teachers, which constitute a new audience. The Committee views training teachers as an excellent approach to broadening the impact of the RCSB PDB's educational efforts.

The Molecular Anatomy Project (MAP) continues to be viewed as one of the most outstanding of the RCSB educational offerings. While the initial attempt to obtain NIH funding for MAP-based summer courses was not successful despite favorable reviews, the Committee is optimistic that a resubmission this fall will succeed, thereby securing funding in mid 2013. There was considerable discussion of the development of MAP-based on-line courses modeled on offerings from existing e-learning platforms such as Coursera. The Committee strongly encourages the RCSB to work with one of these entities in developing on-line courses. In addition, the committee urged the RCSB to seek additional external funding for these on-line offerings and to consider a broader variety of funding mechanisms including SBIR grants. In facing their perennial challenge in carrying out educational and outreach efforts with limited personnel and financial resources, the Committee urges a periodic reevaluation of individual expenditures and their relative importance to the overall education and outreach mission.

Journal Interactions

A review of Journal Interactions was presented by Ms. Christine Zardecki. The Committee was pleased to learn that Science and Nature Communications have joined the list of journals that notify the PDB when papers reporting structures appear on-line, and endorses continued efforts to get all journals to comply. The Committee reiterates its strong support for the wwPDB efforts

to persuade all journals to supply structure validation reports to authors for inclusion with manuscript submissions. The Journal of Biological Chemistry and iUCR journals are already on board, and discussions with PNAS and eLife are ongoing. Most journals are still reluctant to act unilaterally out of fears that this will slow down the reviewing process and put off authors who are fearful of competition. The Committee encourages continued discussion with journals and hopes that increasing familiarity with pre-publication reports will lead to broader acceptance, as has historically been the case with prior changes in coordinate and data release policies.

Improvement of database searches

Dr. Helen Berman presented plans for facilitating searches for information that is currently difficult to retrieve from PDB entries using current search tools. Some examples given were queries about experimental methods, phasing, crystallization conditions or identification of proteins in a particular pathway. A plan was presented for identifying query types, optimizing data for searches, making additional post-processed data available and establishing external reference files containing additional data and classifications. There was also recognition of the need to make it possible to construct more complex searches as well as enable additional forms for representing results. The Committee commends the RCSB for being proactive on this issue and for their responsiveness to the user community.

Plans for financial support

There was a discussion of the 2013 non-competitive grant renewal with Dr. Peter McCartney of the NSF, who participated by telephone. The strong message to the RCSB leadership and the Advisory Committee was the necessity to collect information demonstrating the broad impact of the PDB on different communities. It was clear that demographics on database users, including usage by undergraduate colleges, institutions that serve underrepresented minorities, K-12 institutions (both public and private) and the pharmaceutical industry will be critical to the success of the grant renewal. The Committee strongly recommends that the RCSB survey the community of users and obtain both quantitative and qualitative information so that they are well positioned to make a persuasive case for continued funding.

There was additional discussion of other prospects for funding from additional grants for projects that leverage the information and tools provided by the PDB. It was agreed that these should involve principal investigators outside the PDB and projects or training grants that synergize with PDB core activities.