RCSB Protein Data Bank Advisory Committee
Report of October 31st 2011 Annual Meeting
Rutgers University, New Brunswick, New Jersey

Chair: Stephen K. Burley

Membership: R. Andrew Byrd, Jack Chirikjian (absent), Wah Chiu (absent), Paul Craig, Roland Dunbrack (absent), Andrzej Joachimiak, Ann C. Palmenberg (absent), Sue Rhee, Andrej Sali, David B. Searls, Brian Shoichet (absent), 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:
sburley@lilly.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, shoichet@cgl.ucsf.edu, cwolberg@jhmi.edu, 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 31st October 2011 to consider management and enhancement of the Protein Data Bank (PDB).

Agenda items included
(1) Responses to 2010 RCSB PDB AC Recommendations;
(2) State of the PDB;
(3) Data In: Deposition, Annotation, and Quality Assessment;
(4) Data In: Common Tool Development;
(5) Data Out: Website and Database;
(6) Education and Outreach; and
(7) Journal Interactions.

The meeting was held in the just-completed Rutgers University Center for Integrative Proteomics and opened by Professor Ken Breslauer, Dean of Life Sciences and Vice-President for Health Science Partnerships. The scale and quality of the building and Berman's role as founding center director speak volumes re the institution's commitment to the PDB and the scientific and programmatic importance of the RCSB PDB at Rutgers.

The Overview was presented by Berman, beginning with responses to the 2010 RCSB PDB AC Recommendations. A summary of recent activities was subsequently provided by Young, Zhuravleva, Quesada, Westbrook, Bourne, Rose, Dutta, Quin, and Zardecki.

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 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, a PDBMobile 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, BMRB, PDBe, and PDBj.

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

Finally, the Committee wishes to congratulate the RCSB PDB team on the outstanding success of the PDB40 scientific meeting at Cold Spring Harbor. The quality of the program and the palpable excitement of the multi-generational registrants served to underscore the long scientific tradition that the PDB upholds, together with its continuing importance to the progress of the biological and medical sciences.

Responses to 2010 RCSB PDB AC Recommendations

• Common Deposition and Annotation Tool: Quantitative estimates of improvements in data processing.
RCSB PDB Response: Processing time for ligands cut by 70%, with new interface; benchmarking to continue.

• Data Out: What are the benefits and costs of smart phone/iPad developments.
RCSB PDB Response: Survey of PDB*Mobile* users. Improved website appearance for mobile devices.

• Outreach: Add Education Representative to AC.
RCSB PDB Response: Added Jack G. Chirikjian, Ph.D., Georgetown University.

PDB Metrics

In aggregate, 8,878 (9,088*) depositions were processed between January 1st and December 31st 2010 with a two-week average turnaround (* 2011 projection).

Breakdown of depositions by discipline in calendar 2010 was as follows:

X-ray:	8,186	(92%, up from 7,640 in 2009)
NMR:	604	(~7%, up from 592 in 2009)
EM:	73	(<1%, up from 51 in 2009)
Other:	15	(<1%)

Breakdown of depositions by wwPDB processing site in calendar 2010 was as follows:

RCSB PDB:	5,464	(62%)
PDBj:	2,041	(23%)
PDBe-EBI:	1,373	(15%)

Breakdown of depositors by location in calendar 2010 was as follows:

North America	46%
Europe	29%
Asia	16%
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/May 2011 *versus* April/May 2011:

Website Visits:	+16% (646,682, up from 556,548)
Unique Visitors:	+17% (274,335, up from 235,083)

2011 RCSB PDB AC Discussion

Data In: Deposition, Annotation, and Quality Assessment

A review of current structure deposition/annotation systems and ongoing improvements was presented by Dr. Jasmine Young, Lead Annotator. In 2010, RCSB PDB Annotators processed ~62% of all new PDB entries (significantly exceeding the proportion of entries coming from North America, ~46%). As in previous years, the complexity of PDB depositions continues to increase, both in terms of polymer molecular weight and the proportion containing non-covalently bound ligands. Recently completed Data In improvement activities include provision of a new Anonymous PDF Validation Server and creation of PDB_extract V3.11 and SF Tool, and improved methods for handling ligands (all of which are destined for incorporation into the Common Deposition and Annotation Tool, see below). Recently completed remediation projects include work on Biological Assemblies, B-Factors, Non-standard Crystal Coordinate Frame, Hybrid Neutron/X-ray Structures, Peptide Inhibitors/Antibiotics, and Non-standard Polymer Linkages. Further remediation efforts in progress include work on Carbohydrates and Post-translational Modifications. Work on Chemical Quality Assessment was summarized by Dr. Marina Zhuravleva. Current PDB archive holdings include ~15,000 unique small molecule entities, including carbohydrates, metal-containing complexes, peptide-like molecules, small organics, and ions, each of which is described in the Chemical Components Dictionary. Given the biological and medical importance of the many co-crystal and other complex structures in the PDB, the Committee strongly supports ongoing efforts aimed at providing the most informative depictions of all ligands represented in the archive.

wwPDB Common Deposition and Annotation Tool

A description of progress by wwPDB collaborators towards establishing common, global deposition and annotation tools was presented by Dr. Martha Quesada, Deputy Director. The Committee was pleased to learn that the project's 2011 goals are in hand and that substantial completion in 2012 is likely. Consistent annotation and workload balance among PDB data deposition sites is particularly important given the continued growth of the size and complexity of archive. The flexibility/modularity of the underlying architecture of the common tool is similarly important for the enterprise, as deposition of structures coming from newly established hybrid methods becomes a common occurrence. Thereafter, Dr. John Westbrook provided a review of the System Architecture underpinning the common tool, which was similarly well received.

Once the new functionalities of the Common Deposition and Annotation Tool are adopted across the wwPDB, the Committee reiterates its recommendation that rigorous estimates of speed and throughput be made, with the goal of understanding how best to balance load among the various deposition sites. It will also be important to model the longer term impact of various deposition growth scenarios to plan for future contingencies, including the possibility that one or more of BMRB, PDBe, and PDBj ceases operations.

Data Out: Website and Impact

Dr. Phil Bourne, Associate Director, described progress towards implementing the RCSB's strategic plan for creating more informative "contextual" views of the archive. The Committee was impressed by recent Data Out improvement activities and continues to view the evolving RCSB PDB website as an essential component of its mission. Both the look and feel of the website and its usability engineering should remain front of mind as RCSB PDB leaders determine how best to enhance the public face of the resource and make its value fully commensurate with that of the underlying data archive.

Improving the search/query functions of the website remains critical for the global user community and is equally important for the success of the RCSB PDB in its friendly competition on Data Out with PDBe, PDBj, and BMRB. The demonstration of new search features, conducted by Dr. Peter Rose, exemplified both significant advances and outstanding challenges in optimizing the utility of the query system and access to the underlying archival data. Although this ongoing effort mirrors the challenge confronted daily by the mythic figure, Sisyphus, its importance cannot be overstated. We suggest that the team clearly identify and prioritize the user communities. We also strongly advise additional market research to better understand the needs of high priority user communities.

The Committee continues to be impressed by the ongoing year-on-year growth of website utilization worldwide. The US remains the most active user of www.pdb.org by country, and India and China are now second and third, respectively. Successful development and deployment of the PDB-101 view of the archive can only serve to increase utilization further. Ongoing efforts aimed at creating a "Drug View" of the PDB, adapting the website for use with mobile devices (see below), establishing a structural alignment database, and further development of web services are all strongly supported by the Committee. As work on the "Drug View" progresses, it will be important for the team to achieve synergies with available quality resources, such as the IUPHAR database of receptors, and thereby avoid unnecessary competition/duplication.

Education and Outreach

A review of current educational and outreach activities was presented by Dr. Shuchismita Dutta. The Committee continues to be impressed by the growing maturity of these efforts. Dr. Jack Chirikjian represents a welcome addition to the advisory team. Immediate efforts should be made to solicit his feedback off line, in light of his unexpected absence.

The RCSB philosophy of targeting its approaches to distinct PDB user communities is strongly supported by the Committee. Metrics documenting use of the www.rcsb.org web page continue to show significant year-on-year increases. Given the work being put into creation/development of the PDB-101 "structural view of biology", it is heartening that Molecule of the Month traffic has increased by ~10% as a direct consequence of its introduction. Additional measurement of utilization/impact will be important moving forward, particularly as RCSB leaders work to determine where best to focus limited educational and outreach resources.

The Molecular Anatomy Project (MAP) remains the most compelling of the RCSB educational offerings. Every effort should be made to obtain significant independent funding for this valuable enterprise and to extend the reach of the resource to different geographies and distinct student audiences. MAP lectures could be usefully recorded and packaged to enable access *via* mobile devices (see below).

A detailed review of current mobile outreach activities was presented by Dr. Greg Quinn. The Committee was impressed by the pace of development in this arena over the past year, and fully supports development of "Apps" that facilitate mobile access to the RCSB PDB website. It is remarkable that the "limitations" of various smart phone and tablet form factors are actually proving to be beneficial constraints by driving streamlining improvements in the website.

As discussed in the 2009 and again in 2010, the fundamental challenge facing the RCSB remains the need to leverage limited financial and personnel resources available for education and outreach by broadening the appeal and reach of the PDB and assuming a more strategic role. Further investment is encouraged to maximize mobile device access to PDB-101, Molecule of the Month, RCSB posters, MAP courses/lectures, Science Olympiad exercises, newsletters, etc. Exploitation of Social Networking tools should also be investigated.

Journal Interactions

A review of Journal Interactions was presented by Ms. Christine Zardecki. The Committee strongly endorses ongoing efforts to ensure that definitive PDB entries are released coincident with online publication in advance of the print edition. The Committee is similarly supportive of wwPDB efforts to improve journal refereeing and reduce the incidence of structure publication retractions by providing official structure validation reports to authors for inclusion with manuscript submissions. The Committee reiterates its earlier commitment to work with wwPDB leaders and advisors to the wwPDB, PDBe-EBI, PDBj, and BMRB to develop and circulate a petition among high-profile structural biologists and super users of the PDB archive calling on journal editors to require submission of wwPDB structure validation reports together with scientific manuscripts.