

**RCSB Protein Data Bank Advisory Committee
Report of September 26, 2013 Annual Meeting
Rutgers University, New Brunswick, New Jersey**

Chair: Cynthia Wolberger

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

US Government Representatives: Peter McCartney (NSF representative, present for Skype discussion)

RCSB Leadership: Helen Berman, Phil Bourne, Stephen Burley, Martha Quesada

RCSB PDB AC E-mail Addresses:

cwolberg@jhmi.edu, rabyrd@ncifcrf.gov, jgc@georgetown.edu, wah@bcm.edu, kirk.clark@novartis.com, paul.craig@rit.edu, roland.dunbrack@fcc.edu, acpalmen@facstaff.wisc.edu, andrzej@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

RCSB Leadership E-mail Addresses:

berman@rcsb.rutgers.edu, bourne@sdsc.edu, sburley@proteomics, mquesada@rcsb.rutgers.edu

Executive Summary

The Advisory Committee to the Research Collaboratory for Structural Bioinformatics (RCSB) met in New Brunswick, New Jersey on 26th September 2013 to consider management and enhancement of the Protein Data Bank (PDB).

Agenda items included

- (1) Responses to 2012 RCSB PDB AC Recommendations;
- (2) State of the PDB;
- (3) Education and Outreach;
- (4) Data In: Deposition, Annotation, and Remediation;
- (5) Data In: Development and Implementation of Deposition and Annotation Tool;
- (6) Data Out: Website, Mobile Apps and Hardware;
- (7) Discussion with Dr. Peter McCartney, NSF: and
- (8) Future plans.

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 2012 RCSB PDB AC Recommendations. A summary of recent activities was subsequently provided by Zardecki, Dutta, Young, Westbrook, Rose and Prlic.

The Committee was once again impressed by the ongoing evolution of the PDB as it continues to develop new tools and features that serve the needs of specialists, users in the general scientific community, industry and the K-12 and higher education communities. The ongoing commitment of the RCSB and US Government agencies to maintaining facile, open access to a secure, singular experimental data archive for macromolecular structural biology ensures that this critical resource will be maintained in perpetuity for the public good. The Committee was thus gratified by the highly favorable review this year by the NSF site visit panel and commends Dr. Berman and her team for this well-deserved affirmation of their work. With this non-competing renewal behind them, the Committee encourages the RCSB to begin to think ahead to devise a theme for the next competing renewal. In light of the recent news that the Protein Structure Initiative (PSI) will be phased out by NIGMS, the Committee views this as an opportunity for the RCSB to integrate the PSI Knowledgebase into the PDB and seek new NIH funding to maintain it, as well as partnerships with researchers and funding agencies in Europe and perhaps elsewhere to further enhance molecular modeling capabilities.

New developments in data deposition and curation are ensuring the continued ability of the RCSB PDB to handle increasing numbers of depositions as well as greater file complexity, while enhancing the ability of the PDB to keep up with developments in the field. The completion and initial beta testing of the automated deposition tool, in collaboration with wwPDB global partners, is a major milestone. Moving forward, the Committee hopes that this tool will enable the wwPDB to balance workloads effectively among the three sites. The PDB has also continued its work towards standardizing data entries as well as adopting a new file format that overcomes limitations of the original PDB format. The development of external reference files (ERFs) containing information linked to a PDB entry is a superb addition that enhances the utility of the database while greatly facilitating future additions and modifications to information linked to structures, including additions contributed by external researchers.

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, with decisions of which features to add guided by monitoring of a variety of metrics as well as user feedback. The RCSB leadership recognizes that the web site has grown sufficiently complex to require a new and streamlined design and the Committee strongly supports their plan to make this a priority in the coming year. Outreach efforts are outstanding and have continued to expand, with new additions including an animation contest for high school students, creative teaching tools, teacher-training workshops, and development of on-line courses. Other creative approaches, such as the plan to partner with HIV advocacy groups to teach about the mechanism of anti HIV drug action, will further broaden the impact and use of the PDB.

The Committee endorsed the succession plan presented by Dr. Berman and is confident that continuity in leadership is ensured.

Responses to 2012 RCSB PDB AC Recommendations

- PDBac: Conduct a user survey in preparation for renewal proposal
Response: Conducted November-December 2012

- PDBac: Carefully consider final testing and implementation of the Common Deposition and Annotation System
Response: Detailed transition plan being implemented
- PDBac: Need to improve communication with depositors about biological assembly information
Response: New system captures biological assembly information, including experimental details
- PDBac: Monitor use of the RCSB PDB by mobile devices
Response: Usage addressed in Data Out
- PDBac: Develop of online courses
Response: R25 Proposal under review
- PDBac: Seek external funding for outreach
Response: Success with National Institute on Drug Abuse, Rutgers proposals
- PDBac: Periodically reevaluate outreach expenditures
Response: Print newsletter reviewed via survey; printing discontinued after July 2013

PDB Metrics

In aggregate, 9972 (10,700*) depositions were processed between January 1st and December 31st 2012 with a two-week average turnaround (* 2013 projection).

Breakdown of depositions by discipline in calendar 2012 was as follows:

X-ray:	9,273 (93%, up from 8,550 in 2011)
NMR:	586 (~6%, from 583 in 2011)
EM:	101 (~1%, up from 89 in 2011)
Other:	12 (~0.1%)

Breakdown of depositions by wwPDB processing site in calendar 2012 was as follows:

RCSB PDB:	6409 (64%)
PDBj:	1887 (19%)
PDBe-EBI:	1676 (17%)

Breakdown of depositors by location in calendar 2012 was as follows:

North America	41.6%
Europe	28.7%
Asia	17.8%
Industry	8.1%
South America	1.0%
Australasia	2.6%
Africa	<0.1%

Monitoring of www.rcsb.org continues to show significant year-on-year increases in website traffic for April 2012 *versus* April 2013:

Website Visits: +31% (811,545, up from 617,748)
Unique Visitors: +43% (369,516, up from 258,130)

2013 RCSB PDB AC Discussion

Education and Outreach

A review of outreach and educational activities was presented by Ms. Christine Zardecki and Dr. Shuchismita Dutta. Following the recommendation of the Committee in 2012, a survey was conducted to determine the nature of the user community and thereby gauge the impact of the PDB. Of the 973 respondents, fully 75% had never deposited a structure, indicating the impressive degree to which the PDB has become a resource for scientists outside the structural biology community. More than a quarter were undergraduates, pointing to the importance of the PDB as an educational tool. The number of self-reported underrepresented minorities was, however, quite low, and the RCSB thus intends to step up minority outreach efforts by attending the ABRCMS meeting this year.

An outstanding array of education and outreach offerings were presented. There is now a Spanish language version of selected PDB-101 materials, which represents a new and important direction for PDB outreach and education efforts that reaches a population underrepresented in STEM fields. Overall, PDB-101 is a very popular resource that is highly utilized and will benefit from the proposed creation of an on-line forum for educators and students, which can serve as an educational forum as well as a source of feedback for the RCSB. Creative new additions include an expanded array of cutout models of macromolecules and viruses, including a how-to video on assembling a model. The Committee viewed an impressive new video animation about proteins and protein structure that will be publicized at the upcoming New Jersey Science Teachers Convention. The Committee viewed all of these outreach and education developments highly favorably and suggested that offerings be periodically reviewed for their effectiveness.

The Committee is pleased that an NIH grant to support the Molecular Anatomy Project (MAP) was resubmitted and hopes the application will succeed in securing funding for this outstanding project. The Committee encourages the RCSB team to continue their ongoing efforts to look into alternative models including massive open online courses (MOOCs) that could further broaden the impact of these courses while bringing in additional revenue. 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

Ms. Zardecki updated the Committee on journal outreach efforts, which continue on two fronts: reporting to the PDB when papers describing structures come online and getting journals to agree to require authors to submit a validation report that can be supplied to reviewers. The RCSB PDB continues to dialogue with journals to improve reporting, which remains problematic at some journals. Since last year, eLife is now requiring validation reports and Nature Structural and Molecular Biology is requesting validation reports for relevant articles. The Committee is heartened by progress on this front and encourages the RCSB to continue to lobby journal editors to require this important quality check at the reviewing stage.

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. The Committee was enormously pleased to see that version 1.0 of the new automated deposition tool has been completed in partnership with colleagues at PDBe and PDBj and is currently in beta testing. This was an extremely important milestone, and it is hoped that full implementation will soon solve issues of workload balancing between the three wwPDB sites as well as address issues of data quality and consistency. The introduction of the PDBx/mmCIF format, which can accommodate large structures without splitting coordinates among multiple files as in the old PDB format, is a most welcome development that will benefit the entire community and accommodate increasingly complex structural models produced by integrative methods.

The structure validation reports, which can be obtained by authors prior to manuscript review, have now been fully implemented and have been adopted by several journals (see above). To encourage further adoption by both authors and journals, the Committee recommends that the RCSB allow the option to suppress both author and structure title in the Unreleased Entries list and thus overcome resistance from depositors working in highly competitive areas who may not wish to publicize completion of their structure while a manuscript is under review. Plans are now underway to work in partnership with EMDatabank and BMRB to implement the recommendations of the EM and NMR validation task forces for comparable standards to be applied to structures determined using electron microscopy and NMR, respectively (moreover, similar efforts are underway by the modeling and small angle scattering communities). The Committee salutes Dr. Berman and her team for their leadership in this area and views these as highly important efforts that merit the full support of RCSB and its wwPDB partners.

The Committee was updated on remediation efforts. The ligand module, which does automated searches of the Biologically Interesting molecule Reference Dictionary (BIRD), is currently in production. Current remediation efforts are focused on carbohydrates, post-translational modifications and metal-containing ligands. The analysis of carbohydrates has now been completed, and it is estimated that it will take two years to remediate all PDB entries.

Dr. John Westbrook described the introduction of a new feature called external reference files (ERFs), which can accommodate complementary information such as a chemical component dictionary, as well as facilitate periodic updates without the need to modify the PDB file itself. This development was very enthusiastically received, as were plans for additional ERFs in the future, some of which may be provided by external researchers. ERFs will incorporate additional information that will further enhance the utility of the PDB as a platform that integrates biological information derived from multiple experimental and computational approaches. The Committee viewed this as yet another outstanding example of the ability of the RCSB to meet the evolving needs of the biological community.

Data Out: Website and Impact

Dr. Peter Rose and Dr. Andreas Prlic described the newest additions to the searching and browsing functionality of the PDB web site. Several new and highly useful features have been introduced that further add to the functionality of the web site and address the needs of multiple user communities. These include the Protein Feature View, which incorporates information from UniProt that is presented in a graphical manner, Drug View, Biologically Interesting Molecules (from BIRD) and Protein Symmetry View. The committee was impressed by the growth in PDB-101 usage, which is aimed at K-12 and undergraduates and now receives ~75,000 hits per month. Use of RCSB PDB mobile, which is available for iOS and for which an Android version is now in beta testing, is growing at a very rapid rate. The Committee was pleased to see that the RCSB continues to analyze web site usage and is using this information to inform decisions regarding web site design and addition of features.

The RCSB team is aware that, with so many new tools and features introduced over the past few years, the RCSB PDB homepage has become cluttered and difficult to navigate. There was also some concern expressed that new tools being developed can be difficult to locate and may therefore be underutilized. The Committee agreed that a major redesign of the RCSB PDB web site should be a high priority once the roll-out of the automated data deposition tool is complete. During the discussion, it was noted that the experience with the redesign of the Uniprot web site could supply some useful examples, including the use of professional consultants and an analysis of how outside users utilize the site.

Plans for financial support

There was a discussion of the recent NSF site visit and plans for the next grant renewal in 2018 with Dr. Peter McCartney of the NSF, who participated via video Skype. Dr. McCartney joined the Committee in congratulating the RCSB on their positive view and expressed appreciation to Dr. Berman and colleagues for carrying out the survey of PDB users that was carried out upon Dr. McCartney's recommendation last year. He emphasized, however, the importance of continuing to monitor usage and gather additional metrics of the impact of the RCSB PDB on education, research and drug development, which will be important for the next competing renewal. Useful metrics include how many community college students use the PDB and how many Ph.D. theses and papers depend upon the PDB infrastructure. These data will be important for justifying the assertion that the PDB is an outstanding investment. In the subsequent discussion with Dr. Berman about the next grant renewal, the Committee encouraged Dr. Berman to begin thinking about the theme for the next PDB grant renewal and to consider integrative structural biology as a unifying theme. In light of Dr. Berman's ability to bring together diverse constituencies in the structural biology community employing different techniques such as molecular modeling, small-angle x-ray scattering (SAXS) and cryo-EM, the Committee sees an important opportunity to play a key role in integrative structural biology.

Succession plan

The Committee endorsed the succession plan described by Dr. Berman, which will ensure continuity of leadership when the time comes for a transition.
