RCSB Protein Data Bank Advisory Committee
Report of October 30, 2014 Annual Meeting
Rutgers University, New Brunswick, New Jersey

Chair: Cynthia Wolberger

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

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

RCSB Leadership: Stephen Burley, Helen Berman, Martha Quesada

RCSB PDB AC E-mail Addresses:
cwolberg@jhmi.edu,byrda@mail.nih.gov, jgc@georgetown.edu, wah@bcm.edu, kirk.clark@novartis.com, paul.craig@rit.edu, roland.dunbrack@fccc.edu, acpalmen@facstaff.wisc.edu, andrzejj@anl.gov, srhee@carnegiescience.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:
sburley@proteomics.rutgers.edu, berman@rcsb.rutgers.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 30th October 2014 to consider management and enhancement of the Protein Data Bank (PDB).

Agenda items included
(1) Responses to 2013 RCSB PDB AC Recommendations;
(2) Funding situation;
(3) State of the PDB;
(4) Education and Outreach;
(5) Data In: Deposition, Annotation, and Remediation;
(6) Data In: Roll-out of Deposition and Annotation Tool;
(7) Data Out: Website redesign, Mobile Apps;
(8) Discussion with Dr. Peter McCartney, NSF: and
(9) 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 Dr. Stephen Burley, who took over the leadership of the RCSB PDB from Dr. Helen Berman in July 2014. Burley updated the committee on changes to the

RCSB leadership and outlined the responses to the 2013 RCSB PDB AC Recommendations. There was also a discussion of current and future funding challenges in light of already implemented budget cuts and the imminent sunsetting of the Protein Structure Initiative (PSI). A summary of recent activities was subsequently provided by Zardecki, Dutta, Quesada, Westbrook, Young, Rose and Prlić.

The Committee wishes to express its deep gratitude and admiration for Dr. Helen Berman's leadership of the RCSB PDB from 1998 to 2014. Thanks to her vision and guidance, the PDB grew from a data repository for crystallographers into a sophisticated tool for research, education and drug discovery that is accessed by thousands of users daily world-wide, the great majority of whom are not structural biologists. The Committee is grateful for Dr. Berman's willingness to stay on as Associate Director and for her hard work in ensuring a smooth leadership transition.

The Committee applauds the RCSB PDB for a remarkably productive year that included roll-out of the automatic deposition tool, implementation of the new PDBx file format, a complete redesign of the user web portal, establishment of the wwPDB x-ray deposition and annotation system, establishment of the Small Angle Scattering and Integrative Modeling Task Forces, and continuation of multiple education and outreach activities, including commemoration of 2014 as the International Year of Crystallography. That all this was accomplished during a time that included a planned leadership transition, the departure of Dr. Philip Bourne to become Associate Director for Data Science at the NIH and an unanticipated budget cut is a testament to the outstanding leadership of Drs. Burley and Berman as well as the skill, commitment and hard work of the RCSB Rutgers and UCSD teams.

The Committee was pleased to hear about the successful roll-out of the automated Deposition & Annotation (D&A) tool, in collaboration with wwPDB global partners, which now makes it possible to balance workloads among the three wwPDB partner sites (RCSB PDB, PDBe, PDBj). Now that the D&A tool is fully functional, the Committee recommends terminating the legacy deposition system as soon as possible. The Committee was enthusiastic about plans for upcoming data remediation efforts, particularly for carbohydrate structures, and is hopeful that current and planned grant proposals will succeed in bringing in the necessary additional funding to accomplish this goal.

The Committee commends the RCSB Education and Outreach efforts, which have produced a remarkable array of tools, educational materials and events that reach a diverse audience across the K-12 and higher education community. While the Committee is impressed with the range and creativity of these activities, which include the highly successful PDB-101, a new high school video challenge and World AIDS Day activities, there is concern that these efforts are too broad-reaching and financially unsustainable in light of current funding realities. The Committee urges the RCSB leadership to reassess their education plan, determine how best to measure the impact and effectiveness of individual components, and devise a focused plan that will optimize return on the investment of time and money, even if it reduces the overall education effort.

The Committee expresses its deep concern that the continued ability of the RCSB PDB to serve the scientific, medical, industry and education communities will be compromised by decreased financial support for the RCSB PDB. The Committee commends the NSF, DOE and NIH for their long-standing support of this macromolecular data archive, which has served as an example for management of "big data" and facile open access for all stakeholders. At the same time, the ability of the RCSB to serve a growing community, while at the same time keeping abreast of developments in the increasingly diverse structural biology community, depends upon stable and adequate funding that will ensure that this critical archive for macromolecular structure data is maintained in perpetuity for the public good. While the Committee applauds plans by the RCSB PDB to seek additional sources of funding, both through alternative federal grant mechanisms as well as industry partnerships, these cannot take the place of stable core funding from the NIH, NSF and DOE.

Responses to 2013 RCSB PDB AC Recommendations

- PDBAC: Integrate Knowledge Base (KB) and RCSB PDB.
  Response: Deferred until the fate of KB is determined.

- PDBAC: Review PDB-101 materials for effectiveness, and outreach efforts for cost-effectiveness.
  Response: Fall 2014 survey is informing redesign efforts.

- PDBAC: Develop online courses with external funding
  Response: Submitted proposal is under review

- PDBAC: Monitor use of the RCSB PDB by mobile devices
  Response: Usage addressed in Data Out

- PDBAC: Major redesign of rcsb.org website is needed.
  Response: Was in progress at time of meeting (has since been rolled out).

- PDBAC: Provide option to suppress structure title in unreleased entries.
  Response: Has been implemented in the automated Deposition & Annotation system.

.

PDB Metrics

In aggregate, 10566 (10,500*) depositions were processed between January 1st and December 31st 2013 with a two-week average turnaround (*2014 projection).

Breakdown of depositions by discipline in calendar 2013 was as follows:

X-ray:	9697 (92%, up from 9269 in 2012)
NMR:	590 (6%, from 585 in 2012)
EM:	234 (2%, up from 100 in 2012)
Other:	45 (0.4%, up from 12 in 2012)

Breakdown of depositions by wwPDB processing site in calendar 2013 was as follows:

RCSB PDB:	6652	(63%)
PDBj:	2128	(20%)
PDBe-EBI:	1786	(17%)

Breakdown of depositors by location in calendar 2013 was as follows:

North America	40%
Europe	30%
Asia	18%
Industry	7%
South America	0.6%
Australasia	4%
Africa	0.1%

Monitoring on the RCSB PDB website continues to show growth in usage. The traffic during the busiest months per year is now up to 400,000k unique visitors per month compared to 350,000 previously. During an average month there are now 294,000 unique visitors per month (average over the time period July 2013- June 2014), compared to 284,000 during the corresponding period of the previous year.

2014 RCSB PDB AC Discussion

Management

Dr. Stephen Burley provided the Committee with an account of the leadership transition that took place in July 2014, when he took over as Director from Dr. Helen Berman, who is now Associate Director. Dr. Berman will retain her current roles in the Structural Biology Knowledge Base (SBKB), Nucleic Acids Data Bank (NDB) and Electron Microscopy Data Bank (EMDataBank), and will also oversee external partnerships related to integrative structure determination methods. In addition to overseeing the RCSB PDB Rutgers site, Dr. Burley has also taken over formal responsibility for the UCSD site following the departure of Dr. Philip Bourne for the NIH in early 2014. The Committee agreed with the rationale that having a single director for both sites could integrate and accelerate ongoing efforts in multiple areas.

There was a discussion of the current difficult funding situation, in which funding in real dollars has decreased as usage has continued to rise, doubling over just the past four years. The Committee was distressed to learn that, following 10 years of nearly flat funding, in which purchasing power has declined substantially, there was a 2.7% cut in the 2014 budget as compared to 2013. This decrease limits what the RCSB can accomplish and will require making hard choices so that the RCSB can focus on its core mission of data in/data out. While the new automated D&A tool should increase productivity, this tool currently only handles crystal structures. It will soon expand to handle NMR structure deposition; however, as the number of cryo-electron microscopy structures has increased dramatically and is expected to continue to do so, support is critically needed to extend the use of this tool for all structural modalities. There was a discussion of a number of different options for pursuing additional NSF and NIH funding, as well as creative solutions for obtaining funding from industry and private foundations. Dr. Burley also updated the Committee on new NSF requirements, which include devising a business model, diversity plan, assessment plan, and making adjustments to the Advisory Committee and Charter.

Education and Outreach

A review of outreach and educational activities was presented by Ms. Christine Zardecki and Dr. Shuchismita Dutta. As before, the Committee was greatly impressed by the variety and creativity of outreach and education activities. These include booths at the USA Science and Engineering Festival in Washington, DC, the American Crystallographic Association education Workshop, and IUCR meetings, outreach to minority students at the ABRCMS and SACNAS meetings, a high school video challenge that attracted marvelous entries by many students, and several courses at Rutgers and UCSD, including the undergraduate Molecular Anatomy Project, a new interdisciplinary boot camp and a pilot teacher training workshop. In addition, the RCSB PDB has produced a variety of web-based and "hard copy" materials including the popular PDB-101 site and an updated version of the PDB's Molecular Machinery poster. Additional plans were described for developing high school curricula, materials for Moocs and participation in the Science Olympiad.

Following the recommendation of the Committee in 2013, a survey was conducted to assess the effectiveness of PDB-101 and determine the nature of the user community and how this resource is utilized. Of the ~700 respondents, fully 83% had never deposited a structure and 60% were students (28% undergraduates, 32% graduate students), pointing to the importance of PDB-101 as an educational tool. The most popular feature was the Molecular of the Month, which, combined with the popularity of these images in textbooks and other media, reinforces the impact of this particular activity. The Molecule of the Month page on Ebola virus proteins garnered over 10,000 views in just a few weeks in October 2014, reinforcing the importance of PDB-101 as a resource for educating the public about the molecular basis of disease.

While the Committee found the array of PDB outreach and education activities to be truly impressive, there was concern that these are taking too many person-hours and resources during a period in which the budget has continued to shrink. The Committee fully endorses the importance of these activities and also recognizes that broader impacts is a review criterion for NSF proposals. Educational activities at both Rutgers and UCSD are also important for maintaining strong institutional ties. Given the current funding situation, however, the Committee recommends that the RCSB thoroughly evaluate their education plan, articulate a set of goals and draw up a focused education plan that will maximize impact and return. As part of this process, the Committee recommends that the RCSB leadership determine how to assess the impact of different programs and activities, both to make decisions about which programs to continue as well as to periodically evaluate programs in the future. In light of comments from Dr. Peter McCartney of the NSF (see section below), the Committee encourages the RCSB PDB to seek partners who could help in education and outreach efforts, rather than becoming directly involved.

Data In: Deposition, Annotation, and Quality Assessment

Dr. Martha Quesada gave an overview of the final testing and release of the Common Deposition & Annotation System (D&A), which was developed in collaboration with wwPDB partners. With over 2600 crystal structures deposited and annotated as of September 30, this is an outstanding development that made it possible to balance the workload among the three wwPDB sites. In light of the successful implementation of D&A, the Committee fully endorses the proposal from the RCSB to retire the legacy system for x-ray crystal structures. The Committee was also gratified to see excellent progress in developing and testing D&A versions for NMR and EM. Dr. John Westbrook covered the implementation of the new PDBx file format, a welcome development that obviates the need for split entries to accommodate large structures, as well as the structure validation server for pre-deposition evaluation. Other improvements, including improved sequence and ligand annotation, and new standards for restraints on novel chemical groups and modifications were also presented.

Dr. Jasmine Young described the process of training staff to use the new annotation system, including the successful training of the new RCSB Director. The outstanding instructional materials and tutorials that have been made available to users through wwPDB.org have clearly been instrumental in the smooth rollout of the new D&A tool. The solid plans for soliciting feedback and monitoring user experiences should ensure that needed improvements are identified and implemented in a timely manner. Dr. Young also described future plans for data remediation including correction of residual B factors, translation of coordinates into the crystal frame, update of mmCIF to version 5.0 and remediation of carbohydrate structures. The latter is clearly a major undertaking that will require additional funding. An application has been submitted to the Mizutani Foundation and there was also discussion of pursuing NIH funding through a new Common Fund RFA (RM-14-012) in glycoscience.

A point was raised about annotation of the increasing number of entries deposited in the PDB that are unpublished and therefore are not linked to any abstract describing the structure. A number of solutions were discussed, including the possibility that depositors be given the option of depositing an abstract along with their entry that describes the reasons for doing the structure, what it contains, and what was learned. Other possibilities include deposition of abstracts in BioArXIV or other online resources. The PDBAC recommends that the RCSB consider these and other options and develop plans to make more information available on these unpublished structures than is currently the case. Initially, this may be an optional component of deposition but eventually it may be suitable to require an abstract in one form or another, with the authors able to choose among a number of locations for their abstract. The PDBAC believes that the wishes of publishing houses that charge for publication and do not provide open access to published articles should not be a concern of a public resource like the PDB.

Data Out: Website and Impact

Dr. Peter Rose gave a tour of the new RCSB PDB website home page and navigation, which was completely redesigned from the ground up following a recommendation last year by the Committee. The Committee was impressed by the greatly simplified home page and easily customizable search options based on user interest. The Committee commends the RCSB PDB for the outstanding design and thoughtfulness that went into improving the user experience and website functionality. Plans to continue improving mobile device-friendly versions are encouraged as use of these platforms, while still relatively rare, is increasing.

Dr. Andreas Prlić described recent and planned improvements as well as the use of web site analytics to analyze usage and inform decisions about new features. The new Gene View feature that maps the PDB to specific genes is an attractive and useful new feature that was recently introduced and appears to be garnering interest from users. The extensive use of Google analytics by the RCSB PDB to analyze user metrics and demographics as well as identify areas for change is commendable and will help ensure that the PDB web site continues to meet the evolving needs of the user community.

Plans for financial support

There was a discussion with Dr. Peter McCartney of the NSF, who participated via telephone. Dr. McCartney was pleased to hear that RCSB PDB representatives participated in this year's annual meeting of SACNAS, which targets Hispanic and Native American students, and shared this news with his division director. This activity, together with participation in ABRCMS, which targets African-American students, was a specific previous recommendation of the 2013 site visit panel.

Dr. McCartney was asked to comment on the extent to which the RCSB PDB should devote resources to education and outreach. He responded that the PDB did not need to be directly involved in all outreach activities, but could instead partner with other organizations who are already reaching target audiences. Dr. McCartney was also asked for advice on obtaining funding to develop approaches for managing coordinates and data generated by integrative structure determination methods, which was the topic of a recent wwPDB workshop at Hinxton, U.K. Dr. McCartney suggested using the EAGER (Early Concept Grants for Exploratory Research) or RCN (Research Collaboratory Network) funding mechanisms. These proposals could be submitted to the ABI (Advances in Biological Informatics) division. In the discussion that followed, the Committee agreed that both mechanisms could be pursued to fund pilot projects and world-wide community building, respectively; an EAGER grant proposal on exploring various integrative model representations and archival schemes was in fact already written by Helen Berman and Andrej Sali since the AC meeting, and will be submitted shortly.