RCSB Protein Data Bank Advisory Committee

Report of July 24th 2009 Annual Meeting
Toronto, Ontario, Canada

Chair: Stephen K. Burley

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

US Government Representatives: Peter McCartney (NSF, by telephone), NIH-NIGMS representative (absent), DOE representative (absent), and NIH-NLM representative (absent).

RCSB Leadership/Management: Helen Berman, Phil Bourne, Martha Quesada

RCSB PDB AC E-mail Addresses:
sburley@lilly.com, rabyrd@ncifcrf.gov, 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@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,

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

Executive Summary

The Protein Data Bank Advisory Committee (PDB AC) to the Research Collaboratory for Structural Bioinformatics (RCSB) met in Toronto, Ontario on 24th July 2009 to consider management and enhancement of the Protein Data Bank (PDB).

The agenda included
(1) State of the PDB;
(2) Outreach;
(3) Data In: Deposition, Annotation, Remediation;
(4) wwPDB Common Annotation Tool;
(5) Data Distribution and Query "Data Out"; and
(6) Responses to the RCSB PDB AC 2007 Report.

The Committee is of the opinion that the relational database form of the PDB represents the most important source of experimentally determined macromolecular structure information for research and teaching in biology, biological chemistry, and medicine. The Committee commends the RCSB and US Government agencies for their continuing 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 Committee applauds the RCSB's recent success in winning competitive renewal of funding for the US PDB, which convincingly reflects the quality of the service being delivered to the scientific and educational communities. Both the core mission of the organization and its ramifications are being pursued at the highest level. While efficiently receiving and annotating an ever-increasing number of experimentally-determined macromolecular structures, the RCSB continues to re-engineer its web site, address the challenges posed by structural genomics activities, pursue development of a common deposition/annotation tool under the umbrella of the worldwide PDB (wwPDB) organization, and increase its education, outreach, and public relations activities.

In the wake of the competitive renewal, the Committee looks forward to helping the RCSB leadership conduct a rigorous reassessment of its management/operations, with the goal of refashioning the resource to meet the needs of PDB depositors and users during the next five to ten years.

The Committee continues to view the Director's leadership of the Structural Genomics Knowledgebase as an important adjunct to her RCSB responsibilities.

State of the PDB

RCSB productivity remains impressive (2009 projected statistics are denoted with * and provided in parentheses). In aggregate, 7,073 (8,322*) depositions were processed between January 1st 2008 and December 31st 2008 with a two-week average turnaround. The breakdown of depositions by discipline was as follows: X-ray Crystallography: 6,423 (7,604*); Solution State NMR Spectroscopy: 574 (656*); Electron Microscopy: 63 (44*); and Other: 4 (10*). The breakdown of depositions processed in 2008 by RCSB versus BMRB plus PDBe plus PDBj (aggregate provided in parentheses) was as follows: All Depositions: 4,106 (2,967); X-ray Crystallography: 3,671 (2,752); Solution State NMR Spectroscopy: 390 (184); Electron Microscopy: 41 (22); and Other: 3 (1). Effective ongoing collaborations with each of the NIH-funded structural genomics centers have established efficient protocols/mechanisms for publicizing targets (TargetDB) and pipelining of interim data into the PDB (PepcDB). The RCSB web site and query software have undergone substantial re-engineering to keep pace with the demands of the ever-expanding user community. PDB-User interactions are facilitated by Electronic Help Desks, 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 a Macromolecular Machinery Poster, task forces and local advisory meetings, informal one-on-one discussions, formal interviews of PDB users, and presentations/workshops for educators (kindergartengraduate/professional). The RCSB also continues to strengthen its involvement in collaborative and consultative relationships. Most important among these is the wwPDB, a global partnership of RCSB, BMRB, PDBe, and PDBj.

2009 RCSB PDB AC Discussions

Outreach

A detailed review of current outreach activities was expertly presented by Dr. Shuchismita Dutta, Annotation Specialist, Outreach. The Committee strongly supports current RCSB efforts in education and outreach.

User Communities

The RCSB is continually challenged by the diversity of the PDB user community, which encompasses structural biologists, bioinformatics experts, software developers, biologists, educators, and students of all ages. A wide array of user support services and user outreach activities were described in some detail. The Committee views these activities as being critical to the mission of the RCSB. As anticipated, recent managerial changes at Rutgers have permitted the Director to devote more time to the important role of serving as a PDB ambassador to the manifold constituencies represented within the user community.

Education

Education at all levels, including graduate, undergraduate, and K-12, represents a critical component of the RCSB mission. The education page of the web site links to quality resources on the internet and features home-grown quality online activities for K-12 students. Educational outreach in New Jersey and in Southern California remains appropriately strong. The Committee congratulates the RCSB for its role in fostering the activities of the 2009 National Science Olympiad champions from New Jersey.

The Committee strongly endorses current plans to obtain additional funding to transform RCSB/ Rutgers teaching activities into course materials, etc. that will be broadly available to educators via the website. The Committee also commends the team for creating a PDB short course for macromolecular modelers and urges that a similar offering be provided to biologists in both academe and the pharmaceutical/biotechnology industry.

Industrial Outreach

Previously articulated RCSB plans for an Industrial Associates Program continue to elicit strong support from the Committee. Again, the importance of the RCSB Director as PDB ambassador cannot be overemphasized.

Overall Recommendation

The fundamental challenge facing the RCSB team is the need to leverage limited financial and personnel resources available for outreach by broadening the appeal and reach of the PDB and assuming a more strategic role. The Committee urges the RCSB to embrace this challenge and develop a coordinated five-year plan for outreach that balances costs with benefits, maximizes impact, and establishes productive ties with educator champions of the PDB.

Data In

A detailed review of current structure deposition/annotation systems and ongoing improvements was expertly presented by Dr. Jasmine Young, Lead Annotator. The Committee strongly supports current RCSB efforts to establish a common, global deposition/annotation tools in

collaboration with its wwPDB collaborators. Other important ongoing activities supported by the Committee include full incorporation of EM data, remediation/curation of entries with complex chemistry, and construction of a peptide reference dictionary.

Structure Validation

Considerable discussion was devoted to growing structural biology community concerns re validation of structures in the PDB. Given recent discovery of a dozen apparently fraudulent X-ray structure depositions from a single source, the Committee concurs with the RCSB/wwPDB commitment to more intensive validation of PDB entries. Domain experts on the Committee look forward to reviewing reports from the X-ray and NMR Validation Task Forces. The Committee concurs with RCSB/wwPDB plans to implement a comprehensive validation system that will generate a detailed "structure quality" report to be provided to the depositor and made public as an important component of the released deposition. In addition, the Committee recommends that the RCSB/wwPDB work with scientific journal editors to establish a uniform requirement for author submission of the PDB validation report together with the manuscript describing the structure(s).

wwPDB Common Deposition Tool

A detailed description of the wwPDB collaboration to establish common, global deposition/ annotation tools was presented by Dr. Martha Quesada, Deputy Director. The impact of recruiting Dr. Quesada and implementation of a formal project management system at the RCSB were not lost on the Committee. Both the recently completed remediation project and the ongoing common tool collaboration benefited enormously from her managerial skills and technical expertise. The Committee looks forward to receiving semi-annual reports of progress versus goals and timely completion of the project.

Data Distribution and Query "Data Out"

Dr. Phil Bourne, Associate Director, enthusiastically described recent enhancements and future plans for the RCSB website, which receives an impressive number of hits/data downloads that originate from all inhabited continents. Despite its evident popularity, there is a general perception within the community that RCSB website developments have lagged enhancements of the structural archive itself. The Committee views the evolving website as a critical component of RCSB outreach and educational activities, and urges the RCSB to include developments thereof in its five-year strategic plan for outreach/education. Specifically, the Committee recommends that this strategic plan address the look and feel and usability engineering of the website and consider costs versus benefits for inclusion of website elements that go beyond simple PDB entry display/ structure download. iGoogle may represent an informative model for an enhanced PDB user interface. The potential for building user communities with the aid of social networking tools (e.g., Facebook or Second Life) should also be explored.

In its latest remediation of the PDB archive (March 2009), annotations of biological units for ~20,000 entries have been included from the PISA resource maintained at PDBe. The Committee endorses supplementation of depositor biological unit annotations with software-generated biological unit annotations, which may be of particular value when legitimate differences of

scientific opinion arise. Inclusion of software-generated biological unit annotations offers users additional information on the potential biological relevance of oligomeric states of proteins in the PDB. The Committee recommends that the source of each biological unit annotation for each entry be clearly identified on the RCSB website, and that documentation be furnished explaining how the RCSB decides which biological unit annotations to provide from each source (author versus PISA versus PQS ...).

Commentary on Response to the 2007 RCSB PDBAC Report

The RCSB PDBAC 2007 Meeting Report included various recommendations concerning RCSB activities and enhancements (shown below in italics).

Committee Recommendations: Data In

The Committee made the following recommendations regarding these activities:

1. Bottlenecks in current data deposition/annotation processes should be rigorously identified and addressed during formulation of common, global standards and processes.
2. These newly developed structure deposition and annotation processes should be automated wherever practicable and be amenable to facile audits to ensure quality control.
3. Where these processes cannot be readily automated, they should be rendered amenable to cost-effective outsourcing wherever practicable.
4. While undertaking these activities, the RCSB should make appropriate use of expert software and information architects to strengthen the resource and ensure that it can interoperate freely with web services.
5. The RCSB should provide the Committee with a summary timeline for completion of these activities and provide brief reports of progress versus goals every six months.

Committee Recommendations: Data Integration, Distribution, and Query

The Committee made the following recommendations regarding these activities:

1. The RCSB should establish and articulate a clear and compelling vision for the future of the resource in the context of preparing for the competing renewal of the cooperative agreement.
2. The RCSB should provide the Committee with a review of this vision for the purposes of test marketing/refinement.
3. The RCSB should provide the Committee with a summary timeline for implementation of this vision and provide brief reports of progress versus goals every six months.

Committee Recommendations: Education/Outreach

The Committee made the following recommendations regarding these activities:

1. Emphasize development of kits and other educational materials that could be distributed nationally at the expense of financial/staff resources currently devoted to local undertakings.
2. Lobby the Public Broadcasting System and Nova to make a documentary on the wwPDB.
3. Broaden the scope of the Molecular Anatomy Project by recruiting faculty from other institutions to undertake distinct elements of the project using RCSB templates and editorial oversight.
4. Aggressively pursue relationships/collaborations with educational branches of appropriate scientific societies (i.e., ASBMB, the Protein Society, the ISCB, and the ACA).
5. Seek funding from other sources (e.g., the Howard Hughes Medical Institute) to supplement and expand existing activities.
6. Appoint an additional education representative to the PDBAC.
7. Thereafter, create an education task force.
8. Share educational resources/activities with EBI, BMRB, and PDBj, with the goal of

extending the already productive wwPDB collaboration to the educational arena.

Committee Recommendation: Industrial Outreach (from 2006)

RCSB plans for an Industrial Associates Program continue to garner strong support from the Committee. An action plan for initiation of this project should be provided to the Committee as soon as possible. Instead of trying to nucleate the program around one industrial scientist at one company, the Committee recommends that the program be constructed around a subject of strong mutual interest (i.e., protein-ligand complex analysis and visualization).

Following discussions with the Director, the Chair was satisfied with the RCSB responses and considers most of the issues to be well in hand. Where further guidance is required, Committee members will make themselves available to the RCSB leadership to ensure the speediest possible resolution.