

**RCSB Protein Data Bank Advisory Committee  
Report of October 2<sup>nd</sup> 2010 Annual Meeting  
Rutgers University, New Brunswick, New Jersey**

**Chair:** Stephen K. Burley

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

**US Government Representatives:** Peter McCartney (NSF representative, absent), Ward Smith (NIH-NIGMS representative), 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@fcc.edu, acpalmen@facstaff.wisc.edu, andrzej@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/Management 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 2<sup>nd</sup> October 2010 to consider management and enhancement of the Protein Data Bank (PDB).

Agenda items included

- (1) State of the PDB;
- (2) Data In: Deposition, Annotation, Remediation;
- (3) wwPDB Common Annotation Tool;
- (4) Data Out: Website and Database; and
- (5) Outreach and Impact.

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 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 core mission of the organization and its ramifications are being pursued at the highest level. While efficiently receiving and annotating a growing number of experimentally-determined macromolecular structures of ever greater complexity, the RCSB continues to transform its web site to meet the needs of the scientific community, address the challenges posed by structural genomics programs, and increase its education, outreach, and public relations activities. Under the umbrella of the worldwide PDB (wwPDB) organization, the RCSB is pursuing development of a common deposition/annotation tool and redefining validation standards for X-ray, NMR, and EM structure depositions, while exploring the possibility of archiving "hybrid" structural models of macromolecules based on SAXS/SANS and/or other data.

The Committee was similarly impressed by the newly developed strategic plan for transformation of the RCSB PDB, and believes that its execution will make important new contributions to PDB depositors and users alike during the next decade.

The Committee continues to view the Director's leadership of the NIH-funded Protein Structure Initiative Structural Biology Knowledge Base as an important adjunct to her RCSB responsibilities.

### **State of the PDB**

RCSB productivity remains impressive. In aggregate, 8,300 (8,850\*) depositions were processed between January 1<sup>st</sup> and December 31<sup>st</sup> 2009 with a two-week average turnaround (\* 2010 projection).

The breakdown of depositions by discipline in calendar 2009 was as follows:

X-ray Crystallography:	7640	(92%)
Solution State NMR Spectroscopy:	592	(~7%)
Electron Microscopy:	51	(<1%)
Other:	17	(<1%)

The breakdown of depositions processed by each wwPDB site in calendar 2009 was as follows:

RCSB:	5,069	(61%)
PDBj:	2,173	(26%)
PDBe:	1,058	(13%)

Monitoring of [www.rcsb.org](http://www.rcsb.org) showed significant year-on-year increases in website traffic for April/May 2010 *versus* April/May 2009:

Website Visits:	+14%
Page Views:	+20%
Pages/Visit:	+5%
New Visits:	+16%

Ongoing collaborations with each of the NIH-funded structural genomics centers have established efficient procedures for publicizing targets (TargetDB) and pipelining of interim data into the PDB (PepcDB). 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 a Macromolecular Machinery Poster, task forces and local advisory meetings, informal one-on-one discussions, formal interviews of PDB users, college courses, a PDBMobile application for the iPhone, 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.

### **2010 RCSB PDB AC Discussion**

#### **Data In: Deposition, Annotation, and Remediation**

A review of current structure deposition/annotation systems and ongoing improvements was presented by Dr. Jasmine Young, Lead Annotator. The Committee remains strongly supportive of ongoing wwPDB efforts to develop common, global deposition and annotation tools. The new RCSB strategic objective for automating all routine annotation and validation tasks depends critically on the success of the wwPDB

Common Deposition Tool. Once the common tool is operational, RCSB annotators will be expected to annotate increasingly complex depositions, while maintaining the highest possible levels of data uniformity. As the RCSB transitions to this new regime of data handling, the qualifications and competencies of annotators will need to be further augmented and careful monitoring of productivity metrics will be required.

Recently completed Data In improvement activities include Display of Larger Structures, New Validation Reports, handling of NMR Restraint Files and EM Maps, deployment of ADIT 2.0, and Support of New Methods. Ongoing improvement activities endorsed by the Committee include remediation of Biological Assemblies, construction of a Peptide Reference Dictionary, remediation and curation of entries with complex peptide bond chemistry, implementation of NMR Chemical Shifts, design of a new PDB Working Format, correction of Residual B-factors, and the wwPDB Validation Task Forces. The Committee concurs with RCSB plans to correct errors and improve overall data quality, while preserving all original depositions.

Considerable discussion was devoted to persistent community concerns regarding validation of structures in the PDB. The final report from the X-ray Validation Task force was well received by the Committee. Domain experts on the Committee look forward to reviewing reports from the NMR and EM Validation Task Forces, which should build on the excellent foundation provided by the X-ray team. 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 with the released deposition. In addition, the Committee reiterated its commitment to work with RCSB leaders to establish a uniform requirement that an authorized wwPDB validation report made available to structure manuscript reviewers. To this end, the Committee will undertake a letter writing campaign aimed at scientific journal editors to document broad support among structural biologists for mandatory submission of a structure validation report at the time of manuscript review.

### **wwPDB Common Deposition Tool**

A description of progress by wwPDB collaborators towards establishing common, global deposition/annotation tools was presented by Dr. Martha Quesada, Deputy Director. The Committee was pleased to learn that the project’s 2010 goals are well in hand and that substantial completion in 2011 is likely. 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 Tool are adopted across the wwPDB, the Committee recommends 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 Database**

Dr. Phil Bourne, Associate Director, presented the new RCSB strategic plan for creating “contextual views of the archive that will foster awareness of, and insight into, the structural basis of biology”. Among the topics presented thereafter were (i) new Layouts and Views of the home page, (ii) Molecule of the Month, (iii) ligands, (iv) Query and Reporting Tools for chemical components searching, query drill down and tabular reporting, (v) new Sequence and Structure Analysis and Visualization tools, (vi) bi-directional integration with other data resources using links and web widgets, and (vii) a PDBMobile application for the iPhone.

Work Performance Improvements were undertaken during the past year in response to feedback from the Committee regarding slow loading of www.rcsb.org pages. Streamlining of rendering programs and content combined with caching should result in 25-40% improvements in performance when the new

version of the website is released in the immediate future. The Committee is encouraged by these gains and urges the website engineering team to continue looking for such innovation opportunities.

Enumeration of plans for 2011 and beyond provided further explanation of the RCSB's "contextual view" concept, with emphasis on how the resource could use this approach to better meet the needs of users with widely varying skill levels. A series of pragmatic goals involving Infrastructure Changes, New Types of Data Analysis, New Query and Reporting Features, and support for MyPDB and PDBMobile were also presented.

The Committee was impressed by recent Data Out improvement activities and continues to view the evolving RCSB website as a critical component of outreach and educational activities. Both the look and feel of the website and its usability engineering should remain front of mind as RCSB 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. It will be important to consider costs *versus* benefits when planning further enhancements to elements such as PDBMobile and various expert views. When considering smart phone/iPad compatible access, the Committee urges a thorough understanding of where benefits will accrue (educational *versus* research) and at what cost to other RCSB projects. When considering the content of expert views, the Committee urges RCSB management to focus on exploiting its competitive advantage(s) and avoid efforts duplicative of established external resources.

### **Outreach and Impact**

A detailed review of current activities was presented by Ms. Christine Zardecki (Rutgers) and Dr. Andreas Prlic (UCSD). The Committee was most impressed by the growing maturity of RCSB efforts in education and outreach.

During the past ten plus years, RCSB leaders focused successively on two substantive challenges: first, on enhancing and maintaining the integrity of archival contents following their transfer from Brookhaven to Rutgers/UCSD, and second, on streamlining data deposition/annotation and distribution in the face of an enormous growth in the productivity of structural biologists worldwide. With these tasks either complete or well in hand, the RCSB has appropriately redefined its mission to focus on outreach and education.

The new RCSB strategic plan calls for creation of "contextual views of the archive that will foster awareness of, and insight into, the structural basis of biology". Presentation of information and knowledge to a complex, international user community, encompassing structural biologists, bioinformatics experts, software developers, biologists, media, educators, and students of all ages now underpins all activities of the resource.

Recent outreach and educational initiatives include (i) export of the Rutgers Molecular Views of Human anatomy courses to three other universities in 2011, (ii) working with scientific journals to coordinate PDB deposition release with publication and encouraging incorporation of a wwPDB validation report into the review process, (iii) collaborative work with Google-sponsored summer students, and (iv) expansion of current initiatives.

Metrics documenting use of the Help Desk and the [www.rcsb.org](http://www.rcsb.org) web page were presented. These data provide the first quantitative measure of the resource's impact, demonstrating its global reach, the heterogeneity of the user community, and the enormous potential of the PDB as an educational tool. Cited examples of educational impact included interest in the structural biology of influenza virus infection, motivated by news of the 2009 H1N1 pandemic, and perennial favorites such as hemoglobin and insulin. Periodic review of these metrics and other measures of resource utilization/impact will be important moving forward, particularly as RCSB leaders work to determine where best to concentrate their efforts.

In closing, three new "contextual views" of the archive were presented. These exemplars were well received by the Committee, which looks forward to their further elaboration and augmentation.

Postscript: As discussed in the 2009 Committee report, the fundamental challenge facing the RCSB team remains 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 endorses the new strategic plan focused on outreach and education, and urges inclusion of a K-12 education specialist in its number. Finally, the enormous potential for building user communities with the aid of social networking tools (e.g., Facebook or Second Life) should be explored as soon as is practicable.

### **Commentary Regarding Biological Unit Annotation**

The 2009 RCSB PDB AC report included the following recommendation:

“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 ...).”

The Committee returned to this topic at its 2010 meeting. Discussion focused on how the RCSB chooses among annotations of apparently equal validity. Communication with authors regarding how they choose the biological assembly needs to be further improved, because it has been observed that depositor designated biological assembly definitions sometimes conflict with their published descriptions of the biological unit.