

**RCSB Protein Data Bank Advisory Committee
Report of October 26, 2017 Annual Meeting
Web/teleconference**

Chair: Cynthia Wolberger

Membership: Paul Adams (absent), R. Andrew Byrd, Bridget Carragher, Wah Chiu, Kirk Clark, Paul Craig, Roland L. Dunbrack, Jr., Thomas E. Ferrin, Catherine E. Peishoff, Sue Rhee, Andrej Sali, Torsten Schwede, Jill Trewhella and Cynthia Wolberger

RCSB PDB AC E-mail Addresses:

cwolberg@jhmi.edu, PAdams@LBL.gov, bcarr@nysbc.org, byrda@mail.nih.gov, wahc@stanford.edu, kirk.clark@novartis.com, paul.craig@rit.edu, roland.dunbrack@fccc.edu, tef@cgl.ucsf.edu, peishoffc@gmail.com, srhee@carnegiescience.edu, sali@salilab.org, torsten.schwede@unibas.ch, j.trewhella@mmb.usyd.edu.au

RCSB Leadership: Stephen Burley, Helen Berman

RCSB Leadership E-mail Addresses:

sburley@proteomics.rutgers.edu, berman@rcsb.rutgers.edu

Executive Summary

The Advisory Committee (AC) to the Research Collaboratory for Structural Bioinformatics (RCSB) held a virtual meeting on 19 October 2017 to consider management and enhancement of the Protein Data Bank (PDB).

Agenda items included

- (1) Highlights, 2016 - present;
- (2) Response to the 2017 Site Visit Report;
- (3) Questions for the RCSB PDB AC

The meeting was opened by Dr. Stephen Burley, who gave an overview of the past year's activities and current state of the RCSB PDB, followed by a discussion of the June Reverse Site Visit held at the National Science Foundation (NSF) and questions for the committee. Other RCSB PDB participants were Helen M. Berman, Zukang Feng, John Westbrook, Jasmine Young, Christine Zardecki (Rutgers), Cole H. Christie, Jose Duarte and Tara Kalro (UCSD).

The RCSB PDB has continued to develop a variety of tools and web site improvements to meet the needs of both the depositor and user communities. To manage the increasing number of structures determined by integrative/hybrid (I/H) methods, the RCSB PDB and its wwPDB partners have jointly developed a prototype I/H deposition and archiving system called PDB-Dev. The GroupDep system developed by the RCSB PDB, which supports automated deposition of large numbers of structures generated by drug discovery programs, has already attracted a large number of depositions. As the number of structures determined by cryo electron microscopy (cryo-EM) and X-ray free electron lasers (XFEL) continues to grow, it will be critical that the RCSB PDB receive adequate support for curation and tool development to meet the requirements of these new methodologies. The Committee was therefore particularly

concerned to learn that there have been layoffs due to flat funding, as sufficiently skilled personnel will be essential for fulfilling the mission of the PDB.

With the coming submission of a grant application to renew funding in spring of 2018, the RCSB PDB team is focused on the important work ahead. The review panel at a reverse site visit held at the NSF in June provided a list of recommendations that need to be addressed in the formal application. While the panel clearly appreciated the importance of the PDB, both in regard to the breadth and significance of its impact, effective strategies with detailed implementation plans to achieve a sustained funding model are paramount. It will be critical to address recommendations to prioritize activities of the PDB and define those that are essential versus those that add value. The Committee strongly encourages the RCSB leadership to use the renewal as an opportunity to explain the significance of each activity and how as an integrated whole they address the needs of the research, industry and education communities. The Committee also encourages the RCSB PDB to aggressively pursue new sources of support by approaching private foundations, pharmaceutical companies and NIH institutes that utilize RCSB PDB resources but do not currently provide funding. At the same time, the Committee emphasizes the absolutely critical role that the NSF, NIH and DOE have played in supporting this premier data resource that makes it possible to make maximum use of structural information. The diverse missions of the NSF, NIH and DOE remain essential drivers to ensure that the investments in PDB tools and infrastructure can have maximal impact in medicine, biotechnology and the advancement of knowledge. Further, without free access to the structural data stored in the PDB, the scientific enterprise will be significantly less equipped to meet challenges in medical diagnosis and therapeutics, in biotechnology applications for industry and the environment and for the advancement of knowledge.

Responses to 2016 RCSB PDB AC Recommendations

The AC was provided with the following responses:

- PDBAC: The advisory committee hopes that the release of a pre-deposition server will encourage depositors to check coordinates
Response: A link to a pre-deposition server was added to the OneDep interface; Major issues made more visible during deposition; Pre-deposition server usage increased from 2500 -> 3500 runs/month
- PDBAC: Assemble comprehensive metrics that show PDB utility and impact on NSF, NIH and DOE research
Response: Metrics assembled and presented at Reverse Site Visit will be included in a recently accepted manuscript in *Protein Science*. Metrics are also included in a bibliometric analysis (doi: 10.2210/rcsb_pdb/pdb-cit-anal-2017) and economic report (doi: 10.2210/rcsb_pdb/pdb-econ-imp-2017)
- PDBAC: Feature NSF, DoE research at PDB-101 and Molecule of the Month
Response: XFEL highlighted in Photoactive Yellow Protein (March), Adenine Riboswitch (June); Globin Evolution highlighted in February
- PDBAC: Prioritize applications for additional funding
Response: Recent proposals focused on I/HM and educational efforts
- PDBAC: AC gave the preview of the HIV film, Target Zero, a unanimous thumbs up

Response: The documentary is complete and private screenings have been held in Los Angeles, CA and Piscataway, NJ; Currently on the Festival Circuit
<http://targetzerofilm.org/>

PDB Metrics

In aggregate, 11,914 depositions were received and processed between January 1 and December 31, 2016, with an average turnaround of two weeks. This represents a 6% increase from the 10,957 entries deposited in 2014. Based upon the number of entries deposited in 2017 to date, it is estimated that 13,000 entries will be deposited this year.

Breakdown of depositions by discipline in calendar 2016 was as follows:

X-ray:	10,583 (91% of entries deposited, up from 10,1067 in 2015)
NMR:	473 (4%, down from 510 in 2015)
EM:	531 (4.6%, up from 255 in 2015)
Other:	27 (0.2%, up from 25 in 2015)

Breakdown of depositions by wwPDB processing site in calendar 2016 was as follows:

RCSB PDB:	5623 (46%)
PDBj:	2240 (19%)
PDBe-EBI:	4051 (35%)

Breakdown of depositors by location in calendar 2016 was as follows:

North America	33%
Europe	34%
Asia	19%
Industry	9%
South America	1%
Australasia	3%
Africa	<1%

During 2016, RCSB PDB's website at <http://www.rcsb.org> was visited each month by an average of ~395,000 unique visitors, with more than 1 million unique users annually.

During the same period, more than 591 million data files were downloaded from the PDB archive *via* the wwPDB member FTP and websites (RCSB PDB: 454,856,822; PDBe: 74,707,114; PDBj: 62,312,151).

2017 RCSB PDB AC Discussion

Overview of highlights, 2016 - present

Dr. Stephen Burley presented an overview of RCSB PDB activities and developments over the past year:

Data In/Archive Keeping

Once again, the RCSB PDB saw an increase in depositions, with an anticipated further increase in structures determined by cryo-electron microscopy (cryo-EM) as well as by X-ray free electron lasers (XFEL). At present, the RCSB PDB is on track to process more than 6,000 depositions by the end of 2017. For the first time, more structures determined by cryo-EM were deposited than structures determined by NMR, reflecting the growing use of cryo-EM. In addition, there has been an annual increase in the number of structures determined by X-ray free electron lasers (XFEL), from five structures deposited in 2013 to 66 structures deposited in 2016. The complexity of deposited structures also continues to increase rapidly, with nearly 450 structures containing more than 500 kDa in the asymmetric unit deposited in 2016.

To handle structures that are determined by integrative/hybrid (I/H) methods, the RCSB PDB has continued to work with its wwPDB partners to develop appropriate tools. A prototype I/H deposition and archiving system called PDB-Dev was announced in the September 5, 2017 issue of the journal, *Structure* (Burley et al. (2017) *Structure* 25: 1317-1318), and is an important step in the right direction. Eight integrative structures have been deposited into PDB-Dev as of November 2017.

Structure determination using I/H methods increasingly depends upon data from a range of experimental techniques. One of the recommendations of the wwPDB I/H Methods validation task force (Sali et al., (2015) *Structure* 23:1156-67) is to establish a federation of inter-operable model/data archives. The federation concept aims at ensuring efficient and cost-effective management of I/H model archiving with the necessary complete accompanying data and meta-data. Data not currently archived by the wwPDB is to be managed by the individual specialist communities that generate the contributing data. To this end, the small-angle scattering (SAS) community has established the SAS Biological Data Base (SASBDB, <https://www.sasbdb.org/>) using a PDB-compatible dictionary of terms (sasCIF). In a collaborative project between the wwPDB and SASBDB, a protocol in the wwPDB OneDep system has been established for hybrid NMR/SAXS structure depositions, where the SAS data and meta-data are held in the SASBDB and the models in the wwPDB. This approach to I/H model archiving is cost effective for the PDB, which ensures open access to all of the contributing data, while expert communities maintain responsibility for data management and development of data validation tools.

Several changes discussed at previous AC meetings have now been implemented. The Committee was pleased to hear about the successes of the new GroupDep system, which supports automated deposition of large numbers of structures. This feature, which is aimed at facilitating deposition of structures from drug discovery programs, has already attracted more than 1,300 depositions. Other improvements described include collection of data supporting the identification of biological assemblies and ligand data and validation. Implementation of PDB archive versioning, which allows users to update and correct structures, will be rolled out in the coming year.

Data Out: Access, Exploration and Metrics

As reflected in the statistics above, the number of data downloads from the RCSB PDB continues to rise. With more than 395,000 users monthly and more than 1 million users annually, the RCSB PDB has seen a 3% annual growth in the number of non-bounce unique users. Since 2010, there has been 35% growth in the number of sessions at the RCSB.org web

site. These numbers give a valid reflection of how much the site is used, given the high average session duration of approximately 6 minutes and low proportion of 0-second “bounce” sessions.

Dr. Burley described a number of improvements to the web site. All services are now running on a private cloud, thus improving security as well as cost and operational efficiency. Other improvements to the website include enhancements to the design of the user interface and new search algorithms, both of which improve the user experience. The web site also has a new modular architecture, including new REST Web Services. The Committee was also informed that RCSB is now responsible for EPPIC (Evolutionary Protein-Protein Interface Classifier, <http://eppic-web.org>) following the passing of its founder, Guido Capitani.

Transitions, Milestones

The Committee was informed of a number of changes in the RCSB PDB team. A number of team members in Data In, Data Out and Archive Keeping have been promoted. There is currently active recruitment of a Senior Scientist and postdoctoral fellows at the UCSD site to replace two members who have left. The Committee was troubled to hear that two people at Rutgers had to be laid off due to flat funding. It is imperative that the RCSB PDB receive future funding increases, at least commensurate with inflation and increasing submissions, to continue to serve its broad community of users.

The RCSB PDB continues to be recognized for its educational impact. David Goodsell's beautiful rendering of Zika virus infecting a cell won the 2017 People's Choice award in the NSF “Vizzies” Visualization Challenge. Genetic Engineering & Biotechnology News recently highlighted the RCSB PDB educational resource, PDB-101, in its “Best of the Web” list. Maria Voigt and David Goodsell won a 2016 FASEB BioArt award for their rendering of HIV enzymes. These awards are a testament to the outstanding quality of educational offerings from the RCSB PDB.