RCSB Protein Data Bank Advisory Committee
Report of November 19, 2018 Annual Meeting
Web/teleconference

Chair: Cynthia Wolberger

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

RCSB PDB AC E-mail Addresses:
cwolberg@jhmi.edu, PDAdams@LBL.gov, pandolfa@Princeton.edu, Judith.Blake@JAX.org bcarr@nysbc.org, byrda@mail.nih.gov, wahc@stanford.edu, kirk.clark@novartis.com, paul.craig@rit.edu, Robert.Darnell@Rockefeller.edu, roland.dunbrack@fccc.edu, falko@marine.rutgers.edu, tef@cgl.ucsf.edu, peishoffc@gmail.com, srhee@carnegiescience.edu, sali@salilab.org, torsten.schwede@unibas.ch, jill.trewhella@sydney.edu.au

RCSB PDB Leadership: Stephen Burley (Director), Helen Berman (Director Emerita)

RCSB PDB 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 2018 to consider management and enhancement of the Protein Data Bank (PDB).

Agenda items included

(1) Renewal Status
(2) Highlights, 2017 - Present;
(3) Discussion

The meeting was opened by Dr. Stephen Burley. Other RCSB PDB participants were Helen M. Berman, John Westbrook, Jasmine Young, Christine Zardecki (Rutgers), and Cole H. Christie (UCSD).

The signature event in the past year was the successful renewal of the federal grant supporting the RCSB PDB, which was awarded a 10% increase in budget over previous years. While the Advisory Committee (AC) celebrated this increase, the overall budget falls short of what is needed to support the RCSB PDB's initiatives aimed at handling data generated by cutting edge approaches including cryo-electron microscopy (cryo EM) integrative/hybrid methods (I/HM) methods and X-ray free electron lasers (XFEL). The AC therefore supports the RCSB PDB leadership in its efforts to raise additional funds from pharmaceutical and biotech companies, as well as the Howard Hughes Medical Institute, to support critical projects that are not funded by the current grant. The AC recommends that the RCSB PDB develop specific project proposals that could be supported by one or more of these entities. Obtaining additional funding is

absolutely essential to enable the PDB to continue to meet the challenges of the dynamic and expanding field of structural biology.

The RCSB PDB continues to remain at the cutting edge in maintaining the growing structural database and evolving to meet new challenges. Continued increases in efficiency thanks to the OneDep deposition tool, as well as updates in data handling, curation and archive management, have enabled the RCSB PDB team to handle a greater number of depositions, many with increasing complexity. The roll-out of a redesigned web site in the coming year will further enhance usability. The new organizational plan of the World Wide PDB (wwPDB), which formally encompasses the PDB, Biological Magnetic Resonance Bank (BMRB), and Electron Microscopy Data Bank (EMDB) as core resources and other databases, such as Small Angle Scattering Biological Data Bank (SASBDB), as federated resources, presents new opportunities and challenges that will require new technical and financial support. It will therefore be essential for the RCSB to obtain additional grant funding, as well as support from industry and non-profits, in order to fulfill its mission and best serve the scientific community during this golden age of structural biology.

Responses to 2017 RCSB PDB AC Recommendations

The AC was provided with the following responses:

PDBAC: 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

Response: A series of posters and flyers were produced documenting RCSB PDB impact and support for federal funding agency goals. In addition, analyses of the impact of the RCSB PDB were published in *Protein Science* (Burley et al. *Prot. Sci.* 27:316-330), *Scientific Data* (Markosian et al. 2018 *Sci. Data* 16;5:180212). An article on the impact of the PDB "Impact on recent US FDA Drug Approvals" is now in press (Westbrook, Burley 2018 *Structure*, https://doi.org/10.1016/j.str.2018.11.007).

PDBAC: 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.

Response: Conversations have been initiated with the Howard Hughes Medical Institute (HHMI), National Cancer Institute (NCI), Science Philanthropy Alliance, and Science Gateways Community Institute. This effort is ongoing.

PDB Metrics

In aggregate, 13,049 depositions were received and processed between January 1st and December 31st, 2017, with an average turnaround of two weeks. This represents a 12%

increase from the 11614 entries deposited in 2016. Based upon the number of entries deposited in 2018 to date, it is estimated that 12,100 entries will be deposited this year.

Breakdown of depositions by discipline in calendar 2017 was as follows:

X-ray:	11889 (12% of entries deposited, up from 10583 in 2016)
NMR:	460 (3% down from 473 in 2016)
EM:	658 (29% up from 512 in 2016)
Other:	22 (27% down from 30 in 2016)
Multi-Method:	20 (%25 up from 16 in 2016)

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

RCSB PDB:	6208	(46%)
PDBj:	2797	(19%)
PDBe-EBI:	4044	(35%)

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

North America	30%
Europe	38%
Asia	21%
Industry	7%
South America	1%
Oceania	3%
Africa	<1%

During 2017, RCSB PDB's website at http://rcsb.org was visited each month by >395,000 unique visitors, with more than 1 million unique users annually.

During the same period, more than 679 million data files were downloaded from the PDB archive *via* the wwPDB member FTP and websites (RCSB PDB: 513,708,636; PDBe: 105,588,074; PDBj: 60,124,490).

2018 RCSB PDB AC Discussion

Overview of highlights, 2017 - present

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

Renewal of the RCSB PDB grant

The advisory committee (AC) celebrated the good news that the grant supporting the RCSB PDB, which is jointly supported by the National Science Foundation (NSF), National Institute of General Medical Sciences (NIGMS), the National Cancer Institute (NCI) and the Department of Energy (DOE), was successfully renewed for the period 2019-2023. The very positive comments of the review panel and favorable outcome are a testament to the superlative effort put in by the RCSB PDB team. The AC was especially gratified by the fact that there was a 10% increase in the budget, which is the first increase in over a decade. At the same time, the

funding still falls short of supporting the three challenges addressed in the renewal proposal: (1) to keep up with the continued growth of the PDB and increased complexity of the data deposited; (2) to develop capabilities to handle data from free electron x-ray lasers (XFELS) and cryo-EM; and (3) to develop tools to handle structures determined by integrative/hybrid methods (I/HM). The original execution plan, therefore had to be modified, and the revised plan approved by all funders. Given the impact of the RCSB PDB on a broad range of NIH institutes, the AC expressed the hope that future funding could come from the Office of the Director rather than just from the budgets of NIGMS and NCI.

The revised plan leaves the RCSB PDB with a need to pursue additional funding to completely support challenge 1 and to take on challenges 2 and 3. There was a discussion of possible opportunities for grants from the NSF, DOE and other NIH institutes for specific projects. In addition, the AC was informed of ongoing discussions with leaders in pharma, which have been facilitated by AC member, Catherine Peishoff. Other possibilities include instrument and software companies as well as private philanthropes.

Deposition, Biocuration and Archive Management

There was significant growth in the total number of PDB depositions, with a projected total for 2018 of ~12,100. While cryoEM depositions still represent a small proportion of deposition, this field is growing exponentially and is expected to constitute a growing proportion of structures in the future. The OneDep deposition system has continued to evolve, including a new requirement that depositors provide an ORCiD, support for SFX/XFELS entries and better software management through the use of GitHub. Thanks to the OneDep system, the number of entries handled by each biocurator has increased dramatically, and continues rise each year. The AC was apprised of other developments including initiation of a carbohydrate remediation project and enhancement of ligand validation. A welcome change that will be implemented in 2019 is the ability to replace coordinates with an updated version number, rather than a new PDB ID. There are also plans to allow chemical component versioning.

Dr. Burley described a number of software and archive management developments that will keep the database at the cutting edge. These include upgrading the Archive Management data storage system and cloud migration of the weekly database update operations. The implementation of the PDBx/mmCIF data schema across all four RCSB PDB services is a positive development that will increase efficiency.

Data Exploration

The AC was updated on developments designed to enhance the user experience and data access. A number of new features and tools have been implemented in the RCSB web site, with more planned for 2019. Improvements already implemented include Solr text search, which was first piloted on PDB101, and NGL visualization for electron density maps, ligand-protein interactions and structure quality measures from validation reports. Dr. Burley also outlined web developments planned for 2019 that will utilize APIs (application programming interfaces) to deliver data to users and facilitate batch data downloads. A full redesign of the web site is underway, with a plan to roll out the beta version in 2019. The Committee looks forward to providing feedback on the RCSB web site design as plans progress.

The RCSB PDB continues to deliver outstanding educational materials that are quite popular, as evident in the more than 620,000 users of the PDB-101 web site (http://pdb101.rcsb.org/) in 2017. The curricular materials developed by the RCSB PDB, educational videos and annual

high school Video Challenge engage students in learning about the structural underpinnings of health and disease (https://pdb101.rcsb.org/events/video-challenge/the-challenge).

Report on annual wwPDB meeting

The RCSB leadership attended the annual wwPDB meeting, which was held on November 2, 2018, in Cambridge, UK. A central topic of discussion was the new organizational structure of the wwPDB, which includes core archives and federated resources. The current core archives are the PDB and BMRB (Biological Magnetic Resonance Data Bank), with the EMDB (Electron Microscopy Data Bank) joining soon. There was discussion of the need for the RCSB PDB to receive feedback from the NMR community on how to provide information to users on structure quality and on the relative mobility of different segments. With EMDB also joining as a core resource, the committee felt it would be important to add a member to the RCSB PDB AC with expertise in cryo EM. The federated resources under the new organizational plan are SASBDB (Small Angle Scattering Biological Data Bank), MX images (raw data frames from macromolecular crystallography) and EMPIAR, the Electron Microscopy Public Image Archive for raw EM images. Drs. Jill Trewhella, Helen Berman, Andrej Sali, and Torsten Schwede will be running a workshop in conjunction with the 2019 Biophysical Society Meeting designed to bring resources together to federate with wwPDB. It was noted that the SAS community has been able to make progress in this effort because they have developed standard publication guidelines and with data and model validation tools (http://scripts.iucr.org/cgi-bin/paper?S2059798317011597) through the efforts of the wwPDB SAS validation task force (vtf) and the IUCr Commission on SAS. A subgroup of the SASvtf is now looking in detail at evaluating modelling protocols for generating scattering profiles from PDB coordinates. It was noted at the Advisory Committee meeting that this was an excellent next step for this community. Establishing community agreed standards will be important for other communities with potential interest in developing repositories and, eventually, federated resources.

Transitions

Dr. Burley announced changes in the Advisory Committee beginning in 2019. Dr. Wolberger will be stepping down after joining the AC in 2009 and serving as chair since 2013. Dr. Paul Adams, who is Division Director of Molecular Biophysics and Integrated Bioimaging at Lawrence Berkeley Laboratory, will replace her as the new chair of the AC. Dr. Mandë Holford of Hunter College and the CUNY Graduate Center has joined the AC as a new member.

As discussed in last year's meeting, the annual AC meeting will be moved to the spring beginning in 2019 in order to eliminate scheduling conflicts with the annual wwPDB meeting, which takes place in the fall.

Discussion

The major challenge facing the RCSB PDB is raising additional money to support efforts that were not funded in the recent grant renewal. Additional funding will be essential if the RCSB PDB is to meet the new challenges and opportunities presented by the growth in the areas of cryo EM, integrative structural biology, and SFX/XFEL. Given the importance of the PDB to the pharmaceutical and biotech industries, it will be essential to develop proposals that will be embraced and funded by for-profit users who benefit from the data and resources of the RCSB PDB. The Committee commends the RCSB PDB leadership for reaching out to several companies. The AC recommends that the RCSB PDB leadership develop specific project proposals that are more detailed than the three currently outlined core challenges, and to

produce materials that articulate these projects. In addition to assisting the RCSB leadership in raising funds, these proposals could help AC members seize opportunities for fundraising.