

**RCSB Protein Data Bank Advisory Committee**  
**Report of April 25, 2024 Annual Meeting (Videoconference)**

**Chair:** Paul Adams

**Membership:**

Paul Adams, Peter Andolfatto, Bridget Carragher, Wah Chiu, Kirk Clark, Robert B. Darnell, Roland Dunbrack, Paul Falkowski, Thomas Ferrin, Kevin H. Gardner, Mandë Holford, Cathy Peishoff, Seung Yon (Sue) Rhee, Torsten Schwede, Lance Stewart, Takita F. Sumter

**Present:** Paul Adams, Wah Chiu, Kirk Clark, Roland Dunbrack, Paul Falkowski, Thomas Ferrin, Cathy Peishoff, Torsten Schwede, Lance Stewart, Kevin H. Gardner

**Absent:** Peter Andolfatto, Bridget Carragher, Robert B. Darnell, Mandë Holford, Sue Rhee, Takita F. Sumter

**RCSB PDB AC E-mail Addresses:**

pdadams@lbl.gov, pa2543@columbia.edu, bcarragher@czii.org, wahc@stanford.edu, kirk.clark@novartis.com, darnelr@rockefeller.edu, roland.dunbrack@gmail.com, falko@marine.rutgers.edu, tef@cgl.ucsf.edu, kgardner@gc.cuny.edu, mholford@hunter.cuny.edu, peishoffc@gmail.com, heeseu6@msu.edu, torsten.schwede@unibas.ch, ljs5@uw.edu, sumtert@winthrop.edu

**RCSB PDB Leadership:** Stephen Burley (Director), Andrej Sali (UCSF Site Head)

**RCSB PDB Leadership E-mail Addresses:** sburley@proteomics.rutgers.edu, sali@salilab.org

**Other RCSB PDB Participants:** Jose Duarte, Henry Chao, Zukang Feng, Jasmine Young, Yana Rose, Christine Zardecki

**Executive Summary**

The Advisory Committee (AC) to the Research Collaboratory for Structural Bioinformatics (RCSB) held a virtual meeting on April 25th, 2024 to review recent progress and provide feedback on specific questions.

Agenda items included:

- Welcome
- 2023 Proposal Status and Updates
- PDBx/mmCIF Transition Update, S1-2 Roadmap Highlights
- Computed Structure Models (CSMs) at RCSB.org, S3 Roadmap Highlights
- Recruiting Updates and Team Transitions
- Responses to 2023 Report and Questions to Committee
- Additional Materials: Service 4 Highlights

The meeting was opened by Dr. Stephen Burley. Other RCSB PDB participants were Helen M. Berman, Andrej Sali (UCSF), Jose Duarte (Scientific Software Lead and UCSD Manager), Henry Chao (DevOps Lead East), Zukang Feng (Principal Scientific Software Developer), Yana Rose (S3 Lead; Scientific Software Developer & Data Architect), Jasmine Young (S1-2 Lead; RCSB

PDB Biocuration Team Lead & wwPDB Global Project Lead), Christine Zardecki (Deputy Director, S4 Lead).

The committee were provided with updates on aspects of RCSB activities over the last year, followed by some specific questions for the committee. Appendix 1 provides a summary of the RCSB responses to the 2023 Advisory Panel meeting recommendations. Appendix 2 provides a summary of global PDB deposition and data access statistics in 2023.

#### *Overall Comments from the Advisory Panel*

The team is congratulated on their outstanding work over the last 12 months, which has culminated in the successful renewal of the RCSB funding. The importance of this achievement cannot be overstated. Parallel with these renewal activities there has been significant progress in all service components. This last year has been a record for curation of structure deposits, and structures served. The committee greatly appreciated the very focused presentation materials, and the time provided for discussion.

#### *Recommendations*

Focus on delivery of the goals articulated in the renewal proposal, while investing some effort into seeking funds to support important activities that are not part of the renewal.

### **Advisory Panel Comments and Feedback**

#### ***2023 Proposal Status and Updates***

The renewal has secured funding at a level that will appropriately support core RCSB activities for the coming years. This will leave the team time to focus on seeking funds for other synergistic activities. We were pleased to hear that there are several such activities in progress including managing groups of depositions (joint NSF/BBSRC grant submitted), and a joint proposal to be submitted (NSF/DFG) to support IHM efforts with Germany. The committee feels that the team responded well to the review critique, in particular the creation of the new Service 0 (IT Infrastructure), and the appointment of Henry Chao as S0 lead. This service will strengthen the RCSB and support all activities. The pivot to using no-cost high-performance computing resources from DOE National Energy Research Scientific Computing Center (NERSC) for our calculation intensive workloads is strongly supported by the committee.

#### *Recommendations*

Seek firm assurances about NERSC availability and guaranteed level of access. Will a formal MOU of some kind be established?

The committee also heard about plans to accelerate PDB-Dev unification with the PDB archive to ensure capture and assessment of important IHM structure data, including increasing the originally proposed FTE effort in years 1 through 4. This would bring forward a number of goals from later years (3 and 5) into year 1. The committee supports the general aims of the revised timeline but has concerns about the implementation details and communications with the research community. We heard that PDBj and PDBe will contribute to the maintenance and development of the PDB-Dev system, but it is unclear how this will interoperate with the existing systems. The implementation plan and timeline are also unclear at this point.

### Recommendations

Provide a deeper dive on the timeline and proposed plan at the next AC meeting. It will also be important to make sure these activities are coordinated and well communicated to not confuse users. The committee urges the wwPDB to implement a solution that maintains a single deposition interface for users.

### ***PDBx/mmCIF Transition Update, S1-2 Roadmap Highlights***

The committee heard about the activities around extending the Chemical Component Dictionary IDs (residue IDs) to 5-characters. The team is congratulated on their efforts to make this transition, which was clearly essential as the 3-character codes have been exhausted. It should be recognized that supporting this change is significant work for the software developer community, and the wwPDB might see some delay in full support for the longer IDs. The committee also heard about plans for transition to extended PDB IDs and PDBx/mmCIF. We consider the plan and timeline appropriate and are supportive. Making sure that all relevant groups are fully invested and aware will be important. Although maybe implied, one group that was not specifically mentioned was the software developer community.

### Recommendations

Socialize the plans for transition to extended PDB IDs and PDBx/mmCIF early and often, ensuring broad community engagement.

The committee also heard about ongoing activities to improve deposition, validation and biocuration. We are very supportive and can see the value these deliver both internally and externally. Some care will be needed to ensure that updated validation software is consistent with the goals of supporting PDBx/mmCIF and communicating changes to depositors. The latter are often very concerned about changes in validation statistics and discrepancies between programs.

### Recommendations

Engagement with the software developer community to ensure that updated validation programs are as aligned as possible with the wwPDB goals of transitioning to PDBx/mmCIF. Early communications with depositors to raise awareness of possible changes in validation results.

### ***Computed Structure Models (CSMs) at RCSB.org, S3 Roadmap Highlights***

There has been a lot of activity by the team around computed models, which the committee greatly appreciates. The improvements in display of CSM information and the inclusion of links to external annotations provide users with useful information. The UXD review and community survey are helpful in learning how users are interacting with the CSMs and places for improvements. The committee supports the use of training webinar on CSMs to reach more of the user community and was pleased to hear that these would be recorded for broader dissemination. We did wonder about the CSMs being provided to users and whether there would be a mechanism to include specific focused collections, such as viral proteins or proteins outside the biomedical area (e.g. CRISPr). The committee was supportive of the 2024 goals for service 3. We wonder if there are opportunities to bring in AI/ML methods such as agents that

could greatly enhance the user experience, complementing the efforts to update the advanced search interface.

#### Recommendations

Provide a plan and timeline for improvements to the Advanced Search feature at the next AC meeting. In parallel invest some time in researching how AI/ML methods could impact search and delivery of results.

#### **Recruiting Updates and Team Transitions**

The committee heard about several staffing changes, both departures and new hires. It is encouraging to see that these are well balanced, and it is proving possible to hire staff in what is still a challenging job market. However, we recognize that hiring in the areas of IT, computing and software development is not easy, and do wonder if there might be other modes of engagement to meet RCSB needs, such as strategic partnerships with other organizations. The other major topic of discussion was the status of Director search. We were disappointed to hear that the top, and only viable, candidate had recently withdrawn their application very late in the process. This clearly leaves the RCSB and the current Director in a difficult position and will require the search to be reinvigorated after a lot of very hard work.

#### Recommendations

Leverage the advisory committee to investigate new ideas for recruiting a new Director. We will work offline to identify potential candidates and commit to meeting with Stephen to strategize about the recruitment. We also suggest working with the committee to develop a communication that can be disseminated to both the academic and industrial communities to surface more potential candidates.

#### **Responses to 2023 Report and Questions to Committee**

The committee was pleased with the responses and recognizes that many of these have been actioned already.

#### **Additional Materials: Service 4 Highlights**

Although this material wasn't presented at the meeting, the committee did have a chance to consider two specific topics: future training events in 2025, and future Molecule of the Month articles.

#### Recommendations

For future training events, consider pollution/remediation (e.g., microplastics, PFAS chemicals), alternative energy and environmental cleanup. These are all outside of the more typical biomedical focused topics. They also relate directly to the topics of climate change and environmental resilience, which are grand challenge problems facing the world. Future Molecule of the Month articles could also follow this lead.

#### **Responses to Other Specific Questions to the Committee**

The feedback above addresses many of the specific questions posed to the committee. We have the following feedback on the remaining questions:

***Are there resources we should be developing to help the fight against climate change?***

This is an enormous topic and there are probably many routes to making links with PDB models. Specific suggestions were: enzyme structures involved in carbon management with an emphasis on photosynthesis, carbon utilization and processing (RuBisCO, carbonic anhydrase), carbon sequestration (e.g. through biomineralization), and biofuels (lignocellulosic processing and bioproducts synthesis).

***Other concerns raised***

The committee had some discussions about cryo-EM, in particular validation of single particle studies, and the longer-term plans for supporting cryo-electron tomography (cryo-ET) studies. There is a concern that the validation tools for cryo-EM depositions are lagging and need to be implemented on a fast timeline.

*Recommendations*

The topic of cryo-EM validation tools provided by the wwPDB for depositions should be discussed at an upcoming wwPDB AC meeting.

There is uncertainty about the roadmap for supporting cryo-ET studies at the RCSB and wwPDB. This is becoming pressing as the method is rapidly developing and gaining wider use. It presents challenges in terms of data volumes and the potential for very large atomic models. There is also likely to be an intersection with IHM approaches.

*Recommendations*

At a future AC meeting, the RCSB team presents their plans for supporting cryo-ET depositions.

## Appendix 1: Response to 2023 RCSB PDB AC Report

Summer 2023

The RCSB PDB Advisory Committee Meeting was held April 12, 2023. The advisors provided feedback via their report ([PDF](#)). Responses to advisor suggestions (non-Site Visit-related) are addressed below.

### *AC Recommendations*

*Prepare a concise and compelling presentation for the reverse site visit, with backup slides on topics and questions that might arise during the review. The committee is very willing to review materials prior to the review if that would be helpful.*

Response: Advisor feedback provided during the AC meeting and this report were extremely helpful in our preparations and presentations.

### *AC Recommendations*

*The RCSB is encouraged to increase the adoption of community-based feedback where it will help refine services and the RCSB user experience.*

Response: RCSB PDB will continue to collect feedback at meetings, through our customer service help desk, user surveys, and UXD reviews. In 2024, we will host virtual “office hours” and assess feedback collected.

### *AC Recommendations*

*The committee suggests engaging with the User Experience Design group to get feedback on what would be reasonable approaches to providing CSMs to the community. The committee also recommends more user engagement to inform about CSMs, how they are made available and their strengths and weaknesses.*

Response: CSMs at RCSB.org will be reviewed as part of the Spring 2024 UXD Review. Additionally, we plan to host a Training Event focused on CSMs and to publish supporting materials.

### *AC Recommendations*

*The committee recommends creating a series of web-based short video guides to inform users about new features, which can then be a lasting source of information for anyone new to the RCSB. A YouTube channel provides an easy mechanism for users to find content.*

Response: RCSB PDB plans to explore options and best practices for video guides in 2024, with a goal of publishing videos in 2025 about RCSB.org.

RCSB PDB will continue to collaborate with our wwPDB partners on depositor-focused videos published at wwPDB.org and YouTube.

### *AC Recommendations*

*In addition, it is probably timely to think about a redesign of the RCSB web site, which has had the same look and feel for several years and could benefit from adopting current best practices for website design, including links to the short video guides on feature use.*

Response: RCSB PDB plans to start this process by improving the home page at RCSB.org as well as the Advanced Search interface later in 2024.

*AC Recommendations*

*The RCSB crash courses are an effective way to reach the user community and a place to continue to invest effort if resources permit. The committee felt this was an opportunity to inform and train users on a variety of topics, including advanced searches, programmatic interfaces to access RCSB resources, computed structure models, and graphical tools such as Mol\*.*

Response: RCSB PDB has begun to focus on offering more crash courses and webinars, which in 2023 included a focus on PDBx/mmCIF; another focused on APIs; and another on using RCSB.org resources that was hosted by ASBMB. 2024 training activities will similarly include crash courses/webinars that target data depositors and data consumers.

## Appendix 2: PDB 2023 Metrics

In aggregate, 17063 depositions were received and processed during 2023, an increase from the 16,344 entries deposited in 2022.

Breakdown of depositions by discipline was as follows:

X-ray: 10197 (60% of entries deposited, down from 10,624 in 2022)

NMR: 332 (1.9%, up from 287)

EM: 6507 (38%, up from 5407)

Other: 27 (<1%, up from 26)

Breakdown of depositions by wwPDB processing site was as follows:

RCSB PDB: 6697 (39%)

PDBj: 4104 (24%)

PDBe-EBI: 4990 (29%)

PDBc: 1272 (7%)

Breakdown of depositors by location was as follows:

North America 4952 (29%)

Europe 6272 (36.7%)

Asia 5378 (31.5%)

South America 91 (<1%)

Oceania 358 (2%)

Africa 12 (<1%)

For 2023, RCSB PDB Usage Analytics recorded RCSB.org access by ~8.2 million unique IP addresses, making ~3.5 billion requests/interactions (*i.e.*, >400 data downloads, service usage, web page content views, etc. per unique IP address). During the same period, Google Analytics estimated that RCSB.org hosted ~5.2 million unique users (~20% from USA) viewing ~63 million web pages.

In 2023, 2.6 billion data files in various file formats, including structure files, experimental data files, chemical and molecular reference data files, FASTA sequence files, and validation reports, were downloaded and/or viewed from RCSB PDB-hosted FTP and websites.

Additional data were downloaded from wwPDB partners PDBe and PDBj, for a cumulative total of more than 3 billion data file downloads in 2023.