

RCSB Protein Data Bank Advisory Committee
Report of July 29, 2025 Annual Meeting (Videoconference)

Chair: Paul Adams

Membership:

Paul Adams, Mohammed AlQuraishi, 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, Mohammed AlQuraishi, Wah Chiu, Kirk Clark, Roland Dunbrack, Paul Falkowski, Kevin H. Gardner, Cathy Peishoff, Sue Rhee, Torsten Schwede, Takita F. Sumter

Absent: Peter Andolfatto, Bridget Carragher, Robert B. Darnell, Thomas Ferrin, Mandë Holford, Lance Stewart

RCSB PDB Advisory Committee E-mail Addresses:

pdadams@lbl.gov, ma4129@cumc.columbia.edu, 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@fas.harvard.edu, peishoffc@gmail.com, rheeseu6@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, Jasmine Young, Dave Raymond, Yana Rose, Brinda Vallat, Christine Zardecki, Helen Berman

Executive Summary

The Advisory Committee (AC) to the Research Collaboratory for Structural Bioinformatics (RCSB) held a virtual meeting on July 29, 2025 to review recent progress and provide feedback on specific questions.

This annual review focused on operational performance, progress toward strategic goals, and responses to recommendations from the 2024 AC report.

Agenda items included:

- Brief Welcome; Status of Core Operations and Funding; Other Updates
- PDB-IHM Unification and Updates to RCSB.org
- S0 Update: IT Infrastructure and NERSC
- PDBx/mmCIF Transition Update, S1 (Deposition/Biocuration) and S2 (Archive Management and Access) Roadmap Highlights
- S3 (Data Exploration) Highlights
- S4 (Training, Outreach, Education) Highlights

- Responses to 2024 Report, Questions for Committee

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), David Raymond (S0 Lead; IT Infrastructure), Yana Rose (S3 Lead; Scientific Software Developer & Data Architect), Brinda Vallat (PDB-IHM), Jasmine Young (S1-2 Lead; RCSB PDB Biocuration Team Lead & wwPDB Global Project Lead), Christine Zardecki (Associate 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 2024 Advisory Panel meeting recommendations. Appendix 2 provides a summary of global PDB deposition and data access statistics in 2024.

Overall Comments from the Advisory Panel

The team is again congratulated on their outstanding work over the last 12 months, with significant progress in all service components. This past year set a record for both structure deposit curation and structures served. The committee greatly appreciated the focused presentation material and the time provided for discussion. The primary concern of the committee is the slow progress on recruiting a new RCSB Director.

Recommendations

Continue the excellent work delivering the RCSB to the research community.

Intensify efforts for recruiting the next RCSB Director.

Detailed Advisory Panel Comments and Feedback

Status of Core Operations and Funding

The recent renewal of the core RCSB funding has enabled the team 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 recently awarded). The current uncertainty in Federal funding is recognized as a challenge, but the RCSB seems well placed to navigate this given their central role in enabling research in both academia and industry. Although there has been some staff turnover within the RCSB, the team has done well in recruiting new people. The committee was particularly impressed with the early career additions to the team, and noted that this brings an obvious energy, which was reflected in several of the presentations.

Recommendations

Continue to seek funding opportunities to go beyond the core operations, especially in collaboration with others.

The committee also received an update on the status of the search for a new RCSB Director. It was encouraging to hear that efforts continue on the recruitment including the notion of using a search firm, but of great concern to the committee and the current RCSB Director that there has been little forward motion. It is acknowledged that funding uncertainties contribute to the situation, but the committee feels this shouldn't be a reason for stalling progress.

Recommendations

Push Rutgers to engage the search firm now, irrespective of the federal funding situation (much of which has recently been positively clarified for the RCSB).

Develop a succession plan that leverages the internal strengths of the RCSB so that the current Director can move forward with his own plans.

Strategize with Rutgers leadership about how to best formulate the recruitment, given the Federal focus on AI, and the goal of the new Rutgers President to expand research more broadly.

Given the effort needed, we recommend that strong consideration be given to appointing a new lead for the search, freeing up valuable time for the current Director.

PDB-IHM Unification and Updates to RCSB.org

The committee heard about the work that has gone into integrating the integrative hybrid models (IHM) into the main PDB Archive and across the RCSB activities. This has included making these models available through searches at the main RCSB.org web site. The committee supports this integration, and the high-level plans for the next phases that embrace streamlined deposition, validation and further improvements in the IHM RCSB.org experience for users. There are some concerns about the current lack of a community consensus on the requirement for depositing IHM structures. The committee heard that several journals are aware of the ability of the PDB to accept IHMs, but there has been no push to make this mandatory for publication (as is the case with crystallographic and cryo-EM structures). In addition, the committee heard about the plans to develop a unified Bayesian data and model validation approach. This initiative is promising; however, the committee sought clarification regarding its funding support and prioritization.

Recommendations

Continue to work with the research community and the scientific journals to arrive at a consensus for IHM deposition (models and data) that will promote growth of the IHM database at RCSB.

At the next committee meeting, discuss the funding and priority of the Bayesian data and model validation effort, and whether this is something that would benefit from a concerted effort to obtain additional funding.

S0 Update: IT Infrastructure and NERSC

There has been strong progress in service 0 over the last year, including an effort to transition to containers with Kubernetes. The committee supports this and recognizes that it provides multiple benefits. The team has also worked to leverage the DOE National Energy Research Scientific Computing Center to perform some of the weekly update calculations, reducing the time from days to hours. The committee recognizes the power of this collaboration. While the availability of NERSC time isn't guaranteed, the ERCAP process is flexible and lightweight enough that this arrangement seems sustainable.

Recommendations

Communicate the needs for continued NERSC time to the BER program manager, especially in light of a change in how ERCAP requests are being reviewed by BER.

PDBx/mmCIF Transition Update, S1 (Deposition/Biocuration) and S2 (Archive Management and Access) Roadmap Highlights

There has been great progress in services 1 and 2 over the last year, including the adoption of extended PDB IDs, improvements in the validation software pipeline, and further automation of annotation. These, and other improvements, have continued to increase the number of depositions that are processed by each annotator each year. This has doubled in the past 10 years since OneDep was launched. The committee congratulates the OneDep team for their continued efforts and success, and also Jasmine for her leadership. The plans for supporting the growing area of cryo-EM/ET are good, and the committee was pleased to see progress already in improving the deposition process. The challenges in validating cryo-EM/ET models and data are recognized, and the goal of incorporating CaBlam is supported by the committee.

Recommendations

With the broader wwPDB, develop plans for more community engagement to identify new validation metrics for adoption in the short to medium term.

The committee appreciates that extending the IHM data types presents some challenges in terms of scalability, need for increased annotator expertise, and the need to create new mmCIF dictionary definitions.

Recommendations

Engage with the research community to enlist their help in defining the standards for new experimental types, with an eye to identifying those of highest priority. It is also good to bring users of the IHM models into that discussion.

S3 (Data Exploration) Highlights

Service 3 has implemented several new features in the last year that enhance the user experience. Highlights include the ability to view validation information in the context of a 3D structure, the improved support for SX/XFEL data, and a new statistics page. The committee judged the plans for the coming months to be well chosen and impactful for the users of RCSB.org. The plans for expanding IHM tools and use of CSMs would both benefit from community engagement, including professional society meetings. The committee was very pleased to see that AI is being integrated into the RCSB.org site, and await the implementation of the improved advanced search user interface.

Recommendations

Undertake outreach to the user community to gather input on use of IHMs and CSMs, to help develop prioritized plans for improving the RCSB.org site further. This should include user input on which CSMs from the model archive are added/available through the PDB.

S4 (Training, Outreach, Education) Highlights

Service 4 is congratulated on successfully transitioning its molecule of the month to Janet Iwasa, and adding several new resources to the PDB-101 site. The training webinars have been extremely successful and well attended, and the plans for the coming months are very strong. The committee had some suggestions for Molecules of the Month and other resources: a retrospective on COVID, and materials in the context of biotechnology - degradable, high strength, soft materials.

Recommendations

Consider how to identify topics that highlight translation of basic research to applied impact, as this might be a priority for the current administration.

Responses to 2024 Report

The committee thanks the team for their responses to last year's report.

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:

All Services: Suggestions for exploring other funding sustainability options?

The committee feel that AI is a topic that can provide a rallying point across the RCSB. It already is having a positive impact across the services, and this can be extended. It has the potential to help greatly with deposition and annotation tasks. It is also an area for algorithm development by the team, which can help attract new talent. Finally, it can also be a central part of the recruitment of the Director, and help position the RCSB within the broader Rutgers community.

Appendix 1: Response to 2024 RCSB PDB AC Report

December 2024

The RCSB PDB Advisory Committee Meeting was held April 25, 2024. The advisors provided feedback via their report ([PDF](#)). Responses to advisor suggestions are described below.

AC Recommendations Overall

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.

Response: We are focusing on the projects described in the renewal proposal. We are additionally exploring additional funding mechanisms for additional synergistic activities.

2023 Proposal Status and Updates

AC Recommendations:

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.

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

Response: As of Fall 2024, RCSB PDB has applied for NERSC access via standard requests going through the official NERSC allocation mechanism. Annual requests to date have been granted in full. (N.B.: We have received repeated indications that they view our needs as being modest compared to other NERSC users.) Technical support provided by the NERSC IT teams during our initial exploratory project phase has been very responsive so far, and multiple members of their team have expressed willingness to accept feedback and suggestions for changes to better suit our project needs.

We will initiate discussions with Amy Swain (RCSB PDB Program Officer at DOE) and Ramana Madupu (overseer of biology access to NERSC) regarding the possibility of a longer term commitment in the form of an MOU.

As we have worked more with NERSC, we have learned that there will be occasional times when NERSC will be unavailable. Therefore, in order to be able meet our goals of both on-time weekly data delivery and using other resources to perform needed compute-intensive activities of the PDB weekly update, we are looking to establish a parallel update system that enables facile access to alternate providers, e.g., NSF ACCESS.

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.

AC 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.

Response: A detailed discussion will be held at the next AC meeting. Activities are being developed in collaboration with the wwPDB partners and announced to the public.

At the time of writing, the wwPDB has announced that the [PDB Archive Serves Structures Determined by Integrative and Hybrid Methods \(IHM\)](#) and that users can [Access IHM structures at wwPDB DOI landing pages](#). PDB-Dev has been rebranded as PDB-IHM.

PDBx/mmCIF Transition Update, S1-2 Roadmap Highlights

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.

AC Recommendations

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

Response: RCSB PDB is working to engage the community at scientific society meetings (e.g., [ACA](#)), with supporting documentation (e.g., [PDBx/mmCIF User Guide](#)), and office hours and webinars. In addition, wwPDB will continue communicating the plan and progress with the broader community including community software developers via community mailing lists (e.g., CCP4, CCPn, CCP-EM, 3DEM, BMRB-I, pdb-I).

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.

AC 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.

Response: wwPDB continues working with community software developers to remove PDB file format dependency on 3rd party software in the wwPDB validation software package. Recent updates to remove PDB format dependency include replacing legacy code (e.g., Cyrange) in NMR validation in consultation with NMR community (doi: [10.1016/j.str.2024.02.011](https://doi.org/10.1016/j.str.2024.02.011)) and the EDS upgrade.

wwPDB plans to update the wwPDB validation package in 2025 to allow MolProbity to use PDBx/mmCIF format files. BMRB will invest effort into refactoring NMRClust to remove legacy PDB format dependency.

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

The committee supports the use of training webinars 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.

AC 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.*

A comprehensive plan and timeline for enhancing the Advanced Search UI will be presented at the next AC meeting. This plan will incorporate feedback from various sources, including User Experience Design (UXD), user input gathered through the customer service Help Desk, and insights from internal brainstorming sessions. This plan will outline the specific improvements, milestones, and expected delivery dates.

In parallel, we are conducting research on how AI techniques can be applied to RCSB PDB's scientific searches (currently: structure, sequence, sequence motif, structure motif). In particular, we are developing 3D structure search functionality,

which has the greatest potential to benefit from AI-driven methods. We hope that this research will eventually open the way to wider application of AI-methods to our other scientific search challenges. We intend to make the new structure search system available for users sometime in 2025.

Additionally, we will start to explore readily-available language models customized to our data for internal use to assess their usefulness in aiding everyday tasks such as helpdesk correspondence. This would inform possible later applications for the wider public.

Recruiting Updates and Team Transitions

AC 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.

Response: We thank the committee for their post-AC meeting discussions and feedback. As a result, the [job description has been updated](#) and additional candidates have been contacted.

Responses to 2023 Report and Questions to Committee

Additional Materials: Service 4 Highlights

AC 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.

Response: We thank the advisors for these suggestions.

We will continue to expand PDB-101 content in the [Biotechnology category](#).

Our Q1 2025 Training Events will include webinars on *Seeing Bird Flu in 3D* and *Streamlining API Access with Python*; and a virtual office hour on *Extended PDB IDs*.

cryo-ET studies

AC Recommendations

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

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

Response: Initiatives for cryo-EM validation and cryo-ET deposition are part of the wwPDB collaboration and will be discussed at that meeting in October 2024.

The 2025 wwPDB roadmap includes enhancing cryo-EM validation with a slider for model-map fitness in 2025 and creating longer-term plans to support the growth of cryo-ET.

A summary will be shared at the 2025 AC meeting.

Appendix 2: PDB 2024 Metrics

In aggregate, 19,322 depositions were received and processed during 2024, an increase from the 17,063 entries deposited in 2023.

Breakdown of depositions by discipline was as follows:

X-ray: 10865 (up from 10197 in 2023, 56% of entries deposited)

NMR: 311 (down from 332, 1.6%)

EM: 8127 (up from 6507, 42%)

Other: 19 (down from 27, <1%)

Breakdown of depositions by wwPDB processing site was as follows:

RCSB PDB: 7651 (39%)

PDBj: 3535 (18%)

PDBe-EBI: 5092 (26%)

PDBc: 3044 (15%)

Breakdown of depositors by location was as follows:

North America 6175 (32%)

Europe 6029 (31.2%)

Asia 6657 (34.4%)

South America 82 (<1%)

Oceania 369 (2%)

Africa 10 (<1%)

For 2024, RCSB PDB Usage Analytics recorded RCSB.org access by ~14.5 million unique IP addresses, making ~5.0 billion requests/interactions (*i.e.*, >300 data downloads, service usage, web page content views, etc. on average per unique IP address). During the same period, GA estimated that RCSB.org hosted ~8.7 million unique users (~13.5% from USA) viewing ~64 million web pages (>7 per estimated unique user).

In 2024, ~3 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 archives and websites.

Additional data were downloaded from wwPDB partners PDBe and PDBj, for a cumulative total of more than 3.6 billion data file downloads in 2024.