

Worldwide Protein Data Bank Advisory Committee (wwPDB-AC) Report of the October 14th 2022 Meeting

Advisory Committee Chair:

Jennifer Martin (IUCr Representative) jlm@griffith.edu.au

Advisory Committee Co-Chair:

Paul Adams (RCSB-PDB) PDAdams@lbl.gov

Advisory Committee Members

PDB Site Representatives Nominated by wwPDB partners:

Kirk L. Clark (RCSB-PDB) kirk.clark@novartis.com

Linda Columbus (BMRB) lc4zs@virginia.edu

David Hsiao (PDBj) hsiao@gate.sinica.edu.tw

Juha Huiskonen (EMDB) juha.huiskonen@helsinki.fi

Martin Noble (PDBe) martin.noble@ncl.ac.uk

Toshiya Senda (PDBj) toshiya.senda@kek.jp

Corinne Smith (EMDB) corinne.smith@warwick.ac.uk

Martin Walsh (PDBe) martin.walsh@diamond.ac.uk

Apology: Angela Gronenborn (BMRB) amg100@pitt.edu

ISMAR Representative:

Tatyana Polenova tpolenov@udel.edu

3DEM Representative:

Alexis Rohou rohousa@gene.com

Institutional Representatives:

Denise Hien (Rutgers University) denise.hien@smithers.rutgers.edu

Gerard Kleywegt (EMBL-EBI) gerard@ebi.ac.uk

wwPDB Principal Investigators:

Stephen K. Burley (RCSB-PDB) stephen.burley@rcsb.org

Sameer Velankar (PDBe) sameer@ebi.ac.uk

Jeffrey C. Hoch (BMRB) hoch@uchc.edu

Genji Kurisu (PDBj) gkurisu@proein.osaka-u.ac

Apology: Ardan Patwardhan (EMDB) ardan@ebi.ac.uk

wwPDB Regional Representatives (Associate member and candidate):

Debasisa Mohanty (India) deb@nii.ac.in

Wenqing Xu (China) xu.wq@zjlab.org.cn

wwPDB-AC Meeting October 14 2022:

The Worldwide Protein Data Bank Advisory Committee (wwPDB-AC) and the leadership of the Research Collaboratory for Structural Bioinformatics (RCSB-PDB), the Biological Magnetic Resonance Data Bank (BMRB), the Protein Data Bank in Europe (PDBe), the Electron Microscopy Data Bank (EMDB), and the Protein Data Bank Japan (PDBj) met together in a 3-hour online meeting organized and hosted by RCSB.

wwPDB Vision Statement

Sustain freely accessible, interoperating Core Archives of structure data and metadata for biological macromolecules as an enduring public good to promote basic and applied research and education across the sciences.

wwPDB Mission Statement

- Manage the wwPDB Core Archives as a public good according to the FAIR Principles.
- Provide expert deposition, validation, biocuration and remediation services at no charge to Data Depositors worldwide.
- Ensure universal open access to public domain structural biology data with no limitations on usage.
- Develop and promote community-endorsed data standards for archiving and exchange of global structural biology data

Table of Contents

1. Agenda
2. Introduction
3. Funding
4. Governance
5. Outreach
6. OneDep
7. PDB Archive
8. EMDB Archive
9. BMRB Archive
10. Joint Projects
11. Responses to wwPDB PI questions posed to wwPDB-AC
12. Conclusion

1. Agenda

The planned agenda mirrored that of the 2021 meeting, though with longer timings for sessions 4 and 5, and with a total meeting time of 3 hours compared with 2.5 hours at the 2021 meeting. In practice, the 2022 meeting had one long Executive session, followed by a longer combined Discussion/Q&A/Feedback session of ~95 min. The papers had been provided on-line in advance to the wwPDB-AC, with on-line questions and feedback from individual members of the wwPDB-AC collated. The pre-meeting questions and feedback have been incorporated into this report.

Planned:

1	8:30-8.40 (US EDT)	Welcome and Introductions (10')	Denise Hien/Stephen K. Burley
2	8:40-9:00	Executive Session No. 1 (20')	AC only Breakout room via host
3	9:00-10:00	Discussion/Questions for AC (60')	All
4	10:00-10:45	Executive session No 2 (45') (30' scheduled in 2021)	AC only Breakout room via host
5	10:45-11:15	Feedback to wwPDB Leadership (30') (20' scheduled in 2021)	wwPDB PIs
6	11:15-11.25	Acknowledgements (10')	Stephen K. Burley
7	11:25-11.30	Group Photo/Meeting Close (5')	All, Photo <i>via</i> Host

Actual:

1	8:30-8.40 (US EDT)	Welcome and Introductions (10')	Denise Hien/Stephen K. Burley
2	8:40-9:45	Executive Session No. 1 (65')	AC only Breakout room via host
3	9:45-11:20	Discussion/Questions/Feedback to wwPDB Leadership (95')	wwPDB PIs and AC
4	11:20-11.25	Acknowledgements (5')	Stephen K. Burley
5	11:25-11.30	Group Photo/Meeting Close (5')	All, Photo <i>via</i> Host

Information provided by wwPDB PIs prior to the wwPDB-AC meeting is shown in black text. *Notes and responses from the wwPDB-AC are provided in blue italic text.*

2. Introduction

Members of the wwPDB-AC were welcomed online by Denise Hien, (Vice Provost for Research, Rutgers, The State University of New Jersey) and by Stephen Burley, the host wwPDB PI for this meeting. Stephen Burley then introduced the wwPDB PIs and the advisory committee members, before handing the meeting over to the wwPDB-AC chair, Jenny Martin.

3. Funding

3.1 wwPDB Core Member Funding Status

- RCSB PDB: Joint NSF/NIH/DOE funding (2019-2023) *no change*
- BMRB: NIH NIGMS funding (2019-2023)
 - Inadequate budget: still need to find additional support
 - UConn funding 25% administrative assistant and 25% project manager positions
 - NIH U24 submitted but not funded. *Re-submitted 10/2022 for 2nd time.*
- PDBe: EMBL-EBI, Wellcome Trust (2021-2025) *no change*
- PDBj: NBDC-JST and AMED funding (2022-2027) *new funding*
 - Additional budget from S. Korea (decision pending due to COVID-19)
- EMDB: EMBL-EBI, Wellcome Trust (2019-2024) *no change*
- RCSB PDB/PDBe: Joint NSF/BBCRC NextGen Archive funding (2020-2022) *no change*
- RCSB PDB/PDBe: Joint NSF/BBCRC Mol* Visualization funding (2021-2024) *no change*

3.2 wwPDB Associate Member Funding

Basic support of PDB China (PDBc) (basic PDBc data-in and data-out operations) can be provided by guaranteed funds for the National Facility for Protein Science (NFPS). "Data-base and Computation" for protein science, with approved support for 9 FTEs (personnel), is one of these nine technical systems in NFPS.

For the current funding period (years 2020-2022), this basic yearly grant is CNY 57,130,000 (~USD \$9,021,000/year).

On top of this basic grant, PDBc also receive 0-15% merit-based bonus operation fund every year, and have supplementary salary support for personnel from the Chinese Academy of Sciences (CAS).

The wwPDB-AC congratulates wwPDB partners in securing current funding, including the new funding for PDBj. We reiterate 2021 wwPDB-AC comments that: the entire activity of wwPDB is underfunded in comparison to community need and wwPDB potential contributions; the BMRB situation is concerning (see response to Q6 below); challenges posed by the growth in and changes to structural biology may represent funding opportunities; collaboration between partners is an asset when seeking funding; and we encourage the wwPDB PIs to engage with the wwPDB-AC for help in securing funding.

4. Governance

PDB charter updated. PDB China admitted as first Associate Member. Preparing to admit PDB India as Associate wwPDB Member.

PDBc Update

PI/co-PI: Profs. Wenqing Xu, James Z. Liu, Quan Wang
National Facility for Protein Science in Shanghai (NFPS) and iHuman Institute and SIAIS, Shanghai Tech University, Pudong, Shanghai, China

- The NFPS in Shanghai has recruited/formed the Data-in team (5 people) and the joint Data-out team (5 people, in collaboration with the ShanghaiTech University).
- The Advisory Committee for PDBc has been formed, with balanced scientific expertise, gender and regional distribution. The PDBc AC member list has been reviewed by wwPDB-AC.

- The NFPS, the primary host of PDBc, has upgraded the cooling and security systems for the PDBc computation-facility; and with support of the Shanghai municipal government, has purchased computational hardware dedicated to PDB China (with >\$1.2m USD) and performed related infrastructure set up.
- PDBj obtained funding to purchase the computers for PDBc's first OneDep data-in system (to be sited in Osaka); these equipment are used for on-site training.
- The first version of the PDBc data-out system, with primary search functions, has been developed, and is undergoing testing and further development.
- Remote data-in training for PDBc biocurators started in March 2022. Biocurators from RCSB, PDBe and EMDB have given talks on data format, wwPDB policies and EM annotation. On-site data-in training for two PDBc biocurators started in late August at PDBj.

PDBi Update

PI: Prof. Manju Bansal, Prof. K. Sekar, Dr. Debasisa Mohanty Co-investigators Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India

- The first phase funding (Oct 2019-2022) received from NSM (India) was used for hiring PDB India (PDBi) staff, currently working at IISc, Bangalore & NII, New Delhi. NSM (India) has extended funding for salaries of PDBi staff till March 2024. Efforts are ongoing to secure funding from DBT after 2024 for 5 years.
- PDBi PIs have made arrangements to utilize servers/storage of the Indian Biological Data Center (IBDC), a facility funded by DBT, India. IBDC has recently (Sept 2022) provided access to PDBi PIs for utilizing IBDC servers/storage for PDBi activity.
- Six PDBi staff are currently working on development of structural bioinformatics software for comparative analysis of Protein-DNA complexes in PDB, structure-based analysis of protein-protein interaction network of M.Tb./Plasmodium. A prototype data-out server with very basic search interface has been set up by PDBi staff.
- PDBe (Dr. Velankar) has provided online remote training of four PDBi staff on annotation/curation of structural data. Three of them will visit EBI in person soon for further training on data curation (Funded by PDBe).
- National Advisory Committee for PDBi will be constituted in consultation with NSM & DBT.

The wwPDB-AC congratulates the wwPDB PIs and their teams on the progress made over the past 12 months on both these Associate memberships.

The wwPDB-AC is pleased to note that PDBc has been admitted as an Associate Member. This development should alleviate pressure on biocurators at other sites (especially PDBj) who curate a large proportion of current China depositions. See also response to Q3 below.

Regarding the PDBi National Advisory Committee, we remind the wwPDB PIs that committee membership should be as diverse as possible, with a good gender balance and consideration of different levels of appointments (junior and senior).

We reiterate and expand on some comments from the 2021 report:

– wwPDB-AC would like to be kept informed of timelines and milestones for inclusion of potential new wwPDB associate members. Of particular importance are issues of load balancing, quality control, data security.

– wwPDB-AC requests that AC reports of the individual wwPDB associate members be made available.

5. Outreach

- BMRB Poster and Booth/Biophysical Society: February 19-23
- BMRB Workshop, Poster, and Booth/ICMRBS: August 21-25

- RCSB PDB-PDBe Poster/ISMB: July 10-14
- RCSB PDB Talk/ISMB: July 10-14
- RCSB PDB Poster, Talks, and Booth/ACA*: July 29-August 2
- RCSB PDB Workshop and Talks/ASCB: December 6
- EMDB Talk and Posters/CCPEM Symposium: May 3-5
- EMDB Posters/EMBL Imaging Centre Symposium: May 31
- EMDB Talk/South-West Electron Microscopy Plymouth: July 18-19
- EMDB Talk/CCPEM Ickniel workshop: September 9
- EMDB Poster/Tomography Congress 2022: September 11-14
- PDBe Posters and Booth/ECM: August 23-27
- PDBj Talk/AsCA: October 30-November 2
- * In-Person/Virtual Hybrid Event

The wwPDB-AC continues to be impressed by outreach activities of wwPDB partners. These activities are important to raise awareness and engage the community about this critical global organisation. Our congratulations on continuing the great outcomes from the PDB50 celebrations.

6. OneDep

Ad hoc projects:

- Supported EBI server upgrade
- Improved support for extended PDBx/mmCIF structure factor files
- Made the types of validation reports (For Manuscript Review) more prominent
- Added map-model fitness using Q-score in the 3DEM validation reports
- PDB India training

wwPDB biocurator productivity:

- 14573 depositions in 2021
- Better automation for biocurating incoming depositions
 - Based on existing structure annotations (where possible)
 - with > 100 ligand instances
 - Median processing time/entry reduced for large and/or complex entries (3-4 days to <2 days)
- Better automation for citation update based on DOI or PubMed ID

Additional information was provided on collaboration resources (Nov21-Oct22), Progress versus goals, resources for Nov22-Oct23, goal-setting and roadmap.

The wwPDB-AC recognizes that OneDep is a key uniting factor of the wwPDB. Once again we congratulate the OneDep team on the progress that has been made during a time of rapid and often unanticipated global change (e.g. new developments in structural biology, new wwPDB partners, COVID19 pandemic). As outlined in the 2021 report, the wwPDB-AC considers that continuing to build positive OneDep experiences for users by making the deposition system easier will be a driver of progress. Community engagement through existing workshops in the structural biology community also continues to be strongly encouraged by the wwPDB-AC.

We noted that several OneDep roadmap items continued to be delayed from previous years. These were listed in the pre-meeting papers as 1.1 (EDS), 1.2 (Chemical shift validation), 4.3 (Support depositor-annotated assembly) and 5.1 (X-ray validation reports – dependent on 1.1). The wwPDB provided reasons for the delay and outlined in more detail the progress that had been made. We anticipate that these items will be resolved in the coming year.

7. PDB Archive

Archive Growth in 2021

Year-end holdings 185,472

- 12,592 new entries released
 - ~10% increase *versus* 2019
 - ~7% growth in the archive
- Archival entries growing in both size and complexity
- Record 2,952 new 3DEM entries released
 - ~103% increase *versus* 2019!
 - ~23% of new 2021 entries!
- Total Entries=194,820 (as of 9/02/2022)
- Anticipate 200,000 structures in early 2023.

Core Archive Storage

- OneDep Sessions: ~38.8 TB
- ftp(legacy + versioned): ~1.3TB
- EMDB ftp: ~6.3 TB

CoreTrustSeal certification renewal in progress for 2023-2025 (CoreTrustSeal.org)

PDB archival data now being housed and delivered by Amazon Web Services (AWS) with no storage or egress fees

Global Biodata Coalition

wwPDB is seeking formal recognition of PDB as a Global Core Biodata Resource

- Expression of Interest submitted 04/13/2022
- Full Application submitted 08/04/2022
- Positive feedback received
- Awaiting decision

wwPDB plans to seek formal recognition for EMDB and BMRB in response to subsequent GBC calls

PDB Data Delivery by AWS

Amazon Web Services (AWS) Open Data Sponsorship Program now housing and delivering PDB data

- No storage fees charged to RCSB PDB
- No egress fees charged to PDB users

Current AWS holdings include:

- Annual PDB Archive Snapshots
- Current PDB FTP Archive (updated weekly)

AWS can deliver PDB data faster than RCSB PDB, PDBe, or PDBj!

The wwPDB-AC congratulates the wwPDB on the AWS arrangement, and looks forward to a positive response to the GBC application.

The wwPDB-AC noted a record new number of entries in 2022, and the continued growth in the chemical component dictionary with 2,500-3200 new compounds per year.

8. EMDB Archive

Improvements to 3DEM validation reports

- Reports now display the full model assembly in the 3DEM map
- Atom inclusion and Q-score mapped to the coordinate model is now shown in the validation reports
- A summary table containing information on chain and full structure atom inclusion and Q-score is presented in the reports

Q-score

- Provided for entries with maps that are below 1.25 Å in resolution
- The Q-score per residue is displayed on the atomic coordinate model
- The Q-score for the whole structure and per chain is presented in a summary table.

Atom Inclusion

- For entries with a map and model residue atom inclusion is displayed on the coordinate model is displayed
- Atom inclusion for the whole structure and per chain is presented in a summary table

Improvements to 3DEM deposition

- Mandatory deposition of half-maps for relevant modalities
- Improved support for large files during file upload
- Improved metadata cross-checks during deposition
- Collection of ORCIDiDs for 3DEM entries
- Improved composite map support
- Better 3DEM-data-related integrity checks
- Tomography deposition is undergoing rounds of review and improvement

Improvements to 3DEM annotation

- New metadata checks during annotation
- Molstar has been implemented to allow annotator review of 3DEM maps and models from within the OneDep annotation workflow

Awaiting Community Feedback

- Composite Map Deposition
 - First stage of implementation fully planned
 - Awaiting community feedback on the map naming conventions
- New Content Types
 - Planned and some implementation complete
 - Awaiting community feedback on names used for content types

Social Media

Twitter - Group of EMDB and EMPIAR employees responsible for weekly tweets

- 5600+ followers
- 3+ tweets a week

YouTube channel - hosting tutorials, talks and structure shorts launched June 2022

- 1300+ views
- 29 watch hours
- 59 Subscribers

20,000th EMDB entry released!

- Reached in May 2022 after 20 years
- Twitter posts organised to celebrate the occasion
- Over 22,000 entries by early September 2022
- 4,488 new entries released in 2021
- 4,033 released in 2021 (until early September)

The wwPDB-AC noted the report, and progress and achievements made over the past 12 months. We congratulate the EMDB on reaching the significant milestone of 20,000 entries.

In response to wwPDB-AC pre-meeting questions, we noted that the process of soliciting community feedback on composite map deposition had not yet started. The wwPDB-AC also noted that the per residue Q score is currently in production in OneDep, and that full recalculation of 3DEM validation reports was expected to be complete by end of Oct22.

9. BMRB Archive

Developments since 2021 Meeting

- NMR Restraints validation software was integrated into the OneDep system.
- OneDep system now accepts all experimental data (Chemical shifts, restraints and peak lists) in NMR-STAR v3.2 or NEF v1.2.
- The chemical shift validation code has been refactored and is now under testing.

- Working with PDBj on restraints remediation project.
- Deployed 1000-core CS-Rosetta server.
- Modernized the BMRB website (bmr.io).
- BMRBdep upgraded to newest version of Angular.
- ReBoxitory Data Lake (on NMRbox Platform-as-a-Service, nmrbox.org) now contains monthly snapshots of BMRB, PDB, NCBI, and the full AlphaFoldDB archive.
- Contribution to NAR Database Issue submitted September 22 2022.
- Invitation from *Science Advances* to contribute a perspective on biomolecular NMR (and BMRB) in the 21st Century.

Network for Advanced NMR: update

- First 1.1 GHz instrument in US is scheduled for delivery fall 2023.
- NSF awarded a planning grant to U Delaware (Hoch, co-PI) to develop a new \$100M Midscale-RI2 proposal.
- Tentative target is five 1.2 GHz NMR spectrometers, distributed geographically across the US, and connected to the NAN cyberinfrastructure for automatically harvesting experimental data.

The wwPDB-AC noted the update and developments of the BMRB over the past 12 months. We provide additional notes in response to Q6 below.

10. Joint Projects

Two joint NSF/BBSRC projects. Three year projects.

US funding provided by NSF; UK funding provided by BBSRC

Project 1: OneDep and NextGen PDB archive development

Project 1 started in Oct 2020 at RCSB and Jan 2021 at PDBe

Provides additional resources for OneDep development

Coordination with OneDep team

SA1. To improve fidelity and completeness of 3D structure data deposited into the PDB by

- Harvesting data automatically from structure determination software packages; and
- Streamlining the wwPDB data deposition, validation, and biocuration system.

Ongoing work towards enabling automated creation of and upload to deposition sessions

Will make deposition of investigations possible *via* OneDep

SA2. To improve “FAIRness” of PDB data for researchers, educators, and students by

- Extending chemical metadata for small-molecule ligands (*e.g.*, bound cofactors and inhibitors);
- Incorporating enhanced descriptions of macromolecular assemblies;
- Grouping related PDB structures into investigations for more efficient, parallel data delivery; and
- Creating a “Next Generation” PDB data repository with up-to-date metadata.

Background

- Each wwPDB site adds additional data to support their data out activities
- Aggregating of PDB entries is only possible after the PDB weekly release
- A NextGen FTP will be introduced which will make this additional data and add aggregated data available to users

First example of additional data

- Cross-reference to other data resources and residue level mapping to UniProt (based on the PDBe SIFTS data)
 - Added new categories to PDBx/mmCIF dictionary for SIFTS data
 - Includes
 - Per residue mapping to external resource (*i.e.* UniProt, Pfam, CATH, SCOP)

- UniProt numbering of residues (where appropriate)
- mmCIF files with additional SIFTS data are initially available from PDB FTP area

Three new PDBx/mmCIF categories added, representing segment-wise and residue-wise SIFTS annotations, mapping PDB residues to various data resources.

- `_atom_site` is also modified to indicate the best mapped UniProt sequence.

SA3. To modernize the IT infrastructure to futureproof PDB data management and weekly PDB archive release to the public domain by

- Developing new application programming interfaces (APIs) and microservices infrastructure; and
- Updating existing mechanisms for synchronization of data and software across wwPDB data centers in the US, Europe, and Asia.

Installation procedure of wwPDB OneDep documented and simplified - from 133 manual steps to single installation script

- OneDep is comprised of the following components:
 - System configuration
 - Core utilities
 - Web application and software
- OneDep system configuration
 - On-going work on simplification
 - All OneDep apps changed to support a simpler configuration
 - Required to support additional wwPDB sites
- Core utilities published to PyPi Python repository
- Web applications and software
 - Docker containers created for all public facing modules
 - Docker containers automatically built using CI/CD
 - To be deployed on Kubernetes
- File handling API to provide
 - Single access point for data in OneDep
 - Data exchange mechanism for OneDep
 - Groundwork for data exchange for SA2
 - Resumable upload functionality
 - Supporting asynchronous transfers
 - Handle file compression
- Would be rolled out initially on the deposition side.
- OneDep workflow engine
 - Currently bespoke software which is tied to individual servers
 - Investigated open source replacements which are suitable for containerisation
- Initial implementation developed to
 - Add resilience to regularly scheduled jobs (crons)
 - Monitor memory usage for OneDep processes
- Will initially prototype deposition workflow and assess before making a decision about using the same approach for all of OneDep

Project 2: Visualisation Infrastructure

Project 2 was started in December 2021

PDBe will need to complete the administrative processes at BBSRC

Will provide resources for development of Mol*, web-based components for displaying annotations and efficient data delivery mechanism

SA1 - To augment Mol-star for seamless operation across length scales ranging from atoms to cells with multiscale structure capability by

1. Extending existing data standards to combine atomic coordinates with multiscale structure information from integrative/hybrid methods and in-cell molecular details from advanced electron or visible light/super-resolution imaging techniques;

2. Building tools that enable superposition/simultaneous interactive display of multiscale structures; and
3. Enabling comparisons and analyses of multiscale structures.

SA2 - To extend 3D visualization beyond structure to include display of associated biological and functional annotations by

1. Developing a library of web-based components for displaying structural, biological and functional annotations;
2. Integrating Mol-star 3D display tools with complementary web-based components, such as those displaying 2D topology and 1D sequence data.

SA3 - To extend and strengthen existing IT infrastructure underpinning Mol-star by

1. Developing an agile data delivery system that supports high-speed interactive web-based visualization of 3D structure data; and
2. Enable comparison and analyses of multiscale structures across length scales ranging Å units to microns.

Project 2 Major Year One Activities

SA1 - Augment Mol-star for seamless operation across length scales ranging from atoms to cells with multiscale structure capability

- a. Data standards to be developed
 - Current Mol* uses BinaryCIF to transmit structure data, volume/density data is compressed using the same technology. Coarse-grained representations of multiscale structures are supported throughout Mol*.
- b. API to support multiscale data and multi-structure alignment

SA2 - Extend 3D visualization beyond structure to include display of associated biological and functional annotations

- a. API to facilitate efficient mapping and data integration for chemical, polymer sequence, and 3D structure comparison.
- b. 2D topology viewer for proteins and RNA is in place at PDBe. Investigating its inclusion into the project.

SA3 - To extend and strengthen existing IT infrastructure underpinning Mol-star by developing an agile data delivery system that supports high-speed interactive web-based visualization of 3D structure data for comparison and analyses of multiscale structures across length scales ranging Å units to microns.

- a. Views specification and infrastructure design are being developed

The wwPDB-AC congratulates the joint project partners on the progress made in 2021 to these improvements to OneDep. Regarding Project 1, SA1, we note that completion of this aim is still some way from being fulfilled. We appreciate that this requires complex negotiations since wwPDB is dependent in large part on structure determination software providers to generate output files with the new and improved data items defined by the wwPDB. Nonetheless, given the huge positive impact of this aim for the user community who deposit structures, we encourage active engagement by wwPDB with software providers to explore how this goal could be expedited. The mmCIF/PDBx working group continues to be an excellent forum for developing and approving new data items, and the wwPDB-AC supports its continued use. It may also be helpful to consult with wwPDB-AC members who are involved in community software development. Given the scale of the project, and that we are now 2/3rd of the way through the grant period, the wwPDB-AC asks whether wwPDB PIs could provide an updated roadmap/timeline towards achieving the outstanding specific aims within the remaining grant period (ending October 2023 and January 2024).

11. Responses to wwPDB PI questions posed to the wwPDB-AC

Questions posed by wwPDB are shown in bold text

Information provided by wwPDB is shown in normal font

wwPDB-AC responses are shown in blue italic font

Q1. Does the Advisory Committee have any questions or concerns regarding the performance of the organization since the 2021 wwPDB-AC Meeting (noting that the wwPDB-AC report arrived shortly after the slide deck was reported)? (SKB)

The wwPDB-AC discussion on this question focused on three areas: (i) the 2021 wwPDB-AC report, (ii) the EM validation white paper, and (iii) the format of wwPDB-AC meetings.

(i) The wwPDB-AC acknowledged the late provision of the 2021 wwPDB-AC report and undertook to provide the 2022 report in a more timely manner. The wwPDB-AC recognizes that the wwPDB-AC annual report and the subsequent wwPDB PI response to the report helps everyone keep track of issues, and assists the wwPDB-AC in providing helpful feedback and recommendations to the wwPDB. Notwithstanding the late delivery of the 2021 wwPDB-AC report, the 2022 wwPDB-AC members ask that the wwPDB provide responses to the 2021 recommendations and questions, some of which have been listed in the 2022 report above. The 2022 report format follows the same format used in past reports though in the future, under the current Chair, will focus on responses to the PI Questions. To reduce the time burden of preparing the report, it will not repeat information provided in the publicly available PI documentation, but may refer to page numbers in that documentation.

(ii) The wwPDB-AC noted that the EM validation white paper, referred to in the 2021 meeting report, is still listed as under preparation in the 2022 information. This paper also provides a public peer-reviewed record of the community effort that led to recent updates to the EM validation reports. It would be prudent to synchronize these activities in the future as much as possible. The wwPDB-AC recommends that this paper be submitted as soon as possible so that its publication can support the changes already implemented by the wwPDB and can demonstrate community endorsement of these changes. The wwPDB-AC would also appreciate an update on the plans to establish a working group focused on 3DEM content in both the PDB and EMDB archives (keeping in mind diversity and inclusion).

(iii) The wwPDB-AC discussed the purpose and format of the wwPDB-AC meeting with the overall aim to enable the wwPDB-AC to be of the greatest assistance and to provide the best advice possible to the wwPDB PIs. The wwPDB-AC agreed to provide suggested changes to the proposed format of the next meeting, shortly after the Oct 2022 meeting. Further information (including these suggestions) is provided in response to Q8.

Q2. Does the Advisory Committee have any questions/concerns regarding individual wwPDB member 2021 AC reports? (SKB/SV/GK/JCH/AP)

The wwPDB-AC received the individual wwPDB member 2021 AC reports, which have informed our responses throughout this report.

The wwPDB-AC recommends that all individual AC reports include the names of wwPDB members who were in attendance.

Q3. Does the Advisory Committee have any questions/concerns regarding training of PDBc Biocurators? (GK/WX)

wwPDB is Training PDB China Biocurators:

- PDBc agreement to the terms of the extant wwPDB Charter became effective, which enabled PDBc biocurators to access OneDep via the PDBc training instance physically located at Osaka (PDBcOneDep@Osaka).

wwPDB Plan: wwPDB Core Members provided two types of formal training; (i) remote on-line orientation and policy training provided by RCSB PDB and PDBe (done), and (ii) on-site hands-on biocuration training in Osaka provided by PDBj (on-going until November).

- PDBc biocuration performance will be assessed during an initial six month evaluation period that commences after completion of on-site training including the production processing assigned step-by-step by PDBj lead annotator.

wwPDB Plan: wwPDB Core Members will determine whether or not PDBc biocuration performance meets the global standard, defined in section 4 of the wwPDB Charter, and PDBc should be allowed to install and operate a OneDep system locally in Shanghai. During remote processing using PDBcOneDep@Osaka, depositions for PDBc biocurators will be assigned by the lead PDBj biocurator. Formal geographic direction of depositions to the PDBc server will only begin once the PDBcOneDep@Shanghai is in production.

The wwPDB-AC congratulates PDBc on the milestone of achieving the first 2 released structures.

We noted that China has heavily invested in EM infrastructure and is likely to provide a large proportion of EM structures to the PDB in the future. The wwPDB-AC would appreciate more information on how this specific challenge is being managed in the training of PDBc biocurators.

The wwPDB-AC recommends that wwPDB PIs provide a more detailed timeline with qualitative and quantitative measures for how PDBc (and PDBi in future) will be assessed as meeting the global standard to attain full wwPDB membership.

Q4. Does the Advisory Committee concur with wwPDB plans to collaborate with structure prediction/docking challenges (e.g. CASP, CAMEO, CELPP)? (SV)

wwPDB wishes to work more closely with CASP, CAPRI, etc.:

- wwPDB partners have worked with community efforts such as CASP and CAPRI to encourage depositors to contribute target structures. Support provided has included amplifying the request for targets using wwPDB channels and providing a mechanism in OneDep for depositors to identify a submission as a CASP/CAPRI target.

wwPDB Plan:

Explore the possibility of a more active role in supporting these community efforts by

- Working with the CASP and CAPRI teams to identify requirements for target selection;
- Exploring ways to extend OneDep functionality to identify targets (rather than depend on depositors to nominate their depositions);
- Developing joint funding opportunities with the CASP and CAPRI teams to implement additional support; and
- Using CASP and CAPRI related activities to establish the process to support community driven efforts.

The wwPDB-AC recognises the value of a collaboration between wwPDB and structure prediction/docking challenges but also notes difficulties that will need to be addressed. On the positive side, wwPDB interactions with CASP and CAPRI have contributed to dramatic breakthroughs in the accuracy of protein structure predictions which is revolutionising how structural biologists interpret and use experimental data. Further collaboration could benefit software developers who could continue to improve structure prediction from amino acid sequence, and improve prediction of (for example) protein:ligand interactions, membrane protein structures, protein conformational changes and multi-protein complexes. In addition, PDB depositors may benefit through raising the profile of their research. On the downside, such a collaboration could spread already limited wwPDB resources even more thinly. The wwPDB-AC was keen to understand the motivation for this proposal, how it aligns with the wwPDB Mission and Vision, and how and why it would be prioritized relative to other urgent tasks given current resource constraints. We note from the discussion at the wwPDB AC meeting that there is no funding available to implement thus planned collaboration at present, and that external funding would need to be sought if the plan is to proceed.

Q5. Does the Advisory Committee concur with wwPDB plans for managing grouped depositions as Investigations? (SKB)

wwPDB wishes to improve the OneDep depositor experience and group related structures together as “Investigations:”

- Multiple related structures have been reported in one publication (e.g., fragment screening study, PanDDA method, etc.).
wwPDB Plan: Enable single deposition of grouped structures as an Investigation defined by depositors.
- Refinement software packages now support PDBx/mmCIF formatted files for PDB depositions (e.g., Phenix, CCP4, Global Phasing, CCP4EM, etc.).
wwPDB Plan: Work with structure determination software developers to further automate file generation for investigations to support parallel deposition of multiple related structures

The wwPDB-AC members were excited by this plan, which seems to strongly align with the vision and mission of the wwPDB and could be a powerful means of representing complex sets of data. The plan is timely, given that the issues of grouped depositions have been discussed during the validation task force meeting of January 2020, and in the literature and on social media during 2022. The EM and fragment screening communities of structural biologists are likely to be highly supportive of this plan. Indeed, it seems likely that group depositions will become routine and indispensable in the future. However, whilst there was general support for this proposal, the wwPDB-AC felt there was insufficient detail to provide a definitive response to the specific question posed.

As the Investigations Project is at an early stage, the wwPDB-AC would welcome a more detailed implementation plan, that incorporates for example a timeline, prioritisation, coordinated wwPDB response, as well as community, synchrotron, software developer input.

Concurrently, IUCr is considering establishing a working group to investigate issues that have arisen around fragment screening depositions. Should this working group proceed, the wwPDB PIs undertook to provide an ex officio representative, on the understanding that this representative could provide feedback to and answer questions from the working group but would not provide explicit or implicit endorsement of recommendations or commit wwPDB to any specific course of action.

The wwPDB-AC recommend that there be a single solution to this problem, rather than a variety of community developed solutions which do not interoperate. The wwPDB is well placed to be the coordinator of a single, community accepted solution. The wwPDB-AC is supportive of the approach described by wwPDB PIs during the AC meeting, wherein fragment screening will be used as a first “investigation type” or prototype for investigation, and the architecture developed to be flexible allowing for multiple investigation types to be defined over time. We recommend that each investigation type be defined and designed in close collaboration with the structure solution software developers and representatives of the community most likely to deposit such investigations, to develop a workable plan that can then be implemented. Given the significant effort that will be required to develop and implement a system, the wwPDB-AC recommend that joint funding be sought with the software development community (we note the NSF/BBSRC model has proved fruitful for funding activities between RCSB and PDBe).

Q6. Does the Advisory Committee have feedback regarding the ongoing campaign for renewal of BMRB funding? (JCH)

wwPDB is responding to BMRB Funding Renewal Challenges:

- BMRB re-submitted a U24 proposal on October 3, 2022
- More than 100 letters were submitted in support of the proposal, including from the wwPDB PIs, two Nobel Prize winners, editors of major journals, scientists and

executives from biotechnology companies, early-career researchers, trainees, and students.

- wwPDB is collaborating with BMRB on an NSF Research Collaboration Network funding proposal to develop community standards for statistically robust structure validation.

The wwPDB-AC recognizes the continued importance of NMR datasets. We are concerned about the reported funding renewal challenges and the possibility of losing a community data repository for this substantial domain of science. We are acutely aware that funding support is needed to maintain the current resource and to support the use of NMR-derived data for understanding structure and function in biology in particular, as the focus moves away from structure determination to dynamics and function.

We understand that BMRB are actively exploring all possible funding options. We appreciate that an additional effort in highlighting the essential contribution of NMR to the structural biology landscape may be needed. wwPDB AC will draft a letter to the editor at Science or Nature supporting BMRB and emphasizing that NMR: fills an important gap characterizing and quantifying dynamics in ways that complement macromolecular crystallography and electron microscopy; that NMR permits simple, rapid experiments to validate computed structural models, enabling novel workflows in structural biology and affording atomic-level descriptions of disorder; and provides an important complement to the principal empirical methods as well as machine learning.

Q7. Does the Advisory Committee have any questions/concerns regarding deprecation of FTP for distributing PDB archive content? (SKB)

wwPDB is responding to changes in internet browser technology:

- The FTP protocol has been losing its popularity and the support by main web browsers (Chrome and Firefox)
- HTTP/S is recommended for the speed, statelessness, security, and is well-supported by a broad array of clients.

wwPDB plans to deprecate FTP protocol effective 11/1/2024, following multiple warning announcements (beginning 11/1/2022)

*The wwPDB-AC's pre-meeting on-line feedback included:
What is the current volume of FTP traffic?*

*Response: it has declined over past few years with rsync usage growing dramatically);
Has community input been sought?*

*Response: to be sought after the wwPDB-AC meeting
Would Globus GRIDFTP be considered?*

*Response: both Globus and Aspera are currently under consideration
In the BMRB AC notes, reference is made to use of ftp for archive of remediated data. How would this impact the proposed deprecation of FTP?*

Response: this is an error in the notes - BMRB has already deprecated ftp. Rsync and globus are already being used instead

Overall, the wwPDB-AC had no major concerns about the planned deprecation of FTP for distributing the PDB archive content.

Q8. Next Advisory Committee Meeting Hosted by PDBe: format/scheduling? (SV)

2023 wwPDB AC will be hosted by PDBe

wwPDB Proposed AC Meeting schedule for October 2023 (dates tbc)

Wednesday: wwPDB PIs Arrive at EBI

Thursday: Virtual Pre-meeting with AC Chair/Co-Chair (half-day)

Senior Biocurators Arrive at EBI
Friday: Virtual AC Meeting (half-day)
Saturday: wwPDB PI and Senior Biocurator Team Building
Sunday: wwPDB PI and Senior Biocurator Face-to-Face Meetings
Monday: wwPDB PI and Senior Biocurator Face-to-Face Meetings
Tuesday: Visitors Depart EBI

Questions:

Would the wwPDB-AC prefer a face-to-face meeting in Hinxton, UK (pandemic permitting)?

Which date would the wwPDB-AC prefer - Friday October 13th or Friday October 20th?

Which format would the wwPDB-AC prefer?

Virtual/Hybrid-appears to be the wwPDB-AC Preference.

As a consequence of the COVID19 pandemic the 2020, 2021 and 2022 wwPDB-AC meetings have all been held online. The wwPDB-AC recognises that it is important to hold regular face-to-face meetings, even if these are not held annually. A fully face-to-face meeting has many benefits: it can build stronger relationships, support improved communication, minimise the impact of timezone differences and allow extension of the meeting timeframe. Virtual/hybrid meetings also have a number of benefits: significantly lower travel and accommodation costs, minimized time lost due to travel, significantly reduced carbon footprint, increased participation by those who cannot travel, and decreased risk of contracting an infectious disease during long-haul travel.

Keeping in mind all these factors we recommend that the 2023 wwPDB-AC meeting be a hybrid F2F/virtual meeting and arranged with suitable systems in place to ensure those on-line are fully included in discussions. We agree with the wwPDB PIs that, where possible, those attending the 2023 wwPDB-AC meeting in person and traveling internationally, should coordinate their travel with other reasons for being in the UK/Europe. In future, we recommend that in-person AC meetings be timed to facilitate attendance at other international meetings of interest to PIs and AC members (eg ACA/AsCA/ECA/IUCr). This may require changing the customary October timing of wwPDB AC meetings.

Regarding the proposed schedule:

The wwPDB-AC members found the 3-hour virtual meeting for the 2022 meeting to be rushed, and this limited our ability to properly/fully discuss the issues raised in the wwPDB papers. However, extending the meeting to 4 hours to enable further discussion risks losing those in far-distant timezones. The proposed 2023 hybrid meeting would involve people attending from across the globe, some joining virtually.

We make the following suggestions for the 2023 hybrid meeting:

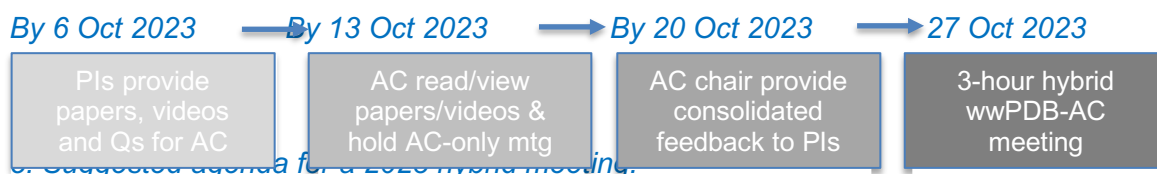
1. At least 3 weeks prior to the scheduled hybrid wwPDB-AC meeting (i.e., by early Oct), each wwPDB PI provide a 5-10 min pre-recorded video to accompany their pre-meeting papers. To ensure that feedback and responses to wwPDB-AC reports are documented appropriately we recommend that the pre-meeting papers include any summaries/updates to recommendations/questions raised in the previous wwPDB-AC report.

2. At least 2 weeks prior to the scheduled hybrid wwPDB-AC meeting (i.e., by mid-Oct), schedule a 2-hour wwPDB-AC-only virtual meeting to allow the advisory committee to discuss the pre-meeting papers/videos and the questions.

3. wwPDB-AC members will need to schedule time in early Oct 2023 to read the papers and view the videos before the virtual AC-only pre-meeting. We anticipate that this wwPDB-AC-only pre-meeting will reduce the time needed for wwPDB-AC executive sessions in the hybrid meeting.

3. At least a week prior to the scheduled hybrid wwPDB-AC meeting (by 20th Oct), the wwPDB-AC chair provide a consolidated list of feedback and questions to the wwPDB PIs.

4. Schedule a 3-hour hybrid wwPDB-AC meeting on Friday 27th October (to accommodate new points 1, 2 and 3).



Suggested agenda for a 2023 hybrid meeting.

- Welcome and Introduction (10 min)
- Discuss questions/answers any other issues - ALL (90 min)
- Executive Session – wwPDB-AC (35 min)
- Summarise wwPDB-AC feedback and further discussion – ALL (35 min)
- Acknowledgements and photo – ALL (10 min)

13. Conclusion

The wwPDB-AC meeting concluded on time.

The wwPDB-AC commends the wwPDB partner PIs and their teams on their coordinated and collaborative approach to support this incredibly valuable and unique global scientific resource. We acknowledge the continued growth of the archives and the significant work required for maintenance and development, which is even more impressive given the ongoing impact of the COVID19 pandemic. During 2022, PDBc was formally approved as an associate member, and PDBi continues to progress towards associate membership. This year has also seen further rapid growth in EM depositions and predicted protein structures. Funding for individual members remains challenging, though there could also be new opportunities in this space.

Overall, the wwPDB partners are to be congratulated on the remarkable job they do to maintain a hugely complex operation, that requires a collective and shared responsibility to implement the wwPDB charter, mission and vision. We applaud you and thank you.

The next wwPDB AC meeting will be hosted by PDBe at Hinxton in the UK, and is proposed to be a hybrid meeting on 27 October 2023.