

**Worldwide Protein Data Bank Advisory Committee (wwPDB-AC)
Report of October 19th, 2021 Meeting
PDBe/EMDB organized online meeting**

Chair: Peter Rosenthal

Co-Chair: Corinne Smith

PDB Site Representatives (Nominated by wwPDB partner): Paul Adams (RCSB-PDB), Kirk Clark (RCSB-PDB), Angela Gronenborn (BMRB), R. Andrew Byrd (BMRB), Martin Noble (PDBe), Martin Walsh (PDBe), Toshiya Senda (PDBj), Daisuke Kohda (PDBj), Corinne Smith (EMDB), Juha Huiskonen (EMDB), Alexandre Bonvin (EMDB)

Ex Officio Community Stakeholder Representatives: Jenny Martin (IUCr), Tatyana Polenova (ICMRBS/ISMAR), Peter Rosenthal (Macromolecular EM)

wwPDB Members: Stephen K. Burley (RCSB-PDB), Sameer Velankar (PDBe), Jeffrey C. Hoch (BMRB), Genji Kurisu (PDBj), Ardan Patwardhan (EMDB)

wwPDB Regional Representatives (Associate member candidates): Debasisa Mohanty (India), Wenqing Xu (China)

Institutional Representative: Gerard Kleywegt (EMBL-EBI)

wwPDB-AC E-mail Addresses:

PDAAdams@lbl.gov, kirk.clark@novartis.com, amg100@pitt.edu, byrdra@mail.nih.gov, martin.noble@ncl.ac.uk, martin.walsh@diamond.ac.uk, toshiya.senda@kek.jp, kohda@bioreg.kyushu-u.ac.jp, corinne.smith@warwick.ac.uk, juha.huiskonen@helsinki.fi, A.M.J.J.Bonvin@uu.nl

Ex Officio Community Stakeholder E-mail Addresses:

jlm@griffith.edu.au, tpolenov@udel.edu, peter.rosenthal@crick.ac.uk

wwPDB Member E-mail Addresses:

sburley@proteomics.rutgers.edu, sameer@ebi.ac.uk, hoch@uchc.edu, gkurisu@proein.osaka-u.ac, ardan@ebi.ac.uk

wwPDB-AC Chair E-mail Address:

peter.rosenthal@crick.ac.uk

wwPDB Regional Representatives E-mail Addresses:

deb@nii.ac.in, xu.wq@zjlab.org.cn

Institutional Representative E-mail Addresses:

gerard@ebi.ac.uk

wwPDB AC Meeting, October 19, 2021:

The Worldwide Protein Data Bank Advisory Committee (wwPDB-AC) and the leadership of the Research Collaboratory for Structural Bioinformatics (RCSB-PDB), the BioMagResBank (BMRB), the Protein Data Bank in Europe (PDBe), the Electron Microscopy Data Bank (EMDB), and the Protein Data Bank Japan (PDBj) met for an online meeting hosted by EMDB on October 19th 2021.

Agenda

14:00	Welcome and Introductions	All
14:10	Executive Session No. 1	SAB
14:30	Discussion/Questions for AC	All
15:30	Executive session No 2	SAB
16:00	Feedback to wwPDB Leadership	SAB + wwPDB PIs
16:20	Acknowledgements	Ardan Patwardhan
16:30	Group Photo/Meeting Close	All, Photo <i>via</i> Host

Introduction and Overview of the wwPDB

Members of the AC were welcomed online by the wwPDB PIs.

Summary: The past year has been an exciting one for the wwPDB. The AC congratulates the partners on the occasion of the 50th anniversary of the PDB as an essential resource for the scientific community. Furthermore, the wwPDB has continued to thrive through a second year of the pandemic, making essential progress, and continuing to underline the importance of the wwPDB to science and human health. The recent success in protein structure prediction (e.g. alphaFold2) is a significant development for biology, and the wwPDB is responding quickly to the new challenges posed by these exciting developments. This is clearly a time of significant change and there is much work to do. A new Memorandum of Understanding (MOU) was signed in 2021. New PDB partners have become active. Overall funding continues to be an area requiring attention, and efforts will be required to cover both existing requirements and future needs arising from the current growth in structural biology. The perseverance and success of the wwPDB PI's and their teams through the pandemic has been inspiring.

The wwPDB Architecture (**Figure 1**) consists of core archives, core members, associate members, and federated resources as in the past. EMDB became a core member/core archive.

The new wwPDB Charter is available as a pdf document on the wwPDB.org website (<https://www.wwpdb.org/about/agreement>). Figure 1 also provides a future framework, as communicated by the wwPDB PI's for integrating new sources of information such as structure prediction and modelling into a resource for both specialists and end users and preserves simplicity and reliability.

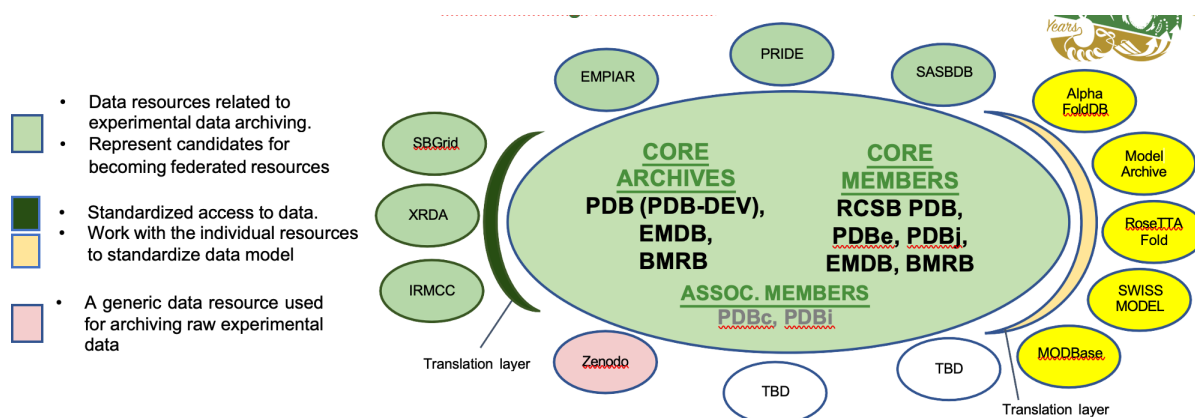


Figure 1

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.

Developments since 2020 AC Meeting

wwPDB New Charter Signed

- Effective Date January 1, 2021
- Admitted EMDB
- Established the framework for admission of PDB China and PDB India as Associate wwPDB Members (with roadmap for transition to Full Membership)
- Adopted Creative Commons CC0 1.0 Universal License for all wwPDB Core Archives

The AC is pleased to see this significant development! In the future, the AC will appreciate receiving the earliest possible circulation of important governance documents so that the AC can be the best possible resource to the wwPDB partners.

Core Archive Updates

PDB Core Archive Update (Stephen K. Burley)

CoreTrustSeal certification renewed through April 2024 (CoreTrustSeal.org)

Archive Growth in 2020

Record 14,044 new entries released (~22% increase versus 2019)

~9% growth in the archive

Record 2,780 new EM entries released (~40% increase versus 2019)

Total Entries=182,418 (as of 9/23/2021)

Anticipate 200,000 structures in early 2023.

PDB Chemical Reference Data: Improved access of chemical references and information about PDB archive holdings

~34K Chemical Component Definitions

3118 new in 2020 (Up 29% over 2019)

~1K Biologically Interesting molecule Reference Dictionary (BIRD) Definitions

174 new in 2020 (Up 9x over 2019)

130 updated (Up 60x over 2019)

BMRB Core Archive Update

OneDep Activity

Completed

- Support for single file upload for NMR data in with NMR-STAR/NEF in OneDep
- NMR distance and dihedral angle restraints analysis is now available through OneDep and wwPDB validation servers.

On-Going

- Installation of OneDep instance at BMRB
Docker method installed.
One script installation, PDBe working on simplifying config files.
- Chemical shift validation code refactoring
- Remediate restraints data available in BMRB restraints grid and copy to wwPDB FTP
- Generate restraints validation report for the whole archive.

BMRB Service Updates

- Completed migration of BMRB services to UConn
- Completed migration of BMRB-related services developed in tandem with NMRFAM (RUNER, GISSMO, etc.) to UConn
- Completed migration of BMRB CS-Rosetta server to CHTC
- Substantial packaging, speed, and error message helpfulness improvements to PyNMR-STAR library
- Progress towards re-implementation of BMRB dictionary management tools in Python (replacing decades-old Visual Basic code)
- Completed and deployed M2MTool to facilitate BMRB depositions directly from inside of NMRbox
- Partnered with NMRbox to add BMRB to the NMRbox “Reboxitory” data lake

Additional news of import to BMRB and wwPDB: Network for Advanced NMR NSF midscale R12 4-year grant to UCONN, UW-Madison Virtual Network of 26 spectrometers. All connected spectrometers will stream experimental data to a central archive with simple opt-in for public accessibility and curation.

EMDB Core Archive Update

Improvements to EM deposition

- Improved upload of large files
- Better EM-data-related integrity checks
- General improvements concerning how EM-related data is collected
- A large number of historical bugs fixed

Over 10,000 EMD entries achieved in February 2020.

Over 16,000 entries by mid-August 2021

EM Data Management Workshop (EM-VTF) January 2020 (over 40 participants):

- Recommendations about improved data capture by wwPDB/EMDB, validation reports. To be published as white paper.
- Recommend to make PDBx/mmCif mandatory for model deposition from 1st July 2021
- EMDB to implement and make available methods where recommendations cannot yet be made to enable archive-wide analysis and expert assessment of performance in individual cases
- Higher level grouping of related entries, requires discussion.
- Single particle half-map deposition is recommended to be mandatory.
- White paper in preparation (Kleywegt *et al.*)

Improvements to EM validation reports

OneDep now generates EM map-only and map model validation reports. Reports generated upon deposition, annotation and release of an entry. A number of EM-VTF recommendations already implemented. Soon the full validation reports will be included for everything, not just the more recent entries.

Added Features:

- If half maps were deposited: calculate raw map, show various slices and projections of the raw map
- Include orthogonal surface views of the raw map
- Rotationally averaged power spectrum of raw map added to RAPS plot.
- Visualisation of map and full assembly (e.g., for viruses).
- Improvements to the FSC plot: FSC plot now only shows '0.143', '0.5' and 'half-bit' criteria, Author-provided and calculated FSC curves in one plot, Resolution-estimate table re-designed for easier viewing.

JOINT PROJECTS DISCUSSION

Three-year projects. The US funding is provided by NSF and the UK funding is provided by BBSRC.

- Project 1 started in Oct 2020 at RCSB and Jan 2021 at PDBe provides additional resources for OneDep development.
- Project 2 was awarded in August 2021 and focuses on visualization infrastructure.

The AC acknowledges the growth of the archives and significant work of the wwPDB partners and their teams in maintaining and developing this important resource. The AC is pleased to see continued community consultation on EM validation and deposition and looks forward to the white paper release.

OneDep 2019/2020 Progress versus Goals

wwPDB Biocurator Productivity

- Biocurator productivity is at an all-time high.
- Better automation for curating incoming datasets
- 2021/2022 roadmap provided but may be subject to adjustment.

The wwPDB-AC recognizes continued progress and success on OneDep and we congratulate Jasmine Young and the OneDep team for their work during this period which has included new developments in structural biology, new wwPDB partners, and stresses associated with the pandemic. For the future, the AC believes that continuing to build positive OneDep experiences during deposition by making the deposition system easier for the user will be a driver of progress. Community engagement from existing workshops in the structural biology community is strongly encouraged.

wwPDB Core Member Funding Status

- RCSB PDB: Joint NSF/NIH/DOE funding renewed: 2019-2023
- BMRB: NIH NIGMS funding: 2019-2023, still needs additional support. UConn funding 25% administrative assistant and 25% project manager positions. NIH U24 submitted but not funded. Re-submitted 9/21.
- PDBe: EMBL-EBI, Wellcome Trust: 2021-2025
- PDBj: NBDC-JST and AMED funding: 2019-2022
Additional budget from S. Korea (decision pending due to COVID-19)
- EMDB: EMBL-EBI, Wellcome Trust: 2019-2024
- RCSB PDB/PDBe: Joint NSF/BBCRC NextGen Archive funding: 2020-2022
- RCSB PDB/PDBe: Joint NSF/BBCRC Mol* Visualization funding: 2021-2024

The AC recognizes the success of the wwPDB partners in securing stable funding. We understand that to some extent the entire activity is underfunded compared to community need and wwPDB potential contributions. BMRB funding may need specific attention. Validation and data deposition may require more funding going forward. There will be additional challenges posed by new initiatives reflecting growth and changes in structural biology, but these changes may represent funding opportunities as well. The AC emphasizes that collaboration between partners is an asset when seeking funding and encourages the wwPDB PI's to engage with the AC for help in this area.

wwPDB Outreach and Expansion

PDB50 Virtual Celebration

wwPDB/ASBMB: May 4-5

RCSB PDB/ACA: July 30-31

RCSB PDB/ACS: August 25

RCSB PDB/BPS: October 6

PDBe/EMBL: October 20-22

RCSB PDB/Rutgers IQB: November 2

PDBe/RCSB PDB/RSC: November 16 and 18

PDBj/BSJ : November 24

wwPDB Foundation

<http://foundation.wwpdb.org>

The wwPDB Foundation is a private entity that works with the wwPDB and acts to support outreach activities of the wwPDB that cannot be supported by the individual partner funding sources. Chair is Celia A. Schiffer, U. of Massachusetts Medical School. Fundraising continues and PDB50 celebrations planned

This has been an important year for Outreach activities celebrating the 50th anniversary of the wwPDB. While live celebrations would have been the preferred way to celebrate this milestone, these were made problematic by the pandemic. Nevertheless, the Outreach celebration activities were an enormous success. The AC suggests maintaining the record of the those celebrations in its current form on the wwPDB website.

wwPDB Associate Members Updates

The associate members (See Figure 1. wwPDB Future Architecture) are PDB China (PDBc) and PDB India (PDBi)

PDBc update

Wenqing Xu, Director

National Facility for Protein Science in Shanghai (NFPS) and iHuman Institute and SIAIS, Shanghai Tech University, Pudong, Shanghai, China

- The National Facility for Protein Science in Shanghai (NFPS), the primary host of PDB China, has purchased computational hardware dedicated to PDB China (with >\$1.2m USD) and performed related infrastructure set up.

- NFPS has recruited/formed the Data-in team (5 people) and the joint Data-out team (5 people, in collaboration with ShanghaiTech University).
- PDB Japan obtained funding to purchase the computers for PDB China's first OneDep data-in system (to be sited in Osaka); equipment just arrived; remote data-in training will start soon.
- The first version of the PDB China data-out system, with primary search functions, has been developed, and is under test and development.
- Dedicated future funding for PDB China of ~\$4.2m USD from MOST (Ministry of Science and Technology of China) can be expected for 2022-2026. More funding from other funding resources, including those from the Shanghai government, the ShanghaiTech University and NFPS, can be expected. No funding concerns.

PDBi update

PDB India, Manju Bansal (PI), Debasisa Mohanty and K. Sekar, Co-investigators
Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India

- First phase funding (Oct 2019-2022) received from NSM(India) for hiring staff, but second phase support for procurement of servers/storage has not materialized yet.
- PDBi PIs have made arrangements to utilize servers/storage of the Indian Biological Data Center (IBDC) which will be operational by December 2021. IBDC is a facility funded by DBT, India and efforts are underway to secure PDBi funding from DBT for 2022-2027.
- 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 PPI network of M.Tb./Plasmodium and integrating other structure analysis software with a local FTP mirror of PDB coordinate files.
- Discussions have been held with PDBe (Dr. Velankar) for online remote training of two PDBi staff on annotation/curation of structural data after securing required permissions from EBI/wwPDB PIs.

The AC welcomes progress on PDBc and PDBi as important to the overall project and the community. The AC understands that the next step is an agreement on Associate Membership. The AC would like to be informed on timelines and milestones for inclusion of partners. Of particular importance are the issues of load balancing and quality control. The AC requests that AC reports of the individual partners be made available. The pandemic has made international interactions difficult, though all partners have risen to the occasion and hopefully the future will afford better opportunities for in-person interactions. In addition to the contribution of specific partners, the OneDep team under Jasmine Young has played an important role in these efforts.

Specific Questions/Actions for the wwPDB AC

1. Does the Advisory Committee have any questions regarding our responses to the 2020 AC report?

- *The AC appreciates responses to the 2020 AC report and also emphasizes that more detailed responses to the AC report will assist the AC in providing feedback and making recommendations.*
- *AC strongly supports wwPDB plans to establish a working group focused on 3DEM data content in both the PDB and EMDB archives in 2022. The AC emphasizes the importance of demographic diversity (including age and gender) in the membership of future task forces as with membership across all wwPDB activities and partners. In addition, a broad communication will encourage innovation. The AC is aware of the need to balance inclusion of younger scientists and more senior colleagues whose experience can support wwPDB on important funding goals.*
- *Some issues in the 2020 report still require discussion in the future (e.g. future role of bioRxiv publication and supporting data release) but have not been center stage at this meeting because of urgent issues surrounding structure prediction. The AC notes the EM validation white paper is under preparation and will be finished soon.*

2. Does the Advisory Committee have any questions/concerns regarding individual wwPDB member 2021 AC reports?

- *The wwPDB AC accepts the individual wwPDB member 2021 AC reports, and they have informed our responses throughout this AC report.*

3. Does the Advisory Committee concur with our updated plans to engage with the relevant structure prediction and experimental data resources to accelerate development of the wwPDB data federation, employing community data standards and distributed archiving?

- *The AC concurs with these plans. We understand that these plans reflect dramatic breakthroughs in the accuracy of 3D structure prediction from amino acid sequence which are likely to significantly affect how the structural biology community and the general user community interprets experimental data. Further similar developments are anticipated and the AC endorses staying ahead of the curve on this. Figure 1 in this report*

(reproduced from the PI presentation) is a schematic for the future wwPDB and federated databases. While similar to past schematics and consistent with the Charter signed in 2021, federated databases now include those for structure prediction. This will be facilitated by adopting common data standards (e.g. mmCIF) as well as uniform APIs through a “translation layer”, as broadly captured in Figure 1, an approach which the AC endorses. We understand this will be a common project for wwPDB members, though each partner may develop unique support infrastructure. The AC emphasizes that all users of the wwPDB should see consistent maps, models, and interpretation despite different access points for the information. The AC feels the wwPDB can provide the leadership in access to predicted structural data while continuing with its main mission to provide a database of structures supported by experimental data. The AC would like more information as these plans are further developed. Community consultation is recommended to help identify the most important priorities.

4. Does the Advisory Committee concur with our goal of engaging the community to develop general, statistically rigorous tools for validation of both experimental structures (e.g., MX, 3DEM, NMR, I/HM) and computed structure models?

- The AC agrees that a Bayesian framework for validation of both experimental and structural models may be an important long-term goal for structural biology. However, the AC requires more information to understand the goals and scale of the proposed engagement. The AC is skeptical given rapid progress and available wwPDB resources that this should be viewed as an urgent focus of development for the wwPDB, and development may come from a broad community. The AC supports a leadership role for the wwPDB in engaging with methods and software developers. The AC asks the wwPDB to discuss the example and possible future role of pdb-DEV and would be pleased to discuss specific sources of expertise for consultation. We encourage the wwPDB to help define validation problems and aspirational goals for structure validation.*

Conclusion

The wwPDB AC meeting concluded with a general discussion with the wwPDB PIs following the executive session.

The next wwPDB AC meeting will be hosted by RCSB PDB at Rutgers University in Piscataway, NJ, USA as a hybrid meeting on October 14th or 21st in 2022.