

Worldwide Protein Data Bank Advisory Committee (wwPDB-AC)
Report of October 2nd 2015 Meeting
PDBj, Institute for Protein Research, Osaka University, Osaka, Japan

Chair: R. Andrew Byrd

PDB Site Representatives (Nominated by wwPDB partner): Paul Adams (RCSB-PDB), Cynthia Wolberger (RCSB-PDB), Angela Gronenborn (BMRB), Masatsune Kainosho (BMRB), Helen Saibil (PDBe), David Brown (PDBe), Genji Kurisu (PDBj), and Daisuke Kohda (PDBj)

Ex Officio Community Stakeholder Representatives: Edward N. Baker (IUCr), R. Andrew Byrd (ICMRBS), Wah Chiu (Macromolecular EM) [unable to attend]

wwPDB Members: Stephen K. Burley (RCSB-PDB), John Westbrook, Jasmine Young (RCSB-PDB), Sameer Velankar (PDBe), Rolf Apweiler (EBI, representing PDBe), John Markley (BMRB), and Haruki Nakamura (PDBj)

wwPDB Associate Members: Manju Bansal (India), Jianping Ding (China)

wwPDB-AC E-mail Addresses:

PDAdams@lbl.gov, cwolberg@jhmi.edu, amg100@pitt.edu, kainosho@tmu.ac.jp, h.saibil@mail.cryst.bbk.ac.uk, D.G.Brown@kent.ac.uk, gkurisu@protein.osaka-u.ac.jp, kohda@bioreg.kyushu-u.ac.jp

Ex Officio Community Stakeholder E-mail Addresses:

ted.baker@auckland.ac.nz, byrdra@mail.nih.gov, wah@bcm.edu

wwPDB Member E-mail Addresses:

sburley@proteomics.rutgers.edu, jasmine@rcsb.rutgers.edu, john.westbrook@rcsb.rutgers.edu, sameer@ebi.ac.uk, apweiler@ebi.ac.uk, gerard@ebi.ac.uk, harukin@protein.osaka-u.ac.jp, markley@nmrfam.wisc.edu

wwPDB Associate Member E-mail Addresses:

mb@mbu.iisc.ernet.in, jpdning@sibs.ac.cn

wwPDB AC Meeting, October 2, 2015:

wwPDB-AC Mission Statement

To help ensure that the Protein Data Bank is maintained for the public good as a secure, single global archive for experimental structural biology data that is freely accessible in perpetuity.

Meeting Summary

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), and the Protein Data Bank Japan (PDBj) met at PDBj, Institute for Protein Research, Osaka University, Osaka, Japan, on October 2nd 2015.

The agenda included

- (1) Overview & State of the wwPDB and Outreach;
- (2) Milestones, Improvements, and Impact of the Common Deposition & Annotation Tool (D&A) on X-ray Depositions;
- (3) NMR;
- (4) EM;
- (5) D&A V2.0 progress;
- (6) Looking ahead, questions, discussion topics; and
- (7) Executive session and feedback

Following a welcome by Prof. Haruki Nakamura, the overview of the current state of the wwPDB and Outreach efforts were presented by Haruki Nakamura.

(1) Overview and Outreach (presented by Haruki Nakamura):

The wwPDB has continued to make outstanding progress in this past year. This remains one of the most impressive and influential databases and services in the world.

Milestones of particular significance include:

Number of total entries exceeds 112,000

Use of data by the scientific community continues to show incredible growth, and it is recognized that the majority of use is by non-structural biologists, emphasizing the impact of the resource and archive on the worldwide scientific and lay community.

Creation of a new web presence for wwPDB that reflects the broadening, non-specialist user base.

Significant improvement in the information describing bound ligands and conditions within the archive/database.

Significant increase in the number of large entries, validating the efforts by wwPDB to shift to new, forward looking formats to support the growth and direction of new methodologies.

Clear and dramatic success for the D&A versions 1.0 and 1.5 supporting deposition of X-ray data. This has come on-line smoothly and has had significant impact in enabling load sharing across the partner organizations for annotation.

Development of new proposals for future expansion and impact, particularly addressing new experimental modalities and the ability to federate data base systems while retaining the single PDB archive under the auspices of the wwPDB.

Leadership in new directions:

Hybrid methods task force, NMR restraints format and working group, and ligand validation workshop.

The expansion in hybrid structure determination is recognized by the wwPDB and they are active partners with the specific method-communities to develop data archiving and sharing tools to meet the coming needs of the global scientific community.

The development of resources by the wwPDB has been quite impressive. The project continues to provide critically important support and leadership for the structural biology community to impact this area of science and its integrated use by the broad scientific and lay community.

FUNDING: The Advisory Committee was very pleased to see that the partners are currently in a relatively stable financial situation. The wwPDB partners are funded individually and work collectively for the scientific community. Currently, all partners have funding through at least the first quarter of 2017, and three partners are funded to 2018 and 2019. Despite this relatively solid footing, the AC supports continued efforts by the partners to establish new and sustainable funding sources for the future. The previous precarious position of BMRB has been stabilized with NIH funding; however, future funding will likely require new initiatives.

(2) Report on the D&A tool (presented by Stephen Burley)

The D&A tool has been in production since January 2014 (V1.0) and a review of its impact and progress was presented.

Initial operation was paired with development and bug corrections, leading to deployment of V1.5 in September 2014. Phasing out of the prior deposition system, ADIT, began in January 2015. For a number of technical reasons, the system was operating primarily at the Rutgers site with annotation support from PDBj, RCSB (Rutgers), and PDBe. In April 2015, full deployment was completed for UCSD and PDBj sites, and in July 2015, the ADIT phase out was complete. At the present time, **all** X-ray depositions are made using the D&A tool. Improvements to the annotation pipeline continued throughout 2015. Utilizing the D&A, improvements were made to sequence annotation and cross-referencing, ligand assessment, and management of the workflow for both individual depositions and across the partnerships of wwPDB.

A review of the metrics for data deposited to date (>9500 structures) was impressive in terms of the ability to accomplish effective load balancing of the annotation task. The AC was most impressed the increased efficiency and enhanced annotation for large and complex structures. The median time for a deposition through annotation decreased from 16.5 days to 1.6 days. The efficiency of annotator function is clearly improved, as was evident from the statistics presented. This increased efficiency bodes extremely well for the ability of the wwPDB teams to keep up with the increasing volume and complexity of large

structural depositions and will free manpower to address the growing input of structural data from new and hybrid methods. This is a critical point for the future of the wwPDB: to handle higher volumes and more complex data with near constant resources.

An important concern by the wwPDB AC was the increase in coordinate replacement during deposition. Overall, this was viewed by the AC as a byproduct of success that may shift and mature with time. The improvement in the Validation Report and the ability to iterate refinement while monitoring the validation reports appears to be driving this problem. The partners are encouraged to track these metrics and determine if the situation worsens or subsides with broader recognition and use of the validation suite. This may induce greater pressure for a public release of the validation software suite, as has been discussed in previous AC meetings.

(3) Report on the BMRB (presented by John Markley)

The news of stabilized funding was greeted warmly and enthusiastically by the AC. The current NIH funding through NIGMS is at a reduced level but will support the effort through March 2019. The Committee supports active efforts by BMRB to strengthen this base by developing new initiatives with respect to ongoing activities in validation, the common exchange format for NMR restraints, and hybrid methods developments.

The AC remains enthusiastic and strongly supportive of the role of BMRB within the wwPDB. This becomes critical in the completion of D&A V2.0 and the forward looking approaches to hybrid structure determinations and providing critical data for both developing and validating structures of large complexes using EM methods.

Trends in data deposition indicate approximately steady deposition rates for structures based primarily on NMR data. The AC feels that the BMRB partner needs to play an expanded role in leading the community with regard to the NMR Validation Task Force and the forthcoming NMR Exchange Format (NEF) effort. Future trends are expected to see NMR data play an increasingly significant role in studies of complexes through hybrid structural determinations, and these expectations require an active partner working to establish data deposition, archiving and validation procedures.

The committee feels that the NMR Validation Report and server need to accompany release of D&A 2.0. There were concerns that strains within the NMR community are a consequence of parallel activities that are insufficiently coordinated. These include the roles of the NMR Validation Task Force (VTF), the NMR Exchange Format (NEF) activities, and coordination with BMRB. It is imperative that the leadership of wwPDB takes a decisive lead to resolve any NEF-associated data dictionary or compatibility (NMR-Star format) issues. As a first

step, a strong and actively engaged NMR task force is necessary. Activities that could develop into a parallel route to D&A, not planned for by BMRB, should be addressed in a timely manner and every attempt should be made to have a united and unified approach. It was noted that D&A 2.0 is not dependent on NEF; however, NEF will clearly impact NMR depositions in the next 1-2 years and resolution of issues and a coordinated, smooth deposition must be addressed. Indeed, the committee believes that the NMR depositions could serve as a model to develop the future integrated methods-derived data, given the diverse nature of input data currently already inherent to NMR depositions.

NMR issues with the VTF and NEF need to be addressed by leadership to establish proper authority and the path forward. There are viable models from other VTFs to follow for more efficient progress. The proposed meeting in 2016 at the ICMRBS is deemed too late for these actions and a more aggressive timeframe should be undertaken.

(4) Report on the state of Electron Microscopy (EM) (presented by Sameer Velankar)

The developments for EM are very impressive and the outlook for supporting this community is outstanding. The wwPDB is being very proactive in recognizing the explosion in this area of structural biology. The efforts to cooperate and lead in establishing standards and validation tools are critical to the broad scientific community.

The parallel efforts ongoing in the EM community involve the wwPDB, the EMDataBank project (involving RCSB-PDB, PDBe, and the National Center for Macromolecular Imaging at Baylor College of Medicine), and the EMPIAR effort at EBI. The committee strongly supports the activities and the continuing example set by wwPDB partners to be active in the community and provide leadership in data protection, validation, and archiving activities as the frontiers of this community expand at a rapid pace.

The rapid growth in depositions makes it all the more urgent to complete and release D&A V2.0, which will support the methods of NMR and EM. Depositions are already being annotated at multiple sites, which should improve with D&A V2.0. Progress is good for the initial version of validation reports, and continued development of validation procedures is anticipated in cooperation with the community.

An exciting development in the EM area is the EMPIAR project. This activity is conducted at PDBe with separate funding from BBSRC and MRC. The goal of EMPIAR is to store raw 2D image data associated with EMDB and PDB entries. This effort may be a concept-leader for storing and archiving other large data types. EMDB and EMPIAR present models for the proposed 'federated data resources'.

Validation is viewed as a critical component of the rapidly growing EM field. It was noted that the last meeting of the VTF for EM was in 2010, and the AC suggests that wwPDB consider convening another meeting and discussion by this VTF. There are significant activities in the EMDataBank project associated with validation, including the 2015 Map and Model Challenges. Progress is strong, and the development of validation reports that mirror those developed for X-ray and NMR data (with respect to entries with atomic models in addition to the density maps) will contribute to a unified view of structural biology by the non-expert scientific community. These efforts are essential to the future impact of the field and are strongly supported by the committee. While they are funded separately from the wwPDB, the tightly integrated awareness and participation by wwPDB partners will insure cooperation and parallel development for the archive.

(5) D&A V2.0 Status (reported by Stephen Burley)

A significant portion of the meeting revolved around the status of Version 2.0 of the D&A tool. This Version represents a significant milestone for the overall re-design and integration of the software systems used by PDB partners across the wwPDB for deposition and annotation of submissions. Version 2.0 expands the system to include the methods of NMR and EM, along with X-ray, using the new format of D&A. It represents the key stage to having a fully multi-disciplinary scheme that paves the way for the future integrative structural biology.

This stage of the D&A development was discussed in the 2014 Advisory Committee meeting, with testing anticipated in February 2015. Regrettably, this timeline was not met, nor were milestones for test evaluations met throughout 2015. At the time of the present AC meeting, Version 2.0 was still not ready for production. Hurdles were presented regarding progression from V1.5 to V2.0 that prevented operation with NMR and EM data. There are also difficulties within the current design scheme that limit the ability to have functional user deposition operation on the PDB site, despite development of significant portions of the system at PDB. This status reveals apparent difficulties in the development plan and strain among the wwPDB partners. This situation was cause for considerable concern by the Advisory Committee. The AC recognizes that this problem will be addressed following release of V2.0 and encourages the partners to address it promptly.

These problems notwithstanding, there has been considerable progress in V2.0, particularly with deposition protocols for all EM methods and validation reports largely complete for both NMR and EM. Considerable internal testing has improved the deposition-annotation-back to deposition pipeline, which is viewed as critical to the production release of V2.0.

As a consequence of the current status, discussion prior to the AC meeting between the wwPDB leadership and the AC chair, and discussions with the full AC,

a plan was developed for completing V2.0. A complete V2.0 system is ready for internal testing at the RCSB Rutgers site, effective October 2, 2015.

The partners committed to an aggressive plan that would (i) resolve current issues, (ii) begin limited public beta testing by mid-November, 2015, (iii) assess and remediate matters arising from testing, (iv) port the system to UCSD PDB and PDBj, (v) complete the Validation Report systems for NMR and EM, and (vi) result in public release of V2.0 in January 2016. The issues that impact porting the system to EBI/PDBe will be addressed following public release and resolved as soon as possible.

The wwPDB partners recommend a joint session, following release of V2.0, to review operating procedures and cooperation between the partners. The AC strongly endorses this action and is willing to participate in this discussion.

AC Recommendations:

1. A clear development plan for V2.0 needs to be defined with accountability and periodic feedback to the Advisory Committee.
2. A single project leader should be appointed and tasked with providing reports to the AC Chair, preferably monthly.
3. The NMR VTF should be tasked with completing the Validation Report, which will enable the wwPDB to complete and have ready the validation server for implementation with V2.0.
4. Establish a renewed management plan for D&A, which may also impact interactions between all of the wwPDB partners.

(5) Future Plans and Questions from the wwPDB Partners

A meeting of the X-ray VTF 2.0 is scheduled for November, 2015. This group has set an excellent example for other VTFs, and the AC recommends that the wwPDB partners work actively to encourage and lead other VTFs to be as effective.

The AC was very enthusiastic about the forthcoming activities in 2016, which will see important milestones for the wwPDB, beginning with the phase out of legacy deposition and annotation systems for NMR and EM. The wwPDB plans to begin carbohydrate remediation, a task that has been mentioned previously by the partnership and postponed due to competing priorities. It is an important and complex task that will enhance the archive in the future.

The wwPDB will host a Symposium on Data Mining and Advanced Analytics at Rutgers. This will assist in developing the future perspectives and goals for the wwPDB, beyond archiving the structural biology data. The AC strongly supports these leadership activities by the wwPDB.

Projections by the wwPDB partnership suggest that expansion of the wwPDB into qualified sites in China and India may begin in 2017. The AC supports

evaluation of potential sites and work with interested parties in these regions. The AC recommends that this topic be discussed again at the 2016 Advisory Committee meeting, which will be hosted by BMRB in October 2016.

The Committee was asked to consider **five questions** and provide input to the wwPDB.

First, the wwPDB has proposed to add a personal identification index to each contact author by utilizing the ORCID system. **The AC supports this action and recommends that instructions and/or a link be provided to depositors to learn how to sign-up for an ORCID ID number.**

Second, the wwPDB proposes a set of Principles Guiding Federation of Data Resources with the PDB Archive. This approach begins the groundwork for interacting with external Data Resources that will support the PDB Archive. It is of particular impact for integrative and hybrid 3D structural models in the future. The AC supports this approach with the proviso that further feedback is provided to the AC for consideration of commitment of resources. **There was strong overall encouragement to follow this development.**

Third, the wwPDB AC was asked to endorse the wwPDB Partners collaboration with the Integrative/Hybrid Methods Working Group to establish a Federation of Data Resources to enable inclusion of Integrative/Hybrid 3D structural models in the PDB Archive together with certain primary data and metadata (sufficient for validation). **The AC supports the inclusion of 3D structural models in the archive with the proviso, which has been long-standing, that a sample was indeed made and used to provide some experimental data used to generate the model.** The model may go into PDB and the data should be included in a Federated Database. The AC feels that validation to a high standard remains a key requirement in this process.

Fourth, the wwPDB AC was asked to support the adoption of the Principles Guiding Versioning of PDB Archival Entries outlined in an Appendix 3, provided to the AC. The policy will enable the wwPDB to allow the original depositors of an entry to deposit a newer entry and retain the original PDB ID with a version identifier added. This will enable appropriate reference to the original citation and will retain both data sets in the Archive. **The AC supports the proposal put forward and recommends careful review of the implementation pathway being considered.**

Fifth, the wwPDB AC was asked to concur with adoption of the Principles Guiding Limited Inclusion of wwPDB Annotations by the wwPDB Partners outlined in Appendix 4.

The wwPDB raised the issue of adding annotations to an entry based on the knowledge and expertise of annotators, for example when the deposition contact author provides input that is counter to analyses by automated methods embedded in the Common Deposition and Annotation System. The most common occurrence relates to the correct annotation of Quaternary Structure. The AC was uncertain of the impact of such a policy change, based on the limited example data provided. ***The AC would be willing to review a selection of example entries, and the AC would then be able to provide a more reasoned response. In the absence of this information, the AC does not support changes to policy at this time.***

Discussion and Critique by the Advisory Committee

Overall, the Advisory Committee is extremely impressed with the wwPDB's achievements, the management of the burgeoning archive, and the impact of reaching the user-base, which is largely the non-structural biology/non-specialist community.

The achievements and metrics observed from 1.5 years of using the Common D&A System (V 1.0/1.5) for crystallography are impressive, and the improved efficiency is encouraging for the future of the PDB, particularly with other methodologies coming onboard.

The developments for EM are also very impressive and the outlook is exciting to support this community. The wwPDB partners are commended for their leadership roles and involvement in the areas of data formats and validation.

Throughout the review of wwPDB activities, the role that wwPDB partners, either acting individually or jointly, provide to the community in the form of leadership for validation tasks and community integration is highly commended and supported by the AC.

In the area of NMR issues, the AC felt that issues involving the NMR VTF and NEF groups need to be addressed by leadership and establish proper authority. These actions are necessary to establish a clear path forward. There are viable models from other VTFs to follow for more efficient progress. The proposed meeting of the NMR VTF and NEF groups in 2016, coinciding with the ICMRBS, is deemed too late to have appropriate impact on issues regarding the Validation Report and NEF. The AC encourages renewed efforts by the VTF and NEF-group to coincide with the push for release of D&A V2.0.

The principal concern arising in the AC meeting of 2015 relates to the interactive relationship between the partners, and the potential impact that this may have in the coming years, all of which was revealed in the delayed status of D&A V2.0. The AC was encouraged by the strides made immediately preceding and during the AC meeting, and the commitment to an implementation schedule for D&A V2.0 exemplified these efforts. The AC strongly supports the plan for the wwPDB partners to renew their working relationships following production release of D&A V2.0. Moving forward, the AC

recommends the development of a plan with more defined accountability and feedback to the AC. The AC further recommends assigning an overall project leader for D&A V2.0 and that this project leader provide monthly reviews to the wwPDB leadership and the AC Chair. The AC supports the firm commitment for D&A V2.0 made at the meeting. The partners are encouraged to communicate progress with the AC monthly until V2.0 is released. The AC requests that if deadlines for V2.0 are missed by 8 January 2016, then there should be a WebEx meeting of the AC with the partners to review the situation.

There was considerable discussion regarding how the AC could improve its contribution to the wwPDB. We suggest that there may be significant benefit to more frequent interaction with the project leaders and that such interaction could assist in the re-establishment of working relationships and interactions (relevant to the present D&A situation). These suggestions are made to engage the AC as a path to avoiding potential problems and/or to act as a more robust approach to conflict resolution in the future. Some elements that the wwPDB leadership may consider in formulating how the AC may be more engaged are as follows:

- a. Consider sharing the SAC reports from individual projects with the wwPDB AC prior to the annual meeting.
- b. Consider having the report of the wwPDB AC presented at the SAC of the partners.
- c. Quarterly videoconference with the wwPDB leadership and the AC Chair.