

Worldwide Protein Data Bank Advisory Committee (wwPDB-AC)
Report of October 12th 2012 Meeting
Institute for Protein Research, Osaka University, Osaka, Japan

Chair: Soichi Wakatsuki

PDB Site Representatives: Cynthia Wolberger (RCSB-PDB), Stephen K. Burley (RCSB-PDB), Wayne Hendrickson (RCSB-PDB), Michael G. Rossmann (RCSB-PDB), Titia K. Sixma (PDBe), Helen Saibil (PDBe), Gaetano Montelione (BMRB), Masatsune Kainosho (BMRB), Genji Kurisu (PDBj), and Soichi Wakatsuki (PDBj)

Ex Officio Community Stakeholder Representatives: Edward N. Baker (IUCr), R. Andrew Byrd (ICMRBS), and Keiichi Namba (Macromolecular EM)

wwPDB Members: Helen M. Berman (RCSB-PDB), Martha Quesada (RCSB-PDB), Gerard J. Kleywegt (PDBe), John Markley (BMRB), and Haruki Nakamura (PDBj)

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

wwPDB-AC E-mail Addresses:

cwolberg@jhmi.edu, StephenKBurleyMD@gmail.com, mgr@indiana.bio.purdue.edu,
wayne@convex.hhmi.columbia.edu, guy@cabm.rutgers.edu, kainosho@tmu.ac.jp,
rjr27@cam.ac.uk, h.saibil@mail.cryst.bbk.ac.uk, soichi.wakatsuki@kek.jp,
gkurisu@protein.osaka-u.ac.jp

Ex Officio Community Stakeholder E-mail Addresses:

ted.baker@auckland.ac.nz, byrdra@mail.nih.gov, keiichi@fbs.osaka-u.ac.jp

wwPDB Member E-mail Addresses:

berman@rcsb.rutgers.edu, mquesada@rcsb.rutgers.edu, 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, jpding@sibs.ac.cn

wwPDB-AC Mission Statement

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

Meeting Summary

The Worldwide Protein Data Bank Advisory Committee (wwPDB-AC) to 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 the Institute for Protein Research, Osaka University, Osaka, Japan on October 12th 2012.

The agenda included

- (1) Overview and State of the PDB;
- (2) Common Deposition and Annotation (D&A) Tool;
- (3) Format, Raw Data and Validation;
- (4) NMR-Specific Activities;
- (4) EM, SAS, Remediation;
- (5) Hybrid Methods;

- (6) Transition;
- (7) wwPDB Organizational Update; and
- (8) Discussion/Advice Requested

An Overview of the State of the wwPDB was presented by Nakamura, beginning with responses to the 2011 wwPDB-AC Recommendations. A summary of recent activities was subsequently provided by Quesada, Berman, Markley, and Kleywegt.

S. Burley was thanked for his long-time support as chair of the wwPDB-AC, after which S. Wakatsuki was introduced as interim chair.

Overview (Haruki Nakamura)

H. Nakamura provided an overview of the past year.

The wwPDB-AC commends the wwPDB partner organisations for their dedication to providing a stable, state-of-the-art biomacromolecular structure archive to the scientific community. Data depositions have steadily increased with impressive records of website access and structure downloads. The PDB has become an important archive supporting a broad array of scientific disciplines, including bioinformatics, structural biology, biochemistry, cell and molecular biology and translational medicine, as well as biologics and pharmaceutical companies and the food industry.

Nonetheless, the biggest concern of the wwPDB-AC is sustainability of the funding scheme. It is crucial to the success of each partner that the wwPDB collaboration remain strong and successful.

Renewal of the NSF grant supporting the RCSB PDB is due at the beginning of 2014. A non-competitive renewal application must be submitted for review in the spring of 2013. Funding is expected to be flat, with no increase or decrease projected. The wwPDB-AC hopes for a successful renewal of the NSF grant and will provide a strong letter of support for the upcoming application.

PDBe is funded in part by EMBL and in addition holds a number of competitive external grants (Wellcome Trust, NIH, EU, and BBSRC). During the past several years, PDBe has managed to secure more EMBL funding for key personnel thanks to the commitment of the EBI management. Unfortunately, EMBL posts are no longer permanent, and PDBe's EMBL funding is likely to decrease somewhat over the next several years. However, the leadership of PDBe is hopeful that it will be allowed to apply for another Wellcome Trust grant (probably early in 2014), which would cover a number of core staff, and also provide funds for international workshops (e.g., task-force meetings) and collaborative travel.

PDBj last year secured a new 3-year funding scheme from the Japan Science & Technology Agency, which commenced in April 2011. PDBj is now part of the National Bioscience Database Center, which will be reorganized in 2014. At present, it is not clear how this change will play out and there is a strong need to ensure that PDBj will be a central component of the future national database center. The wwPDB-AC will provide its strongest support in appropriate forms, for example a support letter or a visit to relevant Ministries, backing the international resource efforts. There has been a recent concern in the Life Science Committee of the MEXT about the lack of a strong Bioinformatics presence in Japan and there will be an effort to improve this.

The wwPDB Foundation is dedicated to outreach and education. This year it organized public lectures in Osaka on October 13 following the wwPDB-AC meeting. Haruki Nakamura, Keichi Namba (Osaka University) and Stephen Burley gave lectures, with simultaneous interpretation for Burley's lecture. The lectures were attended by 52 people. Big Pharma is the target for the next

Fundraising efforts. PDB40 funding was an international effort including the Wellcome Trust, the Japan Society for Promotion of Science, and the NIH –supporting student attendance– plus many commercial donations. The wwPDB-AC was asked to give input on strategies, which would be discussed by email/phone calls and will be provided to the wwPDB partners in due course.

It was noted that the UN and UNESCO will support celebration of the International Year of Crystallography in 2014. The resolution was introduced by the Moroccan delegation to the UN. It will be an ideal occasion for the wwPDB to raise the public knowledge of the wwPDB activities. There are no specific plans for wwPDB activities at this point, but the wwPDB-AC strongly recommends that wwPDB take some leadership and the wwPDB partners brainstorm about special activities that wwPDB can sponsor, for instance exhibitions.

The wwPDB-AC recognizes that confusion still persists regarding wwPDB *versus* the individual wwPDB partners. Clarification should be one of the primary goals of the wwPDB in the coming years. It is most welcome that the wwPDB partners have agreed to move to the common "pdb.org" website.

NMR validation will be included in the Common D&A system in 3 phases: (1) validation protocols that are already performed as part of the deposition procedure currently, (2) those which are not currently included can be added if they are considered to be reliable, informative and community-accepted standards, and (3) in a later stage, new methods can be recommended for inclusion in the D&A system for consideration as future standards.

Common Deposition and Annotation (D&A) Tool (Martha Quesada)

Concerning the Common D&A deposition system, which is nearing readiness for testing by external users, the wwPDB-AC considers that it will be important to find appropriate testers who will represent the relevant communities.

The Common D&A will also include procedures for handling ligands. Real-space R-values of ligands will be calculated and presented in the X-ray validation module of the D&A system. The validation module will be run initially in the deposition system before submission. The annotators would be aware of the deposition validation report prior to running the ligand module during annotation. Within the ligand module a 3D view of the ligand structure including the electron density map will be available.

Data Format and Validation (Gerard Kleywegt)

As reported at the 2011 wwPDB-AC meeting, the mmCIF-based PDBx format will replace the current PDB format, which is severely limited, as the main format for distribution of the archive entries. Kleywegt clarified the meaning and evolution of the term/format PDBx with a proposed timeline for the transition. The D&A tool supports upload of PDB-format files and PDB entries will continue to be available in a best-effort PDB format.

The wwPDB-AC was briefed on the outcome of discussions regarding raw data archiving within the IUCr working group (John Westbrook participated as a representative of the wwPDB). Based on cost and resource estimates, this measure does not appear to be feasible and is not recommended. Archiving of unmerged intensities, however, will be considered in the future. Further discussion with IUCr to consider this requirement is recommended.

The distributions from which the percentile scores in the validation reports are calculated will be repeated annually, which means that scores of some older structures will slowly deteriorate as overall average quality improves.

NMR (John Markley)

The Common D&A user interface will process NMR data files with accompanying atomic coordinates, and “NMR-only” components will also be incorporated into BMRB. The Common D&A tool will not accept NMR data in the absence of atomic coordinates – instead, depositors and files will be redirected to the BMRB.

Concerning data from hybrid methods, it will make sense to have a common entry point. There are two different concepts of hybrid methods: (A) one set of coordinates from multiple methods such as X-ray and neutron diffraction data, and (B) two or more different kinds of data for one common macromolecular structure (e.g., X-ray crystallographic coordinates of domain structures with a SAXS envelope and NMR chemical shift data). Currently, X-ray/neutron hybrid structures are handled as a single entry. The Common D&A system has been designed with the expectation that additional protocols will be incorporated to handle hybrid methods. The wwPDB-AC recommends that discussions with the wider community continue to reach consensus on how to best archive complex hybrid data and to define appropriate validation methods.

The potential impact on NMR data archiving and its remediation were discussed in light of the anticipated cessation of funding for CCPN. BMRB has worked with CCPN to be able to import data into BMRB in the CCPN format. Ideally, CCPN will apply for and receive funding from new sources, but this outcome is by no means certain. BMRB remains able to accept NMR data from other common platforms, including SPARKY, NMRPIPE, and NMRView.

The wwPDB-AC welcomes the news that the leadership of BMRB has met with NIGMS, and that Ward Smith (NIGMS) encouraged BMRB to submit an RO1 application to continue support for its activities beyond 2014. While there is no specific RFA (Request for Applications) for the kinds of tasks that BMRB will propose, the success of a recent EMDB grant represents an encouraging precedent. It will require a highly significant, innovative research plan. BMRB has the opportunity to develop integration/validation of NMR data with other methods such as SAXS.

Broad exploration of other BMRB funding possibilities is strongly encouraged, for example, consultation with NSF Biological Sciences Advisory Committee on the possibility of applying to the BIGDATA program of NSF (http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504767). The issue with wwPDB eligibility for BIGDATA funding needs to be clarified as soon as possible.

3DEM-related issues (Helen Berman)

3DEM data is archived in two related resources: 3D volume data in EMDB and atomistic models in the PDB. EMDB is managed by the EMDataBank partners (analogous to wwPDB) and has its own advisory committee and validation task force, both made up of international experts.

In the past year, the ftp archive of EMDB has been placed inside the same ftp tree as the PDB ftp archive, to make retrieval of all 3D biomacromolecular structure data easier for users. In addition, many procedures have been harmonised and work is on-going to map the EMDB and PDB data models to each other. As the new D&A tool will handle depositions of both PDB and EMDB data, PDBj has been invited to start annotating EMDB data as well. PDBj has agreed to do this and the

PDBj annotation staff has been trained in the curation of 3DEM volume data. PDBj is expected to start accepting and processing EMDB depositions in 2013.

The wwPDB-AC feels that EMDB has somewhat low visibility on some of the wwPDB partner websites and recommends strongly that this issue be addressed in the immediate future.

The wwPDB-AC understands that the 3DEM validation functionality (of maps, models and their mutual fit) will evolve over the next 5 years from current visual “sanity checks” into EM VTF-recommended tests for depositors to report on the plausibility of their data, and to evaluate the fit of structure models and maps. A proposal to NIH by EMDatabank for developing validation methods received an excellent rating and is likely to be funded. The proposal has come up with a procedure for using selected raw data sets for developing methods for map validation.

Small Angle Scattering (SAS) Task Force

The wwPDB SAS Task Force met in July 2012 with Jill Trehwella as chair. A manuscript containing recommendations is being prepared for submission to the journal Structure. There was general agreement that if models that are derived in part from SAS data are deposited in the PDB, those data should be stored in the PDB.

Hybrid Methods and Remediation (Helen Berman)

The wwPDB-AC is pleased to see continued collective efforts by the various wwPDB task forces for various disciplines as it is important to share knowledge and to facilitate discussions on the deposition, description, and validation of data and models from hybrid methods. One or more hybrid methods task forces should be convened as early as possible.

The Common D&A tool is based on the PDBx format. Improvements in efficiency will occur once all of the refinement packages produce processed data in the PDBx format.

The wwPDB-AC recognizes that annotators are committed to testing the Common D&A tool and will postpone significant remediation efforts until the testing is complete. There was a discussion on the way remediation can be carried out. The Committee now appreciates the importance of carrying out both forward and backward remediation in order to keep the resulting archive searchable.

The Committee recognizes the importance of remediation for carbohydrates and post-translational modifications, but cautions that it must be done very carefully in relation to the overall wwPDB activities, and in particular recommends that rules and strategy for remediation processes be developed in 2013, and actual remediation processes not take place until 2014, following implementation of the Common D&A tool.

Concerning interactions with the glycomics community, it was reported that RCSB PDB staff are collaborating with them, through various routes including attending their meetings. There appears to be some room for further strengthening of these interactions based on some of the committee members' experiences with expert groups in glyco-bioinformatics.

At the last wwPDB-AC meeting, questions were raised regarding the status of the Common D&A timeline and whether the wwPDB partners would manage to meet the challenges of accelerating the development for timely delivery. In response, Martha Quesada, the project manager for the D&A project, clarified that the challenge to the team at the last wwPDB-AC meeting was that the

project be substantially complete by this fall. In fact, the annotation pipeline is ready for internal testing and the final stages of refinement and debugging. The D&A team expects the deposition pipeline to be completed shortly. The immediate goal is to begin external testing in mid to late first quarter of 2013. NMR and 3DEM will follow with the expectation that V1.0 beta pipelines will be in place by the first quarter of 2013 for internal testing.

New wwPDB-AC Terms of Reference

New terms of reference were presented by the wwPDB leadership. The wwPDB partners proposed a scheme in which they would solicit nominations of new representatives from each of the communities from which they will select future members of the wwPDB-AC. The current wwPDB-AC supports the idea of soliciting community nominations. An important question to address is what kind of experimental methods will be supported in the future. The user communities played important roles in deciding what data should be collected in the initial phase of the PDB forty years ago when there was strong pressure from the structural biologists to put crystallographic data into the PDB. It is anticipated that the wwPDB will see a continuing expansion of experimental methods and data relevant for inclusion in the archive.

Under the new wwPDB-AC terms of reference, each of the Heads of the partner organizations will nominate two members who have a deep knowledge of the data. How does one select a fair representation of the broader user community? There is some concern with opening the wwPDB-AC up to the larger community since expanding the wwPDB-AC would be a challenge both in terms of the size of the group engaged in in-depth technical discussions, and the cost of wwPDB-AC operation.

Another question is which organizations are to be represented. In this regard, it is important to maintain strong ties with the IUCr as an external organization that has played a role in setting standards and helps to lobby for the wwPDB. There is a concern that while X-ray remains the dominant discipline and has a clear society to represent it, the challenge is how the wwPDB represents the other experimental methods that do not possess similarly strong advocacy organizations yet. Overall, the wwPDB-AC recommends a flexible blend of having both community representatives nominated by organizations and those recommended by the wwPDB partners. Also it might be worthwhile considering inviting observers (guests) recommended by other organizations for particular topics of interest.

Specific points on the proposed new wwPDB Charter

Section 4.3 - There is a need to clarify the second sentence: “one member of the wwPDB to be responsible for maintaining the master copy of the archive”.

Section 4.7 - The second sentence may be too weak – the wwPDB-AC recommends to rephrase it to “members agree to avoid”.

Finally, the draft charter lacks a clause specifying that the agreement will survive the cessation of operations of a signatory partner organization.

Transition strategy

Validation PDF: Preliminary validation report examples were presented to the Committee. Some detailed discussions followed. For example, while clustering outliers would be useful, it would be very difficult for complex structures to prepare reports in an easily understandable form. There was a strong suggestion of improving the visual indication of electron density issues – red dots

indicating problems with the electron density of residues in contact with ligands are not clear enough on the charts. This is the most important piece of information for reviewers and needs to be expanded, with short footnotes given on the same page of the reports. Indeed, there was a strong preference among the wwPDB-AC members for a simplified legend on each page to explain the colors on the graphics.

The wwPDB-AC recognizes that the Common D&A tool project has made significant progress from the last AC meeting, particularly in the last 6 months judging from the project report and video demonstration presented during the meeting. The Committee commends the extensive collaborative work by the D&A team, and suggests at the same time that the team should have tighter timelines governing the project completion and implementation. There was clear consensus among the wwPDB-AC members that the wwPDB partners prioritize the release of the Common D&A tool over other efforts, and that necessary tasks be streamlined as much as possible:

1. First complete X-ray and NMR phase one of the validation proposal, etc. Pay attention to the progress of the other developers of refinement software, which programs are compliant with the PDBx format.
2. Complete internal and preliminary external testing for the X-ray components as soon as possible, but no later than six months from now.
3. Inform the community of the impending changes as soon as possible – the sooner and the more often it is communicated the better. To this end, the wwPDB PIs will produce a paper that describes all the changes surrounding the PDB over the next few years.

In addition, launching a new wwPDB website should be scheduled in a timely manner, ideally prior to the Common D&A tool release, although the wwPDB-AC recognizes the difficulty in doing so since common deposition tools must be ready before then.

General discussions

The wwPDB-AC wanted to know what role the wwPDB plans to have in the various re-refinement activities such as PDBREDO. For example, it will be important for the community to know how the wwPDB will deal with the re-refinement of the archived data. Since there are a number of groups who perform the task in different ways, the community needs to agree on the best way and periodicity of re-refinement of the archived records. Some of the automated re-refinement programs have problems with the ligands and binding sites. No matter how and what will be done in terms of re-refinement processes, the original deposition related to the original structure should not be changed. Some on the Committee felt that if the wwPDB partners were to re-refine a structure, such results would need to be subjected to peer review and published.

The wwPDB partners explained that this concerns new efforts to routinely re-refine the entire archive; however, the results of various programs to date are inconsistent. Validation tools could be used to assess the re-refinements. Currently, large-scale re-refinement activities are performed by Phenix, PDBREDO, Global Phasing and others on a weekly basis. Ligands are a particular issue in these exercises.

The wwPDB-AC would like to see evolution and convergence of the well-tested methods for re-refinement procedures. Consensus will need to be reached on the method for re-refinement, and a new re-refinement task force might be formed in future to guide a process involving

well-controlled re-refinement procedure(s) administered by the wwPDB partners. This issue should be discussed again at the next wwPDB-AC meeting.

Final remarks by the wwPDB-AC

The Committee makes the following recommendations:

1. wwPDB Common D&A – Maintain momentum and establish clear timelines. Report interim progress to the Committee in 6 months. This project should be accorded highest priority among the four transition activities.
2. Format PDBx – Move fast with this effort to maintain momentum.
3. Hybrid Methods Task Force – Form as soon as possible.
4. Joint wwPDB website – Keep it simple and move quickly.
5. Make EMDB more visible on some of the wwPDB partner websites.
6. Archive Remediation – Prepare implementation of new standards for carbohydrates and post-translational modifications.
7. wwPDB Charter – Survivability clause should be included.
8. Reassure the depositor community that original depositions will be kept in the archive while new revisions are clearly marked.
9. The wwPDB-AC requests that presentation files be provided to the Committee a week or two prior to wwPDB-AC meetings in the future so that the Committee members will have more time to review the materials and could be more helpful in providing advice.

Appendix: PDB Metrics

In aggregate, 9250 (9768*) depositions were processed between January 1st and December 31st 2011, with a two-week average turnaround (*2012 projection).

Breakdown of depositions by discipline in calendar year 2011:

X-ray:	8550	(92%, up from 8186 in 2010)
NMR:	582	(6%, down from 604 in 2010)
3DEM:	89	(1%, up from 73 in 2010)
Other:	29	(<1%)

Breakdown of depositions by wwPDB processing site in calendar year 2011:

RCSB-PDB:	5938	(64%)
PDBj:	1816	(20%)
PDBe:	1496	(16%)

Breakdown of depositors by location in calendar year 2011:

North America	46%
Europe	27%
Asia	17%
Industry	7%
Australasia	2%
South America	<1%

Africa

<1%