

# world wide Protein Data Bank Advisory Committee (wwPDBAC) Report of August 30<sup>th</sup> Meeting Firenze, Italy

**Chair:** Stephen K. Burley (RCSB)

**PDB Site Representatives:** Wayne A. Hendrickson (RCSB), Gerard Kleywegt (MSD), Brian Matthews (RCSB, excused), James H. Naismith (MSD, excused), Kei Yura (PDBj), Soichi Wakatsuki (PDBj)

**Ex Officio Community Stakeholder Representatives:** Edward N. Baker (IUCr), R. Andrew Byrd (ICMRBS), Jose-Maria Carazo (Macromolecular EM)

**PDB Site Leaders:** Helen M. Berman (RCSB), Kim Henrick (MSD), Haruki Nakamura (PDBj)

**Funding Agency Representatives:** Toshiyuki Koike (JST)

**wwPDBAC E-mail Addresses:**

[sburley@stromix.com](mailto:sburley@stromix.com); [wayne@convex.hhmi.columbia.edu](mailto:wayne@convex.hhmi.columbia.edu); [soichi.wakatsuki@kek.jp](mailto:soichi.wakatsuki@kek.jp);  
[gerard@xray.bmc.uu.se](mailto:gerard@xray.bmc.uu.se); [naismith@st-and.ac.uk](mailto:naismith@st-and.ac.uk); [brian@uoxray.uoregon.edu](mailto:brian@uoxray.uoregon.edu);

**Ex Officio Community Stakeholder E-mail Addresses:**

[ted.baker@auckland.ac.nz](mailto:ted.baker@auckland.ac.nz); [rabyrd@ncifcrf.gov](mailto:rabyrd@ncifcrf.gov); [carazo@cnb.uam.es](mailto:carazo@cnb.uam.es);

**Funding Agency Representative E-mail Addresses:**

[koike@jst.go.jp](mailto:koike@jst.go.jp);

**PDB Site Leadership E-mail Addresses:**

[berman@PDB.rutgers.edu](mailto:berman@PDB.rutgers.edu); [henrick@ebi.ac.uk](mailto:henrick@ebi.ac.uk); [harukin@protein.osaka-u.ac.jp](mailto:harukin@protein.osaka-u.ac.jp);

## wwPDBAC 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 world wide Protein Data Bank Advisory Committee (wwPDBAC) to the leadership of the Research Collaboratory for Structural Bioinformatics (RCSB), the Macromolecular Structure Database (MSD), and Protein Data Bank Japan (PDBj) met in Florence, Italy on August 30<sup>th</sup> 2005. The agenda included

- (1) Progress Report on wwPDB Activities;
- (2) PDB Data Uniformity Project;
- (3) wwPDB and NMR;
- (4) wwPDB and EM;
- (5) Models in the PDB;
- (6) Impact of Increasing Depositions;
- (7) wwPDBAC Advocacy with Funding Agencies; and
- (8) Other matters.

**The Committee considered various issues and provides the following unanimous commentary and recommendations:**

### Report on wwPDB Activities

### Commentary:

The three wwPDB member organizations are working well together, as evidenced by important achievements on many fronts of common interest. The Committee continues to be impressed by the high level of cohesion and the frequency/professionalism of wwPDB outreach activities.

### **PDB Data Uniformity Project**

#### Commentary:

The Committee was encouraged by recent progress towards the goal of establishing a fully-remediated PDB, while retaining archival access to original PDB entries. Moreover, the Committee concedes that the timeline for completion suggested earlier (i.e., December 31<sup>st</sup> 2005) was unrealistic, given the enormity of the task.

#### Recommendations:

- Provide access to the remediated PDB no later than December 31<sup>st</sup> 2006 in mmCIF and XML formats.
- Provide access ongoing to minimal PDB format files, including only protein name, unit cell information, and atomic coordinates.

### **wwPDB and NMR**

#### Commentary:

The Committee was impressed by recent progress in further improving integration of NMR data with the PDB. Closer cooperation with BMRB was seen as a critical next step in the evolution of the wwPDB organization.

#### Recommendations:

- Conclude formal induction of BMRB into the wwPDB as soon as possible.

### **wwPDB and EM**

#### Commentary:

The Committee was impressed by recent progress in further improving integration of EM data with the PDB. Closer cooperation between RCSB and EBI was seen as a critical next step in the evolution of the wwPDB organization.

#### Recommendations:

- The Committee concurs with plans to establish a joint RCSB/EBI EM database-data deposition to fully integrate EM maps with fitted PDB coordinates.

### **Models in the PDB**

#### Commentary:

While many Committee members continue to believe that macromolecular models without direct basis in experiment have no place in the PDB, developments over the past year make resolution of the “models question” a priority for wwPDB. The Committee concedes that its earlier recommendation to purge the PDB of models merits revisiting. In addition, the Committee recognizes that there are numerous ambiguities with respect to the way models are developed, described, and used, some of which require urgent clarification.

#### Recommendations:

- Establish the necessary software tools and hardware by which the PDB can interoperate with a comprehensive separate database or repository of macromolecular structure models.
- Use the planned November 19-20 2005 modeling workshop at RCSB with key opinion leaders in the modeling community to resolve how best to bridge between their efforts

and the PDB, while preserving the distinction between experimental structures and models with no “direct basis in experiment.”

- Use the November modeling workshop to address the issue of how best to accommodate “hybrid models,” which use experimental data such as EM, NMR, mutational analyses, etc. to predict the structure of a complex of two or more experimentally derived structures.
- Provide the workshop report, related materials, and a wwPDB recommendation on next steps to the Committee as soon as possible thereafter, no later than June 30<sup>th</sup> 2006.

### **Impact of Increasing Depositions**

#### Commentary:

The Committee recognizes that the ever increasing number of depositions represents a challenge for RCSB, the MSD, and PDBj. The Committee is impressed by recent streamlining of the deposition/annotation process to permit the same number of staff professionals to field yet more and more depositions. The Committee was disturbed to learn that some depositors are using the deposition/annotation process to validate their structures, and thereby creating considerable extra work for already overburdened annotation teams.

#### Recommendation:

- The RCSB, the MSD, and PDBj work together to further streamline the deposition/annotation process, with the goal of reducing and possibly eliminating human intervention for all but the most complicated depositions.
- The wwPDB both provide the necessary tools and expand educational efforts to help ensure that depositors completely validate their structures prior to formal deposition.

### **wwPDBAC Advocacy with Funding Agencies**

#### Commentary:

The Committee listened with concern to descriptions of funding challenges encountered by the RCSB, the MSD, and PDBj.

#### Recommendations:

- The RCSB, the MSD, and PDBj shall coordinate with the wwPDBAC to obtain formal letters of support when seeking funding for their respective operational activities.
- The wwPDB shall establish a coordinated plan to both educate and lobby funding agency representatives, involving individual Committee members as needed.

### **Other Matters**

#### Commentary:

After our meeting, the MSD representative (Kleywegt) provided a commentary of issues raised at a recent meeting of the MSD advisors. His commentary is provided as an Annex to this report.

#### Recommendations:

- As there was no discussion of these issues, the Committee has no recommendations at this time.

**wwPDB 2005 Report Annex**  
**MSD Advisory Commentary provided by Kleywegt**

**Report from the MSD-SAC meeting of February, 2005 (GJK)**

(1) The MSD-SAC discussed the wwPDB-AC suggestion concerning cross-membership of the SACs of the three wwPDB members and decided it would not implement this suggestion (apart from the current cross-membership of Prof Kaptein, who is on the SAC of both the RCSB and the MSD).

(2) With respect to the distribution of cleaned-up entries, the MSD-SAC is of the opinion that remediated entries must be made available in PDB format for the foreseeable future. In fact, the remediated file should be available as the default PDB-formatted entry, i.e., before the original, unremediated PDB-formatted file. [Note GJK: it is my understanding that wwPDB has now agreed to provide "minimal" remediated PDB-formatted files for all entries.]

(3) The MSD-SAC noted that there were problems in the exchange of files between MSD and RCSB. These had the effect that the PDB file as approved by the depositor at MSD differed from the PDB file that was generated from the mmCIF file produced by RCSB from the MSD entry. [Note GJK: it is my understanding that these problems have now been resolved.]

(4) The years of delay of NMRstar 3.0, to be defined by BMRB, is a problem for the MSD. The MSD-SAC therefore asked the wwPDB to consider this difficult situation and monitor its progress.

(5) Since the use of taxonomy strings is problematic (e.g., in the PDB there are 68 different spellings of the name of the organism commonly known as "E. coli"), the MSD-SAC suggested that the wwPDB moves to using (unequivocal) taxonomy IDs instead of taxonomy strings.

**Issues that could not be addressed during the wwPDB-AC meeting (GJK)**

(1) An operational issue: I would be interested to find out why the ftp site

is going to be reorganised such that the PDB entries will be subdivided, at the top level of the hierarchy, by experimental technique. Not only does this change cause a lot of people a lot of unnecessary work (everybody who mirrors the PDB and everybody who uses those mirrors), it also makes it impossible to "predict" where a file for a given PDB entry can be found in the directory tree (because the experimental technique cannot be derived from the PDB code).

This means that either "catch-all" directories with soft links to the real files will have to be created, or that all possible locations have to be generated and tried, or that external index files will have to be used. This causes lots of problems with existing scripts and programs. In return, I cannot see any benefits at all with the new directory structure.

(2) An important issue for the future: although with the current scheme there is room for several hundred thousand more unique PDB codes, we are eventually going to have to face the inevitability of an extension of the 4-character code. Since many programs, servers, databases and websites have the 4-character code hard-wired (often with a one-character extension to identify a particular chain), I would suggest to give developers a very generous amount of time to prepare for the new scheme (in newly developed applications) and to fix existing software etc. (at least 5 years). Hence, I think we should ask the wwPDB members to start thinking about a new scheme that is backward compatible with the old one, is practical (i.e., we should avoid identifiers that can have "any length"), and that also allows for an unambiguous referencing of chains (e.g., with some separator token).