

**world wide Protein Data Bank Advisory Committee (wwPDBAC)**  
**Report of October 27<sup>th</sup> 2006 Meeting**  
**Tokyo, Japan**

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

**PDB Site Representatives:** Wayne A. Hendrickson (RCSB, excused), Neil Isaacs (MSD), Rob Kaptein (MSD), Gerard Kleywegt (MSD, excused), Brian Matthews (RCSB), Masatsune Kainosho (BMRB), Hideo Akutsu (BMRB), 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), John Markley (BMRB), Haruki Nakamura (PDBj)

**Funding Agency Representative:** 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); [N.Isaacs@chem.gla.ac.uk](mailto:N.Isaacs@chem.gla.ac.uk); [brian@uoxray.uoregon.edu](mailto:brian@uoxray.uoregon.edu);  
[kainosho@nmr.chem.metro-u.ac.jp](mailto:kainosho@nmr.chem.metro-u.ac.jp); [akutsu@protein.osaka-u.ac.jp](mailto:akutsu@protein.osaka-u.ac.jp)

**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);  
[markley@nmrfam.wisc.edu](mailto:markley@nmrfam.wisc.edu);

**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 BioMagResBank (BMRB), the Macromolecular Structure Database (MSD), and the Protein Data Bank Japan (PDBj) met in Tokyo, Japan on October 27<sup>th</sup> 2005. The agenda included

- (1) Overview of Recent Progress;
- (2) Induction of the BMRB into the wwPDB;
- (3) Policy on Theoretical Models in the PDB;
- (4) Remediation Changes/Rollout Plans;
- (5) Inclusion of SAXS Data in the PDB;
- (6) Four Character PDB ID Code;
- (7) wwPDB Plans for 2007;
- (8) Long Term Funding/Stability; and
- (9) Other Matters.

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

**Overview of Recent Progress**

Commentary:

The three founding 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.

## **Induction of the BMRB into the wwPDB**

### Commentary:

The Committee welcomes induction of the BioMagResBank (BMRB) into the wwPDB. Dr. John Markely provided a succinct, useful summary of BMRB activities, which served to underscore the importance of this resource to the global mission of the wwPDB. It was already clear that the newly expanded wwPDB leadership team is working well together.

## **Theoretical Models in the PDB**

### Commentary:

The Committee was impressed by recent progress on this technically complicated issue and the inclusive manner with which resolution was sought. Following review of the outcome of the November 19-20 2005 modeling workshop at RCSB (Berman *et al. Structure 14*, 1211-1217), the Committee concurred with the recommendations of the "Berman report" in its entirety.

### Recommendations:

- Fully implement the recommendations of "Berman report" without delay.

## **Remediation Changes/Rollout Plans**

### Commentary:

The Committee was pleased to learn that the long anticipated goal of remediating the PDB, while retaining archival access to all original depositions, has been achieved. The Committee was also pleased to learn that the remediated database will, for the first time, conform to IUPAC nomenclature standards, especially for hydrogen atoms. This advance is particularly important for the PDB to fully serve the data archiving needs of the NMR community. Other major changes include: differentiation of RNA and DNA, removal of unusual atom names in, for example, co factors such as NAD, unique three letter codes for all modified nucleotides, and the use of ' rather than \* for atoms found in sugar moieties. There was considerable discussion regarding the impact of remediated PDB release on various super users and software resources (e.g., RASMOL, PyMOL, CCP4, Phenix). The Committee fully recognizes that release of the remediated PDB will entail some recoding of input/output routines for these resources. This inconvenience is, however, viewed as being modest compared to the enormous benefits that will accrue to PDB depositors and users alike, and should not impede definitive rollout of the remediated PDB by July 1<sup>st</sup> 2007. Committee members who are themselves responsible for managing software resources will work closely with the RCSB, the MSD, the PDBj, and the BMRB early in the rollout process to provide leadership by example to the entire PDB user community.

### Recommendations:

- Initiate individually supported rollout of the remediated PDB in mmCIF and XML formats to selected super users and software resource managers beginning no later than December 31<sup>st</sup> 2006.
- Conclude rollout of the remediated PDB in mmCIF and XML formats to all users no later than July 1<sup>st</sup> 2007.
- Provide access to PDB formatted files following the most current format.

## **Inclusion of SAXS Data in the PDB**

### Commentary:

The Committee was impressed by recent progress in handling small-angle X-ray scattering data from biological macromolecules in solution, and concurs with wwPDB leadership as to the desirability of capturing this information in the PDB.

### Recommendations:

- Work with SAXS community to create appropriate representation of these data, and circulate progress reports to the Committee as appropriate.

## **Four Character PDB ID Code**

### Commentary:

The Committee understands that the current PDB ID Code (1-9; A-Z/0-9; A-Z/0-9; A-Z/0-9 → ~400,000 unique combinations) will be exceeded in the foreseeable future, particularly with the impact of structural genomics programs worldwide. Expansion of the first character from nine possibilities to thirty

five effectively resolves the problem (>1,000,000 unique combinations), without the need for any software adjustments.

Recommendations:

- Expand the four character PDB ID Code before the number of depositions reaches 400,000.

**wwPDB Plans for 2007**

Commentary:

The Committee reviewed and endorsed plans for the wwPDB in 2007, which included the following:

- Rollout of the remediated PDB;
- Begin planning for the implementation of a uniform worldwide PDB deposition system;
- Increased exchanges among wwPDB member organizations, particularly at the annotator level;
- Implementation of systematic effort to capture co-crystal structures from industry; and
- Implementation of a uniform worldwide PDB annotation system.

**Long Term Funding/Stability**

Commentary:

The Committee was pleased to learn that the PDBj had secured significant long-term funding renewal during the past year. A major portion of the funding for the MSD will be reviewed in March of 2007. The RCSB will undergo a National Science Foundation site visit in February 2007 as a prelude to a final decision regarding the 2008 funding review. There was a full discussion of the pros and cons of open competition for funding the PDB.

Notwithstanding the recent successes of the PDBj, the Committee does have concerns regarding the paucity of support for PDB activities at these sites. Specifically, much of MSD funding is assembled from a “patchwork” of small grants and contracts, which is far from ideal and places the organization at considerable ongoing risk. PDBj funding levels for the next five years, were the same as compared to the preceding five years, despite increased demands on the organization.

The Committee learned that there is no specific funding for wwPDB activities at present. Global integration of the four wwPDB member organizations and the PDB does have real, albeit modest, costs that are currently covered from individual wwPDB partner budgets.

Recommendations:

- The RCSB, the MSD, the BMRB, and the 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.
- The wwPDB shall establish a charitable organization to serve as a conduit for receipt of both grant funding and gifts from pharmaceutical and biotechnology companies, involving individual Committee members as needed.

**Other Matters**

Commentary:

During the course of various discussions, it became clear that the definition of the purview of the PDB is not sufficiently precise at present. There was general agreement that linked chains of carbohydrates, amino acids, and nucleotides (minimum of three residues joined to one another via standard polymeric linkers) fall within core definition of an appropriate PDB entry. There was, however, no consensus on whether or not to include NMR structures of natural products falling outside this core definition. Such submissions almost invariably involve complicated covalent bonding patterns that cannot be handled by automated annotation software, thereby creating considerable burdens on the annotation staff. The issue appears to obtain only for NMR structures, because the Cambridge Crystallographic Data Center already serves as a repository for X-ray structures of such natural products.

Recommendation:

- The RCSB, the MSD, the BMRB, and PDBj shall develop and present a formal recommendation to the wwPDBAC regarding the purview of the PDB at our September 2007 meeting in Princeton, N.J.

There was discussion of the need to come up with better representations for structures with statistical disorder, and for very large structures with many subunits. Work on these representations is currently in progress by the wwPDB partners. Progress reports will be circulated to the Committee as appropriate.

There was also discussion about the amount of time necessary to secure approval for release of structures that have been ON HOLD more than one year. wwPDB partners proposed that if structures are free of problems they should be released at the end of the one year hold period; if there are problems with the structure the depositors should be notified by email and the deposition withdrawn from the PDB without prejudice. This proposal was accepted unanimously by the Committee and should be implemented forthwith.