

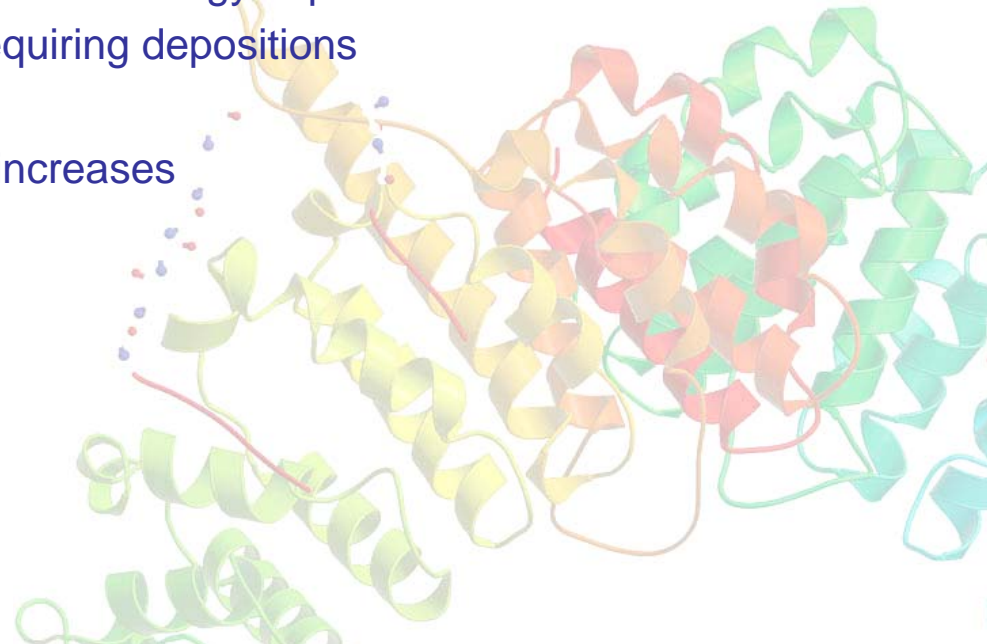
W O R L D W I D E  
 P D B  
P R O T E I N D A T A B A N K

**Worldwide Protein Data Bank**

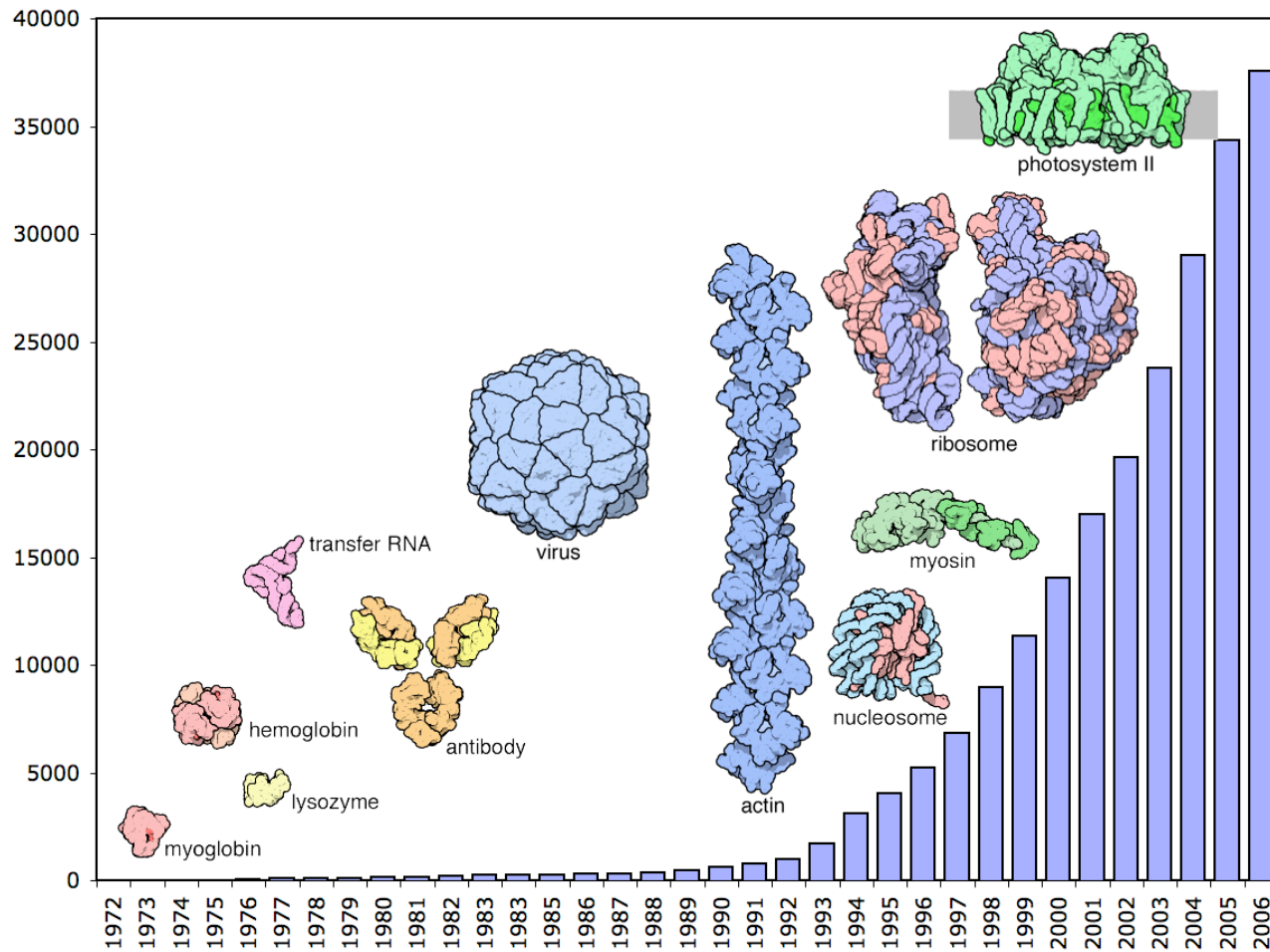
**[www.wwpdb.org](http://www.wwpdb.org)**

# History of the PDB

- *1970s*
  - Community discussions about how to establish an archive of protein structures
  - Cold Spring Harbor meeting in protein crystallography
  - PDB established at Brookhaven (October 1971; 7 structures)
- *1980s*
  - Number of structures increases as technology improves
  - Community discussions about requiring depositions
  - IUCr guidelines established
  - Number of structures deposited increases
- *1990s*
  - Structural genomics begins
  - PDB moves to RCSB
- *2000s*
  - wwPDB formed



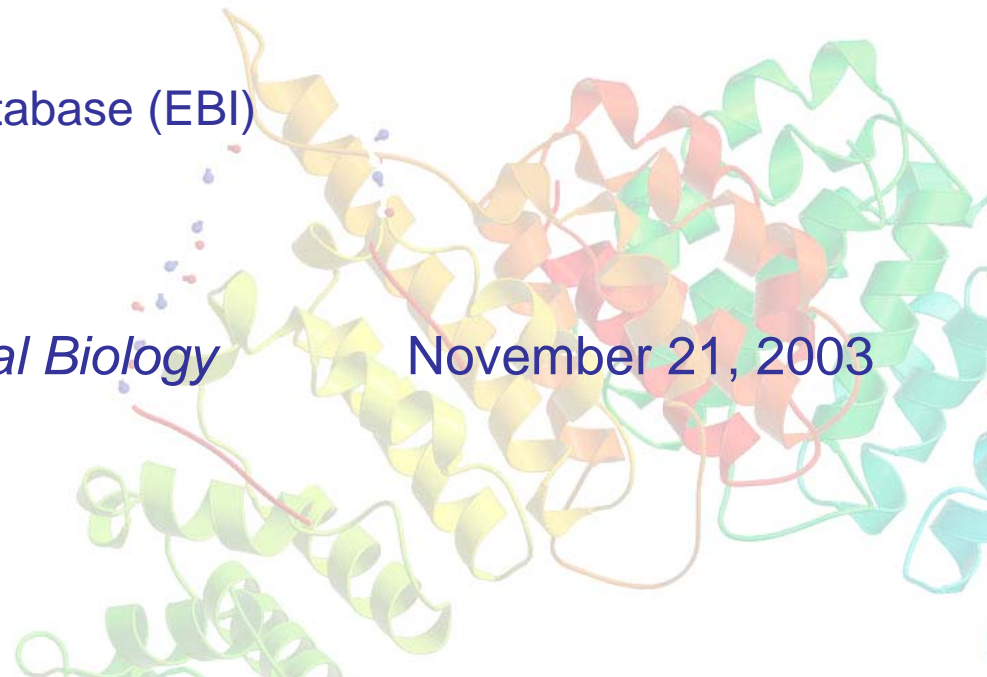
# Number of Structures Available



# wwPDB

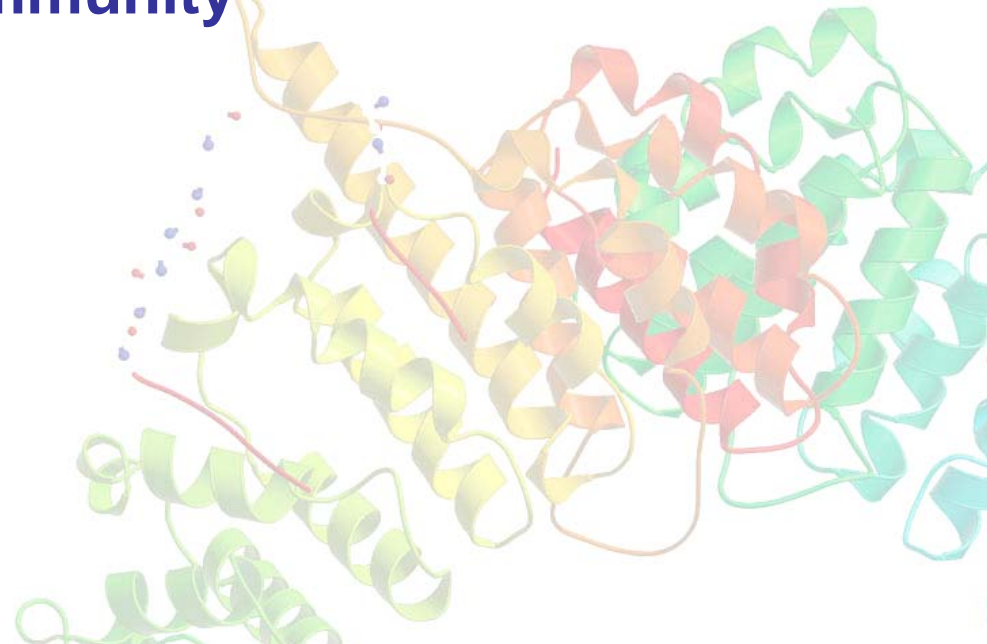
- Formalization of current working practice
- Members
  - RCSB (Research Collaboratory for Structural Bioinformatics)
  - PDBj (Osaka University)
  - Macromolecular Structure Database (EBI)
  - BioMagResBank *joined 2006*
- MOU signed July 1, 2003
- Announced in *Nature Structural Biology*

November 21, 2003



# Mission

**Maintain a single archive of macromolecular structural data that is freely and openly available to the global community**



# Guidelines and Responsibilities

- All members issue PDB ID's and serve as distribution sites for data
- One member is the archive keeper (RCSB)
  - Manage entry ID's
  - Sole write access
- All format documentation publicly available
- Strict rules for redistribution of PDB files
- All sites can create their own web sites
- International advisory board meets yearly



# Maintain Format Standards

- PDB
- PDB Exchange (mmCIF)
  - Mechanism for extension based on new demands
    - Westbrook, J., Henrick, K., Ulrich, E.L. and Berman, H.M. (2005) In Hall, S. R. and McMahon, B. (eds.), International Tables for Crystallography. Springer, Dordrecht, The Netherlands, Vol. G. Definition and exchange of crystallographic data, pp. 195-198.
- PDBML
  - Direct derivative from mmCIF
    - Westbrook, J., Ito, N., Nakamura, H., Henrick, K. and Berman, H.M. (2005) PDBML: The representation of archival macromolecular structure data in XML. *Bioinformatics*, 21, 988-992.



## Activities

- Promote standards
- Publications
- Joint exhibit stands
- Web site with pointers to member groups
- Yearly time-stamped copies of the archive
- Remediated entire archive for release in 2007





WORLDWIDE  
  
 PROTEIN DATA BANK



Welcome to the Worldwide Protein Data Bank

Home
wwPDB Agreement
Statistics
FAQ
News
Contact Us

**Access the PDB FTP:**

- [RCSB PDB](#)
- [MSD EBI](#)
- [PDBj](#)

**Deposit Data to the PDB:**

- [RCSB PDB](#)
- [MSD EBI](#)
- [PDBj](#)
- [BMRB](#)

**PDB Archive Snapshots:**

<ftp://snapshots.rcsb.org>

The Worldwide Protein Data Bank (wwPDB) consists of organizations that act as deposition, data processing and distribution centers for PDB data. The founding members are [RCSB PDB](#) (USA), [MSD-EBI](#) (Europe) and [PDBj](#) (Japan)<sup>1</sup>. The [BMRB](#) (USA) group joined the wwPDB in 2006. The mission of the wwPDB is to maintain a single Protein Data Bank Archive of macromolecular structural data that is freely and publicly available to the global community.

This site provides information about services provided by the individual member organizations and about projects undertaken by the wwPDB.

Questions? [info@wwpdb.org](mailto:info@wwpdb.org)

1. H.M. Berman, K. Henrick, H. Nakamura (2003): Announcing the worldwide Protein Data Bank. Nature Structural Biology 10 (12), p. 980









© 2006 wwPDB



## CORRESPONDENCE

### Announcing the worldwide Protein Data Bank

In recognition of the growing international and interdisciplinary nature of structural biology, three organizations have formed a collaboration to oversee the newly formed worldwide Protein Data Bank (wwPDB; <http://www.wwpdb.org/>). The Research Collaboratory for Structural Bioinformatics (RCSB), the Macromolecular Structure Database (MSD) at the European Bioinformatics Institute (EBI) and the Protein Data Bank Japan (PDBj) at the Institute for Protein Research in Osaka University will serve as custodians of the wwPDB, with the goal of maintaining a single archive of macromolecular structural data that is freely and publicly available to the global community.

The wwPDB represents a milestone in the evolution of the Protein Data Bank (PDB; <http://www.pdb.org/>)<sup>1,2</sup>, which was established in 1971 at Brookhaven National Laboratory as the sole international repository for three-dimensional structure data of biological macromolecules. Since July 1, 1999, the PDB has been managed by three member institutions of the RCSB: Rutgers, The State University of New Jersey; the San Diego Supercomputer Center at the University of California, San Diego; and the Center for Advanced Research in Biotechnology of the National Institute of Standards and Technology.

The wwPDB recognizes the importance of providing equal access to the database—both in terms of depositing and retrieving data—from different regions of the world. Therefore, the wwPDB members will continue to serve as deposition, data processing, and distribution sites. Deposition procedures will not be altered by the formation of the wwPDB; data can still be deposited using ADIT at the RCSB and PDBj or by using AutoDep at the EBI.

To ensure the consistency of PDB data, all entries will be validated and annotated following a common set of criteria. All processed data will be sent to the RCSB, which distributes the data worldwide. All format docu-

mentation will be kept publicly available and the distribution sites will mirror the PDB archive using identical contents and subdirectory structure. However, each member of the wwPDB will be able to develop its own web site, with a unique view of the primary data, providing a variety of tools and resources for the global community.

An Advisory Board consisting of appointees from the wwPDB, the International Union of Crystallography and the International Council on Magnetic Resonance in Biological Systems will provide guidance through annual meetings with the wwPDB consortium. This board is responsible for reviewing and determining policy as well as providing a forum for resolving issues related to the wwPDB. Specific details about the Advisory Board can be found in the wwPDB charter, available on the wwPDB web site.

The RCSB is the 'archive keeper' of wwPDB. It has sole write access to the PDB archive and control over directory structure and contents, as well as responsibility for distributing new PDB identifiers to all deposition sites. The PDB archive is a collection of flat files in the legacy PDB file format<sup>3</sup> and in the mmCIF<sup>4</sup> format that follows the PDB exchange dictionary (<http://deposit.pdb.org/mmcif/>). This dictionary describes the syntax and semantics of PDB data that are processed and exchanged during the process of data annotation. It was designed to provide consistency in data produced in structure laboratories, processed by the wwPDB members and used in bioinformatics applications. The PDB archive does not include the websites, browsers, software and database query engines developed by researchers worldwide.

The members of the wwPDB will jointly agree to any modifications or extensions to the PDB exchange dictionary. As data technology progresses, other data formats (such as XML) and delivery methods may be included in the official PDB archive if all the wwPDB members concur on the alteration. Any new formats will follow the naming and

description conventions of the PDB exchange dictionary. In addition, the legacy PDB format would not be modified unless there is a compelling reason for a change. Should such a situation occur, all three wwPDB members would have to agree on the changes and give the structural biology community 90 days advance notice.

The creation of the wwPDB formalizes the international character of the PDB and ensures that the archive remains single and uniform. It provides a mechanism to ensure consistent data for software developers and users worldwide. We hope that this will encourage individual creativity in developing tools for presenting structural data, which could benefit the scientific research community in general.

#### ACKNOWLEDGMENTS

The RCSB PDB is supported by funds from the National Science Foundation, the Department of Energy, and the National Institutes of Health. The MSD-EBI is supported by funds from the Wellcome Trust, the European Union (TEMBLOR, NMRQUAL, SPINE, ACTOSTRUCT, and IMS awards), CCP4, the Biotechnology and Biological Sciences Research Council (UK), the Medical Research Council (UK), and the European Molecular Biology Laboratory. PDBj is supported by grant-in-aid from the Institute for Bioinformatics Research and Development, Japan Science and Technology Corporation (BIRD-IST), and the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

**Helen Berman<sup>1</sup>, Kim Henrick<sup>2</sup> & Haruki Nakamura<sup>3</sup>**

<sup>1</sup>RCSB, Piscataway, New Jersey, Rockville, Maryland and La Jolla, California, USA.  
<sup>2</sup>MSD-EBI, Hinxton, UK. <sup>3</sup>PDBj, Institute for Protein Research, Osaka University, Osaka, Japan.

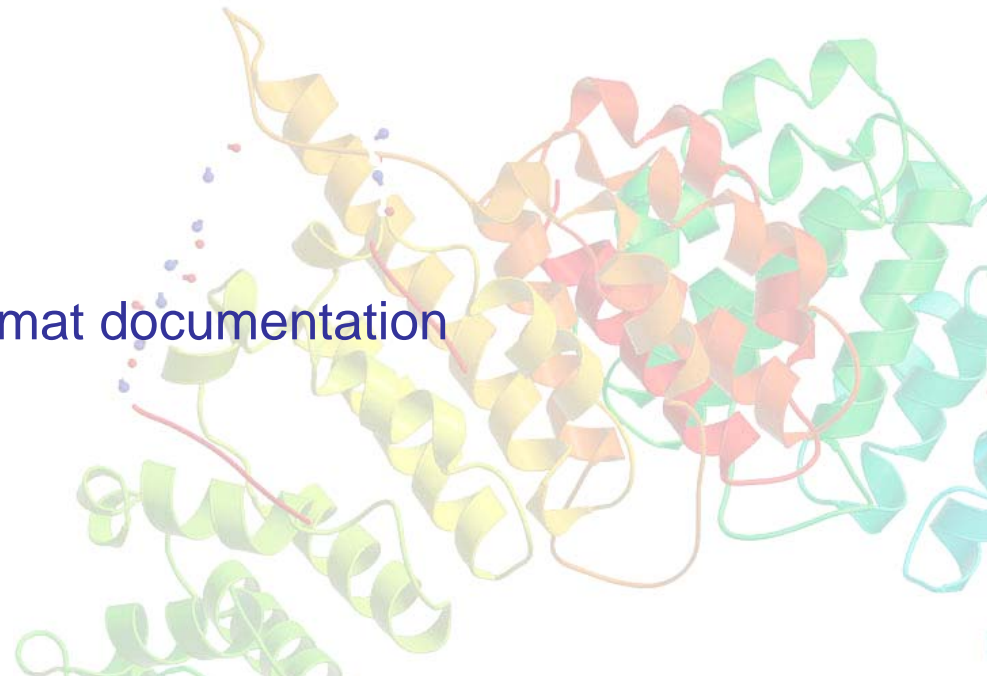
e-mail: [berman@rceb.rutgers.edu](mailto:berman@rceb.rutgers.edu)

1. Berman, H.M. et al. *Nucleic Acids Res.* **28**, 235-242 (2000).
2. Bernstein, F.C. et al. *J. Mol. Biol.* **112**, 535-542 (1977).
3. Callaway, J. et al. *Protein Data Bank Conventions Guide: Atomic coordinates entry format description*. (Brookhaven National Laboratory, 1996).
4. Bourne, P.E., Berman, H.M., Watarasigah, K., Westbrook, J.D. & Fitzgerald, P.M.D. *Methods Enzymol.* **277**, 571-590 (1997).



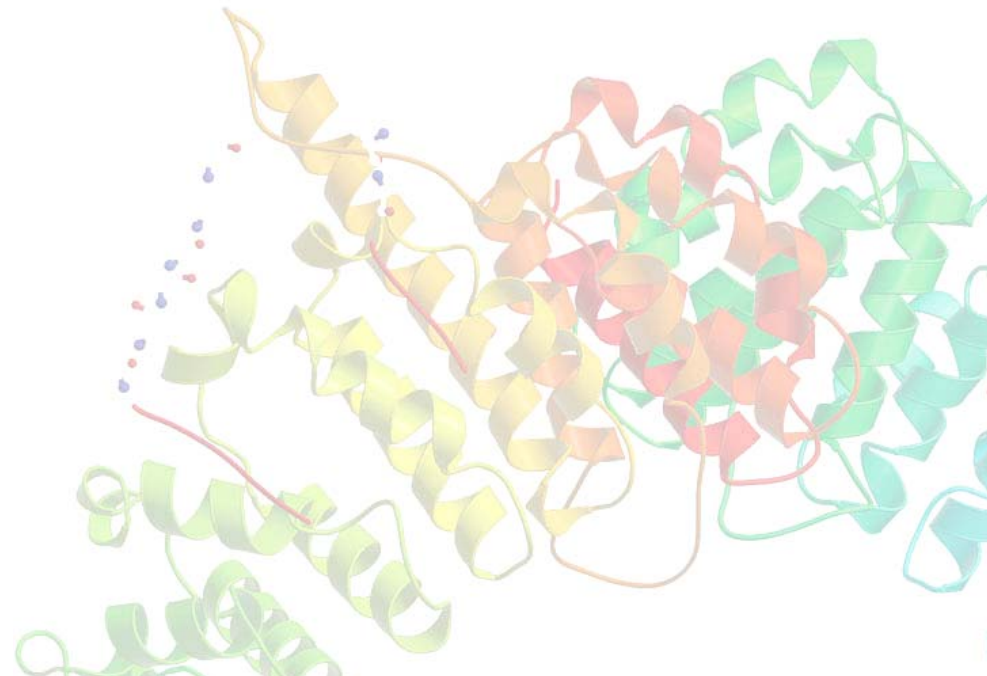
# Time-stamped Copies of the Archive

- 24 Gbytes of data in 2006 release
- Includes:
  - PDB format entries
  - mmCIF format entries
  - PDBML format entries
  - Experimental data
  - Dictionary, schema and format documentation



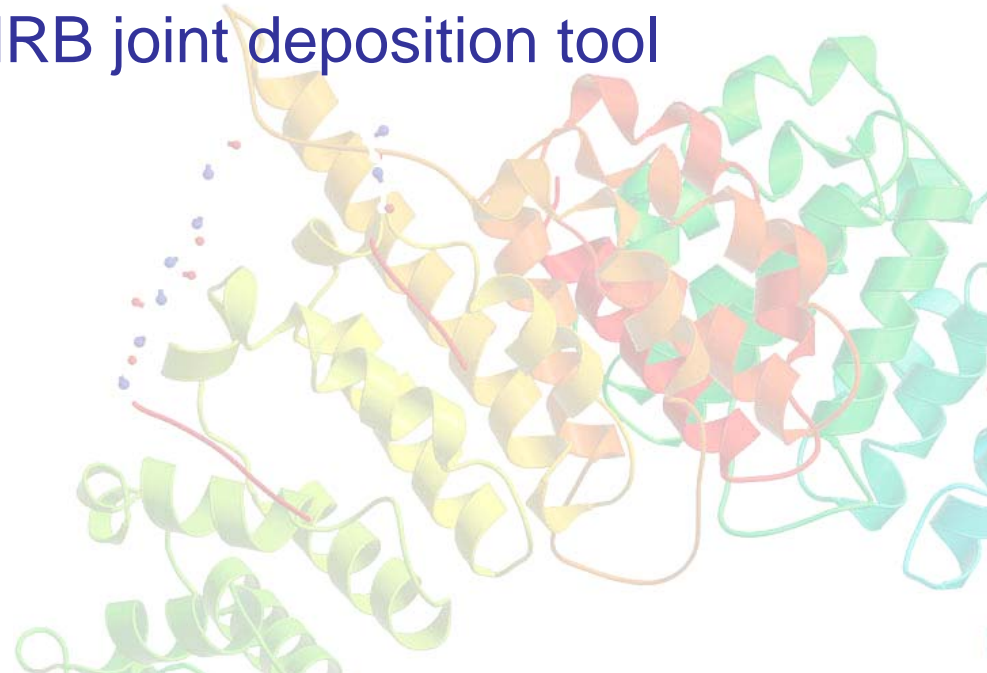
# PDB Uniformity

- **Ligands:** RCSB
- **Sequence, taxonomy, entities:** MSD
- **Citations:** PDBj



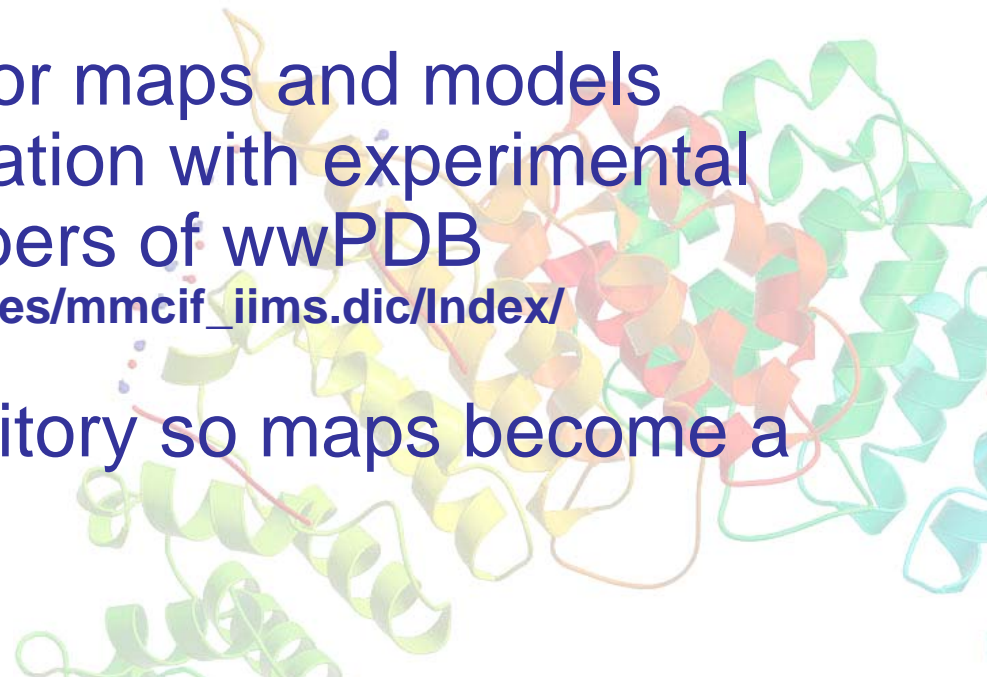
## NMR Data

- Chemical shifts and other primary experimental data deposited to BMRB; PDBj mirror
- Coordinate and meta data deposited to all wwPDB sites - RCSB/BMRB joint deposition tool



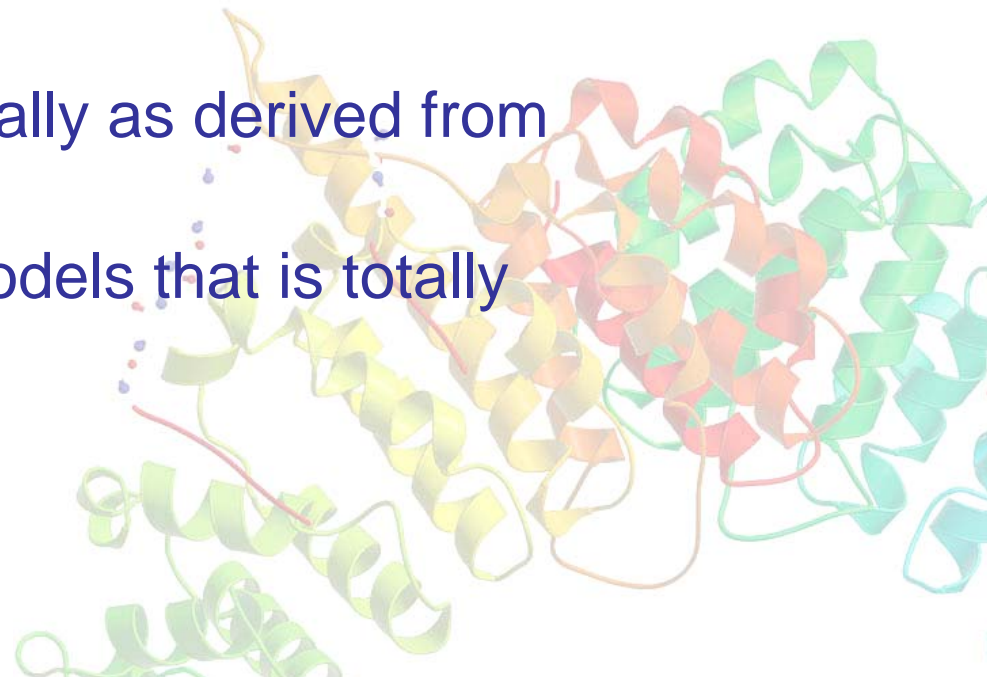
## Electron Microscopy

- Maps deposited at the EBI
- Coordinates deposited at three wwPDB sites (RCSB, MSD, PDBj)
- Joint data dictionary for maps and models developed in collaboration with experimental community and members of wwPDB  
[http://mmcif.pdb.org/dictionaries/mmcif\\_iims.dic/Index/](http://mmcif.pdb.org/dictionaries/mmcif_iims.dic/Index/)
- Plans for single repository so maps become a part of the wwPDB



# Models

- Define line between “pure” models and models based on data
- Large experimental spectrum *e.g.* X-ray, NMR, EM, SAX, FRET models
- Homology models especially as derived from structural genomics
- Need a way to archive models that is totally compatible with PDB



# Defining a policy for models

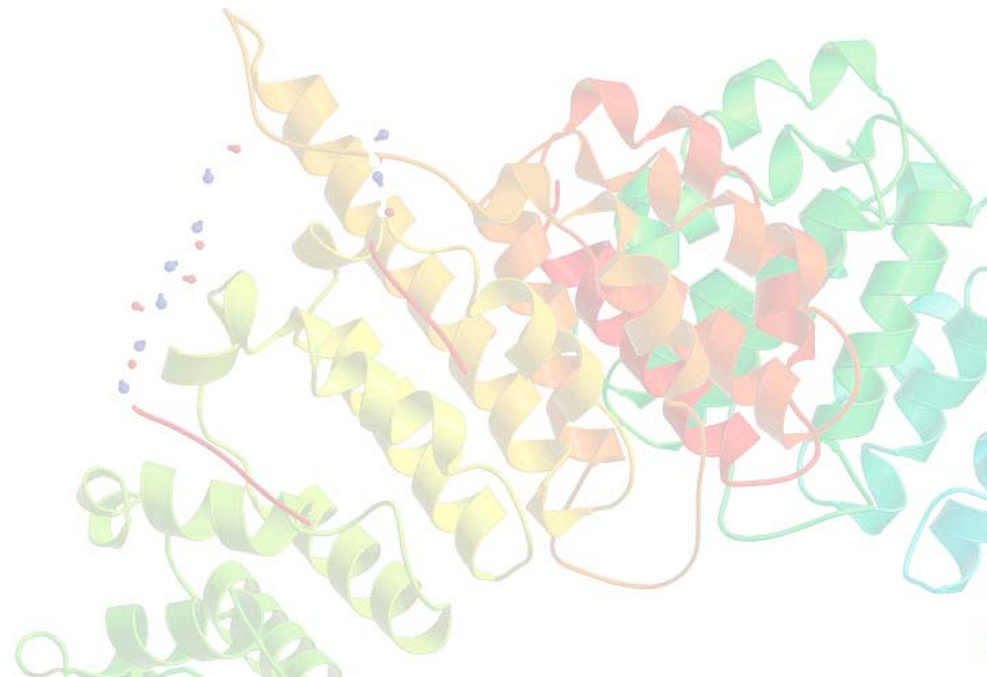
- Workshop at the RCSB PDB to develop a white paper on models (November 19-20, 2005)
- Policy published in Structure
  - Only coordinates derived from experiments on a physical sample should be in PDB
  - Implementation of this policy in progress now
  - “Outcome of a Workshop on Archiving Structural Models of Biological Macromolecules”, Helen M. Berman, Stephen K. Burley, Wah Chiu, Andrej Sali, Alexei Adzhubei, Philip E. Bourne, Stephen H. Bryant, Roland L. Dunbrack, Jr., Krzysztof Fidelis, Joachim Frank, Adam Godzik, Kim Henrick, Andrzej Joachimiak, Bernard Heymann, David Jones, John L. Markley, John Moulton, Gaetano T. Montelione, Christine Orengo, Michael G. Rossmann, Burkhard Rost, Helen Saibil, Torsten Schwede, Daron M. Standley, John D. Westbrook, Structure, 2006 14/8:1211-1217.





# Future

- More merging of data deposition and processing tools



# Acknowledgements



RCSB PDB is supported by grants from the National Science Foundation, National Institute of General Medical Sciences, the Office of Science-Department of Energy, the National Library of Medicine, the National Cancer Institute, the National Center for Research Resources, the National Institute of Biomedical Imaging and Bioengineering, the National Institute of Neurological Disorders and Stroke, and the National Institute of Diabetes & Digestive & Kidney Diseases. RCSB PDB operates out of two sites: Rutgers, The State University of New Jersey and the University of California-San Diego.



E-MSD is supported by grants from the Wellcome Trust, the EU (TEMBLOR, NMRQUAL and IIMS), CCP4, the BBSRC, the MRC and EMBL. E-MSD operates out of the European Bioinformatics Institute (EMBL-EBI), which is part of the European Molecular Biology Laboratory (EMBL).



PDBj is supported by grant-in-aid from the Institute for Bioinformatics Research and Development, Japan Science and Technology Agency (BIRD-JST), and the Ministry of Education, Culture, Sports, Science and Technology (MEXT). PDBj operates out of the Protein Research Institute at Osaka University.



The BMRB is supported by NIH grant LM05799 from the National Library of Medicine. The BMRB operates out of the University of Wisconsin-Madison.