

24. CRYSTALLOGRAPHIC DATABASES

Table 24.5.9.1. *PDB information sources*

Source	Information content
http://www.rcsb.org/pdb/ and http://www.pdb.org/	Main PDB web site
http://rutgers.rcsb.edu/pdb/ (Rutgers)	RCSB member institution PDB web sites
http://nist.rcsb.org/pdb/ (NIST)	
http://www.rcsb.org/pdb/mirrors.html	List of all RCSB PDB mirrors
http://pdb.rutgers.edu/adit/	<i>ADIT</i> web site (Rutgers)
http://pdbdep.protein.osaka-u.jp/adit/	<i>ADIT</i> web site (Osaka University, Japan)
http://pdb.rutgers.edu/validate/	<i>ADIT</i> validation server
http://www.rcsb.org/pdb/newsletter.html	RCSB PDB newsletter
http://www.rcsb.org/pdb/linking.html	Enzyme classifications and nomenclature
http://www.rcsb.org/pdb/ftpproc.final.html	FTP mirroring information
http://www.rcsb.org/pdb/cdrom.html	CD-ROM ordering information
info@rcsb.org	General help desk
deposit@rcsb.rutgers.edu	Data processing correspondence

contents of the PDB as it existed on a specific date; preservation of software, derived data, ancillary data and all other computerized and printed information; automatic archiving of all depositions and the PDB production resource; and maintenance of the PDB correspondence archive that documents all aspects of deposition. During the transition period, all physical materials including electronic media and hard-copy materials were inventoried and stored, and are being catalogued.

24.5.6. Maintenance of the legacy of the BNL system

One of the goals of the PDB has been to provide a smooth transition from the system at BNL to the new system. Accordingly *AutoDep*, which was developed by BNL (Brookhaven National Laboratory, 1998) for data deposition, has been ported to the RCSB site and enables depositors to complete partial depositions as well as to make new depositions. In addition, the EBI accepts data using *AutoDep*. Similarly, the programs developed at BNL for data query and distribution (*PDBLite*, *3DB Browser* etc.) are being maintained by the remaining BNL-style mirrors. The RCSB provides data in a form usable by these mirrors. Finally, the style and format of the BNL ftp archive is being maintained at <ftp://bnlarchive.rcsb.org>.

Links to the PDB at BNL were automatically redirected to the RCSB after BNL closed operations on 30 June 1999 using a network redirect implemented jointly by RCSB and BNL staff. External resources linking to the PDB are advised to change any URLs from <http://www.pdb.bnl.gov> to <http://www.rcsb.org>.

24.5.7. Current developments

An important role of the PDB is to foster new standards and technologies important to researchers and educators using macromolecular structure data. To this end, the following are under development at the PDB.

The RCSB is leading the Object Management Group Life Sciences Initiative's efforts to define a CORBA interface definition for the representation of macromolecular structure data. This is a standard developed under a strict procedure to ensure maximum input by members of various academic and industrial research communities. At this stage, proposals for the interface definition, including a working prototype that uses the standard, are being accepted. For further details refer to <http://www.omg.org/cgi-bin/doc?lifesci/99-08-15>. The finalized standard interface will facilitate

the query and exchange of structural information not just at the level of complete structures, but at finer levels of detail.

As multimedia become more common, the opportunity exists to use them to deliver information on structure and function to a broad PDB user community *via* the web. To date we have developed prototype protein documentaries (Quinn, Taylor *et al.*, 1999) that explore these new media in describing structure–function relationships in proteins. It is also possible to develop educational materials that will run using a recent web browser (Quinn, Wang *et al.*, 1999).

Finally, it is recognized that structures exist both in the public and private domains. To this end we are planning on providing a subset of database tools for local use. Users will be able to load both public and proprietary data and use the same search and exploratory tools used at the PDB resources.

24.5.8. PDB advisory boards

The PDB has several advisory boards. Each member institution of the RCSB has its own local PDB Advisory Committee. Each institution is responsible for implementing the recommendations of those committees, as well as the recommendations of an international advisory board. Initially, the RCSB presented a report to the advisory board previously convened by BNL. At their recommendation, a new board has been assembled which contains previous members and new members. The goal was to have the board accurately reflect the depositor and user communities and thus include experts from many disciplines.

Serious issues of policy are referred to the major scientific societies, notably the International Union of Crystallography (IUCr). The goal is to make decisions based on input from a broad international community of experts. The IUCr maintains the mmCIF dictionary as the data standard upon which the PDB is built.

24.5.9. Further information

The PDB seeks to keep the community informed of new developments *via* weekly news updates to the web site, quarterly newsletters and an annual report. Users can request information at any time by sending an e-mail to info@rcsb.org. Finally, the pdb-l@rcsb.org listserv provides a community forum for the discussion of PDB-related issues. Changes to PDB operations that may affect the community, for example data-format changes, are posted here and users have 60 days to discuss the issue before changes are made

according to major consensus. Table 24.5.9.1 indicates how to access these resources.

24.5.10. Conclusion

These are exciting and challenging times to be responsible for the collection, curation and distribution of macromolecular structure data. Since the RCSB assumed responsibility for data deposition in February 1999, the number of depositions has averaged approximately 50 a week. However, with the advent of a number of structure genomics initiatives worldwide, this number is likely to increase. We estimate that the PDB, which at writing contains approximately 10 500 structures, could triple or quadruple in size over the next five years. This presents a challenge of timely distribution while maintaining high quality. The PDB's approach of using modern data-management practices should permit us to accommodate a large data influx.

The maintenance and further development of the PDB are community efforts. The willingness of others to share ideas, software and data provides a depth to the resource not obtainable otherwise. Some of these efforts are acknowledged below. New

input is constantly being sought and the PDB invites comments at any time by e-mail to info@rcsb.org.

Acknowledgements

The continuing support of Ken Breslauer (Rutgers), John Rumble (NIST) and Sid Karim (SDSC) is gratefully acknowledged. Current collaborators contributing to the future development of the PDB are the BioMagResBank, the Cambridge Crystallographic Data Centre, the HIV Protease Database Group, The Institute for Protein Research, Osaka University, The National Center for Biotechnology Information, the ReLiBase developers, the Swiss Institute for Bioinformatics/Glaxo and the European Bioinformatics Institute.

The cooperation of the BNL PDB staff is also gratefully acknowledged.

Parts of this chapter have appeared in *Nucleic Acids Research* (Berman *et al.*, 2000) and are reproduced here with permission of Oxford University Press.

This work is supported by grants from the National Science Foundation, the Office of Biology and Environmental Research at the Department of Energy, and two units of the National Institutes of Health: the National Institute of General Medical Sciences and the National Library of Medicine.

References

24.1

- Abola, E. E. (1994). *PDB-SHELL*. Available at ftp://pdb.bmc.uu.se/pub/databases/pdb/pdb_software/pdbshell/.
- Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987). *Protein Data Bank*. In *Crystallographic databases – information content, software systems, scientific applications*, edited by F. H. Allen, G. Bergerhoff & R. Sievers, pp. 107–132. Bonn: International Union of Crystallography.
- Abola, E. E., Sussman, J. L., Prilusky, J. & Manning, N. O. (1997). *Protein Data Bank archives of three-dimensional macromolecular structures*. *Methods Enzymol.* **277**, 556–571.
- Bairoch, A. (1994). *The ENZYME data bank*. *Nucleic Acids Res.* **22**, 3626–3627.
- Bairoch, A. & Boeckmann, B. (1994). *The SWISS-PROT protein sequence data bank: current status*. *Nucleic Acids Res.* **22**, 3578–3580.
- Baker, E. N., Blundell, T. L., Vijayan, M., Dodson, E., Dodson, G., Gilliland, G. L. & Sussman, J. L. (1996). *Crystallographic data deposition*. *Nature (London)*, **379**, 202.
- Bloom, F. E. (1998). *Policy change*. *Science*, **281**, 175.
- Cambell, P. (1998). *New policy for structure data*. *Nature (London)*, **394**, 105.
- Commission on Biological Macromolecules (2000). *Guidelines for the deposition and release of macromolecular coordinate and experimental data*. *Acta Cryst.* **D56**, 2.
- Editorial Board (1998). *New policy on release of structural coordinates*. *Proc. Natl Acad. Sci. USA*, **95**, iii.
- Jiang, J., Abola, E. & Sussman, J. L. (1999). *Deposition of structure factors at the Protein Data Bank*. *Acta Cryst.* **D55**, 4.
- Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1998). *Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody*. *Nature (London)*, **393**, 648–659.
- Lin, D., Manning, N. O., Jiang, J., Abola, E. E., Stampf, D., Prilusky, J. & Sussman, J. L. (2000). *AutoDep: a web-based system for deposition and validation of macromolecular structural information*. *Acta Cryst.* **D56**, 828–841.
- Madden, D. R., Garboczi, D. N. & Wiley, D. C. (1993). *The antigenic identity of peptide–MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2*. *Cell*, **75**, 693–708.

- Peitsch, M. C., Stampf, D. R., Wells, T. N. C. & Sussman, J. L. (1995). *The Swiss 3D-image collection and Brookhaven Protein Data Bank browser on the World-Wide Web*. *Trends Biochem. Sci.* **20**, 82–84.
- Phillips, D. C. (1971). *Protein crystallography 1971: coming of age*. *Cold Spring Harbor Symp. Quant. Biol.* pp. 589–592.
- Rizzuto, C. D., Wyatt, R., Hernandez-Ramos, N., Sun, Y., Kwong, P. D., Hendrickson, W. A. & Sodroski, J. (1998). *A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding*. *Science*, **280**, 1949–1953.
- Sayle, R. A. & Milner-White, E. J. (1995). *RASMOL: biomolecular graphics for all*. *Trends Biochem. Sci.* **20**, 374–376.
- Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). *Crystal structure of a CAP–DNA complex: the DNA is bent by 90 degrees*. *Science*, **253**, 1001–1007.
- Seavey, B. R., Farr, E. A., Westler, W. M. & Markley, J. L. (1991). *A relational database for sequence-specific protein NMR data*. *J. Biomol. Nucl. Magn. Reson.* **1**, 217–236.
- Stampf, D. R., Felder, C. E. & Sussman, J. L. (1995). *PDBBrowser – a graphics interface to the Brookhaven Protein Data Bank*. *Nature (London)*, **374**, 572–574.
- Sussman, J. L. (1997). *Bridging the gap*. *Nature Struct. Biol.* **4**, 517.
- Sussman, J. L. (1998). *Protein Data Bank deposits*. *Science*, **282**, 1991.
- Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O. & Abola, E. E. (1998). *Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules*. *Acta Cryst.* **D54**, 1078–1084.

24.2

- Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R. & Watson, D. G. (1979). *The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information*. *Acta Cryst.* **B35**, 2331–2339.
- Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R. & Schneider, B. (1992). *The Nucleic Acid Database – a comprehensive relational database of three-dimensional structures of nucleic acids*. *Biophys. J.* **63**, 751–759.