24.5. THE PROTEIN DATA BANK, 1999–

Table 24.5.3.2. *Static cross-links to other data resources currently provided by the PDB*

Resource	Information content
3Dee (Siddiqui & Barton, 1996)	Structural domain definitions
BMCD (Gilliland, 1988)	Crystallization information about biomacromolecules
CATH (Orengo *et al.*, 1997)	Protein fold classification
CE (Shindyalov & Bourne, 1998)	Complete PDB and representative structure comparison and alignments
DSSP (Kabsch & Sander, 1983)	Secondary-structure classification
Enzyme Structures Database (Laskowski & Wallace, 1998)	Enzyme classifications and nomenclature
FSSP (Holm & Sander, 1998)	Structurally similar families
GRASS (Nayal *et al.*, 1999)	Graphical representation and analysis
HSSP (Dodge *et al.*, 1998)	Homology-derived secondary structures
Image (Sühnel, 1996)	Image library of biological macromolecules
MMDB (Hogue *et al.*, 1996)	Database of three-dimensional structures
MEDLINE (National Library of Medicine, 1989)	Direct access to MEDLINE at NCBI
NDB (Berman *et al.*, 1992)	Database of three-dimensional nucleic acid structures
PDBObs (Weissig *et al.*, 1998)	Obsolete structures database
PDBSum (Laskowski *et al.*, 1997)	Summary information about protein structures
SCOP (Murzin *et al.*, 1995)	Structure classifications
STING (Neshich *et al.*, 1998)	Simultaneous display of structural and sequence information
Tops (Westhead *et al.*, 1998)	Protein structure motif comparisons topological diagrams
VAST (Gibrat *et al.*, 1996)	Vector Alignment Search Tool (NCBI)
Whatcheck (Hooft *et al.*, 1996)	Protein structure checks

24.5.4. Data distribution

Data are distributed to the community in the following ways:

(1) From primary PDB web and ftp sites at UCSD, Rutgers and NIST that are updated weekly.

(2) From complete web-based mirror sites that contain all databases, data files, documentation and query interfaces, updated weekly.

(3) From ftp-only mirror sites that contain a complete or subset copy of data files, updated at intervals defined by the mirror site. The steps necessary to create an ftp-only mirror site are described at http://www.rcsb.org/pdb/ftpproc.final.html.

(4) Quarterly CD-ROM.

Data available for distribution include PDB files, mmCIF files, derived information, structure factors, NMR restraints, documentation, data dictionaries and software.

The RCSB has been responsible for distribution of PDB data since 3 February 1999. Data are distributed once a week. New data officially become available at 1 a.m. Pacific Standard Time each Wednesday. This follows the tradition developed by BNL and has minimized the impact of the transition on existing mirror sites. Since May 1999, two ftp archives have been provided: ftp:// ftp.rcsb.org, a reorganized and more logical organization of all PDB data, software and documentation; and ftp://bnlarchive.rcsb.org, a near-identical copy of the original BNL archive which is maintained for purposes of backward compatibility. RCSB-style PDB mirrors have been established in Japan (Osaka University), Singapore (National University Hospital), Brazil (Universidade Federal de Minas Gerais Brazil) and in the UK (the Cambridge Crystallographic Data Centre). Plans call for operating mirrors in Australia, Canada, Germany and possibly India.

The first PDB CD-ROM distribution by the RCSB contained the coordinate files, experimental data, software and documentation as found in the PDB on 30 June 1999. Data are currently distributed as compressed files using the compression utility program *gzip*. Refer to http://www.rcsb.org/pdb/cdrom.html for details of how to order CD-ROM sets. There is presently no charge for this service.

24.5.5. Data archiving

The PDB is establishing a central master archiving facility. The master archive plan is based on five goals: reconstruction of the current archive in the case of a major disaster; duplication of the

Table 24.5.3.3. *Web query statistics for the primary RCSB site (www.rcsb.org)*

Month	Daily average			Monthly totals		
	Hits	Files	Sites	Kbytes	Files	Hits
August 1999	63768	47675	34928	31781561	1477927	1976818
July 1999	75693	54427	38698	35652864	1687265	2346495
June 1999	33256	27054	11586	11164410	622264	764894
May 1999	26890	22085	12405	12463441	684650	833597
April 1999	21140	17099	12261	9925351	512990	634224
March 1999	8406	6911	6292	3560629	214255	260610
February 1999	2944	2433	2246	844536	68133	82453
January 1999	1563	1353	1153	92014	35202	40641

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/	*ADIT* web site (Rutgers)
http://pdbdep.protein.osaka-u.jp/adit/	*ADIT* web site (Osaka University, Japan)
http://pdb.rutgers.edu/validate/	*ADIT* 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*, 3*DB 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 listserver 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

references