

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 <i>et al.</i> , 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 <i>et al.</i> , 1999)	Graphical representation and analysis
HSSP (Dodge <i>et al.</i> , 1998)	Homology-derived secondary structures
Image (Sühnel, 1996)	Image library of biological macromolecules
MMDB (Hogue <i>et al.</i> , 1996)	Database of three-dimensional structures
MEDLINE (National Library of Medicine, 1989)	Direct access to MEDLINE at NCBI
NDB (Berman <i>et al.</i> , 1992)	Database of three-dimensional nucleic acid structures
PDBObs (Weissig <i>et al.</i> , 1998)	Obsolete structures database
PDBSum (Laskowski <i>et al.</i> , 1997)	Summary information about protein structures
SCOP (Murzin <i>et al.</i> , 1995)	Structure classifications
STING (Neshich <i>et al.</i> , 1998)	Simultaneous display of structural and sequence information
Tops (Westhead <i>et al.</i> , 1998)	Protein structure motif comparisons topological diagrams
VAST (Gibrat <i>et al.</i> , 1996)	Vector Alignment Search Tool (NCBI)
Whatcheck (Hooft <i>et al.</i> , 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