Table 24.1.3.7. *PDB structure-factor submissions, as of November 1998*

Year	No. of X-ray structure submissions	No. of structure-factor submissions (%)
1994	804	205 (25.0)
1995	963	343 (36.0)
1996	1124	546 (49.0)
1997	1484	932 (62.8)
1998	1616	868 (53.7)
Total	5991	2894 (48.3)

member, who completes the annotations and returns the entry to the author for comment and approval. Table 24.1.3.6 summarizes the checks included in our current data-validation suite. Corrections from the author are incorporated into the entry, which is reanalysed and validated before being archived and released. Most of this work covers issues not now fully delegated to automatic software. The resulting entry, after author approval, replaces the LAYER-1 entry in the archive. We strongly believe that such thorough checking and annotation is essential for ensuring the long-term value of the data.

The PDB has long made available the experimental data that were used to determine the three-dimensional structures in the database. In recent years, more and more depositors and users of the PDB have come to appreciate the importance of reliable access to such fundamental data. The deposition of the experimental data, along with the coordinates, is essential for the following reasons.

(1) Rigorous validation of the structure-determination results can only be carried out using both atomic parameters and experimental structure-factor amplitudes.

(2) Archiving of these data will ensure their preservation and continued accessibility.

Whether or not to require that the experimental data be deposited concomitantly with the structure data has recently been hotly discussed in the scientific press (Baker *et al.*, 1996) and on the internet (*EBI/MSD Draft Consultative Document for Deposition of Structure Factors*, http://msd.ebi.ac.uk/sf/sf.html).

At present, more than 50% of the X-ray diffraction submissions are being deposited with their associated structure factors (see Table 24.1.3.7), compared with 25% four years ago. This increase is probably partly due to the ease of uploading the files *via* our web-based submission tool, *AutoDep*, which is available at the EBI (http://autodep.ebi.ac.uk). The PDB strongly encourages all researchers to deposit their structure factors at the time of coordinate submission. Furthermore, we actively encourage journals to require their submission as a prerequisite for publication [see *Protein Data Bank Quarterly Newsletter* (1996), **75**, p. 1, *What's New at the PDB* at http://www.rcsb.org/pdb/newsletter.html).

In order to facilitate the use of deposited structure factors, we at the PDB, together with a number of macromolecular crystallographers and the IUCr Working Group on Macromolecular CIF, developed a standard interchange format for structure factors [*PDB Structure Factor mmCIF* at http://ndb-mirror-2.rutgers.edu/NDB/ftp/PDB/structure_factors/cifSF_dictionary; *Protein Data Bank Quarterly Newsletter* (1995), **74**, p. 1, *What's New at the PDB* at http://www.rcsb.org/pdb/newsletter.html]. This standard is the mmCIF format, *i.e.*, the IUCr-developed macromolecular Crystallographic Information File. It was chosen for its simplicity of design and for being clearly self-defining. The format is also easy to expand as new crystallographic experimental methods or concepts are developed, by simply adding additional tokens. The entire mmCIF crystallographic dictionary (http://ndb.rutgers.edu/NDB/mmcif) has recently been ratified by the IUCr's Committee for the Maintenance of the CIF Standard (COMCIFS).

The PDB has written a program to quickly and easily convert structure factors, as output by the most frequently used crystallographic programs, into mmCIF format. This tool, which also converts binary CCP4 MTZ files, will be accessible through the *AutoDep* program following final testing. MTZ files, which are useful in individual laboratories, are not appropriate for archival purposes. This is because particular groups arbitrarily attach different labels to the MTZ columns.

During the past year, the PDB has converted virtually all the old structure-factor files to this standard format and is keeping up-to-date on all new submissions. As of November 1998, there are about 2000 structure-factor files released in structure-factor mmCIF format (Jiang *et al.*, 1999; PDB mmCIF structure-factor files can be found at ftp://ftp.rcsb.org/pub/data/structures/divided/structure_factors/), with about an additional 1300 'on hold'. The current IUCr policy states that 'The IUCr also urges crystallographers to use their influence to ensure that all journals that publish articles on macromolecular three-dimensional structure require the deposition of both atomic parameters and structure-factor amplitudes.' and 'Authors are urged to release the atomic parameters and structure-factor amplitudes immediately after the publication date. This should be the normal practice. They can, however, request a delay of up to six months in the release of the atomic parameter data and the structure-factor amplitudes.' (Commission on Biological Macromolecules, 2000). The structure factors are also available *via* 3*DB Browser* (http://pdb-browsers.ebi.ac.uk/pdb-bin/pdbmain or http://bioinfo.weizmann.ac.il:8500/oca-bin/ocamain). This can be seen on the browser's Atlas page for each structure.

The ready availability of structure-factor files in a standard format has made it possible for any scientist to validate a structure in the PDB *versus* its experimentally observed data. There are now some excellent tools available for this, such as the Uppsala Electron Density Server (http://alpha2.bmc.uu.se/valid/density/form1.html) and the program *SFCHECK* (http://www.iucr.org/iucr-top/comm/ccom/School96/pdf/sw.pdf). The PDB has also observed that one of the most popular uses for these stored structure factors is for the crystallographer who did the experiment to be able to retrieve their own misplaced data.

24.1.4. Examples of the impact of the PDB

There are numerous examples in molecular biology, medicine and drug discovery where the PDB is playing an increasingly important role, as can be seen in the many sites related to the PDB (see Table 24.1.4.1).

One key example is the impact that structural information is having on the design of new drugs to combat diseases such as AIDS. At present, the three-dimensional structures of eight HIV proteins have been determined, one of which is illustrated in Fig. 24.1.3.1. These three-dimensional structures have aided researchers in the design of several drugs that have one of these proteins as their targets. Other examples can be seen in our basic understanding of the immune system (Madden *et al.*, 1993), Fig. 24.1.4.1, and the interaction between proteins and DNA (Schultz *et al.*, 1991), Fig. 24.1.4.2.

The PDB is a major international resource used by scientists, educators and students throughout the world. During the past few years, we at the PDB, in collaboration with many others, have greatly enhanced this resource into a powerful user-friendly tool for bridging the gap between the three-dimensional structure and the genome worlds (Sussman, 1997). Some examples follow.

(1) The PDB's *AutoDep* procedure (Lin *et al.*, 2000) has made

Table 24.1.4.1. *Key web sites related to three-dimensional structures of biological macromolecules*

Description	URL
PDB home page	http://www.rcsb.org
3*DB Browser*	http://pdb-browsers.ebi.ac.uk/pdb-bin/pdbmain or http://bioinfo.weizmann.ac.il:8500/oca-bin/ocamain
SwissProt database	http://www.expasy.ch/sprot/sprot-top.html
Entrez system	http://www3.ncbi.nlm.nih.gov/Entrez/
PubMed	http://www.ncbi.nlm.nih.gov/PubMed/
SCOP	http://scop.mrc-lmb.cam.ac.uk/scop/
CATH	http://www.biochem.ucl.ac.uk/bsm/cath/
DALI	http://www2.ebi.ac.uk/dali/
Nucleic Acid Database	http://ndbserver.rutgers.edu/
BioMagResBank	http://www.bmrb.wisc.edu/
Biological Macromolecule Crystallization Database and the NASA Archive for Protein Crystal Growth Data	http://wwwbmcd.nist.gov:8080/bmcd/bmcd.html
Archive of obsolete PDB entries	http://pdbobs.sdsc.edu/PDBobs.cgi
EBI PDB submission site	http://autodep.ebi.ac.uk

deposition of structural data much easier. More importantly, the data are much richer in information content and more accurately checked before release. *AutoDep* has also made uploading coordinates, structure factors and NMR restraint files very simple for the depositors.

(2) Results of the layered-release protocol have exceeded our best expectations, with the number of new entries being requested to be 'on hold' now down to only about 20% (and still decreasing), compared with well over 75% just a year ago (Sussman, 1998).

(3) The PDB is now receiving structure factors for a very high percentage of the structures determined by X-ray crystallography (Jiang *et al.*, 1999).

(4) There is now a close interaction between the PDB and most journals relevant to structural studies to ensure coordinate deposition in the PDB (and release) as a prerequisite for acceptance of manuscripts, as seen in editorials in several prominent scientific journals (Bloom, 1998; Cambell, 1998; Editorial Board, 1998).

Numerous close interactions and/or collaborations with scientists from around the world have yielded beneficial results for the entire community. This has resulted in the PDB becoming a truly international endeavour. Some examples follow.

(1) The first remote PDB deposition site has been established in Europe at the EBI (http://autodep.ebi.ac.uk).

(2) Improvement in handling of ligands and het groups for both deposition and retrieval of information has been achieved using programs developed by M. Hendlich (University of Marburg, Germany) and the CCDC (Cambridge, England).

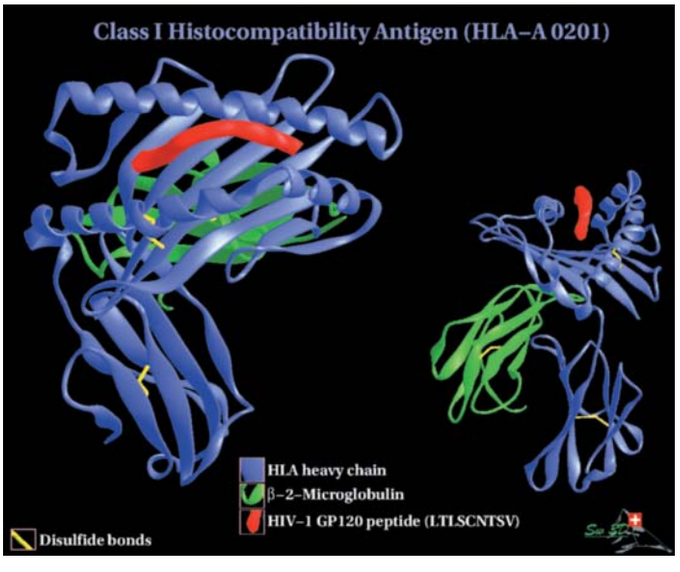

Fig. 24.1.4.1. Crystal structure of a complex of a peptide from an HIV-1 protein bound to the human class I MHC molecule HLA-A2 (Madden *et al.*, 1993), PDB ID code 1HHG, as illustrated in the SwissProt images available on the WWW (Peitsch *et al.*, 1995).

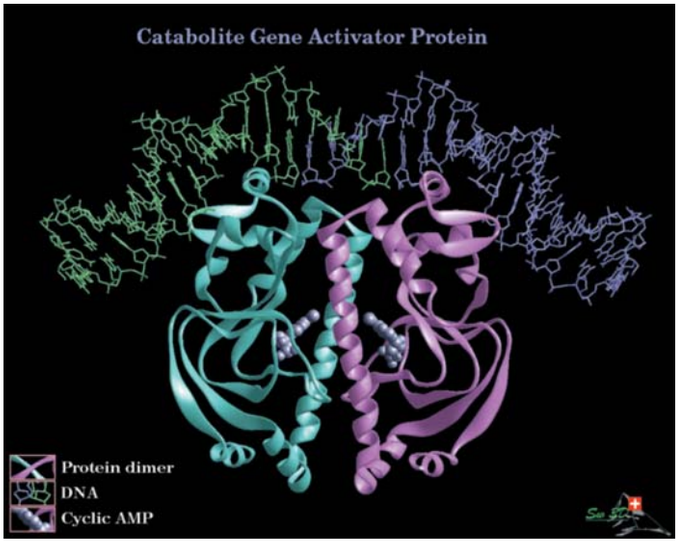

Fig. 24.1.4.2. Crystal structure, at 3 Å resolution, of the *E. coli* catabolite gene activator protein (CAP) complexed with a 30-base-pair DNA sequence. It shows that the DNA is bent by 90°. This bend results almost entirely from two 40° kinks that occur between TG/CA base pairs at positions 5 and 6 on each side of the dyad axis of the complex (Schultz *et al.*, 1991), PDB ID code 1CGP, as illustrated in the SwissProt images available on the WWW (Peitsch *et al.*, 1995).

(3) Tools to improve access and examination of three-dimensional structural information, such as *PDB Lite* and *Noncovalent Bond Finder* (E. Martz, University of Massachusetts, USA) have been developed.

(4) The user-friendly way of accessing the PDB *via* 3*DB Browser* (developed in close collaboration with Dr Jaime Prilusky, Bioinformatics Unit, Weizmann Institute of Science, Israel) has already become the standard for several online journals pointing to the PDB Atlas pages of structures.

(5) There is close interaction with the BioMagResBank (BMRB, University of Wisconsin) for the handling of NMR structural data.

(6) The fact that industrially determined three-dimensional structures are now being deposited with the PDB, even without publication, has been made possible *via* the close collaboration between the PDB and the HIV Protease Database (developed by Alexander Wlodawer at NCI, Frederick, MD, USA, and Jiri Vondrasek at IOCB, Prague, Czech Republic: see http://www.ncifcrf.gov/CRYS/HIVdb).

(7) The 17 official mirror sites in 13 countries around the world now provide easy and fast local access to the PDB web pages and database archives.

Acknowledgements

This work has been carried out by a most dedicated and talented staff at the PDB, including Frances Bernstein, Betty Deroski, Arthur Forman, Sabrina Hargrove, Mariya Kobiashvili, Pat Langdon, Michael Libeson, John McCarthy, Christine Metz, Otto Ritter, Regina Shea, Janet Sikora, Lu Sun, Subramanyam Swaminathan and Dejun Xue. In addition, John Rose (University of Georgia), Mia Raves (Utrecht University), Simone Botti, Meir Edelman, Clifford Felder, Kurt Giles, Harry Greenblatt, Gitay Kryger, Michal Harel, Marilyn Safran, Israel Silman, Vladimir Sobolev (Weizmann Institute of Science), Kim Henrick (EBI), Gert Vriend (EMBL-Heidelberg), Barry Honig (Columbia University) and Axel Brünger (Yale University) have provided invaluable support throughout the years. The PDB Advisory Board and the BNL administration together with the BNL Chemistry and Biology Departments have been an invaluable resource over the years. We wish to express our great appreciation and respect for the members of this team, who have constantly shown enormous initiative and professional capability in all their endeavours.

references