letters to the editor

Deposition of structure factors at the Protein Data Bank

Jiansheng Jiang,[a] **Enrique Abola**[a] **and Joel L. Sussman**[a,b]*†

[a]Protein Data Bank, Biology Department, Building 463, Brookhaven National Laboratory, Upton, NY 11973-5000, USA, and [b]Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

† Head, Protein Data Bank.

Correspondence e-mail: sussman@pdb.pdb.bnl.gov

Table 1
PDB structure-factor (SF) submission.

Year	Number of X-ray structure submissions	Number of SF 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)

† As of November 24 1998.

The Protein Data Bank (PDB) has long made available the experimental data which 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.

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

(ii) Archiving of this 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 been hotly discussed recently in the scientific press [Baker, Blundell, Vijayan, Dodson, Dodson, Gilliland & Sussman (1996). *Nature (London)*, **379**, 202] and on the internet (EBI/MSD Draft Consultative Document for Deposition of Structure Factors http://croma.ebi.ac.uk/msd/Policy/sf.html).

At present more than 50% of the X-ray diffraction submissions are being deposited with their associated structure factors (see Table 1), compared with 25% four years ago. This increase is probably partly due to the ease of uploading the files *via* our WWW-based submission tool, *AutoDep*, and the fact that this tool is available both in the USA at BNL (PDB deposition site at http://www.pdb.bnl.gov) and in Europe at the EBI (EBI deposition site at http://www2.ebi.ac.uk/pdb). 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 [Sussman (1996). *Protein Data Bank Quart. Newslett.* No. 75, p. 1, at ftp://pdb.pdb.bnl.gov/newsletter/newsletter96jan/newslttr.txt].

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 ftp://pdb.pdb.bnl.gov/pub/pdb/structure_factors/cifSF_dictionary; *Protein Data Bank Quart. Newslett.* No. 74, p. 1 (1995), at ftp://pdb.pdb.bnl.gov/newsletter/newsletter95oct/newslttr.tx]. 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 COMCIFS committee.

The PDB has written a program to quickly and easily convert structure factors, as output by the most frequently used crystallographic programs, into the mmCIF format. This tool, which also converts binary *CCP*4 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 ~2 000 structure-factor files released in the structure-factor mmCIF format (PDB mmCIF structure-factor files can be found at ftp://pdb.pdb.bnl.gov/pub/pdb/structure_factors/CIF_format), with an additional ~1 300 'on-hold' for up to four years according to the IUCr policy (see IUCr deposition policy at http://www.iucr.org/iucr-top/journals/acta/actad_notes.html). The structure factors are also available through the PDB's WWW-based *3DB Browser* (http://www.pdb.bnl.gov/pdb-bin/pdbmain). 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 *SFCHECK* (http://www.iucr.org/iucr-top/comm/ccom/School96/pdf/sw.pdf) and the *Uppsala Electron Density Server* (http://alpha2.bmc.uu.se/valid/density/form1.html). 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 his/her own data which have been misplaced in their laboratory.