3.6. CLASSIFICATION AND USE OF MACROMOLECULAR DATA

## A3.6.2.2. Extensions for structural genomics

An International Task Force on Deposition, Archiving, and Curation of Primary Information for Structural Genomics was formed under the auspices of the International Structural Genomics Organization (ISGO) in 2001 (Berman, 2001) and was asked to develop specifications for data from structural genomics projects to be deposited with the PDB. The recommendations from this working group are summarized at http://deposit.pdb.org/mmcif/sg-data/xstal.html and http://deposit.pdb.org/mmcif/sg-data/nmr.html. For data from crystallography-based projects, the content extensions are largely focused on a more detailed description of phasing, tracing and density modification. All of the ISGO recommendations have been incorporated into the PDB exchange dictionary.

## A3.6.2.3. Noncrystallographic methods

The IUCr-sponsored development of data dictionaries has been focused exclusively on crystallographic methods. As the repository for all three-dimensional macromolecular structure data, the PDB accepts structures determined using noncrystallographic techniques such as NMR and cryo-electron microscopy. The description of noncrystallographic methods is beyond the remit of the IUCr, so the PDB has worked with the NMR and cryo-electron microscopy communities to develop data dictionaries that describe these techniques within the mmCIF framework.

### A3.6.2.3.1. *NMR*

The PDB exchange dictionary includes a description of NMR sample preparation, structure solution methodology, refinement and refinement metrics. These extensions were developed in collaboration with the BioMagResBank (BMRB; Ulrich *et al.*, 1989). The BMRB is the archive for experimental NMR data for biological macromolecules and has played an active role in the development of the mmCIF data dictionary. In selecting a format for archiving NMR data, the BMRB opted to use the STAR syntax (Hall, 1991) rather than the more restrictive CIF syntax. Despite this difference in syntax, the conceptual representation of macromolecular structure in the NMR dictionary (NMRStar) has remained semantically very close to the mmCIF representation. This has facilitated the exchange of data and dictionaries between the BMRB and the PDB, the sharing of software tools, and the development of a common platform for depositing data.

### A3.6.2.3.2. *Cryo-electron microscopy*

Cryo-electron microscopy (as a technique for the determination of the structure of large molecular assemblies) is also described in the PDB exchange dictionary. The data extensions for cryo-electron microscopy include a description of the sample preparation, raw volume data (Henrick *et al.*, 2003), structure solution and refinement. These extensions have a prefix of `em_` (http://mmcif.pdb.org/dictionaries/mmcif_iims.dic/Index/).

### A3.6.2.3.3. *Protein production*

The International Task Force on Deposition, Archiving, and Curation of Primary Information for Structural Genomics (Section A3.6.2.2) has also provided recommendations for the deposition of information about protein production. These recommendations are summarized at http://deposit.pdb.org/mmcif/sg-data/protprod.html. These data extensions have been used as the foundation for the Protein Expression Purification and Crystallization database (PEPCdb, http://pepcdb.pdb.org/) and for the protein production process model developed to support the Structural Proteomics in Europe initiative (SPINE; http://www.spineurope.org/).

## A3.6.2.4. Supporting software

The RCSB/PDB has developed a set of software tools which support the PDB exchange dictionary framework (Chapter 5.5). These include *PDB_EXTRACT*, a tool to extract data from the output files of structure determination applications; *ADIT*, a web-based editor for data files based on the PDB exchange dictionary; and *CIFTr*, a translator from mmCIF to PDB format. These applications and other supporting utilities can be downloaded from http://sw-tools.pdb.org/.