MSWK.CF.13 CIF DATA FOR INTERNATIONAL TABLES FOR CRYSTALLOGRAPHY; VOL: A1 [SUBGROUP DATA] (TO APPEAR) AND VOL: A. Wondratschek, H.; Institut fuer Kristallographie, Universitaet, D-76128 Karlsruhe, Germany; Madariaga, G. and Aroyo, M. I., Dpto. de Fisica de la Materia Condensada, Facultad de Ciencias, Universidad del Pais Vasco, Apdo. 644, 48080 Bilbao, Spain

This contribution deals with the inclusion of:

(i) the complete information on subgroups of space groups [International Tables for Crystallography Vol. A1], (to appear)
(ii) the symmetry information of International Tables for Crystallography, Vol. A (IT A) into the CIF format.

The proposed block structure follows closely the arrangement of the symmetry information in IT A. To each subgroup a number of blocks is assigned such that each block contains some aspect of the space-group data. Each subgroup is described in one separate block. Provisions are made for a future expansion of the data, e.g. for information on supergroups of space groups which is not available at present.

For the extension of the existing CIF-core dictionary a list of data names is being proposed which refer to the different entries of IT A and the specific requirements of the subgroup tables.

The complete list of subgroups of the space groups Cc (No. 9) and P3221 (No. 154) as well as part of the data of IT A will be displayed in the proposed CIF format. The information concerning these two examples has been extracted from a CIF master file which already exists. Some simple tools for applications will be demonsated.

This version of an extension of the CIF Dictionary should be considered as a draft presented for approval to the IUCr Committee for the Maintenance of the CIF Standard. When this proposal has been accepted the data files can be supplied in due course.

MSWK.CF.14 STAR/CIF MACROMOLECULAR NMR DATA DICTIONARIES AND DATA FILE FORMATS. Eldon L. Ulrich, David Argentar, Amy Klimowicz, William M. Westler, and John L. Markley, Dept. of Biochemistry, University of Wisconsin- Madison, Madison, WI 53706

BioMagResBank (BMRB), a database for information on macromolecular structure derived from NMR spectroscopy, has begun using STAR files and STAR and CIF data dictionaries for archiving and exchanging pertinent data. So that overlapping features of NMR and crystallographic data will be represented in a common format, we have collaborated with the Protein Data Bank (PDB) and the mmCIF authors in developing the STAR and CIF data dictionaries (NMRIF dictionary). One of the challenges has been to extend existing protocols to handle the wide variety of spectral, kinetic, thermodynamic, and structural data that can be derived from NMR spectroscopy. The STAR format has lent itself to precise descriptions of the kinds of complex systems studied by NMR spectroscopy, which may consist of a single macromolecular species in a solution of particular composition, heterogeneous molecular aggregates, or molecules that undergo dynamic processes such as chemical reactions or conformational interconversions. The data in a single report frequently are derived from the results of multiple experiments where the variables typically are the chemical composition of the sample or parameters in the NMR experiments employed for data collection. Authors may wish to report the primary data in a way that captures the conditions and protocols of individual experiments. We have endeavored to develop a file format that can be easily understood and manipulated by the domain scientist (specialist in macromolecular NMR). The STAR format, in particular the "save frame" construct, has allowed us to encapsulate information pertaining to unique entities (molecules, samples, experimental procedures, sets of results, etc.) and to link related entities in a relatively efficient manner. In addition, the ability to limit the scope of a data tag within a save frame tremendously reduces the number of redundant data tags that are needed within a single file. The content of the STAR and CIF NMR data dictionaries and their use in designing the BMRB NMR data deposition form and in exporting data from the BMRB database will be described.

Supported by a grant from the National Library of Medicine LM04958- 05S3, a Dept. of Energy subcontract from the Protein Data Bank, and the graduate school of the Univ. of Wisconsin.

MSWK.CF.15 TOWARDS A REFERENCE LIBRARY OF MONOMERS. Jean Richelle and Alexei Vagin, Unit de Conformation des Macromolcules, Biologiques Universit Libre de Bruxelles, av. FD Roosevelt 50, CP 160/166, B-1050 Brussels, Belgium

When dealing with an archive composed of a large set of files of macromolecular structures, the correctness, completeness, and coherence of the information in the data files are critical. One of the roles of the mmCIF dictionary is to ensure correctness and completeness. The 'Reference Library of Monomers' is aimed at providing the description of the templates of all the monomers that can be used to guarantee the coherence of the whole set of CIFs. The Library is composed on one hand of a Dictionary, defining the categorie in CIF format, and on the other of the Reference Files, describing the monomer templates as well as parameters for energy calculation.

The talk will describe the conceptual schema of the Library; the main design decisions which lead to this schema will be discussed and the different CIF categories will be explained.

Next the contents of the Reference Files, i.e. types of monomers - amino acids, carbohydrates, ligands, etc. - and energy parameters will be reviewed.

Finally the sources of information used to build the Reference Files will be discussed. Currently the data files of the Protein DataBank (PDB) are the primary source of monomer description; the tools to extract this information will be described. Other sources will be discussed as well.