18-Data Bases and Communications

18.01 – Crystallographic Data Bases

MS-18.01.01 STRUCTURAL DATA IN A CHEMICAL CONTEXT: SEARCH AND RESEARCH USING THE CAMBRIDGE STRUCTURAL DATABASE. Frank H. Allen, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, England.

The CSD currently holds information on over 100,000 organo-carbon crystal structures. This information is not only of interest to crystallographers, but also to a very broad spectrum of chemists. Search systems have been devised that employ the graphical language of chemistry for query construction and for display of hits. The system also permits the systematic analysis of geometric structure, a process that leads to the acquisition of new structural knowledge from the data accumulated in the CSD. The study of conformational preferences, the mapping of structural interconversions and reaction pathways and their relation to the potential energy hypersurface, and the systematic study of hydrogen-bonded and non-bonded interactions will be illustrated. The possibilities for storing this derived information in a computerised knowledge base will be discussed.

MS-18.01.02 THE PROTEIN DATA BANK OF THE FUTURE[+] By T.F. Koetzle*, E.E. Abola, F.C. Bernstein, J.A. Callaway, J.J. Christian, B.R. Deroski, P.A. Esposito, A. Forman, P.A. Langdon, J.E. McCarthy, N.E. Oeder, R.K. Shea, J.G. Skora, K.E. Smith and D.R. Stampf, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973, USA.

The Brookhaven Protein Data Bank (PDB) is very much a resource in transition. In fulfilling its mission to provide the international community with access to information on 3-D structures of biological macromolecules, PDB is developing increasingly automated new procedures for data entry, validation, and distribution. PDB is also actively planning to expand its role as a center for coordination and development of the new database management, manipulation, and analysis tools required in the future if we are to effectively utilize the rapidly growing mass of information available in structural biology.

Coverage of PDB is now up-to-date and the complete collection of approximately 2,000 structures is available from Brookhaven and from affiliated centers located in Europe, Japan, and North America. Increasingly, distribution is *via* Internet. PDB also is now available on CD ROM as well as in the traditional magnetic tape formats, and the database has been incorporated into a number of commercial molecular graphics, simulation, database, and computer-assisted molecular design packages.

Important new developments now under way at PDB include:

- In the future, structure entry and validation will, as much as possible, be the responsibility of depositors. PDB will set standards for an open, extensible AUTHORIN data entry, annotation, and validation system incorporating PDB and community-supplied software modules.
- PDB has officially adopted the IUCr CIF as its future data interchange format. Support of the traditional PDB fixed field format will continue in parallel with CIF for several years.
- An on-line PDB database will be developed which will support intelligent interactive query and analysis from Brookhaven *via* Internet. This on-line system will be the primary access vehicle for the PDB of the future.

[+] We thank our many PDB consultants and collaborators for their numerous valuable contributions and the PDB International Advisory Board for their continuing advice and guidance. Financial support for PDB is provided by US DOE, NIH, NSF, affiliated PDB distribution centres, and user fees. Brookhaven National Laboratory operates under contract DE-AC02-76 CH00016 with the US DOE.

MS-18.01.03 NEW DEVELOPMENTS FOR THE INORGANIC CRYSTAL STRUCTURE DATABASE (ICSD). By M.Berndt, K.Brandenburg and G.Bergerhoff*, University of Bonn, Germany.

The Inorganic Crystal Structure Database is now continuously updated jointly by the Fachinformationszentrum Karlsruhe and the Gmelin-Institut, Frankfurt. The 35,000 datasets in the version of January 1993 represent the considerable amount of primary information available in inorganic crystal chemistry. It is now necessary to develop tools by which the relations between structures such as isotypism, subgroup-supergroup relations, etc. may be found.

For this purpose it is necessary to standardize the structures. E. Parthé and L. Gelato (J. Appl. Cryst. 20 (1987) 139-143) have shown the way and M.Berndt implemented STRUCTURE TIDY in such a way that all structures can now in principle be standardized automatically. Following this a COMPARE function puts together similar sites of isopointal structures and calculates the mean of differences between corresponding coordinate values. 0.05 seems to be a good limit to differentiate non-isotypic structures.

Of course in the case of more complicated structures it will not be easy to do such a comparison automatically. Therefore a graphics program CVIS has been designed by K.Brandenburg. For all structures in the ICSD it starts from the asymmetric dataset by applying the symmetry operators of all space group settings using an internal list of atomic distances. The wide variation of interatomic distances in inorganic chemistry can be taken into account by setting upper and lower limits to a display of distance distributions of selected atom pairs. Molecules and other discrete building units are then found and drawn automatically. The infinite lattices very common in inorganic structures can be drawn starting from one arbitrary atom of the framework which is then built up in steps by simple commands. Subsequently two structures on the display can be superimposed for comparison.

To enable chemists to use ICSD and the new programs for their daily work they are now available on CD-ROM to run on a PC. A menue-driven RETRIEVE program (M.Berndt) searches for all descriptors known from CRYSTIN in a user-friendly way. To further develop ICSD on CD-ROM to be a handy, complete and up-to-date catalogue of all (inorganic) crystal structures, users can help by transferring errors, missing structures and, - last but not least, - new structures via e-mail as CASTOR or CIF files to: csd0d@fizvax.kfk.de

MS-18.01.04 PROBLEMS IN THE MAINTENANCE OF A MATURE POWDER DIFFRACTION DATA BASE. By R. Jenkins, International Centre for Diffraction Data, Swarthmore, PA, U.S.A.

The first issue of the Powder Diffraction File (PDF) dates back to the mid 1940's. Over the intervening half a century there has been a steady and ongoing effort on the part of the International Centre for Diffraction data (ICDD) to maintain and update the PDF. Each year, new patterns are added to the file which today stands at about 60,000 entries.

While the *form* of the data in the PDF has changed little over the years, the *quality* of the data certainly has. Many of the early patterns were recorded using Debye-Scherrer film methods - often using molybdenum Kα radiation. With the development of efficient copper anode X-ray tubes, the increasing popularity of powder diffractometers in the 1950's,

