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,

18-Data Bases and Communications

and the advent of the automated powder diffractometers in the 1970's, there has been a growing need for the publication of high quality powder data. Unfortunately, little, if any, of the original experimental data are available, thus a significant task which has faced the ICDD in recent years is the re-evaluation and re-doing of the original data.

Data base maintenance is an expensive and time-consuming proposition, and the ICDD has been especially creative in its attempts to provide data worthy of use with modern instrumentation. This paper reviews the more important of these procedures, and outlines some of the work of the ICDD committees in keeping a watchful eye on the ever-changing needs of the powder diffraction community.

MS-18.01.05 SYSTEMATICS IN MATERIALS DESIGN. By John R. Rodgers*, National Research Council of Canada, Ottawa, Ont. K1A 0S2, Canada, and Pierre Villars, Intermetallics Phases Databank, 6354 Vitznau, Switzerland.

The search for new materials which have new or greatly improved properties is an intrinsic part of materials science. One aspect of such a search is to identify relations, regularities, rules, models and laws within experimentally determined data. In the field of inorganic and intermetallic materials there are over 4,000 distinct types of ordering of atoms, called structure types. Since the crystal structures are a first indicator for the existence of specific intrinsic properties they should be the first element to be analysed. One approach to such studies is the use of structure maps and Quantum Structural Diagrams (QSD). Structure maps order the vast amount of structural information within 2- and 3-dimensional plots, using a Mendeleyevian philosophy, such that materials with a given structure type cluster together. Quantum Structural Diagrams (QSD) systematize the relationship between composition and structure type and employ atomic size difference, electronegativity difference and sum valence electrons as coordinates. We will show, with examples, how such maps are used in the search for superconducting materials, quasicrystals and permanent magnets. The use of other maps showing relationships between mechanical properties and structure will be reviewed. An attempt will be made to explain if violations of structure types lead us to the discovery of new, yet unknown, exotic materials.

MS-18.01.06 NIST MATERIALS SCIENCE DATABASES. Alan D. Mighell* and Vicky Lynn Karen, Reactor Radiation Division, National Institute of Standards and Technology, Gaithersburg, MD 20899.

The NIST Crystal and Electron Diffraction Data Center builds a comprehensive database with chemical, physical, and crystallographic information on all types of well characterized substances. During the year, the database has been significantly augmented with respect to all categories of substances and now contains greater than 183,000 entries. From the central database, two distribution databases are produced: (1) NIST Crystal Data and (2) the NIST/Sandia/ICDD Electron Diffraction Database. In The Materials Science and Engineering Laboratory at NIST, two other databases are being built that have applications in the materials sciences: one contains information on phase diagrams and the other data on superconductors. The database on high-temperature superconductors will include materials specification and characterization information. It is being developed in a collaborative effort with the National Research Institute for Metals of Japan.

As a source of critically evaluated data, the databases can be used as an aid to research or as a basis for research. They play an invaluable role in materials design, characterization and phase identification. Distribution software for identification and lattice analysis is available from Vicky Lynn Karen. NIST Databases are made available to the scientific community through computer oriented modes of dissemination including: magnetic tape, on-line searching, CD-ROM, and scientific instruments.

MS-18.01.07 NEW TOOLS FOR SCIENTIFIC DATA EXCHANGE. N. Spadaccini, Department of Computer Science, University of Western Australia.

The proliferation of scientific data and related information has necessitated the development of automated procedures to speed archiving and retrieval. Many databases have invested years of effort to developing fast access algorithms designed specifically for their own format. Attempts to pool information across database systems is hampered by incompatible formats and access software.

Agreement on a single database format is not practical, however the adoption of a standardised exchange format (e.g. George et al, Protein Seq Data Anal 1, 27-39, 1987) with more straight forward conversion routines to most other database formats has been widely suggested. Many of these exchange formats are specific to particular applications or disciplines. They are generally fixed or pre-defined free formats, both of which require significant prior knowledge to facilitate data access and are inflexible to evolving data requirements.

Hall (JCICS 31, 326-333, 1991) has reported a flexible, self defining file format known as STAR. The file can be easily extended to incorporate new data storage requirements as they evolve without violating the original file structure. The data can be automatically validated against either universally defined data dictionaries (e.g. the CIF dictionary) or can be extended to user defined dictionaries. The crystallographic community has adopted the CIF format, a restricted implementation of STAR, as its basis for data exchange and retrieval. This paper reviews the underlying advantages of the STAR approach and elaborates on its structure.

Spadaccini, Hall & Hall (*Star_Base* Users Manual, 1993) have developed a query language known as *Star_Base* to facilitate data access from STAR files. *Star_Base* provides a number of options including the extraction of data based on conditional statements over a variety of user defined scopes. The language makes it possible to access data as required by the user with little prior knowledge of the file structure. Possible future extensions of *Star_Base* to a full application level language will be presented.

PS-18.01.08 PCPDFWIN - SEARCH/RETRIEVE PROGRAM FOR THE ICDD POWDER DIFFRACTION DATABASE ON CD -ROM. By R. Jenkins, F. Needham*, International Centre for Diffraction Data, Swarthmore, PA, USA; R. Garvey, North Dakota State University, USA; S. Lin, Nankai University, China.

The PCPDFWIN program uses Btree and Bitmap algorithms to search and retrieve the ICDD PDF2 (Powder Diffraction File) database on CD-ROM, which contains 150 MB of information. This information includes interplanar spacings, relative intensities, Miller indices, crystal system, space group symbol, axial lengths, interaxial angles and several chemical and physical properties. Among its many search parameters, the PCPDFWIN program includes Three Strongest Interplanar Spacings, Three Longest Interplanar Spacings, Cell Parameters, Cell Volume, and Density. In addition, it provides over 250 organic functional groups as search criteria for the organic substances in the database. A two