*Acta Cryst.* (2002). A**58** (Supplement), C367

## DEVELOPMENT OF PDBj-ML

N. Ito[1,2] K. Kobayashi[1,2] H. Sakamoto[1,2] H. Nakamura[2,3]
[1]BIRD, Japan Science and Technology Corporation [2]Osaka University Institute for Protein Research 3-2 Yamadaoka SUITA 565-0871 JAPAN
[3]RIKEN Genomic Sciences Center

A large number of structures of biological macromolecules have been accumulated in the Protein Data Bank (PDB) and the number is expected to increase even more rapidly with the advent of the structural genomics. Efficient methods of its storage and handling are, therefore, required to fully exploit the information available. In an effort to enhance the value of the database as a member of international collaboration of PDB, we are developing pdbj-ML, an XML format for PDB.
The PDBj-ML is based on the Macromolecular Crystallographic Information Format (mmCIF), which is already used in the PDB along with the traditional 'PDB format.' As an mmCIF file is a list of pairs (called data item) of 'name' and its 'value,' they are easily converted to 'tag' and 'element' that constitute elements of an XML file. Various restructuring and regrouping, however, are also introduced to make the most of the features of XML, such the hierarchical structure and attributes. The structure of PDBj-ML is described in XML Schema, rather than the conventional DTD.
In addition to the format conversion, we are complementing data missing in the current PDB files and adding new structural information such as functional sites. These should improve the quality of PDB as a database.

**Keywords: PDB XML**

*Acta Cryst.* (2002). A**58** (Supplement), C367

## NEW BRUKER NONIUS' INTEGRATED DATA COLLECTION AND PROCESSING ENVIRONMENT

J. Kaercher[1] R. W. W. Hooft[2] R. G. Lancaster[1] S. M. Leo[1] A. J. Leonard[1] M. Ruf[1]
[1]Bruker AXS Inc. 5465 East Cheryl Parkway MADISON WI 53711 USA
[2]Bruker Nonius BV, Roentgenweg 1, 2600 AV Delft, The Netherlands

Over the past two years Bruker AXS continued their efforts to integrate all steps that are involved in the determination of a single crystal X-ray structure in their data collection and data processing software PROTEUM. The goal is to provide our customers with two highly optimized tools for the small molecule and macromolecular crystallographic worlds.
Both software suites are based on a modular set of plugins. This approach, combined with a simple and easy to use graphical interface makes for two powerful crystallographic tools. They seamlessly integrate the program modules: data collection, indexing, strategy, integration, scaling, structure solution, and structure refinement. In addition to MS Windows NT/2000 as before the software now also runs under Linux. Data storage has been extended to relational databases. Thus, it is now possible to store project related information in either Oracle or PostGreSQL. As an alternative to storing data in a relational database, highly portable XML text files are also provided.

**Keywords: SOFTWARE DATA COLLECTION DATABASES**

*Acta Cryst.* (2002). A**58** (Supplement), C367

## INORGANIC CRYSTAL STRUCTURE DATABASE: NEW DEVELOPMENTS

V.L. Karen[1] M. Hellenbrandt[2]
[1]National Institute of Stadards and Technology 100 Bureau Drive, Stop 8520 GAITHERSBURG MARYLAND 20899-8520 USA
[2]Fachinformationszentrum (FIZ) Karlsruhe, D-76344 Eggenstein-Leopoldshafen, Germany

Crystallographic data models are used on a daily basis to visualize, explain, and predict the behavior of chemicals and materials. Access to reliable information on the structure of chemicals and materials helps researchers concentrate experimental work in directions that optimize the discovery process.
The Inorganic Crystal Structure Database (ICSD) contains full structural and bibliographic information for more than 55,000 structures. Over recent years, the Fachinformationszentrum (FIZ) Karlsruhe and the National Institute of Standards and Technology (NIST) have been building a modern infrastructure for the ICSD. This has included a re-design of the ICSD database structure, conversion and loading of the data into a relational database management system and designing graphical user interfaces to access the data. Features of the new PC-based graphical user interface for the ICSD include a tabular design, allowing for searches in five general categories of Chemistry, Crystal Data, Reduced Cell, and Reference Data.

**Keywords: INORGANIC CRYSTAL STRUCTURE, INORGANIC DATABASE, MATERIALS DESIGN**

*Acta Cryst.* (2002). A**58** (Supplement), C367

## CSDSymmetry: A DATABASE OF POINT GROUP AND SPACE GROUP SYMMETRY RELATIONSHIPS IN CRYSTAL STRUCTURES

W.D.S. Motherwell[1] E. Pidcock[1] J.W. Yao[1] J.C. Cole[1] F.H. Allen[1] J.A.K. Howard[2]
[1]Cambriidge Crystallographic Data Centre 12 Union Road CAMBRIDGE CB2 1EZ UK [2]Chemistry Department, University of Durham, Durham, UK. DH1 3LE

An algorithm that perceives molecular symmetry [1] has been applied to ca. 200,000 entries from the Cambridge Structural Database [2]. for each molecule the perceived point group together with crystallographic properties such as space group, occupied Wyckoff positions and number of residues have been placed in a relational database, csdsymmetry, using Microsoft Access software. Database queries can be constructed easily to find occurrences of any combination of molecular or crystallographic attributes and thereby answer questions on relative distributions. A number of queries will be presented which illustrate the scope and utility of CSDSymmetry, for example: 'Do high symmetry molecules prefer highly symmetric molecules?' It is intended that CSDSymmetry will be made available, free of charge for non-commercial research purposes, in the near future.
References
[1] F.H. Allen & O. Kennard (1993) Chem. Des. Autom. News, 8, 1, 31-37
[2] J.C. Cole, J.W. Yao, G.P. Shields, W.D.S. Motherwell, F.H. Allen, J.A.K. Howard (2000) Acta. Cryst. B57, 88-94

**Keywords: DATABASES, SYMMETRY, RELATIONSHIPS**