P24.01.01

Acta Cryst. (2008). A**64**, C626

The predicted powder diffraction database (P2D2)

Armel Le Bail
Université du Maine, Laboratoire des Oxydes et Fluorures, CNRS UMR 6010, Avenue Olivier Messiaen, Le Mans, Pays de Loire, 72085, France, E-mail：armel.le_bail@univ-lemans.fr

Massive inorganic crystal structure predictions were recently performed. For millions of virtual zeolites or related materials the composition is imposed : SiO_2 or $AlPO_4$ etc. For other predicted inorganic compounds the composition was let to be free, but at least chemical elements were selected and some geometrical rules for organizing them were applied (exclusive corner sharing of polyhedra within the GRINSP software [1]). The explosion of the number of predictions justified the creation of new databases. Among them, the PCOD [2] (Predicted Crystallography Open Database) contains the crystal data of predicted titanosilicates, phosphates, vanadates, niobates, fluoroaluminates (etc). These databases open now the possibility for the identification of a newly synthesized compound by the comparison of its experimental powder pattern with predicted ones. The powder patterns calculated from the >100.000 PCOD entries were gathered into the P2D2 (Predicted Powder Diffraction Database) [3-4], allowing for identification by using any classical search-match software. To be successful, identification attempts require mainly accurate predicted cell parameters. Many improvements, by using empirical or ab initio approaches, will be needed in order to restrict the number of structure candidates to those having really a chance to exist (quite a difficult task). Prediction is obviously a large part of our future in crystallography, the P2D2 represents a small new step in that direction.
[1] A. Le Bail, J. Appl. Cryst. 38 (2005) 389-395.
[2] Predicted Crystallography Open Database - http://www.crystallography.net/pcod/
[3] A. Le Bail, Powder Diffraction (2008) in the press.
[4] Predicted Powder Diffraction Database - http://sdpd.univ-lemans.fr/cod/pcod/P2D2/

Keywords: crystal structure prediction, databases, powder diffraction

P24.01.02

Acta Cryst. (2008). A**64**, C626

Databases for absorption, XAFS and XANES, and future opportunities for research and investigation

Christopher T. Chantler
University of Melbourne, School of Physics, University of Melbourne, School of Physics, Parkville, Victoria, 3010, Australia, E-mail：chantler@unimelb.edu.au

Databases for absorption include the International Union of Crystallography (IUCr) data, particularly of Creagh, Hubbell & Maslen et al.; the XCOM data of Hubbell & Scofield et al.; and the most recent FFAST (Form Factor ans Scattering Tables) of Chantler; together with experimental data using the X-ray Extended Range Technique (XERT) of Chantler. These databases typically have different purposes and uses, so that guidance and explanation is useful for a user to make informed comparisons with his experiments or medical or laboratory applications. Other databases exist, whether particularly for isolated atoms, diffraction, or condensed matter. The needs of X-ray Absorption Fine Structure (XAFS) and X-ray Absorption Near-Edge Structure (XANES) have often been somewhat different, for example needing fine spacing near absorption edges and the ability to take account of hole widths, discrete excitations, shake transitions and chemical environments to name a few. Typically, this relates to a programming environment rather than a database, and two particular examples are provided by packages by John Rehr (FEFF) and the Finite Difference Method of Yves Joly (FDMNES). This talk will discuss some key purposes and applications of these packages, and a little about how they should or should not be used.

Keywords: databases, absorption, XAFS

P24.02.03

Acta Cryst. (2008). A**64**, C626

Modernization of protein crystallography data formats: PDBML as a replacement for PDB?

Howard J Feldman, Paul Labute
Chemical Computing Group, Inc., 1010 Sherbrooke St. W., Suite 910, Montreal, Quebec, H3A 2R7, Canada, E-mail：feldman@chemcomp.com

Recently the Worldwide Protein Data Bank (MSD-EBI in Europe, PDBj in Japan and RCSB in USA) has made available the PDB Exchange Dictionary (http://pdbml.rcsb.org/schema/pdbx.xsd), adapted from the mmCIF dictionary. The data structures provided allow better disambiguation compared to the aging PDB format, for example introducing the concept of entities - unique molecules within the record. Biological assemblies are more easily represented as well. The PDBML format is aligned with modern relational database practices - only store each piece of information once. However the problem remains that most popular software uses the PDB format for both input and output. We have looked at some of the hurdles involved in converting PDB files to PDBML format, and back again, and present a new, scalable database system, Protein SILO (PSILO) which overcomes these. The benefits of using correctly built PDBML files include more accurate interpretation of ligands and small molecules, more precise definitions of experimental conditions, categories for structure validation measures (a protein 'health check'), and far more powerful search capabilities when stored in a relational database.

Keywords: PDBML, structural databases, biomacromolecular structures

P24.02.04

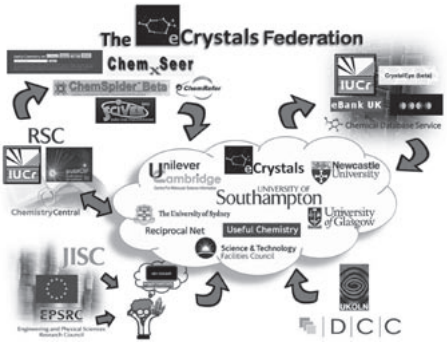
Acta Cryst. (2008). A**64**, C626–627

The eCrystals Federation: Management and publication of small molecule structure data for all

Simon J Coles[1], Michael B Hursthouse[1], Richard A Stephenson[1], Peter D Cliff[2], Elizabeth Lyon[2], Manjula Patel[2], Jim Downing[3], Peter Murray-Rust[3]
[1]University of Southampton, School of Chemistry, School of Chemistry, Southampton, Hampshire, SO17 1BJ, UK, [2]UKOLN, University of Bath, UK, [3]Unilever Centre for Molecular Informatics, University of Cambridge, UK, E-mail：s.j.coles@soton.ac.uk

Modern data collection methods and publication processes cause a bottleneck in the dissemination of crystal structure data, which hinders the growth of databases. There is also increasing pressure from funders to make results available as widely and rapidly as possible. We have established an institutional data repository that supports, manages and disseminates metadata relating to crystal structure data. The eCrystals Federation project (http://wiki.ecrystals.

chem.soton.ac.uk) is establishing a network of such crystallographic data repositories across an international group of laboratories. Data is being harvested by CCDC and the CDS and the project is working with IUCr, RSC, Chemistry Central and Nature to establish protocols for scaleable harvesting mechanisms across the network. By engaging data centres, librarians, researchers, publishers and information providers we are developing approaches to the preservation and curation of scientific data in open repositories (the UK Digital Curation Centre is a partner). A demonstration and strategies for installation and population of eCrystals repositories at new sites will be outlined, based on experiences of early adopter sites.



Keywords: publishing, database preparation, software for crystallography

P24.04.05

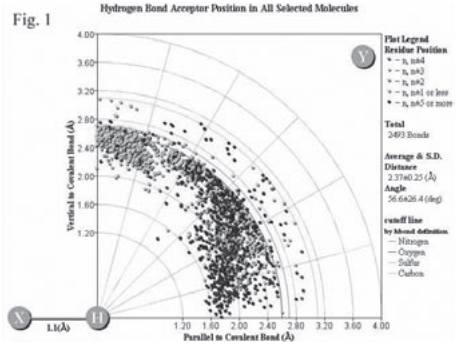
Acta Cryst. (2008). A**64**, C627

Development of hydrogen and hydration database for biomolecules (HHDB)

Nobuo Okazaki[1], Takashi Ohhara[1], Hisao Umino[1], Toshiyuki Chatake[2], Kazuo Kurihara[1], Raul E Cachau[3], Michael Blaber[4], Nobuo Niimura[5], Ryota Kuroki[1]
[1]Japan Atomic Energy Agency, Quantum Beam Science Directorate, Research Group for Molecular Structural Biology, 2-4, Shirakata-Shirane, Tokai, Ibaraki, 319-1195, Japan, [2]Kyoto University Research Reactor Institute, [3]Advanced Biomedical Computing Center, National Cancer Institute at Frederick, [4]Department of Chemistry and Biochemistry, Florida State University, [5]Graduate School of Science and Engineering, Ibaraki University, E-mail：okazaki.nobuo@jaea.go.jp

In protein molecules, key energetic contributors are solvation, desolvation and hydrogen bonding. They contribute protein folding, dynamics and molecular recognition. As a result, more elaborate studies of hydrogen atoms will be great help to recognize protein structures and obtain new findings of them. However, we do not have system which dedicated to characterization and analysis of hydrogen bonding. Therefore, we have developed a database for hydrogen and hydration water molecules. That database named Hydrogen and Hydration Database for Biomolecules (HHDB; http://hhdb.tokai-sc.jaea.go.jp/). Hydrogen bond data stored to HHDB use hydrogen atom coordinates determined directly by neutron diffraction and certain extremely high resolution x-ray diffraction. HHDB provides graphical user interface, users can use it through web browser. HHDB can visualize hydrogen atom positions in protein and solvent, and hydrogen bonding interactions. Figure 1 shows HHDB plot example. In this plot, hydrogen atom is placed at the origin, and each point represents hydrogen bond distance and angle. We are improving the web user interfaces and the performance for usability.

Fig. 1



Keywords: neutron diffraction, protein, hydrogen bond

P24.04.06

Acta Cryst. (2008). A**64**, C627

Protein-protein interactions: Structural features and empirical estimation of free energy of binding

Pinak Chakrabarti, Mainak Guharoy
Bose Institute, Department of Biochemistry, P1/12 CIT Scheme VIIM, Kolkata, W. Bengal, 700054, India, E-mail：pinak@boseinst.ernet.in

An analysis of protein-protein complexes has shown that the average interface area (the accessible surface area, ASA, on the two components that gets buried on complex formation) is ~2000 Å^2 [1] and that the interface can be dissected into core and rim regions with the former being composed of residues that are more conserved than those in the latter [2]. There is an approximate linear relation between the change in the surface area buried (δASA) and the change in free energy of binding (δδG) obtained from alanine scanning mutagenesis. This relationship has now been used to predict the free energy of binding (δG). The experimental δG varies linearly with the interface area (or the number of atoms buried in the complex), but plateaus off beyond ~1600 Å^2, indicating that the energy gained from burying additional surface area is used for conformational changes associated with larger interfaces. Further, the interface can be dissected into secondary structural segments [3]. The β and NR-classes of interfaces have a relatively lower δG than α- and $\alpha\beta$-classes.

References
[1] Chakrabarti, P. and Janin, J. Dissecting protein-protein recognition sites. Proteins, 47, 334-343 (2002).
[2] Guharoy, M. and Chakrabarti, P. Conservation and relative importance of residues across protein-protein interfaces. Proc. Natl. Acad. Sci. USA, 102, 15447-15452 (2005).
[3] Guharoy, M. and Chakrabarti, P. Secondary structure based analysis and classification of biological interfaces: identification of binding motifs in protein-protein interactions. Bioinformatics, 23, 1909-1918 (2007).

Keywords: protein-protein interactions, free energy of binding, molecular recognition

P24.07.07

Acta Cryst. (2008). A**64**, C627–628

Defining a protein: Mining the protein structure database

Boguslaw Stec, B.V.L.S. Prasad, Ying Zhang, Adam Godzik
Burnham Institute for Medical Research, IIDC, 10901 N. Torrey Pines Rd., La Jolla, CA, 92037, USA, E-mail：bstec@burnham.org

Rapid increase of structural information expressed in the explosive growth of the PDB allows for rigorous studies of two links in a central biology paradigm: a sequence-structure and a structure-function relationship. We constructed a database of redundant protein structures by application of sequence- based methods (CD-HIT). Subsequently, we clustered the structures belonging to a single protein sequence by structural methods (BOSv3) and derived a global