**P.24.06.2**
*Acta Cryst*. (2005). A**61**, C483
**The New PDF-4+ 2005: A Relational Database (RDB) with Atomic Structure Information and On-the-fly Total Pattern Analysis Capability**
John Faber, Soorya Kabekkodu, *International Centre for Diffraction Data, Newtown Square, PA 19073 USA*. E-mail: faber@icdd.com

The ICDD is continuing to develop new RDB database capability for the Powder Diffraction File™ (PDF®) [1,2]. The PDF-4+ 2005 will be released in August, 2005. This database will contain approximately 84,000 new entries with complete atomic coordinate information. Using the structure data, powder patterns will be calculated for electron, neutron and x-ray diffraction; these are calculated on-the-fly (as needed). Scattering contrast studies as a function of probe and wavelength can be used to aid experiment design. In addition to the standard peak intensity listing in the PDF, integrated intensity information will be available for all unique hkl's within each entry. Elemental composition data (available for all entries in the PDF) can be used as filter criteria for effective searches in the RDB. We will show how atomic evironment data can be used to help understand classes of materials properties. One of the aims of these initiatives is to enhance our ability to perform materials design studies. Fully digitized powder patterns are a first step in the realization of this process.

[1] Faber J., Fawcett T., *Acta Cryst.*, 2002, **B58**, 325-332. [2] Kabekkodu S. N., Faber J., Fawcett T., *Acta Cryst.*,2002, **B58**, 333-337.
**Keywords: database preparation, powder diffraction analysis, powder patterns**

**P.24.07.1**
*Acta Cryst*. (2005). A**61**, C483
**Carbonyl-Carbonyl Interactions in First-row Transition Metal Complexes**
Hazel A. Sparkes[a], Mary F. Mahon[a], Paul R. Raithby[a], Frank H. Allen[b], Gregory P. Shields[b], [a]*University of Bath, Department of Chemistry, Claverton Down, Bath, BA2 7AY UK.* [b]*Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 IEZ, UK*. E-mail: chphas@bath.ac.uk

Carbonyl-carbonyl interactions in organic species have been previously documented [1]. In theory, the strength of carbonyl-carbonyl interactions should be greater when the carbonyls are directly attached to a transition metal since the magnitude of the carbonyl dipole is greater than in purely organic compounds. The Cambridge Structural Database [2] was used to identify first row transition metal carbonyl species in which two carbonyl groups were separated by less than 3.6 Å and could potentially interact. Only two of the three main carbonyl-carbonyl interaction motifs previously identified for organic carbonyl species are possible in transition metal carbonyl compounds; these are illustrated Figure1.

Data analysis suggests that transition metal carbonyl-carbonyl interactions in motifs A and B are more prevalent than organic carbonyl-carbonyl interactions, with around 33% of transition metal carbonyls suitably orientated to interact while only 8% of organic carbonyls can potentially interact. A comparison of the geometries of the interaction motifs is provided along with an analysis of accompanying theoretical calculations.

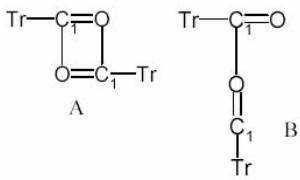

Figure 1 – *interaction motifs in transition metal carbonyl species.*

[1] Allen F.H., Baalham C.A., Lommerse J.P.M., Raithby P.R., *Acta Cryst.*, 1998, **B54**, 320. [2] Allen F.H., *Acta Crystallogr.*, 2002, **B58**, 380.
**Keywords: database manipulation, transition metal complexes, interatomic interactions**

**P.24.07.2**
*Acta Cryst*. (2005). A**61**, C483
**ISOBASEmm: Isostructurality in the Protein Data Bank**
Jan M.M. Smits, R. de Gelder, *Molecular Materials, Institute for Molecules and Materials, Radboud University Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands*. E-mail: R.deGelder@science.ru.nl.

The ISOQUEST program [1,2], a method for the comparison of crystal structures based on pattern matching techniques [3], has recently been applied to analyse the Cambridge Structural Database (CSD) [4]. A full analysis of all entries in the CSD led to ISOBASE, a database which contains all isostructurality relations in the CSD. Investigation of this ISOBASE shows that many classes of identical or closely related structures can be found, although the chemistry within those classes can vary considerably. This information can be used to extract crystal packing rules that otherwise can not be revealed.

To extend the range of applicability of the ISOQUEST method an analogous approach was chosen to analyse the collection of all macromolecular structures present in the Protein Data Bank (PDB) [5]. This poster presents the resulting macromolecular database ISOBASEmm and examples of structural relations in the PDB that can be extracted from this database.

[1] Gelder R. de, Smits J.M.M., *Acta Crystallogr.* , 2004, **A60**, s78. [2] Gelder R. de, Smits J.M.M., 2005, *submitted for publication.* [3] Gelder R. de, Wehrens R., Hageman J.A., *J. Comp. Chem.* , 2001, **22**, 273-289. [4] Allen F. H., *Acta Crystallogr.* , 2002, **B58**, 380-388. [5] Berman, Westbrook, et al., *Nucleic Acids Res.*, 2000, **28**, 235-242.
**Keywords: protein structure comparison, isostructurality, databases**

**P.24.07.3**
*Acta Cryst*. (2005). A**61**, C483
**Powder Diffraction CIFs: Preparation and Review**
Brian H. Toby[a], Nicola Ashcroft[b], [a]*NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8563, USA.* [b]*International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.* E-mail: brian.toby@nist.gov

The IUCr now strongly urges authors to submit observed and computed powder diffraction data in CIF format when publishing Rietveld results [1]. This will offer obvious benefits for the archival of data and for the testing of alternative models where structures are in dispute. The most immediate application is likely to be for manuscript review, as the quality of a Rietveld fit is best judged graphically, not from statistical figures of merit, such as profile R-factors or $\chi^2$ values. Figures submitted for publication are seldom sufficient for close examination.

This paper will present some resources for reporting Rietveld results in CIF format, including information on software available for preparation of a CIF from a Rietveld fit. For the review of Rietveld results from a CIF, the pdCIFplot program will also be discussed [2]. This open-source program runs on all common computer platforms (Windows, Macintosh, & Unix) and allows powder diffraction data and fits to be plotted in a variety of formats directly from a CIF.

[1] http://journals.iucr.org/services/cif/powder.html
[2] http://www.ncnr.nist.gov/xtal/software/cif/pdCIFplot.html
**Keywords: CIF, powder diffraction, Rietveld analysis**

**P.24.07.4**
*Acta Cryst*. (2005). A**61**, C483-C484
**Automating the Identification of Packing Motifs; dSNAP**
Gordon Barr[a], Andrew Parkin[a], W. Dong[a], Chris J. Gilmore[a], Chick C. Wilson[a], [a]*Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK.* E-mail: a.parkin@chem.gla.ac.uk

With the explosion in high quality structural determinations in the area of small molecule crystallography, the problem of efficient and meaningful mining of the data in, for example, the Cambridge Structural Database (CSD) [1] is very relevant for structural chemists.

The available databases represent an enormously powerful resource, but faced with more than 300,000 structures, such as are found in the CSD, the attempt to extract meaningful chemical information can be daunting.

The application of the recently developed program *dSNAP* to inter-molecular interactions and packing motifs is described through some novel examples. Similarity matching within *dSNAP* allows clustering of geometric data extracted from the CSD, which is found to be sensitive to small but significant geometry variations. This method relies solely on the extracted geometric information and is therefore independent of any chemical bias. However, the final interpretation of the different clusters in a chemically sensible way is still the sole responsibility of the structural chemist.

The method is illustrated by several examples, both simple and complex.

[1] Allen F.H., *Acta Crystallogr.*, 2002, **B58**, 380-388.

**Keywords: database mining, cluster analysis, intermolecular packing**

**P.24.07.5**

*Acta Cryst*. (2005). A**61**, C484

**Description of Software for the Planning, Execution, and Refinement of Crystallography Experiments**

Paige N. Vinson, Bryan Greenway, Greg Nordstrom, Dave Riling *Kendro Lab Automation, Brentwood, TN, USA .37027* E-mail: paige.vinson@kendro.spx.com

A modular and highly integrated software package will be described which takes the crystallographer from the planning stage through to the refinement of the experimental process. This includes configuration and control of automation used to execute the experiment. Some key elements include reagent/protein management, screen design, database query tools, imaging, and real-time monitoring of automation and experiment status. Each application is specialized for a specific function but provides input to other applications. For example, while viewing experimental images, a user may choose an image deemed "interesting" and have the conditions for that site sent to the screen designer application for the starting conditions of a fine screen. Data generated from experiments can be mined using a novel, graphical query tool. Query results may be sent to the image viewing and analysis application for further study, as well as to the screen design application for use in designing additional rounds of refined experiments. This technology is highly data-driven and is enabled through the use of a centralized database. This single point of data management promotes efficient viewing, sharing, and mining of information.

These and other features of the software will be presented in a format describing typical scenarios and methods of use.

**Keywords: software for crystallography, application software, databases**

**P.24.08.1**

*Acta Cryst*. (2005). A**61**, C484

**Covariance Correlations from Genome-Wide Homology Sequence Analysis of DHFR**

Vivian Cody, William L. Duax, Robert Huether, *Hauptman-Woodward Medical Research Institute, Buffalo, NY 14203, USA*. E-mail: Cody@hwi.buffalo.edu

To test the hypothesis that active site residue changes among dihydrofolate reductase (DHFR) influence binding specificity, the HSSP alignments of all protein sequences from the SWISS-PROT TrEMBL database that had 30% identity with DHFR were retrieved and resulted in a list of 298 gene products: 177 prokaryotic and 121 eukaryotic entries. Analysis of these profiles at the 70% identity level revealed: (1) 21 residues that are highly conserved in both kingdoms, (2) 13 additional residues whose frequency of occupancy achieves the 70% or greater level of sequence identity in eukaryote species only, (3) 14 sites in which a significant change in the dominant amino acid occupancy occurs between prokaryotic and eukaryotic species, and (4) the precise location of six inserts encompassing from one to seven residues that separate the two gene families. These results suggest that there has been an evolution from prokaryote to lower eukaryote to humans of an increasingly more specific ensemble of residues whose covariance correlates with functional specificity. A preference for ring-ring stacking involving Tyr33 and Tyr179 was noted in human DHFR. The usage profile at positions 33 and 179 respectively are: Y39, F23, H17% / F83, Y8% for eukaryotes and H31, Y3, F3% / F47, Y32% for prokaryotes. In the sequence of *Mycobacterium tuberculosis* DHFR these positions are H30 and L153. Supported in part by GM51670 (VC) and DK026546 (WLD).

**Keywords: genome, dihydrofolate reductase, profile**

**P.24.08.2**

*Acta Cryst*. (2005). A**61**, C484

**The PDB Format in the 21st Century, a Modest Proposal**

Herbert J. Bernstein[a], Frances C. Bernstein[b], [a]*Mathematics and Computer Science Department, Dowling College, Oakdale, NY 11769, USA.* [b]*Bernstein+Sons, Bellport, NY 11713, USA.* E-mail: yaya@bernstein-plus-sons.com

The Protein Data Bank format created in the 1970's is the major user-level interchange format for macromolecular structures but is sorely pressed by the demands of larger structures and the rich detail of information now available for many structures. Newer mmCIF, XML and other formats effectively address these issues, but leave a gap in terms of software support for existing applications. In this presentation we make a modest proposal to help to close this gap and to simplify the adaptation of existing applications to the management of new structures.

**Keywords: PDB format, mmCIF, software**

**P.24.10.1**

*Acta Cryst*. (2005). A**61**, C484

**Protein Crystallization Conditions Database, Crystal T.B.**

Koji Inaka[1], Shigeru Sugiyama[1], Fujiko Shibata[1], Yoshiko Kobayashi[1], Junya Ohori[2], Kaoru Sugimori[2], Jose Martin Ciloy[2], Masato Kitajima[2], Takako Sakamaki[2], Yoko Sato[2], Shigekazu Masumoto[2], [1]*MARUWA Food Industries, Inc.* . [2]*Fujitsu Kyushu System Engineering Ltd*. E-mail: kita@fqs.fujitsu.com

In the past 10 years, technical improvements in protein crystallography have been quite remarkable. Especially, with the improvements in hardware for the collection of diffraction data and in software for structure analysis and refinement, it is now possible to solve the protein structure within a few days of receiving a protein crystal. However, since these ordinary crystallization methods are based on a trial and error screening technique, a great amount of time and sample is necessary. We have developed a database that will help guide the user to a rational crystallization method. This database is composed of all the elements that are essential for protein crystallization experiments. The database contains not only detailed crystallization conditions data extracted from published reports of crystallization and structural analysis, but also biological information for each biological macromolecule. Crystallization conditions related to a specific target protein can be easily searched with the help of just a few keywords. Comparison of the search results readily reveals common parameters that provide an estimate to possible crystallization conditions before any screening experiment is started. It is an efficient approach to crystallization since it helps reduce unnecessary screenings in the process. The database also provides homology searching which is helpful in finding the crystallization conditions for unknown proteins where only the amino acid sequence is known.

**Keywords: database, crystallization, screening**