P.24.01.1
Acta Cryst. (2005). A**61**, C481
COD (Crystallography Open Database) and PCOD (Predicted)
Armel Le Bail, Daniel Chateigner, Xiaolong Chen, Marco Ciriotti, Lachlan M.D. Cranswick, Robert T. Downs, Luca Lutterotti, Alexandre F.T. Yokochi, *COD Advisory Board.* E-mail: cod@crystallography.net

The COD was created in March 2003 and was built on the PDB model of open access on the Internet. It is intended that this database [1] consist of any small or medium crystal structure (inorganic, organic, organometallic). Currently the total entry number is close to 15.000, including 6.600 entries from the American Mineralogist Crystal Structure Database (AMCSD) [2], and CIF files donations from a few laboratories in Europe or from individuals. The distribution is made through an Apache/MYSQL/PHP system that takes queries on chemistry, ranges of cell parameters, volumes, etc, as well as combination of fields, and can download or upload CIF files. A large donation of CIFs is anticipated from the IUCr.

The PCOD, created in December 2003, is a COD subset of crystal structures predicted by the GRINSP computer program [3]. It is growing fast and already contains > 1000 CIF files corresponding to M_2X_3, MX_2, MX_3 or $M_aM'_bX_c$ formulations (X = O, F; M/M' = B, Na, Si, Al, P, Ca, V, Fe, Ga, Re, Zr, etc), including hypothetical zeolites and other binary compounds with N-connected 3D frameworks of M atoms (N = 3, 4, 5, 6) as well as ternary compounds with mixed M/M' frameworks. The PCOD is open for search, download and upload of predicted crystal structures (coming from any prediction computer program, inorganic or small and medium organic molecules).

Crystallographers are invited to deposit their data as CIF files.

[1] http://www.crystallography.net/ [2] http://www.geo.arizona.edu/AMS/amcsd.php [3] http://www.cristal.org/grinsp/
Keywords: crystal structure database, open access, COD

P.24.01.2
Acta Cryst. (2005). A**61**, C481
Hydrogen and Hydration DataBase for Bio-Macromolecules (HHDB)
Ryota Kuroki[1], Toshiyuki Chatake[1], Kazuo Kurihara[1], Takashi Ohhara[1], Hisao Umino[1], Raul E. Cachau[3], Michael Blaber[4], Nobuo Niimura[1,2], *[1]Neutron Science Research Center, Japan Atomic Energy Research Institute (JAERI). [2]Institute of Applied Beam Science, Graduate School of Science and Engineering, Ibaraki University. [3]Advanced Biomedical Computing Center, National Cancer Institute at Frederick. [4]Department of Chemistry and Biochemistry, Florida State University.* E-mail: r-kuroki@neutrons.tokai.jaeri.go.jp

Solvation, desolvation and hydrogen bonding are key energetic contributors to biopolymer folding, dynamics and molecular recognition; however, no systematic, high-resolution stereochemical database dedicated to the characterization and analysis of hydrogen bonding exists. We have created a hydrogen and hydration database for bio-macromolecules (HHDB; http://hhdb.tokai.jaeri.go.jp) that categorizes all hydrogen atoms and hydration water molecules including hydrogen atoms in proteins. The HHDB includes H-bond data only from direct determination of hydrogen atoms by neutron diffraction, and certain extremely high resolution x-ray diffraction data. The HHDB provides a graphical user interface for the visualization of all hydrogen atom positions in proteins and solvent, and all hydrogen bonding (H-bond) interactions. For example, one type of plot featured in the database places the hydrogen atom positions at the origin and plots the direction of the H-bond donor along the horizontal axis, allowing the user to visualize the distribution of the acceptor atoms, distance and angle. The HHDB provides researchers with a much-needed resource for understanding and analyzing hydrogen-bonding.
Keywords: neutron diffraction, protein, hydrogen bond

P.24.01.3
Acta Cryst. (2005). A**61**, C481
Four Years of the EPSRC SRS Service
Ross W. Harrington, W. Clegg, *School of Natural Sciences (Chemistry) University of Newcastle.* E-mail: r.w.harrington@ncl.ac.uk

Since October 2001, the EPSRC national crystallography service in the UK has had a SRS component, run in conjunction with the national crystallographic service based at Southampton University, through the use of station 9.8 at Daresbury laboratory.

Over the course of the past 3.5 years regular monthly visits have taken place, and a total of 385 data collections made, averaging 6 data collections per day of beam time.

This poster reports the many successes of the service, and includes statistics on users and publications benefiting from the service.

The reader will hopefully gain an insight into a service utilizing the most successful small molecule synchrotron beam line in the world.
Keywords: EPSRC, national crystallographic service , statistics

P.24.02.1
Acta Cryst. (2005). A**61**, C481
The Crystallographic Semantic Web
Peter Murray-Rust, Simon Tyrrell, Yong Zhang, *Unilever Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, CB2 1EW, UK.* E-mail: pm286@cam.ac.uk

The semantic Web [Berners-Lee, http://www.w3./2001/sw/] is a vision of a global knowledge network where machines understand and reason from web-based resources. We provide working demonstrations that this technology is ideally suited to create an Open crystallographic knowledge base. The complete information chain - experiment, publication, storage, dissemination, searching and re-use can be completely managed by machines. CIFs can be converted to XML, annotated, and redistributed as WebServices and indexed by conventional search engines. New structures can be announced through crystallographically enhanced RSS feeds.

Our World Wide Molecular Matrix (WWMM, http://wwmm.ch.cam.ac.uk) uses free-text XML indexing in a repository (eXist) to collect hundreds of thousands of compounds. IUPAC InChI and chemical substructure searching (e.g. http://openbabel.sf.net) to provide an instant, Open and freely redistributable crystallographic knowledge base. Data are abstracted from publishers who allow Open access to CIFs and authors can publish CIFs directly into XML (http://www.xml-cml.org). Institutional and national collections ((http://eprints.soton.ac.uk/1633/) with appropriate metadata allow conventional search engines to index the data and effectively create a complete Open database of crystal structures.
Keywords: XML, CIF, semantic web

P.24.02.2
Acta Cryst. (2005). A**61**, C481-C482
ECRYSTALS(.CHEM.SOTON.AC.UK): Open Archive Publication of Crystal Structure Data
Mike B. Hursthouse[a], S. J. Coles[a], J. G. Frey[a], L. Carr[b], C. Gutteridge[b], L. Lyon[c], R. Heery[c], M. Duke[c], M. Day[c], *[a]School of Chemistry, [b]School of Electronics and Computer Science, University of Southampton, [c]UKOLN, University of Bath., UK.* E-mail: mbh@soton.ac.uk

The eCrystals Repository has been developed as part of the eBank-UK project, to serve as an examplar and testbed for open archive publication, linking and aggregation of digital scientific data.

This poster outlines a "pre-print"[1] procedure for the rapid and effective dissemination of structural information "@source", based on the "e-print" concept. An eCrystals record makes available, for assessment and/or re-use, all raw, derived, results and validation data generated during the course of a crystallographic experiment. During the deposition process metadata, comprising bibliographic and chemical identifiers, are associated with a dataset, and key items of data are automatically extracted and displayed on a HTML jumpoff page, which is then offered over the internet by standard Open Archive Initiative (OAI) protocols. The bibliographic, crystallographic and chemical metadata relating to an eCrystals record may be

'harvested' by information providers through the OAI Protocol for Metadata Harvesting (OAI-PMH)[2]. Developing metadata standards allow this information to be linked and aggregated with existing literature and electronic resources to provide 'added value' to the chemical and crystallographic literature.

[1] Garson L.R., Acc. Chem. Res., 2004, ASAP Article, DOI: 10.1021/ar0300017. [2] Heery R., Duke M., Day M., Lyon L., Hursthouse M., Frey J., Coles S., Gutteridge C., Carr L. ESA/ESRIN, 2004, Frascati, Italy, European Space Agency, 8pp. (http://eprints.soton.ac.uk/9705/).

Keywords: structural data publishing, e-science, data harvesting

P.24.03.1

Acta Cryst. (2005). A**61**, C482

Hydrogen Bond Capacity of Organic Functional Groups: a CSD derived Database

W.D. Sam Motherwell, L. Infantes, *Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, UK.* E-mail: motherwell@ccdc.cam.ac.uk

The H-bonding behaviour of organic functional groups is of general interest. We have devised new methodology[1] to build a specialized database of non-bonded contacts extracted from the Cambridge Structural Database (CSD), using the Microsoft Access database program. We extracted 35,056 crystal structures where all hydrogen atoms had 3D-coordinates present, no metal atoms, OH, NH, or SH present, giving the possibility of at least one strong H-bond. The data were processed by the Pluto program, calculating the number of non-bonded contacts for 108 chemical groups, (distance < sum of van der Waals radii + 0.1 Å). Contacts are classified as D (donor bond), A (acceptor bond), X (not H-bond), and U (uncertain). Contacts are both inter- and intra-molecular. The accessible surface area of atoms was also calculated.

This database, CSDContact, can be used to derive average values for H-bond behaviour of functional groups (e.g. OH in COOH D=99% A=4% X=21%; in OH-CH_2-R D=94% A=63% X=19%). We present average figures for the number of donor/acceptor bonds per group, the dependency on available steric surface, total donor/acceptor atom ratio, and some examples of competition effects between groups in specific ratios. More practically, CSDContact can be used to answer questions where we limit the number and ratio of the chemical groups[2], e.g. What happens if the crystals contain just one alcohol OH and one cyano group?

OH-R-CN molecule → crystal OH…OH or OH…NC ? [92%, 4%]

[1] Infantes L., Motherwell W.D.S., *Chem. Commun*, 2004, 1166-1167. [2] Infantes L., Motherwell W.D.S., *Z. Kristallogr*., 2005, **220**, 1-8.

Keywords: hydrogen bonds, databases, functional groups

P.24.04.1

Acta Cryst. (2005). A**61**, C482

Unveiling the ω/ψ Correlation in High Resolution Protein Structures

Luciana Esposito[a], Alfonso De Simone[b], Adriana Zagari[b], Luigi Vitagliano[a], *[a]IBB-CNR, Naples. [b]Dipartimento delle Scienze Biologiche, University of Naples, Italy.* E-mail: lucianaesposito@chemistry.unina.it

The planarity of the peptide group is one of the fundamental features of protein structures. Several investigations on peptide bond distortions have been reported [1]. Here we present a statistical survey of peptide plane deviations analyzed as a function of the local conformation of the backbone. By surveying a dataset of 163 non-homologous protein chains, determined at atomic resolution, we have identified the stereochemical conditions that favor significant deformations of peptide bond planarity. In particular, the values of the ω dihedral angle are found to be strictly correlated to the values of the adjacent ψ angle [2]. This trend is also observed in highly strained states such as those occurring in vicinal disulfide bridges. The dependence of the ω angle on the ψ angle is similar to that already observed for a different type of deviation from peptide planarity: the pyramidalization at the carbonyl carbon atom [3].

These findings provide direct evidence for the mutual influence of the geometrical parameters that describe protein structures.

[1] MacArthur M.W., Thornton J.M., *J. Mol. Biol.*, 1996, **264**, 1180. [2] Esposito L., De Simone A., Zagari A., Vitagliano L., *J. Mol. Biol.*, 2005, **347**, 483. [3] Esposito L., Vitagliano L., Zagari A., Mazzarella L., *Protein Sci*., 2000, **9**, 2038.

Keywords: peptide planarity, statistical analysis, conformation

P.24.04.2

Acta Cryst. (2005). A**61**, C482

Beyond Text-based Queries at the Protein Data Bank, Japan

Daron M. Standley[a,b], Haruki Nakamura[a], *[a]Protein Research Institute, Osaka University, Osaka, Japan. [b]Japan Science and Technology Agency.* E-mail: standley@protein.osaka-u.ac.jp

The Protein Data Bank has traditionally offered only text-based query services. These tools are very powerful in the hands of experts when the data entries have been well annotated. As the database grows through structural genomics projects, however, annotation will likely become limited. Here we introduce a suite of query tools that do not require complex textual input. Starting from a particular entry one may find sequence homologs (Sequence Navigator[1]), structural neighbors (Structure Navigator[2]), or, if the entry is a protein complex, structurally similar protein-protein interfaces (PISup[3]). In addition, alignments may be further optimized and refined using our powerful structural alignment program GASH[4]. All of the above programs utilize the Number of Equivalent Residues (NER[5]), a novel scoring function that detects similarities rather than differences between structures. In this way, even local similarities (i.e., domains, active sites, etc.) can be detected.

[1] http://www.pdbj.org/cgi-bin/run_seq_hom.cgi [2] http:// www.pdbj.org/cgi-bin/run_algn_struc.cgi [3] http://www.pdbj.org/cgi-bin/run_pisup.cgi [4] http://www.pdbj.org/cgi-bin/run_gash.cgi [5] Standley D.M., Toh H., Nakamura H., *Proteins*, 2004, **57**(2), 381-391.

Keywords: protein structure comparison, databases, protein-protein interactions

P.24.06.1

Acta Cryst. (2005). A**61**, C482

Implementation of Calculated Patterns Quality Marks in the Powder Diffraction File

Soorya Kabekkodu, *International Centre for Diffraction Data Newtown Square, PA 19073, U.S.A.* E-mail: suri@icdd.com

Quality mark assignments for the calculated patterns are becoming a necessity considering their growing population in the Powder Diffraction File™ (PDF®). An estimate of the number of calculated diffraction patterns in the Release 2005 is about 400,000. The focus of the quality mark is to determine the confidence level of the structural model used and its impact on the calculated pattern from the phase identification point of view. The major step in the adopted method involves several crystallographic and editorial checks by the International Centre for Diffraction Data (ICDD), followed by the extraction and flagging of the structural database warnings/comments. Resulting calculated patterns will be classified into various categories based on the significance and nature of the warnings/comments. In the final step, a quality mark (QM) will be assigned to a calculated pattern based on its category.

A database analysis of approximately 400,000 calculated diffraction patterns will be presented with special emphasis on phase identification using some case studies. The prime crystallographic checks implemented in the editorial process will be discussed in detail.

Keywords: phase identification, powder diffraction analysis, data checking