'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