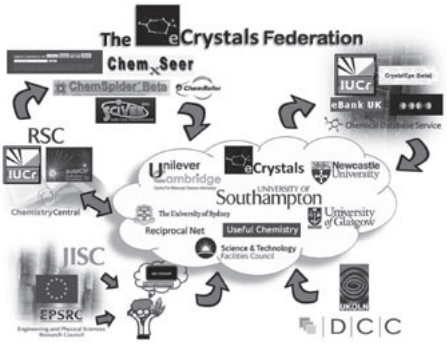
chem.soton.ac.uk) is establishing a network of such crystallographic data repositories across an international group of laboratories. Data is being harvested by CCDC and the CDS and the project is working with IUCr, RSC, Chemistry Central and Nature to establish protocols for scaleable harvesting mechanisms 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

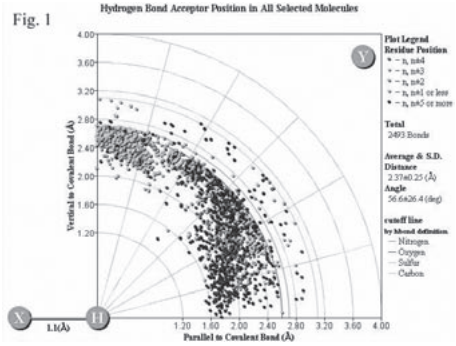
Acta Cryst. (2008). A**64**, C627

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

distribution of conformational states represented in the PDB. A systematic review of multiple deposits shows that a single protein is rarely represented by a single structural conformer. This result sheds light on the first link and demands the reformulation of the protein-folding problem. A vast majority of proteins shows significant number of distinct conformational states with, sometimes large, structural divergence (up to ~24Å). The results suggest that every single protein evolved according to its own optimization principles combining different proportions of rigid (solid-like) and mobile (liquid-like) structural elements. The results suggest further that the optimization process that produced the particular combination of those elements is intricately connected with the function of individual proteins. Therefore, the structural description of the protein, besides the folding class (the architecture represented by the SCOP database), should include the natural structural divergence (width of the distribution) as two main attributes. Additionally, our analysis suggested the principles of functional evolution by use of the Dual Personality sequences (sequences with incomplete representation in the atom records that have distinctive sequence features from regularly folded and intrinsically disordered fragments).

Keywords: redundant structure database, distribution of conformational states, protein folding

P24.06.08

Acta Cryst. (2008). A**64**, C628

Analysis of the organic X-ray powder diffraction database and its use with pharmaceutical substances

Fangling Needham[1], Cyrus Crowder[2], Timothy Fawcett[3]
[1]International Centre for Diffraction Data, Science, 12 Campus Blvd., Newtown Square, PA, 19073, USA, [2]International Centre for Diffraction Data, [3]International Centre for Diffraction Data, E-mail：needham@icdd.com

The crystals of small molecule pharmaceutical substances are usually of low symmetry, often have hydrogen-bonding-induced polymorphs [1], and frequently exhibit anisotropic crystal habits leading to preferred orientation in X-ray powder diffraction (XRPD) experiments. Statistical studies and cluster analyses have been used to show the prevalence of polymorphs and low symmetry space groups for these materials. Such statistical analyses can be performed using permutations of 40 different property and data searches with the PDF-4/Organics database (PDF-4). The XRPD patterns for these materials can exhibit overlapping peaks over small ranges of two theta angles, peak asymmetry at low angles, and preferred orientation. However, the proper choice of diffractometer and specimen configuration can minimize the two latter effects. Examples of XRPD data from experiments with thermodynamically stable polymorphs [2] will be given to illustrate this optimization. Modern crystallographic software can index XRPD patterns, determine unit cell dimensions and assign space group symmetry. Using this information and the PDF-4 searches lead to the model selection from the Cambridge Structural Database (CSD), followed by Rietveld refinement to verify both crystallographic parameters and indexing assignments of the experimental XRPD pattern. The experimental XRPD pattern then becomes a powerful reference for quality control measurements, quantitative analysis and polymorph identification. Systematic analysis examples of data from XRPD experiments for active pharmaceutical ingredients will be given.
[1] Yamamoto K., Uchida T., Yonemochi E., Oguchi T., Terada K., Nakai Y., Chem. Pharm. Bull., 1993, 41, 1632-1635. [2] Needham F., Faber J., Fawcett T., Powder Diffraction, 21, 245-247.

Keywords: X-ray powder diffraction, polymorphs, preferred orientation

P24.07.09

Acta Cryst. (2008). A**64**, C628

Studying conformational preferences for mechanistic purposes: Using database mining and computation

Alexander J Hamilton, Anthony G Orpen, Jeremy N Harvey
University of Bristol, School of Chemistry, Cantocks Close, Bristol, Bristol, BS8 1TS, UK, E-mail：chxah@bristol.ac.uk

It has long been known that phosphine and phosphite ligands can adopt various conformations depending on coordination environment. Containing over 400,000 structures more than half of which are metal complexes, the Cambridge Structural database (CSD)[1] has an extensive amount of information regarding ligands, the attached metal centre and the coordination environment of the complex. By combining the information from database mining with computational (DFT) studies it is possible to explore the response of ligands to different coordination environments. This enables us to have a better understanding of conformational behaviour when exploring mechanistic pathways, which often involve coordination changes. The chosen systems, tribenzylphosphine $P(CH_2Ph)_3$ and triphenylphosphite $P(OPh)_3$, are part of a synthetically relevant series of phosphorus ligands for catalytic studies. Although DFT is known to give reasonable agreement with molecular structure,[2] these systems are far too large to be studied by DFT conformational searching. The combination of DFT studies with database mining results has accurately predicted the conformational preference and lowest energy profile for these ligands across a series of coordination environments, and has been utilised in mechanistic studies undertaken in our group. Further analysis of the data retrieved from the CSD has highlighted conformational interconversion pathways. These have been explored and structures along the pathways have been used to predict transition states between conformers. This furthers our understanding of ligand conformation during reactions.
1. F. H. Allen, Acta Cryst., B58, 380-388, 2002
2. A. Muller, J. Organometallic Chem., 691, 5782, 2006

Keywords: conformational analysis, DFT, Phosphorus ligand chemistry

P24.07.10

Acta Cryst. (2008). A**64**, C628–629

Ligand substructure validation in macromolecular crystallography using the CSD

Judit E. Debreczeni
AstraZeneca UK, Structural Chemistry, 50S38 Mereside, Alderley Park, Cheshire, SK10 4TG, UK, E-mail：judit.debreczeni@gmail.com

Structural studies of protein-ligand complexes have become essential to modern drug discovery processes. The quality of results delivered by X-ray crystallography has a direct effect on downstream computational chemistry studies, therefore thorough validation - especially that of small molecule substructures - should be considered as a key step in structure determination. However, the final geometry of a ligand is influenced primarily by the restraints applied to it during refinement, as a consequence of the typical resolution range of macromolecular crystal structures. Therefore critical assessment of the initial geometrical parameters should carry equal weight, particularly in a high throughput environment. Validation of chemical structures in the context of macromolecular crystallography