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