MS29 IMPROVING STRUCTURES USING BIO-INFORMATICS
Chairpersons: Philip Eric Bourne, R. Sowdhamini

MS29.26.1
Acta Cryst. (2005). A**61**, C42
How the RCSB Validates PDB Structures
Helen M. Berman, *Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey.* E-mail: berman@rcsb.rutgers.edu

The RCSB validates PDB data while considering the evolving nature of data and nomenclature standards. In order to provide the community with high quality data, the RCSB Protein Data Bank (www.pdb.org) has developed a number of tools that support the deposition and processing of X-ray and NMR structures and that are based upon the mmCIF dictionary [1]. A key feature is the Validation Suite [2], which produces a validation report highlighting close contacts, bond and angle deviations, chirality problems, missing and extra atoms and residues, and distant waters. Authors are encouraged to validate their structures before data deposition.

As the number of structures being determined is constantly increasing, the automation of data validation is extremely important. To reach this end, there needs to be consistency in the syntax and representation of incoming structure data. Tools are being developed to aid this process.

The RCSB also collaborates with wwPDB members to validate the entire PDB archive, and to distribute these data in a way that is most useful to members of the community.

The RCSB PDB is funded by NSF, NIGMS, Office of Science DOE, NLM, NCI, NCRR, NIBIB, and NINDS.

[1] Bourne P.E., Berman H.M., Watenpaugh K., Westbrook J.D., Fitzgerald P.M.D., *Meth. Enz.*, **1997**, 277, 571. [2] Westbrook J., Feng Z., Burkhardt K., Berman H.M., *Meth. Enz.* **2003**, 374, 370.
Keywords: structure validation, structure analysis, mmCIF

MS29.26.2
Acta Cryst. (2005). A**61**, C42
A Systematic Study of Flexibility in Protein Structures and its Implications in Protein Structure Prediction
Adam Godzik, *The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA.* E-mail: adam@burnham-inst.org

Despite some level of overall similarity, proteins sharing the same fold usually display a significant structure variation that prohibits effective use of more distantly related proteins in detailed structure prediction, such as done in comparative modeling.

New generation of proteins structure alignments allow describing and classifying differences in structures between related proteins. The broad survey of structure variations within fold groups performed using FATCAT and POSA algorithms shows that proteins sharing a common fold display strong regularities in how their structure changes in response to mutations and/or substrate or inhibitor binding. Most of the structural variation within any given fold can be described by a small number of parameters, usually a position of a pivot point(s) and an angle(s) of rotation around it.

The results of this survey have important implications for comparative modeling and structural genomics. Flexible templates, rearranged according to the rules independently discovered for a given fold can be used for more accurate comparative modeling. At the same time, relatively small number of structures can be used to characterize structural divergence of large protein families. Specific examples for both applications are discussed.
Keywords: protein structures, structure prediction, flexibility

MS29.26.3
Acta Cryst. (2005). A**61**, C42
Bioinformatics Approach to Characterization of SGNH/GDSL-hydrolases
Biserka Kojić-Prodić[a], Filip Kovačić[a], Ivana Leščić[a], Susanne Wilhelm[b], Sanja Tomić[a], Karl-Erich Jaeger[b], [a]*Rudjer Bošković Institute, 10002-Zagreb, POB 180, Croatia.* [b]*Institute for Molecular Enzyme Technology, Heinrich-Heine-University Düesseldorf, Research Centre Jüelich, D-52428 Jülich, Germany*. E-mail: kojic@irb.hr

The present analysis is aimed to recognize structural elements of SGNH/GDSL family of enzymes with a novel folding type using bioinformatics tools on data of primary and secondary structures. Out of 770 proteins sequences deposited, data of seven different structures of GDSL hydrolases are solved, only; those of the best resolution were selected among twenty available in PDB (including mutants): rhamnogalacturonan acetyleserase from *Aspergillus aculeatus*, thioesterase I from *E. coli*, platelet-activating factor acetylhydrolase IBγ from *Bos taurus*, platelet-activating factor human acetylhydrolase IBβ, and esterase from *Streptomyces scabies*. Two novel enzymes of our interest, esterase from *Pseudomonas aeruginosa* and lipase from *Streptomyces rimosus*, were included in the analysis and compared with GDSL hydrolases of known three-dimensional structures. These two enzymes were recognized as the members of the SGNH/GDSL family with a fold being different from the common α/β hydrolase fold. Alignment of amino acid sequences of SGNH/GDSL hydrolases studied reveals similarity about 20%. However, four blocks of conserved sequence, with one conserved residue in each block (S,G,N,H) are common characteristics.
Keywords: databases, bioinformatics, novel hydrolase fold

MS29.26.4
Acta Cryst. (2005). A**61**, C42
Validation and Classification of Protein Structures
Manfred J. Sippl**,** Christian Weichenberger, Markus Wiederstein, Stefan Suhrer, *Center of Applied Molecular Engineering, University of Salzburg, Austria.* E-mail: sippl@came.sbg.ac.at

We discuss and summarize several new developments regarding the automated validation of protein structures, the decomposition of protein structures into domains, and the classification of protein domains.. As a specific example we present results on the assignment problem of oxygen and nitrogen atoms in the side chains of Glutamine (GLN) and Asparagine (ASN) in some detail. These atoms are difficult to distinguish in the interpretation of electron densities and it is known [1] that approximately 15% to 20% of the assignments in all known structures are incorrect.

We demonstrate how mean force potentials [2] derived from a set of high resolution PDB [3] protein structures can be used to recognize and correct erroneous N/O assignments. Since the potentials are derived from erroneous data sets this is an interesting and challenging problem for the development of potential functions. We show that within a few cycles of potential compilation and error correction the potentials converge to a stable functional form. The detected erroneous assignments fully agree with expert curated assignments [4]. The ASN/GLN flipper is available as a WEB service at http://services.came.sbg.ac.at/flipper.

[1] Hooft R.W.W et al, *Proteins*, 1996, **26**, 363. [2] Sippl, M.J., *Proteins*, 1993, **17**, 355. [3] Berman, H.M. et al, *Nucleic Acids Res., 2000,* **28**, 235. [4] Word, J.M. et al, *J. Mol. Biol.*, 1999, **285**, 1735.
Keywords: structure validation, domain assignment, structure comparison, potential of mean force

MS29.26.5
Acta Cryst. (2005). A**61**, C42-C43
ESPript/ENDscript: Sequence and 3D Information from Protein Structures
Patrice Gouet[a], Xavier Robert[a], Emmanuel Courcelle[b], [a]*Laboratoire de BioCristallographie, IBCP IFR128, France.* [b]*Laboratoire de Biologie Moléculaire et des Relations Plantes Microorganismes, Castanet Tolosan, France.* E-mail: p.gouet@ibcp.fr

The fortran program ESPript has been created to display on a single PostScript figure, multiple sequence alignments adorned with secondary structure elements [1]. A web server is available at http://espript.ibcp.fr/ESPript/ESPript. It has been linked to three web tools: ProDom which identifies protein domains, PredictProtein which predicts secondary structure elements and NPS@ which runs sequence