for both protein kinases (Syk and Gleevec-resistant BCR-ABL) and proteases (Factor VIIa).

Keywords: fragment based drug discovery, structure guided drug discovery, protein kinase drug discovery

MS85.30.5

Acta Cryst. (2005). A**61**, C109

Structural Basis of Multi-functional lipocalin-type Prostaglandin D_2 Synthase

Masashi Miyano[1], Hideo Ago[1#], Kosuke Aritake[2#], Daisuke Irikura[2#], Takashi Kumasaka[1#], Tsuyoshi Inoue[3], Masaki Yamamoto[1], Yoshihiro Urade[2], Osamu Hayaishi[2], [1]*RIKEN Harima Institute at SPring-8, Hyogo 679-5148, Japan.* [2]*Osaka Bioscience Institute, Osaka, 565-0874, Japan.* [3]*Department of Engineering, Osaka University, Osaka 565-0871, Japan.* [#]*Equally contributed authors*. E-mail: miyano@spring8.or.jp

Prostaglandin (PG) D_2 is a natural somnogen inducing non-rapid eye moving (NREM) sleep and an immuno-modulator. PGD synthase (PGDS) is responsible for the production of PGD_2. We determined the crystal structures of lipocalin-type PGDS (L-PGDS) as the first enzymatic lipocalin by using SeMet-MAD phasing at 2.1 Å resolution [1]. L-PGDS has a catalytic architecture similar to the phylogenetically independent PGDS, hematopoietic PGDS, which belongs to a sigma class glutathione S-transferase [2]. L-PGDS is a multi-functional protein which also acts as a hydrophobic ligand-binding protein. The structures with different conformations in two crystal forms suggest the structural basis of the multi-functionalities as well as the mode of the catalytic action [3]. These proposed mechanisms were consistent with the extended site-directed mutagenesis. We present the structural and functional basis of L-PGDS as a multi-functional protein relevant to the biological actions including NREM sleep promotion in the prostanoid cascade [4].

[1] Irikura D., Kumasaka T., Yamamoto M., Ago H., Miyano M., Kubata K.B., Sakai H., Hayaishi O., Urade Y., *J. Biochem. (Tokyo)*, 2003, **133**, 29-32. [2] Kanaoka Y., Ago H., Inagaki E., Nanayama T., Miyano M., Kikuno R., Fujii Y., Eguchi N., Toh H., Urade Y., Hayaishi O., *Cell*, 1997, **90**, 1085–1095. [3] Kumasaka T., Irikura D., Aritake K., Ago H. et al., submitted. [4] Hayaishi O., Urade Y., *Neuroscientist*, 2002, **8**, 12–15.

Keywords: prostaglandin D synthase, lipocalin, protein crystallography

MS86 PROGRAMMING ROBUST CIF AND XML INTO CRYSTALLOGRAPHIC SOFTWARE

Chairpersons: Herbert J. Bernstein, Brian McMahon

MS86.30.1

Acta Cryst. (2005). A**61**, C109

CIF2CML - Automatic Processing of Chemical Crystallography in XML/CML

Peter Murray-Rust[a], Simon Tyrrell[a], [a]*Unilever Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, CB2 1EW, UK.* E-mail: pm286@cam.ac.uk

The CIF data structure and hierarchy (CIF, dictionaries, DDL) is largely isomorphic with XML (document/DOM, schema, XMLSchema) and XML tools can therefore be configured to process data from CIFs. First our CIF2CML toolkits convert CIF documents to XML. The data are then validated structurally and semantically (against the dictionaries) and further converted to Chemical Markup Language (CML) (http://www.xml-cml.org).

The crystal structures in CML can then be stored, chemically validated and transformed using the JUMBO CML library and other CML-aware tools. Among the steps are (a) checks on chemical composition (b) treatment of disorder (c) application of symmetry (d) assignment of bonds and (e) molecules (f) unique chemical identification (IUPAC InChI) (g) calculation of 2D coordinates (h) storage in XML repository to create a structural knowledge base which can be searched for chemical and geometrical concepts.

The approach is highly modular with many hundred interoperable components, designed for use with WebServices (http://wwmm.ch.cam.ac.uk/gridsphere/gridpshere) and workflows such as Taverna (http://taverna.sf.net) and institutional repositories (http://eprints.soton.ac.uk/1633/) with Open data. We argue that Open source and Open data provide a robust high-throughput crystallographic semantic web whose prototype will be demonstrated.

Keywords: CIF, XML, CML

MS86.30.2

Acta Cryst. (2005). A**61**, C109

The Role of Data Ontologies in CIF Deposition and Access

Sydney Hall[a], Nick Spadaccini[b], John Westbrook[c], [a]*School of Biomedical and Chemical Sciences, University of Western Australia.* [b]*School of Computer Science and Software Engineering, University of Western Australia.* [c]*Protein Data Bank, Rutgers University, NJ.* E-mail: syd.hall@bcs.uwa.edu.au

For well over a decade crystallographic data have been routinely submitted to journals and databases as CIFs. During the deposition process, data contained within a CIF can be automatically checked and validated using electronic CIF dictionaries that contain the precise definitions of individual data items. When employed, these ontologies also serve an important role when archiving or accessing deposition data, and within or between crystallographic software applications.

This talk will describe how existing ontologies are employed in CIF deposition and access processes, and what software is currently available to utilize the DDL1 and DDL2 dictionaries during the reading and writing of CIFs. We will describe how the concept of ontological definitions can be extended to automatically provide executable functionality to validation and evaluation processes.

Many aspects of this talk are covered in detail in *International Tables for Crystallography Volume G*[1], which is being launched at this congress.

[1] International Tables for Crystallography, Volume G, *Definition and exchange of crystallographic data*, edited by S.R. Hall & B. McMahon. Heidelberg: Springer, 2005.

Keywords: CIF processing, ontologies, software design

MS86.30.3

Acta Cryst. (2005). A**61**, C109

CIFFOLD: Managing Long Lines in CIF

Kostadin Mitev, Georgi Todorov, Herbert J. Bernstein, *Department of Mathematics and Computer Science, Dowling College, Oakdale, NY 11769, USA.* E-mail: kostadin_m@yahoo.com

Until recently, information in Crystallographic Information File (CIF) format [1] was limited to 80 characters per line and there was no way to represent longer data items and comments faithfully. With the release of CIF version 1.1 [2], the maximum line size has been increased to 2048 characters and a protocol has been specified for folding and unfolding text fields and comments that exceed any given maximum line size. The C/C++ program CIFFOLD implements this line folding/unfolding protocol without loss of the semantic information in the files. This allows new, long-line CIF 1.1 files to be converted to a form suitable for processing by existing software for 80-character line CIF 1.0 files and to recover long-line CIF 1.1 files from CIFs produced by CIF 1.0 software. In addition to folding and unfolding, the software performs logical integrity checks and allows the user to set a variety of options providing control over the tradeoff between faithful versus compact representations. CIFFOLD is part of a package of CIF software for managing IUCr publications that is being upgraded from CIF 1.0 to CIF 1.1 specifications. All the new software in this package will be released under open-source software licenses. Parsers for CIF 1.1 written in C and in Fortran are included in this package.

Work supported in part by funding from the IUCr.

[1] Hall S.R., Allen F.H., Brown I.D., *Acta Cryst.*,1991, A**47**, 655-685. [2] http://www.iucr.org/iucr-top/cif/spec/version1.1

Keywords: CIF, mmCIF, software