reviewers, editors and readers, with practical advice on how to get the best out of the journals. So, if you ever wondered what happens to your article during the submission, review and production stages, want to know more about current publication and editorial procedures, or wish to discover quick ways of accessing and searching the journals, then you should attend this session. Various live demonstrations of the work of the IUCr Editorial Office will be included, with opportunities to ask questions throughout the session.

The topics will include: (*a*) article submission tips; (*b*) figure and scheme preparation for publication - figure resolution, use of colour, accurate colour reproduction; (*c*) using the submission and review system; (*d*) using checkCIF; (*e*) demonstration of tools for editing and viewing CIFs, and advice on preparing CIFs for publication; (*f*) demonstration of editorial systems and production processes; (*g*) article viewing and navigation; (*h*) searching **Crystallography Journals Online**; (*i*) linking - description of the creation of links to bibliographic and structural databases; (*j*) article distribution - e-mail alerting, metadata delivery to third parties, search engines and databases, RSS feeds; (*k*) tracking your paper; (*l*) future developments.

Keywords: journal publishing, IUCr journals, Crystallography Journals Online

OCM03 THE CURRENT STATUS AND FUTURE PROSPECTS OF CIF

Coordinator: D. Brown

OCM03.26.1

Acta Cryst. (2005). A**61**, C127

mmCIF and Modern Macromolecular Structure Determination Software: Status and Perspectives

Ralf W. Grosse-Kunstleve, Paul D. Adams, *Lawrence Berkeley National Laboratory, Berkeley, California, U.S.A.* E-mail: rwgrosse-kunstleve@lbl.gov

As developers in the *Phenix* project [1], we are confronted with *mmCIF* in two ways. Firstly, our algorithms produce results that need to be archived. Secondly, access to information stored in databases, such as the PDB [2], is often invaluable in the development and testing of new methods. In contrast to most traditional, static file formats, *mmCIF* is highly flexible. Therefore we have the opportunity to export parameters and results of ever more complex algorithms in a uniform framework. However, it is non-trivial to import information from *mmCIF* files since their processing requires very sophisticated tools. Unfortunately, in many contexts adequate practical tools are not available. The limitations of traditional software development technology are probably the most important factors giving rise to this situation. Fortunately, many in the crystallographic methods development community have begun a transition to modern software technology. Database developers, most notably at the PDB, have already published comprehensive *mmCIF* libraries. Further development of such libraries in a collaborative effort with an open two-way exchange between the communities has the potential to stimulate a much wider use of *mmCIF* in the future.

[1] Adams P.D., Gopal K., Grosse-Kunstleve R.W., Hung L.-W., Ioerger T.R., McCoy A.J., Moriarty N.W., Pai R.K., Read R.J., Romo T.D., Sacchettini J.C., Sauter N.K., Storoni L.C., Terwilliger T.C., *J. Synchrotron Rad.*, 2004, **11**, 53-55. [2] Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E., *Nucleic Acids Research*, 2000, **28**, 235-242.

Keywords: mmCIF, structure database, software development

OCM03.26.2

Acta Cryst. (2005). A**61**, C127

A Dictionary Approach to Translate Memory Variables from Crystallography Software to mmCIF Items

Heping Zheng, Wladek Minor, *Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22908, USA.* E-mail: dust@iwonka.med.virginia.edu

A major obstacle in building CIF output from crystallography software is to address the relation between the information software can supply and the information mmCIF required. Generally the building process is time-consuming and even more effort is necessary to maintain the source code due to constant changes of the software. We present here, a dictionary-based approach, and the tool used to build such a dictionary. In this approach, the memory-to-mmCIF relation is classified as equivalence, conversion, constant, source conversion, comment, pending or unknown. Each mmCIF item is subject to classification by the developer's examination with the assistance from a domain expert. The CIF Translator Dictionary (CTD) builder is utilizing a dump of all global variables with its value in memory as source of information. This memory dump is in STAR format and allow the CTD developer to do realtime tracking of related variables in memory. Generally it is possible to fetch related variable names in 2 to 5 memory scan by a domain expert. Andfter addressing the relationship between these variables with mmCIF item, a CTD entry will be generated automatically for simple relation, or more infomation will be acquired for complicate relation.

To test the effectiveness of this approach, HKL2000 CTD is built in its initial stage. Automatic completion from HKL2000 memory is performed without human intervention. For more specific tuning toward publication quality CIF after autofill, HKL2000-CIF is also designed as a CIF editor featuring entities editing and providing an evolving amount of wizard procedures that assist further manual examination, and validation before final submission.

Keywords: CIF, CTD, mmCIF

OCM03.26.3

Acta Cryst. (2005). A**61**, C127

Analysis and Visualization of TLS Motion in Proteins using the mmLib Toolkit

Ethan A. Merritt, Jay Painter, *Biomolecular Structure Center and Dept of Biochemistry, University of Washington, Seattle.* E-mail: merritt@u.washington.edu

We have developed a programming library, mmLib [1], which provides a rich set of tools for the import, manipulation, and export of macromolecular structural models described in CIF and mmCIF. Using this toolkit, we are developing higher level tools for visualization and structural/functional analysis. We are in particular working to infer and model functionally important modes of protein flexibility directly from single crystal structures.

TLS (Translation/Libration/Screw) models describe rigid-body vibrational motions of arbitrary objects. A single-group TLS model can be used to approximate the vibration of an entire protein molecule within the crystal lattice. More complex TLS models are broadly applicable to describe inter-domain and other internal vibrational modes of proteins. We are developing a web-based analysis tool, **TLSMD**, that generates optimal multi-segment TLS models. These may be used to analyze the presence and physical significance of TLS motion in existing structures, to guide additional crystallographic refinement, or to generate target models of protein flexibility for use in computational protein-protein or protein-ligand docking.

The interactive graphics program **TLSVIEW** [2] allows visualization of these and other models for rigid-body motion in proteins, using animation and a variety of static representations.

Both tools are applicable to protein structures at any resolution.

[1] Painter J., Merritt E.A., *J. Appl. Cryst*.,2004, **37**, 174-178. [2] Painter J., Merritt E.A., *Acta Cryst*.,2005, **D61**, 465-471.

Keywords: graphics, dynamics, docking computation

OCM03.26.4

Acta Cryst. (2005). A**61**, C127-C128

CIF Operations and Applications at the CCDC

Frank H. Allen, Clare F. Macrae, Sam Motherwell, Lucy H. Purkis, Gregory P. Shields, Robin Taylor, *CCDC, 12 Union Road, Cambridge CB2 1EZ, UK.* E-mail: allen@ccdc.cam.ac.uk

Most major journals require CIF deposition to the CCDC, usually during the submission process, and more than 98% of raw input to the CSD now arrives in CIF form. The CCDC maintains a Supplementary Data Archive of deposited CIFs and, after publication, individual CIFs are made freely available via a simple Web-based request form. The CCDC program enCIFer is available for Web download to check, edit

and visualize CIF format, syntax and content, while the CCDC visualiser, Mercury, is fully CIF-enabled, and the most recent version will represent atomic displacement parameters, a feature that will also be incorporated into enCIFer.

We are now developing software tools that assist the processing of CIFs to the CSD: (a) We use knowledge-based and bond valence sum data to establish crystal connectivity, and then assign chemical bond types algorithmically. The algorithm has a success rate of 86.4% when validated against a test set of 1104 structures, including a significant proportion of challenging metal-organics. (b) Heuristic analysis of CIFs permits resolution of disordered molecules/ions into their discrete components, using atomic occupancy factors, together with any CIF 'group' and 'assembly' fields that may be available. (c) We are generating chemical diagrams directly from the CIF using a variety of software tools and measures of chemical similarity with existing CSD structures.

Keywords: CIF applications, cambridge structural database, data processing and visualisation

OCM03.26.5

Acta Cryst. (2005). A**61**, C128

PDBML: the XML-based Database and its Applications

Haruki Nakamura[a], Nobutoshi Ito[a], Reiko Yamashita[b], Daron, Standley[b], Arno Paehler[b], Atsuro Yoshihara[b], [a]*School of Biomedical Science, Tokyo Medical and Dental University.* [b]*PDBj, BIRD, Japan Science and Technology Agency.* [c]*Institute for Protein Research, Osaka University.* E-mail: harukin@protein.osaka-u.ac.jp

A canonical XML for PDB, PDBML [1], has been developed by PDBj (Protein Data Bank Japan) [2] and RCSB. Its structure is defined in XML schema (http://deposit.pdb.org/pdbML/pdbx.xsd) and all the contents in the PDBML files are now well validated.

We have built a native XML database with this new formant and made it available to public through a web service called xPSSS (http://www.pdbj.org/xpsss). In addition to simple keyword search full XPath searches with the SOAP interface are also implemented for complicated searches and large-scale analyses. The contents of the database are also enhanced; additional data such as the biological and biochemical functions and the experimental details extracted from the literatures and other databases are included.

A few applications have also been developed around the database; (i) a molecular graphics viewer, jV, which can parse the PDBML data [3], (ii) electron density maps for evaluation of the structure quality, (iii) the sequence and structure neighbors defined by maximizing the number of equivalent residues (NER) rather than minimizing conventional RMSD [4].

[1] Westbrook J., et al., *Bioinformatics*, 2005, *in press*. [2] Berman H., et al., *Nat. Struct. Biol.*, 2003, **10**, 980. [3] Kinoshita K., Nakamura H., *Bioinformatics*, 2004, **20**, 1329. [4] Standley D., et al., *Proteins*, 2004, **57**, 381.

Keywords: databases, database manipulation, data representation

OCM03.26.6

Acta Cryst. (2005). A**61**, C128

Use of mmCIF in the Publication of Macromolecular Crystallography Communications

Brian McMahon[a], Peter R. Strickland[a], Howard M. Einspahr[b], J. Mitchell Guss[c], Louise E. Jones[a], [a]*IUCr, 5 Abbey Square, Chester CH1 2HU, UK.* [b]*PO Box 6395, Lawrenceville, NJ 08648-0395, USA.* [c]*School of Molecular and Microbial Biosciences, University of Sydney, NSW 2006, Australia.* E-mail: bm@iucr.org

Acta Crystallographica Section F: Structural Biology and Crystallization Communications was launched by the IUCr in 2005 as a rapid-publication electronic-only journal for communications on the crystallization and structure determination of biological macro-molecules. Structure reports are expected to come initially from structural genomics and protein–ligand studies. For such reports, the IUCr is collaborating with the Protein Data Bank (PDB) to facilitate the deposition and publication procedures by extracting as much information as possible from program output files and capturing additional information in a standard exchange mechanism. The mechanism that is used is the macromolecular crystallographic information file (mmCIF). Authors may send the mmCIF generated during the PDB deposition process directly to the journal office, where automated scripts use it for preparation of validation reports for scrutiny by the referees, and for generation of summary tables of information, suitable for embedding within the article, about the structure determination and refinement. Details of these procedures will be given. Further work will be undertaken to refine the mmCIF dictionary and facilitate its use for publication purposes.

Keywords: IUCr journals, data definition, mmCIF

OCM03.26.7

Acta Cryst. (2005). A**61**, C128

mmCIF and Dictionary Driven Software with the MSD Database Production Pipeline

Kim Henrick, Sameer Valenkar, Harry Boutselakis, Ayzaz Hussain, John Ionides, Adamandia Kapopoulou, Peter Keller, Richard Newman, Jorge Pineda, Antonio Suarez, Jawahar Swaminathan, John Tate, *European Bioinformatics Institute, EMBL Outstation Hinxton, Wellcome Trust Genome Campus Hinxton Cambridge UK.* E-mail: henrick@ebi.ac.uk

The Macromolecular Structure Database (MSD) group, http://www.ebi.ac.uk/msd, based at the European Bioinformatics Institute (EBI) (an outstation of the European Molecular Biology Laboratory EMBL) is a member of the wwPDB (http://www.wwpdb.org) and provides one of the PDB deposition sites. The MSD uses AutoDep4 (http://www.ebi.ac.uk/msd-srv/autodep4/) for PDB depositions and carries out not only the annotation tasks required to produce PDB entries it also provides a stable and clean repository of macromolecular structure data services that allow users to access, search and retrieve structural data. In addition the MSD handles the deposition and archive for Maps from cryo-electron microscopy through the separate deposition interface EMDep, http://www.ebi.ac.uk/msd-srv/emdep/. mmCIF and XML are used throughout the data processing pipeline including the mapping of PDB entries to an in-house extended mmCIF for loading intio the Oracle databases and in the numerous processes that are run on the deposition data to enrich the PDB data with extensive derive information. The MSD export the PDB entries to the RCSB in the standard pdbx dictionary format. This talk will outline part of the processing pipeline used.

Keywords: mmCIF, relational databases, processing pipeline

OCM03.26.8

Acta Cryst. (2005). A**61**, C128

mmCIF Applications at the RCSB Protein Data Bank

Zukang Feng, Helen M. Berman, Huanwang Yang, John D. Westbrook, *RCSB Protein Data Bank, Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey.* E-mail: jwest@rcsb.rutgers.edu

The RCSB Protein Data Bank (www.pdb.org) has developed a variety of software tools to manage mmCIF data. These applications include validating parsers, format translators, database loaders, and data extraction tools. The latter suite of extraction tools, pdb_extract[1], automates the collection of key data items from most popular X-ray and NMR structure determination applications. These data are assembled into an mmCIF data file ready for PDB deposition that can be submitted using the AutoDep Input Tool (ADIT). The use of these tools in automating the PDB deposition and validation process will be described.

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

[1] Yang H., Guranovic V., Dutta S., Feng Z., Berman H.M., Westbrook J.D., *Acta Cryst*., 2004, **D60**, 1833.

Keywords: mmCIF, ontologies, protein structure representation