

MS5-04 Automated Identification of Ligands in Macromolecular Electron Density Maps. Ciaran Carolan, Victor Lamzin. *European Molecular Biology Laboratory, Hamburg outstation, Germany.*
E-mail: ciaran.carolan@embl-hamburg.de

The task of model completion in macromolecular X-Ray Crystallography (MX) is often complicated by the presence of a variety of ligands, either endogenous and associated with the protein or otherwise introduced during sample preparation and crystallisation, that must be identified and subsequently modelled into the density in the most appropriate valid manner. Furthermore, the identification of protein-bound ligands can be used to assign molecular function in structural genomics projects. [1] A number of automated methods exist for the actual modelling of ligands into electron density clusters, (e.g. [2]) but the identification of the entity giving rise to signal in a typical difference density map produced after macromolecular modelling is more difficult. Decisions as to which structures to model into such clusters are most often made by the individual researcher, and thus require significant expertise and experience on his/her part while all the while being subjective and possibly biased. An accurate automated routine for the identification of crystallographic ligands will obviate such issues while also making structure modelling more efficient. It should be particularly useful for distinguishing ambiguous density associated with ligands derived from protein production, purification and crystallisation routines, as well as for researchers carrying out multiple ligand soaks as part of fragment based drug design efforts, for example.

We introduce herein a new technology - to our knowledge the first of its kind in a publicly-available software for macromolecular X-Ray crystallography - for the identification of ligands in macromolecular electron density maps. The method makes use of a weighted combination of mathematical features that match the density cluster to each of the molecules in a database of the 200 most-common ligands in the PDB. The density is considered at multiple contour levels and the full conformational variability of each ligand is explored. The methodology has been incorporated into the recently-released ARP/wARP version 7.3. As the features are rapid to calculate and even faster to compare, results are produced immediately and output to the ARPnavigator Graphical User Interface for analysis. No user input is required aside from specification of the coordinates of the electron density cluster to be analysed. We demonstrate the results of tests on a variety of experimental density maps at various resolutions that highlight the utility and accuracy of the method and discuss likely future developments, including the setup of the software to permit screening of a user's own database of compounds.

- [1] Zarembinski, T. I., Hung, L. W., Mueller-Dieckmann, H. J., Kim, K. K., Yokota, H., Kim, R. & Kim, S. H. (2012) *PNAS* **95** (26), 15189-15193.
[2] Langer, G. G., Evrard, G. X., Carolan, C. G. & Lamzin, V. S. (2012). *J. Mol. Biol.* **419** (3-4), 211-222.

Keywords: identification, crystallographic ligands, automation

MS5-05 New Features and Improvements in Carbohydrate 3D Structure Validation. Thomas Lütteke,^a Denis Mokros,^b Robbie P. Joosten,^c Andreas Dominik,^b Gert Vriend,^d ^a*Justus-Liebig-University Giessen, Germany,* ^b*University of Applied Sciences, Giessen-Friedberg, Germany,* ^c*NKI, Amsterdam, The Netherlands,* ^d*Radboud-University Nijmegen Medical Centre, Nijmegen, The Netherlands*
E-mail: thomas.luetteke@vetmed.uni-giessen.de

More than 5000 entries in the Protein Data Bank (PDB, www.pdb.org), the largest collection of biomolecular 3D-structures) contain carbohydrates. This makes the PDB a valuable resource not only for proteomics but also for glycoscience. Unfortunately, the quality of the carbohydrate moieties is significantly lower than that of the protein parts of glycoproteins or protein-carbohydrate complexes in the PDB; many entries contain errors [1]. There are several reasons for this, one of them being the lack of validation software [2]. Only recently crystallographers became aware of this problem and started to use tools such as PDB Carbohydrate Residue check (pdb-care) [3] to examine the carbohydrate parts. Here we present an updated version of pdb-care, which in addition to the residue notation checks already performed by the former version also detects other problems such as invalid residues within the N-glycan core structure (e.g. α -D-GlcNAc instead of β -D-GlcNAc), missing LINK records, which often result in "1-deoxy" sugars, or superfluous atoms within glycosidic linkages. The results are presented via a new, clearly arranged web interface for human readability, or as a computer-readable xml file to aid automatic validation routines. Suggestions how to correct errors are also included in many instances. These are used e.g. within the PDB_REDO project (www.cmbi.ru.nl/pdb_redo/) [4] to enable an automatic correction of some of the problems. The interface to pdb-care is available at www.glycosciences.de/tools/pdb-care/.

- [1] Lütteke, T. (2009). *Acta Cryst.* **D47**, 655-685. [2] Read, R.J.; *et al.* (2011). *Structure* **19**, 1395-1412, [3] Lütteke, T., von der Lieth, C.W. (2004) *BMC Bioinformatics* **5**, 69, [4] Joosten RP, Joosten K, Murshudov GN, Perrakis A. (2011). *Acta Cryst.* **D68**, 484-496.

Keywords: heteroatom validation; software development; carbohydrate structures