

Poster Presentation

MS22.P19

A new generation of CCP4 monomer library based on Crystallography Open Database

F. Long¹, S. Grazulis², A. Merkys², G. Murshudov¹

¹MRC Laboratory of Molecular Biology, Structural Studies Division, Cambridge, UK, ²Vilnius University Institute of Biotechnology, Vilnius, Lithuania

The use of prior chemical knowledge such as bond lengths, bond angles about constituent blocks of macromolecules and ligands is an essential part of macromolecular crystal structure analysis. One of the reliable sources of such chemical knowledge is small molecule database where small molecule crystal structures have been analysed against high-resolution, high-quality experimental data. Furthermore, vast amount of data in small molecule database provide comprehensive coverage of flexible chemical environment and enable proper statistical analysis to avoid biased representation of those chemical properties. This presentation describes our work on organization of the data from open-access and daily-updated small molecule database, Crystallography Open Database(COD) [1], into a new generation of CCP4 monomer library (Dictionary), a container of prior chemical knowledge [2]. In order to describe specific environment atoms are in, they are classified into different atomic types based on local graphs and some basic chemical properties of atoms. This scheme can be applied to any small molecule databases. The atom types, and values of bond lengths and bond associated with them, are further clustered into a hierarchical tree and an isomorphism-mapping algorithm is implemented to facilitate fast search among a large number of atom types (typically several millions). This also provides a mechanism to derive reliable values for bond lengths and angles of novel ligands. Metal and non-organic atoms are treated differently with organic ones. The original data in COD are curated using several criteria and further statistical analysis on derived values of bond lengths and angles are allow to extract reliable chemical information from such databanks as COD. There are several software tools associated with new dictionary including 1) generate "ideal" bond lengths and angles for unknown ligand; 2) generate starting coordinates to represent one of the conformation of the ligand under consideration.

[1] S. Grazulis, D. Chateigner, R.T. Downs, et al., "Crystallography Open Database – an open-access collection of crystal structure", *J. Appl. Cryst.* 42, (2009), pp 726-729., [2] A. Vagin, R. Steiner, A. A. Lebedev, et al., "REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for it use", *Acta Cryst. D60*, (2004), 2184-2195.

Keywords: CCP4 monomer library, Refinement, small molecule database