The Improvements of polymer representation in the COD using quotient graphs

Y. Rozdobudko[1], A. Merkys[1], A. Vaitkus[1], S. Gražulis[1,2]

[1]Vilnius University, Life Sciences Center, Institute of Biotechnology, Sector of Crystallography and Chemical Informatics, Saulėtekio av. 7, LT-10257 Vilnius, Lithuania, [2] Vilnius University, Faculty of Mathematics and Informatics, Naugarduko 24, LT-03225 Vilnius, Lithuania

yaroslav.rozdobudko@gmc.vu.lt

The Crystallography Open Database (COD) [1] contains a number of structures that are infinite in their nature. So-called *polymers*, more precisely *coordination polymers* and *Metal-Organic-Frameworks (MOF)* [2]. Such molecules are used in design and development of new materials, and it is important to understand and analyze them. One such tool *cif_molecule* [3] reconstructs a molecule from an asymmetric unit (AU) and can distinguish if the molecule is a *polymer*. To better analyze, study, and apply existing algorithms [4], we need to have a finite representation of these infinite molecules. The theory for such representation was first introduced in [5] through the use of *quotient graphs*, which captures translationally equivalent bonds and atoms. While known for a long time, it is rarely used in crystallographic software. This work introduces *quotient graphs* to the *cif_molecule* algorithm and to its output, which guarantees *polymer* detection and dimensionality.

Implementation of *cif_molecule* was revised and *quotient graphs* are now created during the reconstruction of a molecule, which allows the correct determination of polymer molecules, and their dimensionality. *Quotient graphs* are added to the output of *cif_molecule* with data items from the *topoCIF* dictionary, which allows us to use it in further analysis, and it is accepted by other software like *Jmol [6]*. Additional software was developed that utilizes *quotient graphs:* to convert quotient graph in *CIF* to other formats – *cif2qg* converts to *graphviz [7]* dot format, and to *cgd* format for *Systre* [8] input. *cif_polymer_multiplicity* to calculate polymer multiplicity [4] (number of self-penetrated molecules in the crystal).

cif_molecule (release 3.11.0) with *quotient graphs* was run on the *COD* revision 291294, all 511329 structures were processed. In addition, *cif2qg* was run on the results to convert them to the *dot* format and to *cgd*, *Systre* was run to determine the *RCSR symbols* [9], and cif_polymer_multiplicity was run to calculate the multiplicity. The repository was created that contains the results of the calculations, and rules defined as Makefiles, which allow reproducible calculations on different revisions of *COD*, or other COD-like databases. A website was created to search and display the results, available at https://polymers.crystallography.net/polymers-viewer.

Quotient graphs are a useful tool for creating a finite representation of infinite molecules. Now we can guarantee the correct representation of *polymers.* Expressing *quotient graphs* with data items defined in the official *topoCIF* dictionary allows calculation results to be easily used by other programs.

[1] Gražulis, S., Daškevič, A., Merkys, A., Chateigner, D., Lutterotti, L., Quirós, M., Serebryanaya, N. R., Moeck, P., Downs, R. T., & Le Bail, A. (2011). Crystallography open database (cod): An open-access collection of crystal structures and platform for world-wide collaboration. Nucleic Acids Research, 40(D1), D420–D427. https://doi.org/10.1093/nar/gkr900

[2] Batten, S. R., Champness, N. R., Chen, X.-M., Garcia-Martinez, J., Kitagawa, S., Öhrström, L., O'Keeffe, M., Suh, M. P., & Reedijk, J. (2013). Terminology of metal–organic frameworks and coordination polymers (IUPAC recommendations 2013). Pure and Applied Chemistry, 85(8), 1715–1724. https://doi.org/10.1351/pac-rec-12-11-20

[3] Gražulis, S., Merkys, A., Vaitkus, A., & Okulič-Kazarinas, M. (2015). Computing stoichiometric molecular composition from crystal structures. Journal of Applied Crystallography, 48(1), 85–91. https://doi.org/10.1107/s1600576714025904

[4] Gao, H., Wang, J., Guo, Z., & Sun, J. (2020). Determining dimensionalities and multiplicities of crystal nets. npj Computational Materials,6(1). https://doi.org/10.1038/s41524-020-00409-0

[5] S. J. Chung et al. Acta Crystallographica Section A Foundations of Crystallography, 40(1):42–50, 1984. doi:10.1107/s0108767384000088.

[6] R. M. Hanson. Journal of Applied Crystallography, 43(5):1250–1260, 2010. doi:10.1107/s0021889810030256.

[7] J. Ellson et al. Graphviz and Dynagraph — Static and Dynamic Graph Drawing Tools, pages 127–148. Springer Berlin Heidelberg, 2004.

[8] O. Delgado-Friedrichs et al. Acta Crystallographica Section A Foundations of Crystallography, 59(4):351–360, jun 2003. doi:10.1107/s0108767303012017.

[9] M. O'Keeffe et al. Accounts of Chemical Research, 41(12):1782–1789, 2008. doi:10.1021/ar800124u.