Improvements of the polymers representation in COD using quotient graphs

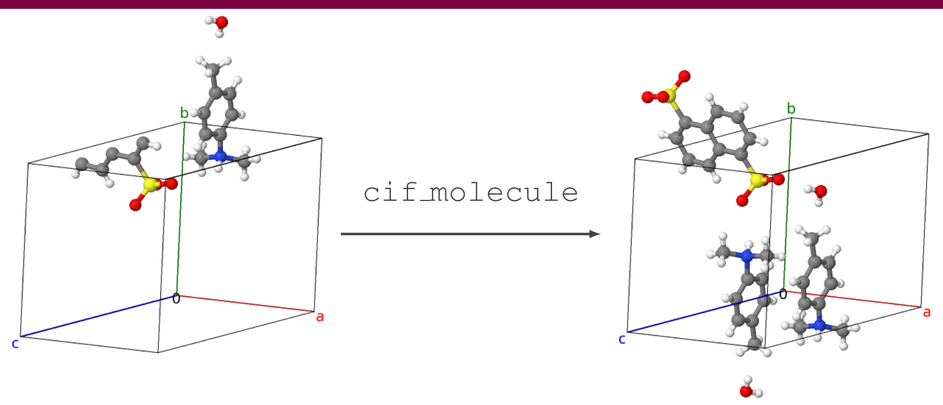
Yaroslav Rozdobudko[1], Andrius Merkys[1], Antanas Vaitkus[1], Saulius Gražulis[1,2]

[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

Email: yaroslav.rozdobudko@bti.stud.vu.lt

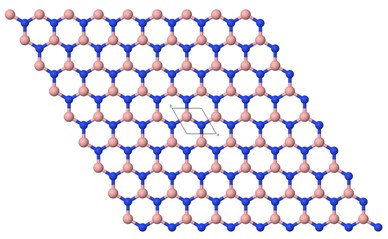
`cif_molecule` molecule reconstruction



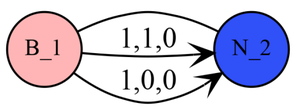
Asymmetric unit *Stoichiometrically correct molecule*

- Reconstructs molecule from the *asymmetric unit (AU)*.
- Preserves stoichiometry.
- Part of the open-source `cod-tools` package [1].

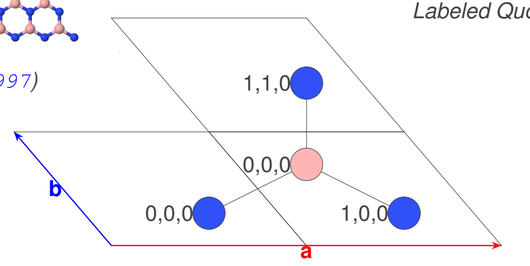
Infinite net converting to labeled quotient graph



Infinite net (COD ID 9008997)



Labeled Quotient Graph (COD ID 9008997)



Constructing Labeled Quotient Graph (LQG)

- Polymer molecules can be represented as an infinite net that spans across crystal.
- We do not know the size of the *AU* and cannot search infinitely.
- To create a finite representation of an infinite net, we use *Labeled Quotient Graphs (LQG)* [2].
- Algorithm for constructing *LQG* from an infinite net:
  1. Choose a coordinate system, select basis vectors.
  2. Assign initial labels to the original vertices $(0, 0, 0)$.
  3. If the net has an edge from point $P(m, n)$ to point $R(p, q) \Rightarrow$ there is a directed edge in *LQG* with label $(p - m, q - n)$ from P to R.

Implementation of the *LQG* in `cif_molecule`

- Polymer representation as *LQG* was implemented in the *cif_molecule* and was made available in *cod-tools* release *3.11.0* [3].
- *LQG* information was added to the *cif_molecule* output expressed using *topoCIF* dictionary data items. This allows *LQG* recognition by existing software like *Jmol* [4].
- *LQG* is used to calculate dimensionality of infinite molecules [5].
- Additional programs introduced:
  - *cif2qg* to convert *LQG* from *CIF* to other formats like *graphviz* [6].
  - *cif_polymer_multiplicity* to calculate *multiplicity* [5] of a polymer (how many identical molecules in the crystal are self penetrated).
- Stoichiometry is preserved in reconstructed molecules. No more need to specify `--max-polymer-span` and `--max-polymer-atoms` parameters for infinite molecules.

References

[1] S. Gražulis, A. Merkys, A. Vaitkus, and M. Okulič-Kazarinas, "Computing stoichiometric molecular composition from crystal structures," *Journal of Applied Crystallography*, vol. 48, no. 1, pp. 85–91, 2015. [Online]. Available: https://doi.org/10.1107/S1600576714025904

[2] S. J. Chung, T. Hahn, and W. E. Klee, "Nomenclature and generation of three-periodic nets: the vector method," *Acta Crystallographica, Section A: Foundations of Crystallography*, vol. 40, no. 1, pp. 42–50, 1984. [Online]. Available: https://doi.org/10.1107/S0108767384000088

[3] Y. Rozdobudko, A. Vaitkus, A. Merkys, and S. Gražulis. (2025) cod-tools, version 3.11.0. svn://www.crystallography.net/cod-tools/tags/v3.11.0 Last accessed: 2025-05-08.

[4] R. M. Hanson, "Jmol – a paradigm shift in crystallographic visualization," *Journal of Applied Crystallography*, vol. 43, no. 5, pp. 1250–1260, 2010. [Online]. Available: https://doi.org/10.1107/S0021889810030256

[5] H. Gao, J. Wang, Z. Guo, and J. Sun, "Determining dimensionalities and multiplicities of crystal nets," *npj Computational Materials*, vol. 6, no. 1, 2020. [Online]. Available: https://doi.org/10.1038/S41524-020-00409-0

[6] J. Ellson, E. R. Gansner, E. Koutsofios, S. C. North, and G. Woodhull, *Graphviz and Dynagraph — Static and Dynamic Graph Drawing Tools*. Springer Berlin Heidelberg, 2004, pp. 127–148. [Online]. Available: https://doi.org/10.1007/978-3-642-18638-7_6

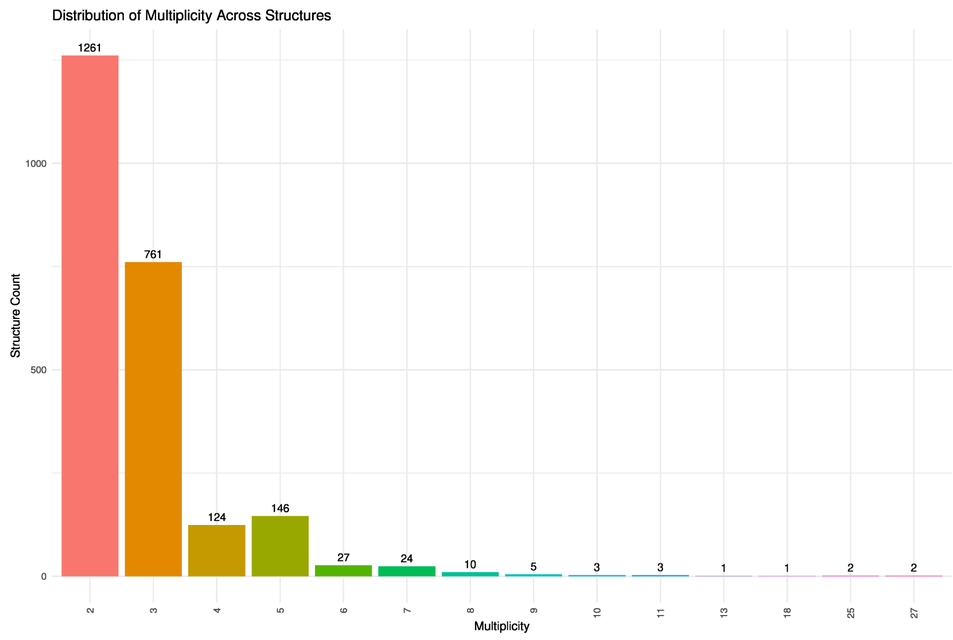
[7] M. O'Keeffe, M. A. Peskov, S. J. Ramsden, and O. M. Yaghi, "The reticular chemistry structure resource (RCSR) database of, and symbols for, crystal nets," *Accounts of Chemical Research*, vol. 41, no. 12, pp. 1782–1789, 2008. [Online]. Available: http://yaghi.berkeley.edu/pdfPublications/09RCSRaccChem.pdf

[8] O. Delgado-Friedrichs and M. O'Keeffe, "Identification of and symmetry computation for crystal nets," *Acta Crystallographica, Section A: Foundations of Crystallography*, vol. 59, no. 4, pp. 351–360, 2003. [Online]. Available: https://doi.org/10.1107/S0108767303012017

Polymer detection in the *COD*

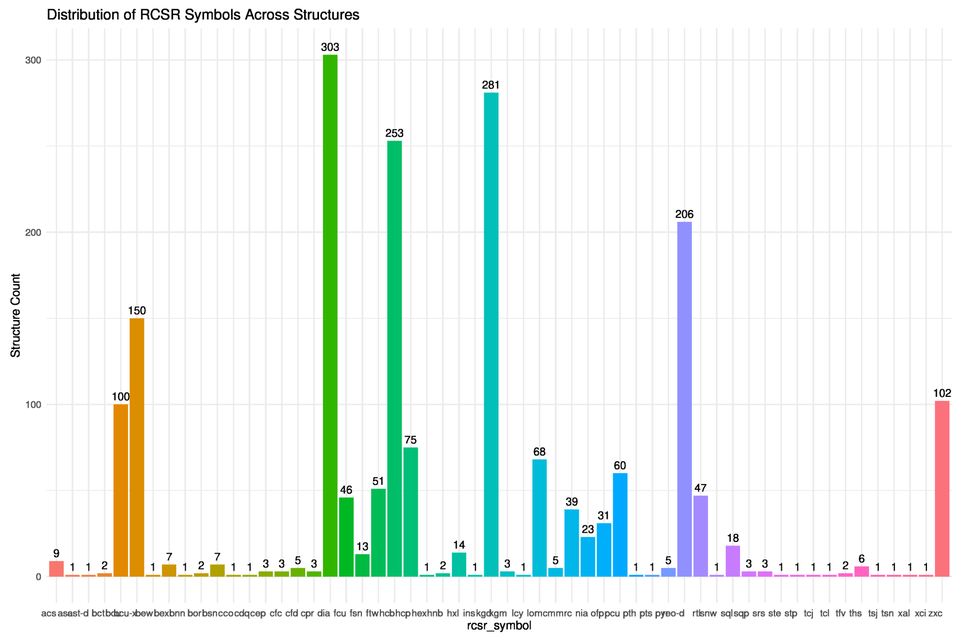
- Calculations were performed on the *Crystallography Open Database (COD)* revision *291294*. Polymers were detected and expressed as *LQG* using *cif_molecule*, *multiplicities* and *RCSR* symbols [7] were calculated using *Systre* [8]. Repository for calculations was created at `svn://databases.crystallography.lt/quotient-graphs`. Website to view results `http://polymers.crystallography.net/polymers-viewer`.
- All *CIF* outputs converted to *graphviz dot* format.
- Comparison made with calculations done on the *COD* with *cif_molecule* without *LQG* at `svn://crystallography.net/polymers-in-COD`.
- For *cif_molecule* 5 structures were not calculated (it takes long time to run).
- Multiplicity calculations were not run for 4669 files.

Multiplicities in the COD



- 2370 polymers were identified as *self-penetrated*, majority of structures have 2 molecules in the crystal.

RCSR symbols in the COD



- *Systre* identified 56 unique *RCSR* symbols. Most common are **hcb**, **dia**, and **kgm**.

Conclusions

- *LQG* allows creating finite representation of infinite polymers.
- Polymers and their dimensionality are correctly identified using *cif_molecule* with *LQG*.
- Output using *topoCIF* dictionary data items allows using *LQG* in other software.

Yaroslav Rozdobudko has no conflict of interest.
Antanas Vaitkus has no conflict of interest.
Andrius Merkys has no conflict of interest.
Saulius Gražulis has no conflict of interest.

On-line version of the poster:
`https://bit.ly/4j4wO6x`