The Journal of Open Source Software

# GEMMI: A library for structural biology

**Marcin Wojdyr**[1]

**1** Global Phasing Ltd., Cambridge, UK

**DOI:** 10.21105/joss.04200

**Software**

- Review
- Repository
- Archive

**Editor:** Kelly Rowland

**Reviewers:**

- @jamesrhester
- @dominiquesydow

**Submitted:** 19 January 2022
**Published:** 30 November -001



## Summary

GEMMI is a library, accompanied by a set of small programs, developed primarily for use in the field of macromolecular crystallography (MX). Parts of this library are useful also in structural bioinformatics and in chemical crystallography.

The library covers three main areas, which overlap and contain common elements, such as handling of the crystallographic symmetry.

The first area is working with structural models of macromolecules. This includes reading and writing files in the PDB and mmCIF formats, analyzing and modifying models and working with restraint dictionaries. The dictionaries are used to restrain geometry of a model using prior knowledge about monomers in the model.

The second area is working with crystallographic data – experimentally observed reflections. This includes reading and writing files in the MTZ and mmCIF formats and performing various operations on the reflections.

The third area is working with electron density maps – real or complex values on a 3D grid. Electron density can be calculated from both the structural model and experimental data. The functionality here includes reading and writing files in the MRC/CCP4 map format, analysing and modifying the density, and using the fast Fourier transform to switch between the so-called direct space and the reciprocal space.

GEMMI is written in C++. It has Python bindings and, for selected functions, also C and Fortran bindings. Interestingly, the library can be compiled to WebAssembly for use in web applications. For example, UglyMol (Wojdyr, 2017) uses it to read MTZ files inside the web browser.

## Statement of need

GEMMI is funded by two organizations that develop MX software: CCP4 and Global Phasing Ltd. The aim is to deliver functionality needed in other projects of these organizations. Initially, the focus was on working with the PDBx/mmCIF file format, then the scope was expanded to other areas.

The library has a significant overlap with other libraries used in this field: CCTBX (Grosse-Kunstleve et al., 2002) and Clipper (Cowtan, 2003). But even when two functions from different libraries have similar purpose, they usually differ in some aspects, for example, by making a different trade-off between the speed of calculations and the accuracy of results, or between the simplicity of the code and the number of provided options.

GEMMI is used in a number of projects, including autoBUSTER (Bricogne et al., 2020), CCP4i2 (Potterton et al., 2018), CCP4 Cloud (Krissinel et al., 2018), Servalcat (Yamashita et al., 2021), an analysis of covalent linkages (Nicholls et al., 2021), reciprocalspaceship (Greisman et al., 2021), and many others.

## Acknowledgements

The library contains contributions from Keitaro Yamashita, Claus Flensburg and other users. It uses third-party libraries: PocketFFT for Fast Fourier Transform, KSW2 (Li, 2018) for sequence alignment, QCProt (Liu et al., 2010) for structure superposition, Cromer-Liberman routine from Larch (Newville, 2013), PEGTL for creating PEG parsers, as well as sajson, stb_sprintf, fast_float, tinydir, zlib and pybind11.

This project would not be possible without Eugene Krissinel, Gérard Bricogne and Garib Murshudov, who initiated it, and without many discussions with users and with colleagues from Global Phasing and CCP4.

## References

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vonrhein, C., & Womack, T. O. (2020). *autoBUSTER*. Global Phasing Ltd., Cambridge, UK. https://www.globalphasing.com/buster/

Cowtan, K. (2003). The Clipper C++ libraries for X-ray crystallography. *IUCr Computing Commission Newsletter*, *2*(4), 9.

Greisman, J. B., Dalton, K. M., & Hekstra, D. R. (2021). reciprocalspaceship: a Python library for crystallographic data analysis. *Journal of Applied Crystallography*, *54*(5), 1521–1529. https://doi.org/10.1107/S160057672100755X

Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W., & Adams, P. D. (2002). The Computational Crystallography Toolbox: crystallographic algorithms in a reusable software framework. *Journal of Applied Crystallography*, *35*(1), 126–136.

Krissinel, E., Uski, V., Lebedev, A., Winn, M., & Ballard, C. (2018). Distributed computing for macromolecular crystallography. *Acta Crystallographica Section D*, *74*(2), 143–151. https://doi.org/10.1107/S2059798317014565

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100.

Liu, P., Agrafiotis, D. K., & Theobald, D. L. (2010). Fast determination of the optimal rotational matrix for macromolecular superpositions. *Journal of Computational Chemistry*, *31*(7), 1561–1563.

Newville, M. (2013). Larch: An analysis package for XAFS and related spectroscopies. *Journal of Physics: Conference Series*, *430*, 012007.

Nicholls, R. A., Wojdyr, M., Joosten, R. P., Catapano, L., Long, F., Fischer, M., Emsley, P., & Murshudov, G. N. (2021). The missing link: covalent linkages in structural models. *Acta Crystallographica Section D*, *77*(6), 727–745. https://doi.org/10.1107/S2059798321003934

Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., Jenkins, H. T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S. J., Nicholls, R. A., Noble, M., Pannu, N. S., Roth, C., Sheldrick, G., Skubak, P., Turkenburg, J., … Wojdyr, M. (2018). CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D*, *74*(2), 68–84. https://doi.org/10.1107/S2059798317016035

Wojdyr, M. (2017). UglyMol: A WebGL macromolecular viewer focused on the electron density. *Journal of Open Source Software*, *2*(18), 350. https://doi.org/10.21105/joss.00350

Yamashita, K., Palmer, C. M., Burnley, T., & Murshudov, G. N. (2021). Cryo-EM single-particle structure refinement and map calculation using Servalcat. *Acta Crystallographica Section D*, *77*(10), 1282–1291. https://doi.org/10.1107/S2059798321009475