MS46 Reproducibility in crystallography

MS46-01
The Role of Archiving Raw Diffraction Data in Ensuring Reproducibility in Crystallography
L.M.J. Kroon-Batenburg [1]
[1]*Utrecht University - Utrecht (Netherlands)*

Abstract
Structure determination with (X-ray) crystallography from single crystals has become a seemingly routine technique. The workflow is rather well established in the community. Its starts with recording diffraction images on an area detector, finding positions of diffraction spots and indexing of the unit cell, predicting and integrating all spots (to a certain resolution) while subtracting the (incoherent) background, data reduction including scaling of the Bragg intensities by matching independent observations, solving the structural model (basically solving the phase problem) and refining (by least-squares or maximum likelihood methods) the model against the intensities or structure factors, improving the models through chemical knowledge or rebuilding in the electron density. A range of metrics is used to show the reliability of the results and the final model is validated through checkCIF or a PDB validation report. In all these steps choices are made by the software and the researcher, based on prior knowledge and experience. However, the choices are rarely documented. Why did the researcher decide to cut off the data at a certain resolution? Could the limit be chosen differently? Did the researcher index the reflection correctly? Is the lattice symmetry correct? Do the structure factors have this symmetry? Did the researcher miss additional lattices, so dealing in fact with a non-merohedral twin? Are some features in the diffraction pattern neglected, like diffuse scattering or incommensurate modulation? In chemical crystallography many efforts were made to check the validity of the model and refinement protocols of the (unmerged) structure factor data through checkCIF, ensuring its reliability when permanently archived, while in macromolecular crystallography PDB validation reports and the OneDep system guarantee the correct version of record of the publication. However, the validation metrics may not show the problems mentioned above.
True reproducibility in crystallography can only be reached if all steps and choices in the workflow can be repeated. Open Science policies require that no research data should be lost but should be made available to the research community according to the FAIR principles. These are compelling reasons for the deposition of raw data in freely accessible repositories at the moment of publication. See also an editorial on FAIR diffraction data coming to Macromolecular Crystallography in Acta Cryst. D, Acta Cryst. F, IUCrJ and J. Appl. Cryst. in 2019 [1]. Recent developments in archiving capabilities such as large-scale repositories have made this feasible.
IUCrData journal will be opening a new section: Raw Data Letters. The papers in this section will describe interesting features in raw data sets [2]. The publication promotes retrieval by other scientists for purposes such as reanalysis by newer methods or protocols and for software and methods development and supports our aim of achieving true reproducibility in crystallography.

References
[1] Editorial on FAIR diffraction data coming to Macromolecular Crystallography in Acta Cryst. D, Acta Cryst. F, IUCrJ and J. Appl. Cryst. in 2019 (https://doi.org/10.1107/S2053230X19005909).
[2] https://iucrdata.iucr.org/x/services/journal_news.html