Integrated Resource for Reproducibility in Molecular Crystallography: experiences of the first five years

Marek Grabowski[1], Marcin Cymborowski[2], David Cooper[3], Przemyslaw Porebski[4], Wladek Minor[5]

[1]University of Virginia [2]University of Virginia, [3]University of Virginia, [4]Molecular Physiology and Biological Physics, Univ of Virginia, [5]University of Virginia

marek@iwonka.med.virginia.edu

Preservation and public accessibility of primary experimental data are cornerstones necessary for the reproducibility of empirical sciences. In the field of molecular crystallography, many journals now recommend that authors of manuscripts presenting a crystal structure should deposit their primary experimental data (X-ray diffraction images) to one of the dedicated resources created in recent years. We present the Integrated Resource for Reproducibility in Molecular Crystallography (IRRMC) and describe several examples of a crucial role that diffraction data can play in improving previously determined protein structures. In its first five years, several hundred crystallographers have deposited data from over 5700 diffraction experiments performed at over 60 different synchrotron beamlines or home sources all over the world. In addition to improving the resource and curating submitted data, we have been building a pipeline for the extraction or, in some cases, reconstruction of the metadata necessary for seamless automated processing. Preliminary analysis shows that about 95% of the archived data can be automatically reprocessed. A high rate of reprocessing success shows the feasibility of automated metadata extraction and automated processing as a validation step for the correctness of raw diffraction images. The IRRMC is guided by the Findable, Accessible, Interoperable, and Reusable data management principles. Data from IRRMC have enabled several novel research projects.