

Integrated management, analysis and interpretation of data along the crystallographic experimental workflow

Clemens Vonnrhein¹, Raphael Bourgeois², Claus Flensburg¹, Peter Murphy², Rasmus Fogh¹, Peter Keller¹, Andrew Sharff¹, Ian Tickle¹, Jose A. Marquez² & Gérard Bricogne¹.

¹ *Global Phasing Ltd, Sheraton House, Castle Park, Cambridge, CB3 0AX, UK.*

² *EMBL Grenoble, 71 Avenue de Martyrs, 38000. Grenoble, France*

Abstract

The systematic introduction of automation into structural biology over the last decades has enabled the study of ever more demanding targets. However, this has created challenges for data management and integration. Typically, crystallography experiments are carried out across multiple sites, including home laboratories and synchrotron facilities, and relatively long periods of time elapse before results and data are deposited in one of the worldwide PDB sites. While some of these facilities use data management systems, they are only poorly interconnected. Moreover, new applications like fragment screening and serial crystallography, producing large amounts of data in a short time, are becoming mainstream. Here we will show how integrated data management and sample tracking along the crystallographic experimental workflow can be performed by implementing automated communication systems between home laboratory and synchrotron information systems, like CRIMS1 and ISPyB2, that can then be connected with automated data processing and analysis tools using Global Phasing's Pipedream3 pipeline to support the entire high-throughput compound screening process. Finally, we will discuss how deposition tools and recent additions to the PDBx/mmCIF dictionary can be exploited to produce enriched and more consistent descriptions of the crystallographic experiment, of the data processing results and of the final refined model for depositions into databases such as the PDB.

References:

- 1) Cornaciu, I. et al. The Automated Crystallography Pipelines at the EMBL HTX Facility in Grenoble. *Journal of Visualized Experiments*. (172), 62491, doi: 10.3791/62491 (2021).
- 2) Delageniere, S. et al. ISPyB: an information management system for synchrotron macromolecular crystallography. *Bioinformatics*. 27 (22), 3186–3192, doi: 10.1093/bioinformatics/btr535 (2011).
- 3) Sharff A, Keller P, Vonnrhein C, Smart O, Womack T, Flensburg C, Paciorek W and Bricogne G (2011). Pipedream, version 1.3.1, Global Phasing Ltd, Cambridge, United Kingdom