

Online automated structure solution from multiple datasets

Santosh Panjekar¹, Manfred Weiss², Victor Lamzin³

¹Macromolecular Crystallography, Australian Synchrotron, Clayton, Australia, ²Macromolecular Crystallography (HZB-MX), Helmholtz-Zentrum Berlin für Materialien und Energie, Berlin, Germany, ³EMBL Hamburg Outstation, Hamburg, Germany
E-mail: santosh.panjekar@synchrotron.org.au

With the advent of the high brilliance and small spot sizes beam available from the modern synchrotron sources, the fast detector (i.e PILATUS, EIGER) and automated sample changer has accelerated the collection of X-ray data from protein crystals significantly. This provides possibility to collect large number of datasets from multiple protein crystals and to improve multiplicity and anomalous signal using appropriate scaling, which is beneficial not only for SAD phasing but also for MAD experiments. Keeping this in mind, two multiple datasets evaluation modes have been developed for various experimental phasing.

The online automated structure solution from multiple datasets at the Auto-Rickshaw (AR) [1] server can be carried out using either the 'multiscale' or 'multidata' mode. When a number of scaled datasets are available in mtz format, 'multiscale' mode can be chosen. Other mode ('multidata') can be opted when raw images of several datasets have been integrated but not scaled. In either case, multiple AR job is launched using fast-version or advanced version of AR [2,3] by providing data directory upload for each set (Peak, inflection, high energy remote and/or low energy remote) and by providing necessary input parameters. In 'multidata' mode, back-end python engine makes decision for selection of useful data sets for combination based on its isomorphism and radiation damage for scaling and anomalous signal to noise ratio for phasing. In order to streamline, the above process has been fully automated that just requires directory path for the processed intensity data to upload (powered by dropbox). Selection of the data is performed based on its unit cell and space group further followed by analysis of each dataset based on radiation damage and anomalous signal. If the anomalous signal of individual dataset is above the threshold, then the dataset is pushed for post data evaluation at the AR server using fast or advanced version otherwise the dataset is selected for clustering, merging and scaling. Similarly, each clustered and scaled data is analysed and pushed for data evaluation at the AR server for SAD phasing. Where as for SIRAS or MAD phasing, decision is made based on isomorphous/dispersive and anomalous signal. Complete data analysis is documented on the fly as a PPTX file for readymade use for presentation.

The experimental phasing protocols of AR have been updated with number of features such as substructure determination using MR method where single or multiple atoms (i.e I3C or Tantalum bromide cluster) is used as a search model. In difficult cases, both approaches direct method and MR method are used for substructure determination. Automated model building is performed with SAD refinement.

AR development and workflow of the data analysis will be presented along with multiple datasets for S-SAD/SIRAS/MAD phasing.

[1] <http://www.embl-hamburg.de/Auto-Rickshaw>

[2] Panjekar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M. S. & Tucker, P. A (2005) Acta Cryst., D61, 449-457.

[3] Panjekar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M. S. & Tucker, P. A (2009) Acta Cryst. D65, 1089-1097.

Keywords: [Auto-Rickshaw](#), [experimental phasing](#), [SAD](#)