MS11-05 | MACHINE LEARNING FOR EXPERIMENTAL PHASING IN MX

Vollmar, Melanie (Diamond Light Source Ltd, Didcot, GBR); Parkhurst, James (Diamond Light Source Ltd, Didcot, GBR); Jaques, Dominic (Diamond Light Source Ltd, Didcot, GBR); Elliott, Jenna (Univeristy of Oxford, Oxford, GBR); Basle, Arnaud (Institute for Cell and Molecular Biosciences, Newcastle upon Tyne, GBR); Murshudov, Garib (MRC Laboratory of Molecular Biology, Cambridge, GBR); Waterman, David (CCP4, Didcot, GBR); Winn, Martyn (STFC, Didcot, GBR); Evans, Gwyndaf (Diamond Light Source Ltd, Didcot, GBR)

Large-scale facilities such as synchrotrons face an increasing demand for computational resources due to rapid increases in data rates, data volumes, and growing interest in real-time feedback to researchers during their experiments. The efficient use of these high performance computing resources is an important factor in the overall efficiency of the facility. Machine learning applications geared towards user assistance offer great opportunities to improve experiment outcomes. These can be trained using a large archive of previously-collected data, enabling more efficient experiments to be performed by rapidly providing information that can be used for decision making.

Here we describe a method to estimate the likelihood of success when determining a protein structure by experimental phasing using X-ray crystallography. Such a predictive tool has the potential to accelerate the structure determination process, as data can be rapidly assessed regarding its usefulness and either discarded or extended until an appropriate data set of sufficient quality is available. In this way, computational and synchrotron beamtime resources can be focused on the most promising data sets.

We have developed applications based on statistical and machine learning methods to predict the success of X-ray crystallographic experimental phasing outcomes ~95% accuracy and to assist in identifying electron density maps of sufficient quality for model building.