Contribution ID: **33**   Type: **not specified**

Development of new data processing methods for serial time-resolved crystallography

Monday 22 September 2025 18:00 (10 minutes)

To obtain the highest quality electron density maps, from a time-resolved serial crystallography experiment, it is crucial to accurately group the observed intensities according to the structure of the protein that generated those intensities. In an ideal world, each unique protein crystal structure would give rise to a set of well defined intensities, which would enable the intensities to be grouped easily. Unfortunately, due to many real world limitations, this is not the case: 1) The crystal is an average of many protein structures 2) Experimental effects such as unequal soaking, unequal laser/ x-ray exposure etc. 3) A range of crystal sizes causing inconsistent intensities which require rescaling. These effects give rise to uncertainty in the observed intensity values. Furthermore, only a minor proportion of hkls are affected by the change in protein structure, and of those hkls, the intensity changes are often subtle.

Despite all the challenges, a statistically rigorous approach is required to accurately group observed intensities according to the structure of the protein that generated those intensities. One approach utilises Bayesian statistics, a probability based approached used in many areas such as weather forecasting, econometrics and natural language processing. In principle, some Bayesian methods such as naive Bayes and maximum likelihood estimator is able to assess the probability a set of intensities were derived from a protein structure. In practice, the intensities have high uncertainty values and there are many nuances as to how this data processing pipeline should be set up, sometimes sacrificing flexibility for increased confidence in the results.

Co-authors: James Beilsten-Edmands, Mike Hough, Graeme Winter

Presenter: TANG, Rachel (Diamond Light Source)

Session Classification: Posters