Structural Models and Refinement for Multi-Dataset Experiments

Lena Lassinantti, Nicholas M Pearce

Physics, Chemistry and Biology (IFM), SciLifeLab, Linköping University, Sweden.

Structural studies of macromolecules in their apo forms or in complexes with ligands have long been used in biomedical applications. However, we are now entering an era of high-volume data generation with the advent of experiments like time-resolved (TR) crystallography, where we can observe complex chemical processes directly *in crystallo*. Yet, the computational pre- and post-processing of structural data gathered from these methods remains underdeveloped and continues to utilize methods from a branching one-model-per-dataset paradigm, which is infeasible for even moderately sized datasets. We have reframed the modelling approach in TR experiments to a core-model paradigm. Where a single "core" model is refined against all time points, maintaining the consistency in the model compositions across datasets, while still adjusting parameters to accommodate for experimental variation. This approach requires generic and flexible merging tools to create arbitrarily complex multi-state models, and corresponding refinement protocols. Exhaustive sampling of state occupancies combined with electron density validation allows the identification of occupancies for each dataset. This makes modelling and refinement tractable for arbitrarily large multi-dataset experiments, where there is no compositional heterogeneity between data, as in TR experiments. Our approach offers a novel way to think about structural approaches in general and opens new possibilities for future method development.