

Discrete conformations of an enzyme are discernible via hierarchical clustering of X-ray diffraction intensities derived from multiple crystals

D. Kreitler¹, J. Jakoncic², H. Bernstein³, A. Soares⁴, T. Nguyen⁵, R. Sweet⁶, K. Patel⁷, A. Gulick⁸

¹National Synchrotron Light Source II ²Brookhaven National Laboratory, ³Ronin Institute for Independent Scholarship, c/o NSLS-II, Brookhaven National Lab, Bldg 745, ⁴Brookhaven National Laboratory, ⁵NA, ⁶Brookhaven Nat'l Lab, ⁷University at Buffalo, ⁸University at Buffalo
dkreitler@bnl.gov

Hierarchical clustering of X-ray diffraction intensities has recently been applied to multi-crystal datasets to determine the optimal combination of reflections for merging into final composite datasets. These methods have seen good success in macromolecular crystallography (MX), typically when target molecules that are recalcitrant to crystallization yield microcrystals which give rise to incomplete, partial datasets that represent small wedges in reciprocal space. In the present work we hypothesized that small deviations in atomic coordinates due to substrate binding in an enzyme protein crystal (relative to apo enzyme coordinates) would manifest as small differences in reflection intensities that would be discernible with agglomerative clustering techniques. To test this hypothesis we prepared apo enzyme crystals of a type C NRPS independent siderophore (NIS) synthetase, DesD, in which apo enzyme crystals were soaked with an ATP substrate for varying time intervals. X-ray diffraction datasets (100 deg) were collected with a microbeam (5x7 um) from 64 individual enzyme crystals, and each dataset was segmented into either N=1,4,10,20 partitions corresponding to individual wedges of 100, 22.5, 10, or 5 degrees respectively. An experiment model was refined separately for each individual wedge (autoPROC; C2221), e.g. in the case of an N=20 partition a single 100 degree dataset gave rise to 20x5 degree separate datasets and thus 1280 total wedges for hierarchical clustering. For each set of partitions we used Ward's algorithm within the KAMO package to cluster (CC-clustering; Wilson B-scaling) and merge (xscale) reflections. Subsequently, models corresponding to individual reflection clusters were auto-refined with phenix.refine followed by manual inspection to verify the presence or absence of substrate. We observed that partitioning of individual datasets into N=20 wedges representing the smallest volume of reciprocal space (5 degrees) was required to observe the isolation of non-substrate bound structures under a single node of the dendrogram. In addition, within the N=20 partition cluster we also observed clusters that yielded substrate bound structures which produced objectively higher quality electron density maps compared to maps derived from single crystal datasets. Together these results suggest that agglomerative clustering may become a useful technique for observing multiple conformations of proteins that undergo modest conformational changes within a crystal lattice.