A simple technique to classify diffraction data from dynamic proteins according to individual polymorphs

Alexei Soares[1], Thu Nguyen[2], Kim Phan[3], Sandra Gabelli[4], Dale Kreitler[5]

[1]Brookhaven National Laboratory [2]No affiliation given, [3]No affiliation given, [4]Johns Hopkins University, [5]National Synchrotron Light Source II

soares@bnl.gov

One often observes small but measurable differences in diffraction data measured from different crystals of a single protein. These differences might reflect structural differences in the protein and potentially reflect the natural dynamism of the molecule in solution. Partitioning these mixed-state data into single-state clusters is a critical step to extract information about the dynamic behavior of proteins from hundreds or thousands of single-crystal data sets. Mixed-state data can be obtained deliberately (through intentional perturbation) or inadvertently (while attempting to measure highly redundant single-crystal data). To the extent that different states adopt different molecular structures, one expects to observe differences in the crystals; each of the polystates will create a polymorph of the crystals. After mixed-state diffraction data are measured, deliberately or inadvertently, the challenge is to sort the data into clusters that may represent relevant biological polystates. Here we address this problem using a simple multi-factor clustering approach that classifies each data set using independent

observables, thereby assigning each data set to the correct location in conformation space. We illustrate this method using two independent observables – unit cell constants and intensities – to cluster mixed-state data from chymotrypsinogen (ChTg) crystals. We observe that the data populate an

arc of the reaction trajectory as ChTg is converted into chymotrypsin.