

W0415 Relationships between Protein Properties and Crystallization. Qizhi Zhang*, Y. Kim, A.Joachimiak, Midwest Center for Structural Genomics, Argonne National Laboratory, Argonne, IL, USA.

The process of obtaining diffraction quality crystals is one of the most difficult bottlenecks in high-throughput structural genomics projects. A protein structure determination pipeline has been implemented and operating at the Midwest Center for Structural Genomics (MCSG). The MCSG pipeline includes protein purification, concentration and crystallization.

From this pipeline, we have collected a wealth of experimental data that is stored in our cloning, purification and crystallization databases. By far the data includes 14278 selected, 4872 cloned, 1907 purified, 624 crystallized proteins and 250 structures deposited in PDB. We have built a data mining technique to capture the patterns between protein properties and success rates of each step along our pipeline. A program was developed to automatically collect property information from published online resources for the proteins. Five properties (PI, MW, charge, hydrophobicity and unfoldability) were obtained and saved into a database. Information was retrieved to investigate the property distributions of proteins cloned, purified, crystallized and with quality diffraction and to analyze the trend of cloning, purification, crystallization and quality diffraction rates corresponding to their properties. The methodology and data mining results will be presented.

This work was supported by the by grants from the National Institutes of Health (GM62414) and the U.S. Department of Energy, Office of Biological and Environmental Research, under Contract W-31-109-ENG-38