

## W0503

**Structural Genomics Experimental Pipelines - Insights from the First Five Years.** O. Kirillova, M. Grabowski, H. Zheng, Z. Otwinowski, W. Minor, Dept. of Molecular Physiology and Biological Physics, Univ. of Virginia, Charlottesville, VA.

Worldwide structural genomics (SG) initiatives, aiming to rapidly elucidate the structures of a large number of strategically selected proteins, have produced more than 2000 structures out of a target list of around 96,000. About two thirds of SG structures were non-redundant (<30% sequence identity in the PDB at the time of deposition), which constituted close to a third of all non-redundant depositions. In 2005 the SG efforts contributed 45% of all non-redundant PDB deposits.

To estimate the potential of the second phase, we examined the dynamics of the SG experimental pipeline based on target data released by various centers, using the techniques of survival analysis.

The comparison of overall target progression rates at the stages of expression, purification, crystallization, and diffraction has shown that the pipeline efficiency varied significantly for different genomes and was influenced by protein characteristics and degree of similarity to proteins of known structure. For eukaryotic targets, most attrition was due to failure of expression and/or purification, while for prokaryotes it was due to failure of obtaining diffraction-quality crystals.

The high-throughput approach did not compromise structure quality. The quality of SG deposits has been steadily improving over time, despite the fact that the average time between data collection and completion of refinement was shortened from 242 days in 2003 to 175 in 2005.