

## W0349

**A Protein Family View of 3D Protein Structure Space Coverage.** Dawei Lin, Yuchao Zhou, Jeremy Praisman, John Rose and B.C. Wang, Southeast Collaboratory for Structural Genomics, Dept. of Biochemistry and Molecular Biology, Univ. of Georgia, Athens, GA.

Proteins can be grouped into families based on their sequence features. A protein family is a group of proteins that come from the same ancestor, and will generally have similar three-dimensional structures. Since it is more sensitive to infer a protein's function from its three dimensional structure than from its sequence, it is significant to have at least one structural representative from each known family in order to understand the functions of the protein universe.

A recent survey based on the protein family coverage by structure shows that although the number of structures solved is increasing exponentially in the Protein Data Bank, the number of structure solved for a structurally unrepresented family has remained constant and even decreased in recent years. A database has been built to investigate structure coverage of those unrepresented Pfam families (a protein family database) by complete genomes and the protein targets of the ongoing NIH PSI structural genomics projects. The analysis of the information in the database offers some suggestions on how to develop an optimal strategy for efforts that are specifically targeting new families.

Supported by NIGMS, IBM, Georgia Research Alliance and University of Georgia.