Poster I-8

Genome-wide de novo structure prediction for over 100 genomes and automatic integration of structure prediction with Gene Ontology database.

Authors:
Lars Malmström (*University of Washington*)
David Baker (*University of Washington*)
Richard Bonneau (*NYU, Dept of Computer Science*)

Short Abstract: We describe application of Rosetta de novo structure prediction to over 100 genomes and our novel method for automatic integration of that structural compendium with information contained in GO (using grid-computing, wcgrid.org). We highlight Rosetta+GO (validated and novel) function predictions in several organisms.

Long Abstract:
Recent progress in de novo structure prediction has resulted in methods with increased accuracy that are applicable to greater numbers of proteins. Rosetta de novo structure prediction can now contribute to systems biology in several ways when combined intelligently with other structure prediction methods, other sequence based bioinformatics techniques and preexisting functional information. Such novel applications include 1) comprehensive structural annotation on a genome wide scale, 2) contribution of structure prediction derived functional information to existing annotation schema such as the Gene Ontology database (GO) and/or metabolic and signaling pathway databases (KEGG) and 3) synergy with experimental approaches to structural genomics such as the derivation of distance constraints from mass spectroscopy.
We describe the recent application of Rosetta de novo structure prediction to over 100 complete genomes (including human) and the integration of the resulting comprehensive structural compendium with information contained in GO. This work contributes to overcoming three major challenges: 1) the massive computational cost of genome wide de novo predictions, 2) integration of higher resolution methods into the structure generation pipeline and 3) automatic methods for combination of structure prediction with previously existing function, process and localization data. This work was completed in collaboration with IBM (World Community Grid) and will employ over 1 million CPU years. We have used our automatic framework for integrating structure predictions with GO annotations to predict function. We provide examples of these predictions in several organisms, describing both cases where these predictions have been subsequently validated experimentally and novel predictions. We will also focus on our findings in Yeast and show that de novo structure predictions combined with pre-existing, non-specific, process and/or component annotations allow us to make confident specific molecular function and structure predictions for proteins playing roles in several key cellular processes and complexes. All data resulting from this project will be freely available in multiple formats and via multiple user interfaces (Cytoscape-BioNetBuilder and web-based).