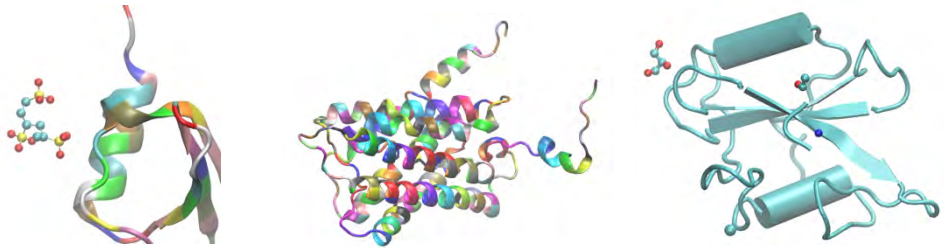
Structure to Function: Bringing Protein Structure and Ligand Screening to KBase

Ada Sedova[1]* (sedovaaa@ornl.gov), Tianhao Gu[2], Loukas Petridis[1], Brian H Davison,[1]
Christopher S. Henry[2], and **Julie Mitchell**[1]
[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee
[2]Argonne National Laboratory, Argonne, Illinois
http://cmb.ornl.gov/index.php/research/bioenergy/dynamic-visualization-of-lignocellulose

Project Goals: The annotation of gene function benefits greatly from information on protein structure and binding. We have been developing new functionality for structural analysis and virtual ligand screening in KBase, in support of the Dynamic Visualization of Biological Structures SFA at Oak Ridge National Laboratory. The new structural branch of KBase includes tools with which to upload and view protein three-dimensional structure, and to create co-evolutionary analysis-based contacts maps for use as restraints in protein folding algorithms. An open-source protein folding tool has been developed, and a protein-ligand docking workflow allows for the use of the AutodockVINA[1] software to screen databases of metabolites, connecting metabolic-level function via physical molecular interactions to genomic information.



The annotation of gene function benefits greatly from information on protein structure and binding. We are developing a new structural branch for KBase, in support of the Biofuels SFA at Oak Ridge National Laboratory. We have created the software infrastructure in KBase that uploads a model protein structure or downloads an experimental structure from the RCSB Protein Data Bank, and a web-based three-dimensional viewer to display the protein. This infrastructure paves the way for a number of structure-based analysis tools to be used in KBase that can add important information to computational biology research. The ligand screening application we have developed uses the AutodockVINA[1] protein-ligand docking program to allow the protein structure to be screened against a set of small molecule metabolites. A batch downloader to interface with the ZINC[2] database allows for the import of three-dimensional mol2 ligand files, providing robust starting geometries for the ligands. This application can be used as a part of a functional annotation of uncharacterized proteins, as ligand screening will help researchers anticipate likely functions. In addition to ligand docking, we have developed applications for protein structure prediction using co-evolutionary analysis. This method infers protein residue contacts from a multiple sequence alignment. The CCMPredPy[3] method has

been chosen for this application. The contacts can then be used as restraints in protein folding algorithms to generate a three-dimensional structural model for the protein from sequence. Information from the contact map an also provide value independent of any structure prediction, as the results indicate amino-acid residues that drive folding stability which is useful for experimental strategies in protein redesign and can also be used to refine the MSA and to create a phylogenetic tree. Finally, we have been developing the only open-source protein folding tool to take a set of distance restraints and arrive at a final three-dimensional structure.

Lignocellulosic biomass is a complex substrate that requires the synergistic action of a variety of enzymes for its efficient deconstruction. Biomass pretreatment generates byproducts, including solubilized lignin-derived aromatics, that inhibit enzymatic hydrolysis of cellulose[4]. Which bioproducts are formed depends on the biomass feedstock as well as the details of the pretreatment process. Applying the Ligand Screening App to predict which specific byproducts affect which particular enzymes can lead to an optimal selection of cellulolytic enzyme cocktails that minimize inhibition. A further barrier to biofuels and bioproduct production is that fermentation products, pretreatment solvents and byproducts can be toxic to microorganisms. Determining which proteins, and which protein residues, the small molecules bind to may lead to rational genetic engineering of those proteins and to microbes exhibiting improved tolerance to toxic pretreatment byproducts and solvents.

1) Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry. 2010 Jan 30;31(2):455-61.

2) Sterling T, Irwin JJ. ZINC 15–ligand discovery for everyone. Journal of chemical information and modeling. 2015 Nov 23;55(11):2324-37.

3)Vorberg S, Seemayer S, Söding J. Synthetic protein alignments by CCMgen quantify noise in residue-residue contact prediction. PLoS computational biology. 2018 Nov 5;14(11):e1006526.

4) Ximenes E, Kim Y, Mosier N, Dien B, Ladisch M. Deactivation of cellulases by phenols. Enzyme and microbial technology. 2011 Jan 5;48(1):54-60.

Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This program is supported by the Office of Biological and Environmental Research in the DOE Office of Science.