AkiraProt: An Ensemble Workflow for Proteome-Wide Structural Analysis

Erica T. Prates[1,2]* (teixeiraprae@ornl.gov), Manesh Shah[1,2], B Kirtley Amos[2,3], Allison Werner[1,4] Gregg T. Beckham[1,4], Daniel Jacobson[1,2] and **Gerald A. Tuskan[1]**

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Computational Systems Biology Lab, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [3]Department of Horticulture, University of Kentucky, Lexington, KY; [4]National Bioenergy Center, National Renewable Energy Laboratory, Golden CO

https://cbi.ornl.gov

Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* **CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by** *in planta* **modifications and biological funneling of lignin to value-added chemicals.**

The design of plants and microbes for cost-effective biomass conversion involves understanding the complex genetic basis of targeted traits. Functional screening techniques and statistical methods for genotype-phenotype mapping are continuously evolving, but their results are often limited to data correlations without capturing the specific roles of genes. Knowledge of intermediate omics layers (e.g., transcriptomics, proteomics and metabolomics), in turn, can help elucidate the molecular pathways of the information propagation from genes to phenotypes. The structural characterization of proteins is of paramount importance as they are the functional manifestation of genes.

Here, we present AkiraProt, a flexible and robust pipeline for high-throughput protein structure prediction, and its use for integration of structural proteomics in our Systems Biology studies. The workflow finely parses protein sequences and determines the optimum combination of state-of-the-art methods to be employed for protein structure modeling. Case-by-case generated protocols are selected to provide both the highest possible local resolution and maximum information about the overall shape of each protein. AkiraProt has been effectively used for structural analysis of eukaryotic and prokaryotic proteomes across several studies conducted within different focus areas of the Center for Bioenergy Innovation. Here we show its application leveraging the supercomputers at the Oak Ridge Leadership Computing Facility to solve the structural proteome of *Pseudomonas putida* KT2440, a widely studied host for biomass lignin valorization. Detailed analysis of specific proteins of interest contributed, for example, to the functional characterization of *P. putida* transport proteins involved in the optimum intake and the tolerance towards lignin-derived compounds. Other current applications of models generated with AkiraProt include prediction of key sites in high-throughput targeted mutagenesis for trait selection and ligand screening for protein functional mapping.