Development of Predictive Software Tools to Construct and Analyze Large Dynamical Networks for Systems Biology Knowledgebase

Ravishankar R. Vallabhajosyula* (rrv@cfdrc.com), B. Prabhakarpandian and Kapil Pant
CFD Research Corporation, 215 Wynn Drive, Huntsville AL
http://www.cfdrc.com

Project Goals: Recent biotechnological advances have accelerated the generation of 'omic' data. This has driven the development of computational tools to model the biological systems by inferring mechanisms responsible for response to external stimuli. However, lack of kinetic information for most biochemical interactions limits the predictive capabilities of these tools. CFD Research Corporation (CFDRC) is developing predictive modeling toolkit to overcome this limitation, thereby facilitating rapid and accurate characterization of the effects of the environment on phenotypes. In particular, our toolkit will enable (1) identification of significant biological features from omic datasets, (2) construction of a comprehensive network model of cellular pathways, and (3) simulation of this pathway model using a kinetics-free algorithm to predict the altered phenotypes when selected targets in the network are modified. This methodology is being validated using well-characterized organisms (e.g., yeast) as well as selected microbe-based biosystems of DoE interest (e.g., identification of higher quantity and quality biofuel yielding algal strains).

Recent developments in genetic engineering and biotechnology have enabled the modification of genes in an organism or the introduction of genes from other organisms towards achieving the desired phenotypes. However, these experimental procedures are often carried out without adequate systems-level knowledge of the cellular biology, which can lead to unexpected outcomes. Well-designed computational methodologies can be used to prevent such scenarios with the aid of predictive software tools. A key goal of the DoE Systems Biology Knowledgebase (Kbase) is to facilitate analysis of vast omic datasets for characterizing the response of organisms to various environmental stimuli towards predicting phenotypes. For example, such tools will be able to identify algal strains with improved attributes of biofuel production, while simultaneously overcoming slow growth rates associated with some of these strains. Such computational approaches should be based on a comprehensive understanding of the cellular biology of the organisms of interest, and will be significantly aided by the adaptation and application of novel algorithms and software that can analyze multi-omic data related to the observed response to various external stimuli.

Under DOE sponsored research, CFD Research Corporation (CFDRC) is currently developing predictive computational tools to address the goals of Kbase towards characterizing the response biological organisms to environmental stimuli that serve as inputs and predicting phenotypes most likely to be observed. Figure 1 shows a schematic of the framework being developed by CFDRC. Drawing upon available databases, our approach relies on the construction of mechanistic Systems Biology based and data-driven models of the differentially regulated cellular pathways. The complex pathway models are then analyzed without requiring information on the kinetics of various biochemical interactions. This enables the discovery and ranking of targets (for example genes, proteins or metabolites) for potential modification and the prediction of their response when these modifications are implemented. This approach thus offers the potential to inform experiments for the development of strains efficient at generating the desired phenotype such as algae strains that can produce biofuels at a higher rate.

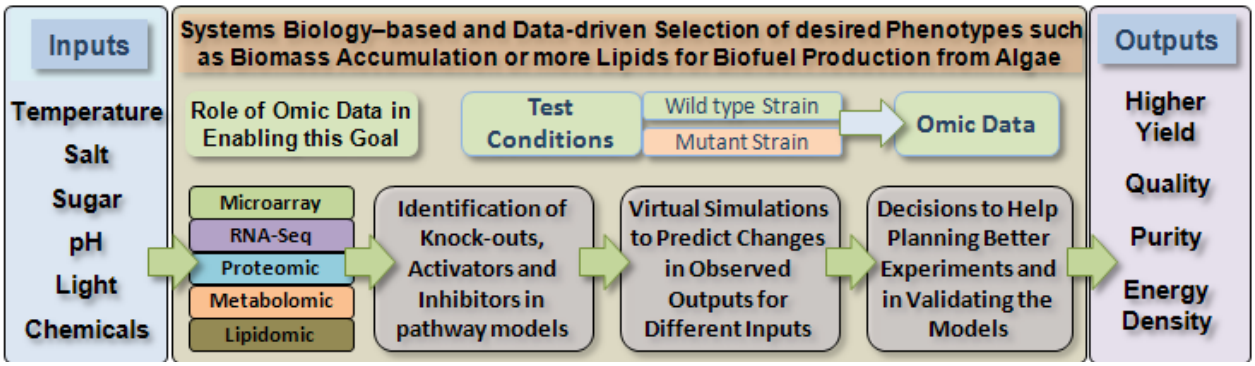

Figure 1: Schematic Detailing the use of Omic Data to Identify Targets towards Predicting Phenotypes

As part of the ongoing Phase I study, we are developing a prototype of the software toolkit using transcriptional data to construct and analyze complex pathway networks in an extensible SBML format (Hucka et al., 2003) that will be enhanced to analyze other omic data types in future. Development of these tools will enable researchers to analyze pathways that play important roles in sensing and responding to the external conditions in an integrated manner. These tools are important to understand the organism's behavior in the modified environment including its survival and in predicting the associated phenotypes. Towards this goal, we are studying the yeast environmental stress response to various external conditions as a test case to test and validate the model. We are also in active discussions with different organizations to demonstrate the technology for microbial systems of DoE interest e.g., identification of targets for genetic engineering of algal strains for higher quantity and quality biofuel production.

References

1. Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., et al., The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models, *Bioinformatics*, **19**(4): 524-531, 2003.

Acknowledgment

This work is being supported through the DOE Office of Biological and Environmental Research under an SBIR Phase I grant (DE-SC0006190).