Metabolic Reconstruction, Analysis and Design of Metabolic Pathways

Charles J. Foster[1], Debolina Sarkar[1], Satyakam Dash[1], Saratram Gopalakrishnan[1], Lin Wang[1], Joshua S.H. Chan[1], Costas D. Maranas[1*] (costas@psu.edu) and **Gerald A. Tuskan**[2]

[1]Department of Chemical Engineering, The Pennsylvania State University of State, University Park, PA 16802, [2] Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, Tennessee

https://cbi.ornl.gov
http://www.maranasgroup.com/

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 will address 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 and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Metabolic models provide a mechanistic description of an organism's metabolic repertoire, by combining reaction stoichiometry information with enzyme availability and metabolite pool levels to form a structured framework that can be analyzed computationally to understand, as well as predict, cellular metabolism. In this study, we reveal that incorporation of kinetic information improved predictive capabilities in a cellulolytic microbe, *C. thermocellum*. We further describe the limitations of existing parametrization procedures for kinetic model development and methods to overcome such limitations using *E. coli* as a model system. We also introduce a recently developed framework (i.e., novoStoic) [1] that seamlessly blends known transformations with reaction rules to construct pathways towards targeted chemicals. This framework is repurposed for prospecting pathways from lignin monomers to value-added chemicals. Finally, we extend our modeling framework from microbes to develop a multi-tissue plant (poplar) model which captures nutrient flow and growth across its tissues.

For *C. thermocellum*, we constructed a core kinetic model "k-ctherm118" [2] to capture the regulatory impact of changes in metabolite pools on reaction fluxes using the Ensemble Modeling paradigm. k-ctherm118 was parameterized by using fermentation yield data in major fermentation pathways for 19 measured metabolites, 19 distinct single and multiple gene knockout mutants along with 18 intracellular metabolite concentration data for a *Δgldh* mutant and ten experimentally measured Michaelis-Menten kinetic parameters. k-ctherm118 captures metabolic perturbations caused by 1) nitrogen limitation leading to increased yields for lactate, pyruvate and amino acids and 2) ethanol stress triggering an increase in intracellular ammonia and sugar phosphate concentrations due to upregulation of cofactor pools. Overall, the *C. thermocellum* case study demonstrates that the developed kinetic model (k-ctherm118) provides greater insight into metabolic pathways and regulations than the stoichiometric model.

We also expand the scope of the kinetic models by incorporating a transcriptional regulatory layer which refines enzyme levels based on a linear combination of log-normalized changes in growth rate (global) and select intracellular metabolite pool (specific) levels. A major computational bottleneck in the kinetic parameterization process is the lack of a fast and efficient algorithm to identify the optimal set of kinetic parameters using MFA-derived steady-state flux distributions. The implementation of a gradient-based

procedure is limited by slow numerical integration. To this end, we have implemented an algorithm [3] to compute steady-state flux distributions for a given set of kinetic parameters that overcome the limitations of numerical integration. In conjunction with an efficient gradient-based scheme for updating kinetic parameters, we have developed a fast and automated algorithm for parametrization of a kinetic model that includes allosteric and transcript regulations. We apply flux datasets from a recent MFA study on *E. coli* (wild-type and 22 knockouts of enzymes in the upper part of central carbon metabolism) and the parametrization procedure to construct a genome-scale kinetic model of *E. coli* containing 787 model reactions, 674 metabolites, and 618 substrate-level regulatory interactions.

The need to identify atom and energy optimized pathways from substrates to terminal products motivates us to go beyond the known repertoire of enzymatic reactions by exploring hypothetical reactions predicted using reaction rules while simultaneously considering all design criteria such as complexity of pathway topology, mass conservation, etc. First, we track and codify all reaction centers as rules, using a novel prime factorization based encoding technique (rePrime). A MILP-based algorithm, novoStoic [1], then allows for the efficient integration of both reaction rules and reactions in the search for pathways that carry out the efficient conversion of the source to target molecules. We demonstrate the use of novoStoic in bypassing existing pathway steps through putative transformations, assembling complex pathways, and blending both known and putative steps from lignin monomers to value-added chemicals.

We also extend the metabolic modeling framework beyond microbes by developing whole-plant metabolic model of poplar (*P. trichocarpa*) to identify key genes responsible for controlling growth, yield, degradability, and biomass composition. Thus, a genome-scale model for poplar was constructed containing over 3,459 metabolites, 3,388 reactions, and 8 sub-cellular compartments. The GSM model acts as blueprint to explore organ-specific compartments and reaction networks spanning the root, shoot, and leaf tissues using omics data and ultimately capture carbon and nitrogen flows between tissues as a function of growth stage and sequester biomass in organ-specific ratios.

References

1. Kumar A., Wang L., Ng C.Y., and Maranas C.D., (2017) Pathway design using de novo steps through uncharted biochemical spaces. *Nature Communications*, in press.
2. Dash S., Khodayari A., Zhou J., Holwerda E.K., Olson D.G., Lynd L.R., Maranas C.D., (2017) Development of a core *Clostridium thermocellum* kinetic metabolic model consistent with multiple genetic perturbations. *Biotechnol Biofuels*, 10:108.
3. Gopalakrishnan S., Foster C.J., Dash S., and Maranas C.D., (2017) in preparation.

The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science