**Application of the MetRxn database to highlight multi-tissue/organisms and expansion to include algorithms for predicting novel reactions and pathways**

Chiam Yu Ng[1]* (cun121@psu.edu), Lin Wang[1], Akhil Kumar[2], Siu Hung Joshua Chan[1], Margaret Simons[1], Anupam Chowdhury[1] and **Costas D. Maranas[1]**

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA; [2]The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA

http://www.maranasgroup.com/

**Project Goals: The project aims to apply and extend the MetRxn database. MetRxn was used to create expanded genome-scale models that include complex interactions through multiple tissues and organisms by examining the metabolism in maize and microbial communities. Additionally, MetRxn is being expanded to include algorithms for predicting novel reactions and pathways by examining promiscuous enzymes and designing novel pathways accounting for thermodynamic feasibility and protein costs.**

The developed MetRxn knowledgebase (www.metrxn.che.psu.edu) is a unified repository of 8 databases and 112 metabolic models, resulting in 44,784 unique reactions and a million plus unique metabolites. In addition, the database includes 6,211 reaction rules developed using Canonical Labelling for Clique Approximation (CLCA), which leverages prime factorization. We focus on applying the MetRxn database to expand models to investigate the metabolic behaviors of complex organisms and microbial communities. By reconstructing multi-tissue and multi-organism models, we can use flux balance analysis to determine the interactions between different cell/tissue-types or organisms, examine metabolic changes among growth stages, and predict the community abundance of microorganisms. Additionally, algorithms are being developed to utilize the standardized database and reaction rules in MetRxn predicting novel reactions from enzyme promiscuity and novel pathway combinations accounting for thermodynamic feasibility and protein costs.

The MetRxn database was applied to highlight the interactions between multiple tissues and organisms. First, a whole-plant metabolic model of maize was developed by reconstructing the root, stalk, leaf, kernel and tassel tissues using the phloem to transport metabolites. The whole-plant model was simulated at three growth conditions: vegetative leaf growth, tassel development, and kernel filling. Growth rate and dry weight proportions were used to define the plant biomass congruity and normalize the transport reaction flux between tissues. Using parsimonious analysis, the total number of reactions required during each growth stage was identified with approximately 80% of the reactions in the vegetative leaf growth stage common among all growth stages. The model predicts the expected transport of sucrose from photosynthetic tissues into the root, where sucrose is consumed for energy and produces organic acids that are recycled to the photosynthetic tissues. Additionally, reactions from MetRxn were combined for multiple microbial organisms to create a community model. Steady-state community metabolism was predicted for long-term stability using a novel framework, SteadyCom. In the absence of the steady-state constraint, faster growing organisms will displace other microbes and disrupt the interactions predicted by non-steady-state methods. The SteadyCom framework is scalable to large communities and compatible with existing constraint-based modeling techniques. The algorithm's capability to predict relative abundances and inter-organism relationships was demonstrated in a community of four *E. coli* auxotrophic mutants.

The standardized MetRxn reaction database and reaction rules within the newly developed quasiPath framework were leveraged to design novel reaction pathways in a mass balanced fashion. While designing novel biotransformations from the substrate to product, we

consider design elements such as network size, non-linear pathway topology, mass-conservation, cofactor balance, thermodynamic feasibility and chassis selection. We demonstrate the capability of quasiPath in imputing the biochemical gaps in gut metabolism. In addition, biotransformation models for the aerobic conversion of polycylic aromatic hydrocarbons (PAH), found in industrial effluents to commercially valuable compounds are also presented. We also contrast the philosophies of quasiPath with recent retrosynthesis efforts.

Factors such as stoichiometric and cofactor balancing, thermodynamic feasibility and pathway metabolic burden have to be carefully considered while designing a novel metabolic pathway from the collection of reactions in MetRxn. We have previously developed the MILP-based optStoic/minFlux pathway designing procedure which identifies a minimal network of flux carrying reactions that satisfy the overall design equation while obeying mass, energy and overall thermodynamic balances. In this project, we extended the procedure to explore the design principle of the natural glycolytic pathways. Although there are a plethora of routes through which glucose can be converted into pyruvate, the canonical ED and EMP glycolytic pathways are widely adopted by different species despite their difference in energy efficiency. We thus employed the modified optStoic/minFlux method to prospect for over 37,301 possible routes between glucose and pyruvate at different pre-specified stoichiometric yields of ATP. Subsequently, we filtered all candidate pathways based on their thermodynamic feasibility and quantified the minimal protein cost of the feasible pathways. A trade-off plot between energy efficiency and protein cost for each of the feasible routes revealed that the naturally evolved ED and EMP pathways are among the most protein cost-efficient pathways in their respective ATP yield categories. The pathway design and analysis procedure developed here can be applied to other important bioconversion pathways.

*Supported by funding from the U.S. Department of Energy by grant DE-FG02-05ER25684* and *DE-SC0008091.*