100. Predicting Gene Targets for Optimizing Oil Production in *Yarrowia lipolytica* by Cell-Wide Ensemble Modeling

Yun Lee[1]* (yun.lee@ucla.edu), Jimmy G. Lafontaine Rivera[1], James C. Liao[1,2], and **Gregory Stephanopoulos[3]**

[1]Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, 5531 Boelter Hall, Los Angeles, CA; [2]UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, 611 Young Drive East, Los Angeles, CA; and [3]Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA

Diesel is a valuable fuel due to its high-energy density and better (than gasoline) thermal efficiency as transportation fuel. Additionally, it can be produced from renewable resources and, as such, it can contribute significantly to mitigation of greenhouse gas generation and enhancement of domestic feedstocks for production of transportation fuels. To date, all biodiesel production facilities rely on vegetable oils and animal fats as feedstocks. Such feedstocks, however, are very limited (maximum capacity of ~1B gallons/year). As a result, there is a need for the cost-efficient production of lipids from carbohydrate feedstocks, which are very plentiful throughout the world and rather well distributed in various forms (monosaccharides, starches, various types of cellulosic and hemicellulosic biomass).

Recently, it has been shown that an engineered strain of *Yarrowia lipolytica* is capable of converting, at very high yields, carbohydrates to oils and lipids for biodiesel production. The current conversion yields achieved by this engineered strain suggest that oil for biodiesel can be produced at a cost of approximately $3/gallon, depending on the feedstock used. Moreover, this engineered oleaginous yeast can metabolize not only simple sugars, such as glucose and sucrose, but also crude glycerol (byproduct of current biodiesel operations), and biomass and algae hydrolysates, among other feedstocks. The above figures of merit exceed significantly what has been reported to date on microbial and algal lipid production and could provide the basis of a cost-effective process for lipid production from renewable feedstocks.

However, further improvement of such strains to achieve near-optimal yields often defies biochemical intuition and stoichiometric reasoning. Alteration of enzyme levels, for example, does not change stoichiometry and, as such, cannot be captured in constraints- based models. As such, it needs to be represented by dynamic models involving enzyme kinetics. Unfortunately, the lack of kinetic parameters has hindered the utility of such approaches. Furthermore, such models typically require metabolite time-course measurement for parameter estimation and model validation. *Ensemble Modeling* (EM) was recently introduced to address these problems by mimicking the strategy of high-throughput screening of biological molecules. The approach starts from constructing a library (or ensemble) of models with different parameter sets such that all models are anchored to the same end point (flux distribution) of fermentation using the control strain. Thus, the ensemble of models spans all the allowable kinetic space while having the same steady-state flux distribution. The constraint to the end point (or steady state) significantly reduces the parameter space, and enables effective sampling of parameters. In the next step, the ensemble of models will be screened by fermentation flux measurements after genetic perturbation of enzymes (overexpression or knockout). To screen the models, each model in the ensemble is perturbed the same way as in strain construction experiments. Only the models that match the new experimental data are retained, and the rest are discarded. This model-screening step is repeated in the next round of strain engineering. After a few rounds of

screening, the remaining models in the ensemble become increasingly accurate, and can be used to generate possible targets for future strain engineering.

To apply EM to improve oil production in *Y. lipolytica*, we have reconstructed a cell- wide description of *Yarrowia* metabolism that encompasses all reactions necessary for lipid biosynthesis. Based on this reconstruction, we adopted the EM approach and generated an ensemble of kinetic models that have the same steady-state fluxes but differ in kinetic parameters. In this case, the flux data came from the 13C-labeling experiments and included measurements for various *Yarrowia* strains with distinct genetic backgrounds. Currently, all the models within the ensemble are anchored to the steady- state fluxes measured from a genetically modified strain where acetyl-CoA carboxylase (ACC1) and diacylglycerol acyltransferase (DGA1) are overexpressed. This ACC1+DGA1 strain has been shown to accumulate up to 62% of its DCW as lipids at an overall volumetric productivity of 0.143 g/L/h, making it a desirable starting point for further optimization.

To identify potential gene targets, we perturbed each model by increasing (up to 2-fold) and decreasing (down to 50%) the activity of each enzyme. Since not every model is capable of sustaining a steady state upon perturbations, we retained the models that survived the maximum perturbations and used their collective results to identify perturbations that were shown to improve the lipid accumulation. Interestingly, our results suggest that overexpression of hexokinase, oxoglutarate dehydrogenase, or cytosolic adenylate kinase, as well as knockdown of glutamate synthase or glutamine synthetase, will lead to a greater production of triacylglycerol compared to the control strain. These computational predictions will provide guidance in the next round of metabolic engineering, and the experimental results, whether they are consistent with the computational predictions or not, will be used for further screening of the ensemble.