Construction of genome-scale metabolic models for non-model yeast organisms for biofuel and bioproduct engineering

Siu H. J. Chan (sxc554@psu.edu), Hoang Dinh, Saratram Gopalakrishnan, and Costas D. Maranas

Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802

http://www.maranasgroup.com/

Project Goals: This project aims to reconstruct the genome-scale metabolic models for two non-model yeast organisms, *Rhodosporidium toruloides* and *Issatchenkia orientalis*, which are industrially potential hosts for producing biofuels and organic acids. Using the metabolic model of the model yeast *Saccharomyces cerevisiae* as reference and multiple sources of data, the models for the two organisms were constructed following a rapid reconstruction workflow. The reconstructed metabolic models and the associated atom mapping will act as an important tool for kinetic modeling, ^{13}C metabolic flux analysis, identification of metabolic regulation and integration of physiological and omics data, facilitating the genome-scale analysis and engineering of the two yeast organisms for production of biofuels and bioproducts.

Yeast organisms are promising engineering targets for microbial cell factories with unique industrial advantages compared to bacteria such as higher tolerance against inhibitory compounds and contamination. However, most metabolic engineering and modeling efforts were focused on the model yeast *Saccharomyces cerevisiae* despite the discovery of other non-model yeast organisms with unique biochemical production capabilities. Among them, *Rhodosporidium toruloides* is able to ferment lignocellulose into lipids and can be a potential platform for the production of fatty-acid-derived biofuels. *Issatchenkia orientalis* is another yeast with industrial potential of high-level production of organic acids. We have reconstructed genome-scale metabolic models for the two yeast organisms following a previously established rapid reconstruction workflow [1]. Using *S. cerevisiae* as the reference organism, we identified the homologous genes between *S. cerevisiae* and the two organisms using bidirectional BLAST as well as the Yeast Genome Annotation Pipeline [2] which identifies homology based additionally on synteny conservation. The most recent yeast7 model for *S. cerevisiae* [3] has been updated with the corrected gene-protein-reaction (GPR) associations [4], completely balanced stoichiometries and standardized biomass reactions [5]. Backbone models for the two organisms were extracted from the updated yeast7 model by comparing the homologous genes and evaluating the GPR associations. Additional reactions not in the backbone models were identified from multiple sources of data, including annotations by InterPro [6], KEGG [7], Pfam [8, 9] and UniProt [10]. Gapfilling was performed using reactions from the yeast7 model to ensure that the models are able to produce biomass. Flux elucidation using 13C-MFA requires the availability of a genome-scale carbon mapping model. To this end, the mapping models corresponding to the constructed genome-scale models of *R. toruloides* and *I. orientalis* are constructed using the previously published mapping models for *E. coli* and *Synechocystis* PCC 6803 as the starting point. Elementary metabolite unit (EMU) decomposition is performed using 46 metabolite fragments from 15 central metabolites and 12 amino acids quantified using GCMS and LCMS techniques. The identified carbon transitions are compared and contrasted with *E. coli*. These findings will serve as the foundation for further genome-scale computational analysis, including 13C-MFA, kinetic modeling, identification of metabolic regulation, and integration of experimental data to facilitate the engineering of the two yeast organisms.

References

1. Mueller, T.J., Berla B.M., Pakrasi H.B., and Maranas C.D. (2013) Rapid construction of metabolic models for a family of Cyanobacteria using a multiple source annotation workflow. *BMC Syst Biol*, 7:142

2. Proux-Wéra, E., Armisén, D., Byrne, K. P., and Wolfe, K. H. (2012). A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics*, *13*: 237
3. Aung,H.W., Henry, S.A., and Walker, L.P. (2013) Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Ind Biotechnol*, 9:215–228.
4. Chowdhury R., Chowdhury A., and Maranas C.D. (2015) Using gene essentiality and synthetic lethality information to correct yeast and CHO cell genome-scale models. *Metabolites*, 5:536-570
5. Chan S.H.J., Cai J., Wang L., Simons-Senftle M.N., and Maranas C.D. (2017) Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. *Bioinformatics*, 33:3603-3609.
6. Finn, R.D., et al. (2017) InterPro in 2017 — beyond protein family and domain annotations**.** *Nucleic Acids Res*, 45, D190-D199.
7. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*, 44:D457-D462
8. Finn, R.D., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*, 44:D279-D285.
9. Alborzi, S. Z., Devignes, M. D., and Ritchie, D. W. (2017) ECDomainMiner: discovering hidden associations between enzyme commission numbers and Pfam domains. *BMC bioinformatics*, 18*:*107.
10. The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res*, 45:D158-D169

Supported by funding from the U.S. Department of Energy by grant DE-SC0018260.