Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Understanding Cellular Metabolism using Systems Engineering Approaches

Jin Wang*

**Department of Chemical Engineering, Auburn University, AL, 36849, USA*
(Tel: 334-844-2020; email: wang@auburn.edu)

Abstract: Biological systems and large-scale industrial processes share many similarities at the systems level, which make the integrative systems engineering approaches essential in the understanding, control and optimization of biological systems. However, biological systems also present unique challenges that cannot be readily addressed by available systems engineering tools. In this work, we present our recent progress made in developing new systems engineering tools to understand microbial cellular metabolism at genome-scale. In particular, we focus on genome-scale metabolic network modeling and dynamic transcriptomic analysis. The effectiveness of the developed tools is demonstrated using a xylose fermenting yeast, *Scheffersomyces stipitis*, as the model system.

Keywords: systems biology, systems engineering, genome-scale metabolic network model, dynamic transcriptomic analysis, systems identification, flux balance analysis.

Biological systems and large-scale industrial processes share many similarities at the systems level: they both consist of many individual components; they both have built-in or engineered feedback control/regulation mechanisms; and the properties of the overall systems are both determined by the complex interactions among different components. Clearly, the complex natures of biological systems make the integrative systems engineering approaches essential in understanding, controlling and optimizing these systems. However, biological systems and industrial processes also have their distinct differences: for industrial processes, the feedback control mechanisms are well-understood, while those in biological systems are largely unknown; the number of available samples are usually significantly greater than the number of variables for industrial processes, while it is often the opposite for biological systems; the measurements obtained from industrial processes are usually quantitative with high precision, while those obtained on biological systems are often qualitative or semi-quantitative with large variations. Because of these unique challenges associated with biological systems, rote applications of systems engineering techniques may not yield useful results.

In this work, we present our recent progress made in adopting systems engineering principles and developing new systems engineering tools to understand microbial cellular metabolism. In particular, we focus on genome-scale metabolic network modeling and dynamic transcriptomic analysis. For genome-scale metabolic network model (GEM) development, we have developed a system identification based (SID-based) framework for GEM validation and refinement; to extract useful information embedded in messy big data of dynamic transcriptomic profiles, we have developed new data pre-processing pipeline and novel data analysis approaches to identify key regulatory mechanisms that govern the transient response of the cellular metabolism. Using a xylose fermenting yeast, *Scheffersomyces stipitis*, as the model system, we illustrate how the developed tools can help obtain deeper system-level understanding on the cellular metabolism at genome-scale. *S. stipitis* is an industrially relevant yeast species as it has one of the highest capacities to convert xylose into ethanol. Using *S. stipitis*, we first demonstrate how the SID-based framework can significantly expedite the GEM refinement, then illustrate how the dynamic transcriptomic profiling can be used to identify key metabolic strategies the yeast utilizes to cope with oxygen limitation.

GEMs provide a holistic view of the organism's metabolism, and have been shown to be a powerful tool in gaining genome-wide understanding on cellular metabolism. In essence, a genome-scale metabolic model (GEM) is a comprehensive functional database of an organism's cellular metabolism, which consists of a set of metabolites, metabolic reactions (*i.e.*, stoichiometric matrix), and constraints. GEMs represent the link between the genotype and phenotype of the organism, and can be used to conduct simulation/computations to answer various questions about the capabilities of the organism and its likely phenotypic states.

Similar to models developed for complex industrial processes, the quality of a GEM determines the successfulness of its applications. For any microorganism, after an initial draft model is reconstructed from sequenced genome, significantly more efforts are needed to validate and refine the GEMs (Österlund et al., 2012). Currently, besides assessing the model size and connectivity, the standard approach for GEM validation is to compare model prediction with experimental data under different conditions (King et al., 2015), which we term as "point-matching" approaches. Because each experimental condition represents a single (although high dimensional) point in the phenotypic space. Such point-matching approaches may work well for well-characterized organisms. However, given the fact that a GEM is severely underdetermined, matching experimental data over a few limited conditions does not necessarily indicate a high-quality GEM and can result in very misleading conclusions (Hilliard et al., 2018).

To address the shortcomings of point-matching validation, we developed a system identification (SID) based framework for GEM validation based on "knowledge-matching". In the SID framework, we use systems level knowledge captured by the GEM for validation, instead of numerical predictions generated by the GEM. First, biological knowledge embedded in a GEM is extracted from a series of designed *in silico* experiments. Next, the extracted knowledge, such as how cells respond under a given stimulus, is visualized and compared with the existing knowledge for model validation and analysis. In this way, instead of directly comparing the simulation results with experimental data, the knowledge captured by the model is compared with available knowledge to validate the model. Although rooted in simulations, the SID-based approach is more of a qualitative validation, and offers additional robustness against measurement errors.

For GEM refinement, the biggest challenge is to identify the root cause of an erroneous model behavior. Due to complex interconnectivity in a GEM, many times seemingly unrelated reactions located far away from the "problematic" reactions (*i.e.*, reactions that are not carried out in the expected way) play a key role in changing model behavior, and the point-matching validation does not provide information on such "hidden" relations. In the SID-based GEM refinement, we first conduct *in silico* experiments by forcing desired model behavior, and then use the SID framework to identify the key reactions that are affected the most by the forced correct behavior. These reactions serve as the key candidates that caused the erroneous model behavior, and are examined closely to identify the real root cause for refining the model.

We use *S. stipitis* as the model system to demonstrate how the SID framework works (Damiani et al., 2015; Hilliard et al., 2018). We first apply knowledge-matching based validation to examine two published GEMs on *S. stipitis*, iSS 884 and iBB 814. Our results suggest that the knowledge captured by iBB 814 shows better agreement with available knowledge on the yeast strain, although it showed worse performance in point-matching validation. Then using iBB 814 as the base model, we apply SID guided refinement to obtain the modified GEM, iDH 814, which shows better performance in both knowledge-matching and point-matching validations among different GEMs.

Recently, RNA-seq based transcriptome profiling has been used to enhance the understanding of the genome-scale response of the organism to different stimuli. From a control perspective, as cellular metabolism is a highly complex dynamic system, the transient response could offer significantly more information on the cellular metabolism, particularly on potential gene regulatory mechanisms.

For *S. stipitis*, it has been suggested that redox balance plays a key role in the fermentative response to reduced oxygen availability. To gain better understanding on xylose fermentative metabolism in the strain, we first cultivated *S. stipitis* in a controlled chemostat condition with xylose as the sole carbon source. Once aerobic steady-state (aeroSS) growth was achieved in the reactor, the oxygen supply was significantly reduced and the cells were allowed to transition to the new micro aerobic steady-state (microSS). Cultivation and RNAseq data were obtained from both steady-states as well as the dynamic transition period to further investigate the metabolic shifts that occur in response to the induced oxygen limitation.

Since very limited tools are available to analyze dynamic transcriptomic profiles, we developed our own data pre-processing and analysis pipeline by integrating available bioinformatics tools with systems engineering tools. By integrating the analysis results obtained from dynamic transcriptomic data and the cultivation data with genome-scale modelling (Hilliard et al., 2018), we were able to identify potential short-term and long-term strategies that the cells utilize to cope with oxygen limitation. Specifically, our results suggest that *S. stipitis* utilizes intracellularly stored sorbitol as a short-term response to the induced oxygen limitation which explains the observed overproduction of ethanol during the first half of the transition period (Diano et al., 2006; Shen et al., 2002). In addition, our results suggest that the upregulation of the glyoxylate shunt is involved in the long-term response cells utilize to cope with the oxygen limitation that persists in the reactor (Caspeta and Nielsen, 2013; Terabayashi et al., 2012).

1. ACKNOWLEDGEMENT

This work was supported by National Science Foundation (1069004, 1248388, 1264861), Department of Education (P200A150074), and U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program (DE-SC0019181).

REFERENCES

Caspeta, L., Nielsen, J., 2013. Toward systems metabolic engineering of *Aspergillus* and *Pichia* species for the production of chemicals and biofuels. Biotechnol. J. 8, 534–544. https://doi.org/10.1002/biot.201200345

Damiani, A.L., He, Q.P., Jeffries, T.W., Wang, J., 2015. Comprehensive evaluation of two genome-scale metabolic network models for Scheffersomyces stipitis. Biotechnol. Bioeng. 112, 1250–1262.

Hilliard, M., Damiani, A., He, Q.P., Jeffries, T., Wang, J., 2018. Elucidating redox balance shift in Scheffersomyces stipitis' fermentative metabolism using a modified genome-scale metabolic model. Microb. Cell Fact. 17, 140.

King, Z.A., Lloyd, C.J., Feist, A.M., Palsson, B.O., 2015. Next-generation genome-scale models for metabolic engineering. Curr. Opin. Biotechnol. 35, 23–29.

Österlund, T., Nookaew, I., Nielsen, J., 2012. Fifteen years of large scale metabolic modeling of yeast: Developments and impacts. Biotechnol. Adv. 30, 979–988.

Shen, B., Hohmann, S., Jensen, R.G., Bohnert, and H.J., 2002. Roles of Sugar Alcohols in Osmotic Stress Adaptation. Replacement of Glycerol by Mannitol and Sorbitol in Yeast. Plant Physiol. https://doi.org/10.1104/pp.121.1.45