Project title: Dynamic Metabolic Model Building Based on the Ensemble Modeling Approach

Final Technical Report

DOE Award number: DE-SC0006698

Name of Recipient: James Liao

Principal investigator: James Liao

Team members: Jimmy Lafontaine Rivera, Yun Lee, Matthew Theisen, Po-Wei Chen

Project Period: 9/1/2011-8/31/2015

October 2016

Executive Summary

Ensemble modeling of kinetic systems addresses the challenges of kinetic model construction, with respect to parameter value selection, and still allows for the rich insights possible from kinetic models. This project aimed to show that constructing, implementing, and analyzing such models is a useful tool for the metabolic engineering toolkit, and that they can result in actionable insights from models. Key concepts are developed and deliverable publications and results are presented.

Introduction

Metabolic engineering has advanced from minor alterations of existing pathways to significant re-routing of the metabolic path for better utilization of substrates or formation of non-native products. Non-steady states in metabolic engineering typically result in accumulation or disappearance of intermediate metabolites. This can result in gradual deterioration of performance, or system failure which would be catastrophic for the cell.

This may cause accumulation or depletion of metabolites and the disappearance of a stable steady state. Since a stable steady state disappears after a bifurcation occurs, the bifurcational robustness should therefore be an important criterion for designing non-native pathways. Artificial dynamic controllers are potentially useful, but it is desirable to choose underlying network configurations or parameter ranges that are inherently robust to bifurcation if possible.

The robustness problem calls for a modeling approach that relates kinetic parameters with system-level performance. Kinetic parameters are perturbed in such models to examine the consequences of drifting or perturbations to the system. Unfortunately, key kinetic parameters (e.g., Vmax′s) are system-dependent and usually unknown. We addressed the uncertainty of metabolic parameters through the random sampling of parameters to form an ensemble of models. Various approaches are then used to extract useful information from the ensemble upon large parameter changes, or infinitesimal perturbations that define control coefficients. Since non-native pathway design normally starts with little knowledge of kinetic parameters, it is sensible to investigate bifurcational robustness for an ensemble of models and to quantify it using the probability of system failure.

We summarize the key deliverables of the project as follows:

1) The theoretical development of ensemble modeling to include parameter continuation for determining bifurcational robustness (so-called Ensemble Modeling Robustness Analysis or EMRA). Published 2014. (Lee, Lafontaine Rivera, & Liao, 2014)
2) The application of EMRA to batch *in vitro* enzymatic systems to demonstrate relation between bifurcational robustness and production. Published 2016. (Theisen, Lafontaine Rivera, & Liao, 2016)
3) The development of an entropy index to characterize the instability/robustness of cellular models for model validation and editing. Published 2015. (Lafontaine Rivera, Lee, & Liao, 2015)
4) The implementation of a cellular model of *E. coli* with prediction of isobutanol production. Submitted 2016.

Deliverable 1: Theoretical development

We defined as the bifurcational robustness as the distance away from an unstable region in parameter space when the system is constrained to a steady state. This bifurcational robustness is readily measured using parameter continuation integration until the Jacobian becomes singular. This measures the ability of a dynamical metabolic system to return to a fixed point upon perturbation. Building a theoretical foundation of robustness, measuring it, and in particular defining a simple way to quantify it, represent unique challenges in systems biology. For small perturbations, local stability criteria are well defined using linear stability analysis. For large perturbations, one must explore global properties of the system. It is important to make the distinction between bifurcational robustness, which quantifies the tendency to avoid sudden change in dynamic regime due to parameter changes, and local sensitivity, which quantifies the changes in performance (flux, period of oscillation) as a function of changes in parameters within the same dynamic regime.

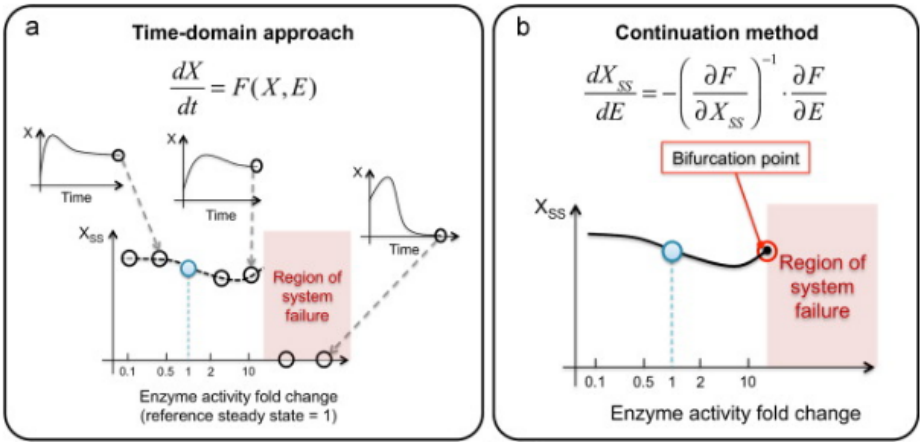

Fig 1. From Lee, Y., Lafontaine Rivera, J. G., & Liao, J. C. (2014). Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. Metabolic Engineering, 1–

9. Comparison of time-domain and parameter-domain integration. The time domain approach **(a)** requires convergence of the system to steady state at each step in parameter space, while parameter domain **(b)**, this can be accomplished in one step of a numerical solver.

We applied the parameter continuation method to an ensemble of models to demonstrate this modeling approach is mathematically and practically viable. Parameter continuation is a method which constrains a dynamic system to steady state as a parameter changes, similar to local sensitivity analysis. Our approach emphasized. This can be computationally more efficient than modeling in time-domain and waiting for steady state convergence. Fig 1, from (Lee et al., 2014) shows the difference between the time- and parameter-domain integration methods.

Deliverable 2: Robustness for batch enzymatic systems

We applied robustness analysis to several in vitro systems which have been characterized in the literature (Theisen et al., 2016). In these literature reports, the performance of the system was found to be reduced by the increase of a certain enzyme or feed rate. We found that the method proved versatile enough to successfully predict these features in three different pathways investigated in different laboratories and powerful enough to do so without a priori knowledge of specific enzyme parameter values. While the characterized pathways were optimized based on intuition during their experimental characterization, it's possible that longer pathways with more enzymes would be much more difficult to optimize without rational balancing methods like those presented here.

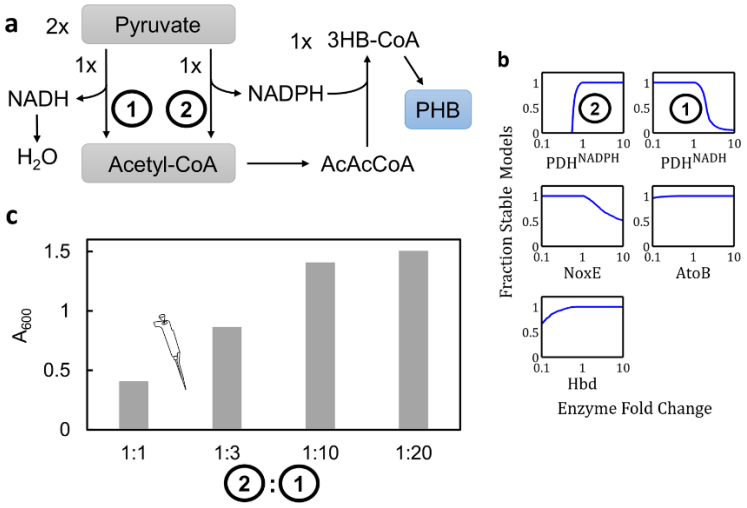

Fig 2. From: Theisen, M. K., Lafontaine Rivera, J. G., & Liao, J. C. (2016). Stability of Ensemble Models Predicts Productivity of Enzymatic Systems. PLOS Computational Biology, 12(3). This pathway shows the analysis of a pathway **(a)** for the production of poly-hydroxybutyrate. The robustness analysis **(b)** showed that enzyme 2 causes instability at low amounts while enzyme 1

causes instability at high amounts. A previous, unrelated literature report **(c)** showed production increased as the enzyme 2:enzyme 1 ratio increased.

Importantly, although some of the phenomena were experimentally determined, it was not necessarily known that instability of the system—causing a step change in the nature of the steady state, rather than a smooth change predictable by sensitivity analysis—could be an underlying reason. Additionally, it was significant that increase of enzyme activity was found to cause instability, since many typical metabolic engineering strategies involve simply overexpressing all enzymes as much as possible. Overall these applications make robustness and stability critical considerations for model and pathway design. Fig 2, from (Theisen et al., 2016), shows the analysis of one of these enzymatic pathways, previously characterized in an unrelated work (Opgenorth, Korman, & Bowie, 2014).

Deliverable 3: Entropy index

Natural metabolic pathways may be presumed to be at least bifurcationally robust against stochastic changes in protein expression levels. Thus, the models of natural metabolic pathways should be similarly robust. We have provided a quantitative way to characterize bifurcational robustness in the presence of random parameter changes. Without a quantitative index, optimization of models for bifurcational robustness becomes difficult, if not impossible. Therefore, our goal here was to develop a quantitative index for bifurcational robustness, and show that such an index enables the optimization of the bifurcational robustness of metabolic models.

The developed index is easy to compute and applies to metabolic systems of various scale and complexity. Interestingly, the mathematical form of our robustness index resembles the definition of entropy in thermodynamics and information theory. We have shown that this entropy-like index, denoted as S, correlates with empirically-measured bifurcational robustness (Lafontaine Rivera et al., 2015). Metabolic systems with a small S are highly robust against bifurcation, and are more likely to retain a steady state under random perturbations affecting every enzyme than systems with a large S.

Deliverable 4: *E. coli* model

Using the presumption that natural pathways should be stable and thus relatively low entropy, we can identify kinetic models and even pathways within models which are likely to have. We have used this approach and shown that stability does indeed identify potential problems during model construction. We demonstrated these capabilities by using an ensemble model of E. coli consisting of 193 reactions which encompass glycolysis, the TCA cycle, the synthesis of all 20 canonical amino acids as well as nucleotides. Starting from a computer generated model using stoichiometric and regulatory data from the EcoCyc database, the model was analyzed

using entropy as a measure of bifurcational robustness. Using this index and on the premise that adequate models of cellular metabolism should be sufficiently stable & robust, pathways with high numbers of high entropy enzymes were inspected for correct incorporation of regulatory information from the EcoCyc database. Three changes were made to the kinetic rate laws, while leaving the network stoichiometry unchanged. After these changes were made, entropy of the system significantly decreased. These changes were verified with reference to literature reports of individual enzyme characterization, which had apparently been wrongly incorporated in the EcoCyc database.

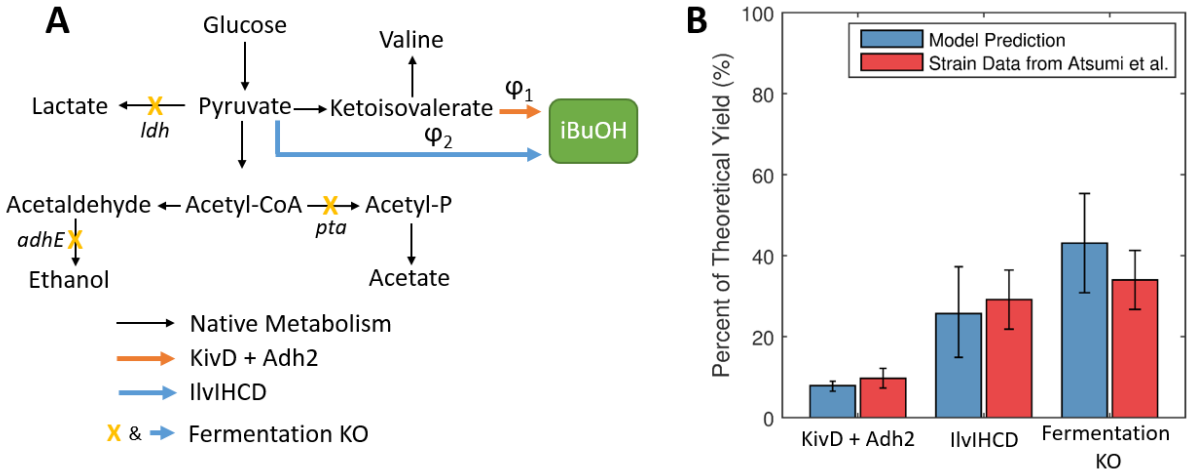

Fig 3. From submitted work. **(A)** shows the genotypes of strains characterized and **(B)** show the yield from each strain determined experimentally (red) as well as from ensemble model prediction using the *E. coli* model.

Following model editing, the model was used to predict isobutanol production in strains that had been characterized in a previous, unrelated work (Atsumi, Hanai, & Liao, 2008) and the simulation results showed good agreement with experiment. Code for the simulation is available at: https://github.com/theis188/Rivera-theisen-paper-code

Comparison with project goals

The project results correspond strongly with the original stated goals. The original goals encompassed the development of kinetic modeling using particular methods for parameter sampling and ensemble construction. The original goals included constructing a model of *E. coli* and using simulation to model biofuel production. All of these have been achieved.

Summary of model mathematics

Rate laws are defined according to Michalis-Menten style enzyme saturation kinetics. The modular rate laws proposed by Liebermeister are used (Liebermeister, Uhlendorf, & Klipp, 2010).

Reference steady state is defined such that:

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{S}\boldsymbol{v}(\boldsymbol{X}, \boldsymbol{p}) = \boldsymbol{F}(\boldsymbol{X}, \boldsymbol{p}) = \mathbf{0} \tag{1}$$

Where S is the stoichiometric matrix, and flux v is a function of metabolite concentration X and parameter values p. To construct an ensemble Model parameters are sampled at random, constrained to the steady state. Dynamic stability is ensured by calculating the eigenvalues of the Jacobian at reference state for each model in the ensemble.

Perturbations in the model are made using the continuation method which is justified as follows:

$$\frac{d\boldsymbol{X_{ss}}}{dt} = \boldsymbol{F}(\boldsymbol{X_{ss}}, \boldsymbol{p}) = \mathbf{0} \tag{2}$$

Where Xss is steady state metabolite concentrations. F = 0 because the system is constrained to steady state. Taking a total derivative yields:

$$\frac{d\boldsymbol{F}}{d\boldsymbol{p}} = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{X_{ss}}} \frac{d\boldsymbol{X_{ss}}}{d\boldsymbol{p}} + \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{p}} = \mathbf{0} \tag{3}$$

From which it follows that:

$$\frac{d\boldsymbol{X_{ss}}}{d\boldsymbol{p}} = -\left(\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{X_{ss}}}\right)^{-1} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{p}} \tag{4}$$

This formula can be used to determine the effect of parameters, p, on steady state metabolite concentration, Xss. This equation can be integrated with respect to p to generate the continuous variation of Xss as a function of p. Once Xss is determined, the steady-state flux can be calculated. The integration stops when the Jacobian matrix, $\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{X_{ss}}}$, becomes singular, or when a metabolite value becomes negative.

Code for this modeling strategy is available at: https://github.com/theis188/Rivera-theisen-paper-code

Conclusion and Next Steps

Ensemble modeling for kinetic metabolic models has been demonstrated as a plausible, important, and convenient method for modeling systems for which limited information is available. The work included theoretical development and practical application, and ultimately culminated in a cell-wide kinetic ensemble model of *E. coli* which successfully represents. Future work can build on these foundations and aim towards the construction of models for further organisms, or organisms under different conditions for specific engineering applications.

Atsumi, S., Hanai, T., & Liao, J. C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*, *451*(7174), 86–9. doi:10.1038/nature06450

Lafontaine Rivera, J. G., Lee, Y., & Liao, J. C. (2015). An entropy-like index of bifurcational robustness for metabolic systems. *Integr. Biol.* doi:10.1039/C4IB00257A

Lee, Y., Lafontaine Rivera, J. G., & Liao, J. C. (2014). Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. *Metabolic Engineering*, 1–9. doi:10.1016/j.ymben.2014.06.006

Liebermeister, W., Uhlendorf, J., & Klipp, E. (2010). Modular rate laws for enzymatic reactions: Thermodynamics, elasticities and implementation. *Bioinformatics*, *26*(12), 1528–1534. doi:10.1093/bioinformatics/btq141

Opgenorth, P. H., Korman, T. P., & Bowie, J. U. (2014). A synthetic biochemistry molecular purge valve module that maintains redox balance. *Nature Communications*, *5*(May), 4113. doi:10.1038/ncomms5113

Theisen, M. K., Lafontaine Rivera, J. G., & Liao, J. C. (2016). Stability of Ensemble Models Predicts Productivity of Enzymatic Systems. *PLOS Computational Biology*, *12*(3), e1004800. doi:10.1371/journal.pcbi.1004800

List of Publication:

Lafontaine Rivera, J. G., Lee, Y., & Liao, J. C. (2015). An entropy-like index of bifurcational robustness for metabolic systems. Integr. Biol. http://doi.org/10.1039/C4IB00257A

Lee, Y., Lafontaine Rivera, J. G., & Liao, J. C. (2014). Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. Metabolic Engineering, 1–9. http://doi.org/10.1016/j.ymben.2014.06.006

Theisen, M. K., Lafontaine Rivera, J. G., & Liao, J. C. (2016). Stability of Ensemble Models Predicts Productivity of Enzymatic Systems. PLOS Computational Biology, 12(3), e1004800. http://doi.org/10.1371/journal.pcbi.1004800