Modeling high-throughput cell-free experiments for metabolic engineering of biofuels

Motivation: In 2017, the Department of Defense used over 85 million barrels of fuel—more than any other entity in the world—at a cost of over $8 billion.[1] Rising energy demands, a finite supply of petroleum-based fuels, and the need for energy security require the development and use of alternative energy sources. In addition to heavy investment in alternative fuels, the Department of Defense recognizes this critical issue through research initiatives; the Army Research Laboratory's Materials Research Campaign's Energy and Power thrust[2] focuses on "novel alternative energy solutions at lower cost" and the Biotechnology thrust focuses on "new biological materials derived through synthetic biology" and generating power from organic sources.[3] Biological fuel production presents a renewable option that requires lower energy costs, minimizes safety production hazards, and utilizes a much wider range of raw materials than traditional fuel sources.[4] For example, the engineered non-model bacterium *Clostridia autoethanogenum* can utilize a mixture of carbon monoxide, carbon dioxide, and hydrogen known as synthesis gas, which is typically inefficiently burned, as its sole feed source to produce ethanol and acetate, both value-added products.[5] However, such non-model microbes are slower growing, less well-characterized, and less compatible with current genetic engineering tools compared to model organisms like *E. coli*, which lack the ability to utilize synthesis gas or most other alternative fuel sources.[5,6] Additionally, the same complexity of metabolic reactions that enable these microbes to produce a vast array of interesting compounds from a diverse, otherwise non-processable set of precursors also makes the task of experimentally testing every possible combination of enzyme levels infeasible. **To address the challenge of creating new sustainable sources of fuels for national defense, I aim to create a computational model that utilizes cell-free experimental data to direct the rapid engineering of bacterially synthesized biofuels.**

Cell-free systems, which have seen recent widespread adoption, present an opportunity to accelerate the "design-build-test" cycle in metabolic engineering[7,8]. Traditionally, this cycle involves identifying a small number of engineering targets through experience or intuition, making these individual changes through genetic engineering, growing the engineered cells, and then testing the efficacy of these changes in the context of background cell behavior. By contrast, cell-free systems, in which cells are split open to extract the machinery essential for protein synthesis, do not need to grow or compete for resources like living cells. Users of cell-free systems can add varying amounts of DNA to this machinery, thereby precisely controlling the levels of each enzyme in a metabolic pathway or network. Thus, these systems enable the optimization of biological engineering objectives without competing with normal cellular processes.[8] Cell-free systems have successfully been applied to the rapid development of small enzymatic pathways for metabolic engineering purposes.[7] However, this strategy still requires that all enzyme concentrations are independently varied to optimize metabolite production, and the large number of enzymes in many pathways quickly renders a true combinatorial approach impossible. In such cases, computational methods offer a high-throughput solution to direct experimental efforts and narrow the search space of potential test conditions.

Kinetic models, which explicitly define reaction rates as functions of enzyme and metabolite concentrations, are ideal for predicting the behavior of systems with varying enzyme levels[9]. Enzyme overexpression data provides a basis for inferring the kinetic parameters of reactions. Furthermore, by representing enzymatic reactions as elementary steps using mass action kinetics, this modeling framework can capture a wide variety of reaction types and enzyme inhibitions without requiring any assumptions as to the true reaction mechanism. The drawback of kinetic models is that they introduce many more model parameters that must be inferred. The strategy of creating and subsequently screening many possible sets of model parameters, known as ensemble modeling, is one of several strategies that has been introduced to address this problem and has been used by the Tyo lab at Northwestern to solve a broad range of engineering challenges, including predicting beneficial enzyme deletions and growth conditions.[5,10]

Research Plan: I will construct a kinetic ensemble model of cell-free metabolic pathways to inform engineering strategies for bacterial production of butanol, a simple biofuel. This model will be validated by predicting optimum conditions for butanol production in cell-free systems (Figure 1). Available data of cell-free n-butanol production, provided by the Jewett group at Northwestern, and cross-validation methods will be used to predict experimental data and suggest possible future experimental conditions. The validated model will then be applied to the metabolism of *Clostridia autoethanogenum* (*C. auto*) to predict metabolic engineering strategies to enhance biofuel production. The specific aims of this project are as follows:

Aim 1: I will develop a kinetic ensemble model of the *n*-butanol pathway. This work will leverage the model-building expertise of the Tyo group at Northwestern and will be trained on data from the Jewett group. This aim will only consider the standalone linear pathway, which will be treated as a baseline result.

Aim 2: I will next add mechanistic details of the system to the modeling framework, including central carbon metabolism, cofactor concentrations, and known enzyme inhibition from substrates. I will then validate the full model against experimental results.

Aim 3: Lastly, I will train the validated model on *C. auto* cell-free systems. Using the final model, I will ultimately suggest engineering targets to optimize biofuel production from synthesis gas.

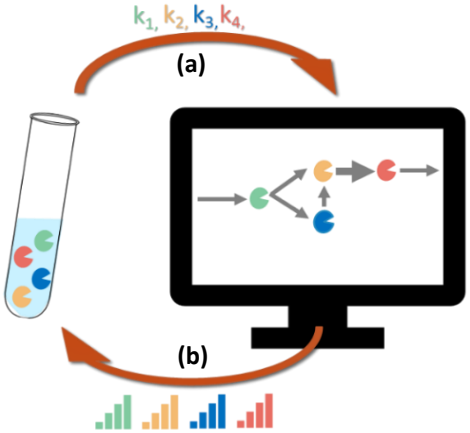

Figure 1: Kinetic models of cell-free metabolism will inform metabolic engineering. **(a)** Cell-free experiments will provide parameter inference and validation for kinetic models, **(b)** which in turn predict optimum enzyme levels for production of metabolites, such as biofuels

Aim 1: Develop a baseline kinetic ensemble model of the standalone *n*-butanol pathway

Overview: Ensemble modeling is a method to address the issue of parameter estimation in kinetic modeling. Because there are multiple possible sets of model parameters that are consistent with the experimental data, this method generates many of these sets into an initial ensemble. Figure 2(a) shows an example of the initial ensemble that I generated for the n-butanol pathway. For each set, the method, developed by Tran, *et al.*,[11] compares the system behavior at some perturbation state, such as at varied levels of enzyme concentration. My model then removes any parameter sets that predict outcomes inconsistent with the perturbation data, reducing the possible parameters sets to that of those shown in Figure 2(b). This procedure is repeated until a final minimal parameter set ensemble remains, each of which correctly predicts all experimental data.

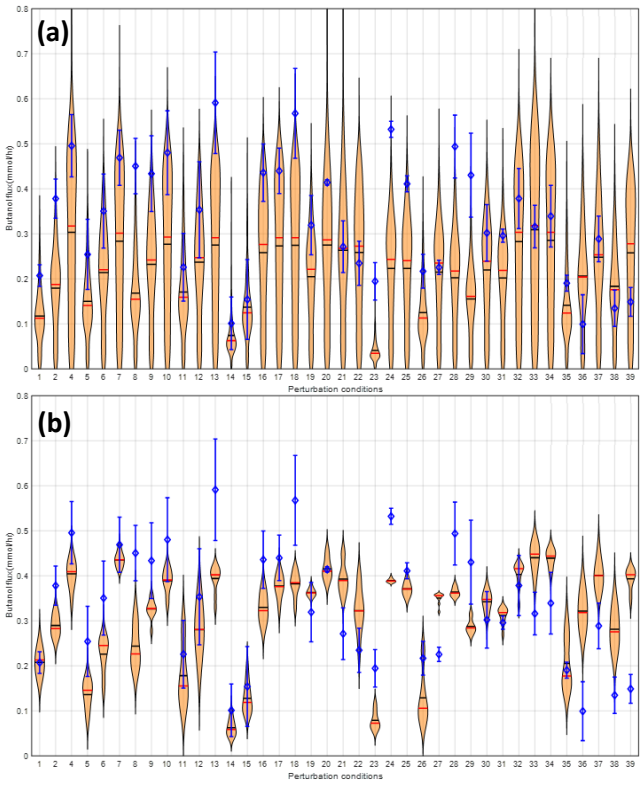

Figure 2: Screening parameter sets against experimental perturbations improves prediction of flux states. **(a)** 1000 parameter sets before screening show a wide range of predicted fluxes (shown by orange violin plots, which are histograms of 1000 predictions from each parameter set) which do not consistently match experimental data (blue box and whisker) across 38 combinations of enzyme levels. **(b)** Screening parameter-sets against experimental data leaves the 10 best parameter sets and improves the butanol flux predictions.

Research Design and Methods: The initial model will consist of only the six reactions that metabolize acetyl-CoA to *n*-butanol. I am currently computationally automating the inclusion of enzyme promiscuity—the ability of some enzymes to act on multiple substrates—which has not previously been used in elementary reaction models, and which I have shown dramatically improves predictions in similar systems.

Expected Outcomes and Alternative Directions: Because this first model includes minimal biological context, it is expected to have only qualitative fit; however, this will provide a baseline for later detailed models. If the model is unable to provide qualitative agreement, I will include statistical tools to account for variation between experimental replicates.

Aim 2: Expand the model to include non-idealities of cell-free systems.

Overview: Because cell-free systems are not purified after cells are split open, they contain some amount of native enzymes that have been shown to support central carbon metabolism.[8] Additionally, pH and cofactor concentrations influence these systems. In this aim, I will incorporate the butanol pathway in the larger context of cell-free metabolism, which will enable more accurate predictions of butanol production by accounting for system-wide dynamics.

Research Design and Methods: Based on the *E. coli* consensus genome-scale model, k-ecoli457, I will use my automated reaction generation tools to build a model of all reactions present in the cell-free system. Using radioisotope-labeling techniques available to collaborators in the Jewett lab at Northwestern, I will track reaction rates of key metabolites and determine the steady-state rate of all metabolic reactions in the cell-free system. Determining these rates is necessary to create the initial ensemble in the ensemble modeling method. The fully developed model will be trained on existing butanol pathway overexpression data, and I will use statistical methods to cross-validate the results at different experimental conditions.

Expected Outcomes and Alternative Directions: The increased context in this aim will likely provide a vastly improved prediction over the baseline model created in Aim 1. Comparison to Aim 1 will also help to determine which effects are due to system-wide dynamics, or due to the butanol pathway alone. This improvement is expected to validate both the ensemble modeling method and the use of the standard set of *E. coli* metabolic reactions for applications in cell-free systems. If the model fails to provide predictive accuracy, I will trim the reactions down to a minimum essential set by employing network-reduction techniques, which have been shown to achieve quicker, more robust solutions while still providing consistent system behavior.[12]

Aim 3: Apply the validated ensemble model method to *Clostridia autoethanogenum* cell-free systems.

Overview: Because *C. auto* is uniquely able to utilize synthesis gas but lacks many of the well-developed genetic engineering tools available for use in *E. coli*, cell-free systems will be indispensable for rapid development and prototyping of enzyme levels to convert renewable feed sources into biofuels. Additionally, *C. auto* only grows in anaerobic conditions, making experiments difficult and modeling solutions more desirable. In collaboration with the Jewett lab, which is currently developing *C. auto* cell-free systems, I will create a model to predict the result of enzyme overexpression on different metabolic pathways in these systems.

Research Design and Methods: In this aim, I will apply our developed *E. coli* model to the set of reactions in the iCLAU786 consensus *C. auto* genome-scale metabolism. Once trained against a set of experimental data at varied enzyme level combinations, I will use the model to predict likely engineering strategies to optimize biofuel production. By performing enzyme sensitivity analysis, wherein I calculate the change in butanol production with respect to each enzyme level, I will determine which enzymes are rate limiting for a pathway or exhibit the strongest influence over a network, thus suggesting beneficial changes in enzyme levels.

Expected Outcomes and Alternative Directions: The work in this aim will predict novel enzyme targets for the increased production of biofuels from synthesis gas. However, if the results of the sensitivity analysis are inconclusive or if the proposed engineering strategies do not have the desired effect, I will carry out a thermodynamic reversibility analysis, which will identify thermodynamic bottlenecks of the system independent of sensitivity analysis.[11] Using this strategy, I will identify enzymes with unfavorable thermodynamics which can then be improved with protein engineering or the substitution of similar enzymes from related organisms.

Defense Applications and Broader Impacts: This research has already produced encouraging preliminary results; I have successfully screened 200,000 parameter sets against 38 experimental perturbations, resulting in a final ensemble of 10 parameter sets that accurately capture butanol flux across all conditions (Figure 2). In this work, I have successfully incorporated the acceleration strategies developed in the Tyo lab at Northwestern, which include identification of conserved metabolites and rejection of unstable flux states.[10] These strategies decrease computational time and ensure that remaining parameter sets exhibit increased robustness to system changes. These results show the capacity of this project to broadly increase the speed and productivity of metabolic engineering efforts by suggesting more informative experiments and expediting the design cycle for cell-free systems, ultimately allowing for rapid biofuel development from alternative feed sources.

The computational design process used to create this model can be readily adapted to other metabolic engineering targets, including cell-free systems derived from other organisms or biosynthetic pathways for products other than biofuels, such as small molecule drugs. Furthermore, a successful mechanistic model of *in vitro* cell-free metabolism would allow comparison and translation of cell-free experimental data to living cell experiments, enabling increased understanding of system-level effects and biological insight in these complex systems. This would help create robust and predictable synthetic biology systems, which is a major thrust of the ARO Synthetic Biology Program[2]. More complex models, aided by increased understanding from this project's model of cell-free systems and their accelerated design-build-test cycles, can be rapidly implemented in a wide range of applications beyond metabolic engineering, including pharmaceuticals, antibiotics, and other high-impact products crucial to the modern warfighter.

References:

1. Operational Energy. *Office of the Assistant Secretary of Defense for Energy, Installations, and Environment* (2018). Available at: https://www.acq.osd.mil/eie/OE/OE_index.html.
2. Army Research Office. *Broad Agency Announcement for Fundamental Research W911NF-17-S-0002-03 01 April 2017 – 31 March 2022*.
3. *Army Research Laboratory Technical Strategy: 2015-2035*.
4. Wook lee, J. *et al.* systems metabolic engineering of microorganisms for natural and non-natural chemicals. (2012). doi:10.1038/nchembio.970
5. Greene, J., Daniell, J., Köpke, M., Broadbelt, L. & Tyo, K. E. J. Kinetic ensemble modelling of gas-fermenting Clostridium autoethanogenum for improved ethanol production.
6. Valgepea, K. *et al.* Maintenance of ATP Homeostasis Triggers Metabolic Shifts in Gas-Fermenting Acetogens. *Cell Syst.* **4,** 505–515.e5 (2017).
7. Karim, A. S. & Jewett, M. C. A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metab. Eng.* **36,** 116–126 (2016).
8. Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J. & Swartz, J. R. An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol.* **4,** (2008).
9. Saa, P. A. & Nielsen, L. K. Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnol. Adv.* **35,** 981–1003 (2017).
10. Greene, J. L., Wäechter, A., Tyo, K. E. J. & Broadbelt, L. J. Acceleration Strategies to Enhance Metabolic Ensemble Modeling Performance. *Biophys. J.* **113,** 1150–1162 (2017).
11. Tran, L. M., Rizk, M. L. & Liao, J. C. Ensemble modeling of metabolic networks. *Biophys. J.* **95,** 5606–5617 (2008).
12. Strutz, J., **Martin, J.**, Greene, J., Tyo, K. & Broadbelt, L. J. Opportunities and Challenges in the Application of Kinetic Metabolic Models. *Curr. Opin. Biotechnol.* (Submitted)