Sequence Specific Constraints Based Modeling of TX-TL Synthetic Circuits

Jeffrey D. Varner*

School of Chemical and Biomolecular Engineering

Cornell University, Ithaca NY 14853

Running Title: Constraints based models of synthetic circuits

To be submitted: *Scientific Reports*

*Corresponding author:

Jeffrey D. Varner,

Professor, School of Chemical and Biomolecular Engineering,

244 Olin Hall, Cornell University, Ithaca NY, 14853

Email: jdv27@cornell.edu

Phone: (607) 255 - 4258

Fax: (607) 255 - 9166

Abstract

In this study, we used sequence specific constraints based modeling to evaluate the performance of synthetic circuits in an *E. coli* TX-TL system. A core *E. coli* metabolic model, consisting of XX metabolites and YY reactions, was developed from literature [REF]. This model, which described glycolysis, pentose phosphate pathway, amino acid biosynthesis and degradation and energy metabolism, was then augmented with sequence specific descriptions of genetic circuits which included mechanistic models of promoter function, transcription and translation. Thus, unlike other synthetic biology modeling efforts, sequence specific constraints based modeling explicitly couples the transcription and translation of circuit components with the availability of metabolic resources. Model parameters were largely taken from literature; our approach had very few adjustable parameters thereby allowing the a first principles prediction of circuit performance. We tested this approach by first simulating σ_{70}-induced deGFP expression and then expanded these studies to more complex multicomponent circuits. First principles predictions of circuit performance were consistent with measurements for a variety of cases. Further, global sensitivity analysis identified the key metabolic processes that controlled circuit performance. Taken together, sequence specific constraints based modeling offers a novel means to *a priori* estimate the performance of cell free synthetic circuits.

Keywords: Synthetic biology, Constraints based modeling, Biochemical modeling

Introduction

Cell free systems offer many advantages for the study, manipulation and modeling of metabolism compared to *in vivo* processes. Central amongst these advantages is direct access to metabolites and the microbial biosynthetic machinery without the interference of a cell wall. This allows us to control as well as interrogate the chemical environment while the biosynthetic machinery is operating, potentially at a fine time resolution. Second, cell-free systems also allow us to study biological processes without the complications associated with cell growth. Cell-free protein synthesis (CFPS) systems are arguably the most prominent examples of cell-free systems used today [**?**]. However, CFPS is not new; CFPS in crude *E. coli* extracts has been used since the 1960s to explore fundamentally important biological mechanisms [**?** **?**]. Today, cell-free systems are used in a variety of applications ranging from therapeutic protein production [**?**] to synthetic biology [**?**]. Interestingly, many of the challenges confronting in-vivo genome-scale kinetic modeling can potentially be overcome in a cell-free system. For example, there is no complex transcriptional regulation to consider, transient metabolic measurements are easier to obtain, and we no longer have to consider cell growth. Thus, cell-free operation holds several significant advantages for model development, identification and validation. Theoretically, genome-scale cell-free kinetic models may be possible for industrially important organisms, such as *E. coli* or *B. subtilis*, if a simple, tractable framework for integrating allosteric regulation with enzyme kinetics can be formulated.

Stoichiometric reconstructions of microbial metabolism popularized by constraint based modeling techniques such as flux balance analysis (FBA) have become standard tools to interrogate biological networks [**?**]. Since the first genome-scale stoichiometric model of *E. coli*, developed by Edwards and Palsson [**?**], stoichiometric reconstructions of hundreds of organisms, including industrially important prokaryotes such as *E. coli* [**?**] or *B. subtilis* [**?**], are now available [**?**]. Stoichiometric models rely on a pseudo-steady-state

assumption to reduce unidentifiable genome-scale kinetic models to an underdetermined linear algebraic system, which can be solved efficiently even for large systems using linear programming. Traditionally, stoichiometric models have also neglected explicit descriptions of metabolic regulation and control mechanisms, instead opting to describe the choice of pathways by prescribing an objective function on metabolism. Interestingly, similar to early cybernetic models, the most common metabolic objective function has been the optimization of biomass formation [**?**], although other metabolic objectives have also been estimated [**?**]. Recent advances in constraint-based modeling have overcome the early shortcomings of the platform, including capturing metabolic regulation and control [**?**]. Thus, modern constraint-based approaches are extremely useful for the discovery of metabolic engineering strategies and represent the state of the art in metabolic modeling [**? ?**].

In this study, we used sequence specific constraints based modeling to evaluate the performance of synthetic circuits in an *E. coli* TX-TL system. A core *E. coli* metabolic model, consisting of XX metabolites and YY reactions, was developed from literature [REF]. This model, which described glycolysis, pentose phosphate pathway, amino acid biosynthesis and degradation and energy metabolism, was then augmented with sequence specific descriptions of genetic circuits which included mechanistic models of promoter function, transcription and translation. Thus, unlike other synthetic biology modeling efforts, sequence specific constraints based modeling explicitly couples the transcription and translation of circuit components with the availability of metabolic resources. Model parameters were largely taken from literature; our approach had very few adjustable parameters thereby allowing the a first principles prediction of circuit performance. We tested this approach by first simulating σ_{70}-induced deGFP expression and then expanded these studies to more complex multicomponent circuits. First principles predictions of circuit performance were consistent with measurements for a variety of cases. Further, global

sensitivity analysis identified the key metabolic processes that controlled circuit performance. Taken together, sequence specific constraints based modeling offers a novel means to *a priori* estimate the performance of cell free synthetic circuits.

Results

Discussion

Materials and Methods

Formulation and solution of the model equations. The flux balance analysis problem was formulated as:

$$\max_{\boldsymbol{w}} \left(w_{obj} = \boldsymbol{\theta}^T \boldsymbol{w}\right)$$

$$\text{Subject to : } \mathbf{Sw} = \mathbf{0}$$

$$\alpha_i \leq w_i \leq \beta_i \qquad i = 1, 2, \ldots, \mathcal{R}$$

where S denotes the stoichiometric matrix, $\mathbf{w}$ denotes the unknown flux vector, $\boldsymbol{\theta}$ denotes the objective selection vector and α_i and β_i denote the lower and upper bounds on flux w_i, respectively. The flux balance analysis problem was solved using the GNU Linear Programming Kit (v4.52) [**?**]. In this study, the objective w_{obj} was to maximize the production of circuit output.

Transcription and translation template reactions.

Global sensitivity analysis. We conducted a global sensitivity analysis, using the variance-based method of Sobol, to estimate which parameters controlled the performance of synthetic circuits [**?**]. We computed the total sensitivity index of each parameter relative to two performance objectives, the peak thrombin time and the area under the thrombin curve (thrombin exposure). We established the sampling bounds for each parameter from the minimum and maximum value of that parameter in the parameter set ensemble. We used the sampling method of Saltelli *et al.* [**?**] to compute a family of $N\,(2d + 2)$ parameter sets which obeyed our parameter ranges, where N was the number of trials, and d was the number of parameters in the model. In our case, N = 10,000 and d = 22, so the total sensitivity indices were computed from 460,000 model evaluations. The variance-based sensitivity analysis was conducted using the SALib module encoded in the Python programming language [**?**].

Acknowledgements

This study was supported by an award from the Army Research Office (ARO #59155-LS).

References