

# Effective Dynamic Models of Metabolic Networks

MICHAEL VILKHOVOY, MASON MINOT, AND JEFFREY D. VARNER

Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14850 USA

CORRESPONDING AUTHOR: J. D. VARNER (jdv27@cornell.edu)

**ABSTRACT** Mathematical models of biochemical networks are the useful tools to understand and ultimately predict how cells utilize nutrients to produce valuable products. Hybrid cybernetic models (HCMs) in combination with elementary modes (EMs) are a tool to model cellular metabolism. However, HCM is limited to reduced metabolic networks because of the computational burden of calculating EMs. In this letter, we develop the hybrid cybernetic modeling with flux balance analysis (HCM-FBA) technique, which uses flux balance solutions instead of EMs to dynamically model metabolism. We show that HCM-FBA has comparable performance to HCM for a proof of concept metabolic network and for a reduced anaerobic *E. coli* network. Next, HCM-FBA is applied to a larger metabolic network of aerobic *E. coli* metabolism, which was infeasible for HCM (29 FBA modes versus more than 153 000 EMs). The global sensitivity analysis further reduces the number of FBA modes required to describe the aerobic *E. coli* data, while maintaining model fit. Thus, HCM-FBA is a promising alternative to HCM for large networks, where the generation of EMs is infeasible.

**INDEX TERMS** Cybernetic models, flux balance analysis (FBA), metabolic models.

## I. INTRODUCTION

**B**IOTECHNOLOGY harnesses the power of metabolism to produce products that benefit society. Constraints-based models are the important tools to understand and ultimately to predict how cells utilize nutrients to produce products. Constraints-based methods, such as flux balance analysis (FBA) [1], and network decomposition approaches, such as elementary modes (EMs) [2] or extreme pathways (EPs) [3], model intracellular metabolism using the biochemical stoichiometry and other constraints, such as thermodynamical feasibility under pseudosteady-state conditions. FBA has been used to efficiently estimate the performance of metabolic networks of arbitrary complexity, including genome scale networks, using linear programming [4]. On the other hand, EMs (or EPs) catalog all possible metabolic behaviors such that any flux distribution predicted by FBA is a convex combination of the EMs (or EPs) [5]. However, the calculation of EMs (or EPs) is computationally expensive and currently infeasible for genome scale networks [6].

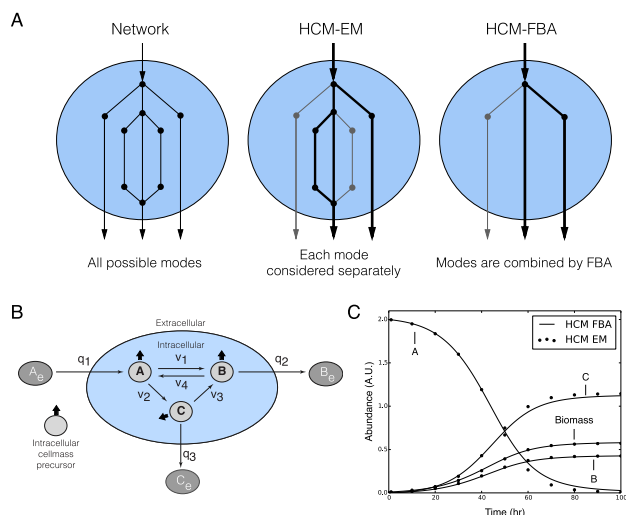
Cybernetic models are an alternative to the constraints-based approach, which hypothesize that the metabolic control is the output of an optimal decision. Cybernetic models have predicted mutant behavior [7], [8], steady-state multiplicity [9], and strain-specific metabolism [10], and have been used in bioprocess control applications [11]. Hybrid cybernetic models (HCMs) have addressed earlier shortcomings of the approach by integrating cybernetic optimality concepts with EMs. HCMs dynamically choose the combinations of biochemical modes (each catalyzed by a pseudoenzyme, whose expression is controlled by an optimal decision) to achieve a physiological objective [Fig. 1(A)]. HCMs gen-

erate intracellular flux distributions consistent with other approaches, such as metabolic flux analysis (MFA), and also describe dynamic extracellular measurements superior to dynamic FBA (DFBA) [12]. However, HCMs are restricted to networks, which can be decomposed into EMs (or EPs).

In this letter, we develop the hybrid cybernetic modeling with FBA (HCM-FBA) technique. HCM-FBA is a modification of the hybrid cybernetic approach of Kim *et al.* [12], which uses FBA solutions (instead of EMs) in conjunction with cybernetic control variables to dynamically simulate metabolism. Since HCM showed superior performance to DFBA, we compared the performance of HCM-FBA with HCM for a prototypical metabolic network, along with two real-world *E. coli* applications. HCM-FBA performed comparably to HCM for the prototypical network and a reduced anaerobic *E. coli* network, despite having a fewer parameters in each case. Next, HCM-FBA was applied to an aerobic *E. coli* metabolic network that was infeasible for HCM. HCM-FBA described cell mass growth and the shift from glucose to acetate consumption with only a few modes. The global sensitivity analysis allowed us to further reduce the aerobic *E. coli* HCM-FBA model to the minimal model required to describe the data. Thus, HCM-FBA is a promising approach for the development of reduced order dynamic metabolic models and a viable alternative to HCM or DFBA, especially for large networks, where the generation of EMs is infeasible.

## II. RESULTS

HCM-FBA was equivalent to HCM for a prototypical metabolic network (Fig. 1). The proof of concept network, consisting of six metabolites and seven reactions [Fig. 1(B)],

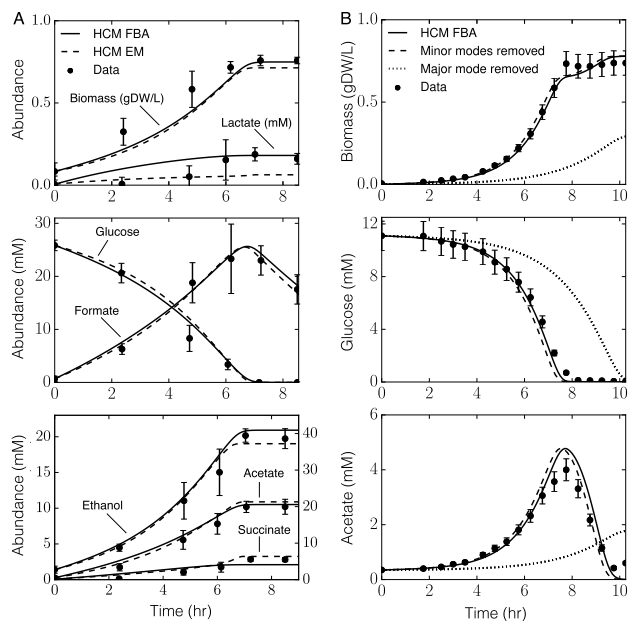


**FIGURE 1.** HCM proof of concept metabolic study. (A) HCMs distribute uptake and secretion fluxes amongst different pathways. For HCM, these pathways are EMs; for HCM-FBA, these are FBA solutions. HCM combines all possible modes within a network; whereas HCM-FBA combines only steady-state paths estimated by FBA. (B) Prototypical network with six metabolites and seven reactions. Intracellular cell mass precursors  $A$ ,  $B$ , and  $C$  are balanced (no accumulation) while the extracellular metabolites ( $A_e$ ,  $B_e$ , and  $C_e$ ) are not balanced (can accumulate). The oval denotes the cell boundary,  $q_j$  is the  $j$ th flux across the boundary, and  $v_k$  denotes the  $k$ th intracellular flux. (C) Simulation of extracellular metabolite trajectories using HCM-FBA (solid line) versus HCM (points) for the prototypical network.

generated three FBA modes and six EMs. Using the EMs and synthetic parameters, we generated test data from which we estimated the HCM-FBA model parameters. The best fit HCM-FBA model replicated the synthetic data [Fig. 1(C)]. The HCM and HCM-FBA kinetic parameters were not quantitatively identical, but had similar orders of magnitude; the FBA approach had three fewer modes; thus, identical parameter values were not expected. The HCM-FBA approach replicated synthetic data generated by HCM, despite having three fewer modes. Thus, we expect that HCM-FBA will perform similar to HCM, despite having fewer parameters. Next, we tested the ability of HCM-FBA to replicate real-world experimental data.

The performance of HCM-FBA was equivalent to HCM for anaerobic *E. coli* metabolism [Fig. 2(A)]. We constructed an anaerobic *E. coli* network [12], consisting of 12 reactions and 19 metabolites, which generated seven FBA modes and nine EMs. HCM reproduced cell mass, glucose, and by product trajectories using the kinetic parameters reported by Kim et al. [12] [Fig. 2(A), points versus dashed]. HCM-FBA model parameters were estimated in this letter from the Kim et al. [12] data set using simulated annealing. Overall, HCM-FBA performed within 5% of HCM (on a residual standard error basis) for the anaerobic *E. coli* data [Fig. 2(A) (solid line)], despite having two fewer modes and four fewer parameters (17 versus 21 parameters). Thus, while both HCM and HCM-FBA described the experimental data, HCM-FBA did so with fewer modes and parameters.

HCM-FBA captured the shift from glucose to acetate consumption for a model of aerobic *E. coli* metabolism that was infeasible for HCM [Fig. 2(B)]. An *E. coli* metabolic network (60 metabolites and 105 reactions) was constructed



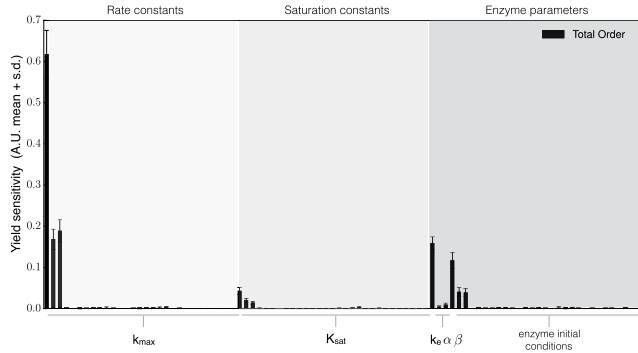
**FIGURE 2.** HCM-FBA versus HCM performance for small and large metabolic networks. (A) Batch anaerobic *E. coli* fermentation data versus HCM-FBA (solid line) and HCM (dashed line). The experimental data was reproduced by Kim et al. [12]. Error bars represent the 90% confidence interval. (B) Batch aerobic *E. coli* fermentation data versus HCM-FBA (solid line). Model performance is also shown when minor modes (dashed line) and major modes (dotted line) were removed from the HCM-FBA model. The experimental data was reproduced by Varma and Palsso [13]. Error bars denote a 10% coefficient of variation.

from the literature [14], [15]. EM decomposition of this network (and thus HCM) was not feasible; 153 000 EMs were generated before the calculation became infeasible. Conversely, FBA generated only 29 modes for the same network. HCM-FBA model parameters were estimated from cell mass, glucose, and acetate measurements [13] using simulated annealing [Fig. 2(B) (solid line)]. HCM-FBA captured glucose consumption, cell mass formation, and the switch to acetate consumption following glucose exhaustion. HCM-FBA described the dynamics of a network that was infeasible for HCM, thereby demonstrating the power of the approach for large networks. Next, we demonstrated a systematic strategy to identify the critical subset of FBA modes required for model performance.

Global sensitivity analysis identified the FBA modes essential to model performance (Fig. 3). Total-order sensitivity coefficients were calculated for all kinetic parameters and enzyme initial conditions in the aerobic *E. coli* model. Five of the 29 FBA modes were significant; removal of the most significant of these modes (encoding aerobic growth on glucose) destroyed model performance [Fig. 2(B) (dotted line)]. Conversely, removing the remaining 24 modes simultaneously had a negligible effect upon model performance [Fig. 2(B) (dashed line)]. The sensitivity analysis identified the minimal model structure required to explain the experimental data.

### III. DISCUSSION

In this letter, we developed HCM-FBA, an effective modeling technique to simulate metabolic dynamics. HCM-FBA



**FIGURE 3. Global sensitivity analysis of the aerobic *E. coli* model.** Total-order variance-based sensitivity coefficients were calculated for the biomass yield on glucose and acetate. Sensitivity coefficients were computed for kinetic parameters and enzyme initial conditions ( $N = 183\,000$ ). Error bars represent the 95% confidence intervals of the sensitivity coefficients.

uses FBA solutions in conjunction with cybernetic control variables to dynamically simulate metabolism. We studied the performance of HCM-FBA on a prototypical metabolic network, along with two *E. coli* networks. First, we showed that the performance of HCM-FBA and HCM was comparable for the prototypical network and a small model of anaerobic *E. coli* metabolism. For the anaerobic case, both approaches described the experimental data. However, HCM-FBA (which was within 5% of HCM and slightly better than HCM for lactate secretion) had fewer modes and parameters. Next, HCM-FBA was applied to an aerobic *E. coli* metabolic network that was not feasible for HCM. EM decomposition of the aerobic network generated over 153 000 EMs. Conversely, the HCM-FBA approach described cell mass growth and the shift from glucose to acetate consumption with only 29 FBA modes. Global sensitivity analysis further showed that only five of the 29 FBA modes were critical to model performance. Removal of these modes crippled the model, but removal of the remaining 24 modes had a negligible impact. These insignificant modes were associated with maintenance; thus, they would likely not impact model predictions for a growing culture. HCM-FBA is an alternative approach to HCM, especially for large networks, where the generation of EMs is infeasible. EMs show the complexity of a cell, displaying the many routes it can take, but mathematically, FBA has an objective superiority for large networks.

HCM-FBA is a promising approach to model large metabolic networks, where EMs calculations are infeasible, and where kinetic models of such systems have intractable identification problems. However, there are additional studies that should be performed. First, the intracellular flux distribution predicted by HCM-FBA should be compared with HCM and to flux measurements calculated using MFA or FBA/DFBA in combination with carbon labeling. HCM predicted intracellular fluxes that were similar to MFA results [12]; however, the fluxes predicted by HCM-FBA have not yet been validated. Next, the performance of HCM-FBA should be compared with lumped HCMs (L-HCMs). L-HCMs, which combine EMs into mode families based

upon metabolic function [10], [16], have been applied to an *E. coli* network with 67 reactions and a *Saccharomyces cerevisiae* network with 70 reactions; both cases had satisfactory fits to extracellular experimental data. However, while L-HCM reduces the dimension of possible alternative modes that must be considered, it still requires the calculation of an initial set of modes. For metabolic networks of even moderate size, EM (or EP) decomposition may not be possible. On the other hand, the generation of flux balance solutions (convex combinations of the EMs or EPs) is trivial, even for genome scale metabolic networks. Thus, HCM-FBA opens up the possibility for dynamic genome scale models of bacterial and perhaps even of mammalian metabolism.

#### IV. MATERIALS AND METHODS

The HCM-FBA approach is a modification of HCM, where EMs are replaced with FBA solutions. Thus, extracellular variables are dynamic, while intracellular metabolites are at a pseudosteady state. The abundance of extracellular species  $i$  ( $x_i$ ), the pseudoenzyme  $e_l$  (catalyzes flux through mode  $l$ ), and cell mass is governed by

$$\begin{aligned} \frac{dx_i}{dt} &= \sum_{j=1}^{\mathcal{R}} \sum_{l=1}^{\mathcal{L}} \sigma_{ij} z_{jl} q_l(\mathbf{e}, \mathbf{k}, \mathbf{x}) c \quad i = 1, \dots, \mathcal{M} \\ \frac{de_l}{dt} &= \alpha_l + r_{El}(\mathbf{k}, \mathbf{x}) u_l - (\beta_l + r_G) e_l \quad l = 1, \dots, \mathcal{L} \\ \frac{dc}{dt} &= r_G c \end{aligned}$$

where  $\mathcal{R}$  and  $\mathcal{M}$  denote the number of reactions and extracellular species in the model, and  $\mathcal{L}$  denotes the number of FBA modes. The quantity  $\sigma_{ij}$  denotes the stoichiometric coefficient for species  $i$  in reaction  $j$ , and  $z_{jl}$  denotes the normalized flux for reaction  $j$  in mode  $l$ . If  $\sigma_{ij} > 0$ , species  $i$  is produced by reaction  $j$ ; if  $\sigma_{ij} < 0$ , species  $i$  is consumed by reaction  $j$ ; if  $\sigma_{ij} = 0$ , species  $i$  is not connected with reaction  $j$ . Extracellular species balances were subject to the initial conditions  $\mathbf{x}(t_0) = \mathbf{x}_0$  determined from experimental data. The term  $q_l(\mathbf{e}, \mathbf{k}, \mathbf{x})$  denotes the specific uptake/secretion rate for mode  $l$ , where  $\mathbf{e}$  denotes the pseudoenzyme vector,  $\mathbf{k}$  denotes the unknown kinetic parameter vector,  $\mathbf{x}$  denotes the extracellular species vector, and  $c$  denotes the cell mass, and  $q_l(\mathbf{e}, \mathbf{k}, \mathbf{x})$  is the product of a kinetic term ( $\bar{q}_l$ ) and a control variable governing enzyme activity. Flux through each mode was catalyzed by a pseudoenzyme  $e_l$ , synthesized at the regulated specific rate  $r_{E,l}(\mathbf{k}, \mathbf{x})$ , and constitutively at the rate  $\alpha_l$ . The term  $u_l$  denotes the cybernetic variable controlling the synthesis of enzyme  $l$ . The term  $\beta_l$  denotes the rate constant governing nonspecific enzyme degradation, and  $r_G$  denotes the specific growth rate through all modes. The specific uptake/secretion rates and the specific rate of enzyme synthesis were modeled using saturation kinetics. The specific growth rate was given by

$$r_G = \sum_{l=1}^{\mathcal{L}} z_{\mu l} q_l(\mathbf{e}, \mathbf{k}, \mathbf{x})$$

where  $z_{\mu l}$  denotes the growth flux  $\mu$  through mode  $l$ . The control variables  $u_l$  and  $v_l$ , which control the synthesis and

activity of each enzyme, respectively, were given by

$$u_l = \frac{z_{sl}\bar{q}_l}{\sum_{l=1}^{\mathcal{L}} z_{sl}\bar{q}_l} \quad v_l = \frac{z_{sl}\bar{q}_l}{\max_{l=1,\dots,\mathcal{L}} z_{sl}\bar{q}_l}$$

where  $z_{sl}$  denotes the uptake flux of substrate  $s$  through mode  $l$ . The model equations were implemented in Julia (v.0.4.2) [17] and solved using SUNDIALS [18]. The model code is available at <http://www.varnerlab.org> under an MIT license.

### A. ELEMENTARY MODE AND FLUX BALANCE ANALYSIS

EMs were calculated using METATOOL 5.1 [19]. FBA modes were defined as the solution flux vector through the network connecting substrate uptake to cell mass and extracellular product formation. The FBA problem was formulated as

$$\begin{aligned} \max_{\mathbf{w}} (w_{obj} = \boldsymbol{\theta}^T \mathbf{w}) \\ \text{s.t. } \mathbf{S}\mathbf{w} = \mathbf{0} \\ \alpha_i \leq w_i \leq \beta_i \quad i = 1, 2, \dots, \mathcal{R} \end{aligned}$$

where  $\mathbf{S}$  denotes the stoichiometric matrix,  $\mathbf{w}$  denotes the unknown flux vector,  $\boldsymbol{\theta}$  denotes the objective selection vector, and  $\alpha_i$  and  $\beta_i$  denote the lower and upper bounds on flux  $w_i$ , respectively. The FBA problem was solved using the GNU linear programming kit (v 4.52) [20]. For each FBA mode, the objective  $w_{obj}$  was to maximize either the specific growth rate or the specific rate of byproduct formation. Multiple FBA modes were calculated for each objective by allowing the oxygen and nitrate uptake rates to vary. For aerobic metabolism, the specific oxygen and nitrate uptake rates were constrained to allow a maximum flux of 10 and 0.05 mM/gDW · h, respectively. Each FBA mode was normalized by the specified objective flux.

### B. GLOBAL SENSITIVITY ANALYSIS

Variance-based sensitivity analysis was used to estimate which FBA modes were critical to model performance. The performance function used in this letter was the biomass yield on substrate. Candidate parameter sets ( $N = 182\,000$ ) were generated using Sobol sampling by perturbing the best fit parameter set  $\pm 50\%$  [21]. Model performance, calculated for each of these parameter sets, was then used to estimate the total-order sensitivity coefficient for each model parameter.

### C. ESTIMATION OF MODEL PARAMETERS

Model parameters were estimated by minimizing the difference between simulations and experimental measurements (squared residual)

$$\min_{\mathbf{k}} \sum_{\tau=1}^T \sum_{j=1}^S \left( \frac{\hat{x}_j(\tau) - x_j(\tau, \mathbf{k})}{\omega_j(\tau)} \right)^2$$

where  $\hat{x}_j(\tau)$  denotes the measured value of species  $j$  at time  $\tau$ ,  $x_j(\tau, \mathbf{k})$  denotes the simulated value for species  $j$  at time  $\tau$ , and  $\omega_j(\tau)$  denotes the experimental measurement variance

for species  $j$  at time  $\tau$ . The outer summation is with respect to time, while the inner summation is with respect to state. The model residual was minimized using simulated annealing implemented in the Julia programming language.

### REFERENCES

- [1] J. D. Orth, I. Thiele, and B. Ø. Palsson, "What is flux balance analysis?" *J. Nature Biotechnol.*, vol. 28, no. 3, pp. 245–248, 2010.
- [2] S. Schuster, D. A. Fell, and T. Dandekar, "A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks," *Nature Biotechnol.*, vol. 18, no. 3, pp. 326–332, 2000.
- [3] C. H. Schilling, D. Letscher, and B. Ø. Palsson, "Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective," *J. Theor. Biol.*, vol. 203, no. 3, pp. 229–248, 2000.
- [4] M. W. Covert, E. M. Knight, J. L. Reed, M. J. Herrgard, and B. Ø. Palsson, "Integrating high-throughput and computational data elucidates bacterial networks," *Nature*, vol. 429, pp. 92–96, May 2004.
- [5] S. J. Wiback, R. Mahadevan, and B. Ø. Palsson, "Reconstructing metabolic flux vectors from extreme pathways: Defining the alpha-spectrum," *J. Theor. Biol.*, vol. 224, no. 3, pp. 313–324, 2003.
- [6] L.-Q. Lee, J. Varner, K. Ko, "Parallel extreme pathway computation for metabolic networks," in *Proc. Int. IEEE CS Comput. Syst. Bioinform. Conf.*, Aug. 2004, pp. 636–639.
- [7] J. Varner and D. Ramkrishna, "Metabolic engineering from a cybernetic perspective: Aspartate family of amino acids," *Metabolic Eng.*, vol. 1, no. 1, pp. 88–116, 1999.
- [8] H.-S. Song and D. Ramkrishna, "Prediction of dynamic behavior of mutant strains from limited wild-type data," *Metabolic Eng.*, vol. 14, no. 2, pp. 69–80, 2012.
- [9] J. I. Kim, H.-S. Song, S. R. Sunkara, A. Lali, and D. Ramkrishna, "Exacting predictions by cybernetic model confirmed experimentally: Steady state multiplicity in the chemostat," *Biotechnol. Prog.*, vol. 28, no. 5, pp. 1160–1166, 2012.
- [10] H.-S. Song and D. Ramkrishna, "Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function," *Biotechnol. Bioeng.*, vol. 108, no. 1, pp. 127–140, 2011.
- [11] K. G. Gadkar, F. Doyle, T. J. Crowley, and J. D. Varner, "Cybernetic model predictive control of a continuous bioreactor with cell recycle," *Biotechnol. Prog.*, vol. 19, no. 5, pp. 1487–1497, 2003.
- [12] J. I. Kim, J. D. Varner, and D. Ramkrishna, "A hybrid model of anaerobic *E. coli* GJT001: Combination of elementary flux modes and cybernetic variables," *Biotechnol. Prog.*, vol. 24, no. 5, pp. 993–1006, 2008.
- [13] A. Varma and B. Ø. Palsson, "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* w3110," *J. Appl. Environ. Microbiol.*, vol. 60, no. 10, pp. 3724–3731, 1994.
- [14] R. Schuetz, L. Kuepfer, and U. Sauer, "Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*," *Molecular Syst. Biol.*, vol. 3, no. 1, p. 119, 2007.
- [15] B. Palsson, *Systems Biology: Properties of Reconstructed Networks*. New York, NY, USA: Cambridge Univ. Press, 2006.
- [16] H.-S. Song and D. Ramkrishna, "Prediction of metabolic function from limited data: Lumped hybrid cybernetic modeling (L-HCM)," *Biotechnol. Bioeng.*, vol. 106, no. 2, pp. 271–284, 2010.
- [17] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. (2014). "Julia: A fresh approach to numerical computing." [Online]. Available: <https://arxiv.org/abs/1411.1607>
- [18] A. C. Hindmarsh et al., "SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers," *ACM Trans. Math. Softw.*, vol. 31, no. 3, pp. 363–396, Sep. 2005.
- [19] A. von Kamp and S. Schuster, "Metatool 5.0: Fast and flexible elementary modes analysis," *Bioinformatics*, vol. 22, no. 15, pp. 1930–1931, 2006.
- [20] (2016). *GNU Linear Programming Kit, Version 4.52*. [Online]. Available: <http://www.gnu.org/software/glpk/glpk.html>
- [21] J. D. Herman, accessed on Feb. 12, 2016. [Online]. Available: <https://github.com/jdherman/salib>