Effective Dynamic Models of Metabolic Networks

Michael Vilkhovoy*, Mason Minot†, Jeffrey D. Varner*

*School of Chemical Engineering, Purdue University, West Lafayette, IN 47907 USA †School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14850 USA

Abstract—**Mathematical models of biochemical networks are useful tools to understand and ultimately predict how cells utilize nutrients to produce valuable products. Hybrid cybernetic models in combination with elementary modes (HCM EM) provides a route to model cellular metabolism. However, this framework is limited to smaller networks due to the computational cost of calculating elementary modes and a high number of solutions for large networks. In this study, we developed a hybrid cybernetic model in combination with flux balance analysis (HCM FBA) modes instead of elementary modes. Flux balance analysis provides a sufficient number of modes (metabolic options) for the network to estimate experimental observations. We show HCM FBA has comparable model performance to HCM EM for a hypothetical proof of concept metabolic network and for a reduced anaerobic *E. coli* network. HCM FBA was applied to a larger metabolic network where 29 FBA modes were computed, in contrast to over 66,000 elementary modes. A sensitivity analysis was performed to reduce the number of FBA modes from 29 to 5. This reduction maintained a robust model performance to fit experimental observations.**

Index Terms—**Dynamic metabolic models, flux balance analysis, cybernetic models**

I. Introduction

Metabolic networks for simple organisms such as *E. coli* contain thousands of reactions that are highly interconnected. These organisms have the potential to be engineered to produce valuable products. Mathematical models of biochemical networks are useful tools to understand and ultimately predict how cells utilize nutrients to produce valuable products. Traditionally, Monod kinetics have been used as unstructured models to describe cell growth, substrate consumption, and product formation, but fail to account for intracellular metabolism [1]. Structured modeling approaches have made great progress to describe metabolic pathways including elementary mode analysis (EM) [2], metabolic flux analysis (MFA), flux balance analysis (FBA) [3], dynamic FBA (DFBA) [4], [5], and hybrid cybernetic modeling (HCM) [6].

Convex and constraints based analysis methods, such as EM, FBA and MFA, describe intracellular metabolism using the stoichiometry of the biochemical networks and the pseudo-steady state assumption [3]. EM analysis is a powerful tool to decompse a metabolic network into independent minimal pathways that are stoihiometrically and thermodynamically feasible. However, elementary mode calculations are computational expensive and infeasible for genome scale networks [7]. FBA is frequently used to study genome-scale networks [3] and often results in good approximations for yields, however it fails to estimate intracellular metabolism due to its

underdetermined solution space. Compared with FBA, MFA has good estimations of intracellular metabolism, but is dependent on experimental data. Just like FBA, HCM approaches cell metabolism as a resource allocation problem towards a specified objective. But HCM has multiple pathways to select from (elementary modes) and can use a combination of modes to direct its resources through the cell. HCM has been shown to have comparable internal flux estimations to MFA and is able to estimate dynamic external fluxes with respect to experimental observations [6]. DFBA is able to account for dynamic external fluxes but fails to incorporate the regulatory processes involved with HCM without the use of boolean rules [8]. Despite the advantages of HCM, it is restricted to small networks due to the exponential increase of the number of elementary modes for large networks. Recent work has been done to reduce the number of elementary modes by lumping modes into groups using complex weighting schemes and experimental data [9], however, HCM is still limited by EM decomposition.

In this study, we developed a hybrid cybernetic model in combination with flux balance analysis modes instead of elementary modes. We show HCM FBA has comparable model performance to HCM EM for a hypothetical metabolic network and a reduced anaerobic *E. coli* network. HCM EM and HCM FBA were then applied to a large metabolic network. The number of elementary modes made it impractical to model with HCM EM, but HCM FBA was computationally feasible and performed well with respect to experimental observations. Taken together, HCM FBA has comparable model performace to HCM EM and may be potentially applied to genome-scale networks which was a shortcoming of HCM EM.

II. Results

The performance of HCM FBA was equivalent to HCM EM for a proof of concept metabolic network (Fig. 1). In the proof of concept metabolic network consisting of 6 metabolites and 7 reactions (Fig. 1A), we calculated 6 elementary modes and 3 FBA modes. Using the HCM EM modeling framework, concentration profiles were generated for the extracellular metabolites and artifical biomass (Fig. 1B). The kinetic parameters of HCM FBA were varied to fit the concentration profiles of the HCM EM model. For this proof of concept model, HCM FBA was able to use fewer modes and kinetic parameters to mimic the simulation profiles of HCM EM.

The performance of HCM FBA was equivalent to HCM EM for a small model of anaerobic *E. coli* metabolism (Fig. 2, left). In the case, we demonstrated that HCM FBA has comparable model performance to a published HCM EM
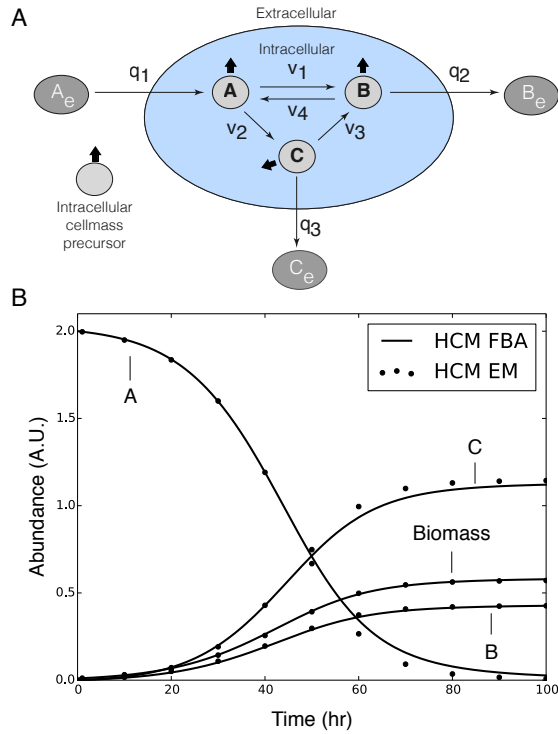
Fig. 1. A: Proof of concept metabolic network with six metabolites and seven reactions. Intracellular cellmass precursors A, B and C are balanced (no accumulation) while the extracellular metabolites (A_e, B_e and C_e) are not balanced (can accumulate). The blue-oval denotes the cell boundary, where q_j denotes the jth flux across the boundaries ([mmol/gdw-hr]) and v_k denotes the kth intracellular flux. B: Simulation of the extracellular metabolites using the FBA hybrid cybernetic approach (solid line) versus elementary modes (points) for the proof of concept metabolic network. FBA modes have comparable model performance compared to elementary modes.
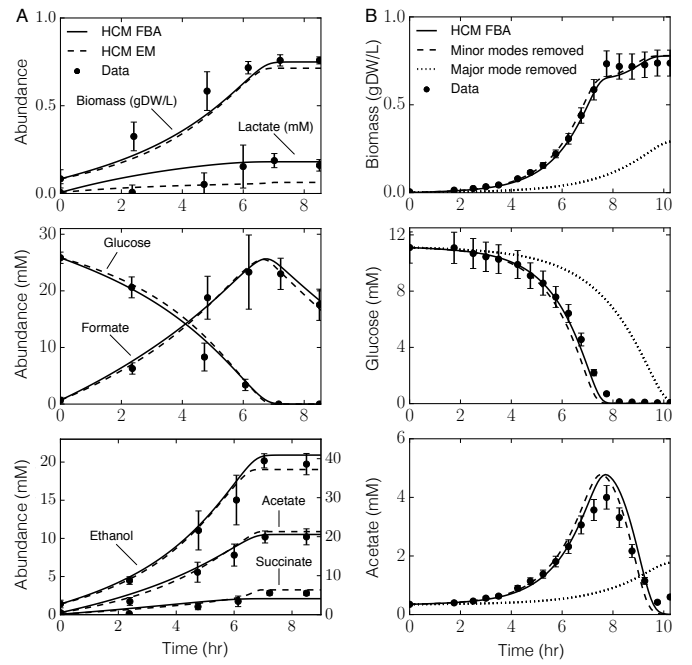

Fig. 2. A: Batch anaerobic *E. coli* fermentation data versus simulation of the hybrid cybernetic model with FBA modes (solid line) and elementary modes (dashed line). (Experimental data from Kim et. al (2008) [6] with error bars representing 90% confidence intervals.) B: Batch aerobic *E. coli* fermentation data of glucose, acetate and cell density measurements versus simulation of the hybrid cybernetic model with FBA modes (solid line). Model performance is shown when minor modes (dashed) and a major mode (dotted) is removed from HCM FBA. (Experimental data from Varma & Palsson (1994) [4] with assumed 10% error bars).

model [6]. The metabolic network consisted of 12 reactions, where we calculated 9 elementary modes and 7 FBA modes. The simulated HCM EM concentration profiles of biomass, substrate and fermentation products (Fig. 2A (dashed line)) were generated using the kinetic parameters and elementary modes reported [6]. The kinetic parameters of HCM FBA were varied to fit the same experimental observations. HCM FBA has 17 kinetic parameters while HCM EM has 21 kinetic parameters and both models resulted in similar fits to the experimental data (Fig. 2A). HCM FBA had a slightly better fit on lactate than HCM EM.

HCM FBA outperformed HCM EM for a large model of aerobic *E. coli* metabolism (Fig. 2, right). In the case, we demonstrated that HCM FBA is applicable for larger metabolic networks where elementary mode calculations increase exponentially. A metabolic network [10] of 98 reactions and 60 metabolites with the biomass reaction from Palsson's *E. coli* core model [11] was used. We calculated over 66,000 elementary modes for this network and determined it was futile to model. For the same network, we computed 29 FBA modes. HCM FBA was simulated to fit cell density, glucose and acetate measurements for an aerobic *E. coli* batch culture [4] (Fig.2B (solid line)).

Variance based sensitivity of aerobic *E. coli* metabolism identified modes that were essential to biomass yield (Fig.

3). The analysis calculated the total order sensitivity indicies of all the kinetic parameters and enzyme initial conditions. We used the sensitivity results as a simple systematic method to eliminate FBA modes to give the smallest number needed while maintaining good model performance. Five FBA modes were determined to be significant by an arbitrary threshold of 0.00218. The remaining 24 FBA modes were identified to be minor modes and insignificant. The removal of the major mode, representing aerobic growth, (Fig, 2B (dotted line)) was detrimental to model performance, whereas the removal of the minor modes (dashed line) had little effect.

III. DISCUSSION

In this study, we developed a hybrid cybernetic model in combination with flux balance analysis modes instead of elementary modes. First, we showed HCM FBA can have comparable model performance to HCM EM for a proof of concept metabolic network. This judgement is based on the quality of fit of the model simulation to the observations. Next, HCM FBA model performance was compared to an established HCM EM model [6] for an anaerobic culture of *E. coli*. HCM FBA had comparable fits to HCM EM on the fermentation data and a slightly better fit on lactate measurements. Finally, HCM was applied to a larger metabolic network where 29 FBA modes were computed, in contrast to over 66,000 elementary modes. HCM FBA was shown to have robust model performance only requiring 5 of the 29 FBA
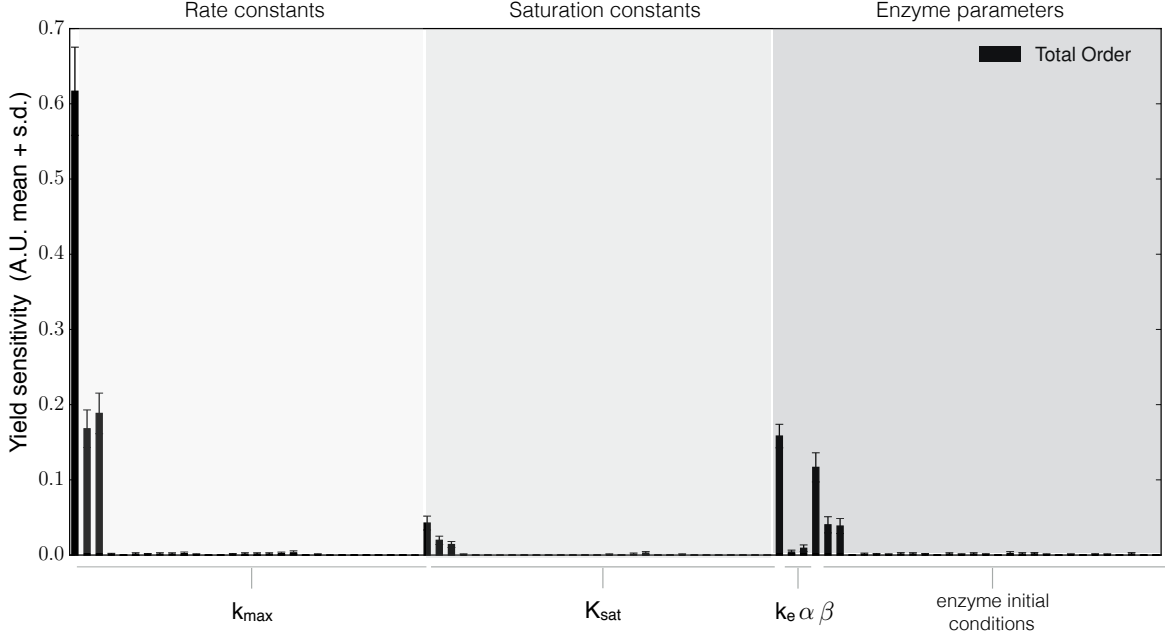
Fig. 3. Total order variance based sensitivity analysis on biomass yield from glucose and acetate. Sensitivity indicies computed for rate constants, saturation constants and enzyme parameters. The most critical parameters are associated with mode 1 representing aerobic growth. (Error bars represent 95% confidence intervals).

modes (selected by the sensitivity analysis) to fit experimental observations.

The hybrid cybernetic model is formulated to maximize substrate uptake based on cybernetic arguments. Towards this objective, HCM uses a combination of modes to describe both the external and internal fluxes. The internal fluxes still follow the pseudo steady state approximation applied in FBA, but HCM uses a combination of pathways. HCM EM has been shown to have comparable internal flux estimates to MFA results [6], this has not been shown for HCM FBA. The HCM framework has been shown to predict dynamic external fluxes using both elementary and FBA modes. While DFBA already has the capacity to estimate dynamic external fluxes, it requires *a priori* knowledge and boolean rules to capture the diauxic phenomena [4], [5], [8]. HCM overcomes this by incorparting the regulatory processes of substrate utilization based on its cybernetic arguments. HCM EM has good model performance for reduced networks, but is impractical for larger networks. HCM EM has been applied to a network with 67 reactions, requiring the elementary modes to be lumped into groups by complex weighting schemes and relying on experimental data [9]. A shortcoming of the HCM EM framework is that EM calculations are computationally expensive and the number of modes exponentially increases for larger networks [7]. In contrast, FBA does not have the computational burden associated with calculating elementary modes. FBA is frequently used to study genome-scale networks [3] and can be used to generate modes for the HCM framework. This opens up the possibility of genome scale cybernetic models.

HCM FBA provides a practical and feasible approach to model large networks where elementary modes fail. The disadvantage of HCM FBA is it requires *a priori* knowledge of the inputs into the metabolic network. FBA modes do not have

the thorough information as elementary modes, but provide sufficient pathways to model experimental observations.

IV. MATERIALS AND METHODS

HCM FBA method.: The model was constructed using the hybrid cybernetic approach of Kim et al. [6]. The abundance of extracellular species i (x_i) is governed by:

$$\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sum_{l=1}^{\mathcal{L}} \sigma_{ij} z_{jl} r_l\left(\mathbf{e}, \mathbf{k}, \mathbf{s}\right) c \qquad i = 1, \dots, \mathcal{M} \quad (1)$$

where $\mathcal{R}$ and $\mathcal{M}$ denotes the number of reactions and extracellular species in the model, and $\mathcal{L}$ denotes the number of modes. The quantity σ_{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, while $\sigma_{ij} = 0$ indicates species i is not connected with reaction j. Extracellular species balances were subject to the initial conditions $\mathbf{x}\left(t_o\right) = \mathbf{x}_o$ determined from experimental data. The $r_l\left(\mathbf{e}, \mathbf{k}, \mathbf{s}\right)$ term denotes the regulated uptake rate of mode k, and was written as the product of a kinetic term ($\bar{r}_l$) and a cybernetic control variable for enzyme activity (v_l), $r_l\left(\mathbf{e}, \mathbf{k}, \mathbf{s}\right) = \bar{r}_l v_l$, where $\mathbf{e}$ denotes the enzyme level for each mode which become additional state variables, $\mathbf{k}$ ($\mathcal{K} \times 1$) denotes the unknown kinetic parameter vector, and $\mathbf{s}$ denotes the substrate utilized. The reaction rate followed Michaelis-Menten kinetics. The abundance of the enzyme species is governed by:

$$\frac{de_l}{dt} = \alpha_l + r_{E,l}\left(\mathbf{k}, \mathbf{s}\right) u_l - \left(\beta_l + r_G\right) e_l \quad (2)$$

where α_l denotes the specific rate of constitutive enzyme synthesis for enzyme l, $r_{E,l}\left(\mathbf{k}, \mathbf{s}\right)$ denotes the specific rate

enzyme synthesis for enzyme l, and u_l denotes the cybernetic variable controlling the synthesis of enzyme l. The term β_l denotes the rate constant governing enzyme degradation, and r_G denotes the growth rate through all modes. The specific rate of enzyme synthesis was modeled using saturation kinetics. All enzyme initial conditions were set to 0.9 for the anaerobic case, and 0.8 for the aerobic case.

$$r_G = \sum_{l=1}^{\mathcal{L}} z_{\mu l} r_l \left(\mathbf{e}, \mathbf{k}, \mathbf{s} \right) \tag{3}$$

where $z_{\mu l}$ denotes the growth flux μ through mode l. The cybernetic control variables, u_l and v_l, control the synthesis and activity of each enzyme for each mode using the matching and proportional laws [12]. They are formulated to maximize the overall substrate uptake rate through each mode.

$$u_l = \frac{z_{s,l} \bar{r}_l}{\sum_{l=1}^{\mathcal{L}} z_{sl} \bar{r}_l} \tag{4}$$

$$v_l = \frac{z_{s,l} \bar{r}_l}{\max_{l=1} z_{sl} \bar{r}_l} \tag{5}$$

where z_{sl} denotes the uptake flux of substrate s through mode l. In the anaerobic case we followed the assumption of Young's thesis [13] and Kim et al. [6] that formate decomposition occurs only outside the network (applicable for strain GJT001). Therefore its enzyme activity was assumed to be at its maximum level of unity. The reaction rate for formate was also modified following Kim et al. [6]

Elementary Flux Mode Analysis: The elementary flux modes were calculated using METATOOL 5.1 [2] for all networks. The input file requires the conditions for reversibility/irreversibility reactions and intracellular/extracellular metabolites to be specified.

Flux Balance Analysis: Flux balance analysis determines an optimal flux vector $\mathbf{w}$ subject to the stoichiometric matrix and flux constraints that maximizes/minimize a specified objective. FBA modes were defined as the solution flux vector through the network. FBA is formulated by the follwoing:

$$\max_{\boldsymbol{w}} \left(w_{obj} = \boldsymbol{c}^T \boldsymbol{w} \right)$$
$$\text{Subject to : } \mathbf{S}\mathbf{w} = 0 \text{ and } \mathbf{S_x}\mathbf{w} = \mathbf{b} \tag{6}$$
$$\alpha_i \leq w_i \leq \beta_i$$

where $\boldsymbol{c}$ is a vector of weights indicating how each flux contributes to the objective; α_i and β_i are the lower and upper limits for the individual flux values, respectively. The stoichoimetric matrix $\mathbf{S}$ denotes the intracellular network where intracellular metabolism has a psuedo steady state approximation. $\mathbf{S_x}$ denotes the extracellular metabolic network and $\mathbf{b}$ denotes the exchange flux across the cell membrane. FBA was solved using linear programming by a GNU Linear Programming Kit, GLPK version 4.52. For each mode, the objective was set to maximize either the growth flux or each fermentation byproduct flux from a specified substrate. Multiple modes were calculated for each objective flux by allowing the oxygen and nitrate uptake rates to be either nonexistent or its maximum. In the aerobic case, oxygen and nitrate uptake rates were constrained to allow a maximum

flux of 10 mM/gDW·hr and 0.05 mM/gDW·hr, respectively. Each flux vector was normalized by the growth flux, or by the specified objective flux.

Sensitivity Analysis: The variance based sensitivity analysis [14] was done using SALib, an open source Python library. The sensitivity analysis identified key parameters that were influential on our specified model output of biomass yield. All model parameters used in the Sobol sensitivity analysis varied between 50% lower and higher than the optimal values obtained from parameter identification, except enzyme initital conditions which varied between 0 and 1. The sample space was set to 1000 and 182,000 parameter sets were generated from the pseudo-random Sobol sequence.

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

$$\min_{\mathbf{k}} \sum_{\tau=1}^{\mathcal{T}} \sum_{j=1}^{\mathcal{S}} \left(\frac{\hat{x}_j \left(\tau \right) - x_j \left(\tau, \mathbf{k} \right)}{\omega_j \left(\tau \right)} \right)^2 \tag{7}$$

where $\hat{x}_j \left(\tau \right)$ denotes the measured value of species j at time τ, $x_j \left(\tau, \mathbf{k} \right)$ denotes the simulated value for species j at time τ, and $\omega_j \left(\tau \right)$ denotes the experimental measurement variance for species j at time τ. The outer summation is with respect to time, while the inner summation is with respect to state. We minimized the model residual using simulated annealing.

REFERENCES

1. Shuler, M. & Kargi, F. *Bioprocess Engineering: Basic Concepts* (Prentice Hall, 2007).
2. Kamp, A. & Schuster, S. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics* **22**, 1930–1931 (2006).
3. Orth, J., Thiele, I. & Palsson, B. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
4. Varma, A. & Palsson, B. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type escherichia coli w3110. *Appl. Environ. Microbiol.* **60**, 3724–3731 (1994).
5. Mahadevan, R., Edwards, J. & Doyle III, F. Dynamic flux balance analysis of diauxic growth in escherichia coli. *Biophys. J.* **83**, 1331 – 1340 (2002).
6. Kim, J., Varner, J. & Ramkrishna, D. A hybrid model of anaerobic e. coli gjt001: Combination of elementary flux modes and cybernetic variables. *Biotechnol. Prog.* **24**, 993–1006 (2008).
7. Lee, L., Varner, J. & Ko, K. Parallel extreme pathway computation for metabolic networks. *Comput Syst Bioinformatics Conf, Int IEEE CS* **0**, 636–639 (2004).
8. Covert, M., Schilling, C. & Palsson, B. Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* **213**, 73 – 88 (2001).
9. Song, H. & Ramkrishna, D. Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function. *Biotechnol. Bioeng.* **108**, 127–140 (2011).
10. Schuetz, R., Kuepfer, L. & Sauer, U. Systematic evaluation of objective functions for predicting intracellular fluxes in escherichia coli. *Mol. Syst. Biol.* **3** (2007).
11. Palsson, B. *Systems Biology: Properties of Reconstructed Networks* (Cambridge University Press, New York, NY, USA, 2006).
12. Young, J. & Ramkrishna, D. On the matching and proportional laws of cybernetic models. *Biotechnol. Progr.* **23**, 83–99 (2007).
13. Young, J. *A system-level mathematical description of metabolic regulation combining aspects of elementary mode analysis with cybernetic control laws*. Ph.D. thesis, Purdue University (2005).
14. Saltelli, A. *et al.* Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Comput. Phys. Commun.* **181**, 259 – 270 (2010).