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 stoichiometrically 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 the HCM FBA approach was equivalent to HCM EM for a proof of concept metabolic network (Fig. 1). The proof of concept network, consisting of six metabolites and seven reactions (Fig. 1A), generated three FBA and six elementary modes. Using the elementary modes 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. 1B). The HCM EM and HCM FBA kinetic parameters were not quantitatively identical, but had similar orders of magnitude; the FBA approach had three fewer modes, thus you would not expect identical parameter values. Taken together, the HCM FBA approach replicated synthetic data generated using HCM EM, despite having three fewer modes. Next, we tested the ability of HCM FBA to replicate experimental data.
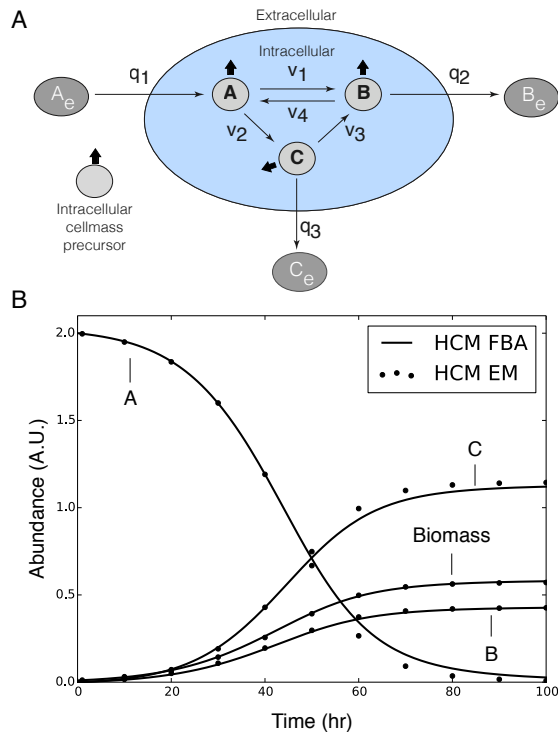
Fig. 1. A: Proof of concept metabolic network with six metabolites and seven reactions. Intracellular cellmass precursors A, B and C are balanced (no accumulation) while the extracellular metabolites (A_e, B_e and C_e) are not balanced (can accumulate). The blue-oval denotes the cell boundary, where 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.
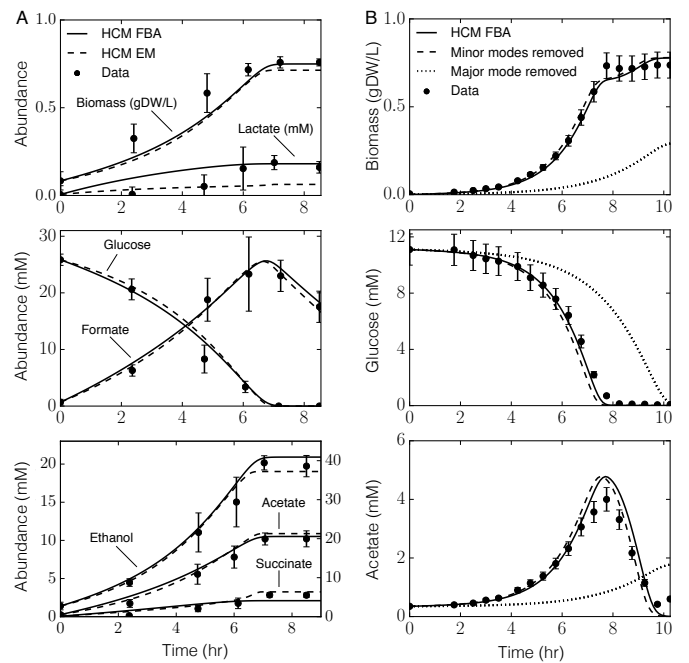

Fig. 2. A: Batch anaerobic *E. coli* fermentation data versus simulation of the hybrid cybernetic model with FBA modes (solid line) and elementary modes (dashed line). (Experimental data from Kim et. al (2008) [6] with error bars representing 90% confidence intervals.) B: Batch aerobic *E. coli* fermentation 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).

The performance of HCM FBA was equivalent to HCM EM for anaerobic *E. coli* metabolism (Fig. 2, left). We constructed an anaerobic *E. coli* network, consisting of 12 reactions and 19 metabolites, which generated nine elementary and seven FBA modes. HCM EM cellmass, glucose and fermentation product trajectories, generated using kinetic parameters reported by Kim et al. [6], were compared with experimental measurements (Fig. 2A, dashed). We estimated HCM FBA model parameters from the Kim et al. experimental data set using simulated annealing. Overall, HCM FBA performed within 5% of HCM EM (on a residual standard error basis) for the anaerobic *E. coli* data (Fig. 2A, solid), despite having two fewer modes and four fewer parameters (17 versus 21 parameters). Taken together, these results suggested HCM FBA was capable of describing experimental data, and is a viable reduced-order alternative to HCM EM.

HCM FBA captured the glucose to acetate metabolic shift for a large model of aerobic *E. coli* metabolism (Fig. 2, right). A core *E. coli* metabolic network, consisting of 60 metabolites and 105 reactions, was constructed from literature [10], [11]. Elementary mode decomposition (and thus HCM EM) was not feasible for this network; 153,000 elementary modes were generated before the calculation prematurely terminated. Conversely, flux balance analysis generated only 29 modes for the same network. HCM FBA model parameters were

estimated from cellmass, glucose and acetate measurements [4] using simulated annealing (Fig. 2B, solid). HCM FBA captured glucose consumption, cellmass formation and the dynamic switch between acetate production and consumption following glucose exhaustion. Taken together, the aerobic *E. coli* results demonstrated that HCM FBA described metabolic dynamics for large networks that were infeasible for HCM EM. However, it was unclear whether all 29 FBA modes were required to capture the experimental data. Next, we considered a systematic approach to identify the critical subset of FBA modes required for model performance.

Global variance based sensitivity analysis of aerobic *E. coli* metabolism identified the FBA modes essential to biomass yield (Fig. 3). Total order sensitivity coefficients were calculated for all kinetic parameters and enzyme initial conditions. Five of the 29 FBA modes were determined to be significant by a threshold test. Removal of the most significant mode (encoding aerobic growth on glucose) destroyed model performance (Fig. 2B, dotted). Conversely, removal of the remaining 24 modes had negligible effect on model performance (Fig. 2B, dashed). Taken together, these results demonstrate that sensitivity analysis can identify critical FBA modes, thereby allowing further reduction to the minimal model structure required to explain the data.

III. DISCUSSION

In this study, we developed a hybrid cybernetic model in combination with flux balance analysis modes instead of
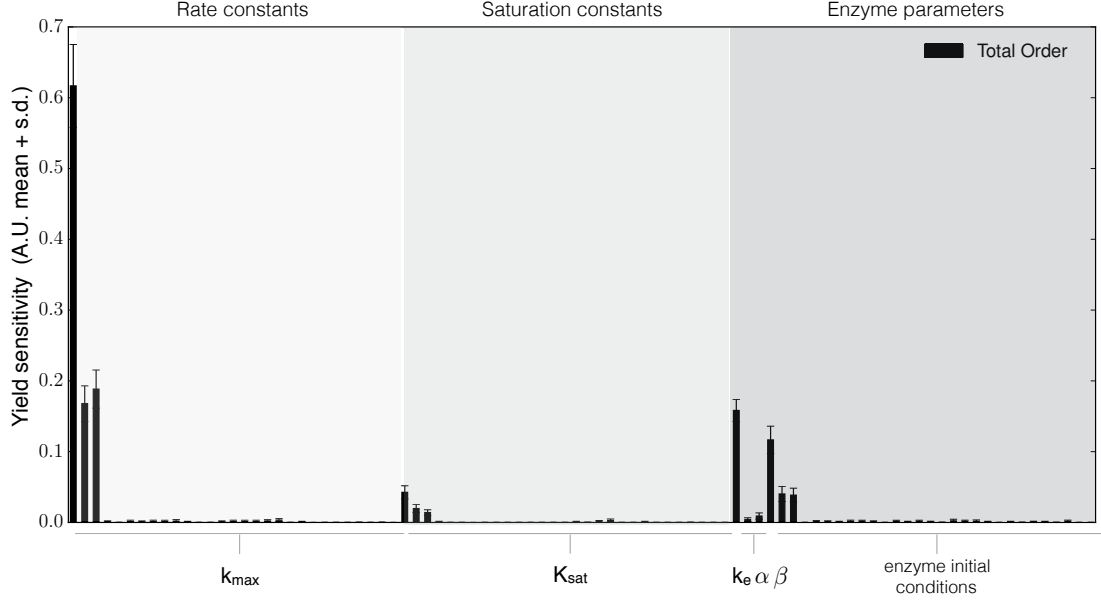
Fig. 3. Total order variance based sensitivity analysis on biomass yield from glucose and acetate. Sensitivity indicies computed for rate constants, saturation constants and enzyme parameters. The most critical parameters are associated with mode 1 representing aerobic growth. (Error bars represent 95% confidence intervals).

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 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

The HCM FBA approach is a modification of the HCM EM strategy of Kim et al. [6]. However, unlike HCM EM, we replaced elementary modes with flux balance analysis solutions. The abundance of extracellular species i (x_i), the pseudo enzyme e_l and cellmass are governed by:

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

$$\frac{de_l}{dt} = \alpha_l + r_{E,l} (\mathbf{k}, \mathbf{x}) u_l - (\beta_l + r_G) e_l \qquad l = 1, \dots, \mathcal{L}$$

$$\frac{dc}{dt} = r_G c$$

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 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 $\mathbf{x}(t_o) = \mathbf{x}_o$ determined from experimental data. The term $r_l(\mathbf{e}, \mathbf{k}, \mathbf{x})$ denotes the specific rate of flux through mode l, and was written as the product of a kinetic term ($\bar{r}_l$) and a cybernetic control variable governing enzyme activity. Flux through each mode was catalyzed by a pseudo enzyme e_l, where enzyme e_l was synthesized at the regulated specific rate $r_{E,l}(\mathbf{k}, \mathbf{x})$ and constitutively at the rate α_l. The term $r_{E,l}$ 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 flux through an FBA mode, and the specific rate of enzyme synthesis were modeled using saturation kinetics. All enzyme initial conditions were set to 0.9 for the anaerobic case, and 0.8 for the aerobic case. Lastly, cellmass was produced at the specific growth rate:

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

where $z_{\mu l}$ denotes the growth flux μ through mode l. The cybernetic control variables u_l and v_l, which control the synthesis and activity of each enzyme respectively, were given by:

$$u_l = \frac{z_{sl}\bar{r}_l}{\displaystyle\sum_{l=1}^{\mathcal{L}} z_{sl}\bar{r}_l} \qquad v_l = \frac{z_{sl}\bar{r}_l}{\displaystyle\max_{\mathbf{l}} z_{sl}\bar{r}_l}$$

where z_{sl} denotes the uptake flux of substrate x through mode l. In the anaerobic case we followed the assumption of 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]. All numerical simulations were conducted in Julia version 0.4.2, a high-level, high performance dynamic programming language [12]. Species balances were solved using a Sundials [13] wrapper for julia.

Elementary Mode and Flux Balance Analysis: Elementary modes were calculated using METATOOL 5.1 [2]. FBA modes were defined as the solution flux vector through the network connecting substrate uptake to cellmass and extracellular product formation. The FBA problem was formulated as:

$$\max_{\mathbf{w}} \left(w_{obj} = \theta^T \mathbf{w}\right)$$
$$\text{Subject to}: \ \mathbf{Sw} = \mathbf{0} \tag{2}$$
$$\alpha_i \leq w_i \leq \beta_i$$

where $\mathbf{S}$ 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, GLPK version 4.52 [14]. For each FBA mode, the objective was to maximize either the growth rate, or the specific rate of byproduct formation from a specified starting substrate. Multiple FBA modes were calculated for each objective flux by allowing the oxygen and

nitrate uptake rates to be either zero or maximal. For aerobic metabolism, the specific 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.

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 study was the biomass yield on substrate. Candidate parameter sets (N = 182,000) were generated using Sobol sampling by perturbing the best fit parameter set $\pm 50\%$ [15]. Model performance, calculated for each of these parameter sets, was then used to estimate the total-order sensitivity coefficient for each model parameter.

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

$$\min_{\mathbf{k}} \sum_{\tau=1}^{\mathcal{T}} \sum_{j=1}^{\mathcal{S}} \left(\frac{\hat{x}_j(\tau) - x_j(\tau, \mathbf{k})}{\omega_j(\tau)}\right)^2 \tag{3}$$

where $\hat{x}_j(\tau)$ denotes the measured value of species j at time τ, $x_j(\tau, \mathbf{k})$ denotes the simulated value for species j at time τ, and $\omega_j(\tau)$ 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. The model residual was minimized 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. Bezanson, J., Karpinski, S., Shah, V. B. & Edelman, A. Julia: A fast dynamic language for technical computing. *CoRR* **abs/1209.5145** (2012).
13. Hindmarsh, A. C. *et al.* Sundials: Suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Softw.* **31**, 363–396 (2005).

14. Makhorin, A. Glpk (gnu linear programming kit). available online: https://www.gnu.org/software/glpk/ .
15. Herman, J. Salib. available online: https://github.com/jdherman/salib .