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 the hybrid cybernetic modeling with flux balance analysis or HCM-FBA technique, which uses flux balance solutions instead of elementary modes, to dynamically model metabolism. 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

Biotechnology harnesses the power of metabolism to produce products that benefit society. Constraints based models are useful tools to understand and ultimately predict how cells utilize nutrients to produce products. Constraints based methods such as flux balance analysis (FBA) [1] and network decomposition approaches such elementary modes (EMs) [2] or extreme pathways (EPs) [3] model intracellular metabolism using the biochemical stoichiometry, and other constraints such as thermodynamically feasibility under pseudo-steady state conditions. Flux balance analysis has predicted the global performance of genome scale networks [4]. On the other hand, elementary modes (or extreme pathways) catalog all possible metabolic behaviors such that any flux distribution predicted by FBA is a convex combination of the elementary modes (or extreme pathways) [5]. However, the calculation of elementary modes is computational expensive and currently infeasible for genome scale networks [6].

Cybernetic models are an alternative to the constraints based approach which hypothesize that metabolic control is the output of an optimal decision. Cybernetic models have predicted mutant behavior [7], [8], steady-state multiplicity [9], strain specific metabolism [10] and have been used in bioprocess control applications [11]. Hybrid cybernetic models (HCM) have addressed earlier shortcomings of the approach

by integrating cybernetic optimality concepts with elementary modes. HCMs dynamically choose combinations of elementary modes to achieve hypothetical physiological objectives. HCMs generate intracellular flux distributions consistent with other approaches such as metabolic flux analysis, and also describe extracellular measurements [12]. However, HCMs are restricted to networks which can be decomposed into elementary modes.

In this study, we developed the hybrid cybernetic modeling with flux balance analysis or HCM-FBA technique. HCM-FBA is a modification of the hybrid cybernetic approach of Ramkrishna and coworkers which uses flux balance analysis solutions (instead of elementary modes) in conjunction with control variables to dynamically simulate metabolism. We compared the performance of HCM-FBA to HCM-EM for a proof of concept metabolic network, along with two *E. coli* networks. HCM-FBA performed similar to HCM-EM for the hypothetical network and a reduced anaerobic *E. coli* network, despite having fewer parameters in each case. Next, HCM-FBA was applied to an aerobic *E. coli* metabolic network that was infeasible for HCM-EM. HCM-FBA described cellmass growth, and the shift from glucose to acetate consumption, with only a few modes. 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-EM especially for large networks where the generation of elementary modes is infeasible.

II. RESULTS

HCM-FBA 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.

The performance of HCM-FBA was equivalent to HCM-EM for anaerobic *E. coli* metabolism (Fig. 2, left). We constructed
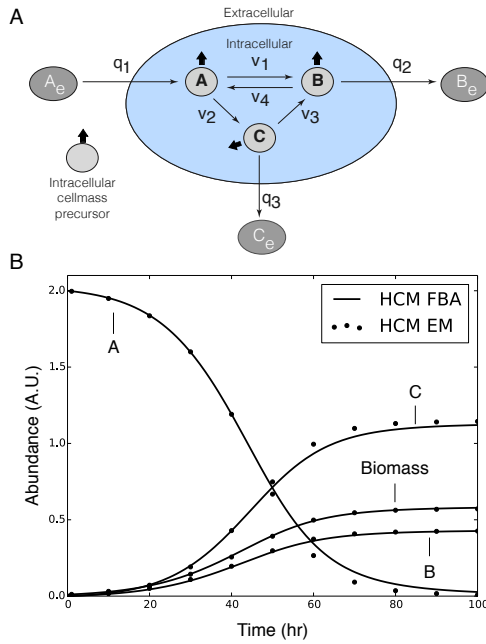
Fig. 1. HCM-FBA proof of concept metabolic study. A: Prototypical network with six metabolites and seven reactions. Intracellular cellmass precursors A, B and C are balanced (no accumulation) while the extracellular metabolites (A_e, B_e and C_e) are not balanced (can accumulate). The oval denotes the cell boundary, q_j is the jth flux across the boundary and v_k denotes the kth intracellular flux. B: Simulation of extracellular metabolite trajectories using HCM-FBA (solid line) versus HCM-EM (points) for the proof of concept metabolic network.
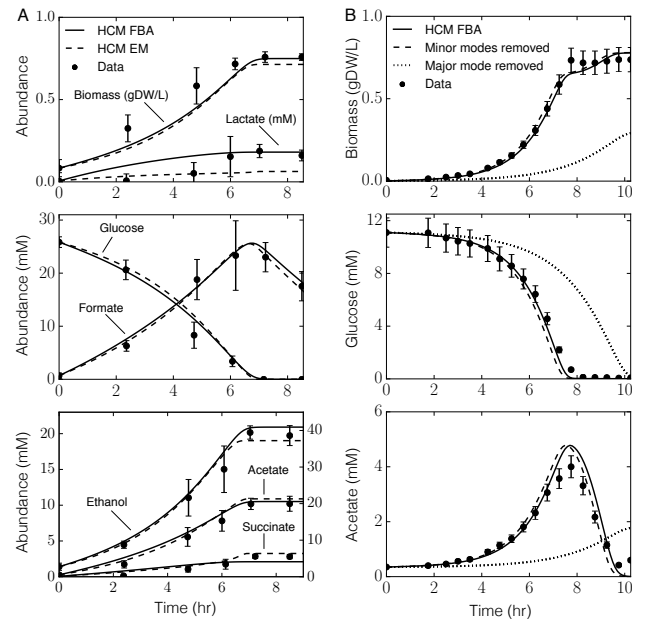

Fig. 2. HCM-FBA versus HCM-EM performance for small and large metabolic networks. A: Batch anaerobic *E. coli* fermentation data versus HCM-FBA (solid) and HCM-EM (dashed). The experimental data was reproduced from from Kim et. al [12]. Error bars represent the 90% confidence interval. B: Batch aerobic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

an anaerobic *E. coli* network, consisting of 12 reactions and 19 metabolites, which generated nine elementary and seven FBA modes. HCM-EM reproduced cellmass, glucose and byproduct trajectories using the kinetic parameters reported by Kim et al. [12] (Fig. 2A, points versus dashed). On the other hand, HCM-FBA model parameters were estimated in this study from the Kim et al. 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). Thus, while both HCM-EM 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-EM (Fig. 2, right). A core *E. coli* metabolic network, consisting of 60 metabolites and 105 reactions, was constructed from literature [14], [15]. Elementary mode decomposition of this network (and thus HCM-EM) was not feasible; 153,000 elementary modes were generated before the calculation became infeasible. Conversely, flux balance analysis generated only 29 modes for the same network. HCM-FBA model parameters were estimated from cellmass, glucose and acetate measurements [13] using simulated annealing (Fig. 2B, solid). HCM-FBA captured glucose consumption, cellmass formation and the switch to acetate consumption following glucose exhaustion. Thus, HCM-FBA described the dynamics of a network that was infeasible for HCM-EM, 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. 2B, dotted). Conversely, removing the remaining 24 modes had a negligible effect upon model performance (Fig. 2B, dashed). Thus, sensitivity analysis identified the minimal model structure required to explain the experimental data.

III. DISCUSSION

In this study, we developed HCM-FBA, an effective modeling technique to simulate metabolic dynamics. HCM-FBA is a modification of the hybrid cybernetic approach of Ramkrishna and coworkers. HCM-FBA uses flux balance analysis solutions (instead of elementary modes) in conjunction with control variables to dynamically simulate metabolism. We studied the performance of HCM-FBA on a proof of concept metabolic network, along with two *E. coli* networks. First, we showed the performance of HCM-FBA and HCM-EM were comparable for a proof of concept metabolic 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-EM and slightly better
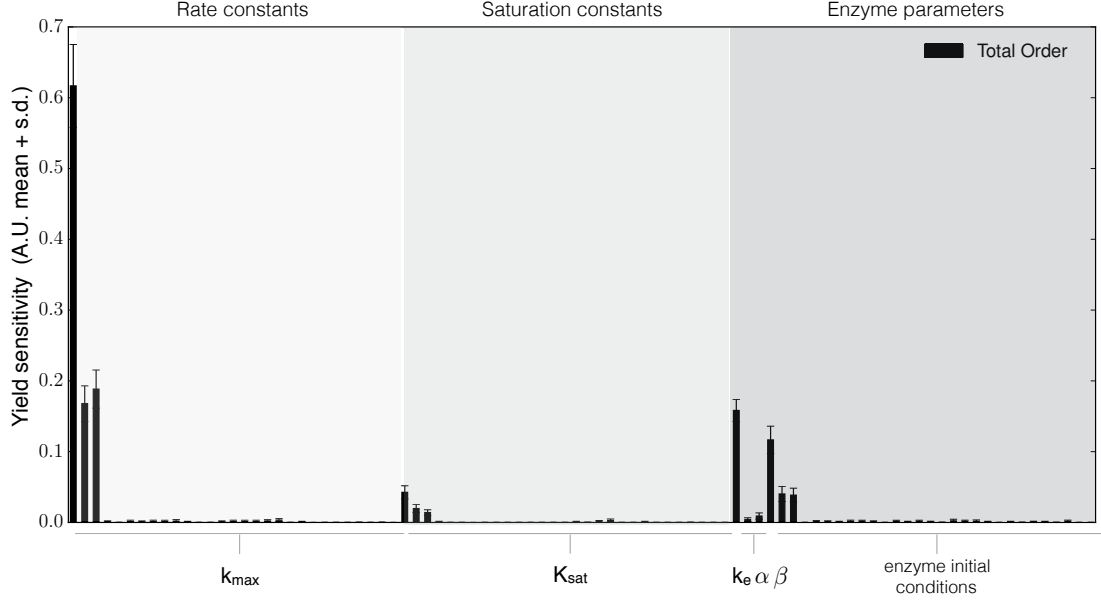
Fig. 3. Global sensitivity analysis of the aerobic *E. coli* model. Total order variance based sensitivity coefficients were calculated for the biomass yield on glucose and acetate. Sensitivity coefficients were computed for kinetic parameters and enzyme initial conditions (N = 183,000). Error bars represent the 95% confidence intervals of the sensitivity coefficients.

than HCM-EM 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. Elementary mode decomposition of the core aerobic network generated over 150,000 elementary modes. Conversely, the HCM-FBA approach described cellmass 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. Thus, HCM-FBA is a viable alternative to HCM-EM, especially for large networks where the generation of elementary modes is infeasible.

Hybrid cybernetic models are formulated to split the uptake flux among different pathways (defined as modes) to satisfy some chosen metabolic objective (i.e. maximum growth or substrate uptake). The HCM framework has been shown to predict dynamic external fluxes of cell growth, substrate utilization and byproduct formation. HCM-EM has good model performance for reduced networks, but is often impractical for larger networks. However, Song and Ramkrishna address this shortcoming by lumping elementary modes into "families" each containing several modes with unifying characteristics based on metabolic function [10], [16]. The elementary modes are lumped into groups by weighting schemes typically based on yields and may be further tuned with experimental data. The framework has been applied to a *E. coli* network with 67 reactions and a *Saccharomyces cerevisiae* network with 53 reactions, both had satisfactory fits to experimental data. The lumped EMs are considered convex basis vectors in yield space since it spans every physiological yield vector of extracellular metabolites. The HCM-FBA framework is a different approach of reducing/lumping EMs, since the FBA mode is

already a convex combination of EMs [5]. The advantage of HCM-FBA is it does not require the computational burden of calculating elementary modes.

HCM-FBA provides a practical and feasible approach to model large networks where elementary modes fail. FBA modes do not have the thorough information as elementary modes, but provide sufficient pathways to model experimental observations. HCM-EM has been shown to have comparable internal flux estimates to metabolic flux analysis results [12], this has not been shown for HCM-FBA. Just as HCM-EM has been applied to *S. cerevisiae*, we can apply the HCM-FBA framework to eukaryotes as long as FBA accounts for compartmentalization. FBA is frequently used to study genome-scale networks [1] and can be used to generate modes for the HCM framework. This opens up the possibility of genome scale cybernetic models.

IV. MATERIALS AND METHODS

The HCM-FBA approach is a modification of the HCM-EM strategy of Kim et al. [12]. 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:

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

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

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

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 $q_l(\mathbf{e}, \mathbf{k}, \mathbf{x})$ denotes the specific rate of flux through mode l, and was written as the product of a kinetic term ($\bar{q}_l$) and a 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 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. Lastly, cellmass was produced at the specific growth rate:

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

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

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

where z_{sl} denotes the uptake flux of substrate x through mode l. All model equations were implemented in Julia (v.0.4.2) [17], and solved using SUNDIALS [18]. The model code is available online:

Elementary mode and flux balance analysis: Elementary modes were calculated using METATOOL 5.1 [19]. 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:

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

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 [20]. For each FBA mode, the objective was to maximize either the specific growth rate, or the specific rate of byproduct formation. Multiple FBA modes were calculated for each objective flux 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 mM/gDW·hr and 0.05 mM/gDW·hr, respectively. Each flux vector was normalized 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 $\pm 50\%$ [21]. 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):

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

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. Orth, J., Thiele, I. & Palsson, B. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
2. Schuster, S., Fell, D. A. & Dandekar, T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol* **18**, 326–32 (2000).
3. Schilling, C. H., Letscher, D. & Palsson, B. O. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol* **203**, 229–48 (2000).
4. Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J. & Palsson, B. O. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**, 92–6 (2004).
5. Wiback, S. J., Mahadevan, R. & Palsson, B. Ø. Reconstructing metabolic flux vectors from extreme pathways: defining the alpha-spectrum. *J Theor Biol* **224**, 313–24 (2003).
6. Lee, L., Varner, J. & Ko, K. Parallel extreme pathway computation for metabolic networks. *Comput Syst Bioinformatics Conf, Int IEEE CS* **0**, 636–639 (2004).
7. Varner, J. & Ramkrishna, D. Metabolic engineering from a cybernetic perspective: aspartate family of amino acids. *Metab Eng* **1**, 88–116 (1999).
8. Song, H.-S. & Ramkrishna, D. Prediction of dynamic behavior of mutant strains from limited wild-type data. *Metab Eng* **14**, 69–80 (2012).
9. Kim, J. I., Song, H.-S., Sunkara, S. R., Lali, A. & Ramkrishna, D. Exacting predictions by cybernetic model confirmed experimentally: steady state multiplicity in the chemostat. *Biotechnol Prog* **28**, 1160–6 (2012).
10. Song, H.-S. & Ramkrishna, D. Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function. *Biotechnol Bioeng* **108**, 127–40 (2011).
11. Gadkar, K. G., Doyle, F. J., 3rd, Crowley, T. J. & Varner, J. D. Cybernetic model predictive control of a continuous bioreactor with cell recycle. *Biotechnol Prog* **19**, 1487–97 (2003).
12. 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).
13. 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).
14. Schuetz, R., Kuepfer, L. & Sauer, U. Systematic evaluation of objective functions for predicting intracellular fluxes in escherichia coli. *Mol. Syst. Biol.* **3** (2007).

15. Palsson, B. *Systems Biology: Properties of Reconstructed Networks* (Cambridge University Press, New York, NY, USA, 2006).

16. Song, H.-S. & Ramkrishna, D. Prediction of metabolic function from limited data: Lumped hybrid cybernetic modeling (l-hcm). *Biotechnology and Bioengineering* **106**, 271–284 (2010).

17. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: A fresh approach to numerical computing (2014). arXiv:`1411.1607`.

18. Hindmarsh, A. C. *et al.* Sundials: Suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Softw.* **31**, 363–396 (2005).

19. Kamp, A. & Schuster, S. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics* **22**, 1930–1931 (2006).

20. GNU Linear Programming Kit, Version 4.52 (2016). URL `http://www.gnu.org/software/glpk/glpk.html`.

21. Herman, J. Salib. available online: https://github.com/jdherman/salib .