

Metabolism

Model implementation. Our challenge in modeling metabolism at the large scale is best reflected in the two major approaches to modeling metabolic networks. Both approaches begin by writing a number of ordinary differential equations to reflect conservation of mass for each of the metabolites in the system. In a kinetics-based approach, the terms of the ODEs are represented in terms of enzyme and small molecule concentrations. The difficulty in implementing this approach at a large scale is that (1) a majority of the parameters are not known, and (2) the link between the parameter values - each measured with their own error - is rigid in this approach, such that it is generally impossible to build a model even at a smaller scale without abandoning most of the experimental data and performing extensive model fitting.

The second approach is called flux balance analysis (FBA), and is a common way to model large-scale metabolic network behavior with a low parameter requirement. However, traditional implementations of FBA are inappropriate for whole-cell modeling due to the dynamic nature of whole-cell simulation and fixed nature of the classic FBA objective function. Moreover, in cases such as *E. coli*, where many parameter values have been measured, it would be unfortunate not to use these measurements.

The method we describe here incorporates the best parts of both kinetics- and flux-balance-based modeling. This enables the model to run stably at a large scale, incorporating known parameter values without additional fitting, and in a manner that is readily integrable with the rest of a whole-cell modeling approach. This is done by using an alternative objective function that involves a multi-objective minimization for homeostatic metabolite composition and reaction kinetics that extends previous work by Birch *et al.*[3]. The effect of this multi-objective function is twofold: (1) to maintain cellular

concentrations of small molecule metabolites and (2) to enforce constraints on metabolic fluxes calculated from Michaelis-Menten kinetics based on metabolite concentrations and curated kinetic parameters. A weighting factor is used to balance the contribution from the two objectives.

We used the metabolic network reconstruction from Weaver *et al.* [?] because it was well-connected to the rest of EcoCyc’s resources and data which we relied on. This network reconstruction was based on the Orth model [5]. Different nutrient conditions (minimal M9, +amino acids, -oxygen, etc.) can be specified by changing bounds on metabolite import reactions, and shifts between these nutrient conditions can be programmatically varied.

Homeostatic objective. The homeostatic objective attempts to maintain small molecule metabolite concentrations at a constant value. For example, if during a time-step the net effect of other **Process** execution transforms ATP to ADP, the concentration of ATP will be lower and ADP higher. The homeostatic objective ensures that the metabolic network will attempt to increase the ATP concentration and decrease the ADP concentration using chemical transformations available in the network.

A total of 140 metabolite set-point concentrations are specified in the objective ($C_{o,i}$ in Equation 1). A list of those metabolites with concentrations in the model is given in Table S7. The homeostatic objective minimizes the deviation from these measured concentrations and can be specified as:

$$\text{minimize } \sum_i \left| 1 - \frac{C_i}{C_{o,i}} \right| \tag{1}$$

where C_i is the concentration of metabolite i and $C_{o,i}$ is the measured set-point concen-

tration for metabolite i . Cytoplasmic concentrations were chosen based on data from Bennett *et al.* [2], and other components of biomass have set-point concentrations specified based on the overall composition of the cell (lipids, metal ions, etc.) [8] and can be dependent on the media environment of the simulation.

This approach is an improvement over the *M. genitalium* simulations, where metabolites were produced in a fixed ratio at every time step regardless of the behavior of the rest of the simulated cell—this could lead to pooling or depletion of metabolites. Furthermore, if one metabolite could not be produced, none of the metabolites could be produced. Our homeostatic objective fixes both of these shortcomings.

Kinetics objective. The *E. coli* model simulates both metabolic enzyme expression via transcription and translation and dynamically maintains 140 metabolite concentrations. This enables the use of Michaelis-Menten kinetic equality constraints on metabolic fluxes using Equation 2:

$$v_{o,j} = k_{cat} \cdot E \cdot \left(\frac{C_1}{C_1 + K_{m,1}} \right) \cdot \left(\frac{C_2}{C_2 + K_{m,2}} \right) \cdots \left(\frac{C_n}{C_n + K_{m,n}} \right) \quad (2)$$

where $v_{o,j}$ is the kinetic target for the flux through reaction j that has n substrates, k_{cat} is the catalytic turnover rate for enzyme E , $K_{m,n}$ is the saturation constant for substrate n , E is the concentration of metabolic enzyme, and C_n is the concentration of substrate n in reaction j .

Kinetics data was reviewed from over 12,000 papers identified from BRENDA [6]. We filtered out papers that did not have a k_{cat} and which did not use a lab strain, or which concerned non-metabolic enzymes. The result was roughly 1200 papers which we manually curated due to our and others' observation that about 20% of the values in the BRENDA database are copied incorrectly from their primary source papers [1]. From this set, we further removed enzymes that were not included in our metabolic model and

consolidated the results of multiple studies (usually preferring experiments conducted at 37°C and/or the report with the highest k_{cat} value). In all, 179 constraints with a K_M and k_{cat} and 216 constraints with only a k_{cat} are used to constrain a total of 380 reactions (with some reactions having multiple constraints). Although some additional constraints were identified, they are not currently being used in the model. In particular, constraints were found for tRNA charging (18 reactions) but not used since tRNA charging is not explicitly included in the model. Additionally, as discussed in the main text, constraints for four reactions (succinate dehydrogenase, NADH dehydrogenase, inorganic pyrophosphatase, and glutathione reductase) were identified that, when included, caused a much higher glucose uptake rate than observed without kinetic constraints and higher than what has been experimentally measured. In related analysis also outlined in the main text, two more reactions (isocitrate dehydrogenase and glyoxylate reductase) were found to be under additional control that is not modeled so their constraints were excluded. Based on this, constraints related to these 24 reactions were excluded from the model. A complete list of the 431 reactions with curated kinetic parameters, including those mentioned above that were excluded from the model, is found in Table S8.

In cases where the enzyme parameters were recorded at non-physiological temperatures, we used the following scaling relationship to adjust the k_{cat} :

$$k_{cat,adj} = 2^{\frac{37-T}{10}} \cdot k_{cat} \quad (3)$$

where T is the reported temperature (in °C) for the experimental conditions—this increases the kinetic rate by a factor of 2 for every 10 °C below 37 °C. The factor 2 comes from an approximation of the change in rates determined by the Arrhenius equation:

$$k = A \cdot e^{-\frac{E_a}{R \cdot T}} \quad (4)$$

where k is the calculated rate, A is a constant for a given reaction, E_a is the activation

energy for a given reaction, R is the universal gas constant and T is the temperature. Assuming $E_a \approx 50,000 \frac{J}{mol}$ and the temperature increases 10 °C near 25 °C ($T_1 = 293$ K to $T_2 = 303$ K), we get an approximate doubling in the rate for the 10 °C increase:

$$\frac{k_2}{k_1} = \frac{A \cdot e^{-\frac{E_a}{R \cdot T_2}}}{A \cdot e^{-\frac{E_a}{R \cdot T_1}}} = \frac{e^{-\frac{50000}{8.314 \cdot 303}}}{e^{-\frac{50000}{8.314 \cdot 293}}} \approx 2.0 \quad (5)$$

Similar to the homeostatic objective, the kinetics objective minimizes deviation from the fluxes determined by the kinetic equations that are calculated at each time step based on the enzyme and metabolite concentrations. Formally:

$$\text{minimize } \sum_j \left| 1 - \frac{v_j}{v_{o,j}} \right| \quad (6)$$

where v_j is the flux through reaction j and $v_{o,j}$ is the target flux for reaction j calculated from Equation 2. This represents a soft kinetic constraint on reactions for which we have kinetic parameters and depends on concentrations of enzymes and metabolites in the model.

Combined objective. The two objectives above are linked in the combined objective by a factor λ , which is chosen to be $\ll 1$ to prioritize the homeostatic concentration objective function’s ability to produce growth at known doubling times, over fit the kinetic data. λ is not a biological parameter such as those found in Table S1 (it is better thought of as a “hyperparameter”), and thus it is not included in that table.

Finally, a hard kinetic constraint of no flux through a reaction is implemented for reactions that have no enzyme present by setting $v_{max} = 0$.

This results in the following linear optimization problem that is solved at each time step with GLPK to determine fluxes for each reaction which are used to update the counts of metabolites through exchange reaction fluxes included in the problem:

$$\begin{array}{ll}
\text{minimize} & (1 - \lambda) \sum_i \left| 1 - \frac{C_i}{C_{o,i}} \right| + \lambda \sum_j \left| 1 - \frac{v_j}{v_{o,j}} \right| \\
\text{subject to} & S \cdot v = 0 \\
& v_j \geq v_{min,j} \\
& v_j \leq v_{max,j}
\end{array}$$

Algorithm 1: Metabolism

Input : C_i concentration for metabolite i

Input : $C_{o,i}$ concentration target for metabolite i

Input : $k_{cat,j}$ turnover number for enzyme j

Input : $K_{m,i,j}$ Michaelis constant for metabolite i for enzyme j

Input : E_j concentration for enzyme j

Input : S stoichiometric matrix for all reactions

1. Set physical constraints on reaction fluxes

For all reactions: $v_{min,j} = -\inf, v_{max,j} = +\inf$

For thermodynamically irreversible reactions: $v_{min,j} = 0$

If required enzyme not present: $v_{min,j} = v_{max,j} = 0$

2. Calculate kinetic target ($v_{o,j}$) for each reaction j based on the enzymes j and metabolites i associated with each reaction for a soft flux constraint

$$v_{o,j} = k_{cat,j} \cdot E_j \cdot \prod_i \left(\frac{C_i}{K_{m,i,j} + C_i} \right)$$

3. Solve linear optimization problem

$$\text{minimize} \quad (1 - \lambda) \sum_i \left| 1 - \frac{C_i}{C_{o,i}} \right| + \lambda \sum_j \left| 1 - \frac{v_j}{v_{o,j}} \right|$$

$$\text{subject to} \quad S \cdot v = 0$$

$$v_j \geq v_{min,j}$$

$$v_j \leq v_{max,j}$$

4. Update concentrations of metabolites based on the solution to the linear optimization problem

Result: Metabolites are taken up from the environment and converted into other metabolites for use in other processes

Associated data

Parameter	Symbol	Units	Value	Reference
Metabolic network	S	-	Stoichiometric coefficients	[4]
Metabolic target fluxes	v_o	$\mu M/s$	[0, 87000]	See GitHub
Metabolic fluxes (validation)	v_v	$\mu M/s$	[82, 1500]	[7]
Enzyme turnover number	k_{cat}	$1/s$	[0.00063, 38000]	See GitHub
Enzyme Michaelis constant	K_m	μM	[0.035, 550000]	See GitHub
Metabolite target concentration	C_o	μM	[0.063, 97000]	[2]
Kinetic objective weight	λ	-	1e-6	See GitHub

Table 1: Table of parameters for metabolism process.

Associated files

wcEcoli Path	File	Type
wcEcoli/models/ecoli/processes	metabolism.py	process
wcEcoli/reconstruction/ecoli/dataclasses/process	metabolism.py	data
wcEcoli/reconstruction/ecoli/flat	biomass.tsv	raw data
wcEcoli/reconstruction/ecoli/flat	enzymeKinetics.tsv	raw data
wcEcoli/reconstruction/ecoli/flat	metaboliteConcentrations.tsv	raw data
wcEcoli/reconstruction/ecoli/flat	reactions.tsv	raw data
wcEcoli/validation/ecoli/flat	toya_2010_central_carbon_fluxes.tsv	validation data

Table 2: Table of files for metabolism.

References

- [1] Arren Bar-Even, Elad Noor, Yonatan Savir, Wolfram Liebermeister, Dan Davidi, Dan S Tawfik, and Ron Milo. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, 50(21):4402–4410, 2011.
- [2] Bryson D Bennett, Elizabeth H Kimball, Melissa Gao, Robin Osterhout, Stephen J Van Dien, and Joshua D Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli. *Nature chemical biology*, 5(8):593–599, 2009.
- [3] Elsa W Birch, Madeleine Udell, and Markus W Covert. Incorporation of flexible objectives and time-linked simulation with flux balance analysis. *Journal of Theoretical Biology*, 345:12–21, March 2014.
- [4] Peter D Karp, Daniel Weaver, Suzanne Paley, Carol Fulcher, Aya Kubo, Anamika Kothari, Markus Krummenacker, Pallavi Subhraveti, et al. The EcoCyc database. *EcoSal Plus*, 2014.
- [5] Jeffrey D Orth, Tom M Conrad, Jessica Na, Joshua A Lerman, Hojung Nam, Adam M Feist, and Bernhard Ø Palsson. A comprehensive genome-scale reconstruction of escherichia coli metabolism2011. *Molecular systems biology*, 7(1):535, 2011.
- [6] Ida Schomburg, Antje Chang, Sandra Placzek, Carola Söhngen, Michael Rother, Maren Lang, Cornelia Munaretto, Susanne Ulas, Michael Stelzer, Andreas Grote, Maurice Scheer, and Dietmar Schomburg. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and

contents in BRENDA. *Nucleic acids research*, 41(Database issue):D764–72, January 2013.

- [7] Yoshihiro Toya, Nobuyoshi Ishii, Kenji Nakahigashi, Takashi Hirasawa, Tomoyoshi Soga, Masaru Tomita, and Kazuyuki Shimizu. ^{13}C -metabolic flux analysis for batch culture of *Escherichia coli* and its *pyk* and *pgi* gene knockout mutants based on mass isotopomer distribution of intracellular metabolites. *Biotechnology progress*, 26(4):975–992, 2010.
- [8] Daniel S Weaver, Ingrid M Keseler, Amanda Mackie, Ian T Paulsen, and Peter D Karp. A genome-scale metabolic flux model of *Escherichia coli* K-12 derived from the EcoCyc database. *BMC Systems Biology*, 8:79, June 2014.