Bulletin of Mathematical Biology manuscript No.
(will be inserted by the editor)

An Approach to Maximize Growth and Entropy Production Rates in Metabolism

Control in Physics-Based Models of Metabolism

Ethan King · Jesse Holzer · Justin A. North · William R. Cannon

Received: date / Accepted: date

Abstract Elucidating cell regulation remains a challenging task due to the complexity of metabolism and the difficulty of experimental measurements. Here we present a method for prediction of cell regulation to maximize cell growth rate while maintaining the solvent capacity of the cell. Prediction is formulated as an optimization problem using a thermodynamic framework that leverages experimental data. We develop a formulation and variable initialization procedure that allows for computing solutions of the optimization with an interior point method. The approach is applied to photoheterotrophic growth of *Rhodospirilium rubrum* using ethanol as a carbon source, which has applications to biosynthesis of ethylene production. Growth is captured as the rate of synthesis of amino acids into proteins, and synthesis of nucleotide triphoshaptes into RNA and DNA. The method predicts regulation that produces a high rate of protein and RNA synthesis while DNA synthesis is reduced close to zero in agreement with production of DNA being turned off for much of the cell cycle.

Keywords Cell Regulation · Maximum Entropy Production · Metabolism

Ethan King
Pacific Northwest National Laboratory, Richland, WA
E-mail: ethan.king@pnnl.gov

Jesse Holzer
Pacific Northwest Natrional Laboratory, Richland, WA
E-mail: jesse.hozer@pnnl.gov

Justin A. North
Department of Microbiology, The Ohio State University, Columbus, OH
E-mail: north.62@osu.edu

William R. Cannon
Pacific Northwest National Laboratory, Richland, WA
E-mail: william.cannon@pnnl.gov

1 Introduction

Biological systems are dissipative systems analogous to tornadoes and hurricanes in that their dynamics act to maximize their entropy production rates as quickly as possible. This is to say that they move to the most probable state possible by taking the least action path. However, biological systems differ from other dissipative systems found in nature in that they have complex mechanisms for adapting to changing environmental conditions. These adaptation mechanisms serve to control the dynamics in such a way that their ability to carry out auto-catalysis [15] is preserved. Adaptation is accomplished through regulation of metabolic reaction pathways such that organisms remain viable in a physico-chemical sense despite rapid or dramatic shifts in the environment.

Because regulation is at the heart of understanding biological behavior, discovery and understanding of the basis for regulation is critical to the field of biology. Regulation may be in the form of just-in-time regulation, graded regulation, switch-like regulation, pre-programmed regulation such as the circadian clock system, or even the choice to not regulate when it is too costly to alter the respective enzyme expression but to instead constitutively turn on an activity [30]. Regulation is an important aspect of fitness in that organisms and even individual cells must regulate themselves so that they act as efficiently and as quickly as possible, otherwise they are outcompeted in their environment. Consequently, the environment over the time of evolution shapes regulation. Cells learn to behave in the most efficient manner in nutrient poor and dramatically shifting environments. In rich environments in which selection pressure is not strong, dysregulation of cells can result in uncontrolled growth, manifest as cancer in metazoans.

Regulation in biological systems has been historically uncovered using experiments carefully designed to test hypotheses, typically employing isotope labeling methods to track reaction fluxes. Such meticulous but labor intensive studies have been quite successful at uncovering metabolic regulation [20, 24, 29, 1, 19, 36, 35, 34, 16, 33, 18, 12], yet we are far from a complete understanding of regulation in even model organisms. High-throughput methods to uncover regulation are highly desirable, and recent work in this area has made progress [11, 28], but we are still a long way from a routine method for determining regulation, either by experiment or prediction.

In a recent study by Hackett, *et al.*, Michealis-Menten steady-state kinetic models with regulation were used to predict reaction fluxes and concentrations. The predictions were compared to reaction fluxes inferred from 13C isotope experiments and concentrations derived from mass spectrometry and NMR measurements [11]. The correlation between simulation-predicted fluxes and experimentally-inferred fluxes were evaluated with and without regulation in the simulation. If the match was better with regulation, then regulation was assumed. The work was a *tour de force* in that 25 chemostat studies were used to carefully measure both absolute and relative metabolomics data while at the same time cover as much of the proteome as possible.

An approach used by Reznick, *et al.*, has less reliance on multimodal experimental designs, and instead used sophisticated informatics to develop a model of small molecule regulatory networks from curated databases of enzymes. They integrated the regulatory network with a metabolic model of *E. coli*, and distilled information on how substrates and inhibitors contribute to metabolic flux regulation [28].

More recently, Britton, *et al.* have developed optimization and reinforcement learning methods that predict which enzymes need to be regulated to maintain metabolite concentrations at a level such that the diffusion process within the cell remains viable. These methods took advantage of a maximum path-entropy/caliber [17,9,10] formulation of the law of mass action in order to predict likely metabolite concentrations. The advantage of the approach is that minimal measurements and few experimentally-derived parameters are required to predict regulation.

Maximum entropy methods are attractive because they use the most likely values of the parameters needed to model a process [17,9,10]. That is, rather than a large search through parameter space to determine the most likely kinetic parameters, a maximum entropy formulation can provide the solution directly, then if needed, reaction parameters can be back calculated. Strictly speaking, these methods maximize the path entropy of a system, meaning that they maximize the entropy subject to the stoichiometric constraints imposed by the reactions and the constraints due to the system boundary conditions. Without the stoichiometric constraints, each molecular species would move to its natural abundance determined by its standard chemical potential. Without non-equilibrium boundary constraints, the reactions would all go to the equilibrium state.

In the study by Britton, et al, regulation was inferred only with regard to which reactions needed to be controlled to keep metabolites at concentrations that are not so high that they would impede diffusion [7]. As mentioned above, in an environmental niche with finite nutrient resources for growth, cells also need to replicate in a timely and efficient manner in order to compete with other species. Cells must regulate themselves to maximize growth, and only produce those enzyme catalysts that contribute to timely and efficient growth.

In this study, we report on an optimization approach to predict regulation that maximizes growth pathways while still constraining metabolite levels to physiological values. We apply this method, referred to as pathway-controlled optimization (PCO), to the metabolism and enzymatic activities of *Rhodospirilium rubrum*, a purple non-sulfur photosynthetic bacterium that is being used as a synthetic biology organism for the purpose of ethylene production [25]. We compare the regulation and reaction fluxes to models with regulation applied to just maintain metabolite concentrations at physiological levels. We characterize the set of solutions obtained from the new approach with regard to sensitivity to initial conditions.

2 Methods

2.1 Maximum Path Entropy Solution without Regulation

For a set of molecular species $\mathcal{I} = \{A, B, C, D\}$ having respective concentrations n_i for each species $i \in \mathcal{I}$, participating in a set of reactions $\mathcal{J} = \{1, -1\}$, with unsigned stoichiometric coefficients $\nu_{i,j}$ for each molecular species i and each reaction $j \in \mathcal{J}$, the reversible reaction is described by the chemical equation,

$$\nu_{\mathrm{A},1} n_A + \nu_{\mathrm{B},1} n_B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \nu_{\mathrm{C},1} n_C + \nu_{\mathrm{D},1} n_D. \tag{1}$$

Here, k_1 and k_{-1} are the reaction rate parameters.

Let $|\mathcal{I}|$ denote the size of the set $\mathcal{I}$ and let $n \in \mathbb{R}^{|\mathcal{I}|}$ be the vector of molecular counts with elements n_i. If S is the stoichiometric matrix of elements $S_{i,j} = \gamma_{i,j}$ where $\gamma_{i,j}$ are the signed stoichiometric coefficients such that $\nu_{i,j} = |\gamma_{i,j}|$, we have according to the law of mass action,

$$\frac{dn}{dt} = SJ(n) \tag{2}$$

where for a given n, $J(n)$ is the net flux of the set of forward and reverse reactions. The chemical species occurring on the left hand side of the equation are known as reactants and belong to the subset $\mathcal{I}_{R_j} \subset \mathcal{I}$, while those on the right hand side are known as products and belong to the subset $\mathcal{I}_{P_j} \subset \mathcal{I}$.

For any reversible reaction with forward and reverse reactions $+j$ and $-j$, the net flux is given by,

$$J_j(n) = k_j \prod_{i \in \mathcal{I}_{R_j}} n_i^{\nu_{i,j}} - k_{-j} \prod_{i \in \mathcal{I}_{R_{-j}}} n_i^{\nu_{i,-j}}. \tag{3}$$

Eqn 3 is a purely kinetic description of the reaction flux. For elementary reactions, thermodynamic terms are introduced into the law of mass action by a simple factorization,

$$J_j(n) = k_{-j} \prod_{i \in \mathcal{I}_{R_{-j}}} n_i^{\nu_{i,-j}} \left(\frac{k_j \prod_{i \in \mathcal{I}_{R_j}} n_i^{\nu_{i,j}}}{k_{-j} \prod_{i \in \mathcal{I}_{R_{-j}}} n_i^{\nu_{i,-j}}} \right) - k_j \prod_{i \in \mathcal{I}_{R_j}} n_i^{\nu_{i,j}} \left(\frac{k_{-j} \prod_{i \in \mathcal{I}_{R_{-j}}} n_i^{\nu_{i,-j}}}{k_j \prod_{i \in \mathcal{I}_{R_j}} n_i^{\nu_{i,j}}} \right) \tag{4}$$

$$= k_{-j} \prod_{i \in \mathcal{I}_{R_{-j}}} n_i^{\nu_{i,-j}} \left(K_j \prod_i n_i^{\gamma_{i,j}} \right) - k_j \prod_{i \in \mathcal{I}_{R_j}} n_i^{\nu_{i,j}} \left(K_{-j} \prod_i n_i^{\gamma_{i,-j}} \right), \tag{5}$$

where $K_j = \frac{k_j}{k_{-j}}$ is the equilibrium constant for reaction j. In Eqn 5, the terms in parentheses are related to the thermodynamic forces on the reaction while

the terms outside the parentheses are the time-dependent kinetic terms. Setting the latter terms to a constant gives the Marcelin Equation [21], in which each forward and reverse reaction occurs on the same timescale by postulate. While this postulate is incorrect for determining the true dynamics of the system, the Marcelin formulation of Eqn 5 does give a way to the maximum path entropy solution [8,7] by assuming that all reactions occur on the same timescale and that their fluxes are proportional to the thermodynamic forces on them. Once these fluxes and concentrations are known from the maximum path entropy solution, they can be used to back-calculate the rate constants that will give the corresponding true dynamics [8,7].

Using the maximum path entropy solution, and representing the thermodynamic reaction forces for each reaction $j \in \mathcal{J}$ as,

$$f_j(n) = K_j \prod_{i \in \mathcal{I}} n_i^{-\gamma_{i,j}}, \tag{6}$$

the flux is then given by

$$J_j(n) = f_j(n) - \frac{1}{f_j(n)} \tag{7}$$

and we can express the time dependence of each metabolite $i \in \mathcal{I}$ as follows

$$\frac{dn_i}{dt} = \sum_{j \in \mathcal{J}} \gamma_{i,j} (f_j(n) - \frac{1}{f_j(n)}). \tag{8}$$

2.2 Steady State

At the maximum path entropy configuration the system is in a steady state, where a set of metabolites, denoted $\mathcal{I}_f$, are assumed to be held at fixed concentrations as boundary conditions for the system. For the rest of the metabolites, $\mathcal{I}_v = \mathcal{I} \setminus \mathcal{I}_f$, their steady state concentration is free and their rate of change is zero. For all $i \in \mathcal{I}_f$, let $\bar{n}_i$ be the fixed concentration of the metabolite. A steady state is defined as a solution to the following system of equations,

$$\begin{aligned} \frac{dn_i}{dt} &= 0 \quad \forall i \in \mathcal{I}_v \,, \\ n_i &= \bar{n}_i \quad \forall i \in \mathcal{I}_f \,. \end{aligned} \tag{9}$$

2.3 Controlling Metabolite Concentrations

Metabolite concentrations predicted using the maximum path entropy approach without regulation will produce values that are much too large to be physiologically reasonable — concentrations may approach the limit of their solubility, causing the cytoplasm to become glass-like [27,14].

That regulation is needed to control concentrations *in vivo* was proposed early in the field of enzymology [2,3]. In a previously reported method to

control concentrations [7], Britton, *et al.* applied regulation to reactions using a scalar valued activity coefficient $\alpha_j \in [0.0, 1.0]$ for each reaction j, that linearly scales the reaction rate such that the time dependence of metabolite concentration n_i is given by,

$$\frac{dn_i}{dt} = \sum_{j \in \mathcal{J}} \gamma_{i,j} J_j(n, \alpha_j)\,, \tag{10}$$

where

$$J_j(n, \alpha_j) = \alpha_j \left(f_j(n) - \frac{1}{f_j(n)} \right)\,. \tag{11}$$

When $\alpha_j = 1.0$, the reaction is fully active and when $\alpha_j = 0.0$, the activity, and hence reaction rate $J_j(n, \alpha_j)$, is zero. Activities are adjusted in a deterministic manner by characterizing the sensitivity of a metabolite concentration to an enzyme activity using Metabolic Control Analysis (MCA). In MCA, the concentration control coefficient $C_{i,j}$ measures the sensitivity of a concentration n_j to an activity coefficient α_j,

$$C_{i,j} = \frac{\partial \log n_i}{\partial \log \alpha_j}. \tag{12}$$

Using this approach, metabolites whose predicted concentrations are furthest from their experimentally observed values, as measured by the log ratio of the predicted to observed concentration, are reduced first. The concentration of a metabolite n_i is reduced by adjusting the activity coefficient α_j that has the largest influence on n_i and all other metabolites that exceed their experimentally observed concentrations, as determined by the sensitivity analysis. This process is iteratively carried out until all concentrations are at or below the experimentally observed concentrations. Details of the approach are provided in the study by Britton, *et al.* [7]. In the cases discussed herein, the maximum value of the experimentally observed metabolite concentrations are taken to be $n_i = 1.0$ mM for $i \in \mathcal{I}_v$ and for $i \in \mathcal{I}_f$ the values are taken from mass spectrometry observed concentrations for bacteria [4,26].

2.4 Controlling Metabolite Concentrations and Maximizing Growth

While controlling metabolite concentrations may be a primary role of metabolic regulation, natural selection also requires that organisms be regulated to grow fast and efficiently - using the available energy from the environment to ensure survival and compete with others. Here we describe a method, referred to as pathway-controlled optimization (PCO), whose goal is to obtain this biological objective.

Let $\mathcal{G} \subset \mathcal{J}$ be the set of reactions corresponding to production of biomass. We formulate the steady state with maximum biomass production as the solution of the optimization problem,

$$\max \sum_{j \in \mathcal{G}} J_j(n, \alpha_j) \tag{13a}$$

subject to:

$$\frac{dn_i}{dt} \;=\; 0 \;\; \forall\, i \in \mathcal{I}_v\,, \tag{13b}$$

$$n_i \;=\; \bar{n}_i \;\; \forall\, i \in \mathcal{I}_f\,, \tag{13c}$$

$$0 \leq n_i \leq n_{\max} \;\; \forall\, i \in \mathcal{I}_v\,, \tag{13d}$$

$$0 \leq \alpha_j \leq 1 \;\; \forall\, j \in \mathcal{J}\,. \tag{13e}$$

The objective seeks to maximize the flux through the growth reactions $\mathcal{G}$ while the constraints (13b) - (13c) ensure that the steady state (9) is satisfied. The activity coefficients and metabolite concentrations are further restricted to physiologically meaningful values with the constraints (13e) and (13d).

The formulation of the PCO problem is simple to express but difficult to solve. The steady state constraints (13b) are nonlinear and non-convex presenting significant challenges to optimization. Values for the flux, activity coefficients, and metabolite concentrations can also vary over many orders of magnitude, which introduces additional difficulty in employing numerical methods to compute solutions. In this work we present a more computationally tractable reformulation of the constraints and present numerical solutions from an interior point solver.

2.4.1 Representing the steady state condition

To simplify solving the optimization, we reformulate the steady state constraint (13b) to be more numerically tractable. Let $S_v \in \mathbb{R}^{|\mathcal{I}_v|,|\mathcal{J}|}$ be the submatrix of S with the rows corresponding to only the variable metabolites. Then at a steady state the reaction fluxes must satisfy

$$S_v J(n, \alpha) = 0\,, \tag{14}$$

that is $J(n, \alpha) \in \mathcal{N}(S_v)$, where $\mathcal{N}(S_v)$ is the nullspace of S_v defined

$$\mathcal{N}(S_v) = \{x \in \mathbb{R}^{|\mathcal{J}|} \;:\; S_v x = 0\}\,. \tag{15}$$

We separate (14) into the identification of a vector of fluxes $y \in \mathbb{R}^{|\mathcal{J}|}$ which satisfy steady state and the construction of a vector of metabolites $n \in \mathbb{R}^{|\mathcal{I}|}$ which achieve those fluxes with the two conditions,

$$\begin{aligned} &y \in \mathcal{N}(S_v)\,, \\ &y_j = J_j(n, \alpha_j) \;\; \forall\, j \in \mathcal{J}\,. \end{aligned} \tag{16}$$

The expression $y_j = J_j(n, \alpha_j)$ can be made more tractable by considering the computation of the reaction flux in terms of the log of the metabolites. For $\eta \in \mathbb{R}^{|\mathcal{I}|}$ let

$$\eta_i = \log(n_i) \;\; \forall\, i \in \mathcal{I}\,, \tag{17}$$

then we have from (7) that the flux as a function of η in the maximum path entropy formulation is given by

$$J_j(\eta, \alpha_j) = \alpha_j K_j e^{-\langle\ (S)_j, \eta\ \rangle} - \alpha_j \frac{1}{K_j} e^{\langle\ (S)_j, \eta\ \rangle}\,, \tag{18}$$

where $(S)_j$ is the jth column of S, and $\langle \cdot, \cdot \rangle$ is the standard inner product. It follows that in terms of the log of the metabolite counts, the conditions (16) are equivalent to

$$y \in \mathcal{N}(S_v)\,, \tag{19a}$$
$$S^T \eta = \hat{y}(y, \alpha)\,, \tag{19b}$$
$$\hat{y}_j(y, \alpha) = \log\left(\frac{K_j}{2\alpha_j}\left(-y_j + \sqrt{y_j^2 + 4\alpha_j^2}\right)\right) \ \ \forall\ j \in \mathcal{J}\,. \tag{19c}$$

Computation of $\hat{y}(y, \alpha)$ is prone to error. For example, in the instance where y_j is positive, rounding may result in evaluating the log of zero. To avoid these issues we use the following equivalent expression

$$\hat{y}_j(y, \alpha) = \log(K_j) + \operatorname{sgn}(y_j)\left(\log(2\alpha_j) - \log\left(|y_j| + \sqrt{y_j^2 + 4\alpha_j^2}\right)\right)\,, \tag{20}$$

with sgn the signum function given by

$$\operatorname{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}\,. \tag{21}$$

A further complication in using the steady state conditions (19) as constraints for optimization is that the value of $\hat{y}(y, \alpha)$ can become extremely sensitive to small changes in the flux. In particular the partial derivative

$$\frac{\partial \hat{y}_j}{\partial y_j} = \frac{1}{\sqrt{y_j^2 + 4\alpha_j^2}} \tag{22}$$

grows unbounded as α_j and y_j go to zero. In this application we consistently found solutions with α_j and y_j on the order of 10^{-9} for some reactions while others were unregulated with values of y_j on the order of 10^3. These discrepancies in scale are challenging for numerical methods.

However, it is not necessary to explicitly compute the activity coefficients in order to characterize a steady state. Rather, it is sufficient to identify metabolite concentrations that produce fluxes that can be reduced to a steady state. Instead of the equality constraints in (16) it is sufficient for the flux and metabolites to satisfy the conditions

$$\begin{aligned} &|y_j| \leq |J_j(n, \mathbf{1})| \ \ \forall\ j \in \mathcal{J}\,, \\ &y_j J_j(n, \mathbf{1}) > 0\,, \end{aligned} \tag{23}$$

where $\mathbf{1}$ is the vector of all ones corresponding to no regulation for all reactions.

If (23) is satisfied then any reaction fluxes $J_j(n, \mathbf{1})$ with magnitude greater than the steady state value y_j can always be reduced to satisfy equality so long as they have the same sign, which is captured by the second condition in (23). This idea can also be captured in our modified formulation (19) for the steady state conditions.

For metabolites η and steady state fluxes y there exists an α such that (19) holds if the following hold

$$\begin{aligned} &|g_j - \log(K_j)| \geq -\left(\log(2) - \log\left(|y_j| + \sqrt{y_j^2 + 4}\right)\right) \ \ \forall\, j \in \mathcal{J}\,, \\ &(\log(K_j) - g_j)y_j \geq 0 \ \ \forall\, j \in \mathcal{J}, \end{aligned} \tag{24}$$

where $g = S^T\eta$. This follows from (19b)-(19c), (20), and the inequality

$$\log(2\alpha_j) - \log\left(|y_j| + \sqrt{y_j^2 + 4\alpha_j^2}\right) \leq \log(2) - \log\left(|y_j| + \sqrt{y_j^2 + 4}\right)\,, \tag{25}$$

for all $y_j \in \mathbb{R}$ and $\alpha_j \in [0, 1]$.

Therefore we can formulate steady state conditions equivalent to (19) without the need to explicitly specify the activity coefficients as follows,

$$y \in \mathcal{N}(S_v)\,, \tag{26a}$$
$$g = S^T\eta\,, \tag{26b}$$
$$h_j = \operatorname{sgn}(y_j)\left(\log(2) - \log\left(|y_j| + \sqrt{y_j^2 + 4}\right)\right) \ \ \forall\, j \in \mathcal{J}\,, \tag{26c}$$
$$|g_j - \log(K_j)| \geq |h_j| \ \ \forall\, j \in \mathcal{J}\,, \tag{26d}$$
$$(\log(K_j) - g_j)y_j \geq 0 \ \ \forall\, j \in \mathcal{J}\,. \tag{26e}$$

2.4.2 Numerical formulation of the steady state constraints

Implementing the conditions (26a) - (26e) as constraints for numerical optimization requires further reformulation. The constraint (26d) is non-convex, and represents an 'or' condition depending on the sign of the flux for each reaction. This switching condition makes the feasible set non-convex and therefore challenging for optimization methods to search over. To find solutions we use a big M relaxation of (26d) with the constraints

$$h_j = \operatorname{sgn}(y_j)\left(\log(2) - \log\left(|y_j| + \sqrt{y_j^2 + 4}\right)\right) \ \ \forall\, j \in \mathcal{J}\,, \tag{27a}$$
$$g_j - \log(K_j) \geq h_j - u_j M \ \ \forall\, j \in \mathcal{J}\,, \tag{27b}$$
$$g_j - \log(K_j) \leq h_j + (1 - u_j)M \ \ \forall\, j \in \mathcal{J}\,, \tag{27c}$$
$$2u_j - 1 = \operatorname{sgn}(y_j) \ \ \forall\, j \in \mathcal{J}\,, \tag{27d}$$

where M is taken to be a large constant. For each reaction the variable u_j is a 'switching' term that relaxes either the constraint (27b) or (27c) depending on the sign of y_j such that only the correct constraint is imposed for each reaction.

In several constraints we utilize the signum function of the flux which is a discrete function, whereas the optimization method we employ here requires continuous functions of all variables for the constraints. Therefore we approximate the signum function with

$$\tilde{\text{sgn}}(x) = \frac{\lambda x}{\lambda|x| + \epsilon} \tag{28}$$

for scaling parameters λ and ϵ.

To implement the constraint $y \in \mathcal{N}(S_v)$ we construct a basis for the nullspace. For m the dimension of $\mathcal{N}(S_v)$, let $B \in \mathbb{R}^{|\mathcal{J}|,m}$ be a matrix with columns given by a basis of $\mathcal{N}(S_v)$. Then we have that

$$\mathcal{N}(S_v) = \{B\beta \ : \ \beta \in \mathbb{R}^m\}\,, \tag{29}$$

and we capture the constraint $y \in \mathcal{N}(S_v)$ with the expression $y = B\beta$ for a $\beta \in \mathbb{R}^m$.

2.4.3 Maximum Growth Optimization Numerical Formulation

Using the steady state conditions as outlined in sections 2.4.2 and 2.4.1 we formulate the maximum growth pathway-controlled optimization problem as

$$\max_{y,\eta} \ \sum_{j \in \mathcal{G}} y_j \tag{30a}$$

subject to:

$$y = B\beta\,, \tag{30b}$$
$$g = S^T\eta\,, \tag{30c}$$
$$h_j = \tilde{\text{sgn}}(y_j)\left(\log(2) - \log\left(|y_j| + \sqrt{y_j^2 + 4}\right)\right) \ \ \forall\ j \in \mathcal{J}\,, \tag{30d}$$
$$g_j - \log(K_j) \geq h_j - uM \ \ \forall\ j \in \mathcal{J}\,, \tag{30e}$$
$$g_j - \log(K_j) \leq h_j + (1-u)M \ \ \forall\ j \in \mathcal{J}\,, \tag{30f}$$
$$(\log(K_j) - g_j)y_j \geq 0 \ \ \forall\ j \in \mathcal{J}\,, \tag{30g}$$
$$2u_j - 1 = \tilde{\text{sgn}}(y_j) \ \ \forall\ j \in \mathcal{J}\,, \tag{30h}$$
$$\eta_i \leq \eta_{\max_i} \ \ \forall\ i \in \mathcal{I}_v\,, \tag{30i}$$
$$\eta_i = \bar{\eta}_i \ \ \forall\ i \in \mathcal{I}_f\,, \tag{30j}$$

where the constraints (30b)-(30h) represent the steady state constraint (13b). Note that the activity coefficients α for reactions are not explicitly included, however for a solution with optimal log metabolite counts η^* and fluxes y^*, the associated activity coefficients can be recovered as

$$\alpha_j^* = \frac{y_j^*}{J_j(\eta^*)} \ \ \forall\ j \in \mathcal{J}\,. \tag{31}$$

2.4.4 Initialization for optimization

Solving (30) using IPOPT, a general nonlinear interior point solver [32], we found that the solution and convergence of the solver was extremely sensitive to the variable initialization, as is common for nonlinear problems. In particular, the switching conditions as the flux changes direction in constraints (30e)-(30h) constitute a significant challenge for optimization. Initialization with numerous fluxes in the direction opposite the optimal solution can lead to a failure of the solver to converge, or convergence to only a local optimum. To initialize the variables we first solve for a feasible point near a maximum path entropy solution. Using random or other initialization methods we found that the solver failed to converge, or converged to points with smaller objective values.

To compute initial variable values we use two steps. First, we solve for a steady state solution that does not constrain the metabolites or regulate reactions by solving the least squares problem

$$\min_{y,\beta,\eta} \ \left|\left|S^T\eta - \hat{y}(y,\alpha)\right|\right|^2 \tag{32a}$$

subject to:

$$y = B\beta\,, \tag{32b}$$
$$\alpha_j = 1 \ \ \forall\ j \in \mathcal{J}\,, \tag{32c}$$
$$\eta_i = \bar{\eta}_i \ \ \forall\ i \in \mathcal{I}_f\,, \tag{32d}$$

We then use the solution of the least squares problem (32) as the initial variable values from which we solve for a steady state solution that satisfies the metabolite upper bounds, by solving the optimization

$$\min \ \left|\left|S^T\eta - g\right|\right|^2 \tag{33a}$$

subject to:

$$y = B\beta\,, \tag{33b}$$
$$h_j = \tilde{\text{sgn}}(y_j)\left(\log(2) - \log\left(|y_j| + \sqrt{y_j^2+4}\right)\right) \ \ \forall\ j \in \mathcal{J}\,, \tag{33c}$$
$$g_j - \log(K_j) \geq h_j - uM \ \ \forall\ j \in \mathcal{J}\,, \tag{33d}$$
$$g_j - \log(K_j) \leq h_j + (1-u)M \ \ \forall\ j \in \mathcal{J}\,, \tag{33e}$$
$$(\log(K_j) - g_j)y_j \geq 0 \ \ \forall\ j \in \mathcal{J}\,, \tag{33f}$$
$$2u_j - 1 = \tilde{\text{sgn}}(y_j) \ \ \forall\ j \in \mathcal{J}\,, \tag{33g}$$
$$\eta_i \leq \eta_{\max_i} \ \ \forall\ i \in \mathcal{I}_v\,, \tag{33h}$$
$$\eta_i = \bar{\eta}_i \ \ \forall\ i \in \mathcal{I}_f\,. \tag{33i}$$

The solution of (33) is then used to supply the initial variable values for solving (30).

2.5 Initialization Experiments

To explore the affect of our initialization procedure in section 2.4.4 we solved one hundred problem instances. For each instance we used a random initialization of the variables for solving (32) with all initial β values uniformly sampled from the interval $[-1, 1]$.

We solved (32) with the SciPy least_squares, implemented with the Levenberg-Marquardt algorithm, using the defaults except with ftol= 2.23×10^{-16} and xtol= 2.23×10^{-16} [31,23]. Solutions of (30) and (33) were computed with IPOPT using Pyomo [13] again using the default optimization parameters except with tol= 1.0×10^{-7} and maximum iterations set to $10,000$. All computations were done on a Macbook Pro with a 2.3 GHz 8-Core Intel Core i9 processor and 16 GB 2667 MHz DDR4 memory.

Of the one hundred trials, 18 trials terminated after hitting the maximum number of iterations. All other trials converged to a solution satisfying the optimization tolerances. The best objective value achieved was 11725.09014. Many trials achieved a similar objective to the maximum, within the top 46 trials the mean objective value was 11724.0901 and the standard deviation was 6E-6. These top 46 solution were near identical, and for all the results reported here we give the average values of these solutions.

3 Results and Discussion

3.1 Biological Results

In order to place the results in the proper context, a brief overview of the metabolism of *Rhodospirilium rubrum*, the subject of the biomass optimization, is necessary. *R. rubrum* is capable of growing both photoautotrophically, meaning that it only uses CO_2 as a carbon source, or photoheterotrophically, meaning that it can carry out photosynthesis while simultaneously assimilating organic carbon. In this study, we are modeling photoheterotropic growth, in which energy is acquired through photosynthesis while carbon is mostly acquired by uptake and assimilation of ethanol.

During photosynthesis, *R. rubrum* uses the Calvin cycle to regenerate chemical precursors for the assimilation of CO_2. The Calvin cycle includes reactions shared with the non-oxidative branch of the pentose phosphate cycle, some of which can act as an alternate route in the Calvin Cycle. The difference between the two routes in the Calvin cycle is that one route produces erythrose-4-phosphate and the other route consumes erythrose-4-phosphate. Erythrose-4-phosphate is a key metabolite for the synthesis of vitamin B6, pyridoxal-5'-phosphate, an essential cofactor in amino acid, and hence protein, synthesis. Hence, it is important that during growth, erythrose-4-phosphate not be entirely used as a Calvin cycle precursor.

We define the biomass pathways to be optimized as the pathways regarding the incorporation of amino acids into proteins and the incorporation of

nucleotide triphosphates (NTPs) into RNA and DNA. The synthesis of amino acids and NTPs from precursors, however, are not included in the biomass pathways. During growth, proteins, RNA and DNA can be simultaneously synthesized, although RNA and DNA synthesis compete for both available energy and molecular precursors, NTPs and deoxynucleotide triphosphates (dNTPs), respectively. During the cell cycle, DNA is synthesized only during the S phase of the cell cycle. While it is not known whether all bacteria have a cell cycle, *R. rubrum* is known to have a cell cycle. Whether DNA is synthesized depends on the redox state of the cell, specifically, the ratio of NADP to NADPH.

Controlling Concentrations. Because some metabolites have highly favorable free energies of formation or standard chemical potentials, their concentrations can rise to dangerously high levels if not controlled [3]. The primary role of control in a biological system must be to maintain viability. Metabolism will not operate if the concentrations of metabolites becomes so great that the cytoplasm becomes too viscous for diffusion-reaction processes to occur. We previously showed that concentrations of metabolites can be controlled by reducing key reaction fluxes through the activities of the respective enzymes identified by concentration control coefficients from Metabolic Control Analysis. In this approach, described briefly in the Methods section and in detail in reference [7], sensitivity analyses are used to characterize the influence of an enzyme's activity on the concentration of any of the models metabolites. If a metabolite concentration is too high, the enzyme with the most influence over that metabolite is chosen to be regulated. This process occurs until all metabolite concentrations are at physiological levels. The approach can be implemented either as an optimal control problem or using reinforcement learning. Analysis of central metabolism showed that both analyses produced similar results and identified known regulators of central metabolism.

In Table 1, we compare the regulation of enzymes that were identified as necessary to be controlled to maintain physiological levels of metabolites in the MCA predictions with those from the pathway-controlled optimization (PCO) predictions. Of course, in the PCO method, reactions are controlled both to maintain physiological levels and to maximize flux to the growth reactions, so any amount of regulation predicted by the PCO method may result from a combination of the two objectives of maximizing growth while constraining metabolite concentrations to physiological levels. Nevertheless, as shown in Table 1, the PCO predicted regulation of reactions is consistent with the MCA predicted regulation of reactions.

The PCO method regulated nine out of 12 of the same reactions as the MCA method, implying that the PCO method reduced the activity of these reactions in order to maintain physiological levels of the metabolites. However, the observation that the flux in some of these reactions is now effectively zero may indicate that this subset of reactions were additionally shutdown to redirect flux to the growth reactions.

In addition, the PCO method may need to control more reactions than the MCA method, since reaction fluxes are also redistributed due to optimizing

the growth reaction rates (discussed below). Decreased or increased reaction flux may result in higher reactant or product levels, respectively, such that the increased levels need to be controlled to keep them at physiological levels. The PCO approach regulated reactions that are not regulated in MCA and yet have non-zero fluxes. These are shown in Table 2. Presumably, these additional reactions need to be regulated to maintain metabolite concentrations at physiological levels in the new regime.

Increasing Growth. In order to increase flux through the growth reactions in accordance with the PCO objective function, additional reactions are regulated to redirect flux into the growth reactions. The cumulative value of all the reaction fluxes for both the MCA-regulated reactions and the PCO-regulated reactions are shown in Figure 1. Some reactions in the PCO model are indeed reduced and some are increased, with the overall effect being a large increase in absolute rates of some reactions.

The reactions with the highest fluxes while using the PCO approach are shown in Table 3. The reaction with the highest absolute flux in the PCO regulated set of reactions is the RNA synthesis reaction, which increased to 8.41×10^3, a 6.5-fold increase as compared to the MCA regulated model for controlling only metabolite levels. Likewise, the reaction for protein synthesis is included in this group. Both of these reactions are part of the growth reaction set used in the objective function. The other ten reactions with the highest flux are all involved in uptake and processing of the carbon source, ethanol. Much of the carbon from ethanol flows into fumarate and then into the pathways for synthesis of the pyrimidine nucleosides, uridine triphosphate (UTP) and cytidine triphosphate (CTP), precursors of RNA synthesis. In contrast the purine nucleosides are readily available from the already high levels of ATP.

Reactions that were effectively shutdown are shown in Table 4. As can be seen at the bottom of the table, DNA synthesis was minimized by significantly down regulating the synthesis of dNTPs, making more NTPs available for RNA synthesis. In addition, the alternate pathway in the Calvin cycle that consumes erythrose-4-phosphate is shutdown (top three reactions in Table 4), making more erythrose-4-phosphate available for synthesis of vitamin B6, a necessary component for production of the branched chain amino acids tryptophan and tyrosine.

While this study uses a limited model (184 reactions and 204 metabolites), that DNA synthesis was turned off, as it is during most of the cell cycle, is interesting, and presumably was done so in order to maximize RNA synthesis. Further evaluation with a more complete model is needed. It is not clear if under low NADP/NADPH conditions that DNA synthesis is likewise maximized by reducing RNA synthesis.

Of course, it will be critical to test the results of the model against experimental observations in order to draw firm biological conclusions. In this regard, previous studies of the metabolic dynamics of *R. rubrum* using Metabolic Flux Analysis [22], an experimental method to derive fluxes from observations of isotope distributions, will provide a good test of the regulation model.

If successful, the incorporation of thermodynamics into the optimization of growth, along with the ability to infer kinetic parameters from maximum path entropy distributions, will provide a way to further explore the relationship between natural selection and thermodynamics in detail. Ideally, these capabilities will allow the study of the expression patterns and activities of enzymes at various stages in the cell cycle.

3.2 Optimization Discussion

One of the biggest challenges in solving the PCO problem is the range of the variable values. For the solutions presented here both metabolites concentrations and activity coefficients for reactions vary over many orders of magnitude.

To make the problem more numerically tractable our reformulation in essence seeks to do computations with respect to the log of the variables to mitigate the large discrepancies in scale. This is relatively straightforward with respect to the metabolite concentrations in part because they must be positive. However, to handle the flux we must separate out its magnitude and sign which complicates the formulation. It is important to note that our approach does not strictly adhere to the requirements of the IPOPT solver, namely that the objective and constraints be twice continuously differentiable. In particular, both constraint (30d) and (30h) violate this condition.

For the constraint (30d), the original expression (19c) it is derived from is twice continuously differentiable but resulted in numerical errors. By introducing the form (20) and using the approximation (28) to the signum function a cusp is introduced into the first derivative at zero and error is introduced in an interval about the origin determined by the parameters λ and ϵ. We choose λ and ϵ to shrink this interval such that the true derivative is closely approximated at all points evaluated by the solver. We found this choice to produce more accurate solutions and better convergence behavior than using twice continuously differentiable approximations of the signum function or other forms for (19c). It may be possible to further improve performance by supplying the solver directly with alternative derivative computations rather than relying on automatic differentiation, particularly if solutions approach the origin.

The relaxed formulation (30) also uses the approximation (28) to the signum function in the constraint (30h) specifying the value of the 'switching' variable u. Here choices of λ and ϵ are made such that the right hand side of the constraint is effectively constant and partials with respect to the flux will evaluate to zero, in order to hold u constant at 1 or -1 so long as the sign of the flux remains the same. This formulation captures the problem well when the sign on the fluxes are in the optimal direction but it is unclear to what extent it constrains the domain reachable by the solver from an arbitrary initial condition. Information on the direction of the fluxes would allow for significant simplification and additional tractability in this setting. We also evaluated solutions including an explicit computation of the activity coefficients as in (20)

but found that convergence of the solver required setting a lower bound on regulation of roughly 1E-9, and solutions had a significantly lower objective value than were found using the implicit regulation formulation (30).

Due to the potential restrictions on the solutions that can be obtained with our approach using the relaxed formulation (30) it would be useful to consider methods to more thoroughly explore the space of solutions. The formulation (30) can also be viewed as a mixed integer nonlinear programming problem, by considering the switching variables u to be binary. A more rigorous exploration of the solution space could be undertaken with branch and bound type methods, however the problem size can quickly become intractable as the cases to check will grow exponentially in the number of reactions. A more fruitful approach may be to use stochastic search methods such as simulated annealing [5] or particle swarm optimization [6]. If stochastic search methods perform similarly to the approach we present here, then it may have a clear practical advantage in that the times-to-solution will likely be much faster.

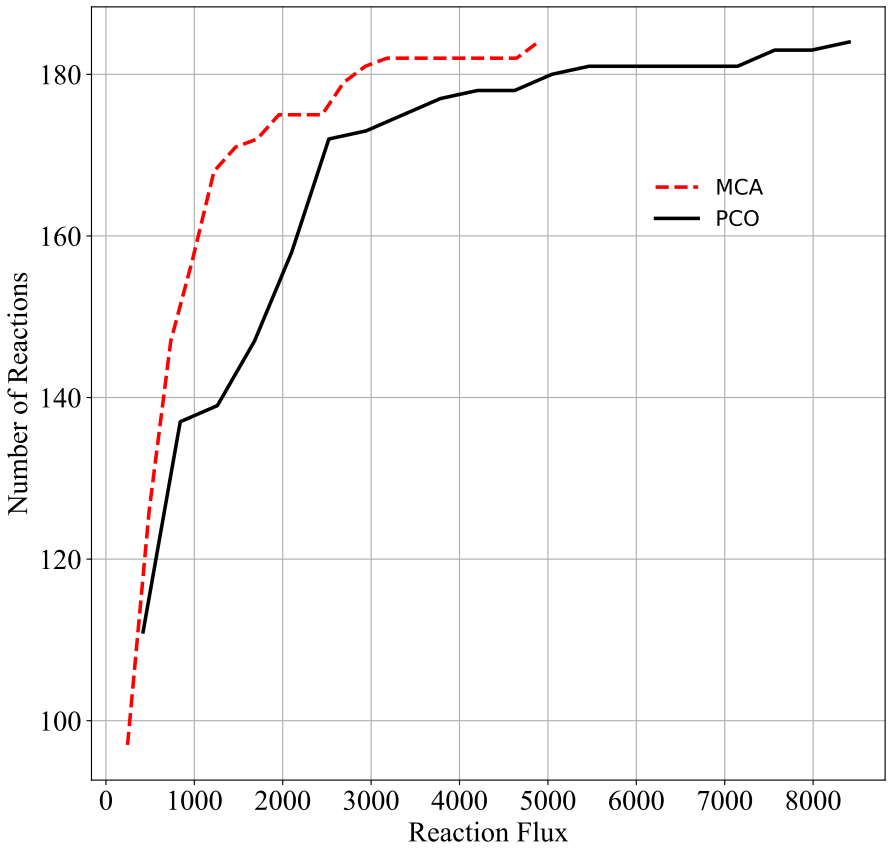

Fig. 1 Comparison of the cumulative distribution of the reaction fluxes for the MCA method and the PCO method. Flux through the growth reactions are maximized by turning reaction fluxes down but not off.

Table 1 Reactions regulated in the MCA approach in which the goal was to control concentrations. Nine out of 12 of these reactions are also regulated in the pathway-controlled optimization approach.

Reaction	J_{MCA}	J_{PCO}	α_{MCA}	α_{OPC}
phosphoenolpyruvate + ADP = pyruvate + ATP	-5.85e+02	-9.96e+02	1.25e-01	1.00e+00
NAD+ + succinate = NADH + fumarate	-1.91e+03	-1.38e-08	1.17e-13	1.70e-29
acetaldehyde + NAD+ + H2O = acetate + NADH	4.32e+02	3.98e+03	1.16e-10	4.21e-13
acetate + ATP + CoA = acetyl-CoA + AMP + diphosphate	1.19e+03	4.87e+03	1.25e-01	1.00e+00
O-acetyl-L-homoserine + H2S = L-homocysteine + acetate	6.01e+01	4.31e-08	4.77e-07	8.91e-25
NH3 + L-glutamate + ATP = L-glutamine + ADP + orthophosphate	2.37e+02	6.57e-09	2.74e-10	1.58e-37
COA + 2.0 AN_oxidized_ferredoxin + pyruvate = acetyl-CoA + CO2 + 2.0 A_reduced_ferredoxin	-2.83e+03	-5.25e+03	1.16e-10	1.00e+00
L-alanine + NAD+ + H2O = NH3 + pyruvate + NADH	-9.76e+01	-1.66e+02	4.88e-04	2.40e-09
ATP + L-aspartate + NH3 = AMP + L-asparagine + diphosphate	6.30e+02	1.66e+02	2.44e-04	2.17e-07
sulfate + ATP = adenosine-5'-phosphosulfate + diphosphate	1.95e+02	3.32e+02	2.47e-32	5.90e-08
ADP + A_reduced_thioredoxin = dADP + AN_oxidized_thioredoxin + H2O	3.49e+02	1.49e-08	1.28e-04	6.61e-18
IMP + NAD+ + H2O = NADH + XMP	6.30e+02	2.10e+03	7.35e-40	5.73e-06

Table 2 Reactions regulated in the PCO method but not in the MCA method and that have significant reaction flux in the PCO method. Presumably these reactions are regulated to control metabolite concentrations.

Reaction	J_{MCA}	J_{OPC}	α_{MCA}	α_{OPC}
L-glutamine + 2-oxoglutarate + NADPH = 2 L-glutamate + NADP+	3.38e+02	-1.33e+03	1.00e+00	4.62e-09
3-phospho-D-glycerate + NAD+ + 3-phosphooxypyruvate + NADH	-3.41e+00	-4.42e+02	1.00e+00	1.64e-02
3-phospho-L-serine + 2-oxoglutarate = L-glutamate + 3-phosphooxypyruvate	3.41e+00	4.42e+02	1.00e+00	9.62e-03
3-phospho-L-serine + H2O = L-serine + orthophosphate	-3.41e+00	-4.42e+02	1.00e+00	1.18e-03
L-serine + L-homocysteine = L-cystathionine + H2O	-3.75e+01	-1.66e+02	1.00e+00	5.23e-02
O-succinyl-L-homoserine + L-cysteine = L-cystathionine + succinate	3.75e+01	1.66e+02	1.00e+00	3.73e-02
L-homoserine + succinyl-CoA = O-succinyl-L-homoserine + CoA	3.75e+01	1.66e+02	1.00e+00	2.89e-02
(2R)-2,3-dihydroxy-3-methylbutanoate = 3-methyl-2-oxobutanoate + H2O	1.95e+02	3.32e+02	1.00e+00	4.75e-02
(2R)-2,3-dihydroxy-3-methylbutanoate + NADP+ = (S)-2-acetolactate + NADPH	-1.95e+02	-3.32e+02	1.00e+00	2.39e-03
2 pyruvate = (S)-2-acetolactate + CO2	1.95e+02	3.32e+02	1.00e+00	2.63e-07
adenosine-5'-phosphosulfate + ATP = 3'-phosphoadenylyl-sulfate + ADP	1.95e+02	3.32e+02	1.00e+00	3.34e-05
3'-phosphoadenylyl-sulfate + A_reduced_thioredoxin = adenosine-3',5'-bisphosphate + sulfite + AN_oxidized_thioredoxin	1.95e+02	3.32e+02	1.00e+00	1.45e-04
H2S + 3 NADP+ + 3 H2O = sulfite + 3 NADPH	-1.95e+02	-3.32e+02	1.00e+00	9.90e-04
ATP + GMP = ADP + GDP	6.30e+02	2.10e+03	1.00e+00	2.12e-04
ATP + L-glutamine + H2O + XMP = AMP + L-glutamate + GMP + diphosphate	6.30e+02	2.10e+03	1.00e+00	5.37e-08

Table 3 Top 12 reactions from the constrained optimization that have the highest fluxes. The reaction that has the highest flux in the constrained optimization is the growth reaction for RNA synthesis. Likewise, the reaction for protein synthesis is included in this group. Other reactions with the highest flux are all involved in uptake and processing of the carbon source, ethanol.

Reaction	J_{MCA}	J_{PCO}	Flux Ratio
oxaloacetate + L-glutamate = L-aspartate + 2-oxoglutarate	-7.31e+02	-2.71e+03	3.71e+00
oxaloacetate = CO2 + pyruvate	-1.66e+03	-3.26e+03	1.96e+00
Amino Acids = protein	1.95e+03	3.32e+03	1.70e+00
GTP + IMP + L-aspartate = adenylo-succinate + GDP + orthophosphate	-1.32e+03	-3.71e+03	2.81e+00
adenylo-succinate = AMP + fumarate	-1.32e+03	-3.71e+03	2.81e+00
acetaldehyde + NAD+ + H2O = acetate + NADH	4.32e+02	3.98e+03	9.21e+00
ethanol + NAD+ = acetaldehyde + NADH	8.23e+02	4.65e+03	5.65e+00
acetate + ATP + CoA = acetyl-CoA + AMP + diphosphate	1.19e+03	4.87e+03	4.09e+00
CoA + 2.0 AN_oxidized_ferredoxin + pyruvate = acetyl-CoA + CO2 + 2.0 A_reduced_ferredoxin	-2.83e+03	-5.25e+03	1.85e+00
(S)-malate = fumarate + H2O	4.89e+03	7.41e+03	1.52e+00
(S)-malate + NADP+ = NADPH + oxaloacetate	-4.89e+03	-7.41e+03	1.52e+00
NTPs = RNA	1.13e+03	8.41e+03	7.46e+00

Table 4 Reactions regulated in the PCO method but not in the MCA method and that have insignificant reaction flux in the PCO method. Presumably these reactions are shut down to redirect mass flow to the growth reactions.

Reaction	J_{MCA}	J_{PCO}	α_{MCA}	α_{PCO}
D-sedoheptulose-7-phosphate + D-glyceraldehyde-3-phosphate = D-ribose-5-phosphate + D-xylulose-5-phosphate	1.84e+01	-3.14e-09	1.00e+00	1.77e-10
D-sedoheptulose-1,7-bisphosphate + H2O = D-sedoheptulose-7-phosphate + orthophosphate	1.84e+01	-3.14e-09	1.00e+00	4.36e-10
glycerone_phosphate + D-erythrose-4-phosphate = D-sedoheptulose-1,7-bisphosphate	1.84e+01	-3.14e-09	1.00e+00	3.72e-10
2-phospho-D-glycerate = phospho-enolpyruvate + H2O	-1.78e-15	-1.53e-12	1.00e+00	1.67e-92
acetyl-CoA + glyoxylate + H2O = COA + (S)-malate	0.00e+00	-2.31e-12	1.00e+00	1.23e-99
S-adenosyl-L-homocysteine + H2O = S-ribosyl-L-homocysteine + adenine	3.68e+00	2.18e-07	1.00e+00	6.26e-08
S-ribosyl-L-homocysteine = L-homocysteine + (4S)-4,5-dihydroxypentan-2,3-dione	3.68e+00	2.18e-07	1.00e+00	6.38e-08
L-glutamine + L-aspartate + ATP + H2O = L-glutamate + L-asparagine + AMP + diphosphate	-5.33e+02	-6.48e-09	5.00e-01	3.72e-26
N-acetyl-L-glutamate + ATP = ADP + N-acetylglutamyl-phosphate	-4.37e-01	-6.88e-09	1.00e+00	3.03e-09
N-acetylglutamyl-phosphate + NADPH = N-acetyl-L-glutamate 5-semialdehyde + NADP+ + orthophosphate	-4.37e-01	-6.88e-09	1.00e+00	3.13e-09
N-acetyl-L-glutamate 5-semialdehyde + L-glutamate = N-acetyl-L-ornithine + 2-oxoglutarate	-4.37e-01	-6.88e-09	1.00e+00	3.08e-09
L-glutamate + N-acetyl-L-ornithine = N-acetyl-L-glutamate + L-ornithine	-4.37e-01	-6.88e-09	1.00e+00	3.13e-09
acetyl-CoA + L-glutamate = N-acetyl-L-glutamate + COA	3.55e-15	5.81e-12	1.00e+00	6.41e-51
ATP + dTMP = ADP + dTDP	3.49e+02	1.49e-08	1.00e+00	3.67e-12
ATP + L-glutamine + UTP + H2O = ADP + CTP + L-glutamate + orthophosphate	-2.56e-01	-5.32e-09	1.00e+00	1.71e-23
ATP + dTDP = ADP + dTTP	3.49e+02	1.49e-08	1.00e+00	5.25e-12
dUTP + H2O = dUMP + diphosphate	3.49e+02	1.49e-08	1.00e+00	2.43e-12
ATP + dADP = ADP + dATP	3.49e+02	1.49e-08	1.00e+00	3.51e-13
GDP + A_reduced_thioredoxin = dGDP + AN_oxidized_thioredoxin + H2O	3.49e+02	1.49e-08	1.00e+00	9.93e-14
ATP + dGDP = ADP + dGTP	3.49e+02	1.49e-08	1.00e+00	1.30e-12
dNTPs = DNA	6.97e+02	2.98e-08	1.00e+00	5.00e-10

Acknowledgements This research was supported by the Data Model Convergence Initiative at Pacific Northwest National Laboratory, under the Laboratory Directed Research and Development Program at PNNL, a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy.

4 Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Aithal, H.N., Walsh-Reitz, M.M., Toback, F.G.: Regulation of glyceraldehyde-3-phosphate dehydrogenase by a cytosolic protein. Am J Physiol **249**(1 Pt 1), C111–6 (1985). DOI 10.1152/ajpcell.1985.249.1.C111. URL https://www.ncbi.nlm.nih.gov/pubmed/4014446
2. Atkinson, D.E.: Limitation of metabolite concentrations and the conservation of solvent capacity in the living cell. In: B.L. Horecker, E.R. Stadtman (eds.) Current Topics in Cellular Regulation, vol. 1, pp. 29–43. Academic Press (1969). DOI 10.1016/B978-0-12-152801-0.50007-9
3. Atkinson, D.E.: Cellular energy metabolism and its regulation. Academic Press, New York (1977)
4. Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., Rabinowitz, J.D.: Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli. Nature chemical biology **5**(8), 593–9 (2009). DOI 10.1038/nchembio.186. URL http://www.ncbi.nlm.nih.gov/pubmed/19561621
5. Bertsimas, D., Tsitsiklis, J.: Simulated Annealing. Statistical Science **8**(1), 10 – 15 (1993). DOI 10.1214/ss/1177011077. URL https://doi.org/10.1214/ss/1177011077
6. Bratton, D., Kennedy, J.: Defining a standard for particle swarm optimization. In: 2007 IEEE Swarm Intelligence Symposium, pp. 120–127 (2007). DOI 10.1109/SIS.2007.368035
7. Britton, S., Alber, M., Cannon, W.R.: Machine learning and optimal control of enzyme activities to preserve solvent capacity in the cell. J. Roy. Soc. Interfaces **17**, 20200656 (2020). DOI 10.1098/rsif.2020.0656
8. Cannon, W., Zucker, J., Baxter, D., Kumar, N., Baker, S., Hurley, J., Dunlap, J.: Prediction of metabolite concentrations, rate constants and post-translational regulation using maximum entropy-based simulations with application to central metabolism of neurospora crassa. Processes **6**(6), 63 (2018). DOI 10.3390/pr6060063. URL https://www.mdpi.com/2227-9717/6/6/63 http://dx.doi.org/10.3390/pr6060063 https://www.mdpi.com/2227-9717/6/6/63/pdf
9. Dewar, R.C.: Maximum entropy production as an inference algorithm that translates physical assumptions into macroscopic predictions: Don't shoot the messenger. Entropy **11**(4), 931–944 (2009). URL ¡Go to ISI¿://WOS:000273028300026
10. Dixit, P.D., Wagoner, J., Weistuch, C., Presse, S., Ghosh, K., Dill, K.A.: Perspective: Maximum caliber is a general variational principle for dynamical systems. Journal of Chemical Physics **148**(1) (2018). URL ¡Go to ISI¿://WOS:000419394500002
11. Hackett, S.R., Zanotelli, V.R.T., Xu, W., Goya, J., Park, J.O., Perlman, D.H., Gibney, P.A., Botstein, D., Storey, J.D., Rabinowitz, J.D.: Systems-level analysis of mechanisms regulating yeast metabolic flux. Science **354**(6311) (2016). DOI 10.1126/science.aaf2786. URL https://science.sciencemag.org/content/354/6311/aaf2786
12. Hallows, W.C., Yu, W., Denu, J.M.: Regulation of glycolytic enzyme phosphoglycerate mutase-1 by sirt1 protein-mediated deacetylation. J Biol Chem **287**(6), 3850–8 (2012). DOI 10.1074/jbc.M111.317404. URL https://www.ncbi.nlm.nih.gov/pubmed/22157007

13. Hart, W.E., Laird, C.D., Watson, J.P., Woodruff, D.L., Hackebeil, G.A., Nicholson, B.L., Siirola, J.D.: Pyomo–optimization modeling in python, vol. 67, second edn. Springer Science & Business Media (2017)
14. Heimlicher, M.B., Bachler, M., Liu, M., Ibeneche-Nnewihe, C., Florin, E.L., Hoenger, A., Brunner, D.: Reversible solidification of fission yeast cytoplasm after prolonged nutrient starvation. J Cell Sci **132**(21) (2019). DOI 10.1242/jcs.231688. URL https://www.ncbi.nlm.nih.gov/pubmed/31558680
15. Hinshelwood, C.N.: On the chemical kinetics of autosynthetic systems. J. Chem. Soc. pp. 745–755 (1952). DOI 10.1039/JR9520000745
16. Holness, M.J., Sugden, M.C.: Regulation of pyruvate dehydrogenase complex activity by reversible phosphorylation. Biochem Soc Trans **31**(Pt 6), 1143–51 (2003). DOI 10.1042/. URL https://www.ncbi.nlm.nih.gov/pubmed/14641014
17. Jaynes, E.T.: Macroscopic Prediction, vol. 31, pp. 254–269. Springer-Verlag, Berlin Heidelberg (1985). DOI 10.1007/978-3-642-70795-7. URL ¡Go to ISI¿://WOS:A19656420500006
18. Jitrapakdee, S., Wallace, J.C.: Structure, function and regulation of pyruvate carboxylase. Biochem J **340 (Pt 1)**, 1–16 (1999). URL https://www.ncbi.nlm.nih.gov/pubmed/10229653
19. Larsson-Raznikiewicz, M.: Kinetic studies on the reaction catalyzed by phosphoglycerate kinase. ii. the kinetic relationships between 3-phosphoglycerate, mgatp2- and activating metal ion. Biochim Biophys Acta **132**(1), 33–40 (1967). URL https://www.ncbi.nlm.nih.gov/pubmed/6030358
20. Lehninger, A.L., Nelson, D.L., Cox, M.M.: Lehninger principles of biochemistry, 4th edn. W.H. Freeman, New York (2005)
21. Marcelin, R.: The mechanics of irreversible phenomenon. Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences **151**, 1052–1055 (1910). URL ¡Go to ISI¿://WOS:000200951000417
22. McCully, A.L., Onyeziri, M.C., LaSarre, B., Gliessman, J.R., McKinlay, J.B.: Reductive tricarboxylic acid cycle enzymes and reductive amino acid synthesis pathways contribute to electron balance in a rhodospirillum rubrum calvin-cycle mutant. Microbiology **166**(2), 199–211 (2020). DOI https://doi.org/10.1099/mic.0.000877. URL https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.000877
23. Moré, J.J.: The levenberg-marquardt algorithm: Implementation and theory. In: G.A. Watson (ed.) Numerical Analysis, pp. 105–116. Springer Berlin Heidelberg, Berlin, Heidelberg (1978)
24. Newsholme, E.A., Start, C.: Regulation in Metabolism. John Wiley & Sons, London (1973)
25. North, J.A., Narrowe, A.B., Xiong, W., Byerly, K.M., Zhao, G., Young, S.J., Murali, S., Wildenthal, J.A., Cannon, W.R., Wrighton, K.C., Hettich, R.L., Tabita, F.R.: A nitrogenase-like enzyme system catalyzes methionine, ethylene, and methane biogenesis. Science **369**(6507), 1094–1098 (2020). DOI 10.1126/science.abb6310. URL https://www.science.org/doi/abs/10.1126/science.abb6310
26. Park, J.O., Rubin, S.A., Xu, Y.F., Amador-Noguez, D., Fan, J., Shlomi, T., Rabinowitz, J.D.: Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. Nat Chem Biol **12**(7), 482–9 (2016). DOI 10.1038/nchembio.2077. URL https://www.ncbi.nlm.nih.gov/pubmed/27159581
27. Parry, B.R., Surovtsev, I.V., Cabeen, M.T., O'Hern, C.S., Dufresne, E.R., Jacobs-Wagner, C.: The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. Cell **156**(1-2), 183–94 (2014). DOI 10.1016/j.cell.2013.11.028. URL http://www.ncbi.nlm.nih.gov/pubmed/24361104
28. Reznik, E., Christodoulou, D., Goldford, J.E., Briars, E., Sauer, U., Segrè, D., Noor, E.: Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity. Cell Reports **20**(11) (2017). DOI 10.1016/j.celrep.2017.08.066
29. Saha, A., Connelly, S., Jiang, J., Zhuang, S., Amador, D.T., Phan, T., Pilz, R.B., Boss, G.R.: Akt phosphorylation and regulation of transketolase is a nodal point for amino acid control of purine synthesis. Mol Cell **55**(2), 264–76 (2014). DOI 10.1016/j.molcel.2014.05.028. URL https://www.ncbi.nlm.nih.gov/pubmed/24981175

30. Sivak, D.A., Thomson, M.: Environmental statistics and optimal regulation. PLoS Comput Biol **10**(9), e1003826 (2014). DOI 10.1371/journal.pcbi.1003826. URL http://www.ncbi.nlm.nih.gov/pubmed/25254493
31. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods **17**, 261–272 (2020). DOI 10.1038/s41592-019-0686-2
32. Wächter, A., Biegler, L.: On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. Mathematical Programming **106**(1), 25–57 (2006)
33. Wang, Y.P., Zhou, L.S., Zhao, Y.Z., Wang, S.W., Chen, L.L., Liu, L.X., Ling, Z.Q., Hu, F.J., Sun, Y.P., Zhang, J.Y., Yang, C., Yang, Y., Xiong, Y., Guan, K.L., Ye, D.: Regulation of g6pd acetylation by sirt2 and kat9 modulates nadph homeostasis and cell survival during oxidative stress. EMBO J **33**(12), 1304–20 (2014). DOI 10.1002/embj.201387224. URL https://www.ncbi.nlm.nih.gov/pubmed/24769394
34. Waygood, E.B., Mort, J.S., Sanwal, B.D.: The control of pyruvate kinase of escherichia coli. binding of substrate and allosteric effectors to the enzyme activated by fructose 1,6-bisphosphate. Biochemistry **15**(2), 277–82 (1976). DOI 10.1021/bi00647a006. URL https://www.ncbi.nlm.nih.gov/pubmed/764863
35. Waygood, E.B., Rayman, M.K., Sanwal, B.D.: The control of pyruvate kinases of escherichia coli. ii. effectors and regulatory properties of the enzyme activated by ribose 5-phosphate. Can J Biochem **53**(4), 444–54 (1975). DOI 10.1139/o75-061. URL https://www.ncbi.nlm.nih.gov/pubmed/236081
36. Waygood, E.B., Sanwal, B.D.: The control of pyruvate kinases of escherichia coli. i. physicochemical and regulatory properties of the enzyme activated by fructose 1,6-diphosphate. J Biol Chem **249**(1), 265–74 (1974). URL https://www.ncbi.nlm.nih.gov/pubmed/4588693