



## OPEN ACCESS

## EDITED BY

Yoram Vodovotz,  
University of Pittsburgh, United States

## REVIEWED BY

Loïc Paulevé,  
UMR5800 Laboratoire Bordelais de  
Recherche en Informatique (LaBRI),  
France

R. Chase Cockrell,  
University of Vermont, United States

## \*CORRESPONDENCE

Ethan King,  
✉ ethan.king@pnnl.gov

## SPECIALTY SECTION

This article was submitted to Multiscale  
Mechanistic Modeling,  
a section of the journal  
Frontiers in Systems Biology

RECEIVED 29 June 2022

ACCEPTED 21 March 2023

PUBLISHED 05 April 2023

## CITATION

King E, Holzer J, North JA and  
Cannon WR (2023), An approach to learn  
regulation to maximize growth and  
entropy production rates in metabolism.  
*Front. Syst. Biol.* 3:981866.  
doi: 10.3389/fsysb.2023.981866

## COPYRIGHT

© 2023 King, Holzer, North and Cannon.  
This is an open-access article distributed  
under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#).  
The use, distribution or reproduction in  
other forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# An approach to learn regulation to maximize growth and entropy production rates in metabolism

Ethan King<sup>1\*</sup>, Jesse Holzer<sup>1</sup>, Justin A. North<sup>2</sup> and  
William R. Cannon<sup>1,3</sup>

<sup>1</sup>Pacific Northwest National Laboratory, Richland, WA, United States, <sup>2</sup>Department of Microbiology, The Ohio State University, Columbus, OH, United States, <sup>3</sup>Interdisciplinary Center for Quantitative Modeling in Biology University of California, Riverside, CA, United States

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 can leverage 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 *Rhodospirillum 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 triphosphates 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

metabolism, regulation, maximum entropy, optimization, modeling

## 1 Introduction

Biological systems can be understood as dissipative systems analogous to tornadoes and hurricanes. Dissipative systems act by taking the most probable path to reduce the energy difference in the environment. These most probable paths result in cyclical patterns of material movement that act to transport the material from regions of high energy to regions of lower energy. In tornadoes, air movement becomes correlated and cyclical in which hot air is driven up and cool air down. Whether one considers entropy from a thermodynamic or information standpoint, the concept is the same—the most probable path allowed by physical constraints is taken. It is in this respect that we expect the dynamics of biological systems act to maximize their entropy production rates to the extent possible. In this regard, biological processes are highly constrained by physical and biological limitations—constraints that are critical for their function. Moreover, for biological systems, the adaptation of their dynamics to maximize their entropy production rates occurs on the time scale of natural selection. This is to say, through natural selection their genome structure and metabolism become updated such that they act to move to the most probable state possible by taking the least action path. On the shorter timescale of a cell's lifetime, those adaptations also include 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 (Hinshelwood, 1952) 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 processes, 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 (Sivak and Thomson, 2014). 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. Those cell phenotypes that operate in the most efficient manner in nutrient poor and dramatically shifting environments will tend towards fixation in the population. 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 (Larsson-Raznikiewicz, 1967; Newsholme and Start, 1973; Waygood and Sanwal, 1974; Waygood et al., 1975; 1976; Aithal et al., 1985; Jitrapakdee and Wallace, 1999; Holness and Sugden, 2003; Lehninger et al., 2005; Hallows et al., 2012; Saha et al., 2014; Wang et al., 2014), 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 (Hackett et al., 2016; Reznik et al., 2017), 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. (2016), 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  $^{13}\text{C}$  isotope experiments and concentrations derived from mass spectrometry and NMR measurements (Hackett et al., 2016). 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 Reznik et al. (2017), 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 *Escherichia coli*, and distilled information on how substrates and inhibitors contribute to metabolic flux regulation (Reznik et al., 2017).

More recently, Britton et al. (2020) 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 build on work by Cannon et al. (2018) that takes advantage of a maximum path-entropy/caliber (Jaynes, 1985; Dewar, 2009; Dixit et al., 2018) formulation of the law of mass action in order to predict likely metabolite concentrations. The advantage of these approaches 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 (Jaynes, 1985; Dewar, 2009; Dixit et al., 2018). 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. (2020), 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 (Britton et al., 2020). As mentioned above, in an environmental niche with finite nutrient resources for growth, for cells to persist they must replicate in a timely and efficient manner in order to compete with other species. Therefore in the context of evolution we expect to see regulation in cells that maximizes growth, and only produces those enzyme catalysts that contribute to growth in a timely and efficient manner.

In this study, we develop an optimization approach that allows for extension of the work by Britton et al. (2020) to be applied for more general constraint based modeling, and demonstrate our method for prediction of regulation that maximizes growth pathways while still constraining metabolite levels to physiological values. While flux balance type analyses optimize a similar objective and may include information about regulation (Covert et al., 2001) or even thermodynamic feasibility constraints (Henry et al., 2007), the method we present is fundamentally different in that we use a thermodynamic perspective to directly infer the most likely distribution of both fluxes and metabolite concentrations subject to given constraints.

We apply our method, that we term pathway-controlled optimization (PCO), to the metabolism and enzymatic activities of *Rhodospirillum rubrum*, a purple non-sulfur photosynthetic bacterium that is being used as a synthetic biology organism for the purpose of ethylene production (North et al., 2020). We compare the regulation and reaction fluxes predicted by this new method to the approach presented in (Britton et al., 2020) that adjust regulation only to maintain metabolite concentrations at physiological levels.

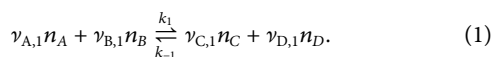
## 2 Methods

We begin this section with a brief introduction to the maximum path entropy approach developed in (Cannon et al., 2018) and

(Britton et al., 2020). Then we present how we extend this approach to account for regulation in cells to optimize growth.

## 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,



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,

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

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,

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

Eq. 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

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

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

where  $K_j = \frac{k_j}{k_{-j}}$  is the equilibrium constant for reaction  $j$ . In Eq. 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 (Marcelin, 1910), 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 Eq. 5 does give a way to the maximum path entropy solution (Cannon et al., 2018; Britton et al., 2020) 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 (Cannon et al., 2018; Britton et al., 2020).

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

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

the flux is then given by

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

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

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

## 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,

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

## 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 (Parry et al., 2014; Heimlicher et al., 2019).

That regulation is needed to control concentrations *in vivo* was proposed early in the field of enzymology (Atkinson, 1969; Atkinson, 1977). In a previously reported method to control concentrations (Britton et al., 2020), 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,

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

where

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

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) (Sauro, 2018). In MCA, the concentration control coefficient  $C_{i,j}$  measures the sensitivity of a concentration  $n_i$  to an activity coefficient  $\alpha_j$ ,

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

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  $\alpha_j$  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. (2020). 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 (Bennett et al., 2009; Park et al., 2016).

## 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 develop a novel approach, we call pathway-controlled optimization (PCO), whose goal is to obtain this biological objective, and derive a formulation that allows for tractable numerical solutions.

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.

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

$$\text{subject to:} \\ \frac{dn_i}{dt} = 0 \quad \forall i \in \mathcal{I}_v, \quad (13b)$$

$$n_i = \bar{n}_i \quad \forall i \in \mathcal{I}_f, \quad (13c)$$

$$0 \leq n_i \leq n_{\max} \quad \forall i \in \mathcal{I}_v, \quad (13d)$$

$$0 \leq \alpha_j \leq 1 \quad \forall j \in \mathcal{J}. \quad (13e)$$

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 non-linear

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  $S_v \in \mathbb{R}^{|\mathcal{I}_v| \times |\mathcal{J}|}$  be the sub-matrix of  $S$  with the rows corresponding to only the variable metabolites. Then at a steady state the reaction fluxes must satisfy

$$S_v J(n, \alpha) = 0, \quad (14)$$

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

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

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,

$$y \in \mathcal{N}(S_v), \\ y_j = J_j(n, \alpha_j) \quad \forall j \in \mathcal{J}. \quad (16)$$

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

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

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

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

where  $(S_j)$  is the  $j$ th 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, conditions (16) are equivalent to.

$$y \in \mathcal{N}(S_v), \quad (19a)$$

$$S^T \eta = \hat{y}(y, \alpha), \quad (19b)$$

$$\hat{y}_j(y, \alpha) = \log\left(\frac{K_j}{2\alpha_j} \left(-y_j + \sqrt{y_j^2 + 4\alpha_j^2}\right)\right) \quad \forall j \in \mathcal{J}. \quad (19c)$$

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 to (19c)

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

with  $\text{sgn}$  the signum function given by

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

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

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

grows unbounded as  $\alpha_j$  and  $y_j$  go to zero. In this application we consistently found solutions with  $\alpha_j$  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. Instead it is sufficient only to identify metabolite concentrations that produce flux values that can be reduced to a steady state. We can capture this criteria with the conditions

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

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  $\eta$  and steady state fluxes  $y$  there exists an  $\alpha$  such that (19) holds if the following hold

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

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

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

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.

$$y \in \mathcal{N}(S_v), \quad (26a)$$

$$g = S^T \eta, \quad (26b)$$

$$h_j = \text{sgn}(y_j) \left( \log(2) - \log(|y_j| + \sqrt{y_j^2 + 4}) \right) \quad \forall j \in \mathcal{J}, \quad (26c)$$

$$|g_j - \log(K_j)| \geq |h_j| \quad \forall j \in \mathcal{J}, \quad (26d)$$

$$(\log(K_j) - g_j)y_j \geq 0 \quad \forall j \in \mathcal{J}. \quad (26e)$$

## 2.4.2 Numerical formulation of the steady state constraints

Implementing 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

$$h_j = \text{sgn}(y_j) \left( \log(2) - \log(|y_j| + \sqrt{y_j^2 + 4}) \right) \quad \forall j \in \mathcal{J}, \quad (27a)$$

$$g_j - \log(K_j) \geq h_j - u_j M \quad \forall j \in \mathcal{J}, \quad (27b)$$

$$g_j - \log(K_j) \leq h_j + (1 - u_j) M \quad \forall j \in \mathcal{J}, \quad (27c)$$

$$2u_j - 1 = \text{sgn}(y_j) \quad \forall j \in \mathcal{J}, \quad (27d)$$

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 use here requires continuous functions of all variables for the constraints. Therefore we approximate the signum function with

$$s\tilde{g}n(x) = \frac{\lambda x}{\lambda|x| + \epsilon} \quad (28)$$

for scaling parameters  $\lambda$  and  $\epsilon$ .

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  $B \in \mathbb{R}^{|\mathcal{J}| \times m}$  be a matrix with columns given by a basis of  $\mathcal{N}(S_v)$ . Then we have that

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

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

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

$$\text{subject to:} \quad (30b)$$

$$y = B\beta, \quad (30c)$$

$$g = S^T \eta, \quad (30d)$$

$$h_j = s\tilde{g}n(y_j) \left( \log(2) - \log(|y_j| + \sqrt{y_j^2 + 4}) \right) \quad \forall j \in \mathcal{J}, \quad (30e)$$

$$g_j - \log(K_j) \geq h_j - u_j M \quad \forall j \in \mathcal{J}, \quad (30f)$$

$$g_j - \log(K_j) \leq h_j + (1 - u_j) M \quad \forall j \in \mathcal{J}, \quad (30g)$$

$$(\log(K_j) - g_j)y_j \geq 0 \quad \forall j \in \mathcal{J}, \quad (30h)$$

$$2u_j - 1 = s\tilde{g}n(y_j) \quad \forall j \in \mathcal{J}, \quad (30i)$$

$$\eta_i \leq \eta_{\max_i} \quad \forall i \in \mathcal{I}_v, \quad (30j)$$

$$\eta_i = \bar{\eta}_i \quad \forall i \in \mathcal{I}_f, \quad (30k)$$

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

**TABLE 1** Reactions regulated in the MCA approach in which the goal was to control concentrations. Ten out of twelve of these reactions are also regulated in the pathway-controlled optimization approach.

Reaction	$J_{MCA}$	$J_{PCO}$	$\alpha_{MCA}$	$\alpha_{PCO}$
phosphoenolpyruvate + ADP = pyruvate + ATP	-5.85e+02	-9.60e+02	1.25e-01	1.00e+00
NAD+ + succinate = NADH + fumarate	-1.91e+03	-1.76e-08	1.17e-13	5.66e-28
acetaldehyde + NAD+ + H2O = acetate + NADH	4.32e+02	3.84e+03	1.16e-10	1.32e-11
acetate + ATP + CoA = acetyl-CoA + AMP + diphosphate	1.19e+03	4.69e+03	1.25e-01	1.00e+00
O-acetyl-L-homoserine + H2S = L-homocysteine + acetate	6.01e+01	4.15e-08	4.77e-07	3.26e-25
NH3 + L-glutamate + ATP = L-glutamine + ADP + orthophosphate	2.37e+02	6.23e-09	2.74e-10	3.98e-43
COA +2.0 AN_oxidized_ferredoxin + pyruvate = acetyl-CoA + CO2 + 2.0 A_reduced_ferredoxin	-2.83e+03	-5.07e+03	1.16e-10	1.24e-01
L-alanine + NAD+ + H2O = NH3 + pyruvate + NADH	-9.76e+01	-1.60e+02	4.88e-04	4.29e-10
ATP + L-aspartate + NH3 = AMP + L-asparagine + diphosphate	6.30e+02	1.60e+02	2.44e-04	2.09e-07
sulfate + ATP = adenosine-5'-phosphosulfate + diphosphate	1.95e+02	3.20e+02	2.47e-32	2.61e-05
ADP + A_reduced_thioredoxin = dADP + AN_oxidized_thioredoxin + H2O	3.49e+02	2.34e-08	1.28e-04	2.55e-16
IMP + NAD+ + H2O = NADH + XMP	6.30e+02	2.03e+03	7.35e-40	2.87e-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	$J_{MCA}$	$J_{PCO}$	$\alpha_{MCA}$	$\alpha_{PCO}$
L-glutamine + 2-oxoglutarate + NADPH = 2 L-glutamate + NADP+	3.38e+02	-1.28e+03	1.00e+00	3.18e-17
3-phospho-D-glycerate + NAD+ + 3-phosphoacetylpyruvate + NADH	-3.41e+00	-4.27e+02	1.00e+00	3.61e-02
3-phospho-L-serine + 2-oxoglutarate = L-glutamate + 3-phosphoacetylpyruvate	3.41e+00	4.27e+02	1.00e+00	2.78e-02
3-phospho-L-serine + H2O = L-serine + orthophosphate	-3.41e+00	-4.27e+02	1.00e+00	1.27e-02
L-serine + L-homocysteine = L-cystathionine + H2O	-3.75e+01	-1.60e+02	1.00e+00	3.07e-02
O-succinyl-L-homoserine + L-cysteine = L-cystathionine + succinate	3.75e+01	1.60e+02	1.00e+00	1.80e-02
L-homoserine + succinyl-CoA = O-succinyl-L-homoserine + CoA	3.75e+01	1.60e+02	1.00e+00	1.56e-02
L-serine + Acetyl-CoA = O-Acetyl-L-serine + CoA	1.35e+02	3.20e+02	1.00e+00	9.25e-02
(2R)-2,3-dihydroxy-3-methylbutanoate = 3-methyl-2-oxobutanoate + H2O	1.95e+02	3.20e+02	1.00e+00	3.93e-02
(2R)-2,3-dihydroxy-3-methylbutanoate + NADP+ = (S)-2-acetolactate + NADPH	-1.95e+02	-3.20e+02	1.00e+00	1.38e-03
2 pyruvate = (S)-2-acetolactate + CO2	1.95e+02	3.20e+02	1.00e+00	1.30e-08
sulfate + ATP adenosine-5'-phosphosulfate + diphosphate	1.95e+02	3.20e+02	1.00e+00	5.82e-08
adenosine-5'-phosphosulfate + ATP = 3'-phosphoadenylyl-sulfate + ADP	1.95e+02	3.20e+02	1.00e+00	2.77e-06
H2S + 3 NADP+ + 3 H2O = sulfite +3 NADPH	-1.95e+02	-3.20e+02	1.00e+00	8.18e-05
oxaloacetate = CO2 + pyruvate	-1.66e+03	-3.15e+03	1.00e+00	1.65e-04
ATP + GMP = ADP + GDP	6.30e+02	2.03e+03	1.00e+00	8.39e-04
ATP + L-glutamine + H2O+ XMP = AMP + L-glutamate + GMP + diphosphate	6.30e+02	2.03e+03	1.00e+00	1.12e-05

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

It is important here to note that when (30) admits solutions they are not necessarily unique. For instance, those reactions that do not directly impact the objective may have a range of choices for the fluxes and constituent metabolite values that satisfy constraints

(30b)–(30h) without changing the objective value. In particular constraints (30e) and (30f) are not necessarily all that restrictive.

#### 2.4.4 Initialization for optimization

The greatest difficulty in solving (30) is in identifying the correct sign for the fluxes. If the correct flux directions are known then the problem simplifies considerably. Using IPOPT, a general non-linear

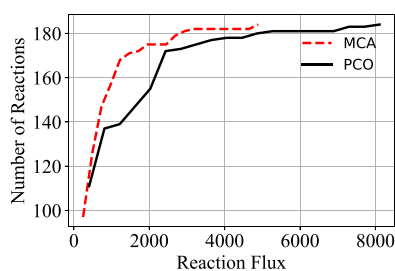


FIGURE 1

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

interior point solver (Wächter and Biegler, 2006), we found that the solution and convergence of the solver was extremely sensitive to the variable initialization, as is common for non-convex problems. In particular, starting with fluxes close to the optimal directions was crucial for convergence to a good solution. To initialize the variables we solve for metabolite values that produces fluxes close to the gradient direction of the objective. 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 first find the projection of the gradient of the objective onto the space of steady state solutions, with

$$y_g = \mathbb{P}_{\mathcal{N}(S_v)} \left( \nabla_y \left[ \sum_{j \in \mathcal{G}} y_j \right] \right) \quad (32)$$

where  $\mathbb{P}_{\mathcal{N}(S_v)}$  is the projection operator onto  $\mathcal{N}(S_v)$ . We then solve the constrained linear least squares problem

$$\min_{\eta} \left\| \text{sgn}(-y_g) \circ (S^T \eta - \log(K)) - s \right\|^2 \quad (33a)$$

$$\text{subject to :} \quad (33b)$$

$$s_j \geq 0, \quad \forall j \in \mathcal{J}, \quad (33b)$$

$$\eta_i \leq \eta_{\max_i} \quad \forall i \in \mathcal{I}_v, \quad (33c)$$

$$\eta_i = \bar{\eta}_i \quad \forall i \in \mathcal{I}_f, \quad (33d)$$

where  $s$  is a vector of positive slack variables and  $\circ$  is the Hadamard product. Consider that if the objective is zero at a solution  $\eta^*$  then  $\text{sgn}(-y_g) = \text{sgn}(S^T \eta^* - \log(K))$  which implies that there exists a  $\gamma > 0$  such that  $(\eta^*, \gamma y_g)$  is a feasible point of (30). Therefore this is a feasible point with fluxes in the directions of  $y_g$ . In practice solutions of (33) may only be close to zero, so we choose initial fluxes  $y^*$  as

$$y^* = -\zeta \mathbb{P}_{\mathcal{N}(S_v)} (S^T \eta^* - \log(K)) \quad (34)$$

for a  $\zeta > 0$ . Larger zeta values move the fluxes further from the origin and was found to improve convergence, likely as it reduces the number of fluxes that will switch sign early in the optimization.

## 2.5 Numerical solution

We solved (33) with the SciPy version 1.7.1 `lsq_linear` routine using the default configuration (Virtanen et al., 2020). A solution of (30) was then computed with IPOPT version 3.14.4 using Pyomo

(Hart et al., 2017) again using the default optimization parameters except with  $\text{tol} = 1.0 \times 10^{-7}$  and maximum iterations set to 10,000.

The parameters for  $\text{sgn}$  in constraint (30d) were  $\lambda = 1.0 \times 10^{50}$  and  $\epsilon = 1.0 \times 10^{-50}$ , while in constraint (30h) they were  $\lambda = 1$  and  $\epsilon = 1 \times 10^{-50}$ . The big  $M$  value was set to 100, and we used  $\zeta = 10$  for the initialization computation.

For all variable metabolites the upper bound was  $\eta_{\max_i} = 13.308$ . Values for the fixed metabolite concentrations, equilibrium constants, and stoichiometric matrix are given in the [Supplementary Material](#). Equilibrium constants were computed using eQuilibrator version 0.3.1 (Beber et al., 2021). Values for the fixed metabolites were taken from mass spectrometry observed concentrations for bacteria (Bennett et al., 2009; Park et al., 2016).

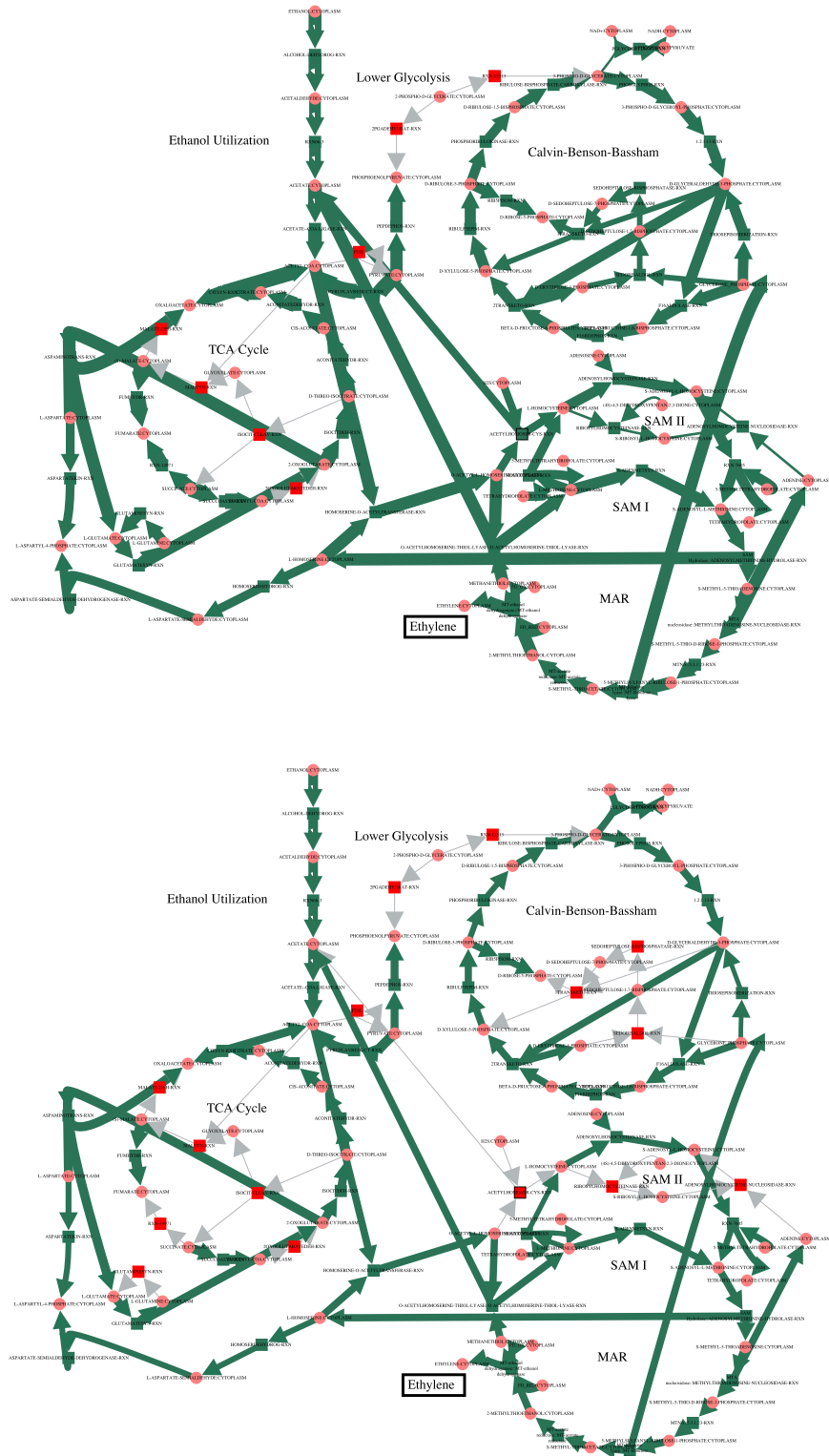
## 2.6 Metabolic model

The metabolic model of *R. rubrum* consisted of 184 reactions and 204 metabolites including central metabolism, the Calvin cycle, the oxidative and reductive TCA cycles, biosynthetic pathways for all amino acids, pyrimidines and purines, the S-adenylsyl methionine (SAM) cycles I and II, the methyl alkane reductase pathway (North et al., 2020), sulfate assimilation, and generic RNA, DNA and protein synthesis pathways. The model represents an engineered species that is predicted to not need the ethylmalonyl CoA pathway due to the inclusion of an oxaloacetate decarboxylase/malic enzyme that converts oxaloacetate to pyruvate and  $\text{CO}_2$  and the inactivation of phospho-glycerate mutase. Consequently, all 3-phosphoglycerate produced by the Calvin Cycle is stays within the Calvin cycle.

Boundary conditions (fixed species concentrations) consisted of concentrations of nutrients (ethanol, diphosphate, orthophosphate, sulfate and ammonia), cell waste products ( $\text{CO}_2$ ,  $\text{H}_2$ , ethylene), ATP derivatives to set up the energy gradient due to photoheterotrophic growth (ATP, ADP, AMP), internal redox pairs ( $\text{NAD}^+$ , NADH, NADP<sup>+</sup>, NADPH, reduced and oxidized ferredoxin, reduced and oxidized thioredoxin), folates (N10-formyltetrahydrofolate, 5-methyltetrahydrofolate, 5,10-methylenetetrahydrofolate, 7,8-dihydrofolate and tetrahydrofolate), and other internal metabolites for which scavenging pathways were not modeled (coenzyme A, (4S)-4,5-dihydroxypentan-2,3-dione, adenine, adenosine, adenosine-3',5'-biphosphate).

The standard free energy of reaction for each reaction was calculated with eQuilibrator, version 0.3.1. The free energies for the methyl alkane reductase pathway, the SAM cycles I & II, the ferredoxin oxidoreductases for 2-oxoglutarate and pyruvate, DNA, RNA and protein polymerization reactions were not available from eQuilibrator and were estimated manually from half-reactions reactions or overall reactions for the respective pathways. If estimates for some individual reactions were available for these pathways, they were adjusted accordingly, otherwise the standard free energy change for the pathways was divided equally among the constituent reactions. Details are included in the computational notebook available as supplemental material.

The model was constructed from a custom *R. rubrum* BioCyc database at PNNL. The draft model is available as a Jupyter notebook along with the necessary auxiliary files for running the notebook in King and Cannon (2023).



**FIGURE 2**

Comparison of the fluxes through central metabolism for the MCA model (top) and the PCO model (bottom). Flux through the growth reactions are maximized by turning reaction fluxes down but not off. The decreased flux through central metabolism in the PCO model reflects the same decrease shown below the red line for the MCA model in [Figure 1](#).



**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	-2.61e+03	3.57e+00
oxaloacetate = CO <sub>2</sub> + pyruvate	-1.66e+03	-3.15e+03	1.89e+00
Amino Acids = protein	1.95e+03	3.20e+03	1.64e+00
GTP + IMP + L-aspartate = adenylo-succinate + GDP + orthophosphate	-1.32e+03	-3.57e+03	2.72e+00
adenylo-succinate = AMP + fumarate	-1.32e+03	-3.57e+03	2.72e+00
acetaldehyde + NAD <sup>+</sup> + H <sub>2</sub> O = acetate + NADH	4.32e+02	3.84e+03	8.88e+00
ethanol + NAD <sup>+</sup> = acetaldehyde + NADH	8.23e+02	4.48e+03	5.45e+00
acetate + ATP + CoA = acetyl-CoA + AMP + diphosphate	1.19e+03	4.69e+03	3.94e+00
CoA + 2.0 AN_oxidized_ferredoxin + pyruvate = acetyl-CoA + CO <sub>2</sub> + 2.0 A_reduced_ferredoxin	-2.83e+03	-5.07e+03	1.79e+00
(S)-malate = fumarate + H <sub>2</sub> O	4.89e+03	7.15e+03	1.46e+00
(S)-malate + NADP <sup>+</sup> = NADPH + oxaloacetate	-4.89e+03	-7.15e+03	1.46e+00
NTPs = RNA	1.13e+03	8.11e+03	7.20e+00

### 3 Results

In order to place the results in the proper context, a brief overview of the metabolism of *R. rubrum*, the subject of the biomass optimization, is necessary. *R. rubrum* is capable of growing both photoautotrophically, meaning that it only uses CO<sub>2</sub> 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<sub>2</sub>. 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 two routes in the Calvin cycle differ in 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 B<sub>6</sub>, pyridoxal-5'-phosphate, an essential cofactor in amino acid, and hence protein, synthesis. Therefore, 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. *R. rubrum* is known to have a cell cycle in which DNA can only be synthesized during the S phase, while RNA is synthesized continuously throughout the cell cycle except during

mitosis. Whether DNA is synthesized depends on the redox state of the cell, specifically, the ratio of NADP to NADPH.

#### 3.1 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 (Atkinson, 1977). 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 (Britton et al., 2020), 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

**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	$J_{MCA}$	$J_{PCO}$	$\alpha_{MCA}$	$\alpha_{PCO}$
D-sedoheptulose-7-phosphate + D-glyceraldehyde-3-phosphate = D-ribose-5-phosphate + D-xylulose-5-phosphate	1.84e+01	-2.90e-09	1.00e+00	1.19e-10
D-sedoheptulose-1,7-bisphosphate + H <sub>2</sub> O = D-sedoheptulose-7-phosphate + orthophosphate	1.84e+01	-2.90e-09	1.00e+00	1.24e-10
glycerone_phosphate + D-erythrose-4-phosphate = D-sedoheptulose-1,7-bisphosphate	1.84e+01	-2.90e-09	1.00e+00	1.30e-10
2-phospho-D-glycerate = phosphoenolpyruvate + H <sub>2</sub> O	-1.78e-15	2.16e-13	1.00e+00	1.30e-15
acetyl-CoA + glyoxylate + H <sub>2</sub> O = COA + (S)-malate	3.55e-15	1.56e-12	1.00e+00	6.97e-07
S-adenosyl-L-homocysteine + H <sub>2</sub> O = S-ribosyl-L-homocysteine + adenine	3.68e+00	2.09e-07	1.00e+00	5.93e-08
S-ribosyl-L-homocysteine = L-homocysteine + (4S)-4,5-dihydropyridin-2,3-dione	3.68e+00	2.09e-07	1.00e+00	6.05e-08
L-glutamine + L-aspartate + ATP + H <sub>2</sub> O = L-glutamate + L-asparagine + AMP + diphosphate	-5.32e+02	-6.16e-09	5.00e-01	9.45e-32
acetyl-CoA + L-glutamate = N-acetyl-L-glutamate + COA	3.55e-15	8.28e-12	1.00e+00	6.68e-44
ATP + dTMP = ADP + dTDP	3.49e+02	2.34e-08	1.00e+00	1.62e-10
ATP + L-glutamine + UTP + H <sub>2</sub> O = ADP + CTP + L-glutamate + orthophosphate	-2.56e-01	-5.10e-09	1.00e+00	4.12e-29
CTP + H <sub>2</sub> O = CDP + orthophosphate	3.32e-02	2.34e-08	1.00e+00	8.97e-10
CDP + A_reduced_thioredoxin = dCDP + AN_oxidized_thioredoxin + H <sub>2</sub> O	3.32e-02	2.34e-08	1.00e+00	9.26e-10
ATP + dCDP = ADP + dCTP	3.32e-02	2.34e-08	1.00e+00	9.68e-10
ATP + dTDP = ADP + dTTP	3.49e+02	2.34e-08	1.00e+00	2.06e-10
UDP + A_reduced_thioredoxin = dUDP + AN_oxidized_thioredoxin + H <sub>2</sub> O	6.97e+02	2.34 e-08	1.00e+00	8.51e-15
ATP + dUDP = ADP + dUTP	6.97e+02	2.34 e-08	1.00e+00	2.35e-14
dUTP + H <sub>2</sub> O = dUMP + diphosphate	3.49e+02	2.34e-08	1.00e+00	1.13e-10
ATP + dADP = ADP + dATP	3.49e+02	2.34e-08	1.00e+00	1.13e-11
GDP + A_reduced_thioredoxin = dGDP + AN_oxidized_thioredoxin + H <sub>2</sub> O	3.49e+02	2.34e-08	1.00e+00	4.33e-12
ATP + dGDP = ADP + dGTP	3.49e+02	2.34e-08	1.00e+00	4.80e-11
dNTPs = DNA	6.97e+02	4.68e-08	1.00e+00	4.88e-09

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 ten out of twelve of the same reactions as the MCA method, suggesting 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.

### 3.2 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 to that end. 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 a subset of reactions. In general, more flux is directed to biosynthetic reactions, with the net result being a redirection of flux away from central

metabolism in comparing the MCA-derived flux distribution to the PCO-derived flux distribution, as indicated in the flux maps for the two models in [Figure 2](#).

The reactions with the highest fluxes from 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  $8.11 \times 10^3$ , a more than 7-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.

## 4 Discussion

### 4.1 Biological discussion

The PCO method we present here agrees with the majority of the regulation imposed by the MCA method for maintaining solvent capacity of the cell, which was shown previously to agree with known points of regulation from experiments as reported in the literature [Britton et al. \(2020\)](#). However, the consideration of growth as an additional objective for regulation in our PCO approach produced significant additional regulation in line with adjusting flux through growth pathways. The inclusion of the consideration of growth is likely to be crucial for a complete understanding of regulation in cells.

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 ([McCully et al., 2020](#)), 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 with the PCO approach developed here,

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.

### 4.2 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 formulation 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 (27a) and (30h) violate this condition.

For the constraint (27a), 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  $\lambda$  and  $\epsilon$ . We choose  $\lambda$  and  $\epsilon$  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  $\lambda$  and  $\epsilon$  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. 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  $1 \times 10^{-9}$ , 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) we will consider methods to more thoroughly explore the feasible set in future work including stochastic search methods. Of course a more complete exploration of the space of solutions can come at significant computational cost and if other 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. Additionally, as discussed in Section 2.4.3 solutions are not necessarily unique. Methods to explore and characterize the set of equivalent solutions, that is solutions that can achieve the same objective value, will allow for a greater understanding of the flexibility that may be possible in controlling reaction fluxes and metabolite concentrations while achieving high growth rates.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## Author contributions

WC conceived and designed the maximum entropy approach to model metabolism and developed the *R. rubrum* metabolic model. EK designed and implemented the optimization method for learning regulation to maximize growth. JH contributed to the optimization development. JN contributed to the development of the metabolic model of *Rhodospirillum rubrum*. The manuscript was written by EK and WC.

## Funding

WC and JN were supported by the U.S. Department of Energy, Office of Biological and Environmental Research through contracts 78266 and award DE-SC0022091, respectively. EK and JH were supported by the Data Model Convergence Initiative, and EK additionally by the Predictive Phenomics Initiative, under the

## References

- Aithal, H. N., Walsh-Reitz, M. M., and Toback, F. G. (1985). Regulation of glyceraldehyde-3-phosphate dehydrogenase by a cytosolic protein. *Am. J. Physiol.* 249, C111–C116. doi:10.1152/ajpcell.1985.249.1.C111
- Atkinson, D. E. (1977). *Cellular energy metabolism and its regulation*. New York: Academic Press.
- Atkinson, D. E. (1969). "Limitation of metabolite concentrations and the conservation of solvent capacity in the living cell," in *Current topics in cellular regulation*. Editors B. L. Horecker and E. R. Stadtman (Academic Press), 1, 29–43. doi:10.1016/B978-0-12-152801-0.50007-9
- Beber, M. E., Gollub, M. G., Mozaffari, D., Shebek, K. M., Flamholz, A., Milo, R., et al. (2021). eQuilibrator 3.0: a database solution for thermodynamic constant estimation. *Nucleic Acids Res.* 50, D603–D609. doi:10.1093/nar/gkab1106
- Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* 5, 593–599. doi:10.1038/nchembio.186
- Britton, S., Alber, M., and Cannon, W. R. (2020). Enzyme activities predicted by metabolite concentrations and solvent capacity in the cell. *J. Roy. Soc. Interfaces* 17, 20200656. doi:10.1098/rsif.2020.0656
- Cannon, W., Zucker, J., Baxter, D., Kumar, N., Baker, S., Hurley, J., et al. (2018). 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, 63. doi:10.3390/pr6060063
- Covert, M. W., Schilling, C. H., and Palsson, B. (2001). Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* 213, 73–88. doi:10.1006/jtbi.2001.2405
- Dewar, R. C. (2009). Maximum entropy production as an inference algorithm that translates physical assumptions into macroscopic predictions: Don't shoot the messenger. *Entropy* 11, 931–944. doi:10.3390/e11040931
- Dixit, P. D., Wagoner, J., Weistuch, C., Presse, S., Ghosh, K., and Dill, K. A. (2018). Perspective: Maximum caliber is a general variational principle for dynamical systems. *J. Chem. Phys.* 148, 010901. doi:10.1063/1.5012990
- Hackett, S. R., Zanotelli, V. R. T., Xu, W., Goya, J., Park, J. O., Perlman, D. H., et al. (2016). Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* 354, aaf2786. doi:10.1126/science.aaf2786
- Hallows, W. C., Yu, W., and Denu, J. M. (2012). Regulation of glycolytic enzyme phosphoglycerate mutase-1 by sirt1 protein-mediated deacetylation. *J. Biol. Chem.* 287, 3850–3858. doi:10.1074/jbc.M111.317404
- Hart, W. E., Laird, C. D., Watson, J.-P., Woodruff, D. L., Hackebeitl, G. A., Nicholson, B. L., et al. (2017). *Pyomo-optimization modeling in python* second edn, 67. Springer Science & Business Media.
- Heimlicher, M. B., Bachler, M., Liu, M., Ibeneche-Nnewiwe, C., Florin, E. L., Hoenger, A., et al. (2019). Reversible solidification of fission yeast cytoplasm after prolonged nutrient starvation. *J. Cell Sci.* 132, jcs231688. doi:10.1242/jcs.231688
- Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2007). Thermodynamics-based metabolic flux analysis. *Biophysical J.* 92, 1792–1805. doi:10.1529/biophysj.106.093138
- Hinshelwood, C. N. (1952). 136. On the chemical kinetics of autotrophic systems. *J. Chem. Soc.* 745, 745. doi:10.1039/JR9520000745

Laboratory Directed Research and Development Program at the Pacific Northwest National Laboratory. PNNL is a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy under contract DE-AC05-76RLO 1830.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fsysb.2023.981866/full#supplementary-material>

**SUPPLEMENTARY DATA SHEET S1**  
Equilibrium constants.

**SUPPLEMENTARY DATA SHEET S2**  
Metabolite concentrations.

**SUPPLEMENTARY DATA SHEET S3**  
Stoichiometric Matrix.

- Holness, M. J., and Sugden, M. C. (2003). Regulation of pyruvate dehydrogenase complex activity by reversible phosphorylation. *Biochem. Soc. Trans.* 31, 1143–1151. doi:10.1042/bst0311143
- Jaynes, E. T. (1985). *Macroscopic prediction*. Berlin Heidelberg: Springer-Verlag, 254–269. doi:10.1007/978-3-642-70795-7
- Jitrapakdee, S., and Wallace, J. C. (1999). Structure, function and regulation of pyruvate carboxylase. *Biochem. J.* 340 (1), 1–16. doi:10.1042/bj3400001
- King, E., and Cannon, W. (2023). Pathway controlled optimization. Available at: [https://github.com/pnnl/pathway\\_controlled\\_optimization](https://github.com/pnnl/pathway_controlled_optimization).
- Larsson-Raznikiewicz, M. (1967). 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, 33–40. doi:10.1016/0005-2744(67)90189-1
- Lehninger, A. L., Nelson, D. L., and Cox, M. M. (2005). *Lehninger principles of biochemistry* 4th edn. New York: W. H. Freeman.
- Marcelin, R. (1910). The mechanics of irreversible phenomenon. *Comptes Rendus Hebd. Des. Seances De. L Acad. Des. Sci.* 151, 1052–1055.
- McCully, A. L., Onyeziri, M. C., LaSarre, B., Gliessman, J. R., and McKinlay, J. B. (2020). Reductive tricarboxylic acid cycle enzymes and reductive amino acid synthesis pathways contribute to electron balance in a rhodospirillum rubrum calvin-cycle mutant. *Microbiology* 166, 199–211. doi:10.1099/mic.0.000877
- Newsholme, E. A., and Start, C. (1973). *Regulation in metabolism*. London: John Wiley & Sons.
- North, J. A., Narrowe, A. B., Xiong, W., Byerly, K. M., Zhao, G., Young, S. J., et al. (2020). A nitrogenase-like enzyme system catalyzes methionine, ethylene, and methane biogenesis. *Science* 369, 1094–1098. doi:10.1126/science.abb6310
- Park, J. O., Rubin, S. A., Xu, Y. F., Amador-Noguez, D., Fan, J., Shlomi, T., et al. (2016). Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat. Chem. Biol.* 12, 482–489. doi:10.1038/nchembio.2077
- Parry, B. R., Surovtsev, I. V., Cabeen, M. T., O'Hern, C. S., Dufresne, E. R., and Jacobs-Wagner, C. (2014). The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. *Cell* 156, 183–194. doi:10.1016/j.cell.2013.11.028
- Reznik, E., Christodoulou, D., Goldford, J. E., Briars, E., Sauer, U., Segrè, D., et al. (2017). Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity. *Cell Rep.* 20, 2666–2677. doi:10.1016/j.celrep.2017.08.066
- Saha, A., Connelly, S., Jiang, J., Zhuang, S., Amador, D. T., Phan, T., et al. (2014). Akt phosphorylation and regulation of transketolase is a nodal point for amino acid control of purine synthesis. *Mol. Cell* 55, 264–276. doi:10.1016/j.molcel.2014.05.028
- Sauro, H. M. (2018). *Systems biology: An introduction to metabolic control Analysis*. Ambrosius Publishing.
- Sivak, D. A., and Thomson, M. (2014). Environmental statistics and optimal regulation. *PLoS Comput. Biol.* 10, e1003826. doi:10.1371/journal.pcbi.1003826
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi:10.1038/s41592-019-0686-2
- Wächter, A., and Biegler, L. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.* 106, 25–57. doi:10.1007/s10107-004-0559-y
- Wang, Y. P., Zhou, L. S., Zhao, Y. Z., Wang, S. W., Chen, L. L., Liu, L. X., et al. (2014). Regulation of g6pd acetylation by sirt2 and kat9 modulates nadph homeostasis and cell survival during oxidative stress. *EMBO J.* 33, 1304–1320. doi:10.1002/embj.201387224
- Waygood, E. B., Mort, J. S., and Sanwal, B. D. (1976). 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, 277–282. doi:10.1021/bi00647a006
- Waygood, E. B., Rayman, M. K., and Sanwal, B. D. (1975). 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, 444–454. doi:10.1139/o75-061
- Waygood, E. B., and Sanwal, B. D. (1974). The control of pyruvate kinases of *Escherichia coli*. *J. Biol. Chem.* 249, 265–274. doi:10.1016/s0021-9258(19)43120-7