Optimal enzyme profiles in unbranched metabolic pathways

Elad Noor[1*] and Wolfram Liebermeister[2*]

[1] Department of Plant and Environmental Sciences, Weizmann Institute of Science, 76100 Rehovot, Israel
[2] Université Paris-Saclay, INRAE, MaIAGE, 78350 Jouy-en-Josas, France
∗ Corresponding authors, equal contributions

Abstract

How to optimize the allocation of enzymes in metabolic pathways has been a topic of study for many decades. Although the general problem is complex and non-linear, we have previously shown that it can be solved by convex optimization. In this paper, we focus on unbranched metabolic pathways with simplified enzymatic rate laws and derive analytic solutions to the optimization problem. We revisit existing solutions based on the limit of mass-action rate laws and present new solutions for other rate laws. Furthermore, we revisit a known relationship between flux control coefficients and enzyme abundances in optimal metabolic states. We generalize this relationship to models with density constrains on enzymes and metabolites, and present a new local relationship between optimal reaction elasticities and enzyme amounts. Finally, we apply our theory to derive simple kinetics-based formulae for protein allocation during bacterial growth.

Keywords
Metabolic pathway, enzymatic rate law, enzyme demand, optimization, protein allocation, bacterial growth law

Abbreviations
ECM – Enzyme Cost Minimization
PSA – Pathway Specific Activity
MDF – Max-min Driving Force method

1 Introduction

The idea that living beings show optimal shapes or behavior has a very long history. A process like evolution, which combines random mutations with a selection for favorable properties, could potentially lead to optimization, but the question of *if* and/or *when* should we expect living beings to function optimally has been widely debated and is far from solved. In practice, it can be useful to invoke optimality principles to seek insights and design principles that might be relevant also in naturally evolved systems [1]. Specifically, cell metabolism has often been studied using this approach [2, 3], thanks to the powerful mathematical models that we have to describe it. But although natural selection has been the main inspiration for this study, the evolutionary aspects of pathway optimization are not discussed here, and are rather left for the reader to reflect upon.

Within cells, protein is arguably the most important and central resource, both in terms of contributing to fitness, but also since protein synthesis requires large amounts of energy, metabolic precursors, and ribosomes, and the proteins themselves occupy a significant portion of cellular space. Therefore, a cell should generally save protein wherever it can. This notion, specifically for enzymes, has been mathematically applied in genome-scale metabolic models [4, 5], models of core metabolism [6], and in direct comparisons between pathways [7, 8]. Since the proteome is a limiting resource, its allocation to different sectors (metabolism, translation, etc.) is a topic of high interest. In some works people assumed optimality. Other papers assumed a general rule based on linear growth rate control. Interestingly, even a very simple linear chain model with two reactions, representing metabolism and protein synthesis, and a bound on the total protein budget has been successful in explaining bacterial growth laws and overflow metabolism [9, 10]. However, these bacterial growth law models did not consider enzyme kinetics.

Here, we focus on a special case of this cost/benefit analysis: the efficient use of metabolic enzymes in unbranched pathways operating at steady-state, giving priority to scenarios that can be solved analytically. We explore several types of kinetic rate laws and introduce the idea of bounding the total metabolite concentration (which is required in some cases for meaningful results). To define states of maximal enzyme efficiency, we can consider two equivalent optimality problems: maximizing a production flux at a given enzyme budget or minimizing protein usage at a given required production flux. In both cases, we maximize the production flux per enzyme usage within the given constraints. If the product of the pathway is directly tied to biomass, the overall enzyme efficiency, called "biomass/enzyme efficiency", can serve as a proxy for cell growth [6]. Furthermore, this optimality problem is also relevant in other contexts, such as metabolic engineering of synthetic pathways using a set of existing and/or new-to-nature enzymes with known kinetic parameters [11, 12].

Another perspective often used to analyze metabolic systems is through their control, e.g. the effect of changes in a level of an enzyme on pathway flux [13]. In optimal metabolic states with a bound on the sum of enzyme levels, each enzyme effectively carries an opportunity cost. This cost must be balanced by a marginal benefit, given by the flux control coefficient – as defined in Metabolic Control Analysis (MCA) [14, 15, 13]). Hence, for systems in optimal states, there are simple relationships between enzyme abundance and flux control [16, 17, 18]. We will recapitulate these results below, generalise them to models with a density constraint on enzyme and metabolite levels and illustrate them using some of our analytic solutions. In addition, we present a simple rule that links optimal enzyme investments around a given metabolite to the reaction elasticities of this metabolite.

Finally, we show how the analytic solutions derived here might be useful for modeling high-level phenomena such as the Monod curve (i.e. the relationship between the concentration of a limiting nutrient and the growth rate of bacteria [19]). We use this model to demonstrate how each kinetic parameter should affect the growth rate under different conditions.

In summary, this paper revisits the question of optimal enzyme allocation and adds to previous results. We focus on unbranched metabolic pathways, extend the optimality problem, discuss new optimality conditions, and present analytic solutions that directly show how different factors determine optimal enzyme levels and fluxes. We discuss general principles, in particular how optimal enzyme levels reflect flux control and local reaction elasticities, and use our theory to derive formulae for kinetics-based bacterial growth.

2 Results

2.1 Optimal states of unbranched pathways

One of the first attempts at analytically solving the enzyme allocation problem was published by Waley [20], who studied short pathways of 2-3 reactions while assuming the concentrations of metabolites (which were denoted

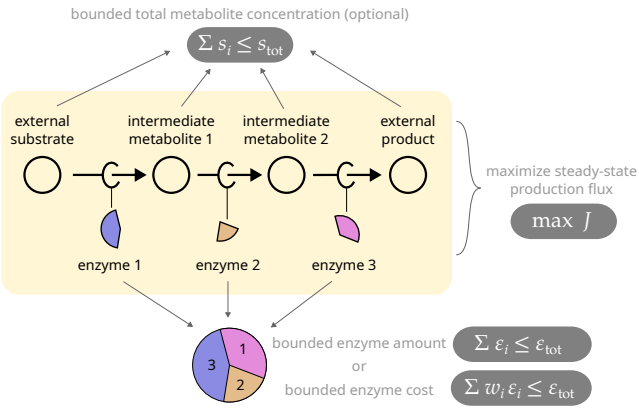


Figure 1: Optimal enzyme levels in unbranched metabolic pathways. In the basic optimality problem in this paper, we consider a chain of reactions and ask how a given protein budget should be spent on metabolic enzymes to achieve a maximal steady-state flux. If all the efficiencies (i.e. the reaction rates per catalyzing enzyme) of all the enzymes were known, the steady-state flux would determine the enzyme levels, and there would be nothing to optimize. Here we assume that the enzyme efficiencies can be adjusted by choosing the metabolite concentrations (not shown), and we search for the optimal metabolite and enzyme profile. The aim is to maximize the flux at a given total enzyme amount (bottom) and possibly under a constraint on metabolite concentrations (top). The flux ratios between reactions are predefined, for example assuming equal fluxes in all the reactions (right).

linking intermediates) are much below the enzymes' K_{M} values, and therefore affect the flux linearly. Given the total amount of catalytically active protein (bounded by ε_{tot}), the relative enzyme concentrations should be such that they maximize rate (see Figure 1). Based on these assumptions, one can derive simple formulae for the optimal enzyme levels and maximal pathway flux. Later studies repeated this result and generalized it to linear pathways of any size [21, 17, 18]. Here, we will revisit this general solution and extend it to other types of rate laws beyond the one considered by Waley [20] (which, from now on, we will refer to as *mass-action*).

Consider the following unbranched pathway [22] (Figure 1):

$$\mathrm{S}_0 \overset{\mathrm{v}_1}{\rightleftharpoons} \mathrm{S}_1 \overset{\mathrm{v}_2}{\rightleftharpoons} \ldots \overset{\mathrm{v}_n}{\rightleftharpoons} \mathrm{S}_n \tag{1}$$

In a kinetic model, each variable s_i represents the concentration of a metabolite i and each variable ε_i represents the level (molar concentration or mass concentration) of the enzyme catalyzing reaction i. Imagine that the total enzyme level in the pathway is bounded by ε_{tot}, i.e.

$$\sum_i \varepsilon_i \leq \varepsilon_{\text{tot}} \,. \tag{2}$$

What would be the optimal strategy for distributing this resource between the reactions in order to maximize the steady-state flux? To answer this question, we need to know how the rate of each reaction depends on the levels of enzyme, substrate, product, as well as kinetic parameters. This is described by rate laws. For some rate laws, we can solve the optimization problem and obtain an analytic solution that describes exactly how much of each enzyme should be allocated. Below, we will also consider a variant of this problem with an extra bound on the sum of metabolite levels or with fixed initial and final metabolites (s_0 and s_n).

Since single analytic solutions are rare but instructive, we explore them in this article. We consider four different rate laws (summarized in Figure 2): the general Haldane rate law (saturable and reversible) which has no analytic solution, and three solvable approximations derived from it. As explained above, the *enzyme levels* ε_i may either refer to molar concentrations or mass concentrations (depending on the modeler's preference). In the case of molar concentrations, the k^{cat} values are catalytic constant (e.g. in units of 1/s). In the case of mass concentrations, the k^{cat} values are specific activities (e.g. in units of μmol × min^{-1} × mg^{-1}).

- **Reversible saturable rate law ("Haldane")** As a general rate law for a reaction S ↔ P, we consider the reversible saturable rate law

$$v = \varepsilon \, \frac{k_+^{\text{cat}} \; s/K_{\text{S}} - k_-^{\text{cat}} \; p/K_{\text{P}}}{1 + s/K_{\text{S}} + p/K_{\text{P}}} \tag{3}$$

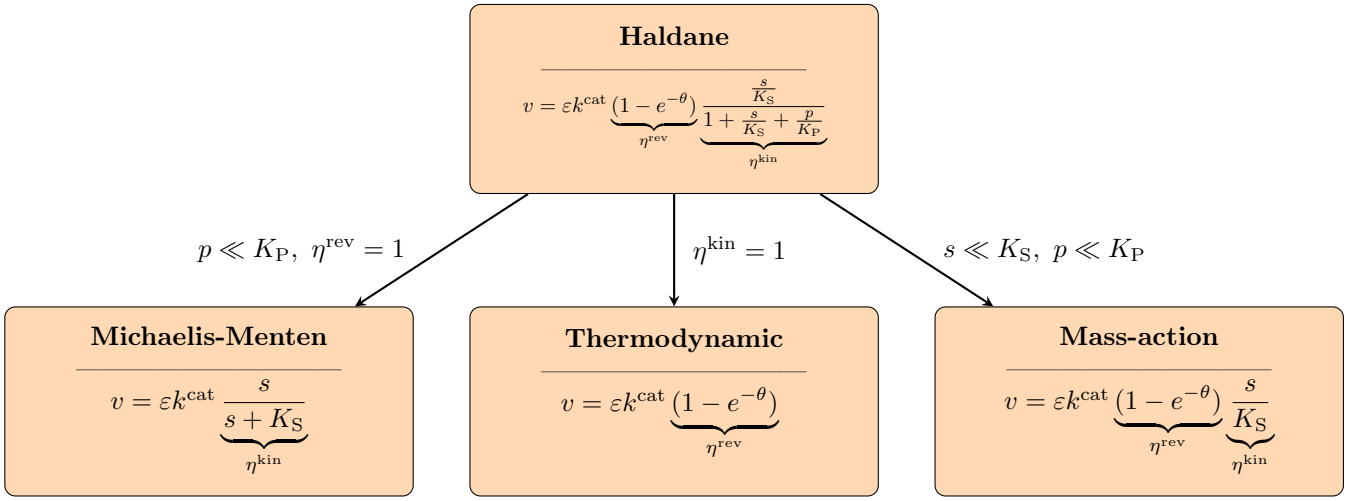


Figure 2: The Haldane rate law and some simplified rate laws. Simplified rate laws are obtained as limiting cases by setting efficiency terms η^{rev} or η^{kin} to 1 or another constant value (see Figure 3) or by assuming that reactant concentrations are small (in the case of the mass-action rate law). For the enzyme optimality problem in unbranched pathways, we do not know of any analytic solutions for the Haldane rate law. We report here solutions for the other, simplified rate laws (where the solution for the thermodynamic rate law contains an unknown auxiliary parameter).

with Michaelis-Mention constants K_S and K_P, which can be factorized into

$$v = \varepsilon\ k_+^{\text{cat}}\ \underbrace{(1-e^{-\theta})}_{\eta^{\text{rev}}(\theta(s,p))}\ \underbrace{\frac{s/K_S}{1+s/K_S+p/K_P}}_{\eta^{\text{kin}}(s,p)}, \tag{4}$$

with the thermodynamic driving force $\theta(s,p) = \ln(K^{\text{eq}}\ s/p)$, the thermodynamic efficiency η^{rev} and the kinetic efficiency η^{kin} (see Figure 3). Note that the driving force can also be defined as $-\Delta G'$ which is equal to $RT\ln(K^{\text{eq}}\ s/p)$, but here we drop the gas constant R and temperature T to have a unitless (positive) variable. The two efficiency terms can only assume values between 0 and 1. This factorized formulation of the Haldane rate law is equivalent to the one in Eq. (3), based on the constraint that $K^{\text{eq}} = k_+^{\text{cat}}/k_-^{\text{cat}} \cdot K_P/K_S$, which is commonly known as the *Haldane relationship* (see [23] for more details). This rate law is the most realistic one that we discuss in this work, and deriving it requires only a few assumptions. However, it is also the most mathematically complex and therefore most questions we raise below do not have analytic solutions. So, in addition, we consider three simplified rate laws as limiting cases.

- **Mass-action rate law** A very common approximation for enzymatic rate laws (the one also made by Waley [20]) is based on the limit of low metabolite concentrations ($s \ll K_S$ and $p \ll K_P$). In this case, the concentration-dependent terms in the denominator (i.e., $s/K_S + p/K_P$) can be neglected, and we get:

$$v = \varepsilon\ (k_+^{\text{cat}}\ s/K_S - k_-^{\text{cat}}\ p/K_P) \tag{5}$$

There are many equivalent ways to write down this rate law. For instance, we can apply the same approximation to the factorized form in Eq. (4) to get $v = \varepsilon\ k_+^{\text{cat}}\ (1-e^{-\theta})\ s/K_S$, and we can further replace θ with its explicit definition based on reactant concentrations and write $v = \varepsilon\ k_+^{\text{cat}}/K_S\ (s - p/K^{\text{eq}})$. Another common form for this rate law is based on the first-order rate constants $k_+ \equiv k_+^{\text{cat}}/K_S$ and $k_- \equiv k_-^{\text{cat}}/K_P$. Using them in Eq. (5) looks like this: $v = \varepsilon\ (k_+ s - k_- p)$. As the rate law resembles mass-action kinetics for non-enzymatic reactions, we will refer to this as the "mass-action" rate law although here the enzyme level appears as a prefactor. Throughout this paper we will switch between these four different notations based on convenience.

- **Thermodynamic rate law** If the kinetic efficiency η^{kin} is approximated by 1 (e.g. in the limit $s \gg K_S$ and $p \ll K_P$), we obtain:

$$v = \varepsilon\ k_+^{\text{cat}}\ \left(1 - \frac{p}{s}\ \frac{1}{K^{\text{eq}}}\right) = \varepsilon\ k_+^{\text{cat}}(1-e^{-\theta}). \tag{6}$$

We will also consider a special case of this rate law where $\theta \to 0$ and therefore $p/s \approx K^{\text{eq}}$, i.e. the reaction is close-to-equilibrium. In this case the rate law becomes $v = \varepsilon\ k_+^{\text{cat}}\ \theta$.

- **Irreversible saturable rate law ("Michaelis-Menten kinetics")** We next assume that both $p \ll s\ K^{\mathrm{eq}}$ (which means that $\theta \to \infty$ and therefore the thermodynamic efficiency η^{rev} can be approximated by 1) and also $p \ll K_{\mathrm{P}}$ (so we can drop p/K_{P} from the denominator in η^{kin}). In this case, we obtain the Michaelis-Menten rate law

$$v = \varepsilon\ k_{+}^{\mathrm{cat}}\ \frac{s}{s + K_{\mathrm{S}}}, \tag{7}$$

  which depends on the substrate, but not on the product concentration. Originally, Michaelis and Menten [24] developed this irreversible rate law by assuming that the rate of enzyme-substrate binding is very fast compared to catalysis, and that the catalytic step is irreversible. The assumptions made here lead to the same result but are less stringent.

In the approximations, instead of setting the thermodynamic or saturation efficiencies to their maximal value of 1, we may also approximate them by constant numbers smaller than 1; we obtain exactly the same rate laws, but instead of the k^{cat} value we obtain a smaller apparent value in the approximated rate law.

2.2 Analytic formulae for unbranched metabolic pathways

How can we characterize metabolic states by using analytic formulae? Here we consider chains of uni-uni reactions. For such an unbranched pathway with a given type of rate laws (e.g. mass-action or Michaelis-Menten rate laws), we are interested in formulae for a number of quantities:

1. **Metabolic steady state** Given the enzyme levels, external metabolite concentrations, and kinetic constants, we can directly compute the stationary fluxes and internal metabolite concentrations. For general metabolic networks, no explicit formulae are known, but for unbranched pathways with some simplified rate laws, explicit formulae exist. Incidentally, this also shows that in these models the steady state concentrations are unique.

2. **Stability of steady state** If the Jacobian matrix in a steady state has positive eigenvalues, the state is asymptotically unstable and is not able to persist under (inevitable) chemical noise in the cell. Stability is a prerequisite for metabolic control coefficients being defined. A sufficient (but not necessary) condition for stable steady states in unbranched metabolic pathways is given in Appendix F.8.

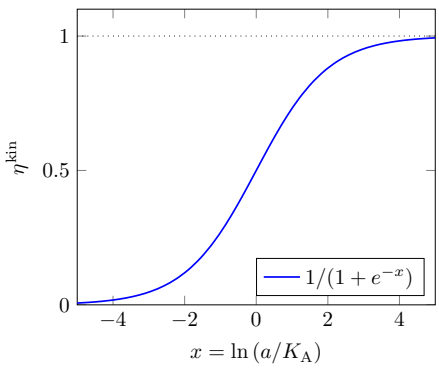
3. **Metabolic control** The response coefficients are defined as the derivatives between steady-state concentrations or fluxes and model parameters (e.g. the enzyme levels). If two model parameters act (exclusively) on the same reaction, all their response coefficients will be the same (except for a proportional scaling). Taking this into account, the control coefficients describe the same type of derivatives, but for a set of hypothetical, reaction-specific parameters. In practice, if reaction rates are proportional to enzyme levels and if each reaction is catalyzed by a single specific enzyme, we can write the enzyme response coefficients as $R^X_{\varepsilon_l} = C^X_l \cdot \frac{v_l}{\varepsilon_l}$. Importantly, throughout this paper we use elasticities, response coefficients, and control coefficients in their unscaled form, for instance unscaled enzyme response coefficients $R^J_{\varepsilon_i} = \partial J / \partial \varepsilon_i$ instead of the common scaled form $\frac{\varepsilon_i}{J} \frac{\partial J}{\partial \varepsilon_i} = \frac{\partial \ln J}{\partial \ln \varepsilon_i}$.

   The metabolic control coefficients can be computed from the stoichiometric matrix and the elasticity matrix, and they satisfy summation and connectivity theorems. They can be computed in two ways: if an analytic formula for the metabolic steady state is known (as is the case in some unbranched pathway models studied below), we may differentiate symbolically by the enzyme levels; otherwise, we may compute the elasticity coefficient matrices by differentiating the rate laws, and then compute the control coefficient matrices from them using a known formula; however, since this formula involves a matrix inversion, writing this down as an analytic formulae may be extremely complicated, and control coefficients are usually computed numerically. Another way to compute control coefficients, which follows from the enzyme-control rule and works only in optimal states, is described below.

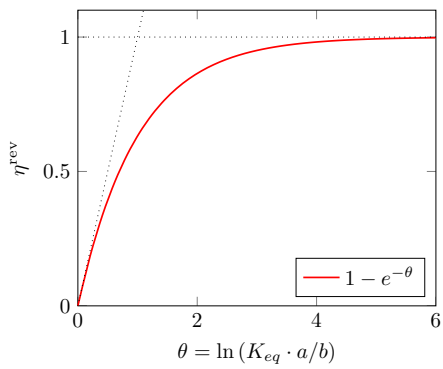
4. **Optimal metabolic states** In our basic metabolic optimality problem, we define optimal states as states in which a given enzyme budget (fixed sum of enzyme levels) is allocated to maximize a production flux. Kinetic constants and external metabolites are given, and we compute the optimal metabolite profile, the optimal enzyme profile, and the optimal flux. If the flux distribution is known (e.g. a steady flux in an unbranched pathway) and can only increase or decrease proportionally, this problem is equivalent to the ECM problem of minimizing the enzyme demand at a given (unit) flux. This convex problem can be solved numerically, but analytic solutions were known only for very few cases. Below we present some new analytic solutions.

$$v = \varepsilon \, \frac{k_+^{\mathrm{cat}} \; a/K_{\mathrm{A}} - k_-^{\mathrm{cat}} \; b/K_{\mathrm{B}}}{1 + a/K_{\mathrm{A}} + b/K_{\mathrm{B}}} = \varepsilon \; k^{\mathrm{cat}} \; \underbrace{\left(1 - \frac{b}{aK^{\mathrm{eq}}}\right)}_{\eta^{\mathrm{rev}}} \; \underbrace{\frac{a/K_{\mathrm{A}}}{1 + a/K_{\mathrm{A}} + b/K_{\mathrm{B}}}}_{\eta^{\mathrm{kin}}}$$

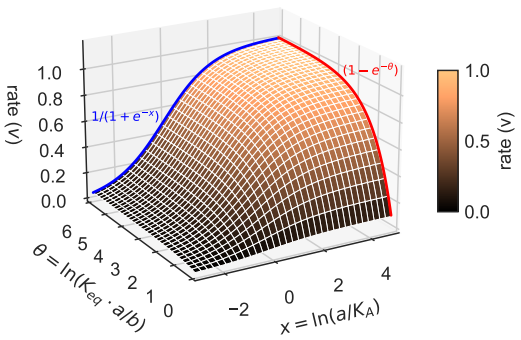
(a) Haldane rate law



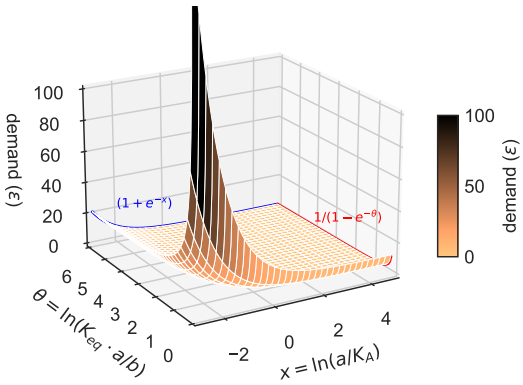
(b) Kinetic efficiency



(c) Thermodynamic efficiency



(d) Rate laws



(e) Enzyme demand

Figure 3: The Haldane rate law and some approximations with constant efficiency terms. (a) The Haldane rate law. In the factorized form (right), thermodynamic and saturation effects are described by separate efficiency factors. (b) The kinetic efficiency as a function of the product of the substrate (in log-scale), assuming $b \ll K_{\mathrm{B}}$. (c) The thermodynamic efficiency as a function of the driving force θ. (d) A surface plot showing the rate v as a function of θ and the substrate concentration (in log-scale, relative to the K_{A}). The parameters are $K^{\mathrm{eq}} = K_{\mathrm{A}} = k^{\mathrm{cat}} = \varepsilon = 1$ and $K_{\mathrm{B}} = 10$. (e) A surface plot showing the enzyme demand for a given rate ($v = 1$). All kinetic parameters are the same as in (d).

Formulae for optimal enzyme levels and the optimal achievable flux are shown below, in Table 1. We also consider a related problem, maximizing the flux under a constraint on the total enzyme plus metabolite amount.

5. **Metabolic control in optimal states** The control coefficients tell us how a metabolic system responds to perturbations (resulting in a new state that is stationary, but probably non-optimal). For optimal states (with a constraint on the sum of enzyme levels), the enzyme-control rule states that enzyme levels and flux control coefficients must be proportional. Since we know (from the summation theorem) that the control coefficients must sum to 1, we can conclude: whenever there is an analytic formula for optimal enzyme levels, we obtain a formula for the control coefficients.

2.3 Optimal metabolic states: analytic solutions

What are the general principles behind optimal enzyme allocation? One important principle, valid in optimal metabolic states, has been shown for pathways with mass-action rate laws [21] and later been confirmed for general rate laws [17]: in optimal states, where the metabolic flux has been maximized at a fixed enzyme budget, enzyme levels and flux control coefficients must be proportional[1]:

$$\varepsilon_i^* \propto C_i^{J\,*}. \tag{8}$$

Here the star * denotes variables in the optimal state. Together with the summation theorem for flux control coefficients [14, 25], $\sum_i C_i^J = 1$, we obtain the conversion formulae

$$C_i^J = \frac{\varepsilon_i^*}{\sum_j \varepsilon_i^*} \qquad \text{and} \qquad \varepsilon_i^* = \varepsilon_{\text{tot}}\, C_i^J. \tag{9}$$

The enzyme-control rule (8) provides a condition for metabolic states, independent of the type of rate laws considered. Importantly, it holds only in states of maximal flux, given a fixed total enzyme amount and no other constraints. More realistic models employ also bounds on metabolite levels [22, 5, 26] for different reasons. First, metabolite levels are bounded in real cells: while metabolite molecules may be small, their concentration is high, and they contribute much more than macromolecules to osmotic pressure. Second, as we will see below, some models without metabolite bounds lead to paradoxical results. We will therefore present a generalized version of the enzyme-control rule that takes metabolite bounds into account.

But what are the general shapes of optimal enzyme profiles, i.e. how do enzyme levels vary across the network? And on what factors (kinetic, thermodynamic, or cost factors) does this depend? To answer this, the enzyme-control rule alone would not be enough (because no kinetic details appear in the rule). Also numerical studies would not be enough (because they apply to single cases and yield no general laws). Hence, to study this, it would be good to consider analytic solutions. Unfortunately, analytic solutions are not known for general metabolic models, but solutions exist for unbranched pathways with simple rate laws (mass-action rate law and saturable kinetics.

We now present analytic formulae for optimal metabolic states with different types of rate laws. We discuss the rate laws in increasing order of difficulty. An overview of all analytic solutions is given in Table 1.

2.3.1 Michaelis-Menten rate law

We first consider the Michaelis-Menten rate law (i.e. irreversible reactions with simple saturation kinetics):

$$v_i = \varepsilon_i \cdot k_i^{\text{cat}}\, \frac{s_{i-1}}{s_{i-1} + K_{\text{M};i}}\,. \tag{10}$$

In an unbranched pathway at steady state, all rates must be equal. To describe them, we introduce a new variable J called the pathway flux, and require that $\forall i\ v_i = J$. Now we can use Eq. (10) to find a relationship between substrate and enzyme levels:

$$\varepsilon_i = \frac{J}{k_i^{\text{cat}}} \cdot \left(1 + \frac{K_{\text{M};i}}{s_{i-1}}\right) = J\left(\alpha_i + \frac{\beta_i}{s_{i-1}}\right) \tag{11}$$

where we defined $\alpha_i \equiv 1/k_i^{\text{cat}}$ and $\beta_i \equiv K_{\text{M};i}/k_i^{\text{cat}}$ for convenience.

Combining this equation with an upper bound on total enzyme from Eq. (2) we get that:

$$\varepsilon_{\text{tot}} \geq \sum_i \varepsilon_i = \sum_i J\ \left(\alpha_i + \frac{\beta_i}{s_{i-1}}\right) = J\left(\sum_i \alpha_i + \sum_i \frac{\beta_i}{s_{i-1}}\right) \tag{12}$$

and by rearranging we obtain:

$$J \leq \frac{\varepsilon_{\text{tot}}}{\sum_i \alpha_i + \sum_i \frac{\beta_i}{s_{i-1}}}. \tag{13}$$

[1] To derive this rule, we note that in an optimal state all enzymes have a marginal cost (contribution to the enzyme budget, per mol of enzyme) which must be balanced by the same marginal benefit (contribution to the production flux, per mol of enzyme). If all enzymes contribute to the enzyme budget with equal weights, their marginal costs are the same. This means that also their marginal benefit must be the same, which are given by the (unscaled) flux response coefficients $R_{e_l}^J = \frac{\partial v^{\text{st}}}{\partial e_l}$. This, finally, means that their flux control coefficients $C_l^J = \frac{\partial v^{\text{st}}}{\partial e_l} / (\frac{v^{\text{st}}}{e_l})$ must be proportional to the enzyme levels e_l.

Name (assumptions)	**Enzymatic rate law:** $v_i/\varepsilon_i =$	**Comments / definitions**	**Optimal enzyme allocation:** $\varepsilon_i{}^* \propto$	**Minimum pathway cost per flux:** $\varepsilon_{\text{tot}}/J^* =$
Trivial (irreversible, saturated)	k_i^{cat}	$\alpha_i \equiv 1/k_i^{\text{cat}}$	α_i	$\|\|\boldsymbol{\alpha}\|\|_1$
Michaelis-Menten (irreversible, saturable)	$k_i^{\text{cat}} \frac{s_{i-1}}{s_{i-1} + K_{\text{M};i}}$	$\beta_i \equiv K_{\text{M};i}/k_i^{\text{cat}}$ $\sum_i s_i \leq s_{\text{tot}}$	$\alpha_i + \sqrt{\beta_i \, \|\|\boldsymbol{\beta}\|\|_{\frac{1}{2}}}/s_{\text{tot}}$	$\|\|\boldsymbol{\alpha}\|\|_1 + \|\|\boldsymbol{\beta}\|\|_{\frac{1}{2}} \big/ s_{\text{tot}}$
Thermodynamic (reversible, unsaturable)	$k_i^{\text{cat}} \cdot \left(1 - \frac{s_i}{s_{i-1} K_i^{\text{eq}}}\right)$	Find Ψ s.t. $-\frac{\theta_{\text{tot}}}{2} = \sum_i \ln\left(\sqrt{\Psi\alpha_i} + \sqrt{1+\Psi\alpha_i}\right)$ where $\theta_{\text{tot}} \equiv -\ln\left(s_0 K_{\text{tot}}^{\text{eq}}/s_n\right) = \text{const}$	$\alpha_i \left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + (\Psi\alpha_i)^{-1}}\right)$	$\sum_i \alpha_i \left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + (\Psi\alpha_i)^{-1}}\right)$
Thermodynamic (approximate solution)	$k_i^{\text{cat}} \cdot \left(1 - \frac{s_i}{s_{i-1} K_i^{\text{eq}}}\right)$	Same as above	No explicit formula	$\|\|\boldsymbol{\alpha}\|\|_1 \left(1 - \exp\left(-\frac{\|\|\boldsymbol{\alpha}\|\|_1}{\|\|\boldsymbol{\alpha}\|\|_{\frac{1}{2}}}\theta_{\text{tot}}\right)\right)^{-1}$
Mass-action [20] (reversible, unsaturable)	$k_i^{\text{cat}} \cdot \left(1 - \frac{s_i}{s_{i-1} K_i^{\text{eq}}}\right) \frac{s_{i-1}}{K_{\text{M};i}}$	$s_0 = \text{const}$, $s_n = \text{const}$ $\gamma_i \equiv \beta_i \cdot \prod_{j=i}^{n} K_j^{\text{eq}}$	$\sqrt{\gamma_i}$	$\|\|\boldsymbol{\gamma}\|\|_{\frac{1}{2}} \big/ \left(s_0 - s_n/K_{\text{tot}}^{\text{eq}}\right)$
Haldane (reversible, saturable)	$k_i^{\text{cat}} \cdot \left(1 - \frac{s_i}{s_{i-1} K_i^{\text{eq}}}\right) \frac{s_{i-1}/K_{\text{S};i}}{1 + s_{i-1}/K_{\text{S};i} + s_i/K_{\text{P};i}}$	No known solution, use convex optimization solver		

Table 1: A summary of all rate laws considered in this paper, along with their solutions for optimal enzyme allocation and the resulting minimum pathway cost per flux. The rate laws are roughly ordered by increasing level of complexity. The "trivial" rate law does not have its own section in the text, but is shown here as the limit of the Michaelis-Menten rate law when $s_{i-1} \gg K_{\text{M};i}$, as well as the thermodynamic rate law when $K_{\text{tot}}^{\text{eq}}\, s_0/s_n \to \infty$. Although the "mass-action" rate law appears in this table in a modified form (for the purpose of using the same set of kinetic parameters as the other rate laws), it is equivalent to the form used in previous studies [20, 16, 17]. In order to calculate the absolute optimal enzyme levels from the relative ones (given in the column titled *Optimal enzyme allocation*), one can use $\varepsilon_i^* = \varepsilon_{\text{tot}} \frac{x_i}{\sum_j x_j}$ where x_i are the relative values.

Maximizing J would mean that we reach the upper bound, and therefore we can also treat this as an equality. Since ε_{tot} is constant and the only free variables are the metabolite concentrations, the maximal flux is reached when the denominator on the right-hand side is minimized. The problem is that it is a monotonically decreasing function in s_i (for each i) and since metabolite concentrations are unbounded, the optimum would be at $s_i \to \infty$ which, formally, is not a mathematically defined optimum state. But even if we consider this as a limit, the resulting state would not be controllable, but instead very fragile against any changes of enzyme levels (see Appendix C.1). In reality, of course, the range of physiological osmotic pressures does impose some constraint on the concentrations of small molecules. As a proxy for this effect, we can add another constraint to bound the sum of all metabolite concentrations, $\sum_i s_i \leq s_{\text{tot}}$. Now, one can show (see Appendix F.1) that the optimal allocation of enzymes will obey:

$$\varepsilon_i{}^* \propto \alpha_i + \sqrt{\beta_i} \cdot \frac{\sqrt{||\boldsymbol{\beta}||_{\frac{1}{2}}}}{s_{\text{tot}}} \tag{14}$$

where $\boldsymbol{\beta}$ is the vector of all β_i and $||\mathbf{x}||_{\frac{1}{2}} = (\sum_i \sqrt{x_i})^2$ is the $\text{l}_{1/2}$ norm (i.e., $||\boldsymbol{\beta}||_{\frac{1}{2}} \equiv \left(\sum_i \sqrt{\beta_i}\right)^2$). In this case, the maximal flux would be:

$$J^* = \varepsilon_{\text{tot}} \cdot \left(||\boldsymbol{\alpha}||_1 + ||\boldsymbol{\beta}||_{\frac{1}{2}}/s_{\text{tot}}\right)^{-1} . \tag{15}$$

Note that the solution looks essentially the same even if we constrain the first metabolite (s_0) to have a fixed concentration (see Appendix F.1.4). We will revisit this case in section 2.5 in the context of a course-grained model of a growing cell.

If the metabolite density constraint is not very tight (i.e. s_{tot} is large enough), we can ignore the second term in Eq. (14), which would be equivalent to assuming all enzymes are substrate-saturated. In this case, the optimal allocation of enzymes will be proportional to α_i (or inversely proportional to k_i^{cat}) and therefore:

$$\lim_{s_{\text{tot}} \to \infty} J^* = \varepsilon_{\text{tot}} \cdot (||\boldsymbol{\alpha}||_1)^{-1} . \tag{16}$$

Interestingly, $||\boldsymbol{\alpha}||_1$, which is equal to the sum of k^{cat} reciprocals, is in fact the inverse of the *Pathway Specific Activity* (PSA) as originally defined by Bar-Even et al. [11]. Indeed, the idealized scenario considered in that study (where all enzymes were irreversible and saturated) provides an upper bound on the maximal flux achievable. Interestingly, if we see the $\alpha_i = 1/k_i^{\text{cat}}$ as minimal enzymatic turnover times and $\alpha_{\text{PW}} = \varepsilon_{\text{tot}}/J$ is the turnover time

of the pathway, we can simply rephrase Eq. (16) by saying that, like in the case of the PSA, turnover times along a pathway are additive. In fact, this holds more generally for any rate law $v_i = e_i\ k_i(\mathbf{s})$, if the enzyme turnover times are defined as $\tau_i = 1/k_i$, and resembles the fact that electric resistances, in a series of resistors, are additive.

One reason that the irreversible Michaelis-Menten model is unrealistic is that it does not capture reactions that are close to equilibrium and therefore suffer from a counter-productive reverse flux. This is yet another reason why it might be impossible for some metabolites to reach very high concentrations: they might be products of unfavorable reactions. In most metabolic networks, about half of the reactions are reversible and therefore it would be more realistic to use a reversible rate law like such as Haldane's. However, using the Haldane equation would typically create a system of equations for which no analytic solution is known. So, first, we will consider rate laws that account for the reverse flux but, for simplicity, ignore substrate saturation.

2.3.2 Thermodynamic rate law

Irreversible rate laws like the Michaelis-Menten kinetics depend only on the substrate concentration; in the formula there is no product-dependent reverse term that decreases the total rate or could make it become negative. However, according to thermodynamics, such laws can only be approximations: thermodynamically feasible rate laws must contain a reverse term and must depend on the thermodynamic imbalance of substrate and product concentrations expressed by the thermodynamic driving force. Some rate laws can even be written as functions of the thermodynamic force alone. Here we describe such a "thermodynamic" rate law, where v is proportional to the thermodynamic efficiency $1 - \mathrm{e}^{-\theta}$, while the kinetic efficiency η^{kin} is assumed to be constant:

$$v_i = \varepsilon_i\ k_i^{\mathrm{cat}} \left(1 - e^{-\theta_i}\right) = \varepsilon_i\ k_i^{\mathrm{cat}} \left(1 - \frac{s_i}{s_{i-1}}\ \frac{1}{K_i^{\mathrm{eq}}}\right) . \tag{17}$$

This type of kinetics approximates cases where all reactions are saturated ($s_i \gg K_{\mathrm{M};i}$), and therefore the thermodynamic efficiency η^{kin} in Eq. (4) becomes 1. The parameters here are the turnover numbers k_i^{cat} and the equilibrium constants K_i^{eq}. However, as we will soon learn, the individual K_i^{eq} do not change the result, and only the overall equilibrium constant ($K_{\mathrm{tot}}^{\mathrm{eq}} = \prod_i K_i^{\mathrm{eq}}$) matters.

As before, we can define the steady-state pathway flux J and apply an upper bound on the total enzyme to get:

$$\begin{aligned} \varepsilon_{\mathrm{tot}} \geq \sum_i \varepsilon_i = \sum_i \frac{J}{k_i^{\mathrm{cat}}\left(1 - e^{-\theta_i}\right)} &= J \cdot \left(\sum_i \frac{1}{k_i^{\mathrm{cat}}} \cdot \frac{1}{1 - e^{-\theta_i}}\right) \\ J \leq \varepsilon_{\mathrm{tot}} \cdot \left(\sum_i \frac{1}{k_i^{\mathrm{cat}}} \cdot \frac{1}{1 - e^{-\theta_i}}\right)^{-1} &. \end{aligned} \tag{18}$$

Unfortunately, there is no closed-form analytic solution for the maximal rate of this thermodynamic rate law. But we do have something very close which requires rather simple computations. First, we need to invert the functional relationship between the overall driving force (θ_{tot}) and an auxiliary variable Ψ which is defined by the inverse function of:

$$\ln\left(\frac{s_0}{s_n} K_{\mathrm{tot}}^{\mathrm{eq}}\right) = \theta_{\mathrm{tot}} = 2 \sum_i \ln\left(\sqrt{\Psi/k_i^{\mathrm{cat}}} + \sqrt{1 + \Psi/k_i^{\mathrm{cat}}}\right) . \tag{19}$$

The expression on the right is analytic and strictly increasing in the range $\Psi \in [0, \infty)$, so there is a unique solution which can be found by simple numerical methods. Then, we can use that value to directly calculate the optimal driving forces, enzyme levels and pathway flux:

$$\begin{aligned} {\theta_i}^* &= 2 \ln\left(\frac{1}{k_i^{\mathrm{cat}}}\left(1 + \sqrt{1 + k_i^{\mathrm{cat}}/\Psi}\right)\right) + \ln\left(\Psi k_i^{\mathrm{cat}}\right) \\ \varepsilon_i^* &= \frac{J^*}{2} \cdot \frac{1}{k_i^{\mathrm{cat}}}\left(1 + \sqrt{1 + k_i^{\mathrm{cat}}/\Psi}\right) \\ J^* &= 2\ \varepsilon_{\mathrm{tot}} \cdot \left(\sum_i \frac{1}{k_i^{\mathrm{cat}}}\left(1 + \sqrt{1 + k_i^{\mathrm{cat}}/\Psi}\right)\right)^{-1} . \end{aligned} \tag{20}$$

The full derivation of this solution can be found in Appendix F.2.2. An example of what the relationship between the driving force and the optimal flux looks like for a pathway with two enzymes is illustrated in Figure 4(a) and 4(b). Furthermore, a comparison between our exact formula presented here and a numerical solution based on convex optimization [8] shows no difference at all.

These formulae cannot be directly evaluated because of the unknown parameter Ψ. To obtain solutions that do not depend on this parameter, we now consider two limiting cases in which Ψ is either infinitely high (very high driving force) or infinitely low (very low driving force).

When the driving forces are very high (i.e. $\theta_{\text{tot}} \to \infty$), the solution for Ψ in Eq. (19) will approach infinity and the optimal flux becomes

$$\lim_{\theta_{\text{tot}} \to \infty} J^* = \lim_{\Psi \to \infty} \frac{2\ \varepsilon_{\text{tot}}}{\sum_i 1/k_i^{\text{cat}} \cdot \left(1 + \sqrt{1 + k_i^{\text{cat}}/\Psi}\right)} = \frac{\varepsilon_{\text{tot}}}{\sum_i 1/k_i^{\text{cat}}} = \varepsilon_{\text{tot}} \cdot (||\boldsymbol{\alpha}||_1)^{-1} \tag{21}$$

with parameters $\alpha_i = 1/k_i^{\text{cat}}$ defined as before. This solution indeed makes sense as it is equivalent to the fully saturated limit in the Michaelis-Menten case (see Section 2.3.1, Eq. (16)).

However, when the driving force in the pathway is limited (Ψ has a finite positive value), this driving force will be split between reactions according to their k_i^{cat} values: enzymes with a higher k_i^{cat} value will have to pay a higher penalty due to their driving force being closer to 0. On the other hand, slow enzymes will obtain more driving force, which will help them by having a smaller fraction of reverse flux. Notably, the distribution does *not* depend on the reaction equilibrium constants (only on the overall $K_{\text{tot}}^{\text{eq}}$). Perhaps this is not that surprising if we consider the fact that the concentrations intermediate substrates and products are unconstrained and therefore we have enough degrees of freedom to adjust to any value of K_i^{eq} as long as the total driving force stays the same.

One can also consider the other extreme where all reactions are close to equilibrium, which means that θ_{tot} is close to 0 (and therefore also each $\theta_i \to 0$). In this case, we can approximate Eq. (18) by:

$$J \leq \varepsilon_{\text{tot}} \cdot \left(\sum_i \frac{1}{k_i^{\text{cat}}} \cdot \frac{1}{1 - e^{-\theta_i}}\right)^{-1} \approx \varepsilon_{\text{tot}} \cdot \left(\sum_i \frac{1}{k_i^{\text{cat}}\ \theta_i}\right)^{-1}. \tag{22}$$

Maximizing J under the constraint $\sum_i \theta_i = \theta_{\text{tot}}$ yields the solution (see the full derivation in appendix F.2.3):

$$\begin{aligned} {\theta_i}^* &\approx \theta_{\text{tot}}\ \frac{\sqrt{1/k_i^{\text{cat}}}}{\sum_i \sqrt{1/k_i^{\text{cat}}}} \\ J^* &\approx \frac{\theta_{\text{tot}}\ \varepsilon_{\text{tot}}}{\left(\sum_i \sqrt{1/k^{\text{cat}}}\right)^2} = \varepsilon_{\text{tot}} \cdot \frac{\theta_{\text{tot}}}{||\boldsymbol{\alpha}||_{\frac{1}{2}}}. \end{aligned} \tag{23}$$

Using the two limiting cases ($\theta_{\text{tot}} \to \infty$ and $\theta_{\text{tot}} \approx 0$), we can approximate the solution from Eq. (20) by a much simpler formula which has the same shape as the thermodynamic rate law for a single reaction:

$$J^* \approx \frac{\varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_1} \left(1 - \exp\left(-\frac{||\boldsymbol{\alpha}||_1}{||\boldsymbol{\alpha}||_{\frac{1}{2}}} \theta_{\text{tot}}\right)\right). \tag{24}$$

Figures 4 compares between the precise and approximate solution for two example cases (a more detailed comparison for metabolic pathways of different lengths and parameter choices is shown in Appendix F.3). One can appreciate that the two curves are nearly indistinguishable, illustrating the good quality of the approximation.

2.3.3 Mass-action rate law

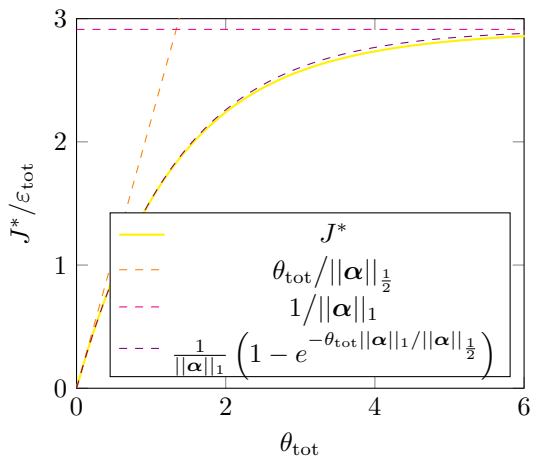
Next, we consider the same unbranched pathway but assuming mass-action rate laws:

$$v_i = \varepsilon_i\ k_i^{\text{cat}}/K_{\text{M};i}\ (s_{i-1} - s_i/K_i^{\text{eq}}) = \varepsilon_i\ \beta_i^{-1}\ (s_{i-1} - s_i/K_i^{\text{eq}}). \tag{25}$$

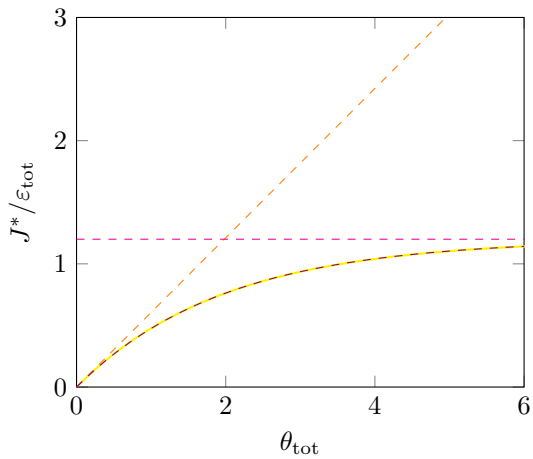
As before, we defined $\beta_i \equiv K_{\text{M};i}/k_i^{\text{cat}}$. Note that instead of Eq. (5) we use a form of the mass-action rate law that does not require the turnover rate and K_{M} of the product (and instead uses K_i^{eq}), to avoid confusing indexation of forward and backward parameters.

Like with the previous rate laws, we define the pathway flux J and apply an upper bound on the total enzyme levels:

$$\begin{aligned} \varepsilon_{\text{tot}} &\geq \sum_i \varepsilon_i = \sum_i \frac{J\ \beta_i}{s_{i-1} - s_i/K_i^{\text{eq}}} = J \sum_i \frac{\beta_i}{s_{i-1} - s_i/K_i^{\text{eq}}} \\ J &\leq \varepsilon_{\text{tot}} \cdot \left(\sum_i \frac{\beta_i}{s_{i-1} - s_i/K_i^{\text{eq}}}\right)^{-1}. \end{aligned} \tag{26}$$



(a) $k_1^{\text{cat}} = 3s^{-1}$ $k_2^{\text{cat}} = 100s^{-1}$



(b) $k_1^{\text{cat}} = 3s^{-1}$ $k_2^{\text{cat}} = 2s^{-1}$

Figure 4: The relationship between the maximal flux per enzyme and overall driving force for the thermodynamic rate law. Even though we do not have a closed-form solution for J^* as a function of θ_{tot}, we can still plot $J^*(\Psi)$ against $\theta_{\text{tot}}(\Psi)$ for varying values of Ψ using Eq. (19) and Eq. (20). Here, we show this relationship for a pathway with 2 steps (yellow). The parameters are: $K_i^{\text{eq}} = 1$, $k_1^{\text{cat}} = 3\ s^{-1}$, and the k^{cat} of the second enzyme is either (a) $k_2^{\text{cat}} = 100\ s^{-1}$ or (b) $k_2^{\text{cat}} = 2\ s^{-1}$. The orange dashed line represents the approximations for very small driving forces based on Eq. (23), and similarly in pink for very large driving forces – Eq. (21). The approximation (dashed purple line) based on Eq. (24) is based on effective parameters chosen to match the slope at $\theta_{\text{tot}} = 0$ and the limiting maximal flux at $\theta_{\text{tot}} \to \infty$.

Again, we maximize the flux at a constrained total enzyme level. Optimization with Lagrange multipliers yields the following expressions for the optimal individual enzyme levels (${\varepsilon_j}^*$) and the maximal flux (J^*) (see full derivation in Appendix F.4):

$$\begin{aligned} \varepsilon_j^* &= \varepsilon_{\text{tot}} \sqrt{\frac{\gamma_j}{||\boldsymbol{\gamma}||_{\frac{1}{2}}}} \\ J^* &= \varepsilon_{\text{tot}} \cdot \frac{s_0\ K_{\text{tot}}^{\text{eq}} - s_n}{||\boldsymbol{\gamma}||_{\frac{1}{2}}} \end{aligned} \tag{27}$$

where γ_j is defined as

$$\gamma_i \equiv \beta_i \prod_{j=i}^{n} K_j^{\text{eq}} \tag{28}$$

Of course, the exact value of this maximum depends on all the different system parameters. However, it is interesting to consider a naïve assumption where all the γ_j parameters are identical. In such a case, the flux in the

pathway would decrease quadratically with the number of steps [13]. Of course, we know that in real metabolic pathways the equilibrium constants are typically not close to 1, and therefore this approximation might not have many applications in biology.

What would happen if a mutation improved the catalytic rate of only one of the enzymes ε_i by a factor of a (i.e. β_i decreases by a factor of a, but the equilibrium constant K_i^{eq} remains the same)? In this case, γ_i would be divided by a but the optimal enzyme concentration for this reaction ${\varepsilon_i}^*$ would only decrease by a factor of $\sqrt{a}$. This saving would then be distributed proportionally among all the n enzymes and would contribute to an increase in the pathway flux J. On the flip side, if for some reason the activity of one enzyme is decreased by a multiplicative factor b, it would need to "pay" (increase the enzyme's concentration) only by a factor $\sqrt{b}$, and this increase would be "funded" by all of the enzymes together in order to keep the same ε_{tot}, thus lowering J.

2.3.4 Haldane rate law

So far we analyzed three cases where all enzymes could be described by the same kinetic rate law: mass-action, Michaelis-Menten, or thermodynamic. Although these rate laws can reliably describe some enzymes in specific conditions, it is very unlikely that such an approximation would apply to *all* the reactions in a single pathway (except for the trivial case of a 1-reaction pathway). A more realistic model would allow all reactions to be reversible and saturable. Here, we will analyze such a case based on the factorized rate law with one substrate and one product. Note that although it is equivalent to Eq. (3), we prefer this formulation because it is easier to separate thermodynamics from saturation effects.

$$v_i = \varepsilon_i \cdot k_i^{\text{cat}} \cdot \left(1 - e^{-\theta_i}\right) \cdot \frac{s_{i-1}/K_{S,i}}{1 + s_{i-1}/K_{S,i} + s_i/K_{P,i}} \,. \tag{29}$$

As always, we can assume that all fluxes are equal to J, and use the total enzyme budget to get an upper bound:

$$\begin{aligned}
\varepsilon_{\text{tot}} \geq \sum_i \varepsilon_i &= \sum_i \frac{J}{k_i^{\text{cat}} \cdot \left(1 - e^{-\theta_i}\right) \cdot \frac{s_{i-1}/K_{S,i}}{1+s_{i-1}/K_{S,i}+s_i/K_{P,i}}} \\
&= J \cdot \left(\sum_i \frac{1}{k_i^{\text{cat}}} \left(1 - e^{-\theta_i}\right)^{-1} \left(1 + \frac{K_{S,i}}{s_{i-1}} + \frac{s_i K_{S,i}}{s_{i-1} K_{P,i}}\right)\right) \\
J &\leq \varepsilon_{\text{tot}} \cdot \left(\sum_i \frac{1}{k_i^{\text{cat}}} \left(1 - e^{-\theta_i}\right)^{-1} \left(1 + \frac{K_{S,i}}{s_{i-1}} + \frac{s_i K_{S,i}}{s_{i-1} K_{P,i}}\right)\right)^{-1}
\end{aligned} \tag{30}$$

where we can now appreciate how this is a generalization of both Eq. (13) (Michaelis-Menten) and (18) (thermodynamic). We can see that the maximal pathway flux would be realized when the term in parentheses is minimized with respect to the s_i, i.e.:

$$\text{minimize}_{\mathbf{s}} \sum_i \frac{1}{k_i^{\text{cat}}} \left(1 - e^{-\theta_i(\mathbf{s})}\right)^{-1} \left(1 + \frac{K_{\text{S,i}}}{s_{i-1}} + \frac{s_i K_{\text{S,i}}}{s_{i-1} K_{\text{P,i}}}\right) \,. \tag{31}$$

Unfortunately, as we already mentioned before, no analytic solution to this problem is known in the general case.

2.4 Optimal metabolic states: insights from Metabolic Control Analysis

2.4.1 Enzyme-control rule and enzyme-elasticity rule

We now continue with some observations about the enzyme-control rule. The enzyme-control rule Eq. (8) states that enzyme levels and flux control coefficients in optimal states are proportional: $\varepsilon_i^* \propto {C_i^J}^*$. When the aim is to compute control coefficients in optimal states (at a given enzyme budget ε_{tot}), the enzyme-control rule comes in handy. The summation theorem tells us that the flux control coefficients must sum to 1; therefore, the optimal enzyme levels are given by the flux control coefficients, multiplied by the (predefined) total enzyme level, and we obtain the simple conversion $C_i^{J*} = \frac{\varepsilon_i^*}{\varepsilon_{\text{tot}}}$ and $\varepsilon_i^* = \varepsilon_{\text{tot}}\ C_i^{J*}$. In the basic form of the rule we put a constraint on the enzyme mass, and assume that enzyme levels are mass densities. If the enzyme levels are molar concentrations, and differently weighted in the constraint, the weights can be taken into account by modifying the rule. To what

types of optimality problems does the rule apply? If the metabolic system is not an unbranched pathway, but a general network with one target flux, the rule will still hold (where the flux control coefficients refer to this target flux. This also holds for models with metabolite dilution.

However, there are some limitations. Obviously, the rule cannot hold in states in which control coefficients are not defined. This condition may seem unproblematic, but it is actually violated in the pathway with Michaelis-Menten rate laws, because in the limit $s_i \to \infty$ required by optimization, any variations of single enzyme levels would break the steady state. We discuss this in Appendix C.1. One way to avoid this problem is to add a bound on metabolite levels. For example, we may consider a generalized density constraint on enzyme levels and metabolite concentrations

$$\text{Maximize } \mathbf{z} \cdot \mathbf{v} \quad \text{s.t.} \quad \mathbf{a} \cdot \boldsymbol{\varepsilon} + \mathbf{b} \cdot \mathbf{s} \leq \rho$$

with a linear flux objective $\mathbf{a} \cdot \boldsymbol{\varepsilon}$ instead of a single target flux, and enzyme weights a_l and metabolite weights b_i, and an upper bound ρ on the molecule density. Problems with this constraint lead to a generalized form of the enzyme-control rule. For an unbranched pathway with equal weights a for enzymes and b for metabolites (e.g. molecular masses), we obtain the new enzyme-control rule

$$\varepsilon_l^* = \varepsilon_{\text{tot}}^* \; C_l^{J*} - J^* \; \frac{b}{a} \; \sum_i C_l^{s_i *}. \tag{32}$$

where $\varepsilon_{\text{tot}}^*$ is the (non-fixed) sum of enzyme levels emerging in the optimal state and the symbols C_l^{J*} and $C_l^{s_i *}$ denote, respectively, the flux and control coefficients in the optimal state. For derivation and details, see Appendix C.2.

The original enzyme-control rule has a direct and useful consequence. Since enzyme levels (in optimal states) are proportional to flux control coefficients, they need to satisfy a connectivity theorem. Connectivity theorems relate the elasticities of a given metabolite i to the control coefficients of reactions around this metabolite. In the case of flux control coefficients, the right-hand side of the theorem is zero, and we obtain a simple equation for the optimal enzyme levels around a metabolite i:

$$\sum_j \varepsilon_j \cdot E_{j,i} = 0\,, \tag{33}$$

where $E_{j,i}$ denotes the unscaled elasticity between metabolite i and reaction j.

We call this formula the enzyme-elasticity rule (see Appendix D). A similar rule for small adaptations of enzyme levels, instead of the enzyme levels themselves, has been shown in [27]. For a linear pathway, the enzyme-elasticity rule yields a simple result: in an optimal state, for each metabolite i and its producing reaction $j = i - 1$ and its consuming reaction i, the ratio of enzyme levels $\frac{\varepsilon_i}{\varepsilon_j}$ must be equal to the absolute inverse ratio of the elasticities $|E_{c_i}^{v_j} / E_{c_i}^{v_i}|$. An example of this, with the mass-action rate law, is derived in Appendix D. In contrast to the enzyme-control rule (which yields only one equation for the entire pathway), the enzyme-elasticity rule yields an equation for every internal metabolite, which determines the ratio of the enzyme levels around this metabolite. Together with the known sum of enzyme levels, these rules determine the enzyme profile completely (of course, given all the elasticities in the optimal state).

2.4.2 Analytic formulae for metabolic control

We learned that enzyme levels in optimal states (with a bound on total enzyme, and maximizing the steady-state flux) satisfy two general laws, the enzyme-control rule and the enzyme-elasticity rule. Hence, if we have a formula for the optimal enzyme levels and if the enzyme-control rule applies, we obtain the flux control coefficients. Moreover, the enzyme-elasticity rule (which comes from the connectivity theorem) relates the optimal enzyme levels to elasticities. These formulae can be trusted, but for didactic reasons we computed control coefficients for some of the rate laws analytically (also demonstrating different ways to compute them) and compared them to optimal enzyme levels to check the rules. Here we summarize this briefly; details can be found in the Appendix.

- **Michaelis-Menten rate law** In metabolic pathways with irreversible rate laws, the first reaction has full flux control and therefore a flux control coefficient of 1, while the remaining reactions of no flux control and therefore flux control coefficients of 0. But under an optimization, this changes. All metabolite levels go to infinity and all elasticities go zero. Mathematically, the optimal state doe not exist (because infinite concentrations are not real numbers), but even if we assume that enzymes can be saturated completely, any enzyme variation would break the steady state. This means that the control coefficients are not defined, and

so the enzyme-control rule does not apply (see Appendix C.1). To obtain meaningful results, we therefore considered an optimality problem with enzyme and metabolite constraints. In this case, metabolite levels remain finite and the control coefficients remain defined.

- **Thermodynamic rate law** For pathway models with the thermodynamic rate law, we can derive a simple formula for the flux control coefficients that contains only the enzyme levels, the k^{cat} values, and the flux. Even if we don't have an explicit formula for the flux, we can verify that the flux control coefficients satisfy summation and connectivity theorems and that the enzyme-control rule is satisfied. If we do not have a closed formula for the flux as a function of enzyme levels, how can we find a closed formula for the flux control coefficients? Such a formula can be obtained with a trick. From the rate law, we obtain a relation between enzyme levels, external metabolite levels, kinetic constants, and flux:

$$\prod_{i=1}^{n}(1 - J/\varepsilon_i k_i^{\text{cat}}) = \frac{s_n}{s_0}\frac{1}{K_{\text{tot}}^{\text{eq}}} \tag{34}$$

(see Eq. (71) in the Appendix for the derivation). While we cannot solve this for the flux J directly, we can obtain the response coefficients $R_{\varepsilon_l}^J = \partial J/\partial \varepsilon_l$ by implicit differentiation. This yields the flux control coefficients (derivation in Appendix F.3.2):

$$C_l^J = \frac{(\varepsilon_l\ k_l^{\text{cat}} - J)^{-1}}{\sum_{i=1}^{n}(\varepsilon_i\ k_i^{\text{cat}} - J)^{-1}}. \tag{35}$$

The control coefficients are proportional to $(\varepsilon_l\ k_l^{\text{cat}} - J)^{-1}$ and normalized to a sum of 1 as required by the summation theorem. For didactical reasons, we checked the connectivity theorem (Appendix F.3.1) and the enzyme-control rule (Appendix F.3.4).

- **Mass-action rate law** With the reversible mass-action rate law, we can analytically compute the flux control coefficients and verify that they satisfy summation and connectivity theorems. To verify that the enzyme-control rule is satisfied in the optimal state, we use the explicit formula (26) for the steady-state flux and take derivatives to obtain the flux control coefficients

$$C^J = \frac{\gamma_l/\varepsilon_l}{\sum_j \gamma_j/\varepsilon_j} \tag{36}$$

with γ_l defined as above (derivation in Appendix F.4.2). The coefficients sum to 1 as required by the summation theorem. Again, we verified the connectivity theorem (Appendix F.4.4) and the enzyme-control rule (Appendix F.4.5) for didactical reasons.

2.5 A cell model with enzyme kinetics and metabolite constraints

As an illustrative example, we apply our formulae to a simple model of a growing cell, describing transport processes, metabolism, and macromolecule synthesis, respectively, by lumped reactions. Obviously, none of the rate laws considered throughout this work can fully capture the dynamics of growing cells. For example, having only one single reaction representing metabolism is a gross oversimplification (and likewise for transport and translation). Nevertheless, we might still be able to draw insights from this model if we make the right assumptions. This approach has been successful in the past: by considering simple cell models and assuming that enzyme efficiencies are constant (i.e. completely independent of growth rate and metabolite concentrations), Basan et al. [10] were able to show how overflow metabolism in *E. coli* corresponds to a proteome-efficient allocation of enzymes. Here we ask how the predictions would change by accounting for enzyme kinetics. Instead of assuming that enzyme efficiencies are constant and given, we consider a model in which protein allocation and metabolite concentrations are always optimized in order to maximize the growth rate.

In our simple cell model (Figure 5) production processes are represented by an unbranched chain of reactions, describing overall transport, overall metabolism, and protein synthesis. The steady-state fluxes are proportional to the cell growth rate and optimized under constraints: an upper bound on the sum of enzyme levels (ε_{tot}) and another one on metabolite levels (s_{tot}). For convenience, in this section we will replace the pathway flux J^* with the standard symbol for growth rate – μ. It should be noted that dilution of the intermediate compounds is ignored even though in reality some of them can have a high total concentration (e.g., the precursors in S_2) and thus a considerable growth dilution rate which affects protein allocation strategies.

For our model, we chose to use the Michaelis-Menten approximation, so that metabolite concentrations (as extra variables) can be adjusted and become part of the optimization problem. We then apply the formulae derived in

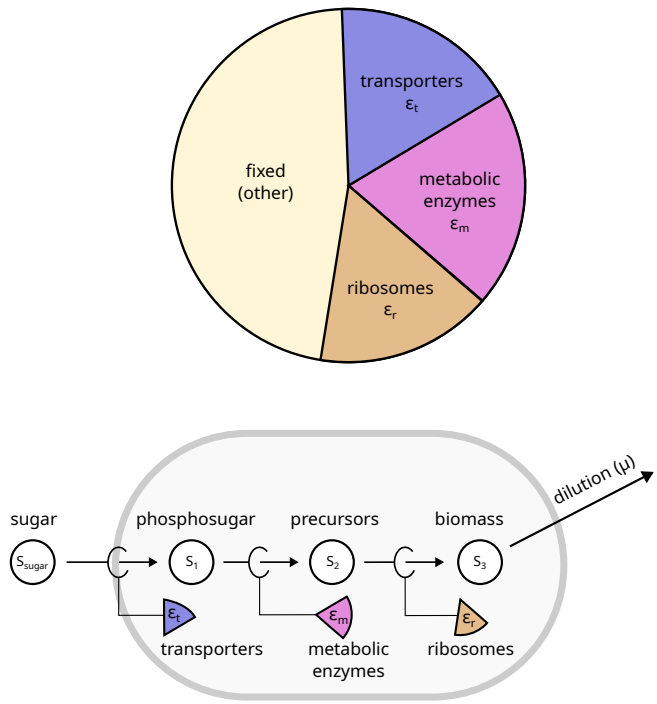


Figure 5: A growing cell described as a chain of reactions. Top: Inspired by the cell model in [9] describing "bacterial growth laws", we assume a fixed protein budget consisting of a fixed protein fraction and variable fractions for transporters, metabolic enzymes, and ribosomal proteins, but with a fixed sum. The proportions of the variable fractions are chosen to maximize growth. Bottom: The growing cell is described as a chain of reactions, each catalyzed by one of the protein fractions. For a constant biomass production, the three reactions must be in steady state.

the previous sections of this paper to find the optimal allocation of enzymes and thereby maximize the growth rate of the cell. Importantly, all calculations are completely based on analytic expressions.

In section 2.3.1, we derived a formula for the optimal allocation and maximal flux in the case where all metabolite levels are free variables (including s_0, which is denoted here by s_{sugar}). However, since in this cell model s_{sugar} represents the concentration of an external nutrient that is not subject to optimization, we would like to treat it as a constant system parameter (and later show how the optimum responds to changes in it). Fortunately, adding this assumption changes the optimal solution only slightly, as described in Appendix F.1.4, and the optimal growth rate as a function of s_{sugar} can be written in the following simple form:

$$\mu = \mu^{\max} \cdot \frac{s_{\text{sugar}}}{s_{\text{sugar}} + K_{\text{Monod}}} \tag{37}$$

where we define

$$\begin{aligned} \mu^{\max} &\equiv \frac{\varepsilon_{\text{tot}}}{1/k_{\text{t}}^{\text{cat}} + 1/k_{\text{m}}^{\text{cat}} + 1/k_{\text{r}}^{\text{cat}} + \left(\sqrt{K_{\text{M;m}}/k_{\text{m}}^{\text{cat}}} + \sqrt{K_{\text{M;r}}/k_{\text{r}}^{\text{cat}}}\right)^2 / s_{\text{tot}}} \\ K_{\text{Monod}} &\equiv \frac{K_{\text{M;t}}/k_{\text{t}}^{\text{cat}}}{1/k_{\text{t}}^{\text{cat}} + 1/k_{\text{m}}^{\text{cat}} + 1/k_{\text{r}}^{\text{cat}} + \left(\sqrt{K_{\text{M;m}}/k_{\text{m}}^{\text{cat}}} + \sqrt{K_{\text{M;r}}/k_{\text{r}}^{\text{cat}}}\right)^2 / s_{\text{tot}}} \,. \end{aligned} \tag{38}$$

As implied by the symbol for K_{Monod}, this form corresponds to empirical observations by Monod [19] (and further followed up by others [28]) which stated that growth rate increases with the nutrient concentration until reaching a saturation level where growth is fastest. Interestingly, the value of K_{Monod} (e.g. the level of s_{sugar} for which growth rate is half of its maximum) is not determined solely by the kinetics of the transporter.

To better understand how changes in the model parameters affect the growth rate, we considered a toy example, where all constants are set to a default value of 1 (except for $K_{\text{M;m}} = 2$). In Figure 6, we plot the value of μ as a function of s_{sugar} (based on Eq. (37-38)), each time changing one of the parameters. In almost all cases, we find a trade-off between growth and binding affinity, namely that improving the kinetics (or relaxing a constraint) improves the maximal growth rate, while increasing K_{Monod} (i.e. making it worse since higher K_{Monod} means higher a concentration of sugar is required to reach the same growth rate). The three exceptions are the response

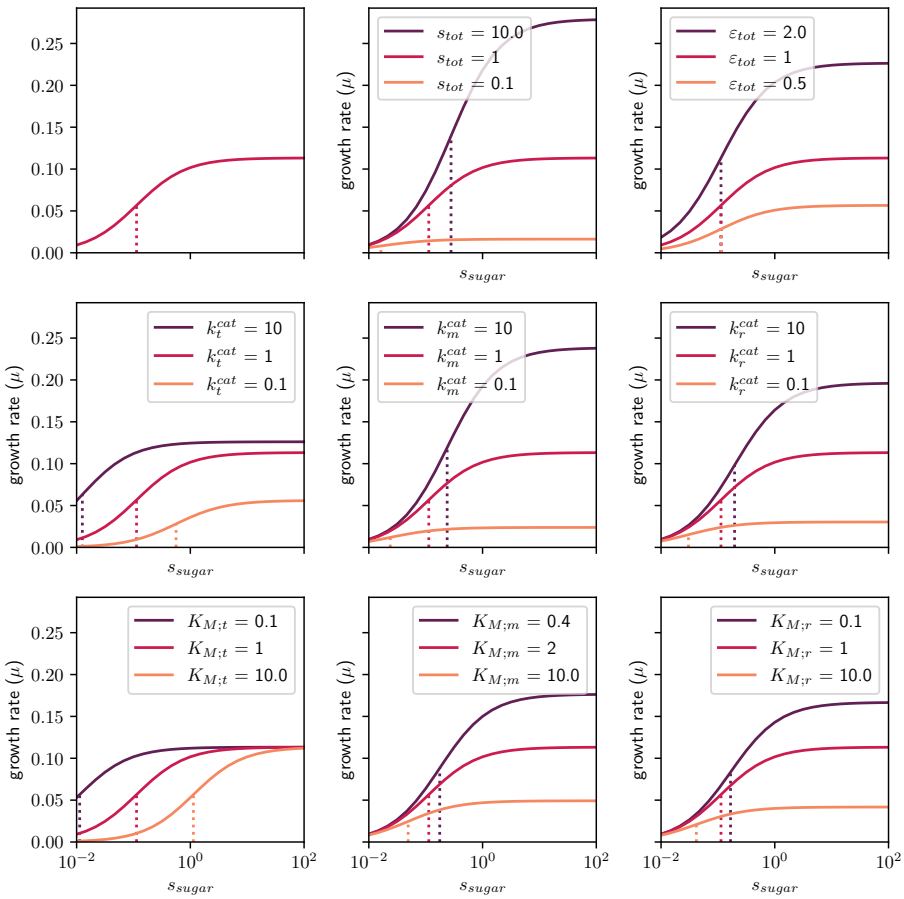


Figure 6: The optimal achievable flux – i.e. the growth rate (μ) – for a given total enzyme ε_{tot}. The model contains three irreversible steps based on Michaelis-Menten kinetics. The default parameters are $k_t^{\text{cat}} = k_m^{\text{cat}} = k_r^{\text{cat}} = K_{\text{M;t}} = K_{\text{M;r}} = 1$, $K_{\text{M;m}} = 2$, and the upper bounds are set to $s_{\text{tot}} = 1$, $\varepsilon_{\text{tot}} = 1$. The panel on the upper-left corner plots the maximal flux as a function of s_{sugar} for the default parameters. In each other panel, only one of the parameters is changed. The vertical dotted lines mark the K_{Monod} (i.e. the concentration of s_{sugar} where the flux is 50% of the maximum). Note that the x-axis is plotted in log-scale, and therefore the curves have a sigmoid-shape (rather than the familiar Monod-style). See Figure S1 for plots of the enzyme demands.

to changes in ε_{tot} which only affects μ^{max}, in K_{t}, which only affects the K_{Monod}, and in $k_{\text{t}}^{\text{cat}}$, which can improve both growth parameters at the same time. Indeed, by observing the formulae in Eq. (38), one can see that all parameters that are only in the denominator should affect μ^{max} and K_{Monod} in the same fashion (which leads to a trade-off), while the three parameters in the numerators have each their own unique effect.

This result might have implications for the design of optimal enzyme regulation. As the rate of sugar transport is not affected by any of the metabolite concentrations inside the cytoplasm, the transport reaction acts as an information buffer in the sense that any changes in the external sugar concentration do not affect the optimal allocation of enzymes (and metabolites) within the cytoplasm. The only thing that changes is the relative abundance of the transporter versus all other (cytoplasmic) enzymes. This rule becomes clear when plotting the optimal enzyme allocations as a function of the pathway flux (i.e. the growth rate), as shown in Figure 7. The two cytoplasmic fractions (metabolic enzymes and ribosomes) change proportionally and appear as straight lines that cross the axis origin. Since ε_{tot} is a constant, the remainder (i.e. the level of the transporter) also decreases linearly with growth rate. These three straight lines extend only up to a point (where $\mu = \mu^{\text{max}}$). This limit represents the fully saturated transporter, which has an efficiency of $k_{\text{m}}^{\text{cat}}$ and therefore an optimal abundance of $\mu^{\text{max}}/k_{\text{m}}^{\text{cat}}$. Higher concentrations of s_{sugar} would not generate any further benefits.

In fact, the principle described above can be shown to be true in a much more general case. As long as one section of the metabolic network is "isolated" from the rest (i.e., connected only by irreversible steps), changes in the upstream parameters will affect the incoming flux, but would not change the optimal allocation of enzymes (and also not the metabolite concentrations). Curiously, in such cases our result predict, based on an underlying kinetic model, that the optimal protein efficiencies are independent of the external glucose concentration and, as

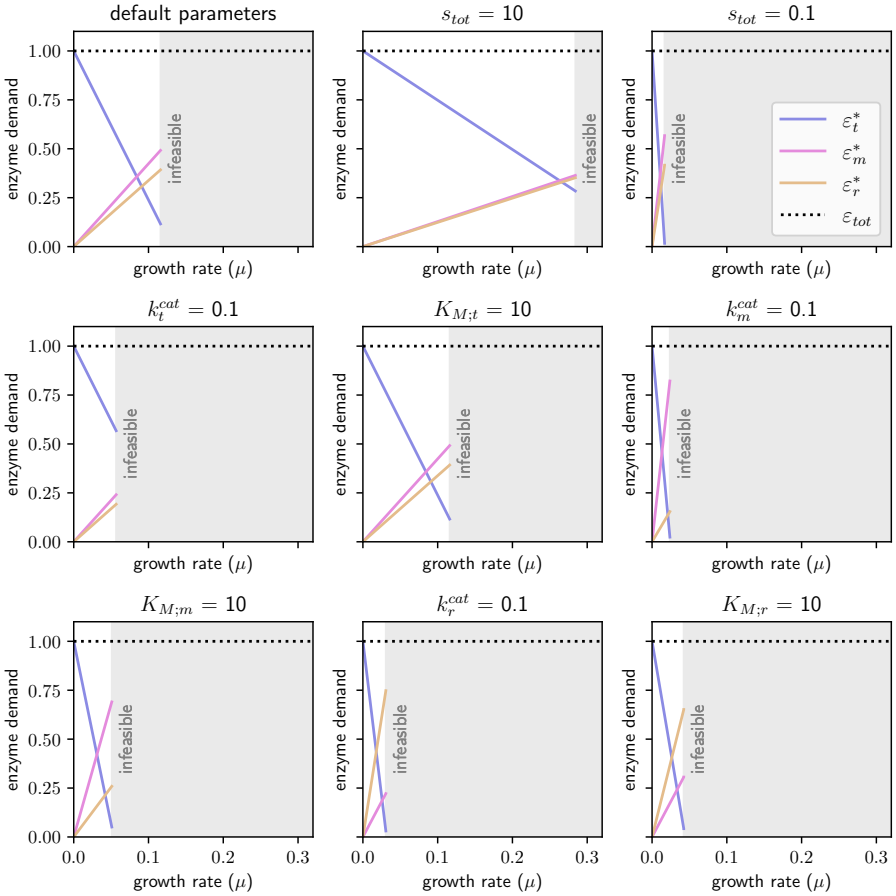


Figure 7: The abundances of the three enzyme sectors ($\varepsilon_{\mathrm{t}}^*$, $\varepsilon_{\mathrm{m}}^*$, and $\varepsilon_{\mathrm{r}}^*$) as a function of the flux (J^*) when the allocation is optimal. The default parameters are $k_t^{\mathrm{cat}} = k_m^{\mathrm{cat}} = k_r^{\mathrm{cat}} = 1$, $K_{\mathrm{M;t}} = K_{\mathrm{M;r}} = 1$, $K_{\mathrm{M;m}} = 2$ (we chose another value in order prevent the curves for ε_{m} and ε_{r} from overlapping), and the upper bounds are set to $s_{\mathrm{tot}} = 1$, $\varepsilon_{\mathrm{tot}} = 1$. In each panel only one of these parameters is changed. The gray shaded area marks growth rates that are higher than the maximum, μ^{max}. The linear protein allocation curves are caused by the irreversible transport step. In the optimization, this leads to constant optimal metabolite concentrations, and the protein fractions $\varepsilon_{\mathrm{m}}^*$ and $\varepsilon_{\mathrm{r}}^*$ must change proportionally with the flux. The line slopes are given by the enzyme efficiencies at the optimal metabolite concentrations, and the transporter fraction – the residual of the protein budget – must be a falling straight line.

a consequence, protein fractions vary linearly. Thus, our model predictions are very similar to ones from previous proteome allocation models that are based on empirical observations.

A series of irreversible steps is only one option for a simple model of cell growth. Alternatively, one might consider cells that live close to chemical equilibrium, in environments that provide only a very small overall thermodynamic driving force. In this case, applying our thermodynamic rate law with the approximation from Eq. (86) yields another simple "growth formula":

$$\begin{aligned} \mu(s_{\mathrm{sugar}}) &= \mu^{\mathrm{max}} \left(1 - \left(\frac{A}{s_{\mathrm{sugar}}}\right)^B\right) \\ \mu^{\mathrm{max}} &\equiv \varepsilon_{\mathrm{tot}} \left(\sum_i 1/k_i^{\mathrm{cat}}\right)^{-1} \end{aligned} \tag{39}$$

where $A \equiv s_n K_{\mathrm{tot}}^{\mathrm{eq}}$ and $B \equiv \left(\sum_i 1/k_i^{\mathrm{cat}}\right) / \left(\sum_i 1/\sqrt{k_i^{\mathrm{cat}}}\right)^2$. Again, the parameters of this curve depend on all the k_i^{cat} values. In contrast to the Michaelis-Menten based growth formula in Equation (38), this formula is also sensitive to product concentrations. This can be interesting because close to chemical equilibrium, product accumulation can decrease growth in a way that cannot be bypassed by any regulatory mechanisms. Instead, cells would profit from "cleaner" strains that remove the accumulating product, which has been proposed as a general mechanism that encourages symbiosis [29].

3 Discussion

In this study we addressed the enzyme allocation problem analytically by considering several different rate-law approximations. Along the way, we learned that each approximation comes with its own idiosyncrasies. For instance, a model with Michaelis-Menten kinetics requires an extra metabolite constraint to obtain a valid optimal state, and the *thermodynamic* rate law requires inverting a function in order to find the value of the Lagrange multiplier. Fortunately, there is a very good approximate solution that works well across the entire range of driving forces.

For historical reasons, we solved the optimality problem under the assumption that there is a bound on the sum of enzyme abundances – $\sum_i \varepsilon_i \leq \varepsilon_{\text{tot}}$ – without explicitly specifying the units. The common interpretation would be that ε_i are *molar* concentrations (where the symbol k_i^{cat} represents the turnover numbers in units of 1 over time). However, the total *mass* concentration of all pathway enzymes may be a better proxy for cost since crowding effects in the cytoplasm are often limiting the total mass concentration of all soluble proteins, including most enzymes. Furthermore, fast growing cells are often limited by the rate of protein elongation, and the molecular mass is nearly proportional to the gene length. Therefore, we usually think of ε_i as *mass* concentrations (e.g. in $\text{g} \times \text{m}^{-3}$) and of k_i^{cat} as specific activities (e.g. in $\text{mol} \times \text{s}^{-1} \times \text{g}^{-1}$). Nevertheless, the analytic derivation and solutions provided here are agnostic to the choice of units and are equally valid for both these interpretations. Moreover, one could imagine a completely different set of linear weights for the enzyme cost function (i.e. $\sum_i w_i \, \varepsilon_i \leq \varepsilon_{\text{tot}}$, as in Figure 5). If one things of the weights as scaling factors for the k^{cat} values, the provided solutions will still hold (while making use of the new "effective" k^{cat} values).

Previous studies focused on the "mass-action" rate law, justifying it by saying that the general reversible form derived by Haldane can be approximated at the limit of low concentrations. However, it is quite rare to have all the substrates and products of an enzyme at concentrations that are way below their K_{M} values [30]. Furthermore, the limit of all reactants concentrations going to zero is not very meaningful because, in the first place, Haldane derived his rate law assuming enzymes are much less abundant than metabolites [31]. Interestingly, just being close-to-equilibrium is not enough for this approximation, since the product can still affect the rate non-linearly via η^{kin} (see Figure 2). Here we tried to consider more comprehensively all the different approximations that yield an analytic solution to optimal enzyme allocation in unbranched pathways (see Table 1).

One of these approximations is the Michaelis-Menten rate law, which is a widely used in enzymatic assays and metabolic modeling of irreversible reactions. Curiously, using it for the simple optimality problem with linear chains leads to paradoxical results: the metabolite concentrations go to infinity, the elasticities vanish, and flux control coefficients are not defined. For solving this problem, we introduced an new upper bound on the total metabolite concentration in order to obtain realistic results and derive analytic solutions based on this rate law. Although it is very reasonable to assume that concentrations of small molecules (and not just enzymes) are restricted in cells, this fact is often ignored in metabolic models. One of the rare cases where this constraint was taken into account is the work of Dourado et al. [32], who found empirical evidence to the fact that there is a balance between enzymes and substrates when minimizing the total mass concentration. In addition to an analytic solution, we also found a new enzyme-control rule for models with a constraint on enzyme and metabolite concentrations: in this case, the enzyme amounts do not only reflect the flux control coefficients, but a sum of flux and concentration control coefficients see Appendix C.2).

Besides "mass-action" and Michaelis-Menten kinetics, we discussed a solution for one other rate law which we call *thermodynamic*, as the rate is only affected by metabolite levels through the thermodynamic driving force (i.e. ignoring any saturation effects). One advantage of this rate law is that it does not require knowing the K_{M} values (which are often difficult to come by). The thermodynamic-only approximation was also used for deriving the Max-min Driving Force (MDF) method F.3.5, which similarly aimed to quantify the efficiency of metabolic pathways [33]. But, unlike the solution presented here, MDF does not explicitly optimize a simplified version of a kinetic rate law, but rather applies a heuristic based on the assumption that the lowest driving forces are mainly responsible for increased enzyme demands, and the total driving force should therefore be as evenly distributed as possible. On the other hand, the advantage of the MDF method is that it takes metabolite concentration bounds into consideration.

All throughout this paper, we only considered pathways with uni-uni reactions (one substrate, one product, with stoichiometric coefficients of 1). In reality, many reactions involve cofactor pairs or other substrates or products. Instead of deriving results for this general case, we assumed that these extra reactants exist, but with known concentrations. In this case, the rate laws contain extra terms, but these terms can be rearranged to yield the same simple formulae as in the uni-uni case, but with effective kinetic constants. We demonstrate this for the case of two substrates and two products following convenience kinetics [34] in Appendix F.6. Notably, this logic also applies to reactions with more than two substrates and products as well as to enzyme activation or inhibition with

constant activator or inhibitor levels.

This paper can be seen as an exercise in solving the enzyme allocation problem and describing the optimal states analytically using MCA. Although some might argue that the required approximations are not realistic, they do represent a step forward compared to the very common approach of assuming that metabolites have no effect on enzyme efficiency at all. On the other hand, adding metabolite concentrations as extra variables greatly increases the complexity of models and typically renders them unsolvable. Therefore, the solutions provided here might be handy, as the assumption of metabolite steady-state combined with the optimality argument give us analytic expressions that only depend on the initial and final metabolites in the pathway. We demonstrate this result using a toy example for a course-grained model of cell growth, and show how the analytic solutions provide valuable insights about the effects of changes in each parameter – all this without the need to simulate the metabolic network or use non-linear solvers. One could imagine a scenario where metabolic engineers, while trying to improve a biochemical pathway, have to decide in which enzyme to invest their time. Although MCA can perfectly serve this purpose while covering more scenarios, it requires tools and language that can seem opaque to people lacking the proper mathematical background. The solutions provided here should be more approachable and serve as simple guidelines while capturing a wider range of cases than previously explored in the literature. We hope that future studies will apply and extend this approach to other, more complex models.

Acknowledgments

This article has originally been conceived as a chapter for the free textbook "Economic Principles in Cell Biology". A shorter chapter version will be published there. We thank Daan de Groot and Avi Flamholz for their insightful comments on our manuscript.

References

[1] Robert Rosen. *Optimality principles in biology.* Springer, 1967.

[2] Erez Dekel and Uri Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436:588–692, 2005. doi: 10.1038/nature03842.

[3] Robert Schütz, Nicola Zamboni, Mattia Zampieri, Matthias Heinemann, and Uwe Sauer. Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–604, May 2012.

[4] Qasim K. Beg, Alexei Vazquez, Jason Ernst, Marcio A. de Menezes, Ziv Bar-Joseph, Albert-László Barabási, and Zoltán N. Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *Proc. Natl. Acad. Sci. U. S. A.*, 104(31):12663–12668, July 2007.

[5] Naama Tepper, Elad Noor, Daniel Amador-Noguez, Hulda S Haraldsdóttir, Ron Milo, Josh Rabinowitz, Wolfram Liebermeister, and Tomer Shlomi. Steady-State metabolite concentrations reflect a balance between maximizing enzyme efficiency and minimizing total metabolite load. *PLoS One*, 8(9):e75370, September 2013. doi: 10.1371/journal.pone.0075370.

[6] Meike T Wortel, Elad Noor, Michael Ferris, Frank J Bruggeman, and Wolfram Liebermeister. Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. *PLoS Comput. Biol.*, 14(2): e1006010, February 2018. doi: 10.1371/journal.pcbi.1006010.

[7] Avi Flamholz, Elad Noor, Arren Bar-Even, Wolfram Liebermeister, and Ron Milo. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci. U. S. A.*, 110(24):10039–10044, June 2013. doi: 10.1073/pnas.1215283110.

[8] Elad Noor, Avi Flamholz, Arren Bar-Even, Dan Davidi, Ron Milo, and Wolfram Liebermeister. The protein cost of metabolic fluxes: Prediction from enzymatic rate laws and cost minimization. *PLoS Comput. Biol.*, 12 (11):e1005167, November 2016. doi: 10.1371/journal.pcbi.1005167.

[9] Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330:1099, 2010.

[10] Markus Basan, Sheng Hui, Hiroyuki Okano, Zhongge Zhang, Yang Shen, James R. Williamson, and Terence Hwa. Overflow metabolism in Escherichia coli results from efficient proteome allocation. *Nature*, 528:99, 2015. doi: 10.1038/nature15765.

[11] Arren Bar-Even, Elad Noor, Nathan E. Lewis, and Ron Milo. Design and analysis of synthetic carbon fixation pathways. *Proceedings of the National Academy of Sciences*, 107(19):8889–8894, 2010. doi: 10.1073/pnas.0907176107.

[12] Hannes Löwe and Andreas Kremling. In-Depth computational analysis of natural and artificial carbon fixation pathways. *BioDesign Research*, 2021, September 2021. doi: 10.34133/2021/9898316.

[13] Reinhart Heinrich and Stefan Schuster. *The Regulation of Cellular Systems*. Springer, New York, NY, 1996 edition, December 2012. doi: 10.1007/978-1-4613-1161-4.

[14] Henrik Kacser and Jaye A. Burns. The control of flux. *Symp. Soc. Exp. Biol.*, 27:65–104, 1973.

[15] Reinhart Heinrich and Tom A. Rapoport. A linear steady-state treatment of enzymatic chains. general properties, control and effector strength. *Eur. J. Biochem.*, 42(1):89–95, 1974.

[16] Guy C. Brown. Total cell protein concentration as an evolutionary constraint on the metabolic control distribution in cells. *J. Theor. Biol.*, 153(2):195–203, November 1991.

[17] Reinhart Heinrich and Edda Klipp. Control analysis of unbranched enzymatic chains in states of maximal activity. *J. Theor. Biol.*, 182(3):243–252, 1996.

[18] Edda Klipp and Reinhart Heinrich. Competition for enzymes in metabolic pathways: implications for optimal distributions of enzyme concentrations and for the distribution of flux control. *BioSystems*, 54:1–14, 1999.

[19] Jacques Monod. The growth of bacterial cultures. *Annual Reviews in Microbiology*, 3(1):371–394, 1949.

[20] Stephen G. Waley. A note on the kinetics of multi-enzyme systems. *Biochem. J*, 91(3):514–517, June 1964.

[21] Jaye A. Burns. *Studies on complex enzyme systems*. PhD thesis, University of Edinburgh, 1971.

[22] Stefan Schuster and Reinhart Heinrich. Minimization of intermediate concentrations as a suggested optimality principle for biochemical networks. *J. Math. Biol.*, 29(5):425–442, April 1991. doi: 10.1007/BF00160471.

[23] Elad Noor, Avi Flamholz, Wolfram Liebermeister, Arren Bar-Even, and Ron Milo. A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects. *FEBS Lett.*, 587(17):2772–2777, September 2013. doi: 10.1016/j.febslet.2013.07.028.

[24] Leonor Michaelis and Maud L. Menten. Die Kinetik der Invertinwirkung. *Biochem. Z.*, 49, 1913.

[25] R. Heinrich and T.A. Rapoport. Linear theory of enzymatic chains: its application for the analysis of the crossover theorem and of the glycolysis of human erythrocytes. *Acta Biol. Med. Germ.*, (31):479–494, 1973.

[26] Hugo Dourado, Matteo Mori, Terence Hwa, and Martin J. Lercher. On the optimality of the enzyme-substrate relationship in bacteria. *PLoS Biol*, 19(10):e3001416, 2021.

[27] Wolfram Liebermeister, Edda Klipp, Stefan Schuster, and Reinhart Heinrich. A theory of optimal differential gene expression. *BioSystems*, 76:261–278, 2004.

[28] Talaat E. Shehata and Allen G. Marr. Effect of nutrient concentration on the growth of Escherichia coli. *J. Bacteriol.*, 107(1):210–216, July 1971.

[29] Tobias Großkopf and Orkun S. Soyer. Microbial diversity arising from thermodynamic constraints. *ISME J*, page 2725–2733, 2016. doi: 10.1038/ismej.2016.49.

[30] Bryson D. Bennett, Elizabeth H. Kimball, Melissa Gao, Robin Osterhout, Stephen J. Van Dien, and Joshua D. Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol.*, 5(8):593–599, August 2009.

[31] Sanderson John Burdon Haldane. Chapter V. The course of enzymatic reactions, and its mathematical theory. *Enzymes*, pages 74–92, 1930.

[32] Hugo Dourado, Veronica G Maurino, and Martin J Lercher. Enzymes and substrates are balanced at minimal combined mass concentration in vivo. *bioRxiv*, page 128009, April 2017. doi: 10.1101/128009.

[33] Elad Noor, Arren Bar-Even, Avi Flamholz, Ed Reznik, Wolfram Liebermeister, and Ron Milo. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput. Biol.*, 10(2):e1003483, February 2014. doi: 10.1371/journal.pcbi.1003483.

[34] Wolfram Liebermeister and Edda Klipp. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor. Biol. Med. Model.*, 3(1):41, 2006. doi: 10.1186/1742-4682-3-41.

[35] Wolfram Liebermeister, Jannis Uhlendorf, and Edda Klipp. Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics*, 26(12):1528–1534, June 2010. doi: 10.1093/bioinformatics/btq141.

[36] Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

[37] Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60, 2018.

[38] Wolfram Liebermeister and Elad Noor. The enzyme cost of given metabolic flux distributions, as a function of logarithmic metabolite levels, is convex. *arXiv: 1501. 02454*, 2015. doi: 10.48550/arXiv.1501.02454.

A Code availability

All the data, scripts, and resulting plots used in this work are freely available at https://gitlab.com/elad.noor/optimal_enzyme_investment under an MIT open-source license.

B Symbols

Mathematical symbols are listed in Table S1.

Parameter name	symbol
Turnover constant	k^{cat}
Turnover constant (forward)	k_{+}^{cat}
Turnover constant (backward)	k_{-}^{cat}
Thermodynamic equilibrium constant	K^{eq}
Michaelis-Menten constant	K_{M}
Michaelis-Menten constant (substrate)	K_{S}
Michaelis-Menten constant (product)	K_{P}
Thermodynamic Driving Force	θ
Reaction rate	v
Pathway flux	J
Enzyme concentration/demand	ε

Table S1: The names of the variables and constants used in this work, along with their mathematical symbol.

C Revisiting the enzyme-control rule

C.1 Why the enzyme-control rule fails with Michaelis-Menten kinetics

The enzyme-control rule fails in a very simple case, the unbranched metabolic pathway with Michaelis-Menten rate laws. We can easily see this. Since the rate law is irreversible, the rate in each reaction depends only on the substrate and enzyme level (and not on anything that happens downstream). In steady state, all reaction rates are fixed by the stationary flux. Taken together, this means that the flux depends only on the initial pathway substrate and the level of the first enzyme, so this enzyme has full flux control and all other enzymes have zero control. However, according to the enzyme-control rule, this would mean that the first enzyme is the only enzyme that should be expressed. This is obviously wrong, because without the other enzymes, no steady flux would be possible. So how can this be?

In fact, optimal steady states in models with irreversible rate laws have some pathological properties. To maximize enzyme efficiency, all metabolite levels must go to infinity, the enzymes are completely saturated, and the metabolite elasticities are all zero. As a result, the flux in every reaction is given by $v_i = \varepsilon_i \; k_i^{\text{cat}}$, so the stationary flux completely determines all enzyme levels. If an enzyme level deviates from its prescribed value, its reaction rate will no longer match the other rates and the steady state breaks down[2]. Now we can see what went wrong: in the optimal state, any small change of an enzyme level will break the steady state; this means that the control coefficients are not even defined, and the enzyme-control rule cannot apply. All this does not only occur with the Michaelis-Menten rate law, but with any irreversible, saturable rate law.

In the case of the Michaelis-Mention rate law, the practical solution to this problem is simple. Since the optimal state requires infinite metabolite concentrations, and since this is not realistic, we need to change our model. We may either consider slightly reversible rate laws (which would penalize high product concentrations) or we introduce an extra density constraint that penalizes high metabolite levels.

[2]In fact, there are a number of similar arguments why the enzyme-control rules fails in this case: first, the state with infinite concentrations does not exist mathematically, so formally we cannot even refer to it as a metabolic state; second, even if this state existed, all elasticities would vanish, and the Jacobian matrix would not be invertible, so the control coefficient would not be defined; third, even if we argued that, logically, the first enzyme must have full flux control, the same argument would also apply to all other enzymes: each of the enzymes would have full flux control, thus violating the summation theorem.

C.2 An enzyme-control rule for models with general density constraints

One way to obtain meaningful optimal states, even with Michaelis-Menten kinetics, is to constrain the metabolite concentrations. The new optimality problem leads to different solutions and to a different enzyme-control rule.

Aside from high enzyme levels, also high metabolite concentrations may put a burden on cells, and considering this burden will lead to different solutions for enzyme levels. Setting up optimality problems with such a constraint raises a couple of questions: should we use separate constraints on enzyme levels and on metabolite concentrations, or just one density constraint on their sum? How should the different compounds be weighted? In fact, the resulting optimality problems are not very different, and we demonstrate here one specific approach: we describe all compounds (enzymes and metablites) by their concentrations and consider a bound on their weighted sum (where weights may have different meanings, for example molecular masses if the aim is to restrict the total mass density in the cell).

We thus consider the optimality problem

$$\text{Maximize } \mathbf{z} \cdot \mathbf{v} \quad \text{s.t.} \quad \mathbf{a} \cdot \boldsymbol{\varepsilon} + \mathbf{b} \cdot \mathbf{s} \leq \rho \tag{40}$$

where $\mathbf{v}$, $\mathbf{s}$, and $\boldsymbol{\varepsilon}$ denote reaction fluxes, metabolite concentrations, and enzyme levels, and where stationarity (of $\mathbf{v}$) and rate laws (between $\mathbf{v}$, $\mathbf{s}$, and $\boldsymbol{\varepsilon}$ must be satisfied as constraints. The maximization objective is a linear benefit function $\mathbf{z} \cdot \mathbf{v}$ scoring the fluxes with linear weights z_i, while the constrained density is linear in all the concentrations, with linear weights[3] a_l (for enzymes) and b_i (for metabolites), e.g. representing effective molecule sizes[4]. To derive the enzyme-control rule, we now reformulate the optimality problem with enzyme levels as the only free variables, and steady-state fluxes and concentrations depending on them (satisfying stationary and the rate laws). We obtain

$$\text{Maximize } \mathbf{z} \cdot \mathbf{v}^{\text{st}}(\boldsymbol{\varepsilon}) \quad \text{s.t.} \quad \mathbf{a} \cdot \boldsymbol{\varepsilon} + \mathbf{b} \cdot \mathbf{s}^{\text{st}}(\boldsymbol{\varepsilon}) \leq \rho \tag{41}$$

This optimality problem leads to a new enzyme-control rule (proof in Appendix F.7)

$$\boldsymbol{\varepsilon}^{*\top} \operatorname{diag}(\mathbf{a}) \operatorname{diag}(\mathbf{v})^{-1} = \left[\frac{1}{\lambda} \mathbf{z}^\top \mathbf{C}^{\mathbf{v}*} - \mathbf{b}^\top \mathbf{C}^{\mathbf{s}*} \right]. \tag{42}$$

In the general case (network with non-uniform flux distribution $\mathbf{v}$, a more complicated flux benefit function with derivatives $\frac{\partial z}{\partial v_i}$, and individual weights a_l and b_i for all enzymes and metabolites, all these vectors would enter the formula as weights.

In a simple linear pathway, our rule can be simplified. We assume the same steady-state flux J in all the reactions, a benefit function given by J, and a density constraint with equal weights a for all enzymes, and equal weight b for all the metabolites. In this case, Eq. (42) simplifies to

$$\varepsilon_l^* \propto C_l^{J*} - \lambda \, b \sum_i C_l^{s_i *} \tag{43}$$

which replaces the enzyme control rule (8). Given the sum of enyzme levels in the optimal state, called $\varepsilon_{\text{tot}}^*$, we can also write this as (see F.7)

$$\varepsilon_l^* = \varepsilon_{\text{tot}}^* \, C_l^{J*} - \frac{b}{a} \, J^* \sum_i C_l^{s_i *}. \tag{44}$$

Since the Lagrange multiplier λ arises from an upper bound in a maximization problem, it is negative in the optimal state, and the second term including the minus sign will be positive.

Compared to the enzyme-control rule (8), the right hand side now contains an extra term: it is given by the metabolite control coefficients, summed over all metabolites, and with a known prefactor. If the ratio of cost weights b/a is small compared to the flux, this term can be neglected and we obtain again our original enzyme-control rule.

[3]Instead of a linear flux objective, a nonlinear function $z(\mathbf{v})$ could be used. In this case, the rule (42) would remain the same, but with the gradient $\nabla_v z(\mathbf{v})$ replacing the weight vector $\mathbf{z}$.. Likewise, in a model with separate density constraints $\mathbf{a} \cdot \boldsymbol{\varepsilon} \leq \rho_\varepsilon$ and $\mathbf{b} \cdot \mathbf{s} \leq \rho_c$, we obtain a similar formulae, but with separate Lagrange multipliers for the two constraints.

[4]For a simple constraint on the total mass, the weights are simply molecular weights. This could be a proxy for excluded volume (which would still ignore, for example, hydration shells). However, a "density constraint" need not represent space demands; it may also be related to osmotic effects or opportunity costs (e.g. energy demand for production of compounds in growing cells). Therefore the meaning of the weights in the density constraint may differ from model to model.

C.3 A sufficient condition for stable states in unbranched pathways

Another possible reason why control coefficients may not be defined are unstable metabolic states. A metabolic steady state can be dynamically stable or unstable. In an unstable state, even small perturbations or chemical noise would drive the system far away from its original state. In MCA we are interested in stable states, because unstable states are not able to persist even under small chemical noise in the cell, and because mathematically, control coefficients for unstable states are not even defined. A metabolic state is asymptotically stable if the Jacobian $\mathbf{N}\ \mathbf{E}_{\mathrm{c}}$ in this state has only negative eigenvalues. But what does this mean in practice? In Appendix F.8, we derive a sufficient (but not necessary) condition for stable steady states in unbranched metabolic pathways, based on Gershgorin's theorem: A steady state in an unbranched metabolic pathway is stable if (but not only if) the first reaction 1 is not completely irreversible (i.e. it has a non-zero product elasticity), the last reaction is not completely saturated (i.e. it has a non-zero substrate elasticity), and in all reactions in between, the substrate elasticity is larger than the (absolute) product elasticity. This condition is not just a chek for our models, but it also describes a potential way in which cells could achieve stability. For a reversible mass-action rate law, the latter condition is satisfied if $K_i^{\mathrm{eq}} > 1$, and the first two conditions are generally satisfied.

D The enzyme-elasticity rule

The connectivity theorem for flux control coefficients

$$\forall i \quad \sum_j C_j^J \cdot E_{j,i} = 0 \tag{45}$$

relates the flux control coefficients C_j^J (between enzyme ε_j and the pathway flux J) to the unscaled elasticities $E_{j,i} = \frac{\partial v_j}{\partial s_i}$ (between metabolite s_i and reaction rate v_j). In optimal state, the enzyme-control rule tells us that control coefficients and enzyme levels must be proportional. Therefore we can replace C_j^J with ε_j in the equation above. For each metabolite i, we obtain an equation

$$\sum_j \varepsilon_j \cdot E_{j,i} = 0\,. \tag{46}$$

In a linear pathway, each metabolite has only one producing and one consuming reaction (with indices i and $i+1$, and we obtain the equality

$$\frac{\varepsilon_{i+1}}{\varepsilon_i} = -\frac{E_{s_i}^{v_i}}{E_{s_i}^{v_{i+1}}} = \frac{|E_{s_i}^{v_i}|}{E_{s_i}^{v_{i+1}}}. \tag{47}$$

In the last step, we used the fact that the backward (product) elasticity $E_{s_i}^{v_i}$ is negative and the forward (substrate) elasticity $E_{s_i}^{v_{i+1}}$ is positive.

Example: Mass-action rate law With the mass-action kinetic rate law in the form $v_j = \varepsilon_j \cdot (k_j s_{j-1} - k_{-j} s_j)$ (see Appendix F.4), Eq. (48) becomes:

$$\forall i \quad 0 = \varepsilon_i \cdot \frac{\partial v_i}{\partial s_i} + \varepsilon_{i-1} \cdot \frac{\partial v_{i-1}}{\partial s_i} = \varepsilon_i \cdot (-\varepsilon_i k_{-i}) + \varepsilon_{i+1} \cdot (\varepsilon_{i+1} k_{i+1}) \tag{48}$$

which by rearranging leads us to the relationship:

$$\forall i \quad \frac{\varepsilon_{i+1}}{\varepsilon_i} = \sqrt{\frac{k_{-i}}{k_{i+1}}} = \sqrt{\frac{\gamma_{i+1}}{\gamma_i}} \tag{49}$$

(where we remind ourselves of the definition $\gamma_i \equiv k_{-i}^{-1} \prod_{j=1}^{i} k_{-j}/k_j$). Applying this equation recursively, we can see that $\varepsilon_i/\varepsilon_1 = \sqrt{\gamma_i/\gamma_1}$, which means that ε_i is proportional to $\sqrt{\gamma_i}$. Since we know that the sum of all enzyme demands gives $\varepsilon_{\mathrm{tot}}$, we can obtain the explicit formula for ε_i:

$$\varepsilon_i = \varepsilon_{\mathrm{tot}} \frac{\sqrt{\gamma_i}}{\sum_j \sqrt{\gamma_j}} = \varepsilon_{\mathrm{tot}} \sqrt{\frac{\gamma_i}{||\boldsymbol{\gamma}||_{\frac{1}{2}}}}\,. \tag{50}$$

This is the exact same result we got by directly optimizing the pathway flux (see Equation (108) in Appendix F.4).

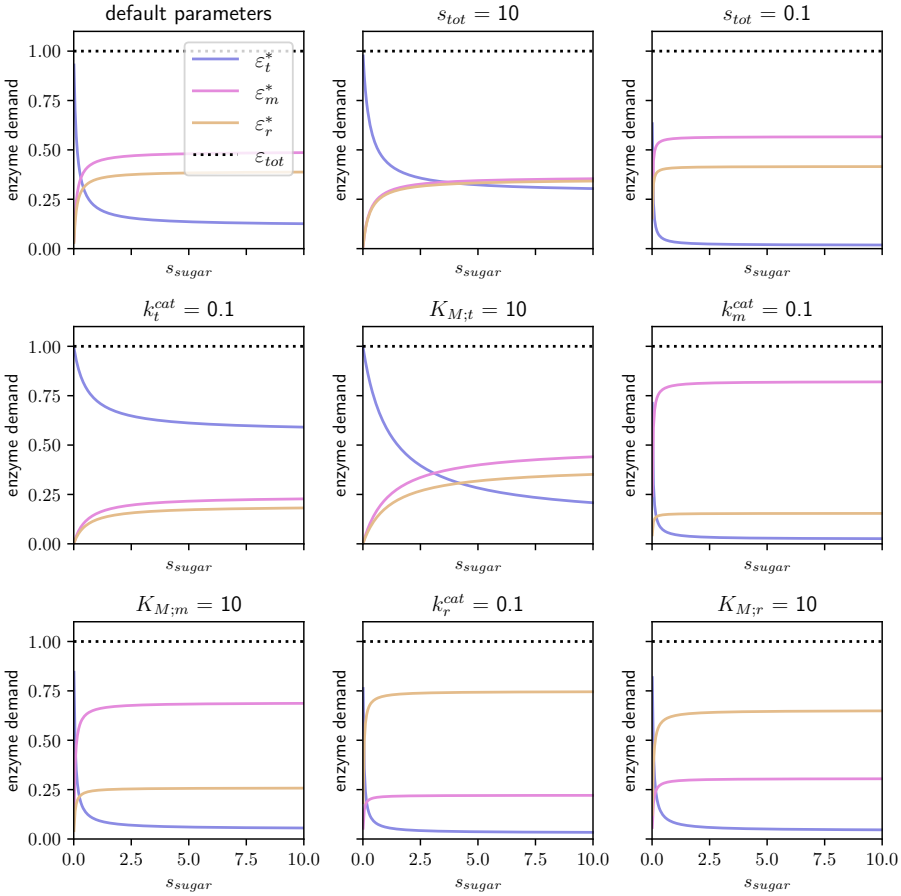


Figure S1: Enzyme demands in a simple cell model, depending on external sugar concentration. The abundances of the three enzyme sectors, when the allocation is optimal, as a function of the external sugar concentration (s_{sugar}). The default parameters are $k_t^{\text{cat}} = k_m^{\text{cat}} = k_r^{\text{cat}} = 1$, $K_{\text{M;t}} = K_{\text{M;r}} = 1$, $K_{\text{M;m}} = 2$, $s_{\text{tot}} = 1$, and $\varepsilon_{\text{tot}} = 1$. In each panel only one of these parameters is changed.

E A cell model with enzyme kinetics and metabolite constraints

The cell model is described by three irreversible reactions:

$$\text{S}_{\text{sugar}} \overset{\varepsilon_t}{\rightleftharpoons} \text{S}_1 \overset{\varepsilon_m}{\rightleftharpoons} \text{S}_2 \overset{\varepsilon_r}{\rightleftharpoons} \text{S}_3 \tag{51}$$

The optimal achievable flux (J^*) for a given total enzyme ε_{tot} is given in Eqs (37-38). Figure S1 shows the individual enzyme demands for the three steps, as a function of the external sugar concentration and J^*:

$$\begin{aligned}
\varepsilon_{\text{t}}^* &= \varepsilon_{\text{tot}} - J^* \cdot \left(1/k_{\text{m}}^{\text{cat}} + 1/k_{\text{r}}^{\text{cat}} + \left(\sqrt{K_{\text{M;m}}/k_{\text{m}}^{\text{cat}}} + \sqrt{K_{\text{M;r}}/k_{\text{r}}^{\text{cat}}} \right)^2 \Big/ s_{\text{tot}} \right) \\
\varepsilon_{\text{m}}^* &= J^* \cdot \left(1/k_{\text{m}}^{\text{cat}} + \sqrt{K_{\text{M;m}}/k_{\text{m}}^{\text{cat}}} \left(\sqrt{K_{\text{M;m}}/k_{\text{m}}^{\text{cat}}} + \sqrt{K_{\text{M;r}}/k_{\text{r}}^{\text{cat}}} \right) \Big/ s_{\text{tot}} \right) \\
\varepsilon_{\text{r}}^* &= J^* \cdot \left(1/k_{\text{r}}^{\text{cat}} + \sqrt{K_{\text{M;r}}/k_{\text{r}}^{\text{cat}}} \left(\sqrt{K_{\text{M;m}}/k_{\text{m}}^{\text{cat}}} + \sqrt{K_{\text{M;r}}/k_{\text{r}}^{\text{cat}}} \right) \Big/ s_{\text{tot}} \right) .
\end{aligned} \tag{52}$$

Note that also ε_{m}^* and ε_{r}^* change with s_{sugar} even though it does not explicitly appear in the equations, because J^* itself is a function of s_{sugar}.

F Mathematical derivations and proofs

This appendix contains many of the detailed mathematical proofs and derivations used in the manuscript for metabolic steady-states, minimum enzyme demand / maximal flux, connectivity theorem, and enzyme control

	$s_i^{\mathrm{st}}(s_0, s_n, \boldsymbol{\varepsilon})$	$J(s_0, s_n, \boldsymbol{\varepsilon})$	$s_i(s_0, J, \boldsymbol{\varepsilon})$	Elasticities	$\mathbf{C}^{\mathrm{s}}$	$\mathbf{C}^J$
Michaelis-Menten	Eq. (55)	$v_1(s_0)$	Eq. (56)	Eq. (54)		Eq. (57)
Thermodynamic (const η^{kin})	Eq. (34)		Eq. (34)		Eq. (94)	
Thermodynamic (substrate sat)						sec F.3.2
Mass-action	Eq. (26)	Eq. (103)		Eq. (111)		sec F.4.2
Haldane	-	-	Eq. (19) in [17]	e.g. [35]	-	-

Table S2: Overview of formulae for steady-state variables and control coefficients. This table summarizes formulae for unbranched metabolic pathways with different rate laws (rows). The columns refer to steady-state concentrations and flux (as a function of enzyme levels and metabolite concentrations); steady-state concentrations as a function of enzyme levels, pathway substrate concentration, and steady-state flux (but not of the product concentration); reaction elasticities; and concentration and flux control coefficients.

rules. The derivations are grouped by the corresponding rate law in the following order: (1) Michaelis-Menten; (2) Thermodynamic; (3) Reversible mass-action; (4) Haldane. An overview of useful equations is given in Table S2.

F.1 Michaelis-Menten kinetics

The Michaelis-Menten rate law is an approximation of the Haldane rate law for irreversible reactions:

$$v_i = \varepsilon_i \; k_i^{\mathrm{cat}} \; \frac{s_{i-1}}{s_{i-1} + K_{\mathrm{M};i}} \tag{53}$$

with reactant elasticities

$$\frac{\partial v_i}{\partial s_j} \quad = \delta_{j,i-1} \; \varepsilon_i \; k_i^{\mathrm{cat}} \; \tfrac{K_{\mathrm{M};i}}{(s_j + K_{\mathrm{M};i})^2} \; . \tag{54}$$

F.1.1 Steady-state concentrations

Computing the steady-state metabolite concentrations in a pathway with Michaelis-Menten kinetics is possible by inverting Eq. (53).

$$\begin{aligned} s_{i-1} &= \frac{v_i}{\varepsilon_i \; k_i^{\mathrm{cat}}} (s_{i-1} + K_{\mathrm{M};i}) \\ \Rightarrow s_{i-1} \left(1 - \frac{v_i}{\varepsilon_i \; k_i^{\mathrm{cat}}}\right) &= \frac{v_i}{\varepsilon_i \; k_i^{\mathrm{cat}}} K_{\mathrm{M};i} \\ \Rightarrow s_{i-1} &= \frac{v_i K_{\mathrm{M};i}}{\varepsilon_i \; k_i^{\mathrm{cat}} (1 - \frac{v_i}{\varepsilon_i \; k_i^{\mathrm{cat}}})} = \frac{K_{\mathrm{M};i}}{\varepsilon_i \; k_i^{\mathrm{cat}} / v_i - 1} \; . \end{aligned} \tag{55}$$

The steady-state concentrations as a function of the pathway flux can be found by replacing v_i with J for all reactions. For any two reactions (i and j), we can relate between their substrate levels by replacing v_i in Eq. (55) with J and using the rate law for $J = v_j$ from Eq. (53)

$$s_{i-1} = \frac{K_{\mathrm{M};i}}{\varepsilon_i \; k_i^{\mathrm{cat}} / J - 1} = \frac{K_{\mathrm{M};i}}{\varepsilon_i \; k_i^{\mathrm{cat}} / \left(\varepsilon_j \; k_j^{\mathrm{cat}} \; \frac{s_{j-1}}{s_{j-1} + K_{\mathrm{M},j}}\right) - 1} = \frac{K_{\mathrm{M};i}}{\frac{\varepsilon_i k_i^{\mathrm{cat}}}{\varepsilon_j k_j^{\mathrm{cat}}} \left(1 + \frac{K_{\mathrm{M},j}}{s_{j-1}}\right) - 1}. \tag{56}$$

In fact, applying this condition to all metabolites is a sufficient condition for having a steady-state. Specifically, if we start with a given concentration s_0 and an enzyme abundance vector $\boldsymbol{\varepsilon}$, there is only one solution for the steady-state flux and for the metabolite levels ($\mathbf{s}$) which support it.

F.1.2 Flux control coefficients

We can compute the flux control coefficients directly by taking derivatives: given that $J = v_1$ and that v_1 depends only on the external substrate, described by a fixed parameter and therefore not dependent on the downstream ε_i, the flux control coefficients read

$$C_l^J = \frac{\partial J}{\partial E_l} \cdot \frac{\varepsilon_l}{J} = \frac{\partial v_1}{\partial \varepsilon_l} \frac{\varepsilon_l}{v_1} = \delta_{l1}. \tag{57}$$

As expected, due to the irreversible rate law, only the first enzyme can have control (except for the pathological case where metabolites concentrations become infinite and flux control is not defined).

F.1.3 Minimum enzyme demand / maximal flux with a metabolite constraint

At steady state (with flux J), Eq. (53) can also be used to express the enzyme demands:

$$\varepsilon_i = \frac{J}{k_i^{\text{cat}}} \cdot \left(1 + \frac{K_{\text{M};i}}{s_{i-1}}\right) = J\left(\alpha_i + \frac{\beta_i}{s_{i-1}}\right) \tag{58}$$

where we defined $\alpha_i \equiv 1/k_i^{\text{cat}}$ and $\beta_i \equiv K_{\text{M};i}/k_i^{\text{cat}}$ as before. Now we can equate the sum of all enzyme demands to ε_{tot}:

$$\varepsilon_{\text{tot}} = J\left(\sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} \frac{\beta_i}{s_{i-1}}\right) = J\left(||\boldsymbol{\alpha}||_1 + \sum_{i=1}^{n} \frac{\beta_i}{s_{i-1}}\right). \tag{59}$$

This inverse relationship between ε_{tot} and the metabolite concentrations means that in order to minimize ε_{tot} for a given flux, all s_i should be as large as possible. Obviously, infinite concentrations are not feasible. To avoid non-physical solutions, we can add a constraint on the total concentration of metabolites $\sum_{i=0}^{n-1} s_i \leq s_{\text{tot}}$. Now, using Lagrange's method, we define $\mathcal{L} = J\left(||\boldsymbol{\alpha}||_1 + \sum_i \beta_i/s_{i-1}\right) + \lambda\left(\sum_i s_i - s_{\text{tot}}\right)$ and take the derivatives:

$$\begin{aligned} 0 = \frac{\partial \mathcal{L}}{\partial s_{i-1}} &= -\frac{\beta_i J^*}{{s_{i-1}^*}^2} + \lambda \\ s_{i-1}^* &= \sqrt{\frac{J^*}{\lambda}}\sqrt{\beta_i}. \end{aligned} \tag{60}$$

To find λ, we can now use the constraint on the sum of all metabolites (assuming the upper bound is hit):

$$s_{\text{tot}} = \sum_{i=0}^{n-1} s_i^* = \sqrt{\frac{J^*}{\lambda}}\left(\sum_{i=0}^{n-1}\sqrt{\beta_i}\right) = \sqrt{\frac{J^*}{\lambda}}\sqrt{||\boldsymbol{\beta}||_{\frac{1}{2}}} \tag{61}$$

where $\boldsymbol{\beta}$ is the vector of all β_i, and $||\cdot||_{\frac{1}{2}}$ is the $\text{l}_{1/2}$ norm. Therefore, we can replace $\sqrt{J^*/\lambda}$ with $s_{\text{tot}}/\sqrt{||\boldsymbol{\beta}||_{\frac{1}{2}}}$ in Eq. (60) and find an explicit expression for s_{i-1}^*:

$$s_{i-1}^* = \frac{s_{\text{tot}}}{\sqrt{||\boldsymbol{\beta}||_{\frac{1}{2}} \cdot \beta_i}}. \tag{62}$$

By combining the results from Eqs (59) and (62), we can see that:

$$J^* = \frac{\varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_1 + \sum_i \frac{\beta_i}{s_{i-1}^*}} = \frac{\varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_1 + \sum_i \frac{\sqrt{\beta_i \cdot ||\boldsymbol{\beta}||_{\frac{1}{2}}}}{s_{\text{tot}}}} = \frac{\varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_1 + \frac{||\boldsymbol{\beta}||_{\frac{1}{2}}}{s_{\text{tot}}}} \tag{63}$$

and similarly we can also find the optimal enzyme levels:

$${\varepsilon_i}^* = J^*\left(\alpha_i + \frac{\beta_i}{s_{i-1}^*}\right) = \varepsilon_{\text{tot}} \frac{\alpha_i + \sqrt{\beta_i\ ||\boldsymbol{\beta}||_{\frac{1}{2}}}/s_{\text{tot}}}{||\boldsymbol{\alpha}||_1 + ||\boldsymbol{\beta}||_{\frac{1}{2}}/s_{\text{tot}}}. \tag{64}$$

F.1.4 Case where the first substrate has fixed concentration

If our pathway model is used to represent a cell growing on an external substrate (S_0), then it would be unrealistic to assume that the concentration s_0 is subject to optimization. Instead, we can assume it is fixed (e.g. based on environmental conditions) and that the cell optimizes only the concentrations of enzymes and internal metabolites.

From Eq. (63) we can see that:

$$J = \frac{\varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_1 + \sum_i \beta_i/s_{i-1}} = \frac{\varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_1 + \beta_1/s_0 + \sum_{i=2}^{n} \beta_i/s_{i-1}}. \tag{65}$$

Since s_0 is fixed, we consider the term β_1/s_0 to be part of the constant part of the denominator (usually just $\sum_i 1/k_i^{\text{cat}}$). Furthermore, we can redefine $||\boldsymbol{\beta}||_{\frac{1}{2}}$ to exclude the value for the first enzyme (β_1) and impose the upper bound s_{tot} only on the sum of internal metabolites. In this case, the solution we got in Eqs (60)-(61) will still be applicable, and therefore we can write:

$$J^* = \frac{\varepsilon_{\text{tot}}}{\beta_1/s_0 + ||\boldsymbol{\alpha}||_1 + ||\boldsymbol{\beta}||_{\frac{1}{2}}/s_{\text{tot}}} = \mu^{\max} \cdot \frac{s_0}{s_0 + K_{\text{Monod}}} \tag{66}$$

where we defined

$$\begin{aligned} \mu^{\max} &\equiv \frac{\varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_1 + ||\boldsymbol{\beta}||_{\frac{1}{2}}/s_{\text{tot}}} \\ K_{\text{Monod}} &\equiv \frac{\beta_1}{||\boldsymbol{\alpha}||_1 + ||\boldsymbol{\beta}||_{\frac{1}{2}}/s_{\text{tot}}} \end{aligned} . \tag{67}$$

As we can see, the solution provides a prediction for the Monod constant – i.e., the concentration of substrate (s_0) where the growth rate reaches half of its maximum. If the transporter turnover constant (k_1^{cat}) is indeed much slower than in all other enzymes,its control coefficient is 1 and we will get $K_{\text{Monod}} = K_{\text{M},1}$, which is often naïvely assumed. However, if the control is distributed along the pathway, the Monod constant can become significantly smaller than $K_{\text{M},1}$.

We also obtain expressions for the individual enzyme levels, using Eq. (64). Because s_0 is fixed, the demand for the first enzyme is can be simply expressed as:

$${\varepsilon_i}^* = J^* \left(\alpha_i + \frac{\beta_i}{s_{i-1}^*} \right). \tag{68}$$

For the first enzyme, the substrate (s_0) is fixed so the expression yields simply ${\varepsilon_1}^* = J^* (\alpha_1 + \beta_1/s_0)$, while for all other enzymes ($\forall i > 0$) we can use the solution for s_{i-1}^* from Eq. (62), i.e. ${\varepsilon_i}^* = J^*(\alpha_i + \sqrt{\beta_i \, ||\boldsymbol{\beta}||_{\frac{1}{2}}}/s_{\text{tot}})$.

F.2 Thermodynamic rate law with fixed kinetic efficiency

The thermodynamic rate law with fixed kinetic efficiency reads:

$$v_i = \varepsilon_i \; k_i^{\text{cat}} \left(1 - e^{-\theta_i}\right) = \varepsilon_i \; k_i^{\text{cat}} \left(1 - \frac{s_i}{s_{i-1} \; K_i^{\text{eq}}}\right) \tag{69}$$

where $\theta_i \equiv \ln\left(\frac{s_{i-1} K_i^{\text{eq}}}{s_i}\right)$.

F.2.1 Steady-state concentrations

Assuming a steady state (i.e. all fluxes are equal to J), we can find a simple recursion formula for the concentrations s_i:

$$s_i = s_{i-1} \; K_i^{\text{eq}} \; (1 - J \cdot (\varepsilon_i k_i^{\text{cat}})^{-1}) \,. \tag{70}$$

By solving the recursion for s_n we get:

$$s_n = s_{n-1} \; K_n^{\text{eq}} \; (1 - J \cdot (\varepsilon_n \; k_n^{\text{cat}})^{-1}) = \ldots = s_0 \prod_{i=1}^{n} K_i^{\text{eq}} (1 - J \cdot (\varepsilon_i \; k_i^{\text{cat}})^{-1}) = s_0 \; K_{\text{tot}}^{\text{eq}} \prod_{i=1}^{n} (1 - J \cdot (\varepsilon_i \; k_i^{\text{cat}})^{-1}) \,, \tag{71}$$

where we used the fact that $K_{\text{tot}}^{\text{eq}} = \prod_{i=1}^{n} K_i^{\text{eq}}$.

The driving force of the entire pathway, which we denote by θ_{tot}, is the sum of all the individual driving forces:

$$\theta_{\text{tot}} = \sum_{i=1}^{n} \theta_i = \sum_{i=1}^{n} \ln\left(\frac{s_{i-1} \; K_i^{\text{eq}}}{s_i}\right) = \ln\left(\prod_{i=1}^{n} s_{i-1}/s_i\right) + \ln\left(\prod_{i=1}^{n} K_i^{\text{eq}}\right) = \ln\left(s_0/s_n\right) + \ln\left(K_{\text{tot}}^{\text{eq}}\right). \tag{72}$$

Since θ_{tot} depends only on $K_{\text{tot}}^{\text{eq}}$, s_0, and s_n – all of which are constant – it is a constant as well. Therefore, we can use this result to rewrite Equation (71) as:

$$\theta_{\text{tot}} = -\sum_{i=1}^{n} \ln\left(1 - J \cdot (\varepsilon_i \; k_i^{\text{cat}})^{-1}\right). \tag{73}$$

F.2.2 Minimum enzyme demand / maximal flux

Directly solving the flux maximization problem using Equation (73) would be difficult, because it cannot be solved analytically to find an expression for J as a function of enzyme levels. However, it is rather simple to do the opposite and express the enzyme levels as functions of J and the other parameters, sum them up and compare to the total (ε_{tot}):

$$\begin{aligned} \varepsilon_i &= \frac{J}{k_i^{\text{cat}}\left(1-e^{-\theta_i}\right)} = J\,\frac{\alpha_i}{1-e^{-\theta_i}} \\ \frac{\varepsilon_{\text{tot}}}{J} &\geq \frac{1}{J}\sum_i \varepsilon_i = \sum_i \alpha_i \left(1-e^{-\theta_i}\right)^{-1} \end{aligned} \tag{74}$$

where we define $\alpha_i \equiv 1/k_i^{\text{cat}}$ for simplicity.

Minimizing $\varepsilon_{\text{tot}}/J$ is an optimization problem that can be solved directly using the s_i as variables, but it can be easier to solve if we consider the θ_i to be the independent variables instead. This variable switch is justified because there is a linear homomorphism between the two sets $\{\ln s_i\}_{i=1}^{n-1}$ and $\{\theta_i\}_{i=1}^{n}$, given fixed s_0 and s_n and under the constraint that $\sum \theta_i = \theta$.

Now, we want to find an optimal set of values for the driving forces, denoted ${\theta_i}^*$, that sum up to θ and maximize J (similarly, J^* and ${\varepsilon_i}^*$ would be the corresponding values of J and ε_i at the optimum).

Lemma F.1. *The values of x_i that minimize the function $\sum_i \alpha_i \left(1-e^{-\theta_i}\right)^{-1}$ under the constraint $\sum_i x_i = x_{\text{tot}}$, satisfy*

$$x_i = 2\,\sinh^{-1}\left(\sqrt{\Psi\alpha_i}\right) \tag{75}$$

for some $\Psi \in \mathbb{R}$.

Proof. Using Lagrange's method, we first define $\mathcal{L} = \sum_i \frac{\alpha_i}{1-e^{-x_i}} + \lambda\left(\sum_i x_i - x_{\text{tot}}\right)$

$$\begin{aligned} 0 &= \frac{\partial \mathcal{L}}{\partial x_i} = \frac{\alpha_i\, e^{-x_i}}{(1-e^{-x_i})^2} + \lambda = \alpha_i\left(e^{x_i/2} - e^{-x_i/2}\right)^{-2} + \lambda \\ \sqrt{-\frac{\alpha_i}{\lambda}} &= e^{x_i/2} - e^{-x_i/2} = 2\,\sinh(x_i/2) \\ x_i &= 2\,\sinh^{-1}\left(\sqrt{\frac{\alpha_i}{-4\lambda}}\right) \end{aligned} \tag{76}$$

and if we define $\Psi = -\frac{1}{4\lambda}$, we can see that it proves the lemma. □

Using Lemma (F.1), one can see that the optimal distribution of driving forces satisfies the following relationship, for some value of Ψ

$${\theta_i}^* = 2\,\sinh^{-1}\left(\sqrt{\Psi\alpha_i}\right) \tag{77}$$

where the exact value of Ψ can be determined by applying the constraint on the sum of θ_i from Eq. (72):

$$\ln\left(\frac{s_0}{s_n}K_{\text{tot}}^{\text{eq}}\right) = \theta_{\text{tot}} = \sum_i {\theta_i}^* = 2\sum_i \sinh^{-1}\left(\sqrt{\Psi\alpha_i}\right) = 2\sum_i \ln\left(\sqrt{\Psi\alpha_i} + \sqrt{1+\Psi\alpha_i}\right). \tag{78}$$

Unfortunately, there is no analytical solution to this equation. Nevertheless, the function on the right-hand side is strictly monotonically increasing with Ψ (in the entire range $\Psi \in \mathbb{R}$), so solving it numerically should be straightforward. Moreover, we can still express $\left(1-e^{-\theta_i}\right)^{-1}$ as a function of Ψ (see Appendix F.2.2):

$$\left(1-e^{-\theta_i}\right)^{-1} = \left(1-e^{-2\cdot\sinh^{-1}(\sqrt{\Psi\alpha_i})}\right)^{-1} = \frac{1}{2} + \frac{1}{2}\sqrt{1+(\Psi\alpha_i)^{-1}}, \tag{79}$$

where we use the fact that $1-e^{-2\cdot\sinh^{-1}(x)} = 2/(1+\sqrt{1+x^{-2}})$. Therefore, the solutions for the maximal flux and individual enzyme allocations would be:

$$\begin{aligned} J^* &= \frac{\varepsilon_{\text{tot}}}{\sum_i \alpha_i\left(1-e^{-\theta_i}\right)^{-1}} = \frac{2\,\varepsilon_{\text{tot}}}{\sum_i \alpha_i\left(1+\sqrt{1+(\Psi\alpha_i)^{-1}}\right)} \\ {\varepsilon_i}^* &= J^*\alpha_i\,\frac{1}{2}\left(1+\sqrt{1+(\Psi\alpha_i)^{-1}}\right). \end{aligned} \tag{80}$$

F.2.3 Limit of small driving forces

If we further assume that the total driving force θ_{tot} is small (and therefore also each one of the reaction driving forces θ_i is even smaller), then we can write:

$$J \leq \varepsilon_{\text{tot}} \cdot \left(\sum_i \frac{1}{k_i^{\text{cat}}} \cdot \frac{1}{1 - e^{-\theta_i}} \right)^{-1} \approx \varepsilon_{\text{tot}} \cdot \left(\sum_i \frac{1}{k_i^{\text{cat}}\ \theta_i} \right)^{-1} . \tag{81}$$

As we did before, the solution for J^* can be found by minimizing $\varepsilon_{\text{tot}}/J$ under the constraint $\sum_i \theta_i = \theta_{\text{tot}}$, using the Lagrange method: $\mathcal{L} = \sum_i \frac{1}{k_i^{\text{cat}}\ \theta_i} + \lambda(\sum_i \theta_i - \theta_{\text{tot}})$.

$$0 = \frac{\partial \mathcal{L}}{\partial \theta_i} = -\frac{1}{k_i^{\text{cat}} \cdot (\theta_i{}^*)^2} + \lambda \tag{82}$$

$$\theta_i{}^* = \frac{1}{\sqrt{\lambda\ k_i^{\text{cat}}}} . \tag{83}$$

By applying the constraint $\theta_{\text{tot}} = \sum_i \theta_i{}^* = \lambda^{-1/2} \sum_i \sqrt{1/k_i^{\text{cat}}} \Rightarrow \sqrt{\lambda} = \theta_{\text{tot}}^{-1} \cdot (\sum_i \sqrt{1/k_i^{\text{cat}}})$, we can write:

$$\begin{aligned} \theta_i{}^* &= \theta_{\text{tot}} \frac{\sqrt{1/k_i^{\text{cat}}}}{\sum_i \sqrt{1/k_i^{\text{cat}}}} \\ J^* &= \varepsilon_{\text{tot}} \cdot \left(\sum_i \frac{1}{k_i^{\text{cat}}\ \theta_i{}^*} \right)^{-1} = \frac{\theta_{\text{tot}}\ \varepsilon_{\text{tot}}}{\left(\sum_i \sqrt{1/k_i^{\text{cat}}} \right)^2} = \frac{\theta_{\text{tot}}\ \varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_{\frac{1}{2}}} \end{aligned} \tag{84}$$

and furthermore, we can see that:

$$\varepsilon_i{}^* = \frac{J^*}{k_i^{\text{cat}}\ \theta_i{}^*} \propto \frac{1}{k_i^{\text{cat}}\ \theta_i{}^*} \propto \frac{1}{k_i^{\text{cat}}\ \sqrt{1/k_i^{\text{cat}}}} = \sqrt{1/k_i^{\text{cat}}} = \sqrt{\alpha_i} . \tag{85}$$

F.3 Approximate analytical formula

Based on the two limits for values of the total driving force ($\theta_{\text{tot}} \to 0$ and $\theta_{\text{tot}} \to \infty$), we can construct an approximation that converges with these limits for extreme values, following the template $a \cdot (1 - e^{-b\ \theta_{\text{tot}}})$.

The solutions for the two limits are:

$$\begin{aligned} \lim_{\theta_{\text{tot}} \to 0^+} J^* &= \frac{\theta_{\text{tot}}\ \varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_{\frac{1}{2}}} \\ \lim_{\theta_{\text{tot}} \to \infty} J^* &= \frac{\varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_1} . \end{aligned}$$

One can see that the approximate formula that has the same limits will be:

$$\begin{aligned} J^* &\approx \frac{\varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_1} \left(1 - \exp\left(-\frac{||\boldsymbol{\alpha}||_1}{||\boldsymbol{\alpha}||_{\frac{1}{2}}} \theta_{\text{tot}} \right) \right) \\ &= \frac{\varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_1} \left(1 - \exp\left(-\frac{||\boldsymbol{\alpha}||_1}{||\boldsymbol{\alpha}||_{\frac{1}{2}}} \ln\left(\frac{s_n}{s_0 K_{\text{tot}}^{\text{eq}}} \right) \right) \right) \\ &= \frac{\varepsilon_{\text{tot}}}{||\boldsymbol{\alpha}||_1} \left(1 - \left(\frac{s_0 K_{\text{tot}}^{\text{eq}}}{s_n} \right)^{\frac{||\boldsymbol{\alpha}||_1}{||\boldsymbol{\alpha}||_{\frac{1}{2}}}} \right) \end{aligned} \tag{86}$$

Figure S2 illustrates how this approximation holds up against the precise solution (which is based on the closed-form formula, but requires inverting Eq. (78) numerically to find Ψ). One can appreciate that the approximation is a very good one across all values of θ_{tot} and for all choices of k^{cat} values tested here. In Figure S3, we compare the approximation to solutions found numerically using convex optimization, i.e. directly optimizing J over the space of metabolite concentrations using a convex optimization solver (CVXPY [36, 37]).

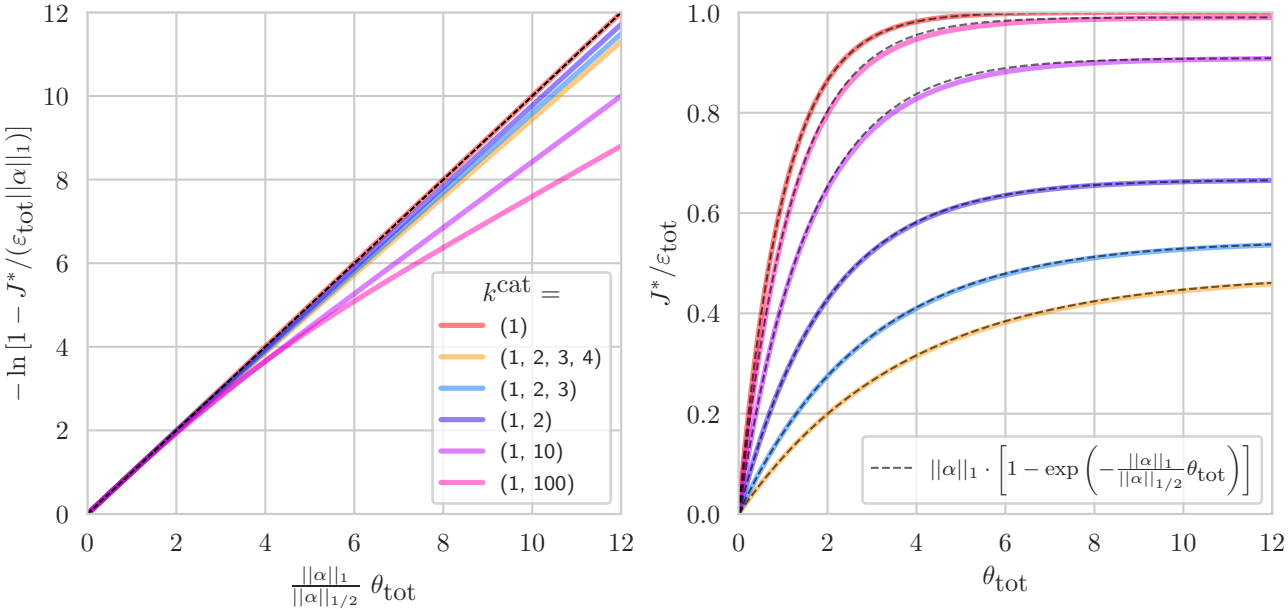


Figure S2: Approximate formula for the thermodynamic rate law case. Left panel: by re-scaling the x-axis by a factor of $||\alpha||_1/||\alpha||_{1/2}$ and transforming the y-axis using $y = -\ln(1 - J^*/\varepsilon_{\text{tot}} \cdot ||\boldsymbol{\alpha}||_1)$, the precise solutions based on Eq. (80) appear almost linear. Each curve represent a different set of k^{cat} values (and a varying number of enzymes in the pathway, between 1-4). The approximation based on Eq. (86) is depicted by the dashed black line, which goes exactly along the diagonal for this set of axes. In the right panel, we plot the same precise solutions and approximations as in the left panel, but without transforming the axes.

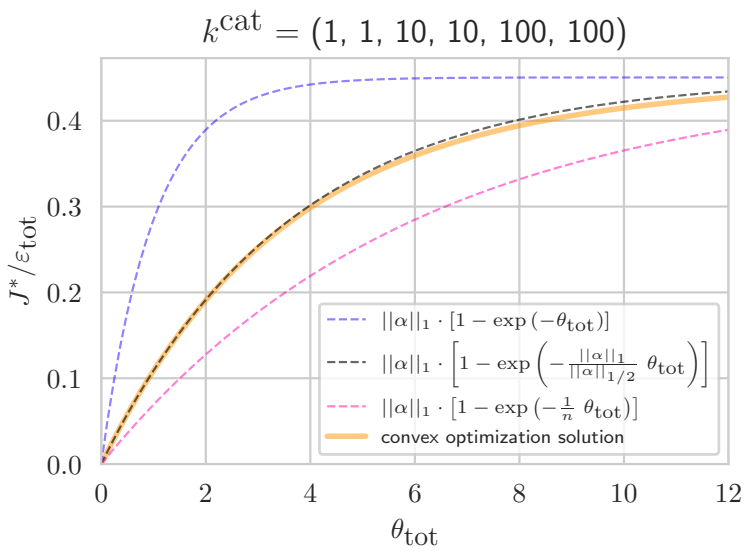


Figure S3: Comparing the approximate solution for the thermodynamic rate law case to the numerical solution based on convex optimization. To arrive at the convex optimization solution (orange curve) we used CVXPY [36, 37]. The precise solution (based on the analytical formula from Eq. (80)) overlaps completely with the one using convex optimization and therefore is not shown. The dashed black line represent the approximate solution. In addition, we plot two other "naïve" approximations: a lower bound (pink dashed line) which represent the rate assuming that the driving force θ_{tot} is uniformly distributed along the pathway – i.e. for every reaction $\theta_i = \theta_{\text{tot}}/n$ – and an upper bound (violet dashed line) which represents a case where all the driving force is concentrated in one step, while all other steps have negligible cost.

F.3.1 Connectivity theorem

The unscaled metabolite elasticities for the thermodynamic rate law read

$$E^{v_l}_{s_i} \equiv \frac{\partial v_l}{\partial s_i} = \frac{\partial}{\partial s_i} \varepsilon_l k_l^{\text{cat}} \left(1 - e^{-\theta_l(\mathbf{s})}\right) = \varepsilon_l \; k_l^{\text{cat}} \; e^{-\theta_l} \; \frac{\partial \theta_l}{\partial s_i} = \left(\varepsilon_l \; k_l^{\text{cat}} - J\right) \frac{\partial \theta_l}{\partial s_i}. \tag{87}$$

In the last step we used the fact that $\varepsilon_l \; k_l^{\text{cat}} \; e^{-\theta_l} = \varepsilon_l \; k_l^{\text{cat}} - \varepsilon_l \; k_l^{\text{cat}}(1 - e^{-\theta_l}) = \varepsilon_l \; k_l^{\text{cat}} - J$. With flux control coefficients proportional to $(\varepsilon_l \; k_l^{\text{cat}} - J)^{-1}$, we can now derive the connectivity theorem for each metabolite s_i,

$$\sum_l C_l^J \; E^{J_l}_{s_i} \propto \sum_l (\varepsilon_l \; k_l^{\text{cat}} - J)^{-1} (\varepsilon_l \; k_l^{\text{cat}} - J) \; \frac{\partial \theta_l}{\partial s_i} = \sum_l \frac{\partial \theta_l}{\partial s_i} = \sum_l \frac{n_{il}}{s_i} = 0. \tag{88}$$

Here we used $\theta_l = \theta_l^\circ + \sum_i n_{il} \ln s_i$ with derivative $\frac{\partial \theta_l}{\partial s_i} = \frac{n_{il}}{s_i}$ (in the second but last step), and the fact that our reactions are uni-molecular, with stoichiometric coefficients of ± 1 for substrate and products (in the last step).

F.3.2 Flux control coefficients

We start by turning Eq. (71) into an equality constraint:

$$\mathcal{Y} \equiv \prod_{i=1}^{n} \left(1 - J(\varepsilon_i \; k_i^{\text{cat}})^{-1}\right) - \frac{s_n}{s_0} \frac{1}{K_{\text{tot}}^{\text{eq}}} = 0. \tag{89}$$

This equation implicitly determines the flux J (given the enzyme levels ε_i and the external concentrations s_0 and s_n). Since $J \leq \varepsilon_i \; k_i^{\text{cat}}$, all product terms are positive and decreasing in J, and so the entire product is a decreasing function of J. Therefore, there can be only one solution with a positive flux. While Eq. (89) cannot be solved for J directly, it determines J implicitly and is sufficient for computing the flux response coefficients. For doing this we use the following lemma:

Lemma F.2. *We consider two variables x and y, constrained by an equality $f(x, y) = 0$. We assume that for each given value of x, the constraint is satisfied by a single value of y, which we call $y = g(x)$. To compute the derivative $\partial g(x)/\partial x$ directly from $f(x, y)$ without an explicit expression for $g(x)$, we insert $g(x)$ and write the constraint as $h(x) = f(x, g(x)) = 0$. With the chain rule, we can write this as $0 = \frac{\partial h(x)}{\partial x} = \frac{\partial f(x,y)}{\partial x}|_{y=g(x)} + \frac{\partial f(x,y)}{\partial y}|_{y=g(x)} \; \frac{\partial g(x)}{\partial x}$, and obtain*

$$\frac{\partial g(x)}{\partial x} = -\frac{\partial f(x, y)/\partial x}{\partial f(x, y)/\partial y}|_{y=g(x)} \tag{90}$$

To compute the flux control coefficients, we now see Eq. (89) as a constraint between an enzyme level ε_i and th eflux J, assuming that all other enzyme levels are given. This constraint implicitly defines a function $J(\varepsilon_i)$. To compute the derivative $\partial J/\partial \varepsilon_i$ (assuming all other enzyme levels are constant), we set $x \to \varepsilon_i$; $y \to J$; $g(x) \to J(\varepsilon_i)$; and $f(x, y) \to \mathcal{Y}(\varepsilon_i, J)$, and obtain:

$$\begin{aligned}
R^J_{\varepsilon_l} &= \frac{\partial J}{\partial \varepsilon_l} = -\frac{\partial \mathcal{Y}}{\partial \varepsilon_l} \left(\frac{\partial \mathcal{Y}}{\partial J}\right)^{-1} \\
&= -\frac{\prod_{i \neq l}(1 - J \cdot (\varepsilon_i \; k_i^{\text{cat}})^{-1})}{\prod_{i=1}^{n}(1 - J \cdot (\varepsilon_i \; k_i^{\text{cat}})^{-1})} \cdot \frac{-J \cdot (k_l^{\text{cat}})^{-1} \cdot \varepsilon_l^{-2}}{\sum_{i=1}^{n} \frac{-(\varepsilon_i \; k_i^{\text{cat}})^{-1}}{1 - J \cdot (\varepsilon_i \; k_i^{\text{cat}})^{-1}}} \\
&= -\frac{1}{1 - J \cdot (\varepsilon_l \; k_l^{\text{cat}})^{-1}} \cdot \frac{J \cdot (k_l^{\text{cat}})^{-1} \cdot \varepsilon_l^{-2}}{\sum_{i=1}^{n} (\varepsilon_i \; k_i^{\text{cat}} - J)^{-1}} \\
&= -\frac{J}{\varepsilon_l} \cdot \underbrace{\frac{(\varepsilon_l \; k_l^{\text{cat}} - J)^{-1}}{\sum_{i=1}^{n} (\varepsilon_i \; k_i^{\text{cat}} - J)^{-1}}}_{C_l^J}.
\end{aligned} \tag{91}$$

The result is simple: the control coefficients (C_l^J) are proportional to $(\varepsilon_l \; k_l^{\text{cat}} - J)^{-1}$ and sum to 1 as required by the summation theorem.

F.3.3 Concentration control coefficients

Using Eq. (71), which we can solve for any s_j in the same way as for s_n, we obtain $s_j = s_0 \prod_{i=1}^{j} K_i^{\text{eq}}(1-J\cdot(\varepsilon_i k_i^{\text{cat}})^{-1})$ and therefore

$$R_{\varepsilon_l}^{s_j} = \frac{\partial s_j}{\partial \varepsilon_l} = \underbrace{s_0 \prod_{i=1}^{j} K_i^{\text{eq}}(1 - J\cdot(\varepsilon_i k_i^{\text{cat}})^{-1})}_{s_j} \cdot \underbrace{\frac{\frac{\partial}{\partial \varepsilon_l}(1 - J\cdot(\varepsilon_l k_l^{\text{cat}})^{-1})}{1 - J\cdot(\varepsilon_l k_l^{\text{cat}})^{-1}}}_{\mathcal{Z}} \tag{92}$$

for all enzyme indices $l \leq j$ (otherwise 0). The first product term is simply s_j. The second product term can be rewritten as:

$$\begin{aligned}\mathcal{Z} &= -\frac{\frac{\partial J}{\partial \varepsilon_l}\cdot(\varepsilon_l\ k_l^{\text{cat}})^{-1} + J\cdot(k_l^{\text{cat}})^{-1}\varepsilon_l^{-2}}{1 - J\cdot(\varepsilon_l k_l^{\text{cat}})^{-1}} = -\frac{R_{\varepsilon_l}^{s_j} + J\varepsilon_l^{-1}}{\varepsilon_l k_l^{\text{cat}} - J} \\ &= \frac{\frac{J}{\varepsilon_l}\frac{(\varepsilon_l\ k_l^{\text{cat}}-J)^{-1}}{\sum_{i=1}^{n}(\varepsilon_i\ k_i^{\text{cat}}-J)^{-1}} - \frac{J}{\varepsilon_l}}{\varepsilon_l k_l^{\text{cat}} - J} = \frac{J}{\varepsilon_l}\cdot\frac{\frac{(\varepsilon_l\ k_l^{\text{cat}}-J)^{-1}}{\sum_{i=1}^{n}(\varepsilon_i\ k_i^{\text{cat}}-J)^{-1}} - 1}{\varepsilon_l k_l^{\text{cat}} - J} = \frac{J}{\varepsilon_l}\cdot\frac{\frac{1}{\sum_{i=1}^{n}(\varepsilon_i\ k_i^{\text{cat}}-J)^{-1}} - \varepsilon_l\ k_l^{\text{cat}} + J}{(\varepsilon_l\ k_l^{\text{cat}} - J)^{-2}}\end{aligned} \tag{93}$$

Putting all this together, we obtain the concentration control coefficients

$$C_l^{s_j} = \frac{R_{\varepsilon_l}^{s_j}}{J/\varepsilon_l} = s_j \left[\frac{\frac{1}{\sum_{i=1}^{n}(\varepsilon_i\ k_i^{\text{cat}}-J)^{-1}} - \varepsilon_l\ k_l^{\text{cat}} + J}{(\varepsilon_l\ k_l^{\text{cat}} - J)^{-2}}\right]. \tag{94}$$

Again, this formula holds only for $l \leq j$ (otherwise $C_l^{s_j} = 0$).

F.3.4 Enzyme-control rule

To show that the control coefficients Eq. (35) satisfy the enzyme-control rule, we assume that the rule is correct, insert our formula for control coefficients, derive from it an expression for the enzyme levels, and compare the result to our formula Eq. (80). We do this now step by step. Starting from the proportionality

$$C_l^J \propto \frac{1}{\varepsilon_l\ k_l^{\text{cat}} - J} \propto \varepsilon_l \tag{95}$$

we obtain

$$\begin{aligned}&\frac{1}{\varepsilon_l^2\ k_l^{\text{cat}} - J\ \varepsilon_l} = \text{const} = 1/a \\ &\Rightarrow \varepsilon_l^2\ k_l^{\text{cat}} - J\ \varepsilon_l - a = 0 \\ &\Rightarrow \varepsilon_l^2 - \frac{J}{k_l^{\text{cat}}}\ \varepsilon_l - \frac{a}{k_l^{\text{cat}}} = 0 \\ &\Rightarrow \varepsilon_l = \frac{J}{2\ k_l^{\text{cat}}} \pm \sqrt{\left(\frac{J}{2\ k_l^{\text{cat}}}\right)^2 + \frac{a}{k_l^{\text{cat}}}} = \frac{J}{k_l^{\text{cat}}}\left(\frac{1}{2} \pm \frac{1}{2}\sqrt{1 + k_l^{\text{cat}}\ \frac{4\ a}{J^2}}\right)\end{aligned} \tag{96}$$

The solution with a minus sign is physically meaningless and can be discarded. Remember that we defined $\alpha_i \equiv 1/k_i^{\text{cat}}$, so we can rewrite Eq. (96) as

$$\varepsilon_i = J\alpha_i\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + (\Psi\alpha_i)^{-1}}\right). \tag{97}$$

Here we also replaced $J^2/(4a)$ with Ψ by comparing this solution to Eq. (80) (and using the fact that the value of this constant is uniquely determined by the constraint $\sum_i \varepsilon_i = \varepsilon_{\text{tot}}$).

F.3.5 A note on Max-min Driving Force

The Max-Min Driving Force method (MDF) is a heuristics for finding realistic metabolite concentrations in metabolic models with known flux directions. It relies on the fact that very small driving forces lead to large enzyme

demands and should be avoided, while very large driving forces do not further improve the enzyme demand. The MDF heuristics can be seen as a way to reduce enzyme cost. To see this, we start from the thermodynamic rate law and consider a lower bound on the enzyme demand:

$$\varepsilon_l = \frac{v_l}{k_l^{\text{cat}}} \frac{1}{1 - \mathrm{e}^{-\theta_l}} \geq \frac{v_l}{k_l^{\text{cat}}} \max(1, \frac{1}{\theta_l}). \tag{98}$$

where we use the fact that $e^{-x} \geq \max(0, 1 - x)$. We now consider again an unbranched pathway with a steady forward flux and a given overall driving force θ. In this case, we obtain the constraints $\theta_l \geq 0$ and $\sum_l \theta_l = \theta$. The forces θ_l are not free variables, but a function $\theta_l = \ln K_l^{\text{eq}} - (\ln c_l - \ln c_{l-1})$ of the metabolite concentrations, which in turn are constrained by physiological ranges. Let us see how the MDF principle can be derived from our present optimality problem. Since pathway flux and total enzyme demand scale proportionally at given metabolite concentrations, a flux maximization at a fixed enzyme budget is equivalent to minimizing the total enzyme demand $\sum_l \varepsilon_l$ at a given flux. With Eq. (98), we can approximate this demand as

$$\sum_l \varepsilon_l \geq \sum_{l=1}^{n} \frac{v_l}{k_l^{\text{cat}}} + \sum_{l:\theta_l<1} \frac{v_l}{k_l^{\text{cat}}} \left(\frac{1}{\theta_l} - 1\right) \geq \sum_{l=1}^{n} \frac{v_l}{k_l^{\text{cat}}} + \max_{l:\theta_l<1} \frac{v_l}{k_l^{\text{cat}}} \left(\frac{1}{\theta_l} - 1\right) \tag{99}$$

where $l : \theta_l < 1$ denotes the indices of reactions with driving forces smaller than 1. The first term describes the constant enzyme demand of a hypothetical model in which all enzymes work at their maximal speed. The second term concerns reactions with driving forces smaller than 1 (if such reactions exist) and denotes the maximal value of $\frac{v_l}{k_l^{\text{cat}}} \left(\frac{1}{\theta_l} - 1\right)$ among these reactions. To minimize the overall enzyme demand, we choose the heuristics of minimizing this term. This is equivalent to maximizing

$$\min_{l:\theta_l<1} \frac{k_l^{\text{cat}}}{v_l} \theta_l. \tag{100}$$

where l runs again over all reactions with a driving force smaller than 1. Thus, we postulate that the smallest driving force in the pathway (weighted by $\frac{k_l^{\text{cat}}}{v_l}$) should be as large as possible. The original MDF driving force method employs some additional simplifications. First, we neglect the prefactor $\frac{k_l^{\text{cat}}}{v_l}$ (assuming that nothing is known about enzyme kinetics). Second, if all the reactions have driving forces larger than 1, we let the index l run over all reactions, assuming that it will still be the smallest driving force that causes the largest enzyme cost.

F.4 Reversible mass-action kinetics

F.4.1 Steady-state flux

Here we will show the full derivation of the optimal solution in the case of reversible mass-action rate laws. We start by reminding ourselves that the rate law for each reaction is given by:

$$v_i = \varepsilon_i \ \beta_i^{-1} \ (s_{i-1} - s_i/K_i^{\text{eq}}), \tag{101}$$

and the pathway flux is thus (as in Eq. (26)):

$$J = \varepsilon_{\text{tot}} \left(\sum_i \frac{\beta_i}{s_{i-1} - s_i/K_i^{\text{eq}}} \right)^{-1}, \tag{102}$$

where we replaced the inequality by an equality since we are looking for the maximal flux.

While J can in theory be maximized by equating the gradient with respect to all metabolite levels to 0, the resulting system of equations would be difficult to solve. Instead, we will use a different approach that involves finding an expression for J as a function of the different enzyme levels (rather than metabolite levels).

First, we use Eq. (101) to obtain a formula for s_i. By equating $v_i = J$ and applying it recursively we get:

$$\begin{aligned}
s_1 &= K_1^{\text{eq}} \left(s_0 - \frac{J\beta_1}{\varepsilon_1} \right) \\
s_2 &= K_2^{\text{eq}} \left(s_1 - \frac{J\beta_2}{\varepsilon_2} \right) = K_2^{\text{eq}} \left(K_1^{\text{eq}} \left(s_0 - \frac{J\beta_1}{\varepsilon_1} \right) - \frac{J\beta_2}{\varepsilon_2} \right) \\
&= s_0 K_1^{\text{eq}} K_2^{\text{eq}} - J \left(\frac{\beta_1}{\varepsilon_1} K_1^{\text{eq}} K_2^{\text{eq}} + \frac{\beta_2}{\varepsilon_2} K_2^{\text{eq}} \right) \\
s_3 &= s_0 K_1^{\text{eq}} K_2^{\text{eq}} K_3^{\text{eq}} - J \left(\frac{\beta_1}{\varepsilon_1} K_1^{\text{eq}} K_2^{\text{eq}} K_3^{\text{eq}} + \frac{\beta_2}{\varepsilon_2} K_2^{\text{eq}} K_3^{\text{eq}} + \frac{\beta_3}{\varepsilon_3} K_3^{\text{eq}} \right) \\
&\vdots \\
s_n &= s_0 \prod_{i=1}^{n} K_i^{\text{eq}} - J \sum_{j=1}^{n} \frac{\beta_j}{\varepsilon_j} \prod_{i=j}^{n} K_i^{\text{eq}} = s_0 \; K_{\text{tot}}^{\text{eq}} - J \sum_{j=1}^{n} \frac{\gamma_j}{\varepsilon_j}
\end{aligned} \tag{103}$$

where we use the fact that $K^{\text{eq}} = \prod_{i=1}^{n} K_i^{\text{eq}}$ (i.e. the equilibrium constant of the pathway net reaction), and substituting $\gamma_j \equiv \beta_j \prod_{i=j}^{n} K_i^{\text{eq}}$. Solving for J we get:

$$J = \frac{s_0 \; K_{\text{tot}}^{\text{eq}} \; - \; s_n}{\sum_{j=1}^{n} \gamma_j (\varepsilon_j)^{-1}} \, . \tag{104}$$

F.4.2 Flux control coefficients

The steady-state flux is given by Eq. (104). By taking the partial derivative with respect to enzyme levels, we obtain the flux response coefficients

$$R_{\varepsilon_l}^{J} = \frac{\partial}{\partial \varepsilon_l} \frac{s_0 \; K_{\text{tot}}^{\text{eq}} \; - \; s_n}{\sum_{i=1}^{n} \gamma_i / \varepsilon_i} = - \frac{s_0 \; K_{\text{tot}}^{\text{eq}} \; - \; s_n}{(\sum_{i=1}^{n} \gamma_i / \varepsilon_i)^2} \cdot \left(- \frac{\gamma_l}{\varepsilon_l^2} \right) = J \; \frac{\gamma_l / \varepsilon_l^2}{\sum_{i=1}^{n} \gamma_i / \varepsilon_i} . \tag{105}$$

After dividing by the enzyme elasticities $E_{\varepsilon_l}^{v_l} = J/\varepsilon_l$, we obtain the flux control coefficients

$$C^J = \frac{R_{\varepsilon_l}^{J}}{E_{\varepsilon_l}^{v_l}} = \frac{\gamma_l / \varepsilon_l}{\sum_{i=1}^{n} \gamma_i / \varepsilon_i} . \tag{106}$$

F.4.3 Minimum enzyme demand / maximal flux

Now, in order to maximize J we need to minimize the denominator (the numerator is a constant). The variables (ε_j) have to satisfy the constraint $\sum_j \varepsilon_j \leq \varepsilon_{\text{tot}}$. To find the minimal value for the denominator, we use the Lagrange method:

$$0 = \frac{\partial}{\partial \varepsilon_j} \left[\sum_{j=1}^{n} \gamma_j (\varepsilon_j)^{-1} - \lambda \left(\sum_j \varepsilon_j - \varepsilon_{\text{tot}} \right) \right]_{\varepsilon_j{}^*} = -\gamma_j (\varepsilon_j{}^*)^{-2} - \lambda \tag{107}$$

which shows that $\varepsilon_j{}^* \propto \sqrt{\gamma_j}$. Since we know the sum of all the enzyme levels, ε_{tot}, we can also find the proper scaling factor, i.e.:

$$\varepsilon_j{}^* = \varepsilon_{\text{tot}} \; \frac{\sqrt{\gamma_j}}{\sum_{j=1}^{n} \sqrt{\gamma_j}} = \varepsilon_{\text{tot}} \sqrt{\frac{\gamma_j}{||\boldsymbol{\gamma}||_{\frac{1}{2}}}} \tag{108}$$

where $\boldsymbol{\gamma}$ is the vector of all γ_j, and $|| \cdot ||_{\frac{1}{2}}$ is the $\mathrm{l}_{1/2}$ norm.

Furthermore, we can calculate the maximal achievable flux J^* by plugging in the $\varepsilon_j{}^*$ values:

$$J^* = \frac{s_0 \; K_{\text{tot}}^{\text{eq}} \; - \; s_n}{\sum_{j=1}^{n} \gamma_j (\varepsilon_j{}^*)^{-1}} = \frac{s_0 \; K_{\text{tot}}^{\text{eq}} \; - \; s_n}{\sum_{j=1}^{n} \gamma_j \; \frac{1}{\varepsilon_{\text{tot}}} \sqrt{\frac{||\boldsymbol{\gamma}||_{\frac{1}{2}}}{\gamma_j}}} = \varepsilon_{\text{tot}} \frac{s_0 \; K_{\text{tot}}^{\text{eq}} \; - \; s_n}{\sqrt{||\boldsymbol{\gamma}||_{\frac{1}{2}}} \sum_{j=1}^{n} \sqrt{\gamma_j}} = \varepsilon_{\text{tot}} \; \frac{s_0 \; K_{\text{tot}}^{\text{eq}} \; - \; s_n}{||\boldsymbol{\gamma}||_{\frac{1}{2}}} \tag{109}$$

An interesting consequence of Eq. (108) is that we can draw this optimization, is looking at optimal ratios between consecutive enzymes:

$$\frac{\varepsilon_j{}^*}{\varepsilon_{j-1}{}^*} = \sqrt{\frac{\gamma_j}{\gamma_{j-1}}} = \sqrt{\frac{\beta_j}{\beta_{j-1}\ K_j^{\mathrm{eq}}}} = \sqrt{\frac{K_{\mathrm{M},j}\ k_{j-1}^{\mathrm{cat}}}{K_{\mathrm{M},j-1}\ k_j^{\mathrm{cat}}\ K_j^{\mathrm{eq}}}}\,. \tag{110}$$

Interestingly, the ratio depends only on the K_{M} and k^{cat} values of the two reactions and is independent of all other enzymes in the system. This might be a design principle supporting for the existence of regulatory mechanisms that ensure these ratios are maintained even if the total expression level changes.

F.4.4 Connectivity theorem for flux control coefficients

The connectivity theorem for flux control coefficients reads $\mathbf{C}^J\ \mathbf{E}_c^v = 0$. Here we verify this explicitly for an unbranched pathway with reversible mass-action kinetics. With the control coefficients Eq. (106) and the metabolite elasticities for the reversible mass action rate law

$$E_{c_i}^{v_l} = \varepsilon_l\ \beta_l^{-1}\ (\delta_{i,l-1} - \delta_{il}/K_l^{\mathrm{eq}}). \tag{111}$$

we obtain

$$\begin{aligned}(\mathbf{C}^J\ \mathbf{E}_c^v)_i &= \sum_l \frac{\gamma_l/\varepsilon_l}{\sum_j \gamma_j/\varepsilon_j}\ \varepsilon_l\ \beta_l^{-1}\ (\delta_{i,l-1} - \delta_{il}/K_l^{\mathrm{eq}}) \propto \sum_l \gamma_l\ \beta_l^{-1}\ (\delta_{i,l-1} - \delta_{il}/K_l^{\mathrm{eq}}) \\ &= \gamma_{i+1}\ \beta_{i+1}^{-1} - \gamma_i\ \beta_i^{-1}\ \cdot\ 1/K_i^{\mathrm{eq}} \\ &= \prod_{j=i+1}^{n} K_j^{\mathrm{eq}} - \prod_{j=i}^{n} K_j^{\mathrm{eq}}\ \cdot\ 1/K_i^{\mathrm{eq}} \\ &= \prod_{j=i+1}^{n} K_j^{\mathrm{eq}} - \prod_{j=i+1}^{n} K_j^{\mathrm{eq}} = 0\,.\end{aligned} \tag{112}$$

F.4.5 Enzyme-control rule

The enzyme levels in optimal state Eq. (27) read

$$\varepsilon_l^* = \varepsilon_{\mathrm{tot}} \sqrt{\frac{\gamma_l}{||\boldsymbol{\gamma}||_{\frac{1}{2}}}}. \tag{113}$$

By inserting this into the formula Eq. (106) for control coefficients, we obtain (where the constant denominator is a normalisation term)

$$C_l^{J^*} \equiv \frac{\gamma_l\ \left(\varepsilon_{\mathrm{tot}} \sqrt{\frac{\gamma_l}{||\boldsymbol{\gamma}||_{\frac{1}{2}}}}\right)^{-1}}{\mathrm{const}} \propto \sqrt{\gamma_l}, \tag{114}$$

which is proportional to the enzyme level ε_l^* itself.

F.5 Haldane rate law

F.5.1 Steady-state concentrations

For this derivation, we first use the notation in the original publication by Heinrich and Klipp [17]. For a unbranched pathway at steady-state (with flux J in all reactions), the Haldane rate law dictates that:

$$J = \frac{v_{+l}^{\max}\ s_{l-1}/K_{\mathrm{S};l} - v_{-l}^{\max}\ s_l/K_{\mathrm{P};l}}{1 + s_{l-1}/K_{\mathrm{S};l} + s_l/K_{\mathrm{P};l}}\,. \tag{115}$$

Using this formula, we can solve for s_l given all other parameters (including s_{l-1}) and get the following recursion:

$$s_l = s_{l-1} \frac{K_{\mathrm{P};l}}{K_{\mathrm{S};l}} \frac{v_{+l}^{\max} - J}{v_{-l}^{\max} + J} - J \frac{K_{\mathrm{P};l}}{v_{-l}^{\max} + J} \tag{116}$$

which can be solved to give:

$$s_l^{\mathrm{st}} = s_0 \left(\prod_{i=1}^{l} \frac{K_{\mathrm{P};i}}{K_{\mathrm{S};i}} \frac{v_{+i}^{\max} - J}{v_{-i}^{\max} + J} \right) - J \sum_{i=1}^{l} \frac{K_{\mathrm{S};i}}{v_{+i}^{\max} - J} \prod_{j=i}^{l} \frac{K_{\mathrm{P}};j}{K_{\mathrm{S}};j} \frac{(v_{+j}^{\max} - J)}{(v_{-j}^{\max} + J)} \tag{117}$$

From now on, we will replace $v_{\pm i}^{\max}$ with $\varepsilon_i \cdot k_{\pm i}^{\mathrm{cat}}$ in order to better match the notation of this paper. If we use the solution for the steady-state concentration of s_n we can write:

$$f(J;\boldsymbol{\varepsilon}) \equiv s_0 \left(\prod_{i=1}^{n} \frac{K_{\mathrm{P};i}}{K_{\mathrm{S};i}} \frac{\varepsilon_i\ k_{+i}^{\mathrm{cat}} - J}{\varepsilon_i\ k_{-i}^{\mathrm{cat}} + J} \right) - J \sum_{i=1}^{n} \frac{K_{\mathrm{S};i}}{\varepsilon_i\ k_{+i}^{\mathrm{cat}} - J} \prod_{j=i}^{n} \frac{K_{\mathrm{P};j}}{K_{\mathrm{S};j}} \frac{(\varepsilon_i\ k_{+j}^{\mathrm{cat}} - J)}{(\varepsilon_i\ k_{-j}^{\mathrm{cat}} + J)} - s_n = 0\,. \tag{118}$$

Solving this equality (i.e., finding the roots of $f(J;\boldsymbol{\varepsilon})$) would give us an expression for $J(\boldsymbol{\varepsilon})$ – the flux as a function of the enzyme levels. This can be used in different ways. First, we can determine the flux response coefficients by implicit differentiation (Lemma F.2: $R_{\varepsilon_i}^J = -\frac{\partial f/\partial \varepsilon_i}{\partial f/\partial J}$), and by dividing by J/ε_i we obtain the flux control coefficients. Second, we can find an optimal metabolic state by minimizing $\sum_i \varepsilon_i$ at a given flux J, and requiring Eq. (118) as a constraint (with Lagrange multiplier λ). Optimization using the Lagrange method would lead to the optimality condition $\partial f/\partial \varepsilon_i = \lambda$. However, since the derivatives are complicated and J and λ are unknown, this is hard to solve.

The most efficient method we currently have for solving this optimization problem is, unfortunately, not analytical. First, we start by writing $\varepsilon_{\mathrm{tot}}/J$ as a function of the metabolite concentrations:

$$\frac{\varepsilon_{\mathrm{tot}}}{J} = \frac{\sum_i \varepsilon_i}{J} = \sum_i \frac{1 + s_{i-1}/K_{\mathrm{S};i} + s_i/K_{\mathrm{P};i}}{k_{+i}^{\mathrm{cat}}\ s_{i-1}/K_{\mathrm{S};i} - k_{+i}^{\mathrm{cat}}\ s_i/K_{\mathrm{P};i}}\,. \tag{119}$$

We showed previously that this function is convex with respect to the metabolite log-concentrations ($\ln \mathbf{s}$) and therefore has a single (global) minimum which is easy to find numerically [38, 8].

F.6 Reactions with cofactors, side substrates, and side products: effective kinetic constants

The pathways considered in this paper consist of uni-uni reactions S $\leftrightarrow$ P. However, all the results apply also to reactions with several substrates and products if only one of the substrates and one of the products have variable concentrations, while all other concentrations are predefined. As an illustrating example, we consider a reaction with stoichiometric coefficients of 1, substrates s (variable) and a (constant), and products p (variable) and b (constant). There are various ways to generalize the Haldane rate law to reactions with several substrates or products. One of them, the convenience kinetics [34], for this reaction reads

$$v = \varepsilon \frac{k_+^{\mathrm{cat}}\ s/K_{\mathrm{S}}\ a/K_{\mathrm{A}} - k_-^{\mathrm{cat}}\ p/K_{\mathrm{P}}\ b/K_{\mathrm{B}}}{(1 + s/K_{\mathrm{S}})(1 + a/K_{\mathrm{A}}) + (1 + p/K_{\mathrm{P}})(1 + b/K_{\mathrm{B}}) - 1}. \tag{120}$$

With the abbreviations $\sigma = a/K_{\mathrm{A}}$ and $\pi = b/K_{\mathrm{B}}$, we obtain

$$\begin{aligned} v &= \varepsilon \frac{k_+^{\mathrm{cat}}\ \sigma\ s/K_{\mathrm{S}} - k_-^{\mathrm{cat}}\ \pi\ p/K_{\mathrm{P}}}{(1 + s/K_{\mathrm{S}})(1 + \sigma) + (1 + p/K_{\mathrm{P}})(1 + \pi) - 1} \\ &= \varepsilon \frac{k_+^{\mathrm{cat}}\ \sigma\ s/K_{\mathrm{S}} - k_-^{\mathrm{cat}}\ \pi\ p/K_{\mathrm{P}}}{(1 + \sigma) + (1 + \pi) - 1 + (1 + \sigma)s/K_{\mathrm{S}} + (1 + \pi)p/K_{\mathrm{P}}} \\ &= \varepsilon \frac{\frac{\sigma}{1+\sigma+\pi} k_+^{\mathrm{cat}} s/K_{\mathrm{S}} - \frac{\pi}{1+\sigma+\pi} k_-^{\mathrm{cat}} p/K_{\mathrm{P}}}{1 + \frac{1+\sigma}{1+\sigma+\pi}\ s/K_{\mathrm{S}} + \frac{1+\pi}{1+\sigma+\pi}\ p/K_{\mathrm{P}}} \end{aligned} \tag{121}$$

and by defining

$$\begin{aligned} \widehat{K_{\mathrm{S}}} &= \frac{1 + \sigma + \pi}{1 + \sigma} K_{\mathrm{S}}, \qquad \widehat{K_{\mathrm{P}}} = \frac{1 + \sigma + \pi}{1 + \pi} K_{\mathrm{P}} \\ \widehat{k_+^{\mathrm{cat}}} &= \frac{\sigma}{1 + \sigma} k_+^{\mathrm{cat}}, \qquad \widehat{k_-^{\mathrm{cat}}} = \frac{\pi}{1 + \pi} k_-^{\mathrm{cat}} \end{aligned} \tag{122}$$

we can write this in the form of our uni-uni rate law:

$$v = \varepsilon \cdot \frac{\widehat{k_+^{\text{cat}}}\ s/\widehat{K_\text{S}} - \widehat{k_-^{\text{cat}}}\ p/\widehat{K_\text{P}}}{(1 + s/\widehat{K_\text{S}}) + (1 + p/\widehat{K_\text{P}}) - 1} \tag{123}$$

with effective ($\hat{\cdot}$) parameters. From the Haldane relationship, we obtain the effective equilibrium constant:

$$\widehat{K^{\text{eq}}} = \frac{\widehat{k_+^{\text{cat}}}}{\widehat{k_-^{\text{cat}}}} \cdot \frac{\widehat{K_\text{P}}}{\widehat{K_\text{S}}} = \frac{\sigma}{\pi} \frac{\cancel{(1+\pi)}}{\cancel{(1+\sigma)}} \frac{k_+^{\text{cat}}}{k_-^{\text{cat}}} \cdot \frac{\cancel{(1+\sigma+\pi)}}{\cancel{(1+\sigma+\pi)}} \frac{\cancel{(1+\sigma)}}{\cancel{(1+\pi)}} \frac{K_\text{P}}{K_\text{S}} = \frac{a}{b} \frac{k_+^{\text{cat}}}{k_-^{\text{cat}}} \frac{K_\text{P} K_\text{B}}{K_\text{S} K_\text{A}} = \frac{a}{b} K^{\text{eq}}, \tag{124}$$

where we used the Haldane relationship for the original reaction: $K^{\text{eq}} = \frac{k_+^{\text{cat}}}{k_-^{\text{cat}}} \frac{K_\text{P} K_\text{B}}{K_\text{S} K_\text{A}}$.

The driving force $\theta = \ln K^{\text{eq}} - \ln \frac{p}{s} \frac{b}{a}$ of a reaction, as a thermodynamic quantity, cannot depend on how the rate law is written. With the original rate law, it depends on the (true) equilibrium constant as well as on all four reactant concentrations. But if we ignore A and B in the rate law and simply define our equilibrium constant as the ratio b^{eq}/a^{eq} in some imagined equilibrium state, the resulting driving force would be different from our original driving force and the term $1 - \mathrm{e}^{-\theta}$ will be wrong or, even worse, the new θ will have the opposite sign and be opposite to the flux direction. To avoid this problem, there are two possibilities: either we use the effective equilibrium constant Eq. (124), which contains the constant ignored concentrations, or we compute the driving force with an "external" extra term,

$$\theta = \ln K^{\text{eq}} - \underbrace{\ln \frac{b}{a}}_{\theta_{\text{ext}}} - \ln \frac{p}{s}. \tag{125}$$

If we compare the thermodynamic force to voltage in an electric circuit, A and B would act like an external voltage source.

F.7 Enzyme-control rule in models with constraints on enzyme and metabolite levels

We assume an unbranched metabolic pathway in steady state. The steady-state fluxes $\mathbf{v} = \mathbf{v}^{\text{st}}(\boldsymbol{\varepsilon})$ and steady-state metabolite concentrations $\mathbf{s} = \mathbf{s}^{\text{st}}(\boldsymbol{\varepsilon})$, differentiated by ε_l, yield the response coefficients $R^{v_i}_{\varepsilon_l} = \partial v_i^{\text{st}}/\partial \varepsilon_l$ and $R^{s_i}_{\varepsilon_l} = \partial s_i^{\text{st}}/\partial \varepsilon_l$. We now consider the optimality problem (same as in Eq. (40)):

$$\text{Maximize } \mathbf{z} \cdot \mathbf{v} \quad \text{s.t.} \quad \mathbf{a} \cdot \boldsymbol{\varepsilon} + \mathbf{b} \cdot \mathbf{s} \leq \rho$$

where $\mathbf{v}$, $\mathbf{s}$, and $\boldsymbol{\varepsilon}$ denote reaction fluxes, metabolite concentrations, and enzyme levels. The maximization objective is a linear benefit function scoring the fluxes, $\mathbf{z} \cdot \mathbf{v}$ with linear weights z_i, while the constrained density is linear in all the concentrations, with linear weights a_l (for enzymes) and b_i (for metabolites), e.g. representing effective molecule sizes. Optimization with a Lagrange multiplier λ yields

$$\text{Maximize} \quad \mathcal{L}(\boldsymbol{\varepsilon}) = \mathbf{z}^\top \mathbf{v}^{\text{st}}(\boldsymbol{\varepsilon}) - \lambda \left[\mathbf{a}^\top \boldsymbol{\varepsilon} + \mathbf{b}^\top \mathbf{s}^{\text{st}}(\boldsymbol{\varepsilon}) - \rho\right], \tag{126}$$

where $\mathbf{z}$ is the flux weight vector, $\mathbf{a}$ and $\mathbf{b}$ are the molecular masses of the enzymes and metabolites (respectively), and ρ is the upper bound on the total density. The Lagrangian yields the optimality condition:

$$\begin{aligned} 0 = \frac{\partial \mathcal{L}}{\partial \boldsymbol{\varepsilon}} &= \mathbf{z}^\top \frac{\partial \mathbf{v}^{\text{st}}}{\partial \boldsymbol{\varepsilon}} - \lambda \left[\mathbf{a}^\top + \mathbf{b}^\top \frac{\partial \mathbf{s}^{\text{st}}}{\partial \boldsymbol{\varepsilon}}\right] \\ &= \mathbf{z}^\top \mathbf{R}^{\mathbf{v}}_{\varepsilon} - \lambda \left[\mathbf{a}^\top + \mathbf{b}^\top \mathbf{R}^{\mathbf{s}}_{\varepsilon}\right] \\ \Rightarrow \mathbf{a}^\top &= \frac{1}{\lambda} \mathbf{z}^\top \mathbf{R}^{\mathbf{v}}_{\varepsilon} - \mathbf{b}^\top \mathbf{R}^{\mathbf{s}}_{\varepsilon}. \end{aligned} \tag{127}$$

By splitting the response matrices into control and enzyme elasticity matrices, $\mathbf{R}^x_\varepsilon = \mathbf{C}^x E^v_\varepsilon = \mathbf{C}^x \operatorname{diag}(\mathbf{v}) \operatorname{diag}(\boldsymbol{\varepsilon})^{-1}$, we next obtain

$$\begin{aligned} \mathbf{a}^\top &= \left[\frac{1}{\lambda} \mathbf{z}^\top \ \mathbf{C}^{\mathbf{v}} - \mathbf{b}^\top \mathbf{C}^{\mathbf{s}}\right] \ \operatorname{diag}(\mathbf{v}) \operatorname{diag}(\boldsymbol{\varepsilon})^{-1} \\ \Rightarrow \boldsymbol{\varepsilon}^\top \operatorname{diag}(\mathbf{a}) \operatorname{diag}(\mathbf{v})^{-1} &= \left[\frac{1}{\lambda} \ \mathbf{z}^\top \mathbf{C}^{\mathbf{v}} - \mathbf{b}^\top \mathbf{C}^{\mathbf{s}}\right]. \end{aligned} \tag{128}$$

In the case of an unbranched pathway, i.e. with the same steady-state $v_i^{\text{st}} = J$ in all reactions, we can arbitrarily chose a benefit function $\mathbf{z}^\top = (1, 0, \ldots, 0)$ which will give use $\mathbf{z}^\top \mathbf{v} = v_1 = J$. In this case $\mathbf{z}^\top \mathbf{C}^{\mathbf{v}}$ will simply be the

first column of the $\mathbf{C}^{\mathbf{v}}$ matrix, which we will refer to as $\mathbf{C}^J$. In addition, we can choose density constraints with uniform weights for all enzymes and for all metabolites ($\mathbf{a} = a\mathbf{1}$, $\mathbf{b} = b\mathbf{1}$). Then, the enzyme profile becomes:

$$\boldsymbol{\varepsilon} = \frac{J}{a}\left[\frac{1}{\lambda}\,\mathbf{C}^J - b\,\mathbf{1}^\top\mathbf{C}^{\mathrm{s}}\right] \tag{129}$$

Our new enzyme-control rule thus reads, as a proportionality (and marking again the optimal state by *)

$$\varepsilon_l^* \propto C_l^{J*} - \lambda\, b\, \sum_i C_l^{s_i*}. \tag{130}$$

In our density constraint in Eq. (40), with a predefined overall density ρ, there is no explicit bound $\varepsilon_{\mathrm{tot}}$ on the sum of enzyme levels. However, given the sum of enzyme levels $\varepsilon_{\mathrm{tot}}^*$ that emerges from the solution, we obtain an explicit enzyme-control rule with no unknown parameters. To derive it, we write Eq. (129) in the form

$$\varepsilon_l^* = \frac{J}{a\,\lambda} C_l^{J*} - \frac{bJ^*}{a}\,\sum_i C_l^{s_i*} \tag{131}$$

We can then sum over l and get:

$$\varepsilon_{\mathrm{tot}}^* = \sum_l \varepsilon_l^* = \frac{J^*}{a\,\lambda}\sum_l C_l^{J*} - \frac{bJ^*}{a}\,\sum_i\sum_l C_l^{s_i*} = \frac{J^*}{a\,\lambda}\cdot 1 - \frac{bJ}{a}\cdot\sum_i 0, \tag{132}$$

where in the last step we used the summation theorems for fluxes and metabolite concentrations. We can solve this for $\lambda = J^*\,(a\,\varepsilon_{\mathrm{tot}})^{-1}$, and the optimal enzyme levels read

$$\varepsilon_l^* = \varepsilon_{\mathrm{tot}}^*\, C_l^{J*} - \frac{b}{a}\,J^*\,\sum_i C_l^{s_i*}. \tag{133}$$

F.8 Sufficient condition for stable states in unbranched pathways

Here we show, that a steady state in an unbranched metabolic pathway is stable if (but not only if) the first reaction 1 is reversible (i.e. it has a non-zero product elasticity), if the last reaction is not completely saturated (i.e. it has a non-zero substrate elasticity), and if in all other reactions the substrate elasticity is larger than the (absolute) product elasticity. Note that for reversible mass-action rate laws, the first two conditions are automatically satisfied and the latter condition is satisfied if $k_{+i} > k_{-(i-1)}$.

Proof A state is asymptotically stable if all the Jacobian eigenvalues are negative (i.e. they have negative real parts). A sufficient (but not necessary) condition for stable metabolic states can be obtained from the Gershgorin's disc theorem, which we recall here:

Theorem F.3. *Let* $\mathbf{A}$ *be a square matrix. Each diagonal element* a_{ii} *is associated with a closed disc* $s_i = D_i(a_{ii}, \sum_j |a_{ij}|))$ *in the complex plane, with center* a_{ii} *and radius* $\sum_j |a_{ij}|$. *According to the theorem, all eigenvalues of* $\mathbf{A}$ *must lie in the union of all these discs. Since the theorem applies both to* $\mathbf{A}$ *and to its transpose, each diagonal element gives rise to two discs with the same center, but different radii* $\sum_j |a_{ij}|$ *and* $\sum_j |a_{ji}|$.

For a real-valued matrix, this means: our system is stable if for all i, $a_{ii}+\sum_j |a_{ij}| < 0$ or if for all i, $a_{ii}+\sum_j |a_{ji}| < 0$.

We now apply this theorem to an unbranched metabolic chain with 4 reactions and 3 internal metabolites A, B, C:

$$S \xrightarrow{1} A \xrightarrow{2} B \xrightarrow{3} C \xrightarrow{4} P \tag{134}$$

The system has no conservation relations. From the stoichiometric matrix

$$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \tag{135}$$

and the metabolite elasticity matrix

$$\mathbf{E}_{\mathrm{c}} = \begin{pmatrix} -g_A^1 & 0 & 0 \\ g_A^2 & -g_B^2 & 0 \\ 0 & g_B^3 & -g_C^3 \\ 0 & 0 & g_C^4 \end{pmatrix}, \tag{136}$$

parameterised by positive values g, we obtain the Jacobian matrix

$$\mathbf{A} = \mathbf{N}\ \mathbf{E}_{\mathrm{c}} \begin{pmatrix} -(g_A^1 + g_A^2) & g_B^2 & 0 \\ g_A^2 & -(g_B^2 + g_B^3) & g_C^3 \\ 0 & g_B^3 & -(g_C^3 + g_C^4) \end{pmatrix}. \tag{137}$$

Applying Gershgorin's theorem, from the condition for matrix columns[5], we obtain the sufficient stability conditions

$$\begin{aligned} -(g_A^1 + g_A^2) + g_A^2 &< 0 \\ -(g_B^2 + g_B^3) + (g_B^2 + g_C^3) &< 0 \\ -(g_C^3 + g_C^4) + g_C^3 &< 0. \end{aligned} \tag{139}$$

To guarantee a stable steady state, all three inequalities must hold. They are satisfied whenever $g_A^1 > 0$ (reaction 1 is not completely irreversible), $g_C^4 > 0$ (reaction 4 is not completely saturated), and $g_B^3 > g_C^3$ (the substrate elasticity of reaction 3 is larger than the (absolute) product elasticity of this reaction). For longer metabolic chains, we obtain the same type of conditions: the steady state will be stable if (but not only if) the first reaction is not completely irreversible, the last reaction is not completely saturated, and in every reaction in between, the substrate elasticity is larger than the (absolute) product elasticity.

[5]Incidentally, from the condition for rows we obtain another (alternative) set of sufficient conditions

$$\begin{aligned} -(g_A^1 + g_A^2) + g_B^2 &< 0 \\ -(g_B^2 + g_B^3) + (g_A^2 + g_C^3) &< 0 \\ -(g_C^3 + g_C^4) + g_B^3 &< 0. \end{aligned} \tag{138}$$

These conditions are satisfied, for example, if all elasticities are non-zero, if the product elasticity g_A^1 of the first reaction is equal or larger than the product elasticity g_B^2 of the second reaction, if the substrate elasticity of the last reaction is equal or larger that the substrate elasticity g_C^3 of the third reaction, and if $g_A^2 - g_B^2 < g_B^3 - g_C^3$. For a reversible mass-action kinetics $v = e[k_+ s - k_- p]$ the latter condition would mean: $\varepsilon_2[k_+^2 - k_-^2] < \varepsilon_3[k_+^3 - k_-^3]$ or $\frac{\varepsilon_3}{\varepsilon_2} > \frac{k_+^2 - k_-^2}{k_+^3 - k_-^3} = \frac{v_2^{\mathrm{std}}}{v_3^{\mathrm{std}}}$ where the "standard velocity" v_l^{std} denotes the rate that a reaction l would show at unit metabolite and enzyme levels.