

# Resource allocation in biochemically structured metabolic networks

Leon Seeger<sup>a</sup>, Fernanda Pinheiro<sup>b</sup>, Michael Lässig<sup>a,1</sup>

March 27, 2024

<sup>a</sup>Institute for Biological Physics, University of Cologne, 50923 Cologne, Germany

<sup>b</sup>Human Technopole, I-20157 Milan, Italy

<sup>1</sup>To whom correspondence should be addressed. E-mail: mlaessig@uni-koeln.de

## Abstract

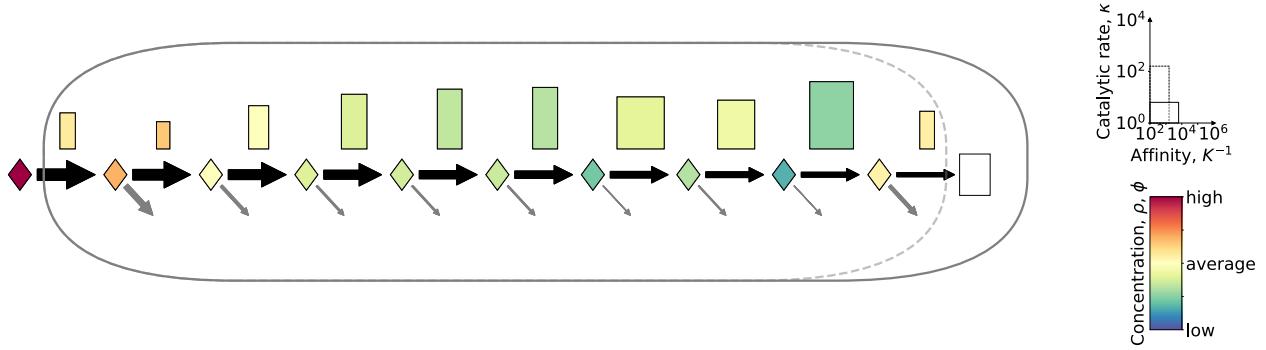
Microbes tune their metabolism to environmental challenges by changing protein expression levels, metabolite concentrations, and reaction rates simultaneously. Here, we establish an analytical model for microbial resource allocation that integrates enzyme biochemistry and the global architecture of metabolic networks. We describe the production of protein biomass from external nutrients in pathways of Michaelis-Menten enzymes and compute the resource allocation that maximizes growth under constraints of mass conservation and metabolite dilution by cell growth. This model predicts generic patterns of growth-dependent microbial resource allocation to proteome and metabolome. In a nutrient-rich medium, optimal protein expression depends primarily on the biochemistry of individual synthesis steps, while metabolite concentrations and fluxes decrease along successive reactions in a metabolic pathway. Under nutrient limitation, individual protein expression levels change linearly with growth rate, the direction of change depending again on the enzyme's biochemistry. Metabolite levels and fluxes show a stronger, nonlinear decline with growth rate. We identify a simple, metabolite-based regulatory logic by which cells can be tuned to near-optimal growth. Finally, our model predicts evolutionary stable states of metabolic networks, including local biochemical parameters and the global metabolite mass fraction, in tune with empirical data.

## Introduction

Bacteria exhibit remarkable plasticity when reacting to changes in their environment (1–2). Understanding and harvesting this plasticity promises advances in medicine, synthetic biology, and biotechnology (3–6). Recent improvements in systems-wide metabolic models progress towards this goal (7–10). However, the complexity of metabolic networks and the vast amount of potential model parameters limit a predictive understanding of microbial physiology on the whole-genome level (11–13).

At the same time, simpler empirical patterns emerge from this complexity. Metabolic flux-balance analysis (14), bacterial growth laws governing proteome resource allocation (15), and enzyme-substrate scaling relationships (16) isolate sets of metabolic variables that can be explained without detailed knowledge of the others. Such emergent principles have been used successfully to model the dose-response to ribosome-targeting antimicrobials in diverse environments (17) and to predict dosage-dependent resistance evolution (18). However, dose-response modeling for target enzymes embedded more deeply in the metabolic network than the ribosome (19, 20) remains difficult in general.

In this paper, we integrate enzyme biochemistry and the topology of metabolic pathways into an analytically solvable model of networks with biochemically heterogeneous constituents. The model predicts resource allocation at a multi-omics level including fluxes, the metabolome, and the proteome, as shown schematically in Fig. 1 for a simple chain topology. We consider multi-component metabolic pathways that produce and consume intermediate metabolites, and we compute the state of the cell generating maximal growth under the constraint of mass conservation in a population of replicating cells. This is a step towards an analytical solution of optimal states in large metabolic networks, including realistic non-linear kinetics, which has so far remained elusive (21, 22). The model predicts broad patterns describing how fluxes, protein expression, and steady-state metabolite concentrations depend on an enzyme's position in the network and on the nutrient environment.



**Figure 1: Biochemical model of cell metabolism.** A set of reaction pathways processes external nutrients into protein production, leading to cell growth. A simple network is a metabolic chain of  $\ell$  enzymes (boxes) and  $\ell$  metabolites (diamonds), shown here for  $\ell = 10$ . Each reaction step is characterized by Michaelis-Menten kinetics with catalytic capacity  $\kappa_i^{\max}$  (box height) and substrate binding affinity  $K_i^{-1}$  (box width) ( $i = 1, \dots, \ell$ ). The first metabolite is an external nutrient of mass density  $\rho_1$ , the output of the network is protein production at rate  $\lambda$ . The model developed in this paper predicts growth-optimal, balanced growth states in nutrient-rich environments and under nutrient limitation. Such states are characterized by steady-state metabolite concentrations,  $\rho_i$ , and protein expression levels,  $\phi_i$  (color shading). The network produces a decreasing sequence of mass fluxes,  $j_i$  (black arrows), with loss of metabolites by growth (gray arrows).

A key network parameter, called the metabolite cost, summarizes the effects of nutrient level and upstream metabolism on a given metabolite-enzyme pair. This constraint combines with local metabolic reaction kinetics to set the resource allocation to individual metabolites and enzymes. Under nutrient restriction, an enzyme is linearly up- or down-regulated with decreasing growth rate. The direction of change depends on a local biochemical parameter negatively correlated with the enzyme's target affinity and positively correlated with its catalytic rate. This pattern explains the heterogeneity of single-enzyme growth responses found in proteomics studies of microbial growth (23–25).

An integrative metabolic model of enzymes and metabolites is key to understand how regulation and evolution shape systems metabolism. Here we introduce a simple regulatory scheme that suggests how cells can be tuned to near-optimal growth by using metabolites as transcriptional signals for protein expression. On longer time scales, evolution alters network biochemistry. Specifically, we show that physiological values of enzyme kinetics emerge from an evolutionary equilibrium. This, in turn, sets the global mass ratio of metabolome and proteome in the cell.

The text is organized as follows. We first introduce the model framework and derive local optimality conditions that determine cellular resource allocation to a given metabolite-enzyme pair. We then solve these equations, first in a nutrient-rich environment, and discuss the resulting system-wide patterns of metabolite levels and protein expression. Next, we analyze the metabolic response to nutrient limitation, explore more complex metabolic networks, and compare our findings to published experimental data (16, 24). Finally, we discuss the implications of the model for regulation and evolution of metabolic networks.

## Metabolic model

**Networks of metabolites and enzymes.** We consider a metabolic network of  $\ell$  enzymes. Each enzyme catalyzes a reaction according to Michaelis-Menten kinetics: in a cell of volume  $V$ , the enzyme species present at mass  $M_i$  acts on a metabolite present at mass concentration  $\rho_i$  with catalytic rate  $\kappa_i^{\max}$  and substrate binding affinity  $K_i^{-1}$ , generating a mass flux

$$J_i = \frac{\kappa_i^{\max} M_i}{1 + \frac{K_i}{\rho_i}}. \quad (1)$$

The metabolite mass flux per unit of protein mass,  $\kappa_i = \kappa_i^{\max}/(1 + K_i/\rho_i)$  is the product of  $\kappa_i^{\max}$  and the saturation level of the enzyme with its substrate. In line with empirical observations (11, 26–28), we assume that the total intracellular protein concentration is constant (15); the cell volume is therefore proportional to the total protein biomass  $M_P = \sum_{i=1}^{\ell} M_i$ . Accordingly, we write concentrations as mass fractions,  $\rho_i = R_i/M_P$ ,  $\phi_i = M_i/M_P$ , and we use mass-normalized fluxes,  $j_i = J_i/M_P$ , and dimensionless affinities,  $K_i^{-1}$  ( $i = 1, \dots, \ell$ ). Details are given in Supporting Information (SI). These affinities set the

ratio of intracellular metabolite biomass to protein biomass,  $M_R/M_P = \sum_{i=2}^{\ell} \rho_i$ , a fundamental global observable of cell metabolism. The metabolic state of the system, denoted by the shorthand  $\mathcal{S} \equiv (\rho, \phi)$ , changes in response to the environment. The set of kinetic constants,  $(\kappa^{\max}, K)$ , is fixed for a given organism but changes by evolution on longer time scales.

We start with the simplest case of a metabolic network, which consists of a single linear pathway where the metabolite produced by one enzyme is the substrate of the next (Fig. 1). Overall, the network converts a single environmental nutrient, present at concentration  $\rho_1$ , into protein biomass with a rate

$$\lambda \equiv \frac{J_\ell}{M_P}. \quad (2)$$

This defines the growth rate of the cell population. In the case of a single pathway, mass conservation on all intracellular metabolites takes the form

$$\frac{d\rho_i}{dt} = \frac{J_{i-1} - J_i}{M_P} - \lambda \rho_i \quad (i = 2, \dots, \ell). \quad (3)$$

We consider balanced exponential growth states [29] ( $\rho_i = \text{const.}$  for  $i = 2, \dots, \ell$ ) in a time-independent environment  $\rho_1$ . Importantly, these states maintain a flux mismatch  $\lambda \rho_i$  between metabolite in- and outflux at each metabolic step, reflecting the dilution of metabolites by growth [8, 22, 30, 33]. With the boundary condition  $j_\ell = \lambda$ , we obtain a decreasing cascade of steady-state mass fluxes,

$$j_i = \lambda \left( 1 + \sum_{k=i+1}^{\ell} \rho_k \right). \quad (4)$$

**Optimal balanced growth states.** In the following, we consider metabolic states that maximize growth rate under given nutrient conditions,  $\mathcal{S}^* = (\rho^*, \phi^*) = \arg \max(\lambda(\mathcal{S}))|_{\rho_1}$ . Under laboratory conditions, growth rate is often closely related to fitness [31], and bacteria allocate their resources close to the growth-optimal allocation [35]. This motivates our assumptions on stationarity and growth-maximization as a reasonable approximation of more complex growth cycles and evolutionary objectives [5].

Finding optimal balanced growth states analytically is challenging because the kinetic rate law is nonlinear and there is a global feedback loop imposed by growth dilution: the dilution rate in each step is set by growth rate, which itself depends on the output of the pathway. Environmental conditions and cell growth thus impose constraints from opposite ends of the metabolic cascade. Partially numerical solutions [30], specific properties of optimal solutions [8, 22], as well as an analytical solution of a pathway with two enzymes [33] were studied previously. Here, starting from the two-enzyme system, we iteratively construct optimal balanced growth states of longer pathways by adding a substrate-enzyme pair to the front end (SI, Fig. S1).

By equation 2, maximal growth is achieved if the protein mass necessary to produce a given protein production flux  $J_\ell$  is minimal [36]. We use the Michaelis-Menten relation 1 and the continuity equation 4 for each reaction to express the enzyme masses  $M_k$  in an arbitrary stationary state in terms of the nutrient level  $\rho_1$ , the current vector  $J = (J_1, \dots, J_\ell)$ , and the set of kinetic constants  $(\kappa^{\max}, K)$ . Thus, we can formulate the maximum-growth condition in differential form,

$$\frac{\partial \sum_{k=1}^{\ell} M_k(\rho_1, J, \kappa^{\max}, K)}{\partial J_i} \bigg|_{J^*, \rho_1} = 0 \quad (i = 1, \dots, \ell-1). \quad (5)$$

When a single flux  $J_i$  is varied, only the metabolite concentrations  $\rho_i, \rho_{i+1}$  and the enzyme masses  $M_i, M_{i+1}$  change (Fig. S1C). Inserting the Michaelis-Menten relationship 1 for  $M_i$  and  $M_{i+1}$ , as well as equation 4 for  $\rho_i$  and  $\rho_{i+1}$ , then yields a recursive relation for growth-optimal fluxes,

$$0 = \frac{1}{\kappa_i^{\max}} \left( 1 + \frac{K_i J_\ell^*}{\Delta J_i^*} \right) + \frac{J_i^*}{\kappa_i^{\max}} \frac{K_i J_\ell^*}{(\Delta J_i^*)^2} - \frac{J_{i+1}^*}{\kappa_{i+1}^{\max}} \frac{K_{i+1} J_\ell^*}{(\Delta J_{i+1}^*)^2}, \quad (6)$$

where  $\Delta J_i^* \equiv J_{i-1}^* - J_i^*$  ( $i = 2, \dots, \ell-1$ ). The initial condition  $(J_{\ell-1}^*, J_\ell^*)$  parameterizes the growth rate  $\lambda$  and the (a priori unknown) protein mass  $M_P$ . Solving this recursion produces the chain of metabolic fluxes,  $J^*$ ; by equations 1–4, we then obtain  $M_P$  and the mass-normalized metabolic state  $S^* = (\rho^*, \phi^*)$ . To study environmental perturbations, we can alternatively express  $S^*$  as a function of the nutrient level  $\rho_1$  or the growth rate  $\lambda$  as the independent control parameter. Details of the solution procedure are given in SI.

## Results

**Local optimality and metabolite cost.** We now study local properties of optimal balanced growth states in more detail. To this end, we consider a variation around the growth-optimal state that alters a single metabolite concentration  $\rho_i$ , whereas all other metabolite concentrations and the protein production flux  $J_\ell = \lambda M_P$  remain constant (Fig. S1D). The optimality criterion (5) takes the form

$$\left( \sum_{k=1}^{i-1} \frac{\partial M_k(J_k, \rho_k, \kappa^{\max}, K)}{\partial J_k} \frac{\partial J_k}{\partial \rho_i} + \frac{\partial M_i(J_i, \rho_i, \kappa^{\max}, K)}{\partial \rho_i} \right) \Big|_{J_\ell, \rho^*} = 0, \quad (7)$$

$$= 0, \quad (8)$$

where  $\partial M_k / \partial J_k = 1/\kappa_k$  by the Michaelis-Menten relation (1) and  $\partial J_k / \partial \rho_i = \lambda M_P$  by mass conservation (4). Thus, a small increase in concentration of the focal metabolite,  $\delta \rho_i > 0$ , is generated by an expression decrease of its cognate enzyme,  $\delta M_i < 0$ , and an increase of all upstream enzymes,  $\delta M_k > 0$  for  $k < i$ . All metabolic currents  $J_i, J_{i+1}, \dots, J_\ell$  and all masses  $M_{i+1}, \dots, M_\ell$  remain constant to first order. We can rewrite this relation for intensive variables,

$$\rho_i^* = \arg \min_{\rho_i} (\beta_i^* \rho_i + \phi_i(\rho_i, j_i^*, \kappa_i^{\max}, K_i)) \Big|_{j_i^*, \beta_i^*}, \quad (9)$$

where the *metabolite cost* parameter  $\beta_i = \sum_{k < i} \lambda^* / \kappa_k^*$  accounts for the upstream proteome resources required to sustain an increased focal metabolite concentration against metabolite dilution in nutrient-rich medium. A more general form of this relation applies when the external nutrient becomes growth-limiting ( $\rho_1 \lesssim K_1$ ) and the nutrient uptake enzyme is constrained ( $\phi_1 < \phi_1^{\max}$ ). This constraint, the biological rationale of which is discussed below, curbs all downstream metabolite concentrations,  $\rho_2, \dots, \rho_\ell$ . It can be described by an additional metabolite cost  $\beta_0(\rho_1, \phi_1^{\max})$  that becomes relevant under nutrient restriction and is negligible in the nutrient-rich regime (SI). Together, we obtain the total metabolite cost

$$\beta_i^* \equiv \beta_0^*(\rho_1, \phi_1) + \sum_{k < i} \frac{\lambda^*}{\kappa_k^*}. \quad (10)$$

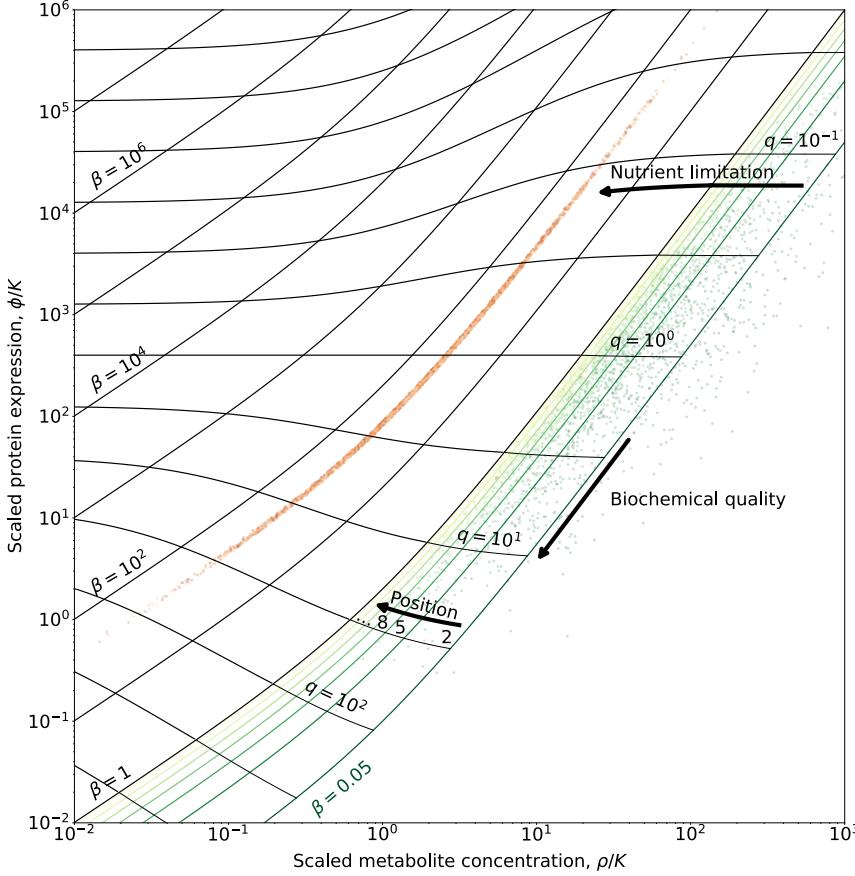
This cost summarizes the influence of the global metabolic network and the nutrient environment on the local balance of substrate and enzyme masses.

**Metabolite-enzyme relations.** Combining the optimality condition (9, 10) with the kinetic rate law (1), we obtain remarkably simple constitutive relations for cellular resource allocation,

$$\rho_i^* = \sqrt{\frac{K_i}{\kappa_i^{\max}}} \sqrt{\frac{j_i^*}{\beta_i^*}}, \quad (11)$$

$$\phi_i^* = \beta_i^* \rho_i^* \left( 1 + \frac{\rho_i^*}{K_i} \right) \quad (i = 2, \dots, \ell). \quad (12)$$

The first relation determines  $\rho_i$  as a function of the local kinetic parameters  $K_i, \kappa_i^{\max}$  and the network-dependent variables  $j_i, \beta_i$ . The second relation links enzyme and metabolite mass fractions. Fig. 2 summarizes the central results of this model; derivations and more detailed analysis will be given in the remainder of this section. In maximum-growth states ( $\rho^*, \phi^*$ ), resource allocation to individual metabolite-enzyme pairs depends on three main factors: network position, nutrient level, and the biochemistry of local metabolic steps (arrows in Fig. 2). This pattern is organized by two key parameters: the metabolite cost  $\beta_i$  and the growth response parameter  $q_i \sim (\kappa_i^{\max} K_i)^{1/2}$ , a summary variable of local reaction kinetics (grid lines in Fig. 2). The cost parameter  $\beta_i$  encodes information on network position and on the environment. Increasing cost generates reduced metabolite levels  $\rho_i$  and a weaker, inhomogeneous response of enzymes: high- $q$  proteins increase and low- $q$  proteins decrease in their expression level  $\phi_i$ . Variation of the growth response parameter  $q$  induces a crossover from a linear enzyme-metabolite relation at low saturation ( $\phi_i \sim \rho_i$  for  $\rho_i/K_i \lesssim 1$ ) to a quadratic relation at high saturation ( $\phi_i \sim \rho_i^2$  for  $\rho_i/K_i \gtrsim 1$ ), as given by equation (12). A similar scaling relation has been derived by Dourado et al. (16) from a mass minimization condition for an individual metabolite-enzyme pair. The network relation (12) differs in two aspects: the metabolite density is independently set by equation (11) and is weighed by the cost parameter  $\beta_i$ .



**Figure 2: Resource allocation to metabolites and enzymes in metabolic networks.** Points show affinity-scaled mass fractions of individual metabolite-enzyme pairs,  $(\rho_i^*/K_i, \phi_i^*/K_i)$ , in optimal balanced states of metabolic networks at high nutrient supply (green,  $\beta_0^* = 0$ ) and under nutrient limitation (orange,  $\beta_0^* = 40$ ). These levels are determined by two key parameters (solid lines): the metabolite cost  $\beta_i$  and the growth response parameter  $q_i \sim (\kappa_i^{\max} K_i)^{1/2}$ , which depends on the kinetic coefficients of the metabolite-enzyme pair (see text for precise definitions). At high nutrient supply, the cost depends on the network position  $i$ , causing higher levels of upstream metabolites (darker green shading), while protein expression depends predominantly on local biochemistry; see also Fig. 3A, B. Nutrient limitation induces a global increase of metabolic costs, generating a strong decrease of metabolite levels and a weaker, inhomogeneous response of protein expression: high- $q$  proteins are up-regulated, low- $q$  proteins are down-regulated with decreasing nutrient supply; see also Fig. 4. Model parameters: chain length  $\ell = 20$ ; kinetic parameters are log-normal random variables,  $\kappa_{\text{ave}}^{\max} = 10^{1.6 \pm 0.4} \text{ h}^{-1}$ ,  $K_{\text{ave}} = 10^{-4.2 \pm 1.1}$ , resulting in networks with dilution strength  $\epsilon = 0.06 \pm 0.04$ , as given by equation 17. Grid lines and nutrient limitation (orange points) are shown for specific realizations with  $\epsilon = 0.05$  and  $\phi_1^{\max} \ll 1$ . In these networks,  $\beta_0^* = 40$  corresponds to a growth reduction by 50%.

**Resource allocation at high nutrient supply.** We now show that the constitutive relations 12 and 11, together with the continuity relation 4 and the cost function 10, determine cellular states of maximum growth. We start with cells in a nutrient-rich environment, which grow at a rate close to their physiological bound. In this growth regime, we find a self-consistent solution

$$\rho_i^* \simeq \frac{s_i}{\sqrt{\beta_i^*}}, \quad (13)$$

$$\phi_i^* \simeq \frac{\lambda^*}{\kappa_i^{\max}} \left( 1 + 2\epsilon(1 - \sqrt{\beta_i^*}) \right) + \epsilon s_i \sqrt{\beta_i^*}, \quad (14)$$

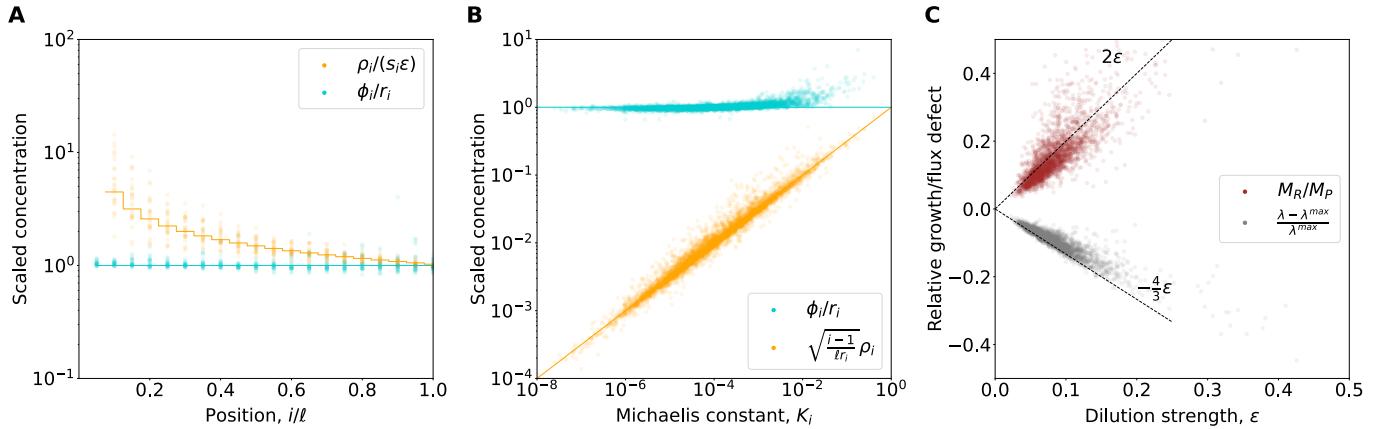
$$\beta_i^* \simeq \frac{i-1}{\ell}, \quad (15)$$

$$\lambda^* \simeq \lambda^{\max} \left( 1 - \frac{4}{3}\epsilon \right) \quad \left( \frac{\rho_1}{K_1} \gtrsim 1 \right), \quad (16)$$

where  $\lambda^{\max} = (\sum_{k>1} (1/\kappa_k^{\max}))^{-1}$  is the maximum growth rate in the absence of metabolite dilution,  $s_i = (K_i/\kappa_i^{\max})^{1/2}/\sum_{k>1} (K_k/\kappa_k^{\max})^{1/2}$  depends on the enzyme specificity  $\kappa_i^{\max}/K_i$ , and

$$\epsilon = \sum_{k=2}^{\ell} \left( \frac{K_k \lambda^{\max}}{\kappa_k^{\max}} \right)^{1/2} \quad (17)$$

is a dimensionless biochemical summary variable, which we term *dilution strength*. To derive these relations, consider first the case  $\epsilon = 0$ , which reproduces the result of standard flux balance analysis: fluxes are conserved throughout the network



**Figure 3: Metabolite effects at high nutrient supply.** (A) Specificity-scaled metabolite and enzyme catalytic rate-scaled protein levels,  $\rho_i/s_i$  and  $\phi_i/r_i$ , as functions of relative network position,  $i/\ell$  (with  $r_i = 1/\sum_k (\kappa_i^{\max}/\kappa_k^{\max})$ ). Points are individual enzyme-substrate pairs in physiologically parameterized cell models in a nutrient-rich environment. Lines are analytical predictions of equations 13–17. (B) Position- and catalytic rate-scaled metabolite and protein levels,  $\rho_i(\beta_i/r_i)^{1/2}$  and  $\phi_i/r_i$ , as functions of inverse substrate affinity,  $K_i$ . (C) Growth rate degradation,  $\lambda/\lambda^{\max}$ , and metabolome-proteome mass ratio,  $M_R/M_P$ , as functions of the dilution strength  $\epsilon$ . Points are complete, physiologically parameterized cell models in a nutrient-rich environment. Stochastic parameters as in Fig. 2.

( $j_i = \lambda$ ), all enzymes run at full saturation with rates  $\kappa_i^{\max}$  and optimal expression levels proportional to  $(1/\kappa_i^{\max})$ , and metabolites are neglected ( $\rho_i = 0$ ). The effects of metabolite dilution can be included by a power series expansion in the parameter  $\epsilon$ . Equations 13–14 reproduce the constitutive relations 11–12 to first order in  $\epsilon$ . In the following steps, we use a mean-field approximation, replacing the heterogeneous kinetic coefficients  $\kappa_i^{\max}$ ,  $K_i$  by position-independent values  $\kappa_{\text{ave}}^{\max}$ ,  $K_{\text{ave}}$ . To evaluate the metabolite cost, we approximate 10 as  $\beta_i \simeq \sum_{k>i} (\lambda^{\max}/\kappa_{\text{ave}}^{\max}) + O(\epsilon)$ , which leads to equation 15. In the same way, we approximate the continuity equation 4 to evaluate the effect of flux degradation on protein expression,  $j_i = \lambda(1 + \epsilon \sum_{k>i} \beta_i^{-1/2}/\ell)$ , which enters equation 14. Finally, the first-order correction to the growth rate, equation 16, is obtained from the normalization condition  $\sum_k \phi_k = 1$ . This relation predicts the degradation of growth by metabolite dilution in typical metabolic networks.

Physiological values of the parameter  $\epsilon$  can be inferred from empirical data. Metabolic networks have typical catalytic rates  $\kappa^{\max} \approx 10^{1.6 \pm 0.4} \text{ h}^{-1}$  and Michaelis constants  $K \approx 10^{-4.2 \pm 1.1}$  in dimensionless units of mass fractions (37). Typical pathways of protein biosynthesis from glucose via a representative amino acid have length  $\ell \approx 20$  (SI). In mean-field approximation, equation 17 then yields typical values  $\epsilon \sim 0.05$  of the dilution strength, justifying the first-order calculus in equations 13–16. We can also use these physiological network data to test the accuracy of the mean-field approximation. We construct a set of metabolic chains with kinetic parameters  $\kappa_i^{\max}$ ,  $K_i$  ( $i = 1, \dots, 20$ ) independently drawn from log-normal distributions centered at the corresponding physiological average values (SI). Consistently, these chains have growth rates of order  $1 \text{ h}^{-1}$ , and match other physiological cell properties (Fig. S2). For each chain, we evaluate maximum-growth metabolic state  $S^* = (\rho^*, \phi^*)$  by numerical solution of equation 6 and compare with equations 13–16. Remarkably, the analytical form reproduces individual metabolite and enzyme levels of heterogeneous chains, as well as chain-specific growth rates, in good approximation (Fig. 3, Fig. S2E).

Together, our model predicts a consistent pattern of metabolic resource allocation in nutrient-rich media. Metabolite levels strongly depend on network position and on enzyme specificity,  $\kappa_i^{\max}/K_i$ : metabolites in upstream or less specific reactions have systematically higher concentrations than those in downstream or more specific reactions (Fig. 3A, B). Enzymes show a weaker dependence on these parameters; their expression levels are predominantly determined by the catalytic rate,  $\phi_i \simeq \lambda/\kappa_i^{\max} + O(\epsilon)$ . The corresponding affinity-scaled levels are predominantly in the high-saturation regime,  $\rho_i/K_i, \phi_i/K_i \gtrsim 1$  (Fig. 2), consistent with empirical data (16, 38).

**Metabolome-proteome mass ratio.** Our model predicts the relative abundance of metabolites and proteins in a metabolic network as a function of its reaction kinetics. We sum up the metabolite densities  $\rho_i$ , using equation 13 in mean-field

approximation ( $\rho_i \simeq \beta_i^{-1/2}$ ), to obtain the metabolome-proteome mass ratio in typical metabolic networks at high nutrient supply,

$$\frac{M_R}{M_P} \simeq 2\epsilon; \quad (18)$$

with correction terms of order  $O(\epsilon^2, \epsilon/\ell)$  (SI). The resulting numerical value,  $M_R/M_P \sim 10^{-1}$ , is in line with empirical data [38, 39], providing a further consistency check for the model. Equations [16]–[17] measure the growth cost of metabolites at high nutrient supply; this cost is caused by dilution and increases with the network size  $\ell$ . Below, we show that physiological values of  $M_R/M_P$  are set by the global evolution of network biochemistry.

**Nutrient limitation.** Cells respond to a reduced density of external nutrients by up-regulation of the cognate uptake machinery. In our simple metabolic chain, we model this response as a nutrient-dependent expression of the first enzyme,  $\phi_1(\rho_1) = \phi_1^0 + (\phi_1^{\max} - \phi_1^0)/(1 + \rho_1/K^{\text{reg}})$ , which varies between a basal level  $\phi_1^0$  at high nutrient concentration  $\rho_1$  and a maximal level  $\phi_1^{\max}$  at low concentration (Fig. S3A). The upper bound  $\phi_1^{\max}$  can be generated by different biological mechanisms, including spatial constraints on the cell membrane, diffusive limits on nutrient uptake at low supply [40], and a maximum fold change achievable by regulation. Such limited overexpression of the uptake machinery has been observed for glucose limitation in chemostat experiments [41]. In turn, this generates a constrained optimal metabolic state,  $\mathcal{S}^*(\rho_1, \phi_1^{\max})$ . In this state, the growth rate is limited by the uptake step,  $\lambda^*(\rho_1, \phi_1^{\max}) \simeq \kappa_1^{\max} \phi_1(\rho_1)/(1 + K_1/\rho_1)$ . Nutrient limitation significantly affects growth in the regime  $\rho_1/K_1 \lesssim 1$ , corresponding to  $\lambda/\lambda^{\max} \lesssim 1 - \epsilon - \phi_1^{\max}$  (SI, Fig. S2F). In the following, we use  $\lambda$  as an independent control parameter. Under strong nutrient limitation, the constitutive relations [11]–[12] take the form,

$$\rho_i^*(\lambda) \simeq \frac{\lambda}{Z} \sqrt{\frac{K_i}{\kappa_i^{\max}}}, \quad (19)$$

$$\phi_i^*(\lambda) \simeq Z \sqrt{\frac{K_i}{\kappa_i^{\max}}} + \frac{\lambda}{\kappa_i^{\max}} \quad \left( \frac{\rho_1}{K_1} \lesssim 1 \right) \quad (20)$$

for  $i = 2, \dots, \ell$ , where the normalization factor

$$Z(\lambda) \equiv \frac{1 - \phi_1^*(\lambda) - \sum_{k=2}^{\ell} (\lambda/\kappa_k^{\max})}{\sum_{k=2}^{\ell} \sqrt{K_k/\kappa_k^{\max}}} \quad (21)$$

is given by the condition  $\sum_k \phi_k = 1$ . Here we have omitted the flux degradation by metabolite dilution ( $j_i \simeq \lambda$ ), which becomes small at low growth rates. As shown by comparison of equations [19]–[20] and [10]–[12], nutrient limitation generates an overall metabolite cost  $\beta_0 = Z^2/\lambda$ . This cost is numerically larger than the network-dependent terms  $\lambda^*/\kappa_k^*$ , which are therefore neglected ( $\beta_i \simeq \beta_0$ ). For simplicity, we also assume that the uptake proteins remain a small fraction of the total proteome ( $\phi_1^{\max} \ll 1$ ), which simplifies equation [21] to  $Z(\lambda) \simeq (\lambda^{\max} - \lambda)/(\sqrt{\lambda^{\max}}\epsilon)$ .

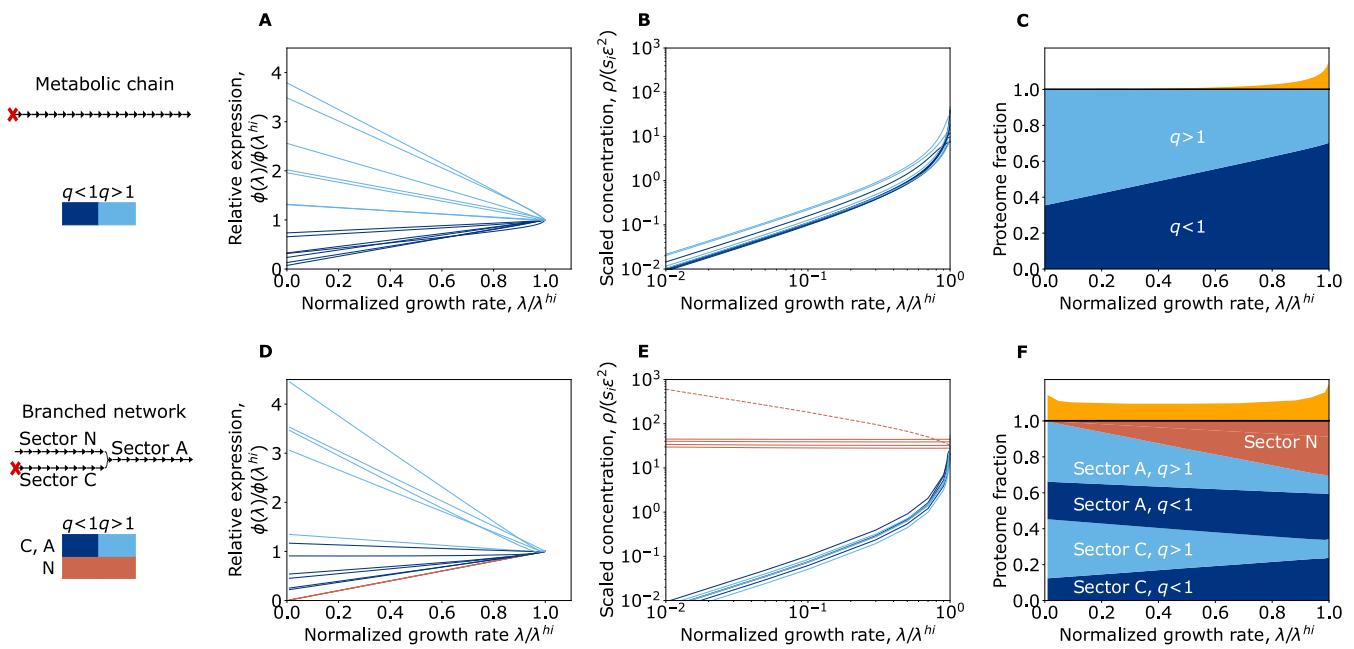
The solution [19]–[21] shows two salient features of metabolic response to nutrient limitation. Given that  $Z$  is a linear function of  $\lambda$ , metabolite levels increase in a strongly nonlinear way with growth rate,

$$\rho_i^*(\lambda) \simeq \frac{\epsilon^2 \lambda}{(\lambda^{\max} - \lambda)} s_i, \quad (22)$$

with  $s_i = \sqrt{K_i/\kappa_i^{\max}} / \sum_k \sqrt{K_k/\kappa_k^{\max}}$ . This equation is correct up to position-dependent pre-factors of order 1 relevant at high growth, and differences in  $\phi_1$  (Fig. 4B). The global metabolome-proteome mass ratio,  $M_R/M_P$ , depends on growth in the same way. In contrast, protein expression levels show a linear response pattern,

$$\phi_i^*(\lambda) \simeq \left( 1 - \frac{\lambda}{\lambda^{\max}} \right) s_i + \frac{\lambda}{\lambda^{\max}} r_i, \quad (23)$$

with  $r_i = 1 / \sum_k (\kappa_i^{\max}/\kappa_k^{\max})$ , which is found by re-arranging terms in [20]–[21]. Hence, the expression of a given enzyme is inversely proportional to its square root specificity at low growth and to its catalytic rate  $\kappa_i^{\max}$  at high growth. Defining

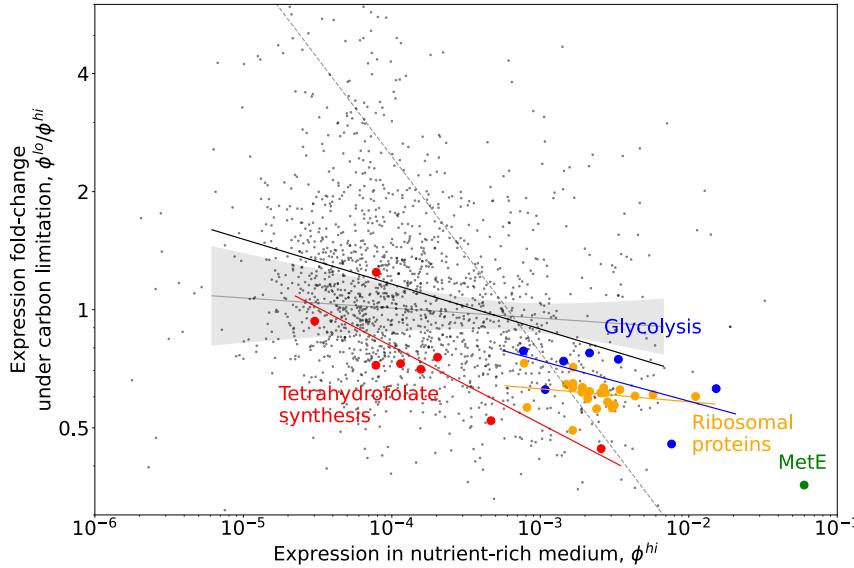


**Figure 4: Metabolic response to nutrient limitation.** (A - C) Metabolic chain. (A) Reduction of protein expression levels,  $\phi_i^*/\phi_i(\lambda^{\text{hi}})$ , as functions of the growth rate, for two enzyme classes: rate-optimized ( $q > 1$ ) and specificity-optimized ( $q < 1$ ). Here  $\lambda^{\text{hi}}$  is the maximum physiological growth rate at high nutrient supply, as given by equation 16. Lines represent randomly drawn enzymes from the ensemble of physiologically parametrized networks (see text; nutrient uptake enzymes were excluded from the analysis). (B) Reduction of metabolite levels,  $\rho_i^*$ , as functions of the growth rate. (C) Ensemble averages of global response of proteome and metabolome. The metabolite mass fraction is shown in units of the proteome mass in orange. (D - F) Three-sector system. (D) Growth-dependent protein levels,  $\phi_i^*/\phi_i(\lambda^{\text{hi}})$ , by growth response class and by sector. (E) Growth-dependent metabolite levels,  $\rho_i^*/(s_i \epsilon^2)$ , by sector. (F) Global response of proteome and metabolome. Chain length:  $\ell = 20$ ,  $\ell_A = \ell_C = \ell_N = 10$ . Model parameters as in Fig. 2 for chains with  $\phi_1^{\text{max}} \ll 1$ .

the ratio  $q_i = s_i/r_i \sim \sqrt{K_i \kappa_i^{\text{max}}}$ , metabolic growth response distinguishes two classes of enzymes: rate-optimized enzymes ( $q_i > 1$ ) are up-regulated, specificity-optimized enzymes ( $q_i < 1$ ) are down-regulated upon nutrient limitation (Fig. 2, Fig. 4A). In networks with a sizeable proteome fraction  $\phi_1^{\text{max}}$ , there is a small but systematic shift of the growth response pattern (SI).

The corresponding global resource allocation pattern is shown in Fig. 4C: a depletion of metabolites comes together with an expression shift from specificity-optimized to rate-optimized enzymes. These analytical predictions capture specific networks with physiological reaction parameters, not just ensemble averages (Fig. S4).

**Growth-dependent patterns in genome-wide protein expression data.** A comprehensive test of our model requires system-wide kinetic data,  $(K_i, \kappa_i^{\text{max}})$ , of metabolic reactions, which are currently not available. However, statistical evidence can be obtained from whole-genome proteomics data. Here we use a recent dataset of protein expression for 1743 genes in *E. coli* growing under different levels of glucose uptake limitation (24). For each gene, we plot the expression level at high nutrient supply,  $\phi_i^{\text{hi}}$ , against the relative change upon nutrient limitation,  $\xi_i = \phi_i^{\text{lo}}/\phi_i^{\text{hi}}$  (Fig. 5). We find a negative correlation,  $\log \xi_i = -0.11 \log \phi_i^{\text{hi}} + \text{const.}$ , that is statistically significant ( $p < 0.03$  under a null distribution with scrambled growth rates). This type of correlation can be explained by our model: high values of  $\phi_i^{\text{hi}}$  are correlated with low values of  $\kappa_i^{\text{max}}$ , signalling predominantly specificity-optimized proteins ( $q_i < 1$ ) that are down-regulated upon nutrient limitation (assuming no overriding correlation between  $\kappa^{\text{max}}$  and  $K$ ). The signal identified in the proteomics data is likely to be diluted by several factors of network heterogeneity not included in the minimal model. Typical growth responses under nutrient limitation affect predominantly specific functional sectors and exclude others, such as metabolically inactive proteins (24, 42) (see also below). Specific functional groups involved in a common reaction, such as the ribosomal genes, are consistent with the model on average but do not show a within-group anti-correlation of  $\xi$  and  $\phi_i^{\text{hi}}$  (Fig. 5). Other groups, including the pathways of tetrahydrofolate synthesis or glycolysis, involve reaction chains and show the predicted anti-correlation with



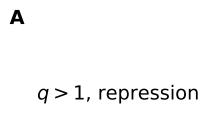
**Figure 5: Protein expression changes under nutrient limitation.** Expression fold change upon glucose uptake limitation,  $\phi_i^{\text{lo}}/\phi_i^{\text{hi}}$  plotted against expression in a nutrient-rich environment,  $\phi_i^{\text{hi}}$ , for 1743 proteins of *E. coli*; data from (24). These data show a negative correlation (solid black line) that is statistically significant ( $p < 0.03$ ; the shaded region shows the range of slopes expected from data with randomized growth rates). The biochemical network model explains this correlation by the growth-dependent expression shift from specificity-optimized to rate-optimized enzymes (see text). Individual single-reaction pathways are consistent with the model in their average expression (orange: ribosomal proteins), multi-step pathways (blue: glycolysis, red: tetrahydrofolate synthesis (43)) show intra-group anti-correlations with pathway-dependent offsets.

pathway-specific offsets.

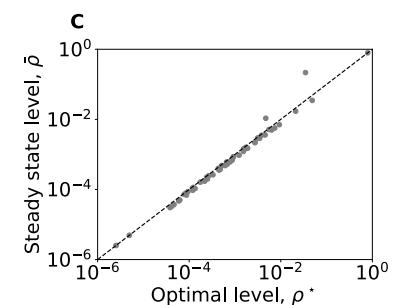
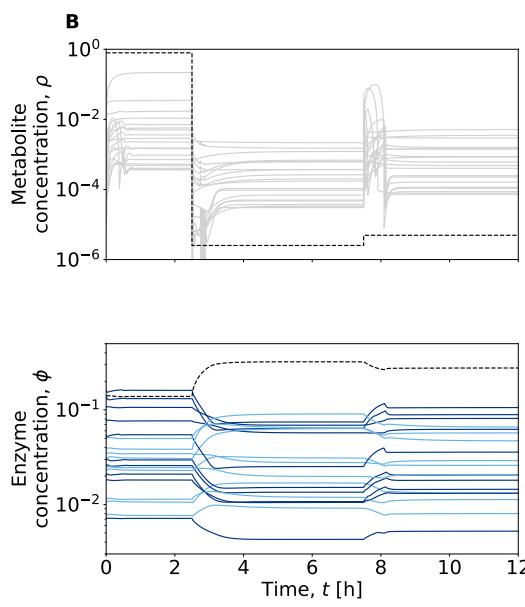
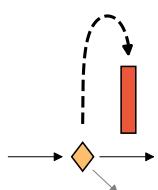
**Growth response of functional sectors.** How does the pattern of growth response generalize to more complex metabolic networks? Here we consider a network with two uptake pathways, C and N, processing distinct nutrients both necessary for growth; their outputs feed into a single downstream pathway, A (Fig. 4D-F). This is one of the simplest networks with differentiation into functional sectors. In previous work, such sectors have been identified as coherent units of proteome resource allocation (44, 45). We study the three-sector network under limitation of nutrient C, assuming again an expression constraint  $\phi_{1,C}^{\text{max}}$  of the cognate uptake machinery. To obtain the growth-optimal resource allocation pattern shown in Fig. 4D-F, we employ Eq. 6 in each pathway, and numerically solve for optimal resource allocation at their junction. Sectors C and A show a common growth response similar to the single-chain system (22, 23): metabolites are depleted, proteins experience an expression shift from specificity-optimized to rate-optimized enzymes. In sector N, metabolites are not depleted and enzymes are uniformly down-regulated, inducing a net resource shift from N to C and A (Fig. 4F).

Together, individual metabolite and enzyme levels depend on their functional sector and on the biochemistry of the local reaction step. Thus, our model produces growth response patterns with coherent shifts between sectors, as well as heterogeneous changes within sectors. This heterogeneity is consistent with available experimental data: multiple groups of proteins characterized by distinct regulatory responses to external perturbations are enriched in the same functional Gene Ontology-terms (24). Local biochemistry can also contribute to coherent shifts if kinetic parameters are broadly correlated with sectors, as suggested by recent empirical observations (46). A specific case in point are ribosomes, which have a small catalytic rate per unit of mass,  $\kappa^{\text{max}} \approx 10^{-1} \kappa_{\text{ave}}^{\text{max}}$  (11) and a small binding constant  $K < 10^{-2} K_{\text{ave}}$  (47, 48), generating a sub-average growth response parameter  $q < 1$  (SI). Hence, by local biochemistry, ribosomes are highly expressed in nutrient-rich media and down-regulated with decreasing growth; a similar example may be methionine synthase (MetE) (Fig. 5). Such low- $q$  anabolic enzymes with large proteome fractions may contribute to the generic down-regulation of the A sector experimentally observed under multiple growth-limiting conditions (24, 42).

**Metabolic response by regulation.** Can a metabolic system come close to an optimal growth state by local regulation of its enzymes (49)? Specifically, we consider the response of a metabolic chain to nutrient limitation. This process requires growth-dependent shifts in metabolite and enzyme levels,  $\rho_i^*(\lambda)$  and  $\phi_i^*(\lambda)$ , as given by equations (22, 23). Protein concentrations change by synthesis with rate  $b_i$  and by dilution,  $\dot{\phi}_i = b_i - \lambda\phi_i$ . In the simplest case of transcriptional regulation, the



$q < 1$ , activation



**Figure 6: Metabolite-mediated regulation.** (A) Logic of regulation: metabolites act as transcription factors for their cognate enzymes, inducing repression for rate-optimized enzymes ( $q_i > 1$ ) and activation for specificity-optimized enzymes ( $q_i < 1$ ). (B) Time-dependent metabolite levels,  $\rho_i(t)$  (gray) nutrient concentration (dashed), and regulated protein concentrations  $\phi_i(t)$  (dark blue: specificity-optimized enzymes,  $q < 1$ ; light blue: rate-optimized enzymes,  $q > 1$ , dashed: uptake enzyme) under time-dependent nutrient limitation,  $\rho_1(t)$  (top, dashed). (C) Regulated late-time stationary levels,  $\bar{\rho}_i$  and  $\bar{\phi}_i$ , plotted against growth-optimal levels,  $\rho_i^*$  and  $\phi_i^*$ .

stationary level

$$\bar{\phi}_i = \frac{b_i}{\lambda} = \phi_i^0 + \frac{a_i}{1 + K_i^{\text{reg}}/\rho}, \quad (24)$$

depends on the equilibrium binding of a transcription factor present at density  $\rho$  (SI and Fig. 6A). This regulation function has three parameters: the basal level  $\phi_i^0$ , the regulatory amplitude  $a_i$ , and the regulatory binding constant  $K_i^{\text{reg}}$ . Here we assume a simple logic: at each metabolic step, the substrate regulates the transcription of its cognate enzyme (Fig. 6A). In this network, growth optimization by regulation depends on a remarkable property: with regulatory parameters depending only on local biochemistry,  $\phi_i^0 = s_i$ ,  $a_i = r_i - s_i = r_i(1 - q_i)$ , and  $K_i^{\text{reg}} = s_i \epsilon^2$ , the stationary expression level 24 follows the optimal growth response,  $\bar{\phi}_i(\rho_i^*(\lambda)) = \phi_i^*(\lambda)$ , up to terms of order  $\epsilon$ . This identity follows by inspection of equations 22 – 24; see SI for details. As required for optimal growth response, rate-optimized enzymes ( $q_i > 1$ ) are repressed, specificity-optimized enzymes ( $q_i < 1$ ) are activated by their cognate metabolite. The adaptive dynamics of the metabolic state,  $S(t) = (\rho, \phi)(t)$ , consists of regulated changes of protein levels and correlated changes of metabolite levels given by mass conservation, equation 3. Here we use metabolite-mediated regulation specified by the set of parameters  $\mathcal{R}_m = (\phi^0, a, K^{\text{reg}})$  for the entire network. Fig. 6B shows these dynamics under time-dependent nutrient limitation,  $\rho_1(t)$ , switching between subsequent stationary levels. Rate-optimized and specificity-optimized enzymes change in opposite directions and settle to a stable stationary state  $\bar{S}(\rho_1)$  with growth rate  $\bar{\lambda}(\rho_1)$  that matches the growth-optimal state in very good approximation,  $\bar{\lambda} \gtrsim 0.98\lambda^*$ ,  $\bar{\phi}_i = \phi_i^*(1 \pm 0.05)$ , and  $\bar{\rho}_i = \rho_i^*(1 \pm 0.5)$  (Fig. 6C).

More broadly, our analysis suggests that metabolite-induced regulation of enzymes can generate partial optimization of metabolic states under diverse conditions. In more complex networks, local regulation of each enzyme is not a plausible scenario. However, few, broad-acting transcription factors can generate coarse-grained optimization of the average expression for co-regulated target genes in pathways or functional sectors.

**Network evolution.** On large time scales, metabolic networks change by protein evolution. Molecular changes affect local reaction parameters ( $\kappa_i^{\text{max}}, K_i$ ), inducing changes in pathway usage and, ultimately, in network topology. Consider an

enzyme mutation changing the affinity to its cognate metabolite,  $K_i \rightarrow K_i(1 + \delta K_i)$ . Its effect on growth is partially buffered by the physiological response of the network, which regulates the protein expression  $\phi_i$  and changes the metabolite density  $\rho_i$ . Our model predicts the strength of selection on network-buffered binding affinity changes,  $\sigma_i = -(K_i/\bar{\lambda})(\partial\bar{\lambda}/\partial K_i)$ , where we assume a regulation function  $\mathcal{R}$  reproducing near-optimal stationary states,  $(\bar{\rho}, \bar{\phi}) \approx (\rho^*, \phi^*)$ , in the wild-type system  $(\kappa_i^{\max}, K)$ . We obtain buffered selection coefficients,

$$\sigma_i \sim c_i \rho_i^*, \quad (25)$$

where  $c_i$  are growth-dependent factors of order 1 (SI). Two important consequences follow. First, given  $\rho_i \sim (K_i/\kappa_i^{\max})^{1/2}$ , selection acts on enzyme specificities. In other words, the selection pressure to increase the affinity  $K_i^{-1}$  is higher for enzyme-metabolite pairs with low catalytic rate  $\kappa_i^{\max}$ , predicting a negative correlation between affinity and rate in evolved networks. Second, we can estimate the order of magnitude of selection by evaluating  $\rho_i$ , as given by equation [22] at intermediate growth rates. In the mean-field approximation introduced above, we find  $\sigma_i \sim \epsilon^2/\ell \sim 10^{-4}$  at physiological network parameters  $K_{\text{ave}} \sim 10^{-4}$  and  $\ell \sim 20$ . Importantly, the buffered  $\sigma_i$  is at least two orders of magnitude lower than the growth effect of an affinity change for a single Michaelis-Menten pair,  $\sigma_0 = (K/J)(\partial J/\partial K) = (1 + \rho/K)^{-1}$  at constant  $\rho$  [1].

The evolution of enzyme-metabolite binding is a mutation-selection-drift process [50] on a biophysical fitness landscape [51, 52], here of the form  $f(\Delta G_i) \sim -C_i \exp[\Delta G_i/2]$ , where  $\Delta G_i = \log(K_i/K_0)$  is the binding free energy and  $C_i$  is a constant. In such landscapes, mutation-selection equilibrium emerges at typical selection coefficients  $\sigma_i \sim 1/N_e$ , where  $N_e$  is the effective population size [58, 60]. Hence, the strength of selection on  $K_i$  is consistent with evolutionary equilibrium at  $N_e \sim 10^4$ , in tune with the effective population sizes obtained for microbial systems evolving under clonal interference of protein phenotypes (in such systems,  $1/2N_e$  is the average number of generations between subsequent selective sweeps) [58]. In turn, this implies that physiological values of the metabolome-proteome mass ratio,  $M_R/M_P$ , are set by an evolutionary equilibrium where the selective cost of metabolite dilution balances with mutations degrading enzyme-metabolite binding.

## Discussion

Here we have developed a model for resource allocation in metabolic networks that integrates the kinetics of local reaction steps, the global structure of the network, and external nutrient constraints to predict enzyme and metabolite levels in growth-optimal network states (Fig. 2). Protein expression levels are set predominantly by local biochemistry: levels at high growth rates are set by catalytic rates,  $\phi_i \sim 1/\kappa_i^{\max}$ ; the response to growth limitation is linear with a slope depending on the response parameter  $q_i = (K_i\kappa_i^{\max})^{1/2}$ . This pattern is in broad agreement with empirical observations in metabolic networks. First, the model explains the heterogeneous growth response of enzymes with similar functions [24]. Second, it predicts finite protein expression levels in the limit of low growth rates, specifically for the ribosome; different mechanisms for this residual expression have been discussed in the recent literature [34, 42, 61, 62]. Third, the linear growth response of protein levels derived here reproduces the linear pattern found in previous work at the level of functional protein sectors [15, 44, 62]. However, the coherent growth response of broad sectors (related, e.g., to limitation of a particular nutrient influx), is not a generic outcome for growth-optimized networks. Our analysis suggests two possible mechanisms generating sector responses: broad correlations of biochemical parameters and network position [46] or deviations from fine-grained optimality by pleiotropic, rather than gene-specific, regulation.

Metabolites are predicted to show a nonlinear, position-dependent depletion of steady-state levels in response to growth limitation. Moreover, we show that metabolite levels depend on a local cost parameter  $\beta_i$  that summarizes constraints affecting the upstream metabolic chain (here, nutrient limitation). Such growth laws for metabolites have not yet been tested systematically by experiment; available data suggest that some but not all metabolites are depleted by growth limitation [25, 63, 64]. Deviations from model predictions may signal toxicity or enhanced efflux of specific metabolites. They may also reflect an increase in  $\beta_i$  caused by bottlenecks in upstream reactions or nutrient conditions, in tune with an observed decrease in metabolite concentrations on poor carbon sources [16]. More broadly, our model can serve as a starting point to describe the quantitative organization of the metabolome, which has remained elusive so far [39].

A mechanistic model of metabolic networks including enzymes and metabolites is also a prerequisite for predicting network dynamics by regulation and, on longer time scales, by evolution of reaction rates and binding constants. Here we have shown that a simple regulatory scheme based on feed-forward regulation – each metabolite regulates the enzyme of the next reaction step – can closely approximate optimal resource allocation under nutrient limitation (Fig. 6). Thus, our model highlights the

role of metabolites as regulatory signals and paves the way to design regulated synthetic networks that maintain optimality under variable input (49, 65).

Evolution acts on network states buffered by regulated physiological response. The resulting selection coefficients for local changes in network biochemistry, which become computable from our model, are lower than in the absence of network buffering. Specifically, we have shown that metabolite dilution induces system-wide selection pressure on enzyme-metabolite binding. Individual enzymes evolve on approximately independent fitness landscapes depending on their target affinity. Evolutionary equilibria on these landscapes can explain the order of magnitude of physiological enzyme-target affinities, which are well below the biophysical bound set by substrate diffusion (37), as well as the global mass ratio of metabolome and proteome. In contrast, interventions like antibiotics can impair specific enzymes in a stronger way, often beyond the limits of regulatory buffering (66), generating stronger selection pressure on evolutionary network changes (18, 19). Understanding this interplay of regulation and evolution is crucial for the design of efficient antibiotic interventions into microbial metabolism.

## Acknowledgements

We thank Tobias Bollenbach, Laura Collesano, Martin Lercher, and all members of the Bollenbach and Lässig groups for helpful discussions. This work was supported by Deutsche Forschungsgemeinschaft Grant SFB 1310 (to ML). FP acknowledges funding by Human Technopole.

## Supporting Information

### Maximization of balanced growth rate by resource allocation

**Scaling of variables.** Bacteria reallocate metabolic resources under environmental challenges. We explore growth-optimal resource allocation in a structured network of biochemically heterogeneous enzymes (Fig. 1) in variable environments. In contrast to previous proteome sector models (44), our model concurrently accounts for both metabolites and enzymes. As a result, our model introduces second-order rate constants that describe reactions with multiple reaction partners, such as enzyme-substrate interactions. The rates and concentrations in our model exhibit a different scaling with the reaction volume when enzyme and substrate amounts are held constant. To include this scale in our model, we introduce the parameter

$$\nu \equiv \frac{V}{M_P}. \quad (26)$$

Here,  $V$  is the total intracellular volume, and  $M_P$  is the total mass of protein. Empirical observations suggest that this fraction remains approximately constant throughout the cell cycle and across different environments (15). The parameter  $\nu$  enables converting mass concentrations  $m/V$  into dimensionless mass fractions  $m/M_P$ . A metabolite mass fraction  $\rho$  is defined as  $\rho = M_S \nu c$ , with the classical concentration  $c$ , in units of  $\text{mol L}^{-1}$ , and the molar mass of the metabolite  $M_S$ . For simplicity, we apply identical unit conversions to the extracellular nutrient concentration  $\rho_1$ . Similarly, a protein mass fraction  $\phi$  corresponds to  $\phi = M_E \nu c_E$  with the classical concentration  $c_E$  (in units of  $\text{mol L}^{-1}$ ) and the molar mass of the protein  $M_E$ . In this unit system protein mass fractions in a cell of  $\ell$  proteins must sum up to 1,

$$\sum_{i=1}^{\ell} \phi_i = 1. \quad (27)$$

Flux units are accordingly chosen to directly express the time derivatives of mass fractions as flux summations,  $d\rho/dt = j_{\text{production}} - j_{\text{consumption}} - j_{\text{dilution}}$ . The conversion also extends to kinetic rate constants. For instance, a Michaelis-Menten reaction depicts an enzyme with a mass fraction  $\phi$  consuming a metabolite  $\rho$  at a rate  $j = \kappa^{\max} \phi / (1 + K/\rho)$ . Here  $K$  is a dimensionless Michaelis constant, linked to the classical Michaelis constant  $\tilde{K}_M$  (in units of  $\text{mol L}^{-1}$ ) via  $K = M_S \nu \tilde{K}_M$ .  $\kappa^{\max}$  is the enzyme's turnover number in this unit system and relates to the classical turnover number  $k_{\text{cat}}$  via  $\kappa^{\max} = (M_S/M_E) \cdot k_{\text{cat}}$ . This unit conversion allows us to write a compact and (apart from time) dimensionless derivation in the following. In this unit system, the growth rate is compactly represented as the mass-normalized flux through the protein biosynthesis reaction, which we take as the last reaction of the metabolic cascade,  $j_\ell$ ,

$$\lambda \equiv \frac{1}{M_P} \frac{dM_P}{dt} = j_\ell. \quad (28)$$

Furthermore, we define the following extensive variables to simplify notation throughout:  $R_i \equiv \rho_i M_P$ ,  $M_i \equiv \phi_i M_P$  and  $J_i \equiv j_i M_P$ .

**Independent variables: metabolite fluxes and nutrient concentration.** We restrict our analysis to balanced growth states. This is exponentially growing, metabolic steady-states that obey metabolite mass conservation in each metabolic reaction. In balanced exponential growth (averaging over cell-cycle dependent and stochastic variation), no intracellular metabolite concentration may change over time (Eq. 3 must be zero for  $\rho_{k \geq 2}$ ). Furthermore, all  $\ell$  reactions must fulfill the kinetic rate law 1. As a result, only a specific subset of all possible metabolic states, denoted by  $\mathcal{S}$ , is viable during balanced growth. We demonstrate that, given a chain of enzymes with kinetic parameters  $\kappa^{\max}$  and  $K$ , any metabolic state during balanced growth is fully determined by its metabolite fluxes  $J$ , and a nutrient environment  $\rho_1$ : inserting Eq. 2 into the balanced growth condition (Eq. 3 being zero) and solving for  $\rho_i$  yields all but the first metabolite concentration as a function of the fluxes (and vice versa 4),

$$\rho_i(J) = \frac{J_{i-1} - J_i}{J_\ell} \quad (i = 2, \dots, \ell). \quad (29)$$

The environmental nutrient concentration  $\rho_1$  is not constrained by balanced growth but by the experimental setting.

We next compute each individual protein mass according to the kinetic rate laws,

$$M_i(J, \rho_1, K, \kappa^{\max}) = \frac{J_i}{\kappa_i^{\max}} \left( 1 + \frac{K_i}{\rho_i(J, \rho_1)} \right). \quad (30)$$

By summation, the total protein mass can easily be computed, and subsequently  $R$  from  $\rho$ ,  $\phi_i$  from  $M_i$ ,  $j$  from  $J$ , as well as the growth rate of the metabolic state in balanced growth. Taken together, growth rate can be expressed as a function of all metabolic fluxes, the nutrient environment, and the kinetic constants.

$$\lambda(J, \rho_1, K, \kappa^{\max}) = \frac{J_\ell}{\sum_{i=1}^\ell M_i(J, \rho_1, K, \kappa^{\max})}. \quad (31)$$

**Optimality of metabolic states.** Akin to flux balance analysis (14), we employ constraints to identify physiological solutions. A key constraint is the constant intracellular protein density  $\nu^{-1}$ , which is well established empirically (1, 15, 26–28). Due to this constraint, we can identify a metabolic state  $\mathcal{S}^*$  that optimizes growth rate  $\lambda$ : if an enzyme's concentration is too low, it limits growth through a metabolic bottleneck. If its concentration is too high, it demands a reduction of other enzyme's concentrations, which also impairs growth.

We focus only on the regulatory problem of finding optimal metabolic states  $\mathcal{S}^*$  for a given set of kinetic constants ( $\kappa^{\max}, K$ ). We assume that kinetic constants evolve on a slower timescale, in contrast to simultaneous evolutionary optimization of kinetic constants and expression levels. To shorten the notation, we drop the kinetic constants from all arguments below. Optimal balance growth states are defined as

$$\mathcal{S}^* \equiv \arg \max(\lambda(\mathcal{S})) \Big|_{\rho_1, \frac{d\rho}{dt}=0}. \quad (32)$$

As shown in the preceding paragraph, metabolic states that fulfill balanced growth and the kinetic rate laws are fully determined by their fluxes  $J$  and a nutrient environment  $\rho_1$  (31). We thus rewrite the previous equation,

$$J^*(\rho_1) \equiv \arg \max(\lambda(J, \rho_1)) \Big|_{\rho_1}. \quad (33)$$

By construction, this optimal state can grow exponentially with time-independent metabolite concentrations. For any system size  $M_P$ , at least one optimal state exists. We here set the system size implicitly by choosing a certain  $J_\ell$ . Maximal growth rate is achieved if the protein mass necessary to produce this biomass production flux  $J_\ell$  is minimal; see Eq. (2). Thus, we can simplify the optimality condition further and obtain Eq. (5) - optimality is achieved only if

$$\frac{\partial \sum_{k=1}^\ell M_k(J, \rho_1)}{\partial J_i} \Big|_{J^*, \rho_1} = 0 \quad (i = 1, \dots, \ell - 1). \quad (34)$$

This principle resembles Enzyme Cost Minimization (36), which has been used to numerically find growth rate maximizing resource allocation in Flux-Balance-Analysis when considering reversible kinetic rate laws but no dilution of metabolites. Here we instead focus on growth maximization in a model that explicitly takes metabolite dilution into account, but no reversibility. We proceed analytically. A restricted analysis of this kind was previously pursued by Ehrenberg and Kurland (30). Beyond their analysis, we quantify the relative influences of enzyme biochemistry and network properties on resource allocation to individual enzymes and metabolites and analyze how these influences change under environmental challenges.

## Iterative construction of the optimal metabolic states in longer chains

**Optimal resource allocation in coarse-grained metabolic models.** A cell model consisting of just two proteome sectors ( $\ell = 2$ ) - with catabolic enzymes  $M_1$  and anabolic enzymes  $M_2$  - has been employed successfully to infer proteome resource reallocation in response to changes in the nutrient environment  $\rho_1$  and the catalytic rate of the anabolic sector  $\kappa_2^{\max}$  (15). For such a short chain, the optimality condition (33) has also been solved analytically, including both the proteome and the metabolome (33).

**Producing the first substrate of the chain from a new environmental precursor.** Longer chains feature many nonlinearities and feedback loops, and can not easily be solved analytically anymore. To find the optimal metabolic states  $\mathcal{S}^*(J_\ell, \rho_1)$  for longer chains, we identify an iterative scheme for constructing optimal metabolic states of a chain with  $\ell + 1$  enzymes from the optimal metabolic states of a chain with  $\ell$  enzymes. Suppose the optimal metabolic state  $\mathcal{S}_\ell^*(J_\ell, \rho_1)$  is known for a given chain defined by its  $2\ell$  kinetic constants in an environment  $\rho_1$ . We extend the given chain by a single enzyme with its kinetic constants, a protein mass  $M'_0$ , catalyzed flux  $J'_0$  and a new environmental precursor concentration  $\rho'_0$  instead of  $\rho_1$ . The new enzyme produces the first substrate of the shorter chain (Fig. S1). We indicate variables of the extended chain by a dash '·', and drop the superscript '\*' to simplify the notation in this section. Below, all state variables denote their growth-optimal values.

**Recursion relations for state variables.** An extension that adds one additional flux  $J'_0$  but conserves all fluxes of the shorter chain  $J'_{i \geq 1} = J_i$ , also conserves all downstream concentrations  $\rho'_{i \geq 2} = \rho_{i \geq 2}$  and thus all downstream protein masses  $M'_{i \geq 2} = M_{i \geq 2}$ , according to Eqs. 29 and 30. An extension of the chain by a single enzyme with arbitrary flux therefore only changes  $\rho_1$  to a new  $\rho'_1$  for which mass conservation now applies, correspondingly  $M_1$  to a new  $M'_1$ , and adds  $\rho'_0$ ,  $M'_0$  and  $J'_0$ .

These five new or modified state variables are linked by five equations: two kinetic rate laws (Eq. 1 for  $i = 0$  and  $i = 1$ ), a mass conservation constraint (Eq. 29 for  $i = 1$ ), and two conditions on optimality (Eq. 34 for  $k = 0$  and  $k = 1$ ). These five equations can be solved for the five state variables. The solution is jointly given by

$$\begin{aligned} \rho'_0 &= \frac{\frac{\kappa_1^{\max}}{\kappa_0^{\max}} K_0 \rho_1 \left( 2 \frac{J_1}{J_\ell} \left( K_1 - \frac{\kappa_1^{\max}}{\kappa_0^{\max}} \rho_1 \right) + K_1 \left( \rho_1 + \sqrt{4 \rho_1 \frac{J_1}{J_\ell} + \rho_1^2} \right) \right)}{2 \frac{J_1}{J_\ell} \left( K_1 - \frac{\kappa_1^{\max}}{\kappa_0^{\max}} \rho_1 \right)^2 - 2 \frac{\kappa_1^{\max}}{\kappa_0^{\max}} K_1 \rho_1^2}, \\ J'_0 &= J_1 \left( 1 + \sqrt{\frac{J_\ell}{J_1} \frac{\kappa_0^{\max}}{\kappa_1^{\max}} \frac{K_1}{(1 + \frac{K_0}{\rho'_0})}} \right), \\ \rho'_1 &= \frac{J'_0 - J_1}{J_\ell}, \\ M'_0 &= \frac{J'_0}{\kappa_0^{\max}} \left( 1 + \frac{K_0}{\rho'_0} \right), \\ M'_1 &= \frac{J_1}{\kappa_1^{\max}} \left( 1 + \frac{K_1}{\rho'_1} \right). \end{aligned} \tag{35}$$

A single step of this iteration is illustrated in Fig. S1. All other conditions on optimality (Eq. 34 for  $k \geq 2$ ) are fulfilled by the extended chain if they were fulfilled by the shorter chain because the protein masses  $M'_0$  and  $M'_1$  do not depend on downstream fluxes  $J'_{k \geq 2}$  in a system of irreversible enzymes ( $\partial M'_0 / \partial J'_{k \geq 2} = 0$ ,  $\partial M'_1 / \partial J'_{k \geq 2} = 0$ ), while the required protein mass downstream  $\sum_{i=2}^\ell M_i$  only depends on downstream fluxes in a manner independent of the extension ( $\partial \sum_{i=2}^\ell M_i / \partial J'_0 = 0$ ). Precisely this property of the model cell allows for a local construction of a new chain start that again yields an optimal chain, rather than requiring a full adjustment of the whole chain.

**Direct iteration of fluxes.** Computationally, it is practical to calculate the concentration of an external nutrient just once. Instead of iteratively determining the five modified state variables in each step, fluxes alone can be iterated according to the optimality criterion of Eq. 34. To derive an explicit formula, we consider a variation of a single flux  $J_i + \delta J_i$  (with  $i = 2, \dots, \ell - 1$ ), while keeping all other fluxes constant, as required by the optimality criterion. Among all mass conservation criteria 29, only those for  $\rho_i$  and  $\rho_{i+1}$  depend on  $J_i$ . We find  $\rho_i + \delta \rho_i = (J_{i-1} - (J_i + \delta J_i)) / J_\ell$  and  $\rho_{i+1} + \delta \rho_{i+1} = (J_i + \delta J_i - J_{i+1}) / J_\ell$ . Together, these changes in metabolite concentrations and the change in flux affect enzyme masses according to the Michaelis-Menten relation 1. Inserting the changes derived above yields mass changes under variation of a single flux around the optimal state,

$$\delta M_i = \frac{\delta J_i}{\kappa_i^{\max}} \left( 1 + \frac{K_i}{\rho_i} \right) + \frac{J_i}{\kappa_i^{\max}} \frac{K_i}{\rho_i^2} \frac{\delta J_i}{J_\ell}, \quad \text{and} \quad \delta M_{i+1} = -\frac{J_i}{\kappa_i^{\max}} \frac{K_i}{\rho_i^2} \frac{\delta J_i}{J_\ell}. \tag{36}$$

To summarize, for  $i \geq 2$ , only  $\rho_i$  and  $\rho_{i+1}$  (see Eq. 29) depend on  $J_i$ . Only  $M_i$  and  $M_{i+1}$  depend on either  $J_i$ ,  $\rho_i$  or  $\rho_{i+1}$ . The respective changes are illustrated in Fig. S1C. The optimality criterion 34 can now be drastically simplified to  $(\partial(M_i(J, \rho_1) + M_{i+1}(J, \rho_1)) / \partial J_i|_{J^*, \rho_1} = 0$  for  $i = 2, \dots, \ell - 1$ . By inserting the derivatives 36, and expressing metabolite concentrations through fluxes 29, we find Eq. 6, which can be solved for  $J_{i-1}$ ,

$$J_{i-1}(J_i, J_{i+1}, J_\ell) = J_i + \frac{J_\ell K_i \kappa_{i+1}^{\max} + \sqrt{J_\ell K_i \kappa_{i+1}^{\max} \left( \frac{4 J_i J_\ell J_{i+1} \kappa_i^{\max} K_{i+1}}{(J_i - J_{i+1})^2} + (J_\ell K_i - 4 J_i) \kappa_{i+1}^{\max} \right)}}{\frac{2 J_\ell J_{i+1} \kappa_i^{\max} K_{i+1}}{(J_i - J_{i+1})^2} - 2 \kappa_{i+1}^{\max}}. \quad (37)$$

The result only depends on downstream fluxes  $J_{q \geq i}$  and kinetic constants. From a starting condition  $J_\ell$  and  $J_{\ell-1}$  all fluxes in the chain can be determined iteratively. Only in the last iteration step, a matching nutrient concentration is computed. It is given analogously to Eq. 35 by

$$\rho_1(J) = \frac{\frac{\kappa_2^{\max}}{\kappa_1^{\max}} K_1 \frac{J_1 - J_2}{J_\ell} \left( 2 \frac{J_2}{J_\ell} \left( K_2 - \frac{\kappa_2^{\max}}{\kappa_1^{\max}} \frac{J_1 - J_2}{J_\ell} \right) + K_2 \left( \frac{J_1 - J_2}{J_\ell} + \sqrt{\frac{4 J_1 - J_2}{J_\ell} \frac{J_2}{J_\ell} + \left( \frac{J_1 - J_2}{J_\ell} \right)^2} \right) \right)}{2 \frac{J_2}{J_\ell} \left( K_2 - \frac{\kappa_2^{\max}}{\kappa_1^{\max}} \frac{J_1 - J_2}{J_\ell} \right)^2 - 2 \frac{\kappa_2^{\max}}{\kappa_1^{\max}} K_2 \left( \frac{J_1 - J_2}{J_\ell} \right)^2}. \quad (38)$$

At this point, Eqs. 29 and 30 can again be used to compute all other metabolic state variables of  $\mathcal{S}^*$ .

**Matching the environmental nutrient concentration.** Starting from a minimal chain that consists of only two enzymes, and all its possible optimal metabolic states  $\mathcal{S}_2^*(\rho_1, J_2)$  we can construct optimal metabolic states of a three-enzyme-chain. By shifting indices, we can iteratively determine optimal metabolic states  $\mathcal{S}_\ell^*(\rho_1, J_\ell)$  for any longer chain.

In many biologically relevant scenarios, a boundary condition to be fulfilled is the environmental precursor concentration  $\rho_1$ . However, we here construct a suitable environmental nutrient concentration  $\rho_1$  iteratively from the environmental nutrient concentration of the initial two-enzyme chain. It can easily be seen from Eq. 35 that if any positive  $\rho_1$  was possible for the shorter chain, any positive  $\rho'_0$  can be constructed for the elongated chain. Thus, the iterative method we present here can construct an optimal resource allocation for any environmental nutrient concentration. Unfortunately, it is not clear from the outset of the iteration, which environmental precursor concentration of the two-enzyme chain yields the environmental precursor concentration for the  $\ell$ -enzyme-chain set by the biological scenario of interest. We solve this question numerically, whenever a defined nutrient environment is required.

**Uniqueness of optimal resource allocation.** Any local extremum of growth rate found in this manner is a global optimum because the solution of the chain of two enzymes contains all optima of the chain of two enzymes. In each step, only one physically meaningful solution (with all metabolic state variables larger than zero) of the set of five equations solved for the extension of the chain exists. Thus, the optimal states  $J^*$  determined here are indeed the only sets of fluxes that fulfill Eq. 5. No physiological growth maxima exist at extremal values of any  $J_i$ . Consequently, the optimal metabolic state  $\mathcal{S}_\ell^*(\rho_1, J_\ell)$  is unique and all optimal metabolic states of linear pathways can be constructed by the iterative method.

## Local properties of optimal resource allocation

**Metabolite costs increase along the metabolic chain.** We next show how optimal resource allocation in a network relates to the individual enzyme-substrate pair. To this end, we switch from a variation of a single flux  $J_i$ , which we used to construct growth-optimal balanced growth states above, to a variation of a single metabolite concentration  $\rho_i + \delta\rho_i$ . We again derive the mass changes under such a variation. Eq. 4 connects metabolite concentrations and fluxes in balanced growth states. It indicates that a variation of a single flux can be constructed if all fluxes upstream of reaction  $i$  are increased by a small  $\delta J$ . We insert all changes in fluxes into the Michaelis-Menten relationship to obtain changes in enzyme masses,

$$\delta J_{k < i} = J_\ell \delta \rho_i, \quad \text{and} \quad \delta M_{k < i} = \frac{J_\ell M_k \delta \rho_i}{J_k}. \quad (39)$$

This variation is illustrated in Fig. S1D. Here we implicitly assumed that the environmental nutrient concentration  $\rho_1$  does not depend on the uptake flux  $J_1$ . This may in general not be true under nutrient limitation - for example in a chemostat

setting, in which increased nutrient uptake reduces the steady-state nutrient concentration - and should only be understood as a limiting case in the nutrient-rich regime. We discuss the effect of nutrient limitation in more detail below. The focal reaction is also affected by the variation in  $\rho_i$  - not by a change in flux, but by the change in substrate concentration. We will investigate  $\delta M_i$  in more detail below. For a constant protein production flux  $J_\ell$  and nutrient environment  $\rho_1$ , the total protein mass in the cell must also be minimal under variation of a single metabolite concentration (that is under a linear combination of changes in fluxes [39]). We simplify based on the changes in masses derived above to obtain Eq. [8],

$$\left( \sum_{k=1}^{i-1} \frac{M_k}{J_k} J_\ell + \frac{\partial M_i(J_i, \rho_i, \kappa^{\max}, K)}{\partial \rho_i} \right) \Big|_{J_\ell, \rho^*} = 0 \quad (40)$$

In the optimal state,  $M_P^*$  must be a constant to first-order approximation in any of the variations discussed here. We can therefore equivalently write

$$\frac{\partial}{\partial \rho_i} \left( \beta_i^* \rho_i + \phi_i(j_i, \rho_i, \kappa^{\max}, K) \right) \Big|_{j_i^*} = 0, \quad \text{with} \quad \beta_i^* = \sum_{k=1}^{i-1} \frac{\lambda \phi_k^*}{j_k^*}, \quad \left( \frac{\rho_1}{K_1} \gtrsim 1 \right). \quad (41a, b)$$

Eq. [41a] corresponds to Eq. [9]. The constant [41b] (which we define explicitly as independent of the variation in  $\rho_i$ ) corresponds to the second term of Eq. [10]. The full form of Eq. [10] will be derived below. We obtain a result similar to the minimization of the joint mass concentration of an enzyme and its substrate at constant flux proposed by Dourado et al. [16],  $(\frac{\partial}{\partial \rho})_J(\rho + \phi) = 0$ , which explains experimentally observed relations between enzyme and substrate mass concentrations. In contrast to Dourado et al. we here minimize the total *protein* mass at a fixed *protein* production flux. Our optimization directly follows from growth-optimality: In balanced exponential growth, each constituent of the cell must double at the same rate. Applied to protein, this means that growth rate is given by the protein production flux  $J_\ell$  divided by the total protein mass  $M_P$  of the cell. At fixed  $J_\ell$ ,  $M_P$  must be minimal to achieve optimality. From the outset, it is not clear how such a minimization of the protein mass relates to the metabolite level. We here show that a joint minimization of protein and weighted metabolite mass follows from the optimization of growth rate. The weight  $\beta_i^*$  must be derived for a given metabolic network, and is generally different from 1. We retrace the derivation of the scaling law discovered by Dourado et al. [16] with the additional cost  $\beta_i^*$ . Eq. [30] can be rewritten intensively as  $\phi_i = (j_i/\kappa_i^{\max}) \cdot (1 + K_i/\rho_i)$ , and inserted into Eq. [41a]. Evaluating the derivative and rearranging yields Eq. [12], and by re-inserting into Eq. [30], we find Eq. [11].

## Physiological model parameters

**Growth costs of metabolites and dilution strength  $\epsilon$ .** Eqs. [11] and [12] must ensure that all proteome mass fractions sum to one. This property can be used to establish a parameter scale, which distinguishes between physiological parameterizations that approximately obey a production-consumption balance in each reaction and non-physiological parameterizations that are dominated by dilution. We insert Eq. [11] into [12] and solve for  $\phi_i^*$ ,

$$\phi_i^* = \frac{j_i^*}{\kappa_i^{\max}} + \beta_i^* \rho_i^*. \quad (42)$$

Summing over all proteome mass fractions [27], and inserting mass conservation [4], we derive

$$1 = \sum_{i=1}^{\ell} \phi_i^* = \sum_{i=1}^{\ell} \frac{\lambda^* (1 + \sum_{k=i+1}^{\ell} \rho_k^*)}{\kappa_i^{\max}} + \sum_{i=2}^{\ell} \beta_i^* \rho_i^*. \quad (43)$$

We insert the definition of metabolite costs [41b] (again in the nutrient-rich regime) to derive  $\sum_{i=2}^{\ell} \beta_i^* \rho_i^* > \sum_{i=1}^{\ell} \frac{\lambda^*}{\kappa_i^{\max}} \sum_{k=i+1}^{\ell} \rho_k^*$ , and establish two bounds on growth rate,

$$1 > \sum_{i=1}^{\ell} \frac{\lambda^*}{\kappa_i^{\max}} > 1 - 2 \sum_{i=2}^{\ell} \beta_i^* \rho_i^*. \quad (44)$$

	$\tilde{K}_M$ *	$k_{\text{cat}}$ *	$K = M_S \nu \tilde{K}_M$	$\kappa^{\max} = \frac{M_S}{M_E} k_{\text{cat}}$
All enzymes	$10^{2.0 \pm 1.2}$ $\mu\text{M}$	$10^{1.1 \pm 1.4}$ $\text{s}^{-1}$	-	-
<i>E. coli</i>	$10^{2.0 \pm 1.1}$ $\mu\text{M}$	$10^{1.1 \pm 1.3}$ $\text{s}^{-1}$	$10^{-4.2 \pm 1.1}$	$10^{2.1 \pm 1.3}$ $\text{h}^{-1}$
Model cell	-	-	$10^{-4.2 \pm 1.1}$	$10^{1.6 \pm 0.4}$ $\text{h}^{-1}$ **

**Table 1: Distributions of enzyme kinetic parameters**

To convert to the described unit system, we estimate  $M_S \approx 10^2 \text{ g mol}^{-1}$ ,  $\nu \approx 6 \text{ mL g}^{-1}$  and  $M_E \approx 3.6 \cdot 10^4 \text{ g mol}^{-1}$ .

\*: based on data from Bar-Even et al. (37).

\*\*: reduced mean and variance to account for the reduced number of enzymes in the model, see below.

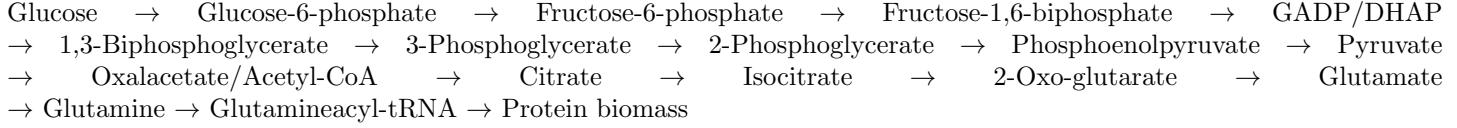
Growth rate must be smaller than the inverse of the sum of all inverse efficiencies, termed  $\lambda^{\max}$ , which would be the growth rate if all enzymes could work at maximal saturation without dilution losses. At the same time, scaled growth rate is reduced by less than  $2 \sum_{i=2}^{\ell} \beta_i^* \rho_i^*$ . By employing again the definition of metabolite costs (10), and  $1 - \sum_{i=2}^{\ell} \beta_i^* \rho_i^* > \sum_{i=1}^{\ell} \frac{\lambda_i^*}{\kappa_i^{\max}}$  (43), we find that  $\epsilon \equiv \sum_{i=2}^{\ell} \sqrt{K_i \lambda^{\max} / \kappa_i^{\max}}$  is an upper bound for  $\sum_{i=2}^{\ell} \beta_i \rho_i$ . Below, we show that physiologically parameterized cells exhibit a small value of  $\epsilon \approx 10^{-1.3}$ . Thus, their growth rate in nutrient-rich environments is predominantly determined by enzyme efficiencies.

**Physiological kinetic parameters of enzymes.** To estimate the value of  $\epsilon$  we use available data for kinetic rate constants. Bar-Even et al. (37) found that enzyme kinetic parameters across numerous species tend to follow a log-normal distribution. Restricting their analysis to *E. coli* enzymes yields distributions of similar mean and variance (see tab. 1). We leverage their results to approximate  $K$  and  $\kappa^{\max}$  distributions for model parameterization. To convert units as described above, we estimate the average molecular weight of substrates  $M_S$  based on the molecular weight of glucose and amino acids (BNID: 104877); the volume per mass ratio  $\nu$  (26) as the inverse of the protein mass per volume density in *E. coli* computed from the product of the dry mass protein fraction of *E. coli* (BNID: 101955) and its dry mass density (BNID: 109049); and the average protein mass  $M_E$  using the average protein length (BNID: 108986) and the average molecular weight of amino acids in *E. coli* (BNID: 104877). We disregarded any molar mass variability for substrates and enzymes on the grounds they are minor relative to the intrinsic variation of the kinetic parameters  $\tilde{K}_M$  and  $k_{\text{cat}}$  and potentially correlated with one another and/or kinetic parameters.

**Variability in  $\kappa^{\max}$  and central metabolism dynamics.** In comparison to a real cell, our model incorporates a reduced spectrum of enzyme types, namely those of a single pathway converting environmental nutrients to new protein biomass. In a real cell, enzymes unaccounted here convert metabolites through parallel, interconnected pathways and generate a variety of intermediary metabolites. In addition to a larger variety in protein precursors, these pathways also synthesize other cellular biomass components such as nucleic acids and lipids. Given the large spectrum of enzymes, the average proteome mass fraction of an individual enzyme type remains small. No singular enzyme mass fraction takes over the entire proteome, despite significant variability in catalytic rates.

To achieve physiological growth rates and prevent a singular enzyme from completely dominating mass allocation in our model, we use a distribution of enzyme efficiencies  $\kappa^{\max}$  with lower mean and variance in comparison to the values found by Bar-Even et al. This choice addresses the reduced spectrum of enzyme types and is further motivated by evolutionary arguments: here we study central metabolism, which covers the conversion of environmental nutrients into new protein biomass. Central metabolism enzymes likely undergo more stringent selection than enzymes in secondary metabolism, resulting in decreased variance in their efficiencies. In contrast, we maintained the natural variance in substrate affinities and thereby maintained a near-physiological variance in enzyme qualities.

To estimate the chain length of the metabolic network in a living cell, we focus on the conversion of an environmental carbon source into new protein biomass. This pathway likely contributes most to the overall protein biomass production of the cell. Environmental carbon sources are converted into different amino acids via different pathways. To estimate the physiological pathway length, we use the production of glutamine from glucose, which consists of 17 enzymatic steps connecting the following metabolites:



Other amino acid biosynthesis pathways include fewer (e.g. Serine), equally many (e.g. Threonine), or more enzymatic steps (e.g. Arginine). Alternative carbon sources may require a different number of enzymatic steps to produce pyruvate. Some enzymes may require additional modifications after protein synthesis. We here use  $\ell \approx 20$  as an order of magnitude estimate for growth in amino-acid-free growth media.

**Stochastic model parametrization.** While kinetic properties of individual enzymes become increasingly available, only a minority has been measured to date. Additionally, enzymes with multiple substrates or otherwise complicated reaction kinetics may not exhibit a clearly defined substrate affinity. For these reasons, we do not parameterize our model based on available data for the specific enzymes. Instead, we parameterize our model stochastically, drawing both substrate affinities and kinetic constants independently from the log-normal distributions given in table 1. We thereby also disregard any bias in biochemical qualities along the metabolic cascade in our model. We note, however, that ribosomes exhibit a particularly high molecular mass but only an average catalytic constant  $k_{cat} \approx 20 \text{ aa s}^{-1}$  (11). Together with the average molecular weight of amino acids, these properties yield a distinctly low catalytic rate of  $\kappa^{\max} = 10^{0.9} \text{ s}^{-1}$ . Furthermore, they operate very close to their maximal speed in the nutrient-rich regime (47, 48) despite typical aminoacyl-tRNA concentrations in the  $\mu\text{M}$ -concentration regime (BNID: 105275). We therefore estimate a strong ribosomal substrate affinity  $K^{-1} > 10^6$ . Both properties are the sign of a specificity-optimized protein. As outlined further in the main text, we hypothesize that the ribosome is part of a coherently regulated group of specificity-optimized enzymes that lie downstream of most nutrient limitations.

**Environmental nutrient concentrations.** Amino-acid-free laboratory media for bacterial growth such as M9 typically contain between  $0.1 \text{ g L}^{-1}$  and  $1 \text{ g L}^{-1}$  glucose. The glucose concentration  $\tilde{K}_{\text{Monod}}$  at which the growth rate of *E. coli* is half-maximal is on the order of  $1 \text{ mg L}^{-1}$  (BNID: 111048). We here assume - and later show that this assumption is in line with our model's predictions - that the environmental nutrient concentration at which growth is half-maximal is on the order of the substrate affinity of the nutrient uptake enzyme,  $\tilde{K}_{\text{Monod}} \approx \tilde{K}_1$ . This suggests that nutrient uptake enzymes are typically highly saturated in nutrient-rich laboratory media. Based on these typical values, we specifically use  $\rho_1 = 1000 K_1$  to describe nutrient-rich media. This value reflects all properties of the limit  $\rho_1 \rightarrow \infty$  well but is numerically more easily accessible than the limit itself.

## Global properties of optimal resource allocation

**Physiological enzyme saturation.** We compute the inactive proteome fraction  $\phi_{\text{inactive}}$  (i.e. enzymes not currently engaged in catalysis due to a lack of substrates) as  $\phi_{\text{inactive}} \equiv \sum_{i=1}^{\ell} \phi_i^* (1 - \kappa_i^* / \kappa_i^{\max})$ . Inserting the definition of  $\kappa_i$  and Eq. 12, this yields

$$\phi_{\text{inactive}} = \sum_{i=2}^{\ell} \beta_i^* \rho_i^* < \epsilon \quad \left( \frac{\rho_1}{K_1} \gtrsim 1 \right). \quad (45)$$

Enzymes of physiologically parameterized cells in nutrient-rich environments are typically highly saturated - on average to a level larger than  $1 - \epsilon$ . Our model thus explains solely from the physiological distribution of enzyme kinetic parameters, why enzymes are typically well-saturated in nutrient-rich environments (38).

**The cellular metabolite mass fraction in the homogeneous chain.** Consider a homogeneous chain (which consists of enzymes with equal kinetic constants  $\kappa^{\max}$  and  $K$  only) in the nutrient-rich regime. In the physiological parameter regime,

such a chain fulfills  $\epsilon \leq \sqrt{(\ell-1)K} \ll 1$ . In the limit of small dilution strength  $\epsilon$ , Eqs. [1], [10], [11], [12] and [29], are then jointly solved by

$$\begin{aligned}
 \rho_i^* &\approx 0 + \epsilon \sqrt{\frac{1}{(i-1)(\ell-1)}} + \mathcal{O}(\epsilon^2), \\
 \phi_i^* &\approx \frac{1}{\ell} (1 + \epsilon \left( \frac{2}{3} - \sqrt{\frac{i-1}{\ell-1}} \right) + \mathcal{O}(\epsilon^2)), \\
 j_i^* &\approx \frac{\kappa_{\max}}{\ell} (1 + \epsilon \left( \frac{2}{3} - 2\sqrt{\frac{i-1}{\ell-1}} \right) + \mathcal{O}(\epsilon^2)), \\
 \beta_i^* &\approx \frac{i-1}{\ell} (1 + \epsilon \left( \frac{2}{3} \sqrt{\frac{i-1}{\ell-1}} - \frac{4}{3} \right) + \mathcal{O}(\epsilon^2)), \\
 \lambda^* &\approx \frac{\kappa_{\max}}{\ell} (1 - \epsilon \frac{4}{3} + \mathcal{O}(\epsilon^2)), \\
 \epsilon &\approx \sqrt{(\ell-1)K}.
 \end{aligned} \tag{46}$$

We here use  $\sum_{k=i+1}^{\ell} (k-1)^{-1/2} \approx 2(\sqrt{\ell-1} - \sqrt{i-1})$  and  $\sum_{i=1}^{\ell} \sqrt{i-1} \approx 2\ell\sqrt{\ell-1}/3$ , which are motivated by the corresponding integrals. With these approximations, Eqs. [46] form a consistent solution of microbial resource allocation to first order in  $\epsilon$ , and the limit of vanishing  $\epsilon$  constitutes the limit of vanishing dilution fluxes. In such a homogeneous chain, we can then compute the cellular mass fraction as  $M_R/M_P = (j_1 - j_{\ell})/j_{\ell} \approx 2\epsilon$  [18], which is itself on the order of  $\epsilon$ . At the same time, the growth effect of metabolite dilution is estimated [46] to be on the order of  $4\epsilon/3$ , which stems approximately equally from proteome resources required for unsaturated enzymes and proteome resources required to balance metabolite dilution fluxes [44].

**Resource allocation in heterogeneous chains.** To extend our analysis to heterogeneous chains, we replace position by the continuous variable  $\beta$ , and sums over position by integrals over  $\ell d\beta$ . For a compact notation we define the relative specificity  $s_{\beta}$ ,

$$s_{\beta} \equiv \frac{\sqrt{\frac{K_{\beta}}{\kappa_{\beta}^{\max}}}}{\int_0^1 \sqrt{\frac{K_{\beta'}}{\kappa_{\beta'}^{\max}}} d\beta'}. \tag{47}$$

The central equations for  $\rho$ ,  $j$  and  $\phi$  then yield a self-consistent solution even in heterogeneous chains (under the additional assumptions that  $\beta$  is not correlated with other variables and fluctuations in other variables average out efficiently even across small intervals in  $\beta$  - i.e. any integral  $\int_L s_{\beta} d\beta \approx L$ ):

$$\begin{aligned}
 \rho_i^* &\simeq \frac{\epsilon}{\ell} \frac{s_i}{\sqrt{\beta_i}}, \\
 \phi_i^* &\simeq \frac{\lambda}{\kappa_i^{\max}} \left( 1 + 2\epsilon(1 - \sqrt{\beta_i}) \right) + \frac{\epsilon}{\ell} s_i \sqrt{\beta_i}, \\
 \lambda^* &\simeq \left( \sum_{k=1}^{\ell} \frac{1}{\kappa_k^{\max}} \right)^{-1} \left( 1 - \frac{4}{3}\epsilon \right) \quad \left( \frac{\rho_1}{K_1} \gtrsim 1 \right).
 \end{aligned} \tag{48}$$

For sufficiently small variations of efficiencies and affinities around their respective mean, this derivation - together with the mean-field approximation for  $\beta_i$  [46] - provides a reasonable mean-field approximation of heterogeneous chains (Fig. S4). Systematic deviations introduced by physiological enzyme heterogeneity are small, as well as typical stochastic deviations. We note however, that this is not true for all heterogeneous parameterizations in the case of the cellular metabolite mass fraction. In heterogeneous chains, no upper limit on the metabolite mass fraction can be established: if the nutrient uptake enzyme has an exceptionally high catalytic rate, its proteome fraction nearly vanishes. Then, metabolite cost  $\beta_2^*$  nearly vanishes, and  $\rho_2^*$  can reach arbitrary values, which on their own may exceed  $2\epsilon$  drastically.

**Comparison to experimental data.** We compute several observables regarding the whole cell under the assumption of optimal resource allocation in our model. For physiological kinetic parameter distributions, they are in good agreement with those observed empirically (Fig. S2), apart from a slight overestimation of the cellular metabolite mass fraction. The metabolic resource allocation patterns identified here are generic and independent of the mean of  $\kappa^{\max}$ . They are also largely independent of the variance in  $\kappa^{\max}$ , as long as individual enzymes do not dominate the proteome. Overall higher efficiencies solely lead to higher growth rates and fluxes. Higher values of  $K$ , on the other hand, lead to higher dilution relative to growth, creating more significant flux gradients. This would also result in stronger position-dependent gradients in optimal metabolite saturation.

## Optimal balanced growth states under nutrient limitation

We consider an additional constraint on the nutrient uptake flux from the environment (which was left unconstrained in the nutrient-rich regime, and jointly optimized with the rest of the metabolic state of the cell), next to the constraint on the cellular protein concentration. We then identify the metabolic state, which maximizes growth rate at the given nutrient uptake flux. In the absence of such a constraint on nutrient uptake flux at low nutrient concentrations, the complete calculus presented above would remain valid at low nutrient concentrations. Metabolite production costs for all intracellular metabolites would remain on the order of 1, and all  $\phi_{i \geq 2}$  would converge to zero with decreasing flux  $j_1$  according to Eqs. 12, 11 and  $j_i < j_1$ . This implies that the uptake protein would become a macroscopic fraction of the total proteome, and eventually the whole proteome. Instead, for the reasons outlined in the main text, we specifically assume that uptake protein expression is constrained, and thereby limits the nutrient uptake flux. Consider any constraint on  $\phi_1(\rho_1)$ . To find optimal resource allocation under the constraint, we use the method of Lagrange multipliers,

$$\mathcal{L}(J, \rho_1, \mu) \equiv M_P(J, \rho_1) + \mu \left( \frac{M_1(J, \rho_1)}{M_P(J, \rho_1)} - \phi_1(\rho_1) \right), \quad (49)$$

with a Lagrange multiplier  $\mu$ . The optimal resource allocation must now fulfill  $\partial \mathcal{L} / \partial J_i|_{J^*, \rho_1} = 0$  for  $i = 1, \dots, \ell - 1$ , in analogy to Eq. 34, and additionally  $\partial \mathcal{L} / \partial \mu = 0$ . For  $i \geq 2$  the uptake mass  $M_1$  does not depend on  $J_i$ , and the partial derivatives are

$$\frac{\partial \mathcal{L}}{\partial J_i} = (1 - \phi_1 \frac{\mu}{M_P}) \frac{\partial M_P}{\partial J_i} \quad (i = 2, \dots, \ell - 1). \quad (50)$$

They are zero in the optimal state if Eq. 5 is fulfilled. Thus the iterative computation of optimal fluxes 37 is still valid.  $J^*$  and all  $M_i^*$  for  $i$  in  $(2, \dots, \ell)$  follow from the iterative computation of optimal fluxes. A different boundary condition  $\rho_1$  now matches a certain starting condition of the iteration  $(J_\ell^*, J_{\ell-1}^*)$  in comparison to unconstrained optimality. For a given constraint  $\phi_1(\rho_1)$ , this boundary condition can be found by solving

$$J_1^* = \frac{1}{1 - \phi_1(\rho_1)} \sum_{k=2}^{\ell} M_k^* \frac{\kappa_1^{\max} \phi_1(\rho_1)}{1 + \frac{K_1}{\rho_1}} \quad (51)$$

for  $\rho_1$ , which corresponds to solving  $\partial \mathcal{L} / \partial \mu = 0$ . The last remaining condition on optimality is  $\partial \mathcal{L} / \partial J_1|_{J^*, \rho_1} = 0$ , which could be solved for the remaining undetermined variable  $\mu$ . However,  $\mu$  is not required for describing the optimal state, and equations given above 37, 51 fully suffice to compute optimal balanced growth states under nutrient limitation. The iterative construction of optimal balanced growth states is applicable under nutrient uptake constraints.

**Metabolite costs under nutrient limitation.** To put this result into the context of metabolite costs, we again consider the variation of a single metabolite concentration (and drop the superscripts  $^*$  below). Based on Eq. 4, we derive

$$j_i = j_1 \frac{1 + \sum_{k=i+1}^{\ell} \rho_k}{1 + \sum_{k=2}^{\ell} \rho_k}. \quad (52)$$

This means that under a constraint on nutrient uptake, which defines a fixed flux  $j_1(\rho_1)$  in a given nutrient environment, variation of a single metabolite concentration violates growth optimality,  $\partial j_\ell / \partial \rho_k \neq 0$ . A variation of a single metabolite concentration that maintains all other metabolite concentrations is ill-defined. Instead, we reformulate the optimality condition. We observe that in an optimal balanced growth state under a nutrient uptake constraint, the sum of metabolite concentrations must be minimal to maximize growth [52]. This criterion can be reformulated further: all variations of any single metabolite concentration must have the same effect on the sum of proteome mass fractions  $\sum_{k=2}^\ell \phi_k$ ,

$$\frac{\partial \sum_{k=2}^\ell \phi_k}{\partial \rho_i} = \frac{\partial \sum_{k=2}^\ell \phi_k}{\partial \rho_{i'}} \quad (i, i' = 2, \dots, \ell). \quad (53)$$

To verify, suppose that this statement was incorrect in a growth-optimal state. A joint variation of  $\rho_i$  and  $\rho_{i'}$  by the same  $\delta \rho$  in opposite directions could reduce the required sum of proteome fractions. Adding the freed proteome fraction to any reaction would reduce dilution fluxes and thereby improve growth rate.

We insert the kinetic growth law  $\phi_k(j_k, \rho_k, \kappa_k^{\max}, K_k)$ , and find

$$\frac{\partial \sum_{k=2}^\ell \phi_k}{\partial \rho_i} = \frac{\partial \phi_i}{\partial \rho_i} \Big|_{j_i} + \sum_{k=2}^\ell \left( \frac{\phi_k}{j_k} \frac{\partial j_k}{\partial \rho_i} \right). \quad (54)$$

We insert [52] and subtract Eq. [53] for  $i' = 2$ ,

$$0 = \frac{\partial \phi_i}{\partial \rho_i} \Big|_{j_i} - \beta_0^* + \sum_{k=1}^{i-1} \frac{\lambda \phi_k}{j_k} \quad \text{with} \quad \beta_0^* \equiv -\frac{\partial \phi_2}{\partial \rho_2} \Big|_{j_2} - \frac{\lambda \phi_1}{j_1}. \quad (55)$$

A single summand is added in comparison to the unconstrained optimality condition [40]. This additional cost associated with nutrient uptake affects all metabolite production costs equally as an additive constant. We obtain the functional form  $\beta_i^* = \beta_0^*(\rho_1) + \sum_{k=1}^{i-1} \lambda^* \phi_k^* / j_k^*$ , which is Eq. [10].

Again, local metabolite production costs are not suitable for determining optimal resource allocation, as they depend on the optimal resource allocation towards other reactions. Instead, we use them to interpret patterns in optimal resource allocation. We therefore do not solve for  $\beta_0^*$  analytically, but simply identify the additional summand as an arbitrary offset, which depends on the nutrient environment and the nutrient uptake constraint. Under variation of  $\rho_1$  we biologically expect to find a one-parametric family of optimal states that fulfill the constraint.

Instead of characterizing optimal states by their nutrient concentration  $\rho_1$ , we choose to characterize them by their growth rate  $\lambda^*$ .  $\beta_0^*$  is an arbitrary function of  $\lambda^*$  in this scenario. We can analogously to the case of unconstrained optimality derive  $\rho_i^* = \arg \min((\beta_i^{\text{rich}} + \beta_0^*(\lambda^*))\rho_i + \phi(\rho_i, J_i))|_{J_i, \beta_i^{\text{rich}}, \beta_0^*}$ . Here,  $\beta_i^{\text{rich}}$  denotes the expression found for  $\beta_i^*$  in the nutrient-rich regime ( $\beta_i^{\text{rich}} \equiv \sum_{k=1}^{i-1} ((\lambda^* \phi_k^*) / j_k^*)$ ). The shape of the equation, including an additive  $\beta_0^*$  is independent of the exact nature of the constraint limiting nutrient uptake, as long as it only affects the flux  $j_1$ . Notably, a diffusive constraint on nutrient uptake leads to the same result: in this case, the mass required for increasing the nutrient uptake flux  $j_1$  would be larger than  $\delta j_1 / \kappa_1$ , because the local nutrient concentration at the cell surface would decrease. This also yields an additional cost  $\beta_0^*$ , which affects all reactions equally. Based on this argument, we here assume that the additional cost  $\beta_0^*$  vanishes in a nutrient-rich environment. In a nutrient-rich environment, the balance between diffusion and uptake has a vanishingly small effect on the metabolite concentration at the cell surface. Additionally, uptake protein expression was observed to increase under nutrient limitation, making it unlikely that the uptake protein expression is constrained below its unconstrained-optimal value in nutrient-rich environments. The effect of a regulatory nutrient uptake constraint on uptake expression, growth and metabolite cost is illustrated in Fig. S3.

**The limit of vanishing nutrient uptake.** Above, we derived that a nutrient uptake limitation affects all reactions equally by an additive cost component  $\beta_0^*$ . Solving for  $\beta_0^*(\rho_1)$  for a given uptake constraint  $\phi_1(\rho_1)$  is challenging due to the same feedback loops concerning optimal balanced growth states discussed earlier. However its form as an additive constant allows for a simplified computation based on the summation constraint on proteome fractions,

$$1 = \sum_{i=1}^\ell \phi_i^* = \phi_1(\rho_1) + \sum_{i=2}^\ell \left( \frac{j_i^*}{\kappa_i^{\max}} + (\beta_i^{\text{rich}} + \beta_0^*)\rho_i^* \right). \quad (56)$$

In the limit of strong nutrient limitation  $j_1 \rightarrow 0$ , this constraint can be solved directly for  $\beta_0^*(j_1)$ : By Eq. 4, all fluxes fulfill  $j_i < j_1$  and vanish in the limit. Thus, each  $\phi_i$  is well approximated by the linear term  $((\beta_i^{\text{rich}} + \beta_0^*)\rho_i^*)$  of Eq. 12. We denote  $\lim_{j_1 \rightarrow 0} \phi_1$  with  $\phi_1^{\max}$ . By inserting Eqs. 11 and 10, we find

$$1 - \phi_1^{\max} = \lim_{j_1 \rightarrow 0} \left( \sum_{i=2}^{\ell} \sqrt{\frac{K_i}{\kappa_i^{\max}}} \sqrt{j_i^*(\beta_0^* + \sum_{k=1}^{i-1} \frac{\lambda_k^*}{\kappa_k^*})} \right). \quad (57)$$

For the reasons outlined in the main text, we assume that the uptake enzyme never constitutes the whole biosynthetic proteome, even in the limit of vanishing  $j_1$ , and  $\phi_1^{\max} < 1$ . Thus, the right-hand-side must also remain finite in the limit. For a given  $j_1$ , growth is maximized if the metabolite mass fraction  $M_R/M_P$  is minimal 52. We show below by construction, that  $M_R/M_P \rightarrow 0$  in the limit of  $j_1 \rightarrow 0$ . We expand in small  $j_1$  and drop all terms of order  $j_1$ . We specifically use  $j_i = j_1 + \mathcal{O}(j_1^2)$ , and that all terms  $\beta_i^{\text{rich}} = \sum \lambda_k^* / \kappa_k^* < \sum \phi_k^* \leq 1$  remain finite in the limit,

$$1 - \phi_1^{\max} = \lim_{j_1 \rightarrow 0} \left( \sum_{i=2}^{\ell} \sqrt{\frac{K_i}{\kappa_i^{\max}}} \sqrt{j_1 \beta_0^*} \right). \quad (58)$$

Under a strong constraint on nutrient uptake,  $\sqrt{j_1 \beta_0^*}$  must therefore converge to a suitable normalization constant  $Z$ , and  $\beta_0^*$  must diverge as  $Z^2 \cdot j_1^{-1}$ . This means that resource allocation in the limit of vanishing nutrient uptake is given by

$$\begin{aligned} \lim_{j_1 \rightarrow 0} \phi_i^* &= \sqrt{\frac{K_i}{\kappa_i^{\max}}} Z, \\ \rho_i^* &\approx \frac{j_1}{Z} \sqrt{\frac{K_i}{\kappa_i^{\max}}} (1 + \mathcal{O}(j_1)), \\ j_i^* &\approx j_1 (1 + \mathcal{O}(j_1)) \quad \text{with } Z = \frac{(1 - \phi_1^{\max})}{\sum_{i=2}^{\ell} \sqrt{\frac{K_i}{\kappa_i^{\max}}}} \quad \text{for } (i = 2, \dots, \ell). \end{aligned} \quad (59)$$

As assumed above, metabolite concentrations vanish in the limit of vanishing nutrient uptake. In this limit, optimal proteome fractions are not anymore set by enzyme efficiencies, but by their specificities.

**Expansion to small nutrient uptake flux.** An optimal balanced growth state in the limit of vanishing nutrient uptake is experimentally unattainable. For this reason, we compute optimal balanced growth states at small but finite growth rates by a first-order expansion of Eq. 56 in small  $j_1 = \lambda^*(1 + \mathcal{O}(\epsilon))$ ,

$$1 - \phi_1(\lambda^*) - \sum_{i=2}^{\ell} \frac{\lambda^*}{\kappa_i^{\max}} (1 + \mathcal{O}(\epsilon)) = \sum_{i=2}^{\ell} \sqrt{\frac{K_i}{\kappa_i^{\max}}} \sqrt{\lambda^* \beta_0^*} (1 + \mathcal{O}(\epsilon)) (1 + \frac{\beta_i^{\text{rich}}}{2\beta_0^*}). \quad (60)$$

As long as  $\beta_i^{\text{rich}} \ll \beta_0^*$ , we obtain

$$\sqrt{\beta_0^*} \approx \frac{1 - \phi_1(\rho_1) - \sum_{i=2}^{\ell} \frac{\lambda^*}{\kappa_i^{\max}} (1 + \mathcal{O}(\epsilon))}{\sum_{i=2}^{\ell} \sqrt{\frac{\lambda^* K_i}{\kappa_i^{\max}}} (1 + \mathcal{O}(\epsilon))}. \quad (61)$$

We specifically find for  $\rho_i$  by Eq. 11, and for  $\phi_i$  by Eq. 56,

$$\rho_i^* \approx \lambda^* \sqrt{\frac{K_i}{\kappa_i^{\max}}} \frac{\sum_{i=2}^{\ell} \sqrt{\frac{K_i}{\kappa_i^{\max}}} (1 + \mathcal{O}(\epsilon))}{1 - \phi_1(\rho_1) - \sum_{i=2}^{\ell} \frac{\lambda^*}{\kappa_i^{\max}} (1 + \mathcal{O}(\epsilon))}, \quad (62)$$

$$\phi_i^* \approx \frac{\lambda^*}{\kappa_i^{\max}} (1 + \mathcal{O}(\epsilon)) + \frac{\sqrt{\frac{K_i}{\kappa_i^{\max}}} \left( 1 - \phi_1(\rho_1) - \sum_{i=2}^{\ell} \frac{\lambda^*}{\kappa_i^{\max}} (1 + \mathcal{O}(\epsilon)) \right)}{\sum_{k=2}^{\ell} \sqrt{\frac{K_k}{\kappa_k^{\max}}} (1 + \mathcal{O}(\epsilon))}. \quad (63)$$

Proteome mass fractions are approximately linear in growth rate in the limit of strong nutrient limitation (apart from nonlinearities introduced by  $\phi_1(\rho_1)$ , and metabolite dilution terms of order  $\epsilon$ ) and converge to well-defined values as derived above. Correction terms are of order  $\mathcal{O}(\epsilon)$  in Eqs. 62 and 63. For physiological parameter values, they can thus be neglected even at higher growth rates, and the approximation above remains valid as long as  $\beta_i^* \approx \beta_0^*$ . For this,  $\beta_0^*$  must be larger than all  $\beta_i^{\text{rich}}$ . We estimate based on Eq. 61, and the definition of  $\epsilon$ ,  $\sqrt{\beta_0^*} \geq (1 - \phi_1(\rho_1) - \sum_{i=2}^{\ell} \lambda^* / \kappa_i^{\max}) / \epsilon$ . Position-dependent metabolite production costs  $\beta_i^{\text{rich}}$  do not exceed 1. If the constraint on  $\phi_1$  is stringent enough to ensure  $\phi_1 < 1 - \sum_{i=2}^{\ell} \frac{\lambda}{\kappa_i^{\max}} - \epsilon$ , the added cost exceeds 1. Thus, proteome fractions are approximately linear in growth rate at growth rates

$$\lambda^* < \lambda^{\max} (1 - \phi_1^{\max} - \epsilon). \quad (64)$$

This constitutes the boundary between the nutrient-rich and the nutrient-poor regime. The boundary translates approximately into  $\rho_1/K_1 \simeq 1$ : Above the boundary, the catalyzed flux through the first reaction is largely unaffected by reductions of  $\rho_1$ , and growth rate is approximately  $\lambda^{\max}$ . Below  $\rho_1/K_1 \simeq \phi_1^{\infty} / \phi_1^{\max}$ ,  $\phi_1$  is not up-regulated meaningfully further, a property discussed in more detail in the SI section on regulatory constraints. Then, any further reduction in  $\rho_1$  will lead to a near-proportional reduction in  $\lambda$ , and  $\beta_0$  quickly exceeds one. The approximate equivalence of  $\lambda \simeq \lambda^{\max} (1 - \phi_1^{\max} - \epsilon)$  with  $\rho_1/K_1 \simeq 1$  was further verified numerically (Fig. S2F).

**Linearity of expression level changes.** Under physiological conditions, both  $\phi_1^{\max}$  and  $\epsilon$  are typically small. Neglecting both, an even stronger approximation can be made to simplify proteome fractions as a function of growth rate 63,

$$\phi_i^* \approx \frac{\lambda}{\kappa_i^{\max}} + \left(1 - \frac{\lambda}{\lambda^{\max}}\right) \frac{\sqrt{\frac{K_i}{\kappa_i^{\max}}}}{\sum_{i=2}^{\ell} \sqrt{\frac{K_i}{\kappa_i^{\max}}}}, \quad (65)$$

which is a linear function in growth rate for each enzyme. We normalize by the expression level in the nutrient-rich regime  $\phi_i \approx \lambda^{\max} / \kappa_i^{\max} + \mathcal{O}(\epsilon)$ . Inserting the definition of  $q$  and the approximate maximal growth rate for negligible  $\phi_1$  and  $\epsilon$ ,  $\lambda^{\max} \approx (\sum_{k=2}^{\ell} 1 / \kappa_k^{\max})^{-1}$ , we find that an enzyme's biochemical quality directly determines the optimal expression fold-change between a nutrient-rich environment and the limit of vanishing nutrient availability,

$$\frac{\phi_i^*(\lambda)}{\phi_i^*(\lambda^{\max})} \approx q_i + (1 - q_i) \frac{\lambda}{\lambda^{\max}}. \quad (66)$$

In a chain of irreversible Michaelis-Menten reactions with physiological parameters, any constraint that restricts the uptake proteome fraction from becoming a macroscopic fraction of the proteome under nutrient limitation leads to an optimal resource allocation pattern approximately linear in growth rate for all enzymes except the uptake enzyme. Enzymes with  $q < 1$  are down-regulated upon nutrient limitation, enzymes with  $q > 1$  are up-regulated. If a finite  $\phi_1$  is to be taken into account, the critical value of  $q$  that differentiates between up- and down-regulation shifts to  $q_C \approx 1 + \phi_1^{\max}$ . Further non-linear corrections arise at high growth rates due to finite  $\epsilon$  and, in particular at very high  $q$ , due to the lowered substrate saturation in the nutrient-rich regime (Fig. 2). However, these corrections are typically small in the physiological parameter regime (Fig. 4).

## Model evaluation against physiological cell characteristics

**Applicability to individual cascades.** Our calculations indicate that the discovered two-dimensional maps, flux- and saturation gradients as well as linear expression level changes upon nutrient limitation are not just properties of averages, but can be recovered for individual cascades. We illustrate this result in Fig. S4. Quantitatively, position and biochemical quality account for about 98 % of the variances in scaled metabolite concentrations and expression levels respectively in the nutrient-rich regime, as well as about 80 % of the corresponding variances under nutrient limitation in our ensemble of optimal resource allocation in physiologically parameterized cascades. These fractions denotes percent reductions in squared residuals before and after applying the re-scaling derived through our analytical calculations (see Fig. S4). We thus expect our results to be applicable to a single physiological parameterization, as would be encountered in a given experiment.

**Predicted correlations in proteomics and metabolomics datasets.** Our model predicts a correlation of protein expression fold-changes and protein expression levels, as both depend on biochemistry. For physiologically small dilution strength, most enzymes are well-saturated, and the quadratic term dominates Eq. 12,

$$\frac{\phi_i}{K_i} \approx \beta_i \left( \frac{\rho_i^*}{K_i} \right)^2 \quad \text{for } \left( \frac{\rho_i}{K_i} \gtrsim 1 \right). \quad (67)$$

We insert Eq. 11 and  $j_i \approx \lambda + \mathcal{O}(\epsilon)$ , to find  $\phi_i/K_i \sim 1/q_i^2$ . Together with Eq. 66, this defines the correlation

$$\log \left( \frac{\phi_i^{\text{lo}}}{\phi_i^{\text{hi}}} \right) \approx -\frac{1}{2} \log(\phi_i^{\text{hi}}) + \frac{1}{2} \log(K_i) + \text{const.} \quad (68)$$

We predict an anti-correlation with log-log-slope of  $-1/2$  between expression levels in the nutrient-rich regime and expression fold-changes in the limit of vanishing growth. We note that the affinity-dependent term  $\log(K_i)/2$  may introduce considerable noise. If no affinities are known, averaging over a large number of enzymes may be required to identify the anti-correlation. We here test whether an anti-correlation is present when taking the entire proteome into account (Fig. 5). We note that this approach is biased for several reasons: the proteome contains not only enzymes involved in the conversion of environmental nutrients into new cell biomass, but for example also structural proteins and stress-response proteins. Our model does not account for such other proteome constituents, and does not predict their growth responses. At the same time, our model suggests that protein complexes should be treated as a joint proteome fraction. Still, we observe an anti-correlation stronger than expected from regression to the mean, both in the entire proteome, and when restricting our analysis to individual metabolic pathways, which consist solely of proteins involved in biomass synthesis.

Our analysis of the proteome could be improved by including additional kinetic or metabolomic data. Unfortunately, accurate metabolomic and kinetic data is not available for many reactions. A small number of reactions was recently analyzed by Dourado et al. 16. In this dataset, quantitative deviations from the original enzyme-substrate relationship appear too small for the considered enzymes for meaningful analysis. We expect that ongoing efforts to expand the availability and accuracy of kinetic data 11 will allow for a direct test of our predictions. Under nutrient limitation, we predict that the added cost of metabolite production should shift the optimal enzyme-substrate relationship to lowered metabolite concentrations. Interestingly, a qualitatively similar deviation from the un-shifted enzyme-substrate relation was observed experimentally when reducing growth rate by switching the carbon source of a growth medium 16.

## Extension to more complex network topologies

Metabolic networks typically exhibit a complex topology. Capturing the full complexity of a real cell lies beyond the scope of this paper, as it would impede analytical solvability, vastly increase the degrees of freedom in the parameterization, and impede the identification of simple patterns. However, we use an exemplary network with more complex topology to illustrate that our framework applies to more complex topologies and that its key predictions are still valid. We focus on one particularly interesting use case that can be directly linked to existing proteomics studies: the consumption of two different, non-interchangeable environmental nutrients by the cell, such as a carbon and a nitrogen source. We specifically model this system as two convergent pathways with two different environmental nutrients (Fig. 6E). To ease the notation, we call one pathway the N-branch and the other the C-branch. At the convergence point, both metabolite types (C-type and N-type derived from the two nutrients) are required for a single reaction. The product of this reaction is further converted by a third pathway, the stem of the metabolic network.

We assume that the enzyme at the convergence point exhibits Michaelis-Menten-kinetics in both of its substrates - i.e. both substrates are needed for the reaction to proceed,

$$J = \frac{\kappa^{\max} M}{1 + \frac{K^N}{\rho_N} + \frac{K^C}{\rho_C} + \frac{K^{CN}}{\rho_N \rho_C}}. \quad (69)$$

This rate law features a kinetic constant  $K^{CN}$  (with units of concentration squared), for which no physiological parameter regime is known. We here assume that  $K^{CN} = K^N K^C$ . We further extend the framework of mass conservation to the

convergence point: we assume that the two metabolite types are used with a fixed mass stoichiometry. The C-branch contributes a mass flux fraction  $\gamma_C$  to the downstream mass flux, the N-branch  $\gamma_N = 1 - \gamma_C$ , with  $0 < \gamma_C < 1$ . Enzymes along the C-branch and the stem form a chain of  $\ell$  enzymes. Its enzymes are numbered as before. On the other hand, enzymes in the N-branch are numbered by their position in the branch. To differentiate, we denote indices in the N-branch by a tilde  $\tilde{y}$ . The N-branch has length  $\tilde{\ell}$ . Let  $y$  be the index of the enzyme with two substrates, so that  $\rho_C \equiv \rho_y$ . Mass conservation demands

$$J_C \equiv J_{y-1} = \gamma J_y + J_\ell \rho_C, \quad \text{and} \quad J_N \equiv J_{\tilde{\ell}} = (1 - \gamma) J_y + J_\ell \rho_N. \quad (70)$$

We can now express  $M_y$  as a function of only fluxes and kinetic constants,

$$M_y = \frac{J_y}{\kappa_{CN}^{\max}} \left( 1 + \frac{K^N J_\ell}{J_N - (1 - \gamma) J_y} + \frac{K^C J_\ell}{J_C - \gamma J_y} + \frac{K^{CN} (J_\ell)^2}{(J_N - (1 - \gamma) J_y)(J_C - \gamma J_y)} \right). \quad (71)$$

All other enzyme masses are still given by Eq. 30. Thus, all fluxes from  $J_\ell^*$  to  $J_y^*$  can still be constructed iteratively according to Eq. 37 from a starting condition of the iteration  $J_{\ell-1}^*$  (and  $J_\ell^*$ , which sets system size). Analogously, we can pick another starting condition  $J_N^* \equiv J_{\tilde{\ell}}^*$  for the N-branch of the cell. This starting condition reflects an additional degree of freedom in space of growth-maximizing metabolic states solution: next to system size (associated with  $J_\ell$ ) and nutrient environment (associated with  $J_{\ell-1}$ ), which were both also present in the linear chain, an optimal state is now also dependent on a second nutrient concentration in the environment (associated with  $J_N$ ). For any  $J_N$ , one can solve  $\partial M_P / \partial J_y = 0$  for  $J_C$  analytically by a suitable computer algebra system.

For all other criteria on optimality, Eq. 71 behaves like a regular Michaelis-Menten kinetic rate law with

$$\kappa_y^{\max} \equiv \frac{\kappa_{CN}^{\max}}{1 + \frac{K^N J_\ell}{J_N - (1 - \gamma) J_y}}, \quad \text{and} \quad K_y \equiv \frac{\kappa_y^{\max}}{\kappa_{CN}^{\max}} (K^C + \frac{K^{CN} J_\ell}{J_N - (1 - \gamma) J_y}). \quad (72)$$

Thus, the iterative construction of the optimal metabolic state can proceed until  $J_1^*$ , where a matching environmental nutrient concentration must be determined. In comparison to the solution of a linear chain with  $\ell$  enzymes, all fluxes  $J_{i < y}$  are scaled by  $\gamma$ . Analogously, all fluxes in the N-branch can be constructed, once  $J_C^*$  is known by viewing the N-branch and the stem together as a linear chain. In this manner, a consistent solution is always found, because the same equation  $\partial M_P / \partial J_y = 0$  is used to determine first  $J_C$  from a given  $J_N$  and then  $J_N$  again from the determined  $J_C$ . Due to the irreversible nature of reactions considered, matching environmental nutrient concentrations  $\rho_1$  and  $\rho_{\tilde{1}}$  can be determined (even for constrained expression levels  $\phi_1$  and/or  $\phi_{\tilde{1}}$ ), analogously to the linear chain. To parameterize the branched network, we assume that the substrate mass  $M_S$  on each branch is given by  $\gamma$  times the substrate mass present in the stem of the network. While  $M_E$ ,  $\nu$ , and the distributions of  $k_{\text{cat}}$  and  $\tilde{K}$  are the same as in the stem, thus both  $K$  and  $\kappa^{\max}$  are on average reduced by a factor of  $\gamma$  in comparison to their values in the stem.

In the nutrient-rich regime, optimal resource allocation in two convergent pathways qualitatively resembles optimal resource allocation in a single chain. Scaled metabolite concentration gradients in each separate pathway are qualitatively unaffected by the altered topology. They exhibit a strong decrease over the first few reactions after nutrient uptake and a weaker decline after. At the convergence enzyme itself, the cost of metabolite production suddenly increases - requiring only one production pathway beforehand, but both in the shared stem. Thus, saturation drops at the convergence.

If one of the two nutrients is available at a growth-limiting concentration, the proteome reallocation within the branch with limited nutrient influx and in the stem resembles the proteome reallocation in a single chain: high- $q$  enzymes are up-regulated while low- $q$  enzymes are down-regulated (Fig. 4D, F). On the other hand, proteome fractions in the unaffected branch are reduced with decreasing growth rate, leading to an overall resource reallocation from the unaffected branch and low- $q$  enzymes in the affected branch and stem to high- $q$  enzymes in the affected branch and stem.

## Regulatory constraints

**Regulation of the uptake enzyme.** In a simple regulatory model, a metabolite can directly act as a transcription factor to influence protein expression. Here we assume that such a regulatory framework leads to a Hill-function governing an

enzyme's expression level as a function of its substrate's concentration, with a basal expression level  $\phi^0$  and an expression level  $\phi^\infty$  at full promoter saturation with the metabolite. At a concentration  $K^{\text{reg}}$ , the promoter is half-saturated,

$$\bar{\phi}_i(\rho_i) = \phi^0 + \frac{a_i}{1 + K_i^{\text{reg}}/\rho_i}, \quad \text{with } a_i = \phi_i^\infty - \phi_i^0. \quad (73)$$

In this model, the metabolite can either induce protein production ( $\phi^0 < \phi^\infty$ ), or inhibit protein production ( $\phi^0 > \phi^\infty$ ). We focus first on the nutrient uptake reaction ( $i = 1$ ). Consider a cell that efficiently converts environmental nutrients into new protein biomass ( $J_1 \approx J_\ell(1 + \mathcal{O}(\epsilon))$ ) and that exhibits a small uptake protein mass fraction  $\phi_1 \ll 1$  in a nutrient-rich environment. The flux that can be efficiently handled by downstream metabolism ( $i = 2, \dots, \ell$ ) is proportional to  $1 - \phi_1$ . For sufficiently small  $\phi_1$  and only moderate reductions in nutrient availability, keeping the nutrient uptake rate approximately constant therefore recapitulates the unconstrained optimal resource allocation strategy,

$$\phi_1^*(\rho_1) \approx (1 + \frac{K_1}{\rho_1})\phi_1^\infty, \quad \text{with } \phi_1^\infty \equiv \lim_{\rho_1 \rightarrow \infty} \phi_1^* \quad \text{for } \left(\frac{\rho_1}{K_1} > 1\right). \quad (74)$$

In the case of the nutrient uptake reaction, we thus expect that the proteome fraction of the uptake enzyme should increase with decreasing nutrient availability (Fig. S3A). For the reasons outlined in the main text, we assume that the approximate form 74 is only valid at high growth rates. Instead, expression levels off to  $\phi^0 = \phi_1^{\max}$  well before the uptake protein becomes the dominant proteome mass fraction in the cell. To achieve qualitatively comparable up-regulation of all parameterizations in the ensemble despite different optimal expression levels in a nutrient-rich environment, we use  $\phi^0 = \phi_1^{\max} \equiv n\phi_1^\infty/(1 + n\phi_1^\infty)$ , where  $n$  determines the fold-change of  $\phi_1$  under nutrient limitation. In this manner,  $\phi_1$  can be up-regulated  $n$ -fold for parameterizations with small  $\phi^\infty$  but does not exceed 1 for larger  $\phi^\infty$ . In all simulations, we use  $n = 5$ . To mimic optimal resource allocation 74,  $K^{\text{reg}}$  must then be  $K^{\text{reg}} = K_1\phi_1^\infty/\phi_1^{\max}$ . This regulatory scheme approximates unconstrained optimal resource allocation to the uptake protein at high nutrient concentrations  $\rho_1/K_1 \gg 1$ , and smoothly interpolates to a cellular state with a constrained nutrient uptake flux at lower nutrient concentrations (Fig. S3).

**Approximating optimal resource allocation by regulation.** We define an extended regulatory scheme for system-wide metabolism. Transcription takes place with enzyme-specific rates

$$b_i^t(\rho) = b_0^t \left( \phi^0 + \frac{a_i}{1 + K_i^{\text{reg}}/\rho_i} \right), \quad (i = 2, \dots, \ell), \quad (75)$$

where  $b_0^t$  is a basal transcription rate. We use metabolite-mediated regulation as introduced in the main text, which is specified by the set of parameters  $\mathcal{R}_m(\kappa^{\max}, K) = (\phi_0, a, K^{\text{reg}})(\kappa^{\max}, K)$  with

$$\phi_i^0 = \frac{\frac{K_i}{\kappa_i^{\max}}}{\sqrt{\frac{K_i}{\kappa_i^{\max}} \sum_{i=2}^{\ell} \sqrt{\frac{K_i}{\kappa_i^{\max}}}}}, \quad a_i = \frac{\frac{1}{\kappa_i^{\max}}}{\sum_{i=2}^{\ell} \frac{1}{\kappa_i^{\max}}} - \phi_i^0, \quad K_i^{\text{reg}} = \frac{\sqrt{\frac{K_i}{\kappa_i^{\max}}} \sum_{i=2}^{\ell} \sqrt{\frac{K_i}{\kappa_i^{\max}}}}{\sum_{i=2}^{\ell} \frac{1}{\kappa_i^{\max}}}. \quad (76)$$

As a second step, we use a minimal model of translation that specifies protein synthesis rates proportional to the corresponding transcription rates,

$$b_i(\rho, \lambda) = \frac{\lambda b_i^t(\rho)}{\sum_{k=1}^{\ell} b_k^t(\rho)}. \quad (77)$$

This form sets the overall normalization of protein synthesis,  $\sum_k b_i(\rho, \lambda) = \lambda$ , as required in a steady state with dilution at rate  $\lambda$ , and leads to stationary expression levels  $\bar{\phi}_i(\rho)$  given by equation 24. The relation between transcription and translation rates reflects competition for a growth-dependent global resource, e.g., RNA polymerases, translating ribosomes, or energy.

The steady-state protein expression levels given by the metabolite-mediated regulatory scheme 75–77 reproduce growth-optimal metabolic states in very good approximation. First, we obtain the consistency relation

$$\bar{\phi}_i(\rho_i^*(\lambda), \mathcal{R}_m) \approx \phi_i^*(\lambda) \quad (i = 2, \dots, \ell), \quad (78)$$

where  $\rho_i^*(\lambda)$  and  $\phi_i^*(\lambda)$  are given by equations 19–21. Second, we numerically study time-dependent metabolic states  $\mathcal{S}(t) = (\rho(t), \phi(t))$  given by the dynamical equations

$$\frac{d\rho_i(t)}{dt} = \frac{\kappa_{i-1}^{\max} \phi_{i-1}(t)}{1 + \frac{K_{i-1}}{\rho_{i-1}(t)}} - \frac{\kappa_i^{\max} \phi_i(t)}{1 + \frac{K_i}{\rho_i(t)}} - \lambda(t) \rho_i(t), \quad (i = 2, \dots, \ell), \quad (79)$$

$$\frac{d\phi_i(t)}{dt} = b_i(\rho_i(t), \lambda(t)) - \lambda(t) \phi_i(t) \quad (i = 1, \dots, \ell) \quad (80)$$

with  $\lambda(t) = \kappa_\ell^{\max} \phi_\ell(t) / (1 + K_\ell / \rho_\ell(t))$  and a time-dependent input signal  $\rho_1(t)$  (Fig. 6). We find stable stationary states  $\mathcal{S}(\rho_1, \mathcal{R}_m)$  with

$$\begin{aligned} \bar{\rho}_i(\rho_1, \mathcal{R}_m) &= \rho_i^*(\rho_1)(1 \pm 0.5), \\ \bar{\phi}_i(\rho_1, \mathcal{R}_m) &= \phi_i^*(\rho_1)(1 \pm 0.05), \\ \bar{\lambda}(\rho_1, \mathcal{R}_m) &= \lambda^*(\rho_1)(0.98 \pm 0.02). \end{aligned} \quad (81)$$

We note that no flux-sensing or growth-sensing mechanism is required to robustly implement growth-dependent gene expression.

## Network evolution

In this subsection\*, we estimate selection on enzyme-metabolite binding in networks with a regulated physiological response specified by a regulation function depending on the reaction parameters,  $\mathcal{R}(\kappa^{\max}, K)$ . We evaluate the selection coefficients, or relative growth effects in the stationary state, of amino acid changes that affect the metabolite binding affinity,  $K_i \rightarrow K_i + \Delta K_i$  ( $i \neq 1$ ),

$$\begin{aligned} \sigma_i(\rho_1, \mathcal{R}, \kappa^{\max}, K) &\equiv \frac{\Delta \bar{\lambda}(\rho_1, \mathcal{R}, \kappa^{\max}, K)}{\bar{\lambda}(\rho_1, \mathcal{R}, \kappa^{\max}, K)} \\ &= \frac{K_i}{\bar{\lambda}(\rho_1, \mathcal{R}, \kappa^{\max}, K)} \frac{\partial \bar{\lambda}(\rho_1, \mathcal{R}, \kappa^{\max}, K)}{\partial K_i} \Delta \Delta G_i, \end{aligned} \quad (82)$$

where  $\Delta \Delta G_i = \Delta K_i / K_i$  is the corresponding change in reduced free energy of binding. We assume uniform effects  $|\Delta \Delta G_i| = 1$ , in line with typical amplitudes of mutations affecting protein evolution. In equation 82, we compute only the direct effect of the mutation on binding, while keeping the regulation function  $\mathcal{R}(\kappa^{\max}, K)$  at the wild-type parameters. We assume that regulation tunes the wild-type metabolic state  $\bar{\mathcal{S}}(\rho_1, \mathcal{R}, \kappa^{\max}, K)$  to near-optimal growth,

$$\bar{\lambda}(\rho_1, \mathcal{R}, \kappa^{\max}, K) \approx \lambda^*(\rho_1, \kappa^{\max}, K); \quad (83)$$

an example of such regulation functions is the metabolite-mediated scheme  $\mathcal{R}_m$  discussed above. This assumption is self-consistent if the evolution of protein binding affinities is followed by compensatory evolution of regulation,  $\mathcal{R}(\kappa^{\max}, K) \rightarrow \mathcal{R}(\kappa^{\max}, K + \Delta K)$ . However, details of the regulation function and its evolution are not relevant for our current purpose, to obtain an order-of-magnitude estimate of selection on protein binding affinities.

First, we compute an upper bound for the selection coefficients  $\sigma_i(\rho_1, \mathcal{R}, \kappa^{\max}, K)$  from a hypothetical system where the mutant is constrained to the same protein expression levels as the wild-type. For this system, we obtain

$$\begin{aligned} \sigma_i^{(1)}(\rho_1, \kappa^{\max}, K) &= -\frac{K_i}{\lambda^*(\rho_1, \kappa^{\max}, K)} \frac{\partial \lambda^*(\rho_1, \kappa^{\max}, K)}{\partial K_i} \Big|_{\phi} \\ &= -\frac{K_i}{j_i(\kappa_i^{\max}, K_i)} \frac{\partial j_i(\kappa_i^{\max}, K_i)}{\partial K_i} \Big|_{j_{i-1}, \phi_i} \\ &= \rho_i(\rho_1, \kappa^{\max}, K) [1 + O(\epsilon)] \end{aligned} \quad (84)$$

which follows from the Michaelis-Menten relation [1], the flux continuity relation [3] and  $\rho_i^2/K_i \lesssim 1$  [11]. We have used  $\Delta\Delta G_i = -1$  for an affinity-increasing mutation and the relation [83] for the wild-type. We can compare this strength of selection with the case of a single enzyme at constant metabolite density,

$$\begin{aligned}\sigma_0(\rho, \kappa^{\max}, K) &= \frac{K}{j(\rho, \kappa^{\max}, K)} \frac{\partial j(\rho, \kappa^{\max}, K)}{\partial K} \\ &= \frac{1}{1 + \rho/K}.\end{aligned}\quad (85)$$

This shows that in a network, moderate affinity changes of bulk enzymes are largely buffered by compensatory changes of the cognate metabolite,  $\Delta\rho_i/\rho_i \approx -\Delta K_i/K_i$ , leaving a residual effect on growth by metabolite dilution. Furthermore, the effects on downstream metabolites, which are omitted in [84], are of order  $\epsilon^2$ .

Second, we compute a lower bound on  $\sigma_i(\rho_1, \mathcal{R}, \kappa^{\max}, K)$  from the variation of optimal growth rates. This approximation couples local enzyme changes  $K_i \rightarrow K_i + \Delta K_i$  and compensatory changes of regulation,  $\mathcal{R}(\kappa_i^{\max}, K_i) \rightarrow \mathcal{R}(\kappa^{\max}, K_i + \Delta K_i)$ , assuming a hypothetical simultaneous evolution of an enzyme's protein and cis-regulatory sequence, which overestimates the mutant fitness. We obtain selection coefficients

$$\sigma_i^{(2)}(\rho_1, \kappa^{\max}, K) = -\frac{K_i}{\lambda^*(\rho_1, \kappa^{\max}, K)} \frac{\partial \lambda^*(\rho_1, \kappa^{\max}, K)}{\partial K_i} \quad (86)$$

$$\approx K_i \frac{\frac{1}{2} \sqrt{\frac{j_i}{K_i \kappa_i^{\max} \beta_i}}}{1 + \sum_{k=2}^{\ell} \sqrt{\frac{K_k j_k}{\kappa_k^{\max} \beta_k}}} \quad (87)$$

$$= \frac{1}{2} \rho_i(\rho_1, \kappa^{\max}, K) [1 + O(\epsilon)]. \quad (88)$$

Here we have used that  $j_1$  is independent of  $K_i$  (which is justified under nutrient limitation, see equation [73], and correct to order  $\epsilon$  otherwise), and neglected position-dependent inhomogeneities and changes of the cost factors (i.e.,  $\beta_i \approx \beta_0 \approx \text{const.}$ , equation [21]), as appropriate under nutrient limitation. In the same rationale, variation of the growth-optimal metabolic state  $\mathcal{S}^*(\rho_1, \kappa^{\max}, K)$  provides estimates of the buffering in regulated networks,

$$\frac{\Delta\rho_i}{\rho_i} \approx \frac{1}{2} \frac{\Delta K_i}{K_i}, \quad (89)$$

$$\frac{\Delta\phi_i}{\phi_i} \approx \frac{1}{2} \left( 1 - \frac{\lambda}{\lambda^{\max}} \frac{\phi_i^{\max}}{\phi_i} \right) \frac{\Delta K_i}{K_i}, \quad (90)$$

with  $\phi_i^{\max} = \lambda^{\max}/\kappa_i^{\max}$ , which are obtained from equations [19] and [20].

The approximations [84] and [88] of the selection coefficients  $\sigma_i(\rho_1, \mathcal{R}, \kappa^{\max}, K)$  are of the same form and magnitude, which establishes equation [25] and the resulting mean field-estimates of  $\sigma_i$  derived in the main text. Furthermore, inserting the metabolite densities given by equation [13] or [19] shows that individual enzymes evolve target binding on exponential fitness landscapes,

$$f(\Delta G_i) \sim -C_i \exp\left(\frac{\Delta G_i}{2}\right), \quad (91)$$

where  $\Delta G_i = \log(K_i/K_0)$  is the reduced free energy of binding,  $f(\Delta G_i) = \lambda(\Delta G) - \lambda(\Delta G_{i,\min})$  is the growth difference to the protein state of strongest binding, and  $C_i$  is a constant. This form is similar to the biophysical fitness landscapes of viable proteins discussed in ref. [58].

## Numerical simulations

Numerical computations were performed using Python. Equations [6], [3] and [1] were used to construct optimal balanced growth states of stochastically parameterized metabolic models from iteration starting conditions  $(J_{\ell-1}, J_\ell)$ . Iteration starting conditions were varied numerically using the Broyden–Fletcher–Goldfarb–Shanno algorithm to match a nutrient-rich

environment (Fig. 1 – 4, S2 – S4), or match an otherwise defined environmental nutrient concentration (Fig. 3, orange points). Subsequently, the ratio  $J_{\ell-1}/J_\ell$  was reduced to construct optimal balanced growth states under nutrient limitation. In case of the branched metabolic network, also the metabolite concentration  $\rho_N$  was varied numerically for each ratio  $J_{\ell-1}/J_\ell$  using the same algorithm to ensure a constant environmental nutrient concentration  $\rho_1^N$  under nutrient limitation of the  $C$  branch. Time evolutions of the feed-forward regulated metabolic cascades (Fig. 6, Eqs. 79 and 80) were computed by means of Euler's method. The code required to construct optimal balanced growth states and all figures displayed in the manuscript is available online in the following repository: <https://github.com/laseeg/ResourceAllocation>.

## Cytoplasmic density affects optimal resource allocation

Throughout the whole paper, we assume that the protein mass density of the cell is approximately  $400 \text{ g L}^{-1}$ , independent of environmental conditions. This assumption is in line with experimental observations (1, 26, 28), which indicate an approximately constant cellular dry mass density under nutrient limitation. However, literature estimates of the cellular protein mass density vary, and we can not exclude that the composition of the cell varies under nutrient limitation which may lead to a modified biosynthetic protein density at constant dry mass density, or exclude that cellular dry mass density changes within the margins of experimental accuracy. A recent study has proposed potential underlying mechanisms to a constant intracellular protein density based on a model of cellular self-replication similar to the one we present here (67). Simultaneous optimization of protein density and proteome allocation under a nutrient limitation may be feasible but lies beyond the scope of this paper.

Generally, we predict that cells exhibit higher optimal metabolite mass fractions and lowered mean enzyme saturation at lowered protein mass density in the nutrient-rich regime. The metabolite mass fraction changes in proportion to the square-root fold change in protein mass density. The same behavior is expected under a nutrient uptake constraint, which concerns  $\phi_1$ . Thus, changes in resource allocation induced by changes in protein mass density are likely negligible in comparison to the predicted changes due to differences in biochemical quality  $q$  at constant protein mass density. Only a constraint on  $j_1$  (but unconstrained  $\phi_1$  relative to all other metabolic proteins in the cell) qualitatively yields a different behavior. In this case, both the cellular metabolite mass fraction and mean enzyme saturation would increase in the nutrient-poor regime with decreasing biosynthetic protein mass density.

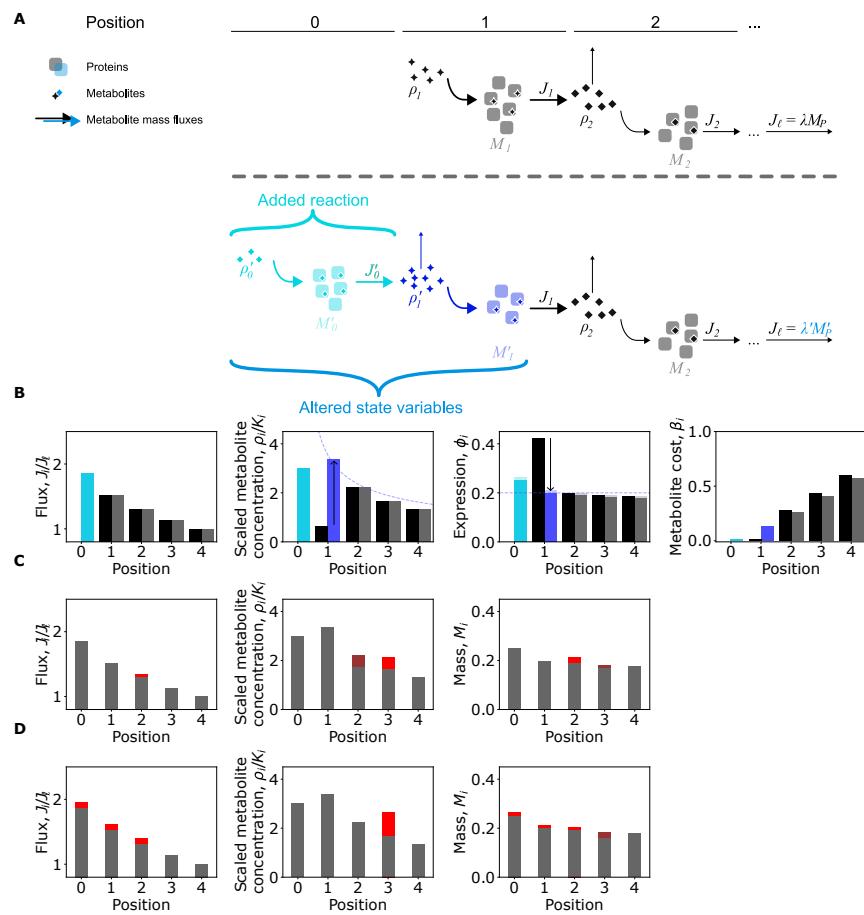
## References

- [1] Hans Bremer and Patrick P Dennis. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3, 9 2008.
- [2] Suckjoon Jun, Fangwei Si, Rami Pugatch, and Matthew Scott. Fundamental principles in bacterial physiology-history, recent progress, and the future with focus on cell size control: A review. *Reports on Progress in Physics*, 81, 2 2018.
- [3] Hans V Westerhoff and Bernhard O Palsson. The evolution of molecular biology into systems biology. *Nature Biotechnology*, 22:1249–1252, 10 2004.
- [4] Karthik Raman and Nagasuma Chandra. Flux balance analysis of biological systems: Applications and challenges. *Briefings in Bioinformatics*, 10:435–449, 2009.
- [5] Douwe Molenaar, Rogier Van Berlo, Dick De Ridder, and Bas Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology*, 5, 1 2009.
- [6] Sushovan Dam, Jean Marie Pagès, and Muriel Masi. Stress responses, outer membrane permeability control and antimicrobial resistance in enterobacteriaceae. *Microbiology (United Kingdom)*, 164:260–267, 3 2018.
- [7] Aarash Bordbar, Jonathan M. Monk, Zachary A. King, and Bernhard O. Palsson. Constraint-based models predict metabolic and associated cellular functions, 2 2014.
- [8] Hugo Dourado and Martin J. Lercher. An analytical theory of balanced cellular growth. *Nature Communications*, 11, 12 2020.
- [9] Hao Wang, Jonathan L. Robinson, Pinar Kocabas, Johan Gustafsson, Mihail Anton, Pierre-Etienne Cholley, Shan Huang, Johan Gobom, Thomas Svensson, Mattias Uhlen, Henrik Zetterberg, and Jens Nielsen. Genome-scale metabolic network reconstruction of model animals as a platform for translational research. *Proceedings of the National Academy of Sciences*, 118, 7 2021.
- [10] Anurag Passi, Juan D. Tibocha-Bonilla, Manish Kumar, Diego Tec-Campos, Karsten Zengler, and Cristal Zuniga. Genome-scale metabolic modeling enables in-depth understanding of big data, 1 2022.
- [11] Avlant Nilsson, Jens Nielsen, and Bernhard O. Palsson. Metabolic models of protein allocation call for the kinetome. *Cell Systems*, 5:538–541, 12 2017.
- [12] Changdai Gu, Gi Bae Kim, Won Jun Kim, Hyun Uk Kim, and Sang Yup Lee. Current status and applications of genome-scale metabolic models. *Genome Biology*, 20, 6 2019.
- [13] David B. Bernstein, Snorre Sulheim, Eivind Almaas, and Daniel Segrè. Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. *Genome Biology*, 22, 12 2021.
- [14] Jeffrey D. Orth, Ines Thiele, and Bernhard O. Palsson. What is flux balance analysis? *Nature Biotechnology*, 28: 245–248, 3 2010.
- [15] Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330:1099–1102, 11 2010.
- [16] Hugo Dourado, Matteo Mori, Terence Hwa, and Martin J. Lercher. On the optimality of the enzyme–substrate relationship in bacteria. *PLoS Biology*, 19, 10 2021.
- [17] Philip Greulich, Matthew Scott, Martin R Evans, and Rosalind J Allen. Growth-dependent bacterial susceptibility to ribosome-targeting antibiotics. *Molecular Systems Biology*, 11:796, 3 2015.
- [18] Fernanda Pinheiro, Omar Warsi, Dan I. Andersson, and Michael Lässig. Metabolic fitness landscapes predict the evolution of antibiotic resistance. *Nature Ecology and Evolution*, 5:677–687, 5 2021.

- [19] João V. Rodrigues, Shimon Bershtein, Annabel Li, Elena R. Lozovsky, Daniel L. Hartl, and Eugene I. Shakhnovich. Biophysical principles predict fitness landscapes of drug resistance. *Proceedings of the National Academy of Sciences of the United States of America*, 113:E1470–E1478, 3 2016.
- [20] S Andreas Angermayr, Tin Yau Pang, Guillaume Chevereau, Karin Mitosch, Martin J Lercher, and Tobias Bollenbach. Growth-mediated negative feedback shapes quantitative antibiotic response. *Molecular Systems Biology*, 18, 9 2022.
- [21] Mohammadreza Yasemi and Mario Jolicœur. Modelling cell metabolism: A review on constraint-based steady-state and kinetic approaches, 2 2021.
- [22] Hugo Dourado, Wolfram Liebermeister, Oliver Ebenhöh, and Martin J. Lercher. Mathematical properties of optimal fluxes in cellular reaction networks at balanced growth. *PLoS Computational Biology*, 19, 6 2023.
- [23] Alexander Schmidt, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer, Luciano Callipo, Kévin Knoops, Manuel Bauer, Ruedi Aebersold, and Matthias Heinemann. The quantitative and condition-dependent escherichia coli proteome. *Nature Biotechnology*, 34:104–110, 1 2016.
- [24] Matteo Mori, Zhongge Zhang, Amir Banaei-Esfahani, Jean-Benoît Lalanne, Hiroyuki Okano, Ben C Collins, Alexander Schmidt, Olga T Schubert, Deok-Sun Lee, Gene-Wei Li, Ruedi Aebersold, Terence Hwa, and Christina Ludwig. From coarse to fine: the absolute escherichia coli proteome under diverse growth conditions. *Molecular Systems Biology*, 17, 5 2021.
- [25] Jianye Xia, Benjamin J. Sánchez, Yu Chen, Kate Campbell, Sergo Kasvandik, and Jens Nielsen. Proteome allocations change linearly with the specific growth rate of saccharomyces cerevisiae under glucose limitation. *Nature Communications*, 13, 12 2022.
- [26] Enno R Oldewurtel, Yuki Kitahara, and Sven Van Teeffelen. Robust surface-to-mass coupling and turgor-dependent cell width determine bacterial dry-mass density. *BIOPHYSICS AND COMPUTATIONAL BIOLOGY*, 118, 2021.
- [27] Herbert E Kubitschek, William W Baldwin, Sally J Schroeter, and Rheinhard Graetzer. Independence of buoyant cell density and growth rate in escherichia coli. *JOURNAL OF BACTERIOLOGY*, 158:296–299, 1984.
- [28] Markus Basan, Manlu Zhu, Xiongfeng Dai, Mya Warren, Daniel Sévin, Yi-Ping Wang, and Terence Hwa. Inflating bacterial cells by increased protein synthesis. *Molecular Systems Biology*, 11:836, 10 2015.
- [29] Itzhak Fishov, Arieh Zaritsky, and N. B. Grover. On microbial states of growth. *Molecular Microbiology*, 15:789–794, 3 1995.
- [30] M Ehrenberg and C G Kurland. Costs of accuracy determined by a maximal growth rate constraint. *Quarterly reviews of biophysics*, 17:45–82, 2 1984.
- [31] Tomer Benyamin, Ori Folger, Eytan Ruppin, and Tomer Shlomi. Flux balance analysis accounting for metabolite dilution. *Genome Biology*, 11, 4 2010.
- [32] Daan H. de Groot, Josephus Hulshof, Bas Teusink, Frank J. Bruggeman, and Robert Planqué. Elementary growth modes provide a molecular description of cellular self-fabrication. *PLoS Computational Biology*, 16, 2020.
- [33] Griffin Chure and Jonas Cremer. An optimal regulation of fluxes dictates microbial growth in and out of steady-state. *eLife*, 12, 3 2023.
- [34] Frank J. Bruggeman, Robert Planqué, Douwe Molenaar, and Bas Teusink. Searching for principles of microbial physiology. *FEMS Microbiology Reviews*, 44:821–844, 11 2020.
- [35] Benjamin D. Towbin, Yael Korem, Anat Bren, Shany Doron, Rotem Sorek, and Uri Alon. Optimality and sub-optimality in a bacterial growth law. *Nature Communications*, 8, 1 2017.

- [36] 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 Computational Biology*, 12, 11 2016.
- [37] Arren Bar-Even, Elad Noor, Yonatan Savir, Wolfram Liebermeister, Dan Davidi, Dan S. Tawfik, and Ron Milo. The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, 50: 4402–4410, 5 2011.
- [38] Bryson D. Bennett, Elizabeth H. Kimball, Melissa Gao, Robin Osterhout, Stephen J. Van Dien, and Joshua D. Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli. *Nature Chemical Biology*, 5:593–599, 2009.
- [39] Ron Milo and Rob Phillips. *Cell Biology by the Numbers*. Garland Science, 12 2015. ISBN 9781317230694.
- [40] Arthur L Koch and Houston Wang. How close to the theoretical diffusion limit do bacterial uptake systems function? *Arch Microbiol*, 131:36–42, 1982.
- [41] Karl Peebo, Kaspar Valgepea, Andres Maser, Ranno Nahku, Kaarel Adamberg, and Raivo Vilu. Proteome reallocation in escherichia coli with increasing specific growth rate. *Molecular BioSystems*, 11:1184–1193, 4 2015.
- [42] Sheng Hui, Josh M Silverman, Stephen S Chen, David W Erickson, Markus Basan, Jilong Wang, Terence Hwa, and James R Williamson. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular Systems Biology*, 11:784, 2 2015.
- [43] Huaiyu Mi and Paul Thomas. Panther pathway: an ontology-based pathway database coupled with data analysis tools. *Methods in molecular biology (Clifton, N.J.)*, 563:123–140, 2009.
- [44] Matthew Scott, Stefan Klumpp, Eduard M Mateescu, and Terence Hwa. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Molecular Systems Biology*, 10:747, 8 2014.
- [45] Matthew Scott and Terence Hwa. Shaping bacterial gene expression by physiological and proteome allocation constraints. *Nature Reviews Microbiology*, 21:327–342, 11 2022.
- [46] Xiao-Pan Hu, Stefan Schroeder, and Martin J. Lercher. Proteome efficiency of metabolic pathways in escherichia coli increases along the nutrient flow. *mSystems*, 8, 10 2023.
- [47] Jes Forchhammer and Lasse Lindahl. Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of escherichia coli 15. *Journal of Molecular Biology*, 55:563–568, 2 1971.
- [48] Michael Yu. Pavlov and Måns Ehrenberg. Rate of translation of natural mrnas in an optimized in vitro system. *Archives of Biochemistry and Biophysics*, 328:9–16, 4 1996.
- [49] Robert Planqué, Josephus Hulshof, Bas Teusink, Johannes C. Hendriks, and Frank J. Bruggeman. Maintaining maximal metabolic flux by gene expression control. *PLoS Computational Biology*, 14, 9 2018.
- [50] Motoo Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, 10 1983. ISBN 9780521231091.
- [51] Ulrich Gerland and Terence Hwa. On the selection and evolution of regulatory dna motifs. *Journal of Molecular Evolution*, 55:386–400, 10 2002.
- [52] Johannes Berg, Stana Willmann, and Michael Lässig. Adaptive evolution of transcription factor binding sites. *BMC Evolutionary Biology*, 4, 10 2004.
- [53] Konstantin B. Zeldovich, Peiqiu Chen, and Eugene I. Shakhnovich. Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proceedings of the National Academy of Sciences*, 104:16152–16157, 10 2007.
- [54] Michael Lässig. From biophysics to evolutionary genetics: Statistical aspects of gene regulation, 9 2007.

- [55] Ville Mustonen, Justin Kinney, Curtis G. Callan, and Michael Lässig. Energy-dependent fitness: A quantitative model for the evolution of yeast transcription factor binding sites. *Proceedings of the National Academy of Sciences*, 105: 12376–12381, 8 2008.
- [56] Michael Manhart and Alexandre V. Morozov. Protein folding and binding can emerge as evolutionary spandrels through structural coupling. *Proceedings of the National Academy of Sciences of the United States of America*, 112:1797–1802, 2 2015.
- [57] Peter B. Chi and David A. Liberles. Selection on protein structure, interaction, and sequence, 7 2016.
- [58] Torsten Held, Daniel Klemmer, and Michael Lässig. Survival of the simplest in microbial evolution. *Nature Communications*, 10, 12 2019.
- [59] D L Hartl and C H Taubes. Towards a theory of evolutionary adaptation. *Genetica*, 102-103:525–33, 1998.
- [60] Armita Nourmohammad, Torsten Held, and Michael Lässig. Universality and predictability in molecular quantitative genetics, 12 2013.
- [61] Eyal Metzl-Raz, Moshe Kafri, Gilad Yaakov, Ilya Soifer, Yonat Gurvich, and Naama Barkai. Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *eLife*, 6, 8 2017.
- [62] Matteo Mori, Terence Hwa, Olivier C. Martin, Andrea De Martino, and Enzo Marinari. Constrained allocation flux balance analysis. *PLoS Computational Biology*, 12, 6 2016.
- [63] Viktor M Boer, Christopher A Crutchfield, Patrick H Bradley, David Botstein, and Joshua D Rabinowitz. Growth-limiting intracellular metabolites in yeast growing under diverse nutrient limitations. *Molecular Biology of the Cell*, 21, 2010.
- [64] Dušica Radoš, Stefano Donati, Martin Lempp, Johanna Rapp, and Hannes Link. Homeostasis of the biosynthetic e. coli metabolome. *iScience*, 25, 7 2022.
- [65] Kirill Sechkar, Harrison Steel, Giansimone Perrino, and Guy-Bart Stan. A coarse-grained bacterial cell model for resource-aware analysis and design of synthetic gene circuits. *Nature Communications*, 15:1981, 3 2024.
- [66] Adam C. Palmer, Remy Chait, and Roy Kishony. Nonoptimal gene expression creates latent potential for antibiotic resistance. *Molecular Biology and Evolution*, 35:2669–2684, 2018.
- [67] Tin Yau Pang and Martin J. Lercher. Optimal density of bacterial cells. *PLoS Computational Biology*, 19, 6 2023.



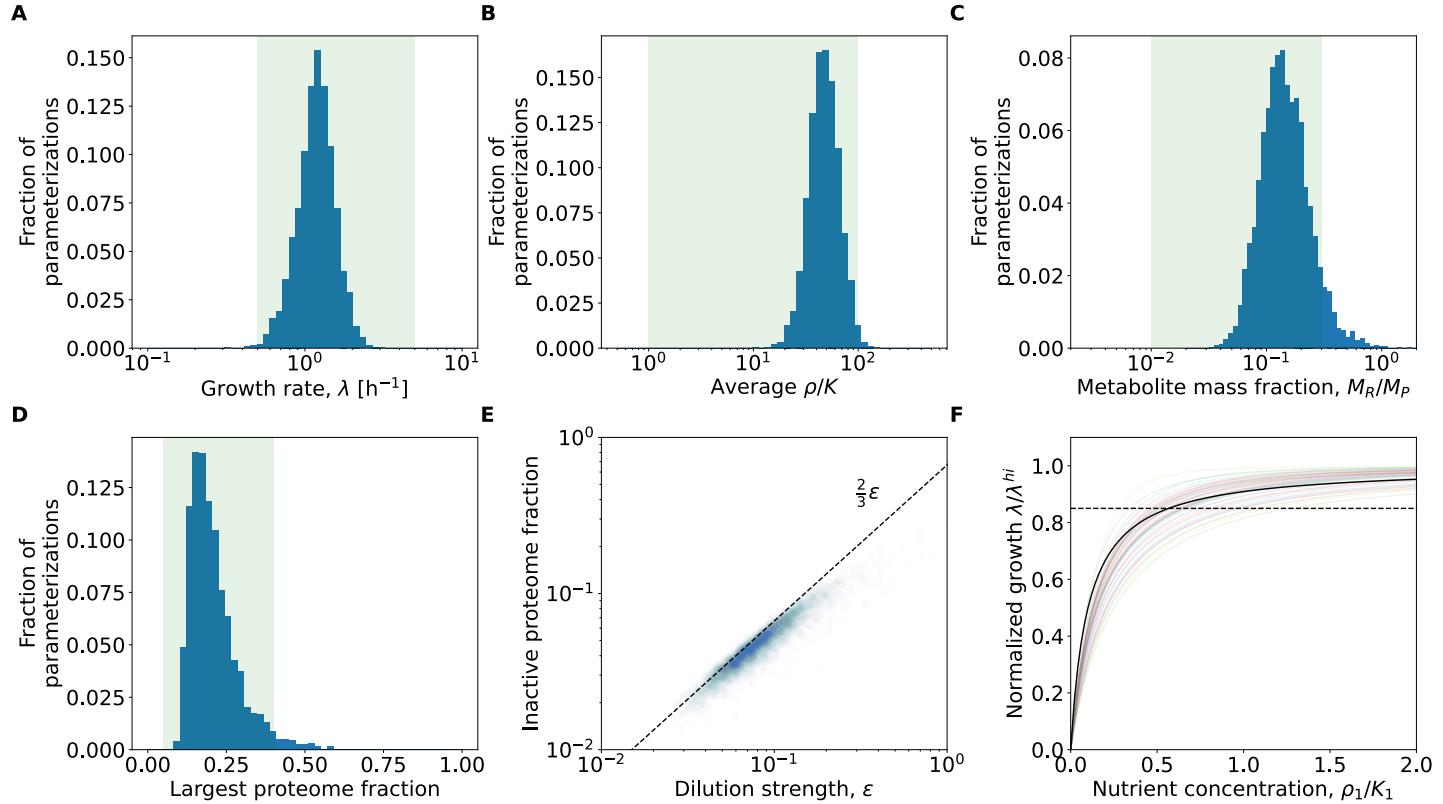
**Figure S1: Optimal metabolic states can be constructed iteratively.**

(A) For a two-enzyme chain, the optimal metabolic state in a given nutrient environment  $\rho_1$  can be determined analytically [33]. Here we construct optimal metabolic states of longer chains, by iteratively adding an enzyme to the chain that produces the previous environmental nutrient. Suppose all growth-optimal fluxes  $J$  of a chain of length  $\ell$  are known. We show (SI) that a new growth-optimal set of fluxes can be constructed by adding a single new flux  $J'_0$  to the chain and finding a suitable concentration of the new environmental nutrient  $\rho'_0$ . Such an extension of the chain defines a protein mass of the newly added enzyme  $M'_1$ , and affects the metabolite mass concentration of the previous environmental nutrient  $\rho'_1$ , as well as the required mass of the consuming enzyme  $M'_1$ , but does not affect further downstream metabolite mass concentrations or enzyme masses, making the problem analytically tractable.

(B) The optimal metabolic state of an exemplary short chain (left, black) and the optimal state of the corresponding elongated chain (right, shaded) are shown next to each other for comparison. For simplicity, we illustrate a cascade with equal substrate affinities  $K_i = 0.1$  and efficiencies  $\kappa_i^{\max} = 1 \text{ h}^{-1}$  for all enzymes. A system of five equations can be solved for the five modified or new state variables  $J'_0$ ,  $M'_0$ ,  $\rho'_0$ ,  $M'_1$  and  $\rho'_1$ . All other fluxes, saturations, and masses are conserved under this elongation. Instead of enzyme masses  $M_i$ , dimensionless enzyme mass fractions  $M_i/M_P$  are shown. Light gray bars represent unchanged enzyme masses before rescaling due to the change of  $M_P \rightarrow M'_P$ , caused by the added mass  $M'_0 + M'_1 - M_1$ . Arrows indicate changes in the metabolite mass concentration  $\rho'_1$  and enzyme mass  $M'_1$  caused by adding a metabolite-producing reaction with index  $i = 0$ . The resulting patterns in fluxes, saturation, and expression can be understood through the cost of metabolite production  $\beta_i$ . The cost reflects how many proteome resources are needed to produce a given metabolite.

For physiological chain parameters, the cost of metabolite production is well approximated by the proteome fraction upstream of the reaction of interest. It increases as a function of position. Thus, the optimal metabolite concentration decreases with position. The resulting gradient is well-approximated by  $(K_i)^{-1/2}$  (curved dashed line). At the same time, dilution of these metabolites induces gradients in mass fluxes. Both gradients in flux and enzyme saturation together result in a very weak gradient in enzyme mass fractions along the metabolic chain. Equal allocation to all five enzymes is shown as a horizontal dashed line for comparison.

(C), (D) Changes in metabolic state variables for the perturbation of a growth-optimal state in a single flux or metabolite concentration, as discussed in the main text. (C) Upon a variation that increases flux  $J_2$  in comparison to the growth-optimal state, the metabolite concentration  $\rho_3$  increases, whereas  $\rho_2$  decreases. At the same time, the enzyme masses required for these fluxes and metabolite concentrations in steady state, change for enzymes 2 and 3. In the optimal state, the total proteome mass must remain unchanged for a small variation in any single flux. (D) A concerted variation of all fluxes upstream of metabolite 3 leads to an isolated change in  $\rho_3$ . All enzyme masses upstream increase, whereas the mass  $M_3$  is reduced. In the optimal state, the total proteome mass must remain unchanged for a small variation in a single metabolite concentration. In all panels, increases are shown in bright red, whereas decreases are shown in dark red.



**Figure S2: Model cells exhibit physiological properties.**

An ensemble of 5000 linear chains of enzymes was parameterized by drawing individual enzyme affinities and efficiencies from log-normal distributions with physiological mean and variance informed by measurements of Bar-Even et al. (37). The environmental nutrient concentration was set to  $\rho_1 = 1000K_1$  to reflect a nutrient-rich environment.

(A) Cells parameterized in this manner exhibit physiological growth rates between  $0.5 \text{ h}^{-1}$  and  $5 \text{ h}^{-1}$ .

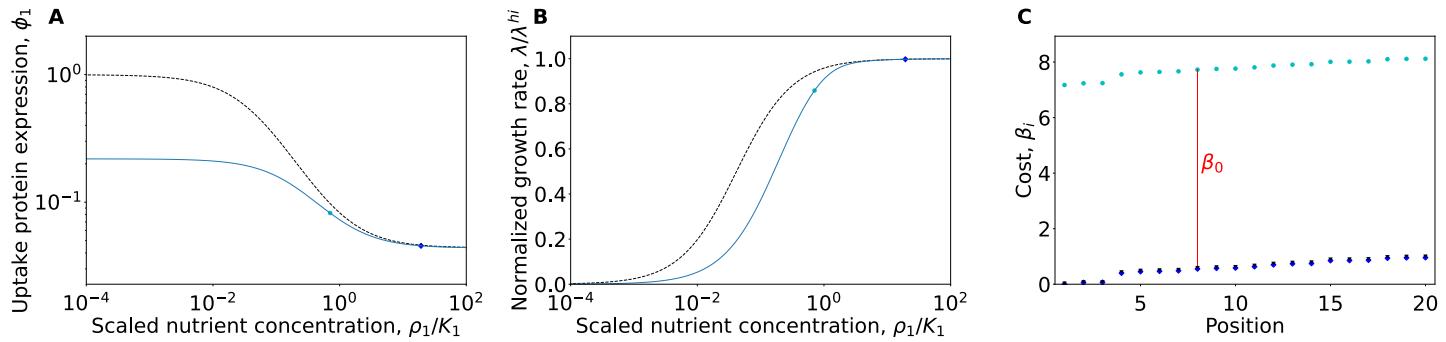
(B) The geometric mean of scaled intracellular metabolite concentrations ( $\rho_i/K_i$  for  $i = 2, \dots, \ell$ ) in physiologically parameterized model cells matches physiological estimates: enzymes are typically operating above their  $K$ , but not at exceedingly high substrate concentrations. Thus, values between 1 and 100 are deemed physiological.

(C) Cumulative metabolite mass fraction of all metabolite species in a cell. Typically, metabolites contribute on the order of 10 % of the proteome mass to the total dry mass in a cell (from BNID: 101436). A small fraction of physiologically parameterized cascades exhibit unphysiologically large metabolite mass fractions, exceeding the single percent range. We hypothesize that this result - as well as a relatively high average average saturation of enzymes in our model - may stem from neglecting additional growth costs associated with metabolites in a real cell.

(D) In a typical bacterial cell, the largest single proteome fraction is given by the ribosome and constitutes about 30 % of the cellular proteome. A small fraction of physiologically parameterized cascades exhibit unphysiologically large largest proteome mass fractions.

(E) Inactive proteome fraction predicted from the simplified resource allocation pattern 45 and obtained from optimal balanced growth states in physiologically parameterized cascades in the nutrient-rich regime.

(F) Predicted, representative nutrient dose-response functions for physiologically parameterized chains with a regulatory constraint on nutrient uptake limiting uptake overexpression to 5-fold (colored thin lines), are well approximated by a Monod-function with  $K_{\text{Monod}} = K_1/10$  (black solid line). The nutrient-rich regime is separated from the nutrient-poor regime by a growth reduction of  $\phi_1^{\max} + \epsilon$  (on average at the black dashed line), which typically corresponds to  $\rho_1 \approx K_1$ .



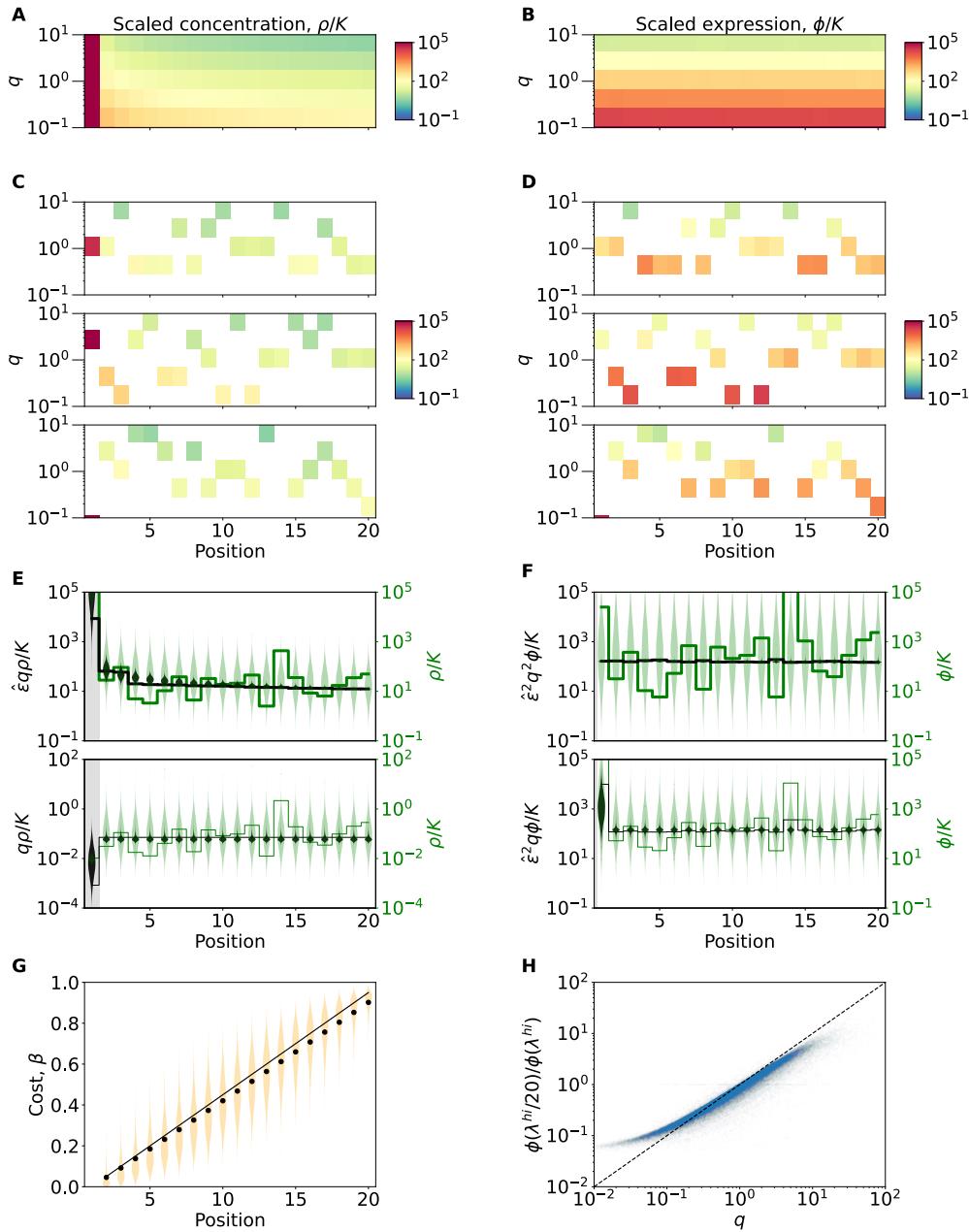
**Figure S3: A nutrient uptake constraint increases the costs of metabolite production.**

(A) Unconstrained optimal resource allocation leads to non-physiological overexpression of uptake proteins under nutrient limitation (black, dashed). This ultimately leads to a cell consisting solely of nutrient-uptake enzyme. In contrast, we constrain uptake protein expression (blue) at low nutrient concentrations. The constraint was chosen such that it mimics unconstrained optimal resource allocation at high growth rates and exhibits the same Hill-like shape (SI).

(B) Growth is reduced in a Monod-like fashion in response to nutrient limitation for both unconstrained (black) and constrained (blue) overexpression of the uptake protein, with a Monod-constant on the order of the nutrient affinity of the uptake protein. The Monod-like growth response results mainly from the Michaelis-Menten kinetics of the nutrient uptake reaction but is modulated by the uptake expression change, as well as the increasing overall efficiency of nutrient conversion under nutrient limitation.

(C) When uptake expression is not constrained, the cost of metabolite production  $\beta_i^*$  increases several-fold along the metabolic chain (dark blue). It is closely approximated by the cumulative upstream proteome fraction. When the nutrient uptake is reduced below its optimal level, all costs increase (cyan dots). The additional cost is the same for all enzymes ( $\beta_0^*$ ). While the cost of metabolite production still increases monotonously, relative changes are reduced. When accounting for differences in  $q$ , this monotonous increase in cost along the chain leads to a monotonous gradient in metabolite saturations along the chain. As relative gradients in cost flatten upon nutrient limitation, concentration gradients flatten. This leads to a position-dependent fold change in metabolite concentration upon nutrient limitation.

All panels illustrate a single, representative, physiological parameterization of a heterogeneous chain of length  $\ell = 20$ .



**Figure S4: Position and biochemical quality-based predictions are accurate for individual enzymes.**

(A), (B) Analytical maps of optimal scaled metabolite concentrations and expression levels. (C), (D) Corresponding maps of optimal scaled metabolite concentrations and expression levels for three individual model parameterizations. (E), (F) To quantify the residual variability, unexplained by position and quality, we rescale growth-optimal concentrations and expression levels according to our theory by  $q$ . The normalization of  $q$  ensures that the mean value of all  $q_i$  is one - this enables a direct comparison of unscaled patterns (green) and scaled patterns (black). Both are shown for a single parameterization of the model cell (shown in the third line of C, D, lines). Distributions of scaled metabolite concentrations and expression levels before and after rescaling by  $q$  and  $\epsilon$  in nutrient-rich and poor environments from the ensemble of physiologically parameterized chains are shown as violins. To conserve the mean of each distribution and allow for a direct comparison of both scaled and un-scaled variables  $\hat{\epsilon} \equiv \epsilon/\langle \epsilon \rangle$  was used for rescaling. Our theory reduces the squared residuals per enzyme of the log concentrations in the nutrient-rich regime by a factor 40, in the nutrient-poor regime by a factor of 6, and the squared residuals of the log expression levels by 50- and 5-fold respectively. (G) Distributions of metabolite production costs in an ensemble of physiologically parameterized chains as a function of enzyme position. The solid line corresponds to the estimate of  $\beta_i^* = (i-1)/\ell$ . Black dots denote the means of the shown distributions. (H) Correlation between enzyme quality and expression fold change upon severe nutrient limitation for all individual enzymes of the ensemble (excluding nutrient uptake enzymes).