An Analytical Theory of Cellular Growth*

Hugo Dourado and Martin J. Lercher†
Institute for Computer Science & Department of Biology, Heinrich Heine University, Düsseldorf, Germany

(Dated: April 12, 2019)

ABSTRACT

The biological fitness of non-interacting unicellular organisms in constant environments is given by their balanced growth rate, i.e., by the rate with which they replicate their biomass composition. Evolutionary optimization of this growth rate occurred under a set of physicochemical constraints, including mass conservation, reaction kinetics, and limits on dry mass per volume (cellular capacity). Mathematical models that account explicitly for these constraints are inevitably nonlinear, and their optimization has been restricted to small, non-realistic cell models. Here, we show that states of maximal balanced growth are elementary flux modes of a related flux balance problem, i.e., deactivating any active reaction makes steady-state growth impossible. For any balanced growth state that corresponds to an elementary flux mode of an arbitrarily sized model, we provide explicit expressions for individual protein concentrations, fluxes, and growth rate; all variables are uniquely determined by the concentrations of metabolites and total protein. We provide explicit and intuitively interpretable expressions for the marginal fitness costs and benefits of individual concentrations. At optimal balanced growth, the marginal net benefits of each metabolite concentration and of total protein concentration equal the marginal benefit of the cellular capacity. Based solely on physicochemical constraints, our work unveils fundamental quantitative principles of balanced cellular growth, quantifies the effect of cellular capacity on fitness, and leads to experimentally testable predictions.

I. INTRODUCTION

The defining feature of live is self-replication. For non-interacting unicellular organisms in constant environments, the rate of this self-replication is equivalent to their evolutionary fitness [1]: fast-growing cells outcompete those growing more slowly. Accordingly, natural selection has optimized the cellular composition of many microbes for maximal balanced growth rate in specific environments [2, 3], i.e., for the fastest possible reproduction of all cellular components in proportion to their abundances [4].

The variables of this evolutionary optimization are the abundances of the cellular metabolites and of the proteins and ribonucleic acids that catalyze their conversion into biomass. The corresponding boundary conditions are provided by the environment and by physicochemical constraints, including mass conservation, the kinetics of enzymatic and spontaneous reactions, and capacity constraints that limit cellular concentrations [3, 5–8].

Molenaar et al. [5] proposed a coarse-grained, mathematical model of balanced growth encapsulating the most important physicochemical constraints and the activity of up to seven biochemical reactions. Numerical optimization resulted in predictions that recovered qualitatively the growth-rate dependencies of cellular ribosome content, of cell size, and of the emergence of overflow metabolism. Similar to other constraint-based approaches [9–11], this modelling paradigm does not consider regulatory constraints, but instead assumes that protein regulation evolved to implement the optimal state.

Due to the inclusion of non-linear enzyme kinetics, maximizing the growth rate is inevitably a non-linear optimization problem; this may explain why no attempts have been made to extend this approach to more detailed biochemical whole-cell models. Instead, "toy models" of 1-3 reactions were solved analytically to gain further qualitative understanding of systems-level effects, such as the increased cellular investment into ribosomes at faster growth [6–8, 12] and optimal gene regulation strategies [3, 7].

Alternative modeling strategies, such as flux-balance analysis (FBA) [9, 13], resource balance analysis (RBA) [10], and ME (metabolism and expression) models [11], can be viewed as simplifications of the balanced growth scheme [14]. These models are also based on the idea of growth rate optimization in a steady-state solution space defined by physicochemical and environmental constraints. However, they consider only linear constraints, and they ignore the influence of metabolite concentrations on reaction kinetics [9, 13] or approximate their effects through growth-rate dependent phenomenological scaling laws [10, 11, 15].

Fig. 1A shows a simple model of balanced cellular growth, where a transporter protein imports a nutrient (G), which is converted into precursors for proteins (AA) through an enzymatic reaction, and where the precursors are converted by a "ribosome" protein into the three pro-

* Main text + SI

† martin.lercher@hhu.de

teins used as catalysts. Fig. 1B shows a slightly more involved model with cofactors, while Fig. 1C sketches an arbitrarily complicated balanced growth model. The boundary conditions and the requirement of balanced exponential growth define the system's solution space; maximizing the growth rate results in quantitative prediction of reaction fluxes and cellular concentrations (biomass composition). Here, we develop a theoretical framework for the analysis of such systems.

II. MODELS OF BALANCED EXPONENTIAL GROWTH

Our model assumes that in balanced growth, the cell increases exponentially in size, while the concentrations of all cellular components remain constant. In particular, we assume that the number of different membrane constituents per cell volume (and thus membrane composition and surface/volume ratio) remain constant in a given environment [5]. We do not explicitly model cell division; thus, our model can also be interpreted as describing the growth of a population of cells, with the simplifying assumption that individual cells have the same molecular composition [7]. To satisfy these assumptions, the net production rate of each molecular constituent must balance its dilution by growth,

$$\frac{dx}{dt} = \mu x \quad , \tag{1}$$

where x denotes the concentration of a given component and μ is the cellular growth rate [7].

The mass conservation in chemical reaction networks such as in Fig. 1 is commonly described through a *stoichiometric matrix* N, where rows correspond to metabolites and each column describes the mass balance of one reaction, with negative entries for consumed and positive entries for produced metabolites [19]. Here, we will focus on matrices A that describe a network of active reactions, i.e., A is a sub-matrix of N that contains all columns j for reactions with flux $v_j \neq 0$ and all rows for reactants i involved in these reactions either as substrates or as products. Note that the activity of each reaction j in A implies a positive concentration $p_j > 0$ for the protein catalyzing j; in our notation, catalytic proteins include not only enzymes, but also transporters and the ribosome.

For the development below, it will be convenient to express each concentration in units of mass concentration (mass per volume), which can be obtained from the corresponding number concentration by multiplying with the molecular mass. Accordingly, the entries of A are not stoichiometric coefficients but are mass fractions. We normalize the columns of A such that the negative entries (the mass concentration fraction consumed) sum to -1 and the positive entries (the mass concentration fraction produced) sum to $+1$; transport reactions do not have to be mass balanced, and thus one of these sums may have a smaller absolute value [19].

The mass conservation of each component in an active reaction network can then be stated in matrix notation as

$$A\boldsymbol{v} = \mu \begin{bmatrix} P \\ \boldsymbol{a} \end{bmatrix} \quad , \tag{2}$$

where a_α is the mass concentration of reactant α and P is the total protein mass concentration, summed over all proteins j,

$$P = \sum_j p_j \quad . \tag{3}$$

The rate v_j of a biochemical reaction j is the product of the concentration of its catalyzing protein p_j and some kinetic function $k_j(\boldsymbol{a})$ that depends on the concentrations a_α of active reactants,

$$v_j = p_j k_j(\boldsymbol{a}) \quad . \tag{4}$$

We assume that the functional form of $k_j(\boldsymbol{a})$ and the (constant) kinetic parameters are known. k_j has units of $[\text{time}]^{-1}$, and v_j has units of $[\text{mass}][\text{volume}]^{-1}[\text{time}]^{-1}$. $k_j(\boldsymbol{a})$ may depend on the mass concentrations of substrates, products, and other molecules a_α acting as inhibitors or activators. In the simplest case of reaction j following irreversible Michaelis-Menten kinetics with a single substrate α,

$$k_j(\boldsymbol{a}) = k_{cat} \left(\frac{a_\alpha}{a_\alpha + K_m} \right) \tag{5}$$

with constant enzyme activity k_{cat} (in units of $[\text{time}]^{-1}$) and Michaelis constant K_m (in units of $[\text{mass}][\text{volume}]^{-1}$).

The final constraint considered here reflects the cellular requirement for a minimal amount of free water to facilitate diffusion [20, 21]. *E. coli* growth decreases when cellular free water content is reduced below standard conditions, eventually stopping altogether when free water disappears [21]. *E. coli*'s buoyant density [22] and cellular water content [21] depend only on external osmolarity; at fixed external osmolarity, buoyant density remained constant even when experimenters induced drastic changes in cell mass and macromolecular composition [23].

Accordingly, we assume here that the cellular dry weight per volume is limited to ρ, where ρ is determined by external osmolarity. We express this capacity constraint as

$$\rho \geq P + \sum_\alpha a_\alpha \quad . \tag{6}$$

For simplicity of notation, we use the following conventions: $\{\alpha\}$ is the set of all reactants in the active stoichiometric matrix A, and $\sum_\alpha$ indicates that we sum over

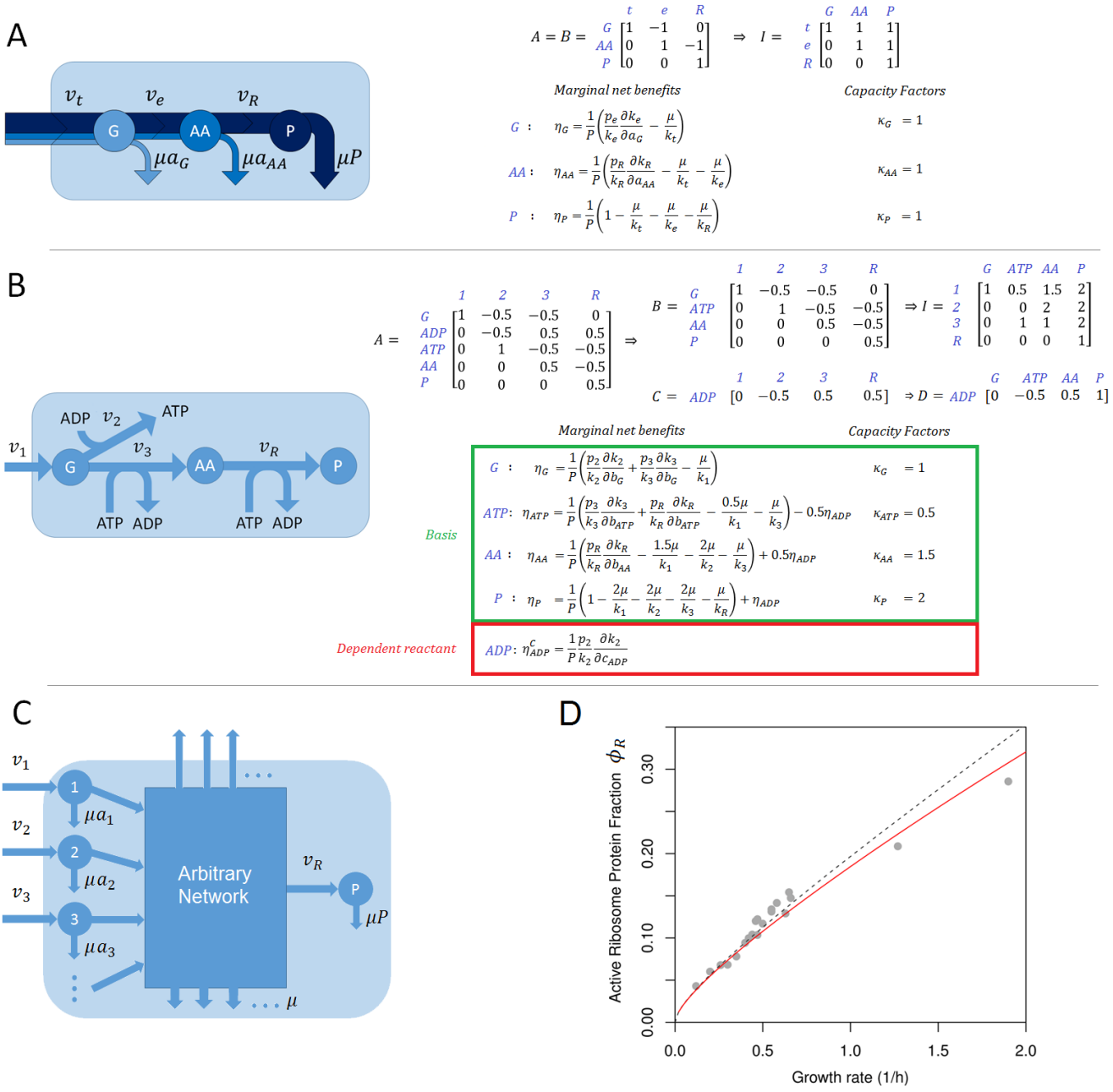

FIG. 1. Examples of balanced growth models (BGMs) and their mathematical description, derived from the active matrix A and the kinetic functions $\boldsymbol{k}(\boldsymbol{a})$: basis matrix B, investment matrix $I = B^{-1}$, closure matrix C, dependence matrix $D = CI$, marginal net benefits η, and capacity factors κ. (A) A model with a simple linear network of irreversible reactions, connecting a single transporter to the final production of proteins; linear networks never have dependent reactants, as the number of reactions equals the number of components. Colors indicate the fraction of flux that is eventually diverted into the dilution of each downstream component. (B) A more elaborate, nonlinear model of irreversible reactions that includes cofactors and a dependent reactant (ADP). (C) Schematic diagram of an arbitrary model with multiple transporters, the excretion of compounds, and a final "ribosome" reaction producing proteins. (D) The prediction of the active ribosomal protein fraction in *E. coli* based on Eq.(27) (red line, no free parameters) agrees with experimental values across 20 different growth conditions [16, 17], with a Pearson's correlation coefficient of $R^2 = 0.96$ ($P = 2.4 \times 10^{-14}$) and geometric mean fold error $= 1.09$. The dashed grey line is the prediction using Eq.(S34) without production costs, which becomes a good approximation of Eq.(27) at lower growth rates (see also Ref. [18]).

all $\alpha \in \{\alpha\}$. We use corresponding notations for the sets of *basis reactants* $\{\beta\}$ and *dependent reactants* $\{\gamma\}$ (see below). For an overview over the symbols used in this manuscript, see **Suppl. Table S2**.

Based on biophysical considerations, we could replace Eq.(6) with separate capacity constraints on the total volume concentration inside each cellular compartment [20] and on the total area occupied by non-lipid membrane components per membrane area [5, 24]. An even simpler capacity constraint imposed in most previous models [3, 5–8, 10–12] is to fix total protein concentration P to a constant value. However, it has been shown that P decreases with increasing growth rate [23, 25]. Thus, while a constant P allows to simplify the presentation, Eq.(6) provides a more meaningful constraint; moreover, Eq.(6) allows us to determine the costs and benefits of varying the total protein concentration.

Each cellular state, defined through the (element-wise) positive concentration vectors $\boldsymbol{p} > 0$ and $\boldsymbol{a} > 0$ (and the corresponding flux vector $\boldsymbol{v}$ uniquely defined by the concentrations), is a balanced growth state if and only if it satisfies Eqs. (2), (4), and (6); the set of all such states forms the solution space of balanced growth. Mass conservation (Eq.(2)) and reaction kinetics (Eq.(4)) relate reaction fluxes to the concentration vector $[P, \boldsymbol{a}]^T$ in two fundamentally different ways; below, we will exploit this fact to eliminate the flux variables and to derive explicit expressions for p_j and for μ.

III. CELLULAR STATE DEFINED BY THE CONCENTRATION VARIABLES

Our first aim is to derive a simple mathematical description of the solution space of balanced growth, with an emphasis on optimal states (i.e, states of maximal balanced growth rate μ). Let A be the active stoichiometric (sub-)matrix, $\boldsymbol{v}^*$ the flux vector, and $\boldsymbol{x}^* := [P^*, \boldsymbol{a}^*]^T$ the concentration vector of such an optimal balanced growth state. In **SI text VIII A**, we use results on constrained FBA problems [26, 27] to show that $\boldsymbol{v}^*$ is an *elementary flux mode* (EFM) [28] of a related FBA problem defined by A together with a constant biomass vector $\boldsymbol{x}^*$.

The derivations below assume that A has full column rank. This will be the case if A is the active matrix of an EFM for any constant biomass vector $\boldsymbol{x}$ [29]; in particular, this is true for the active matrix of optimal balanced growth states (**SI text VIII A**; see also Ref. [30]).

A may have more rows than columns. It is convenient to decompose the linear system of equations represented by Eq.(2) into two parts, rearranging the rows of A into matrices B, C such that $\#rows(B) = rank(B) = rank(A)$,

$$B\boldsymbol{v} = \mu \begin{bmatrix} P \\ \boldsymbol{b} \end{bmatrix} \tag{7}$$

$$C\boldsymbol{v} = \mu \boldsymbol{c} \quad , \tag{8}$$

where $\boldsymbol{b}$ and $\boldsymbol{c}$ are the reactant concentration vectors corresponding to the rows of B and C, respectively. B is identical to the reduced stoichiometric matrix in Ref. [31]. The relationship between A and B, C can be understood in terms of Matroid theory, where the rows of B form a *basis* for the *matroid* spanned by the rows of A, and the set of rows of C is the *closure* for the set of rows of B. If the choice for the partitioning of A into B and C is not unique, some partitionings may be pathological and should be avoided (**SI text VIII B**).

B is a square matrix of full rank, so there is always a unique inverse $I := B^{-1}$. Multiplying both sides of Eq.(7) by I from the left, we obtain

$$\boldsymbol{v} = \mu I \begin{bmatrix} P \\ \boldsymbol{b} \end{bmatrix} \quad . \tag{9}$$

I_{ji} quantifies the proportion of flux j invested into the dilution of component i, and we thus name I the *investment* (or *dilution*) matrix (see Fig. 1 for examples). In contrast to the stoichiometric matrix A, which describes local mass balances (Eq.(2)), I describes the structural allocation of reaction fluxes into the production of cellular components diluted by growth, and thus carries global, systems-level information.

By substituting $\boldsymbol{v}$ in Eq.(8) with Eq.(9), we obtain

$$\boldsymbol{c} = D \begin{bmatrix} P \\ \boldsymbol{b} \end{bmatrix} \quad , \tag{10}$$

where we defined the *dependence* matrix $D := CI$. D describes the linear dependence of the *dependent concentrations* $\boldsymbol{c}$ on P and $\boldsymbol{b}$; it is identical to the link matrix in Ref. [31].

When A is not square, B includes a proper subset of the rows in A, and thus B on its own is not mass balanced. The "missing" mass fluxes are balancing $\boldsymbol{c}$, and hence the flux investment into $\boldsymbol{c}$ is already accounted for by Eq.(9).

We are now in a position to express growth rate as an explicit function of the concentrations $[P, \boldsymbol{a}]^T$. As $k_j(\boldsymbol{a}) \neq 0$, we can use the kinetic equations (4) to express the individual protein concentrations as $p_j = v_j / k_j(\boldsymbol{a})$. Inserting v_j from the investment equation (9) gives

$$p_j = \frac{v_j}{k_j(\boldsymbol{a})} = \mu \frac{I_{jP} P + \sum_\beta I_{j\beta} b_\beta}{k_j(\boldsymbol{a})} \quad . \tag{11}$$

Substituting these expressions into the total protein sum, Eq.(3), we obtain

$$P = \mu \sum_j \frac{I_{jP} P + \sum_\beta I_{j\beta} b_\beta}{k_j(\boldsymbol{a})} \quad . \tag{12}$$

Below, we simplify the notation by writing $k_j := k_j(\boldsymbol{a})$. Solving for μ results in the *growth equation*

$$\mu(P, \boldsymbol{a}) = \frac{P}{\sum_j \frac{I_{jP} P + \sum_\beta I_{j\beta} b_\beta}{k_j}} \quad . \tag{13}$$

Thus, if the active matrix A of a balanced growth state is full rank, there are unique and explicit mathematical solutions for $\boldsymbol{p}$, $\boldsymbol{v}$, and μ. In particular, this is the case for optimal states, as well as for all other states whose active matrix is the active matrix of an EFM for any constant biomass $\boldsymbol{x}$. If all p_j are positive (Eq.(11)), the corresponding cellular state is a balanced growth state; otherwise, no balanced growth is possible at these concentrations.

IV. MARGINAL FITNESS CONTRIBUTIONS OF CELLULAR CONCENTRATIONS

In biological systems, costs and benefits should be expressed in terms of fitness effects. In situations where fitness f is determined exclusively by growth rate, a small change in growth rate $\delta\mu$ translates into a corresponding change in relative fitness (**SI text VIII C**) of

$$\delta f = \frac{\delta\mu}{\mu} \quad . \tag{14}$$

Accordingly, we define the *direct marginal net benefit* η_i^0 of the concentration x_i with $i \in \{P, \beta\}$ (i.e., $x_i \in \{P, b_\beta\}$) as the relative change in growth rate due to a small change in x_i [32],

$$\eta_i^0 := \frac{1}{\mu}\frac{\partial\mu}{\partial x_i} \quad . \tag{15}$$

While we assume that the original state before the change in x_i respected the capacity constraint Eq.(6), we ignore the capacity constraint for the perturbed state in these definitions (i.e., we allow capacity to "adjust" to the change in x_i and in any dependent concentrations c_γ).

For $i \in \{P, \beta\}$, let us define

$$q_i^j := \frac{\mu I_{ji}}{P k_j} \quad . \tag{16}$$

From the growth equation (13), it follows that

$$\eta_P^0 = \frac{1}{P} - \sum_j q_i^j \tag{17}$$

and

$$\eta_\beta^0 = \sum_j (u_i^j - q_i^j) \tag{18}$$

with

$$u_\beta^j := \frac{p_j}{P}\frac{1}{k_j}\frac{\partial k_j}{\partial b_\beta} \quad . \tag{19}$$

The summands in the denominator of the growth equation (13) can be expressed as $p_j/\mu = v_j/(\mu k_j)$ (Eq.(11)). Accordingly, $q_i^j = \frac{1}{P}\left(\frac{\partial p_j}{\partial x_i}\right)_{k_j=\text{const.}}$ quantifies the proportional increase of p_j to help offset the increased dilution of component i, and we thus call this the marginal (relative) *production cost* incurred by the system via protein p_j. If I_{ji} and k_j are both positive, then q_i^j is also positive, i.e., it decreases fitness. The production costs are global, systems-level effects, quantified through the investment matrix I.

Conversely, $u_\beta^j = -\frac{1}{P}\left(\frac{\partial p_j}{\partial b_\beta}\right)_{v_j=\text{const.}}$ quantifies the proportion of protein p_j "saved" due to the change in kinetics associated with an increase in b_β [33]. The benefit u_β^j will generally be positive if β is a substrate of reaction j. We thus call u_β^j the marginal (relative) *kinetic benefit* of reactant β to reaction j. The kinetic benefit is a purely local effect, as it is non-zero only for reactants that affect the kinetics of reaction j. Because fluxes are proportional to the concentrations of the proteins catalyzing the corresponding reactions, the marginal benefit of total protein in terms of relative fitness is simply P^{-1}.

Combining the two relationships for production cost and kinetic benefit, we see that the direct net benefit is the reduction of the protein fraction p_j/P at constant μ facilitated by the increase in x,

$$\eta_i^0 = -\frac{1}{P}\sum_j \left(\frac{\partial p_j}{\partial x_i}\right)_{\mu=\text{const.}} \quad . \tag{20}$$

The definition of η_i^0 ($i \in \{P, \beta\}$) accounts for the production costs of dependent reactants c_γ (as these are embedded in I), but ignores their kinetic benefits. In analogy to Eq.(19), we define these as

$$\eta_\gamma^c := \sum_j u_\gamma^j \quad , \tag{21}$$

with the marginal (relative) *kinetic benefit* of reactant γ to reaction j

$$u_\gamma^j := \frac{p_j}{P}\frac{1}{k_j}\frac{\partial k_j}{\partial c_\gamma} \quad . \tag{22}$$

We can now use the above relations to define the *total marginal net benefit* η_i of concentration $i \in \{P, \beta\}$ as the relative change in growth rate due to a small change in x_i and the resulting change in the concentrations of its dependent reactants c_γ,

$$\begin{aligned} \eta_i &:= \frac{1}{\mu}\left(\frac{\partial\mu}{\partial x_i} + \sum_\gamma \frac{\partial\mu}{\partial c_\gamma}\frac{\partial c_\gamma}{\partial x_i}\right) \\ &= \eta_i^0 + \sum_\gamma D_{\gamma i}\eta_\gamma^c \quad , \end{aligned} \tag{23}$$

where the second equality follows from Eqs.(23), (10), (13) and definitions (21), (22).

A change δx_i of x_i ($i \in \{P, \beta\}$) causes a correlated change of each dependent concentration $\delta c_\gamma = D_{\gamma i}\delta x_i$

(Eq.(10)). Thus, a change by δx_i results in a total change of the utilization of cellular capacity by $\kappa_i \delta x_i$, with the *capacity factor*

$$\kappa_i := 1 + \sum_\gamma D_{\gamma i} \quad . \tag{24}$$

V. OPTIMAL GROWTH AND THE BALANCE OF MARGINAL NET BENEFITS

At maximal growth rate, the cellular components will utilize the full cellular capacity ρ to saturate enzymes with their substrates, and thus the constraint in Eq.(6) will be active, that is, the inequality will become an equality.

To derive necessary conditions for any optimal balanced growth state at constant cellular capacity ρ, we use the method of Lagrange multipliers, which quantify the importance of the capacity constraint, Eq.(6), and of the constraints for the dependent reactants, Eq.(10), for the maximization of the objective function. The Lagrangian $\mathcal{L}$ is a function of P, $\boldsymbol{a}$, and ρ (**SI text VIII D**).

Note that instead of using Lagrange multipliers, one could express the total protein concentration $P = \rho - \sum_\alpha a_\alpha$ (Eq.(6)) and the dependent reactant concentrations $c_\gamma = D_{\gamma P} P + \sum_\beta D_{\gamma\beta} b_\beta$ (Eq.(10)) in terms of ρ and of the independent reactant concentrations $\boldsymbol{b}$. Substituting the resulting expressions in the growth equation (13) would result in an objective function that depends only on ρ and $\boldsymbol{b}$, and that is constrained only by the requirement of positive concentrations. While this would lead to the same balance equations (26) as derived in the Lagrange multiplier framework, this formulation misses important insights that can be derived from the Lagrange multipliers themselves.

The maximal balanced growth rate μ^* will be a function of the cellular capacity ρ. In analogy to the marginal net benefits of cellular components, we define the *marginal benefit* of the cellular capacity as the fitness increase facilitated by a small increase in ρ,

$$\eta_\rho := \frac{1}{\mu^*}\frac{d\mu^*}{d\rho} = \frac{1}{\mu}\frac{\partial \mathcal{L}}{\partial \rho}\bigg|_{\mu=\mu^*} = -\frac{\lambda_\rho}{\mu^*} \quad , \tag{25}$$

where the second equality follows from the envelope theorem [34].

A necessary condition for optimal balanced growth is that all partial derivatives of $\mathcal{L}$ with respect to the concentrations (P, b_β, c_γ) and to the Lagrange multipliers $(\lambda_\rho, \lambda_\gamma)$ are zero. As detailed in **SI text VIII D**, this leads to the balance equations

$$\forall_{i\in\{P,\beta\}} \quad \eta_i = \kappa_i \eta_\rho \quad . \tag{26}$$

The optimal state is perfectly balanced: the marginal net benefit of each independent cellular concentration x_i equals the marginal benefit of the cellular capacity, scaled by κ_i to account for its total utilization of cellular capacity. If i does not have any dependent reactants ($\forall_\gamma \; D_{\gamma i} = 0$) or if $A = B$, then the balance equation simplifies to $\eta_i^0 = \eta_\rho$.

Eq.(26) states that if the dry weight density ρ would be allowed to increase by a small amount, such as $1\mu g/l$, then the marginal fitness gain that could be achieved by increasing protein concentration (plus dependent concentrations) by this amount is identical to that achieved by increasing the concentration of any reactant β (plus its dependent concentrations) by the same amount. In hindsight, this should not be surprising: if the marginal net benefit of concentration x_i (scaled by κ_i) was lower than that of $x_{i'}$, growth rate could be increased by increasing $x_{i'}$ at the expense of x_i.

Eq.(26) together with Eq.(6) describes a system of $m+2$ equations for $m+2$ unknowns (with m the number of basis reactants β; **SI text VIII D**). Any state of optimal growth must satisfy these equations. In realistic cellular systems, this set of equations has a finite number of discrete solutions. Thus, solving the non-linear optimization problem described by Eq.(2) and the corresponding constraints may potentially be accelerated by instead searching for the solution of the balance equations. If the optimization problem is convex, the conditions given by Eq.(26) are necessary and sufficient, and the solution is unique.

VI. QUANTITATIVE PREDICTIONS

A fully parameterized genome-scale balanced growth model could be used to predict all cellular concentrations at maximal growth rate. However, kinetic constants are currently lacking for many reactions even in the best studied model organisms [35]. We can still make quantitative predictions for the ribosome if we consider simplified models where the ribosome produces proteins from a single substrate, a generic ternary complex AA not consumed by other reactions (such as used in Ref. [36] and shown in Fig. 1A). In the balance equation $\eta^0_{AA} = \eta^0_P$, the costs on both sides largely cancel each other, as all reactions that contribute directly or indirectly to AA production contribute proportionally to protein production. The only cost not cancelled is $q^R_P = \mu/(P k_R)$ (see, e.g., the balance equation for reactant "AA" in Fig.1A). Thus, only the kinetics of the ribosome k_R remain relevant, which we obtain from Ref. [36]. Using the mass balance of proteins, $v_R = \mu P$, we thus predict the optimal protein fraction of actively translating ribosomes, $\phi_R = p_R/P$, at each growth rate (**SI text VIII E**),

$$\phi_R(\mu) = \frac{\mu r_P}{k_{cat}}\left[1 + \frac{K_m}{2P}\left(\sqrt{1 + \frac{4P}{K_m}\left(\frac{k_{cat}}{\mu} - 1\right)} - 1\right)\right] \quad , \tag{27}$$

where r_P is the mass fraction of the ribosome made up of protein, and k_{cat}, K_m are the kinetic parameters [36, 37].

Despite the simplicity of this model, the predicted ϕ_R is in good agreement with experimental values [16, 17] (Fig. 1D, solid red line). An approximation that ignores

the dilution of intermediates (production costs) and bases its predictions only on the capacity ρ (Fig. 1D, grey dashed line) results in less accurate predictions. However, this last approximation becomes better at lower growth rates, where the dilution of intermediates μa_α becomes less and less important, and the optimal concentrations are increasingly determined by the capacity constraint. This finding explains why the assumption of a minimal utilization of cellular capacity by individual catalysts and their substrates provides a good approximation for the relationship between their concentrations [18].

To get a rough quantitative estimate of the marginal net benefits η, let us consider the simplest model of a complete cell, consisting of only a transport protein and the ribosome [3, 6] (Suppl. Fig. VIII F). Based on the experimentally observed protein fraction of total dry weight $P/\rho = 0.54$ in *E. coli* [21], we estimate $\rho\ \eta_\rho = 0.69$ (**SI text VIII F**). $\rho\ \eta_\rho$ quantifies the relative change in the maximal growth rate μ^* resulting from a small, relative change in ρ. Thus, we estimate that a decrease in cellular dry weight density ρ of 1% would lead to a 0.69% decrease in growth rate, indicating that the capacity constraint is indeed highly biologically significant.

ρ changes when external osmolarity is modified [21]. $\rho\ \eta_\rho = \frac{\rho}{\mu}\frac{d\mu}{d\rho} = \frac{d\ln\mu}{d\ln\rho}$ is the slope of the log-scale plot of μ vs. ρ at different external osmolarities. Increases in cellular dry weight density may have strong effects on diffusion and may hence change the kinetic constants. In contrast, reductions in ρ due to decreased external osmolarity are within the scope of our model, which assumes constant parameters for $\boldsymbol{k}(\boldsymbol{a})$. The very limited available experimental data (three data points from Ref. [38], Fig. VIII F) suggests $\rho\ \eta_\rho \approx 0.66$, close to our rough estimate from the minimal cell model.

VII. CONCLUSIONS

Our derivations are based on the insight that for any balanced growth state that corresponds to an EFM of the related FBA problem, the inverse I of the active stoichiometric matrix (or a basis thereof) contains global, systems-level information on the contribution of individual fluxes to the production of cellular constituents diluted by growth. Purely through structural constraints, this leads to an explicit dependence of reaction fluxes on the concentrations of the cellular constituents, scaled linearly by the growth rate (Eq.(9)). We combine this description with the complementary kinetic dependence of fluxes on concentrations. This allows us to provide explicit expressions for the individual protein concentrations $\boldsymbol{p}$ and fluxes $\boldsymbol{v}$ and for the growth rate μ as functions of arbitrary (positive) concentrations $[P, \boldsymbol{b}]^T$, and it provides the framework from which we derive the balance equations for the marginal net benefits of cellular concentrations at optimal growth.

Previous work has emphasized the central role of proteins in the cellular economy [3, 5–8, 10–12]. Whereas total protein mass concentration in real biological systems is indeed much higher than the mass concentration of any other cellular constituent a_α, the balance equations show that at optimal growth, their marginal net benefits are in fact equal, emphasizing the importance of explicitly accounting for all cellular constituents.

To make the presentation concise, our equations assume (i) that all proteins contribute to growth by acting as catalysts or transporters; (ii) that there is a 1-to-1 correspondence between proteins and reactions; (iii) that proteins are not used as reactants; and (iv) that all catalysts are proteins. It is straight-forward to remove these simplifications. E.g., assumption (i) can be removed by adding a sector of non-growth related proteins [15, 39] with concentration Q to the r.h.s. of Eq.(3); to remove assumption (iv), we can add different RNA species as cellular components and introduce reactions that combine proteins and RNA into molecular machines such as the ribosome.

An equation analogous to Eq.(9) can be formulated for non-growing cells (or cellular subsystems) that are instead optimized for the production of specific molecules, as is the case for many cell types in multicellular organisms. Instead of multiplying a dilution term, the production matrix I would then multiply a weighted *production vector* representing the desired output. The same strategy might be used to accelerate FBA solutions in a given EFM: FBA studies cellular growth, but assumes a fixed biomass composition $[P, \boldsymbol{a}]^T$.

In principle, exploitation of the balance equations (Eq.(26)) may allow the numerical optimization even for cellular systems of realistic size, encompassing hundreds of protein and reactant species. One remaining obstacle to the accurate formulation of such models, though, is the current incompleteness of the kinetic constants needed to parametrize the functions $\boldsymbol{k}(\boldsymbol{a})$ [35]. The need for the development of high-throughput methods to systematically ascertain these parameters has been recognized [35]; in the meantime, methods from artificial intelligence may provide reasonable approximations [40].

As an alternative to genome-scale models, the balanced growth theory developed here could be applied to coarse-grained cellular models of increasing complexity, parameterized from experimental data. One would start from minimal models with two reactions [3, 6], proceeding to models comprising the six previously described sectors of the cellular economy [39] and beyond [41].

Our work extends the ad-hoc optimizations of toy models [3, 5–8] into a full-fledged theory of balanced growth. We show that the balanced growth framework allows general, quantitative insights into cellular resource allocation and physiology, as exemplified by the growth and balance equations. Application and further development of this theory may foster an enhanced theoretical understanding of how physicochemical constraints determine the fitness costs and benefits of cellular organization. Moreover, the explicit expressions for the (marginal) costs and benefits of cellular concentrations in terms of fitness provide a rig-

orous framework for analyzing the cellular economy. We anticipate that this approach will prove fruitful in the interpretation of natural and laboratory evolution, and in optimizing the design of synthetic biological systems.

ACKNOWLEDGEMENTS

We thank Johannes Berg, Oliver Ebenhöh, Xiao-Pan Hu, Terry Hwa, Michael Lässig, Wolfram Liebermeister, and Deniz Sezer for discussions. This work was funded by the Deutsche Forschungsgemeinschaft (DFG,German Research Foundation) through grants IRTG 1525, CRC 680, CRC 1310, and, under Germany's Excellence Strategy, through grant EXC 2048/1 (Project ID: 390686111).

VIII. SUPPLEMENTARY MATERIAL

A. The active sub-network at maximal growth rate forms an elementary flux mode

Let N be a stoichiometric matrix of a general balanced growth model. Let A be a sub-matrix of N that contains only the columns for reactions that are active in a solution that maximizes the growth rate μ of the general problem, with corresponding active (nonzero) concentrations $\boldsymbol{x} = [P, \boldsymbol{a}]^T$. Let $\boldsymbol{x}^*$ be the concentrations, $\boldsymbol{v}^*$ the fluxes, $\boldsymbol{k}^* = \boldsymbol{k}(\boldsymbol{x^*})$ the values of the kinetic functions, and μ^* the growth rate of the optimal solution. Then these values characterize also the optimal solution to the reduced balanced growth optimization problem, constrained by Eqs. (2), (3), (4), and (6) and summarized here:

$$\begin{aligned} &\max\ \mu \\ &\text{subject to:} \\ &\quad A\boldsymbol{v} = \mu\boldsymbol{x} \\ &\quad P = \sum_j p_j \\ &\forall_j\ \ v_j = p_j k_j(\boldsymbol{x}) \\ &\quad \rho \geq P + \sum_\alpha a_\alpha \quad . \end{aligned} \tag{S1}$$

We now wish to relate the solution of this problem to elementary flux modes (EFMs) [28], which are defined for the types of steady state models used in flux balance analysis (FBA) [9, 13]. We thus convert Eq.(S1) into a corresponding constrained FBA problem by treating the "biomass" vector $\boldsymbol{x}^*$ of the optimal solution as a constant and by relaxing the equality constraint on the individual protein concentrations into an inequality constraint, $P^* \geq \sum_j p_j$. With constant biomass $\boldsymbol{x}^*$, Eq.(2) is equivalent to the standard steady state constraint of flux balance analysis problems, formulated with an explicit balance equation for each biomass component rather than with a separate biomass reaction in A (see, for example, Eq.(2) in Ref.[42]). With constant concentrations, total protein P^* and the kinetic functions $\boldsymbol{k}^*$ also become constant; the constraint relating to the cellular capacity (Eq.(6)) is trivially respected and can be ignored. Thus, the constrained FBA problem for the active stoichiometric matrix becomes:

$$\begin{aligned} &\max\ \lambda \\ &\text{subject to:} \\ &\quad A\boldsymbol{v} = \lambda\boldsymbol{x}^* \\ &\quad P^* \geq \sum_j p_j \\ &\forall_j\ \ v_j = p_j k_j^* \quad , \end{aligned} \tag{S2}$$

with biomass production rate λ and constant $\boldsymbol{k^*}$, P^*, and biomass $\boldsymbol{x}^*$. This is precisely the type of constrained flux balance problem analyzed in Refs. [26, 27], which prove that the solutions $\boldsymbol{v}^{\text{opt}}$ to the optimization problem defined by Eq.(S2) are elementary flux modes (EFMs).

In the optimal solution to the problem defined by Eq.(S2), the protein concentration constraint will be active, that is, $P^* = \sum_j p_j$; if not, the biomass production rate λ could be increased by multiplying the vector of protein concentrations $\boldsymbol{p}$ with a constant > 1 (as $v_j = p_j k_j^*$ for all j). Thus, the optimization problem described by Eq.(S2) is the same as that described by Eq.(S1), except for a reduction in the dimension of the search space due to the fixed concentrations $\boldsymbol{x}^*$. Accordingly, the flux distribution $\boldsymbol{v}^*$ that maximizes the balanced growth rate μ in Eq.(S1) also maximizes the biomass production rate λ of the FBA problem in Eq.(S2); it is hence an EFM of the active stoichiometric matrix A with biomass $\boldsymbol{x}^*$ [26, 27].

Furthermore, it has been shown that the active stoichiometric matrix A of an EFM has full column rank (if A is formulated without an explicit "biomass" reaction) [29]. We conclude that the active stoichiometric matrix A of a balanced growth model at maximal growth rate has full column rank. Consequently, the corresponding basis matrix B is always invertible.

We emphasize that $\boldsymbol{v}^*$ is an EFM of the constrained FBA problem in Eq.(S2), not of the balanced growth problem in Eq.(S1) from which it is derived. EFMs are defined as equivalence classes of minimal admissible steady-state flux distributions, whose members can be converted into each other by multiplication with a positive scalar [28]. This definition can not be generalized to balanced growth models, as multiples of an admissible flux vector generally do not satisfy Eq.(2). For this reason, de Groot *et al.* have generalized the concept of EFMs to equivalence classes of minimal sets of active reactions in balanced growth states, termed *elementary growth modes* (EGMs) [30]. In parallel work to that presented here, these authors have shown that optimal solutions to balanced growth problems are EGMs, and that the active stoichiometric matrix of an EGM has full rank [30].

If instead of a single constraint on cellular capacity, multiple capacity constraints are imposed simultaneously (e.g., to describe separate constraints on cytosolic and membrane capacities), then the solutions may instead correspond to convex combinations of EFMs of the related FBA problem [30, 43]; this appears to be the case in overflow metabolism in *E. coli* [44]. In such cases, it is not guaranteed that A has full column rank, and generalizations of our theory may not be straightforward.

B. Choice of basis and relationship between capacity and dependence constraints

Not every reactant can be considered dependent: a reactant such that the corresponding row in the active matrix A is linearly independent of all other rows will

always be in the basis (equivalently, a reactant that has zero entries in all vectors in a basis for the left null space of A cannot be a dependent reactant).

It is possible for some models that there is one or more choices of basis such that its corresponding dependence matrix has for some $i \in \{P, \beta\}$

$$\sum_{\gamma} D_{\gamma i} = -1 \quad . \tag{S3}$$

In these cases, any marginal change in the mass concentration of component i will cause the exact opposite change in the total mass concentration of its dependent reactants γ. When this is combined with the capacity constraint as defined in Eq.(6), these changes in concentrations result in a perfect cancellation in the capacity utilized by i and its dependent reactants, and thus a zero net change in capacity for any change in the concentration i (i.e., $\kappa_i = 0$ from Eq.(24)). For this reason, the marginal net benefit of i is simply $\eta_i = 0$ (Eq.(26)).

Such a perfect cancellation is highly unlikely if we use a more realistic description of the capacity constraint, where different cellular components i have different specific capacity utilizations σ_i; e.g., if we assume that the capacity constraint limits the total volume occupied by cellular components, then σ_i gives the volume per mass of component i. In this case, the capacity constraint Eq.(6) is replaced by a constraint on the volume of cellular dry mass per volume of cell water, ν:

$$\nu = \sigma_P P + \sum_{\alpha} \sigma_\alpha a_\alpha \quad , \tag{S4}$$

where σ_P is the specific capacity of proteins (almost constant for different proteins [45]) and σ_α is the specific capacity of reactant α, which depends on its chemical properties such as hydrophobicity and charge [46].

C. Definition of relative fitness

In a situation where competition among cells is solely through differential intrinsic growth rates, absolute fitness is equal to growth rate: In a population of cells growing exponentially with growth rate μ, the selection coefficient for a variant with growth rate $\mu + \delta\mu$ is simply $\delta\mu$ [1]. Population genetics models almost always employ relative fitness [47], which we here define as a relative growth rate:

$$f := \frac{\mu + \delta\mu}{\mu} = 1 + \frac{\delta\mu}{\mu} \quad . \tag{S5}$$

Thus, to quantify the effect of a small change δx to some parameter x on relative fitness, we use

$$\frac{\delta f}{\delta x} = \frac{1}{\mu}\frac{\delta\mu}{\delta x} \quad . \tag{S6}$$

Note that population genetics models are frequently defined in terms of discrete generations. With generation time $T = \ln 2/\mu$, the selection coefficient of the variant *per generation* is then [48]

$$s_T = (f-1)\ln 2 = \frac{\delta\mu}{\mu}\ln 2 \quad . \tag{S7}$$

D. Solution for the optimization problem using Lagrange multipliers

Our objective function is given by Eq.(13), which expresses μ as an explicit function of the concentrations $[P, \boldsymbol{a}]^T$. The capacity constraint will be active at maximal growth rate, i.e., Eq.(6) becomes an equality. The capacity constraint can then be expressed as a function g_ρ that also only depends on the concentrations,

$$g_\rho(P, \boldsymbol{a}) = P + \sum_{\alpha} a_\alpha - \rho = 0 \quad . \tag{S8}$$

Finally, the constraints on each dependent reactant γ also only depend on $P, \boldsymbol{a}$, with the entries $D_{\gamma P}$ determining the composition of each γ in terms of P, and $D_{\gamma\beta}$ determining the composition of γ in terms of $\boldsymbol{b}$,

$$g_\gamma(P, \boldsymbol{a}) = D_{\gamma P} P + \sum_{\beta} D_{\gamma\beta} b_\beta - c_\gamma = 0 \quad . \tag{S9}$$

We are now able to define a Lagrange function as the sum of the objective function μ and the constraints $\boldsymbol{g}$ scaled by Lagrange multipliers,

$$\mathcal{L} = \mu + \lambda_\rho g_\rho + \sum_{\gamma} \lambda_\gamma g_\gamma \quad . \tag{S10}$$

The first order necessary conditions for a constrained local maximum are that all partial derivatives of $\mathcal{L}$ with respect to the variables P, b_β, c_γ and to the Lagrange multipliers $\lambda_\rho, \lambda_\gamma$ are zero,

$$\begin{aligned} \forall_{i\in\{P,\beta,\gamma\}} \quad 0 &= \frac{\partial\mathcal{L}}{\partial x_i} \quad , \\ \forall_{\gamma} \quad 0 &= \frac{\partial\mathcal{L}}{\partial\lambda_\gamma} \quad , \\ 0 &= \frac{\partial\mathcal{L}}{\partial\lambda_\rho} \quad . \end{aligned} \tag{S11}$$

For the partial derivative with respect to an independent concentration x_i ($i \in \{P, \beta\}$), we have

$$\frac{\partial\mathcal{L}}{\partial x_i} = \frac{\partial\mu}{\partial x_i} + \lambda_\rho + \sum_{\gamma} \lambda_\gamma D_{\gamma i} = 0 \quad . \tag{S12}$$

With Eq.(23), this results in

$$\mu\ \eta_i^0 + \lambda_\rho + \sum_{\gamma} \lambda_\gamma D_{\gamma i} = 0 \quad . \tag{S13}$$

For the partial derivative relative to a dependent reactant γ', we have

$$\frac{\partial \mathcal{L}}{\partial c_{\gamma'}} = \frac{\partial \mu}{\partial c_{\gamma'}} + \lambda_\rho - \lambda_{\gamma'} = 0 \quad . \tag{S14}$$

With Eq.(21), we obtain

$$\lambda_{\gamma'} = \mu\ \eta^0_{\gamma'} + \lambda_\rho \quad . \tag{S15}$$

Substituting $\lambda_{\gamma'}$ from Eq.(S15) into Eq.(S13) gives (for $i \in \{P, \beta\}$)

$$\mu\ \eta^0_i + \lambda_\rho + \sum_\gamma \left(\mu \eta^c_\gamma + \lambda_\rho\right) D_{\gamma i} = 0 \quad . \tag{S16}$$

Rearranging results in

$$\begin{aligned} 0 &= \mu\ \eta^0_i + \lambda_\rho \left(1 + \sum_\gamma D_{\gamma i}\right) + \mu \sum_\gamma D_{\gamma i} \eta^c_\gamma \\ &= \mu\ \eta_i + \lambda_\rho \kappa_i \\ &= \mu\ \eta_i - \mu\ \eta_\rho \kappa_i \quad , \end{aligned} \tag{S17}$$

where we used $\eta_\rho = -\lambda_\rho/\mu$ (Eq.(25)). We thus obtain the balance equation

$$\eta_i = \kappa_i \eta_\rho \tag{S18}$$

for $i \in \{P, \beta\}$.

If neither P nor β have dependent concentrations ($\forall_\gamma\ D_{\gamma P} = 0 = D_{\gamma \beta}$), Eq.(S18) results in the equality

$$\frac{1}{P} - \sum_j q^j_P = \eta^0_P = \eta^0_\beta = \sum_j \left(u^j_\beta - q^j_\beta\right) \quad . \tag{S19}$$

In a regime where all reaction kinetics are much faster than the growth rate, $k_j \gg \mu$, the production costs q^j_i are close to zero, and Eq.(S19) can be approximated by

$$\begin{aligned} \frac{1}{P} &\approx \sum_j u^j_\beta = \sum_j \frac{p_j}{P} \frac{1}{k_j} \frac{\partial k_j}{\partial b_\beta} \\ \Leftrightarrow \quad 1 &\approx \sum_j \frac{p_j}{k_j} \frac{\partial k_j}{\partial b_\beta} \quad . \end{aligned} \tag{S20}$$

Thus, in this regime, the marginal net benefits of reactants are determined only by their influence on kinetics, and dilution can be neglected.

The second and third equation in Eq.(S11) simply enforce the constraints for the dependent reactants, Eq.(8), and for the cellular capacity, Eq.(6). Thus, the total number of equations is 1 (for ρ) + 1 (for P) + #(independent reactants β) + #(dependent reactants γ) = $m' + 2$, where m' is the total number of reactants α. The number of unknowns is also $m' + 2$ ($m' + 1$ concentrations + η_ρ). To find the state of maximal balanced growth for a given system, it is more convenient to first consider only the m independent reactants β, and determine the dependent reactants γ later from Eq.(8). In this case, Eq.(S18) and the constraint on cellular capacity Eq.(6) provide $m + 2$ equations for $m + 2$ unknowns ($m + 1$ concentrations in $[P, b_\beta]^T$ and η_ρ).

E. Optimal ribosome protein fraction

Here we assume a very simple model for translation [Klumpp 2013]. It accounts only for the elongation phase, where one catalyst (the ribosome plus bound mRNA, with concentration R) converts one substrate (the ternary complex, with concentration T) into protein, following irreversible Michaelis-Menten kinetics (Eq.(5)). As further simplifications, we assume that the model has no dependent reactants ($A = B$) and that the ternary complex is not used in any other reaction. In this case, the same canceling of production costs as in the model depicted in Fig.1A happens, and the balance of net benefits of ternary complex and total protein, $\eta_T = \eta_P$ (Eq.(26)), simplifies to

$$P u^R_T = 1 - \frac{\mu}{k_R} \quad . \tag{S21}$$

Substituting the partial derivative of irreversible Michaelis-Menten kinetics (Eq.(5)), we obtain

$$\frac{R}{a_T(1 + a_T/K_m)} = 1 - \frac{\mu}{k_R} \quad , \tag{S22}$$

where K_m is the Michaelis constant of the ribosome for the ternary complex. Rearranging Eq.(5), we also see that the kinetics determine the concentration a_T uniquely in terms of v_R, R, K_m, and the ribosome's turnover number k_{cat},

$$a_T = \frac{K_m}{\frac{k_{cat} R}{v_r} - 1} \quad . \tag{S23}$$

Substituting this into Eq.(S22) gives

$$\begin{aligned} R &= \left(1 - \frac{\mu}{k_R}\right) \left[\frac{K_m}{\frac{k_{cat} R}{v_r} - 1} \left(1 + \frac{1}{\frac{k_{cat} R}{v_r} - 1}\right)\right] \\ &= \left(1 - \frac{\mu}{k_R}\right) K_m \left[\frac{\frac{k_{cat} R}{v_r}}{\left(\frac{k_{cat} R}{v_r} - 1\right)^2}\right] \quad . \end{aligned} \tag{S24}$$

From the ribosome kinetics and mass conservation of proteins, we have

$$R\ k_R = v_R = \mu P \quad . \tag{S25}$$

Thus, substituting $\mu/k_R = R/P$ and $v_R = \mu P$ in Eq.(S24), we obtain

$$\frac{R}{P} = \left(1 - \frac{R}{P}\right) \frac{K_m}{P} \left[\frac{\frac{k_{cat} R}{\mu P}}{\left(\frac{k_{cat} R}{\mu P} - 1\right)^2}\right] \quad . \tag{S26}$$

Solving for R/P gives

$$\begin{aligned} &\left(\frac{R}{P}\right)^2 + \frac{\mu}{k_{cat}} \left(\frac{K_m}{P} - 2\right) \left(\frac{R}{P}\right) \\ &+ \left(\frac{\mu}{k_{cat}}\right)^2 \left(1 - \frac{k_{cat} K_m}{\mu P}\right) = 0 \quad , \end{aligned} \tag{S27}$$

i.e., a quadratic equation in R/P. Its two solutions are

$$\frac{R}{P} = \frac{\mu}{k_{cat}}\left[1 + \frac{K_m}{2P}\left(\pm\sqrt{1 + \frac{4P}{K_m}\left(\frac{k_{cat}}{\mu} - 1\right)} - 1\right)\right] \quad . \tag{S28}$$

To see which of the two solutions is relevant, we rewrite this as

$$k_{cat}R = \mu P\left[1 + \frac{K_m}{2P}\left(\pm\sqrt{1 + \frac{4P}{K_m}\left(\frac{k_{cat}}{\mu} - 1\right)} - 1\right)\right] \quad . \tag{S29}$$

Because $k_{cat}R > R\ k_R = v_R = \mu P$, the term in square brackets ($[\cdot]$) in Eq.(S29) must be > 1. Only the positive root is compatible with this condition. Thus, the ratio R/P is uniquely determined by

$$\frac{R}{P} = \frac{\mu}{k_{cat}}\left[1 + \frac{K_m}{2P}\left(\sqrt{1 + \frac{4P}{K_m}\left(\frac{k_{cat}}{\mu} - 1\right)} - 1\right)\right] \quad . \tag{S30}$$

To estimate the actual ribosome protein fraction of total protein ϕ_R, we need to scale the previous expression by the fraction r_P of ribosome which is protein, resulting in Eq.(27).

The same procedure can be used to find an equation for ϕ_R that ignores the production costs. Starting from Eq.(S24) without the production cost term μ/k_R, we obtain

$$\frac{R}{P} \approx \frac{K_m}{P}\left[\frac{\frac{k_{cat}R}{\mu P}}{\left(\frac{k_{cat}R}{\mu P} - 1\right)^2}\right] \quad , \tag{S31}$$

which results in a quadratic equation similar to Eq.(S27),

$$\left(\frac{R}{P}\right)^2 - 2\frac{\mu}{k_{cat}}\frac{R}{P} + \left(\frac{\mu}{k_{cat}}\right)^2\left(1 - \frac{k_{cat}K_m}{\mu P}\right) \approx 0 \quad . \tag{S32}$$

Solving for R/P gives

$$\frac{R}{P} \approx \frac{\mu}{k_{cat}}\left(1 \pm \sqrt{\frac{k_{cat}K_m}{\mu P}}\right) \quad . \tag{S33}$$

Again because $Rk_{cat} > \mu P$, the term in parentheses $(\cdot)$ in Eq.(S33) must be > 1, and again only the positive root is compatible with this condition. Thus, the ribosome protein fractionis uniquely determined in this approximation by

$$\phi_R \approx \frac{\mu r_P}{k_{cat}}\left(1 + \sqrt{\frac{k_{cat}K_m}{\mu P}}\right) \quad . \tag{S34}$$

We compared the predictions of the ribosome fraction of total protein, $\phi_R = R/P$, to quantitative proteomics data obtained by Schmidt *et al.* [16]. To obtain molar ribosome concentration, we calculated the median over all reported concentrations of ribosomal proteins. The concentration of ternary complexes was assumed to be identical to the concentration of their protein component, the elongation factor Tu. Molar concentrations of the ribosome and (total) ternary complexes were converted to mass concentrations by multiplying with molar masses derived from the amino acid sequences (for the protein parts) and nucleotide sequences (for the RNA parts). For this, we assumed that each ribosome contained one copy of each of its constituents, with the exception of four copies of RplL [49]. To calculate the mass fraction of total protein occupied by ribosomes, we multiplied ribosome mass concentrations with the mass fraction of ribosomes that is protein ($r_P = 0.58$ [16]), and divided the result by the total protein mass concentration $P = 127.4$ g/l in *E. coli*, assumed to be constant across growth conditions [16].

The concentration of actively translating ribosomes was determined based on total ribosome concentration and the fraction of active ribosome at different growth rates. The latter was estimated by fitting a smooth saturation function $s(\mu) = \mu/(\mu + z)$ over the fractions of active ribosomes estimated in Ref. [17], with the best-fitting parameter $z = 0.124/h$.

We set the Michaelis constant of the ribosome to $K_m = 3 \times 10^{-6}$ mol/l, based on the diffusion limit for ternary complexes calculated in Ref. [36]. We set the ribosome's turnover number to $k_{cat} = 22$ AA/s, the highest elongation rate observed experimentally in Ref. [37]. As we do not distinguish between different ternary complexes and the ribosome only accepts one of the 40 different ternary complex types at any given time, K_m was multiplied by 40 [36]. For consistency of the units with the mass concentration units used throughout our paper, the kinetic parameters had to be converted from molar to mass concentrations. The mean weight ($\pm$ SD) of amino acids across all conditions assayed in Ref. [16] was 132.60 ± 0.09 Da; the ribosome molecular weight is $2,306,967$ Da; and the mean weight of ternary complexes is $69,167 \pm 1,351$ g/mol. With these numbers, we obtain $k_{cat} = 22$ AA/s $\times$ (132.60 Da/AA)/(2,306,967 Da) $\times 3,600$s/1h $= 4.55$/h, and $K_m = 40 \times 3 \times 10^{-6}$mol/l $\times 69,167$ g/mol $= 8.30$ g/l.

F. Minimal whole-cell model and the dependence of maximal growth rate on cellular water content

Cayley *et al.* [21, 38] showed that the internal water content of *E. coli* cells increases when these are grown in environments with reduced osmolarity. This effect corresponds to a decrease of cellular dry weight per volume, ρ, by $\delta\rho$. η_ρ quantifies the associated reduction in relative fitness, $\delta f = \delta\mu^*/\mu^* = \eta_\rho\delta\rho$, with μ^* the maximal growth rate (Eq.(25)). The relative change in the maxi-

mal growth rate as per relative change in ρ is then

$$\frac{\rho}{\mu^*}\frac{d\mu^*}{d\rho} = \frac{d\ln\mu^*}{d\ln\rho} = \rho\ \eta_\rho \tag{S35}$$

From Eq.(26), we know that $\kappa_P \eta_\rho = \eta_P$; if there are no dependent reactants for P (i.e., $\forall_\gamma\ D_{\gamma P} = 0$), this simplifies to

$$\eta_\rho = \eta_P^0 = \frac{1}{P} - \sum_j q_P^j \quad , \tag{S36}$$

and thus

$$\frac{\rho}{\mu^*}\frac{d\mu^*}{d\rho} = \rho\ \eta_\rho = \rho\left(\frac{1}{P} - \sum_j q_P^j\right) \quad . \tag{S37}$$

The mass fraction of total protein in cell dry weight P/ρ has been shown to be approximately constant and equal to 0.54 across growth conditions that result in intermediate or high growth rates [21]. To estimate the total protein production cost $\sum_j q_P^j$, we consider the simplest possible whole-cell model, comprising only a transport reaction and the ribosome reaction (Fig.S1).

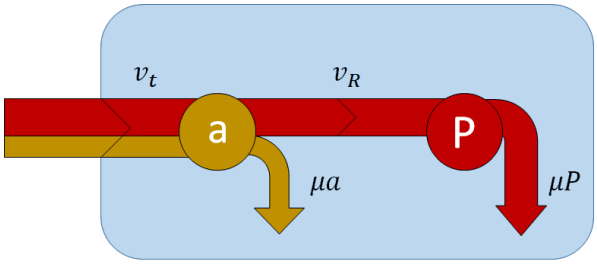

FIG. S1. Minimal whole-cell model, comprising a transport reaction (with rate v_t) and the ribosome reaction (with rate v_R).

The active stoichiometric matrix A of this model and its inverse $I = A^{-1}$ are

$$A = \begin{array}{c} \\ a \\ P \end{array}\begin{array}{c} \begin{array}{cc} t & R \end{array} \\ \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \end{array} \quad , \quad I = \begin{array}{c} \\ t \\ R \end{array}\begin{array}{c} \begin{array}{cc} a & P \end{array} \\ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \end{array} \quad . \tag{S38}$$

The capacity is determined only by its two components,

$$\rho = P + a \quad , \tag{S39}$$

where

$$P = p_t + p_R \quad . \tag{S40}$$

From the inverse I and Eq.(9), we obtain

$$v_t = \mu(P + a) = \mu\rho \tag{S41}$$

$$v_R = \mu P \quad . \tag{S42}$$

From the inverse I and Eq.(4), we get

$$\sum_j q_P^j = \frac{1}{P}\left(\frac{\mu}{k_t} + \frac{\mu}{k_R}\right) = \frac{1}{P}\left(\frac{\mu p_t}{v_t} + \frac{\mu p_R}{v_R}\right) \tag{S43}$$

Combining this with Eq.(S41) and (S42) and with $\phi_R = p_R/P$ and $\phi_t = p_t/P = 1 - \phi_R$ results in

$$\begin{aligned} \sum_j q_P^j &= \frac{1}{P}\left(\frac{\mu p_t}{\mu\rho} + \frac{\mu p_R}{\mu P}\right) \\ &= \frac{(1-\phi_R)}{\rho} + \frac{\phi_R}{P} \quad . \end{aligned} \tag{S44}$$

Combining this equation with Eq.(S37), we obtain

$$\begin{aligned} \rho\ \eta_\rho &= \rho\left(\frac{1}{P} - \frac{(1-\phi_R)}{\rho} - \frac{\phi_R}{P}\right) \\ &= \frac{\rho}{P} - 1 + \phi_R - \frac{\rho}{P}\phi_R \\ &= \left(\frac{\rho}{P} - 1\right)(1 - \phi_R) \quad . \end{aligned} \tag{S45}$$

From Eq.(27), we estimate the mass fraction of ribosomal proteins in total protein ϕ_R at $\mu = 1.0/h$ (growth rate in the reference growth condition of osmolarity Osm = 0.28 in Cayley *et al.* [38]) to be $\phi_R = 0.19$. Substituting this value into Eq.(S45) together with $P/\rho = 0.54$, we estimate the relative change in the maximal growth rate per relative change in ρ as

$$\rho\ \eta_\rho = 0.69 \quad . \tag{S46}$$

Cayley *et al.* [38] report cell growth at reduced osmolarities, summarized in Table S1. The cell free water content $\overline{V}_{free}$ in Table S1 is calculated from the total cell water $\overline{V}_{cell}$ minus the observed constant bound water $\overline{V}_b = 0.40 \pm 0.04$ ml/gCDW [21]. Errors are estimated standard deviations based on error propagation among normally distributed random variables.

Figure S2 plots the natural logarithms of μ and ρ. We estimate the slope of 0.66 at $\mu = 1.00/h$ with a linear regression over these points. This experimental value is close to our estimate of $\rho\ \eta_\rho = 0.69$.

Growth Osmolarity (Osm)	$\overline{V}_{free}$ (ml/gCDW)	ρ (gCDW/ml)	μ (1/h)
0.03	2.56 ± 0.10	0.39 ± 0.02	0.84 ± 0.07
0.10	2.12 ± 0.08	0.47 ± 0.02	0.91 ± 0.04
0.28	2.05 ± 0.11	0.49 ± 0.03	1.00 ± 0.10

TABLE S1. Experimental data from Cayley *et al.* [38], including cell water content $\overline{V}_{cell}$ and growth rate μ across different external osmolarities, together with the respective cellular dry weight per cellular free water volume, calculated as $\rho = 1/\overline{V}_{cell}$.

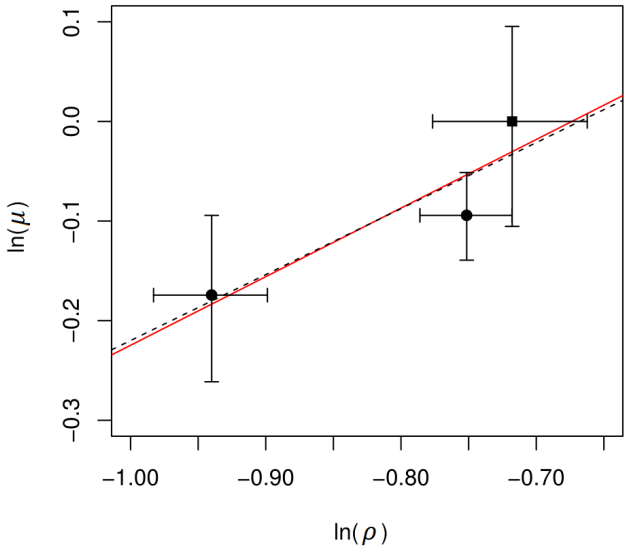

FIG. S2. Dependence of the growth rate ($\ln \mu$) on dry weight per free water volume ($\ln \rho$) in *E. coli* grown at different external osmolarities [38]. The square (■) indicates the usual environmental conditions, which correspond to the maximal growth rate; dots (●) indicate growth at lower osmolarities. The dotted line indicates the linear regression with slope = 0.66. The red line indicates the predicted slope = 0.69, drawn through the center of gravity of the 3 data points. Error bars are based on the reported experimental standard deviations.

Symbol	Definition (units)
A	Active matrix [mass fraction]
B	Basis matrix [mass fraction]
C	Closure matrix [mass fraction]
D	Dependence matrix
I	Investment matrix
v	Reaction rate [mass][volume]$^{-1}$[time]$^{-1}$
μ	Growth rate [time]$^{-1}$
P	Total protein concentration [mass][volume]$^{-1}$
a	Reactant concentration [mass][volume]$^{-1}$
b	Basis reactant concentration [mass][volume]$^{-1}$
c	Dependent reactant concentration [mass][volume]$^{-1}$
α	Reactant index
β	Basis reactant index
γ	Dependent reactant index
j	Reaction index
i	Protein and active reactant index ($\{P, \beta\}$)
k	Kinetic function (in units of k_{cat})
k_{cat}	Turnover number [time]$^{-1}$
K_m	Michaelis constant [mass][volume]$^{-1}$
ρ	Cellular dry weight per volume [mass][volume]$^{-1}$
f	Fitness
η_i^0	Direct marginal net benefit of i [volume][mass]$^{-1}$
η_i	Marginal net benefit of i [volume][mass]$^{-1}$
η_ρ	Marginal benefit of of the cellular capacity [volume][mass]$^{-1}$
η_γ^c	Marginal net benefit of γ [volume][mass]$^{-1}$
q_i^j	Marginal production cost of i relative to j [volume][mass]$^{-1}$
u_β^j	Marginal kinetic benefit of β relative to j [volume][mass]$^{-1}$
u_γ^j	Marginal kinetic benefit of γ relative to j [volume][mass]$^{-1}$
κ	Capacity factor
$\mathcal{L}$	Lagrangian
λ_ρ	Lagrange multiplier of the capacity constraint

TABLE S2. Symbols and definitions. We note that for simplicity we also use P as an index for total protein, and ρ as an index for cellular dry weight per volume.

[1] R. A. Fisher and J. H. Bennett. *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press, 1999.

[2] Rafael U. Ibarra, Jeremy S. Edwards, and Bernhard O. Palsson. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420(6912):186–189, nov 2002.

[3] 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*, 8:14123, 2017.

[4] A. Campbell. Synchronization of cell division. *Bacteriol Rev*, 21(4):263–272, Dec 1957. 13488884[pmid].

[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):323, 2009.

[6] Arijit Maitra and Ken A. Dill. Bacterial growth laws reflect the evolutionary importance of energy efficiency. *Proceedings of the National Academy of Sciences*, 112(2):406–411, 2015.

[7] Nils Giordano, Francis Mairet, Jean-Luc Gouzé, Johannes Geiselmann, and Hidde de Jong. Dynamical Allocation of Cellular Resources as an Optimal Control Problem: Novel Insights into Microbial Growth Strategies. *PLOS Computational Biology*, 12(3):e1004802, mar 2016.

[8] Moshe Kafri, Eyal Metzl-Raz, Ghil Jona, and Naama Barkai. The Cost of Protein Production. *Cell Reports*, 14(1):22–31, jan 2016.

[9] Nathan E. Lewis, Harish Nagarajan, and Bernhard O. Palsson. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4):291–305, 2012.

[10] Anne Goelzer, Jan Muntel, Victor Chubukov, Matthieu Jules, Eric Prestel, Rolf Nölker, Mahendra Mariadassou, Stéphane Aymerich, Michael Hecker, Philippe Noirot, Dörte Becher, and Vincent Fromion. Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic Engineering*, 32:232–243, nov 2015.

[11] Edward J. O'Brien, Joshua A. Lerman, Roger L. Chang, Daniel R. Hyduke, and Bernhard Palsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 2013.

[12] Andrea Y. Weiße, Diego A. Oyarzún, Vincent Danos, and Peter S. Swain. Mechanistic links between cellular tradeoffs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9):E1038–E1047, 2015.

[13] M. R. Watson. Metabolic maps for the Apple II. *Biochemical Society Transactions*, 12(6):1093–1094, dec 1984.

[14] Hidde De Jong, Stefano Casagranda, Nils Giordano, Eugenio Cinquemani, Delphine Ropers, Johannes Geiselmann, and Jean Luc Gouzé. Mathematical modelling of microbes: Metabolism, gene expression and growth. *Journal of the Royal Society Interface*, 14(136):20170502, 2017.

[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(6007):1099–1102, 2010.

[16] Alexander Schmidt, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer, Luciano Callipo, Kèvin Knoops, Manuel Bauer, Ruedi Aebersold, and Matthias Heinemann. The quantitative and condition-dependent escherichia coli proteome. *Nature Biotechnology*, 34:104 EP –, Dec 2015.

[17] Xiongfeng Dai, Manlu Zhu, Mya Warren, Rohan Balakrishnan, Vadim Patsalo, Hiroyuki Okano, James R. Williamson, Kurt Fredrick, Yi-Ping Wang, and Terence Hwa. Reduction of translating ribosomes enables escherichia coli to maintain elongation rates during slow growth. *Nature Microbiology*, 2:16231 EP –, Dec 2016. Article.

[18] Hugo Dourado, Veronica G. Maurino, and Martin J. Lercher. Enzymes and substrates are balanced at minimal combined mass concentration in vivo. *bioRxiv*, 2017.

[19] Reinhart Heinrich and Stefan Schuster. *The Regulation of Cellular Systems*. Springer, 2011.

[20] Daniel E. Atkinson. Limitation of Metabolite Concentrations and the Conservation of Solvent Capacity in the Living Cell. *Current Topics in Cellular Regulation*, 1(C):29–43, 1969.

[21] Scott Cayley, Barbara A. Lewis, Harry J. Guttman, and M.Thomas Record. Characterization of the cytoplasm of escherichia coli k-12 as a function of external osmolarity: Implications for protein-dna interactions in vivo. *Journal of Molecular Biology*, 222(2):281 – 300, 1991.

[22] W. W. Baldwin, Richard Myer, Nicole Powell, Erika Anderson, and Arthur L. Koch. Buoyant density of escherichia coli is determined solely by the osmolarity of the culture medium. *Archives of Microbiology*, 164(2):155–157, Aug 1995.

[23] Markus Basan, Manlu Zhu, Xiongfeng Dai, Mya Warren, Daniel Svin, Yi-Ping Wang, and Terence Hwa. Inflating bacterial cells by increased protein synthesis. *Molecular Systems Biology*, 11(10):836, 2015.

[24] Kai Zhuang, Goutham N Vemuri, and Radhakrishnan Mahadevan. Economics of membrane occupancy and respiro-fermentation. *Molecular Systems Biology*, 7(1), 2011.

[25] Stefan Klumpp, Zhongge Zhang, and Terence Hwa. Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7):1366–1375, Dec 2009.

[26] Meike T. Wortel, Han Peters, Josephus Hulshof, Bas Teusink, and Frank J. Bruggeman. Metabolic states with maximal specific rate carry flux through an elementary flux mode. *The FEBS Journal*, 281(6):1547–1555, 2014.

[27] Stefan Müller, Georg Regensburger, and Ralf Steuer. Enzyme allocation problems in kinetic metabolic networks: Optimal solutions are elementary flux modes. *Journal of Theoretical Biology*, 347:182 – 190, 2014.

[28] Stefan Schuster and Claus Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 02(02):165–182, 1994.

[29] Julien Gagneur and Steffen Klamt. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, 5(1):175, 2004.

[30] Daan H. de Groot, Josephus Hulshof, Bas Teusink, Frank J. Bruggeman, and Robert Planqu. Elementary growth modes provide a molecular description of cellular self-fabrication. *bioRxiv (submitted 12. April 2019)*,

2019.

[31] Christine Reder. Metabolic control theory: A structural approach. *Journal of Theoretical Biology*, 135(2):175 – 201, 1988.

[32] Erez Dekel and Uri Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050):588–592, 2005.

[33] 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):1–29, 11 2016.

[34] S. Afriat. Theory of maxima and the method of lagrange. *SIAM Journal on Applied Mathematics*, 20(3):343–357, 1971.

[35] Avlant Nilsson, Jens Nielsen, and Bernhard O. Palsson. Metabolic models of protein allocation call for the kinetome. *Cell Systems*, 5(6):538–541, Dec 2017.

[36] Stefan Klumpp, Matthew Scott, Steen Pedersen, and Terence Hwa. Molecular crowding limits translation and cell growth. *Proceedings of the National Academy of Sciences*, 110(42):16754–16759, 2013.

[37] Dennis P Bremer H. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 2008.

[38] D. Scott Cayley, Harry J. Guttman, and M. Thomas Record. Biophysical characterization of changes in amounts and activity of escherichia coli cell and compartment water and turgor pressure in response to osmotic stress. *Biophysical Journal*, 78(4):1748 – 1764, 2000.

[39] S. Hui, J. M. Silverman, S. S. Chen, D. W. Erickson, M. Basan, J. Wang, T. Hwa, and J. R. Williamson. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular Systems Biology*, 11(2):e784–e784, 2015.

[40] David Heckmann, Colton J. Lloyd, Nathan Mih, Yuanchi Ha, Daniel C. Zielinski, Zachary B. Haiman, Abdelmoneim Amer Desouki, Martin J. Lercher, and Bernhard O. Palsson. Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nature Communications*, 9(1):5252, dec 2018.

[41] Hans-Michael Kaltenbach and Jörg Stelling. Modular Analysis of Biological Networks. In Bjrn H. Junker and Falk Schreiber, editors, *Analysis of Biological Networks*, volume 736 of *Wiley Series on Bioinformatics: Computational Techniques and Engineering*, pages 3–17. John Wiley & Sons, Inc., Hoboken, NJ, USA, feb 2008.

[42] Tomer Benyamini, Ori Folger, Eytan Ruppin, and Tomer Shlomi. Flux balance analysis accounting for metabolite dilution. *Genome Biology*, 11(4):R43, 2010.

[43] Daan H. de Groot, Coco van Boxtel, Robert Planqu, Frank J. Bruggeman, and Bas Teusink. The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates. *PLOS Computational Biology*, 15(3):1–24, 03 2019.

[44] Markus Basan, Sheng Hui, Hiroyuki Okano, Zhongge Zhang, Yang Shen, James R. Williamson, and Terence Hwa. Overflow metabolism in escherichia coli results from efficient proteome allocation. *Nature*, 528:99 EP –, Dec 2015. Article.

[45] B Lee. Calculation of volume fluctuation for globular protein models. *Proceedings of the National Academy of Sciences*, 80(2):622–626, 1983.

[46] Arren Bar-Even, Elad Noor, Avi Flamholz, Joerg M. Buescher, and Ron Milo. Hydrophobicity and charge shape cellular metabolite concentrations. *PLOS Computational Biology*, 7(10):1–7, 10 2011.

[47] H. Allen Orr. Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*, 10:531 EP –, Aug 2009. Review Article.

[48] Chevin Luis-Miguel. On measuring selection in experimental evolution. *Biology Letters*, 7(2):210–213, Apr 2011.

[49] Alberto Santos-Zavaleta, Alexander G. Shearer, Amanda Mackie, Anamika Kothari, Araceli M. Huerta, Carol Fulcher, Csar Bonavides-Martnez, Deepika Weerasinghe, Ian Paulsen, Imke Schrder, Julio Collado-Vides, Luis Muiz-Rascado, Mario Latendresse, Markus Krummenacker, Martin Peralta-Gil, Mike Travers, Pallavi Subhraveti, Quang Ong, Robert P. Gunsalus, Socorro Gama-Castro, Suzanne Paley, Verena Weiss, Ingrid M. Keseler, and Peter D. Karp. EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Research*, 41(D1):D605–D612, 11 2012.