An analytical theory of balanced cellular growth

Hugo Dourado[1] **and Martin J. Lercher**[1,*]

[1]Institute for Computer Science & Department of Biology, Heinrich Heine University, 40221 Düsseldorf, Germany
[*]to whom correspondence should be addressed: martin.lercher@hhu.de

ABSTRACT

The biological fitness of unicellular organisms is largely determined by their balanced growth rate, i.e., by the rate with which they replicate their biomass composition. Natural selection on this growth rate occurred under a set of physicochemical constraints, including mass conservation, reaction kinetics, and limits on dry mass per volume; mathematical models that maximize the balanced growth rate while accounting explicitly for these constraints are inevitably nonlinear and have been restricted to small, non-realistic systems. Here, we lay down a general theory of balanced growth states, providing explicit expressions for protein concentrations, fluxes, and the growth rate. These variables are functions of the concentrations of cellular components, for which we calculate marginal fitness costs and benefits that can be related to metabolic control coefficients. At maximal growth rate, the net benefits of all concentrations are equal. Based solely on physicochemical constraints, the growth balance analysis (GBA) framework introduced here unveils fundamental quantitative principles of cellular growth and leads to experimentally testable predictions.

Introduction

The defining feature of life 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, we expect that natural selection favoring fast growth in specific environments has played an important role in shaping the physiology of many microbial organisms[2,3].

Conceptually, we can envision a bacterial cell as a volume enclosed by a membrane, filled with a solution of metabolites and of the proteins and nucleic acids that catalyze their conversion into biomass. A state of the cell is characterized by the molecular concentrations, which in turn determine the fluxes of the biochemical reactions. The boundary conditions limiting the concentrations and fluxes are provided by the environment and by physicochemical constraints. Cellular growth has to be balanced over the cell cycle, i.e., all cellular components must be reproduced in proportion to their abundances[4]. Casting these constraints into a mathematical model and characterizing states of optimal growth may provide a detailed understanding of central aspects of bacterial physiology[3,5–8].

Currently, the most popular method to model the physiology of whole cells is flux balance analysis (FBA)[9,10]. FBA maximizes the production rate of a constant biomass concentration vector while balancing the fluxes producing and consuming internal metabolites to account for mass conservation. All constraints in FBA are linear (Fig. 1). The resulting computational efficiency comes at the price of ignoring reactant concentrations, and hence FBA cannot account for reaction kinetics and the resulting demand for catalytic proteins from first principles. Instead, extensions of FBA that consider enzyme concentrations rely on phenomenological kinetic functions that are assumed to be either constant (FBA with molecular crowding[11], metabolism and expression models[12]) or linear functions of the growth rate (resource balance analysis[13]).

Molenaar et al.[5] proposed a small, coarse-grained model of balanced growth with explicit non-linear reaction kinetics. Numerical growth rate optimization predicted qualitatively the growth-rate dependencies of cellular ribosome content, cell size, and the emergence of overflow metabolism. No extensions of this approach to models accounting for more than seven cellular reactions have been proposed, likely because of its inherent nonlinearity. Instead, "toy models" of 1-3 reactions were solved analytically to gain further qualitative understanding of systems-level effects[3,6–8,14], including optimal gene regulation strategies[3,7].

We term this general modeling scheme *Growth Balance Analysis* (GBA); below, we develop an

analytical theory for GBA of arbitrarily complex cellular systems. FBA and its extensions can be viewed as linearizations of the GBA scheme[15]. Fig. 1 shows a schematic comparison of FBA and GBA. While FBA predicts a linear dependence of maximal growth rate on nutrient uptake fluxes, GBA leads to a non-linear (Monod-type) dependence on nutrient concentrations.

Results

Modeling balanced exponential growth

Our model assumes that the cell increases exponentially in size, while the concentrations of all cellular components (including the number of membrane constituents per cell volume) remain constant[5]. We do not explicitly model cell division; thus, our model can also be interpreted as describing the growth of a population of cells[7]. In balanced growth, the net production rate of each molecular constituent must

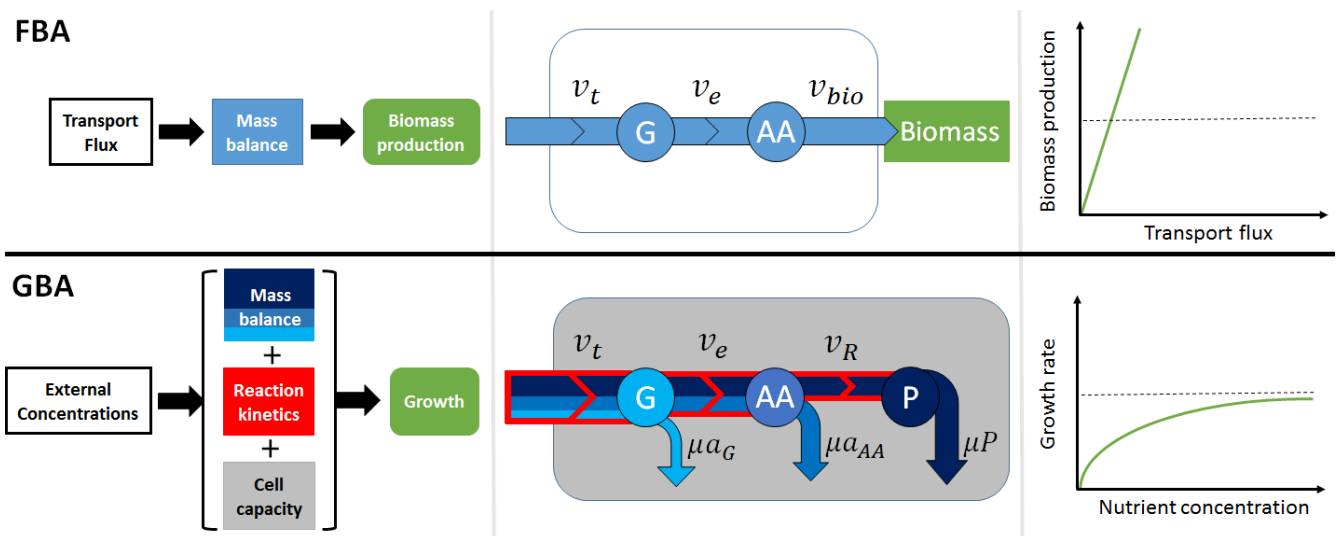

Figure 1. A schematic comparison of flux balance analysis (FBA, top) and growth balance analysis (GBA, bottom) for a simple toy model. A nutrient G is taken up through a transporter t at rate v_t and is then converted by an enzyme e with rate v_e into a precursor for protein synthesis, AA. In FBA, AA is equated with the biomass, the production of which is maximized while enforcing the stationarity of internal concentrations (blue); this leads to a linear dependence of growth rate on uptake flux. In GBA, AA is converted further into total protein P by a ribosome R, where P represents the sum of the three proteins t, e, R. GBA maximizes the balanced reproduction of the cellular composition with growth (blue), constrained by non-linear reaction kinetics (red) and cellular capacity (dry mass per volume, grey); this leads to a non-linear dependence of growth rate on nutrient concentrations.

balance its dilution by growth, $\frac{dx}{dt} = \mu x$, where x denotes the concentration of a given component and μ is the cellular growth rate[5,7].

The mass conservation in chemical reaction networks is commonly described through a *stoichiometric matrix* N, where rows correspond to metabolites and each column describes the mass balance of one reaction[16]. Here, we focus on matrices A 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. A also includes a "ribosome" reaction to produce catalytic proteins, encompassing enzymes, transporters, and the ribosome itself. We express concentrations as mass concentrations (mass per volume); accordingly, the entries of A are not stoichiometric coefficients but are mass fractions. The mass conservation of each component can then be stated as

$$A\boldsymbol{v} = \mu \begin{bmatrix} P \\ \boldsymbol{a} \end{bmatrix} \quad , \tag{1}$$

where $\boldsymbol{v}$ is the flux vector (in units of [mass]/[volume]/[time]), $\boldsymbol{a}$ is the vector of reactant mass concentrations a_α, and P is the sum of the mass concentrations p_j of all proteins $j \in \{1, \ldots, n\}$,

$$P = \sum_j p_j \quad . \tag{2}$$

The first row of A describes the net production of total protein P, which is then distributed among the individual proteins j. The remaining rows describe the net production of the reactants α.

Each reaction rate v_j is the product of the concentration of its catalyzing protein p_j and a kinetic function $k_j(\boldsymbol{a})$ that depends on the reactant concentrations a_α,

$$v_j = p_j k_j(\boldsymbol{a}) \quad . \tag{3}$$

We assume that the functional form and kinetic parameters of $k_j(\boldsymbol{a})$ are known. $k_j(\boldsymbol{a})$ may depend on the mass concentrations of substrates, products, and other molecules a_α acting as inhibitors or activators, and accounts for the system's thermodynamics. The activity of all reactions j ($v_j \neq 0$) implies $p_j > 0$ and $k_j(\boldsymbol{a}) \neq 0$.

For a given concentration vector $\boldsymbol{x} \equiv [P, \boldsymbol{a}]^T$, we define a *balanced growth state* (BGS) as a cellular state (characterized by its flux vector $\boldsymbol{v}$) that satisfies constraints (1), (2), and (3). The set of all such states forms the solution space of balanced growth. On the following pages, we first develop a framework for GBA by characterizing BGSs at a known concentration vector $\boldsymbol{x}$. Formal definitions and theorems are detailed in **SI text A.**; Table S2 lists the symbols used.

Cellular state defined by the concentration variables

We define an *elementary growth state* (EGS) as a BGS $\boldsymbol{v}$ that also represents an elementary flux mode[17] of a linearized problem (**SI text A.**, Def. 3). We can express any BGS as a weighted average of EGSs at the same concentration vector $\boldsymbol{x} = [P, \boldsymbol{a}]^T$ (Theorem 3). Moreover, any optimal BGS under a single capacity constraint (see below) is also an EGS (Theorem 9; see also Ref.[18]). Thus, without loss of generality, we focus on EGSs from here on.

If A is the active stoichiometric matrix of an EGS, it has full column rank (Theorem 4; see also Ref.[18]). A may have more rows than columns, in which case some reactant concentrations are linearly dependent on other concentrations[19]. These dependent concentrations are not free variables, and hence they can be put aside and dealt with separately. For clarity of presentation, we here present only the case without dependent reactants; the generalization can be treated similarly and is detailed in **SI text A.**. Without dependent reactants, A is a square matrix with a unique inverse $I \equiv A^{-1}$. Multiplying both sides of the mass balance constraint (1) by I, we obtain (Theorem 5)

$$\boldsymbol{v} = \mu I \begin{bmatrix} P \\ \boldsymbol{a} \end{bmatrix} \quad . \tag{4}$$

The right hand side of the mass balance constraint (1) quantifies how much of each component needs to be produced to offset the dilution that would otherwise occur through the exponential volume increase. I_{ji} quantifies the proportion of flux v_j invested into offsetting the dilution of component i, and we thus name I the *investment* (or dilution) matrix; see Fig. S1 for examples. In contrast to the stoichiometric matrix A, which describes local mass balances, I describes the structural allocation of reaction fluxes into offsetting the dilution of all downstream cellular components, carrying global, systems-level information.

From the kinetic equations (3), $p_j = v_j / k_j(\boldsymbol{a})$, and inserting v_j from the investment equation (4) gives

$$p_j = \mu \frac{I_{jP} P + \sum_\alpha I_{j\alpha} a_\alpha}{k_j(\boldsymbol{a})} \quad , \tag{5}$$

where $\sum_\alpha$ sums over the set of all reactants (denoted by $\{\alpha\}$; Theorem 6). Substituting these expressions into the total protein sum (Eq. (2)) and solving for μ results in the *growth equation* (Theorem 7)

$$\mu(P, \boldsymbol{a}) = \frac{P}{\sum_j \frac{I_{jP} P + \sum_\alpha I_{j\alpha} a_\alpha}{k_j(\boldsymbol{a})}} \quad . \tag{6}$$

Thus, for any EGS and concentration vector $\boldsymbol{x}$, there are unique and explicit mathematical solutions for $\boldsymbol{v}$, $\boldsymbol{p}$, and μ. If μ and all individual protein concentrations p_j in Eq. (5) are positive, the cellular state is a BGS; otherwise, no balanced growth is possible at these concentrations.

Marginal fitness contributions of cellular concentrations

We now use these relationships to calculate the costs and benefits of concentration changes, which are naturally expressed in terms of relative fitness effects. If fitness is determined predominantly by growth rate[1] (**SI text B.**), we can we define the *marginal net benefit* η_i of concentration x_i ($i \in \{P, \alpha\}$) as the relative change in growth rate[20] due to a small change in x_i (Def. 4),

$$\eta_i \equiv \frac{1}{\mu}\frac{\partial \mu}{\partial x_i} \tag{7}$$

To aid in the interpretation of η_i below, we define the *marginal production cost* incurred by the system via protein j as a consequence of increasing concentration x_i at fixed growth rate μ and kinetics k_j,

$$q_i^j \equiv \frac{1}{P}\left(\frac{\partial p_j}{\partial x_i}\right)_{\mu, k_j = \text{const.}} = \frac{\mu I_{ji}}{P k_j} \quad ,$$

where the second equality arises from the growth equation (6). q_i^j quantifies the proportional increase of p_j to help offset the increased dilution of component i. Thus, q_i^j is related to the *protein control coefficient* from metabolic control analysis (MCA); **SI text F.** summarizes the relationship between GBA and MCA[21–23]. From the growth equation (6), it further follows that

$$\eta_P = \frac{1}{P} - \sum_j q_P^j$$

$$\eta_\alpha = \sum_j (u_\alpha^j - q_\alpha^j)$$

with

$$u_\alpha^j \equiv -\frac{1}{P}\left(\frac{\partial p_j}{\partial a_\alpha}\right)_{v_j = \text{const.}} = \frac{p_j}{P}\frac{1}{k_j}\frac{\partial k_j}{\partial a_\alpha}$$

(Theorem 8). u_α^j is the *marginal kinetic benefit* of reactant α to reaction j and quantifies the proportion of protein p_j "saved" due to the change in kinetics associated with an increase in a_α[24]. It relates directly to the *elasticity coefficients* from MCA (**SI text F.**). The kinetic benefit is nonzero only for reactants that directly affect the kinetics of reaction j, making it a purely local effect. Because fluxes are proportional to the concentrations of the catalyzing proteins, the marginal kinetic benefit of total protein is simply P^{-1}. Overall, the net benefit of component i via reaction j is the reduction of the protein fraction p_j/P at constant μ facilitated by the increase in x_i,

$$\eta_i = -\frac{1}{P}\sum_j \left(\frac{\partial p_j}{\partial x_i}\right)_{\mu = \text{const.}} \quad . \tag{8}$$

This result provides a formal justification for the notion that cellular costs lie predominantly in protein production[3,5–8,12–14,24–27].

Optimal growth and the balance of marginal net benefits

Up to this point, we kept $\boldsymbol{x} = [P, \boldsymbol{a}]^T$ fixed. We will now characterize optimal growth states, i.e., BGSs with maximal growth rate across all allowed concentration vectors. To make this problem well defined, we need to consider an additional constraint that reflects the cellular requirement for a minimal amount of free water to facilitate diffusion[28,29]. We implement this constraint by assuming that cellular dry weight per volume is limited to a maximal density ρ, where ρ is determined by external osmolarity[29,30] but is otherwise constant across growth conditions[31],

$$\rho \geq P + \sum_{\alpha} a_{\alpha} \quad . \tag{9}$$

A BGS is a *capacity-constrained balanced growth state* (cBGS) if it additionally satisfies constraint (9). At maximal growth rate, the cellular components will utilize the full cellular capacity to saturate enzymes with their substrates, and thus the inequality in Eq. (9) becomes an equality.

The maximal obtainable balanced growth rate μ^* will be a function of 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,

$$\eta_{\rho} \equiv \frac{1}{\mu^*} \frac{d\mu^*}{d\rho} \quad .$$

Using the method of Lagrange multipliers with the growth equation (6) as the objective function, we derive necessary conditions at optimal growth, which we term *balance equations*:

$$\forall_{i \in \{P, \alpha\}} \quad \eta_i = \eta_{\rho} \tag{10}$$

(Theorem 10). The optimal state is perfectly balanced: the marginal net benefits of all cellular concentrations x_i are identical. If the dry weight density ρ could increase by a small amount (such as 1 mg/l), then the marginal fitness gain that could be achieved by increasing protein concentration by this amount is identical to that achieved by increasing the concentration of any reactant α by the same amount. This should not be surprising: if the marginal net benefit of concentration x_i was higher than that of $x_{i'}$, growth could be accelerated by increasing x_i at the expense of $x_{i'}$.

Eq. (10) together with Eq. (9) describes a system of $n+1$ equations for $n+1$ unknowns. In realistic cellular systems, this set of equations has a finite number of discrete solutions. Thus, growth rate optimization can be replaced by searching for the solution of the balance equations. If the optimization problem is convex, the conditions given by Eq. (10) are necessary and sufficient, and the solution is unique.

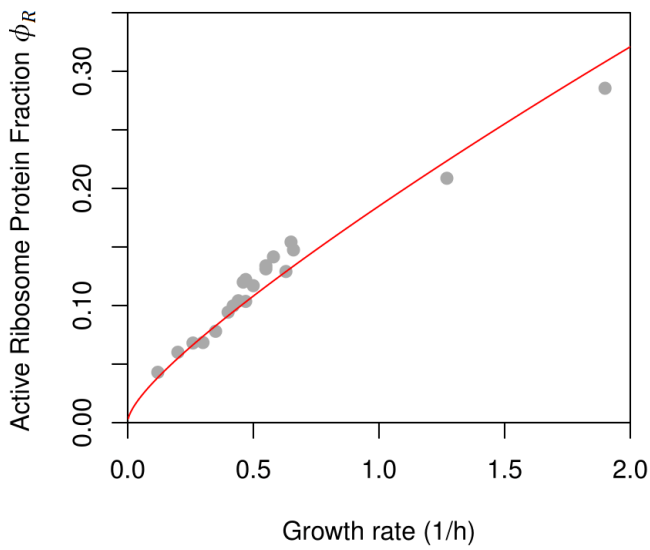

Figure 2. The GBA prediction of the active ribosomal protein fraction in *E. coli* agrees with experimental values. Comparison of GBA predictions (red line, no free parameters) and data across 20 different growth conditions[32,33] (grey dots) results in a Pearson correlation coefficient $R^2 = 0.96$ ($P < 10^{-13}$).

Quantitative predictions

If a substrate i is consumed only by a single reaction that produces i', the non-local dilution terms in the balance equation $\eta_i = \eta_{i'}$ cancel, and we are left with a local problem for which only the kinetic benefits of x_i and $x_{i'}$ must be considered. This is the case for protein production in simplified models where the ribosome (R) produces proteins from a single substrate, a generic ternary complex (T)[26]. In such models, we can calculate the optimal protein fraction of actively translating ribosomes, $\phi_R = p_R/P$ from the balance equation $\eta_T = \eta_P$ (**SI text C.**), which agrees quantitatively with experimental values[32,33] (Fig. 2).

An approximation that ignores the dilution of intermediates and hence production costs ($q_i^j \approx 0$) results in less accurate predictions especially at high growth rates (Fig. S4). In growth on minimal media ($\mu < 1$), the dilution of intermediates μa_α becomes less important. This explains why the relationship between the concentrations of a substrate and its catalysts is well approximated in this regime through minimizing their joint utilization of cellular capacity[34].

To obtain a rough quantitative estimate of the marginal net benefits η_i, we here consider the simplest model of a complete cell, consisting of only a transport protein and the ribosome[3,6] (Fig. S2). Based on

the experimentally observed protein fraction of total dry weight in *E. coli*, $P/\rho = 0.54$,[29] we estimate $\frac{\rho}{\mu}\frac{d\mu}{d\rho} = \rho\ \eta_\rho = 0.69$ (**SI text D.**). Thus, a decrease in cellular dry weight density ρ of 1% would lead to a 0.69% decrease in growth rate, emphasizing the biological significance of the capacity constraint.

The cellular capacity ρ changes when external osmolarity is modified[29]. $\rho\ \eta_\rho = \frac{d\ln\mu}{d\ln\rho}$ is the slope of the log-scale plot of μ vs. ρ at different external osmolarities. While increases in ρ may have strong effects on diffusion and thus on enzyme kinetics, reductions in ρ due to decreased external osmolarity are within the scope of our model. The very limited available experimental data (three data points from Ref.[35], Fig. S3) suggests $\rho\ \eta_\rho \approx 0.66$, close to our rough estimate from the minimal cell model.

Discussion

Our derivations are based on the insight that for any EGS, the inverse I of the active stoichiometric matrix quantifies how individual fluxes offset the dilution of downstream cellular components by growth. These non-local, structural constraints arising from mass balance lead to an explicit dependence of reaction fluxes on the cellular concentrations (Eq. (4), Theorem 5). Independently of this, fluxes also depend on concentrations through reaction kinetics (constraint (3)). Combining these two relationships leads to explicit expressions for the individual protein concentrations p_j and for the growth rate μ as functions of arbitrary concentrations $[P, \boldsymbol{b}]^T$. As any BGS can be expressed as a weighted average of EGSs (Theorem 3), these results allow a general characterization of the solution space of balanced growth. Further, the growth rate equation (6) can be employed to calculate marginal fitness benefits of concentrations and to derive balance equations for marginal benefits at optimal growth.

Previous work has emphasized the central role of proteins in the cellular economy[3,5–8,12–14,24–27], and this is confirmed by Eq. (8). However, 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.

To make the presentation concise, our development of GBA assumes (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; (iv) that all catalysts are proteins; and (v) that cells are optimized for growth. As outlined in **SI text E.**, it is straight-forward to remove these simplifications.

In principle, exploitation of the balance equations (Eq. (10)) may allow the numerical optimization of cellular systems of realistic size, encompassing hundreds of protein and reactant species. One remaining obstacle to the accurate formulation of such models is the current incompleteness of the kinetic constants

needed to parameterize the functions $k_j(\boldsymbol{a})$. Until methods for the high-throughput ascertainment of kinetic parameters[36] are fully developed, artificial intelligence may provide reasonable approximations[37] for the required parameters. 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[27,38].

Our work extends previous ad-hoc optimizations of toy models[3,5–8] into a 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 rigorous 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.

References

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

2. Ibarra, R. U., Edwards, J. S. & Palsson, B. O. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420**, 186–189, DOI: 10.1038/nature01149 (2002).

3. Towbin, B. D. *et al.* Optimality and sub-optimality in a bacterial growth law. *Nat. Commun.* **8**, 14123, DOI: 10.1038/ncomms14123 (2017).

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

5. Molenaar, D., van Berlo, R., de Ridder, D. & Teusink, B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol. Syst. Biol.* **5**, 323, DOI: 10.1038/msb.2009.82 (2009).

6. Maitra, A. & Dill, K. A. Bacterial growth laws reflect the evolutionary importance of energy efficiency. *Proc. Natl. Acad. Sci.* **112**, 406–411, DOI: 10.1073/pnas.1421138111 (2015).

7. Giordano, N., Mairet, F., Gouzé, J.-L., Geiselmann, J. & de Jong, H. Dynamical Allocation of Cellular Resources as an Optimal Control Problem: Novel Insights into Microbial Growth Strategies. *PLOS Comput. Biol.* **12**, e1004802, DOI: 10.1371/journal.pcbi.1004802 (2016).

8. Kafri, M., Metzl-Raz, E., Jona, G. & Barkai, N. The Cost of Protein Production. *Cell Reports* **14**, 22–31, DOI: 10.1016/j.celrep.2015.12.015 (2016).

9. Watson, M. R. Metabolic maps for the Apple II. *Biochem. Soc. Transactions* **12**, 1093–1094, DOI: 10.1042/bst0121093 (1984).

10. Lewis, N. E., Nagarajan, H. & Palsson, B. O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* **10**, 291–305, DOI: 10.1038/nrmicro2737 (2012).

11. Beg, Q. K. *et al.* Intracellular crowding defines the mode and sequence of substrate uptake by escherichia coli and constrains its metabolic activity. *Proc. Natl. Acad. Sci.* **104**, 12663–12668, DOI: 10.1073/pnas.0609845104 (2007).

12. O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R. & Palsson, B. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* DOI: 10.1038/msb.2013.52 (2013).

13. Goelzer, A. *et al.* Quantitative prediction of genome-wide resource allocation in bacteria. *Metab. Eng.* **32**, 232–243, DOI: 10.1016/j.ymben.2015.10.003 (2015).

14. Weiße, A. Y., Oyarzún, D. A., Danos, V. & Swain, P. S. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci.* **112**, E1038–E1047, DOI: 10.1073/pnas.1416533112 (2015).

15. De Jong, H. *et al.* Mathematical modelling of microbes: Metabolism, gene expression and growth. *J. Royal Soc. Interface* **14**, 20170502, DOI: 10.1098/rsif.2017.0502 (2017).

16. Heinrich, R. & Schuster, S. *The Regulation of Cellular Systems* (Springer, 2011).

17. Schuster, S. & Hilgetag, C. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **02**, 165–182, DOI: 10.1142/S0218339094000131 (1994).

18. de Groot, D. H., Hulshof, J., Teusink, B., Bruggeman, F. J. & Planqué, R. Elementary growth modes provide a molecular description of cellular self-fabrication. *bioRxiv* DOI: 10.1101/608083 (2019).

19. Reder, C. Metabolic control theory: A structural approach. *J. Theor. Biol.* **135**, 175 – 201, DOI: https://doi.org/10.1016/S0022-5193(88)80073-0 (1988).

20. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588–592, DOI: 10.1038/nature03842 (2005).

21. Kleijn, I. T., Krah, L. H. J. & Hermsen, R. Noise propagation in an integrated model of bacterial gene expression and growth. *PLOS Comput. Biol.* **14**, 1–18, DOI: 10.1371/journal.pcbi.1006386 (2018).

22. Liebermeister, W. Optimal metabolic states in cells. *bioRxiv* DOI: 10.1101/483867 (2018).

23. Liebermeister, W. The value structure of metabolic states. *bioRxiv* DOI: 10.1101/483891 (2018).

24. Noor, E. *et al.* The protein cost of metabolic fluxes: Prediction from enzymatic rate laws and cost minimization. *PLOS Comput. Biol.* **12**, 1–29, DOI: 10.1371/journal.pcbi.1005167 (2016).

25. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: Origins and consequences. *Science* **330**, 1099–1102, DOI: 10.1126/science.1192588 (2010).

26. Klumpp, S., Scott, M., Pedersen, S. & Hwa, T. Molecular crowding limits translation and cell growth. *Proc. Natl. Acad. Sci.* **110**, 16754–16759, DOI: 10.1073/pnas.1310377110 (2013).

27. Hui, S. *et al.* Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.* **11**, e784–e784, DOI: 10.15252/msb.20145697 (2015).

28. Atkinson, D. E. Limitation of Metabolite Concentrations and the Conservation of Solvent Capacity in the Living Cell. *Curr. Top. Cell. Regul.* **1**, 29–43, DOI: 10.1016/B978-0-12-152801-0.50007-9 (1969).

29. Cayley, S., Lewis, B. A., Guttman, H. J. & Record, M. ...characterization of the cytoplasm of escherichia coli k-12 as a function of external osmolarity: Implications for protein-dna interactions in vivo. *J. Mol. Biol.* **222**, 281 – 300, DOI: https://doi.org/10.1016/0022-2836(91)90212-O (1991).

30. Baldwin, W. W., Myer, R., Nicole Powell, Anderson, E. & Koch, A. L. Buoyant density of escherichia coli is determined solely by the osmolarity of the culture medium. *Arch. Microbiol.* **164**, 155–157, DOI: 10.1007/s002030050248 (1995).

31. Basan, M. *et al.* Inflating bacterial cells by increased protein synthesis. *Mol. Syst. Biol.* **11**, 836, DOI: 10.15252/msb.20156178 (2015).

32. Schmidt, A. *et al.* The quantitative and condition-dependent escherichia coli proteome. *Nat. Biotechnol.* **34**, 104 EP – (2015).

33. Dai, X. *et al.* Reduction of translating ribosomes enables escherichia coli to maintain elongation rates during slow growth. *Nat. Microbiol.* **2**, 16231 EP – (2016). Article.

34. Dourado, H., Maurino, V. G. & Lercher, M. J. Enzymes and substrates are balanced at minimal combined mass concentration in vivo. *bioRxiv* DOI: 10.1101/128009 (2017).

35. Cayley, D. S., Guttman, H. J. & Record, M. T. Biophysical characterization of changes in amounts and activity of escherichia coli cell and compartment water and turgor pressure in response to osmotic stress. *Biophys. J.* **78**, 1748 – 1764, DOI: https://doi.org/10.1016/S0006-3495(00)76726-9 (2000).

36. Davidi, D. *et al.* Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k _{cat} measurements. *Proc. Natl. Acad. Sci.* **113**, 3401–3406, DOI: 10.1073/pnas.1514240113 (2016). arXiv:1404.2263v1.

37. Heckmann, D. *et al.* Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nat. Commun.* **9**, 5252, DOI: 10.1038/s41467-018-07652-6 (2018).

38. Kaltenbach, H.-M. & Stelling, J. Modular Analysis of Biological Networks. In Junker, B. H. & Schreiber, F. (eds.) *Analysis of Biological Networks*, vol. 736 of *Wiley Series on Bioinformatics: Computational Techniques and Engineering*, 3–17, DOI: 10.1002/9780470253489 (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2008).

39. Benyamini, T., Folger, O., Ruppin, E. & Shlomi, T. Flux balance analysis accounting for metabolite dilution. *Genome Biol.* **11**, R43, DOI: 10.1186/gb-2010-11-4-r43 (2010).

40. Zhuang, K., Vemuri, G. N. & Mahadevan, R. Economics of membrane occupancy and respiro-fermentation. *Mol. Syst. Biol.* **7**, DOI: 10.1038/msb.2011.34 (2011).

41. Klumpp, S., Zhang, Z. & Hwa, T. Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139**, 1366–1375, DOI: 10.1016/j.cell.2009.12.001 (2009).

42. Wortel, M. T., Peters, H., Hulshof , J., Teusink, B. & Bruggeman, F. J. Metabolic states with maximal specific rate carry flux through an elementary flux mode. *The FEBS J.* **281**, 1547–1555, DOI: 10.1111/febs.12722 (2014).

43. Müller, S., Regensburger, G. & Steuer, R. Enzyme allocation problems in kinetic metabolic networks: Optimal solutions are elementary flux modes. *J. Theor. Biol.* **347**, 182 – 190, DOI: https://doi.org/10.1016/j.jtbi.2013.11.015 (2014).

44. Gagneur, J. & Klamt, S. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinforma.* **5**, 175, DOI: 10.1186/1471-2105-5-175 (2004).

45. de Groot, D. H., van Boxtel, C., Planqué, R., Bruggeman, F. J. & Teusink, B. The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates. *PLOS Comput. Biol.* **15**, 1–24, DOI: 10.1371/journal.pcbi.1006858 (2019).

46. Basan, M. *et al.* Overflow metabolism in escherichia coli results from efficient proteome allocation. *Nature* **528**, 99 EP – (2015). Article.

47. Afriat, S. Theory of maxima and the method of lagrange. *SIAM J. on Appl. Math.* **20**, 343–357, DOI: 10.1137/0120037 (1971).

48. Orr, H. A. Fitness and its role in evolutionary genetics. *Nat. Rev. Genet.* **10**, 531 EP – (2009). Review Article.

49. Luis-Miguel, C. On measuring selection in experimental evolution. *Biol. Lett.* **7**, 210–213, DOI: 10.1098/rsbl.2010.0580 (2011).

50. Radzikowski, J. L. *et al.* Bacterial persistence is an active σs stress response to metabolic flux limitation. *Mol. Syst. Biol.* **12**, DOI: 10.15252/msb.20166998 (2016).

51. Santos-Zavaleta, A. *et al.* EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* **41**, D605–D612, DOI: 10.1093/nar/gks1027 (2012).

52. Bremer H, D. P. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus* (2008).

53. Lee, B. Calculation of volume fluctuation for globular protein models. *Proc. Natl. Acad. Sci.* **80**, 622–626, DOI: 10.1073/pnas.80.2.622 (1983).

54. Bar-Even, A., Noor, E., Flamholz, A., Buescher, J. M. & Milo, R. Hydrophobicity and charge shape cellular metabolite concentrations. *PLOS Comput. Biol.* **7**, 1–7, DOI: 10.1371/journal.pcbi.1002166 (2011).

[References 39-54 are for SI text only.]

Acknowledgements

We thank Johannes Berg, Oliver Ebenhöh, Daan de Groot, Xiao-Pan Hu, Terry Hwa, Michael Lässig, Wolfram Liebermeister, Elad Noor, 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).

Author contributions

HD and MJL jointly conceived the study, interpreted the results, and wrote the manuscript. HD developed the GBA framework, performed all data analyses, and derived all formal results except Theorems 1-3, which were derived by MJL.

Supplementary Information

A. Growth balance analysis

In this section, we provide a formal description of growth balance analysis (GBA), detailing the formal definitions, theorems, and proofs that form the basis of the main text. 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 all $\alpha \in \{\alpha\}$. We use corresponding notations for the sets of *basis reactants* $\{\beta\}$, with concentrations b_β, and *dependent reactants* $\{\gamma\}$, with concentrations c_γ (see below).

Characterization of balanced growth states

First, we introduce the fundamental definitions that characterize the solution space of balanced cellular growth. We define balanced growth states and generalize the concept of elementary flux modes from linear constraint-based models to elementary growth states (defined as flux vectors). We then introduce several theorems on the characterization and decomposition of balanced growth states.

In the formulation presented here, we assume that proteins do only act as catalysts and not as substrates of reactions. Hence, neither total protein nor individual proteins are considered "reactants".

Definition 1 (Balanced growth states (BGSs))**.** *Let $\boldsymbol{v}' \in \mathbb{R}^{n'}$ be the vector of fluxes through the biochemical reactions that occur in a cell, in units of [mass/volume/time]. Let $\boldsymbol{v} \in \mathbb{R}^n_{\neq 0}$, $n \leq n'$, be the subvector of $\boldsymbol{v}'$ that contains all active fluxes of $\boldsymbol{v}'$ (i.e., all entries $v'_k \neq 0$) . Let $\boldsymbol{x} = [P, \boldsymbol{a}]^T \in \mathbb{R}^{m+1}_{>0}$ be a corresponding vector of total protein concentration P and individual reactant concentrations $a_\alpha, \alpha \in \{1, ..., m\}$., where each a_α is consumed or produced by at least one of the fluxes v_i; x_i is in units of [mass/volume]. Let $A \in \mathbb{R}^{(m+1)\times n}$ be the corresponding active stoichiometric matrix in mass fraction units, i.e., column j of A describes reaction j with flux v_j, row i of A corresponds to the cellular component x_i, and each column is mass balanced. Thus, the sum of negative entries in each column is $S_- = -1$ and the sum of positive entries of each column is $S_+ = +1$; for reactions that involve an external substrate not represented by a row of A, $-1 < S_- \leq 0$, while for reactions that involve an external product, $0 \leq S_+ < 1$.*

Let $\boldsymbol{p} \in \mathbb{R}^n_{>0}$ be the vector of individual protein concentrations (in units of [mass/volume]), where protein j catalyzes reaction j; for simplicity, we assume that "ribosome" catalyzing protein production is also itself a protein. Let $\boldsymbol{k}(\boldsymbol{a})$ be a vector of kinetic functions, $\boldsymbol{k} : \mathbb{R}^m_{>0} \mapsto \mathbb{R}^n_{\neq 0}$, where $k_j(\boldsymbol{a})$ is in units of [1/time].

Then $\boldsymbol{v}$ *is a* ***balanced growth state (BGS)*** *at growth rate* μ *if and only if it fulfills the following three constraints:*

$$A\boldsymbol{v} = \mu \begin{bmatrix} P \\ \boldsymbol{a} \end{bmatrix} \tag{S1}$$

$$v_j = p_j k_j(\boldsymbol{a}) \tag{S2}$$

$$P = \sum_j p_j \quad . \tag{S3}$$

Constraint (S1) implements mass balance, constraint (S2) implements concentration-dependent reaction kinetics, while constraint (S3) implements a constraint on the total proteome concentration.

A BGS $\boldsymbol{v}$ *at growth rate* μ *is a* ***capacity-constrained balanced growth state (cBGS)*** *if it additionally fulfills the constraint on cellular capacity*

$$\rho \geq P + \sum_\alpha a_\alpha \quad . \tag{S4}$$

The kinetic constraint (S2) assumes that the flux through each reaction is linear in the concentration of the catalyzing enzyme, while the dependence on the reactant concentrations a_α will typically be non-linear. For simplicity of notation, we will sometimes make the dependence of kinetics on $\boldsymbol{a}$ implicit, i.e., we will use $k_j \equiv k_j(\boldsymbol{a})$.

In the above definitions, we define BGSs (or cBGSs) as a function of the set of active reactions (corresponding to the columns of A) and the concentration vector. The set of all such states at all concentrations $\boldsymbol{x} = [P, \boldsymbol{a}]^T \in \mathbb{R}^{m+1}_{>0}$ defines the solution space of balanced growth (or capacity-constrained balanced growth) for a given active stoichiometric matrix A.

If we consider the concentration vector $\boldsymbol{x} = [P, \boldsymbol{a}]^T$ as a descriptor of a constant biomass composition, Eq. (1) is mathematically identical to the flux balance equation at the heart of FBA (see., e.g., Ref.[39]).

Based on biophysical considerations, we might replace Eq. (S4) with separate capacity constraints on the total volume concentration inside each cellular compartment[28] and on the total area occupied by non-lipid membrane components per membrane area[5,40]. An even simpler capacity constraint imposed in most previous models[3,5–8,12–14] is to fix total protein concentration P to a constant value. However, it has been shown that P decreases with increasing growth rate[31,41]. Thus, while a constant P allows to simplify

the presentation, Eq. (9) provides a more meaningful constraint; moreover, Eq. (9) allows us to determine the costs and benefits of varying the total protein concentration.

De Groot *et al.* have defined balanced growth states for a similar problem[18]. In their formulation, the dimensions of the concentration vector $\boldsymbol{x}$ include not only total protein P, but all individual protein concentrations p_j. This more general problem formulation comes at the cost of more involved decomposition rules[18] compared to Theorem 2, and does not lend itself to the derivation of explicit expressions for growth rate (Theorem 7), fitness costs of concentrations (Theorem 8), or necessary conditions of maximal balanced growth (Theorem 10).

We now provide the basis for linking BGSs to elementary flux modes, which are defined for FBA-type linear constraint-based problems[17] and which have been extended to proteome-constrained models[42,43].

Definition 2 (Elementary flux modes (EFM))**.** *Let $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{x} = [P, \boldsymbol{a}]^T \in \mathbb{R}^{m+1}_{>0}$, and $A \in \mathbb{R}^{(m+1)\times n}$ be as in Def. 1. Let $\boldsymbol{k}^{(\text{eff})} \in \mathbb{R}^n_{\neq 0}$ be a vector of effective kinetic constants. Then we call $\boldsymbol{v}$ a **feasible flux vector** at biomass production rate v_{bio} if and only if it fulfills the following constraints:*

$$A\boldsymbol{v} = v_{bio} \begin{bmatrix} P \\ \boldsymbol{a} \end{bmatrix} \tag{S5}$$

$$v_j \leq p_j k_j^{(\text{eff})} \tag{S6}$$

$$P = \sum_j p_j \quad . \tag{S7}$$

*A feasible flux vector $\boldsymbol{v}$ is a representative of an **elementary flux mode** if and only if it is non-decomposable, i.e., it fulfills the following additional constraint[17]: There exists no couple of feasible flux vectors $\boldsymbol{v}', \boldsymbol{v}''$ such that $\boldsymbol{v} = \lambda_1 \boldsymbol{v}' + \lambda_2 \boldsymbol{v}''$ with $\lambda_1, \lambda_2 > 0$ and where both $\boldsymbol{v}'$ and $\boldsymbol{v}''$ have at least the same number of zeroes as $\boldsymbol{v}$, while at least one of them contains more zeroes than $\boldsymbol{v}$.*

Constraint (S5) is equivalent to the standard steady state constraint of flux balance analysis problems, formulated with an equation analogous to Eq. (S1) for a fixed biomass vector instead of including an artificial "biomass reaction" in A (see, for example, Eq. (2) in Ref.[39]).

Note that in the definition of EFMs, both the biomass composition $\boldsymbol{x} = [P, \boldsymbol{a}]^T$ and the effective kinetics $\boldsymbol{k}^{(\text{eff})}$ are assumed to be constant; thus, the constraints (S5)-(S7) that define the space of feasible flux

vectors are fully linear. In contrast, constraint (S2) in Def. 1 defines reaction kinetics as a function of the reactant concentrations $\boldsymbol{a}$.

Definition 3 (Elementary growth state (EGS))**.** *A BGS $\boldsymbol{v}$ is an **elementary growth state** (EGS) if and only if it is a representative of a corresponding EFM, i.e., $\boldsymbol{v}$ represents an EFM of the corresponding linear problem with constant biomass $\boldsymbol{x} = [P, \boldsymbol{a}]^T$ and effective kinetic constants $\boldsymbol{k}^{(\mathrm{eff})} = \boldsymbol{k}(\boldsymbol{a})$.*

We emphasize that $\boldsymbol{v}$ is an EFM of the corresponding linearized (FBA-like) problem (see Def. 2), not of the balanced growth problem from Def. 1 from which it is derived. EFMs are defined as equivalence classes of minimal feasible steady-state flux distributions, whose members can be converted into each other by multiplication with a positive scalar[17]. This definition cannot be generalized to balanced growth models, as multiples of an admissible flux vector generally do not satisfy constraint (S1). 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)[18].

Theorem 1 (Existence of solutions)**.** *Let $\boldsymbol{x} = [P, \boldsymbol{a}]^T$ be a concentration vector, and $\mu > 0$ be a growth rate. For any flux vector $\boldsymbol{v}'$ that satisfies the mass balance constraint (S1), there exists a unique BGS $\boldsymbol{v} = \lambda \boldsymbol{v}'$ with $\lambda > 0$ if all fluxes run in the direction compatible with the reaction kinetics (i.e., $\forall_j\, k_j v_j > 0$), and no such BGS otherwise.*

Proof. From constraint (S2), it is clear that if $k_j v_j \leq 0$, no BGS with $p_j > 0$ exists. For $k_j \neq 0$, the concentration of protein j is uniquely defined by $p_j = v_j / k_j$ (constraint (S2)). Let $P' = \sum_j v'_j / k_j$ be the total protein concentration associated with $\boldsymbol{v}'$. Then setting $\lambda \equiv P/P'$ results in the only flux vector that fulfills all constraints of Def. (1). □

Next, we use this result to show that any weighted average of BGSs is itself a BGS.

Theorem 2 (A weighted average of BGSs is a BGS)**.** *Let $(\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(k)})$ be an ordered set of BGSs for the concentration vector $\boldsymbol{x} = [P, \boldsymbol{a}]^T$ with growth rates $(\mu^{(1)}, \ldots, \mu^{(k)})$, but with potentially different active stoichiometric matrices $A^{(l)}$. Let A be the stoichiometric matrix that combines all reactions represented in $(A^{(1)}, \ldots, A^{(k)})$, i.e., the columns of A consist of all unique columns of $(A^{(1)}, \ldots, A^{(k)})$. Let $(\boldsymbol{v}'^{(1)}, \ldots, \boldsymbol{v}'^{(k)})$ be a representation of the individual BGSs $\boldsymbol{v}^{(l)}$ in the flux space defined by A, i.e., $v'^{(l)}_j = 0$ for all columns (reactions) of A not represented in $A^{(l)}$. Then any weighted average $\boldsymbol{v} = \sum_l w_l \boldsymbol{v}'^{(l)}$ of these extended flux vectors (with weights $w_l > 0$ and $\sum_l w_l = 1$) is itself a BGS for $\boldsymbol{x}$, with a growth rate that is the weighted average of the individual growth rates, $\mu = \sum_l w_l \mu^{(l)}$.*

Proof. The mass balance constraint (S1) is linear in the fluxes and growth rates, and is hence also fulfilled for the weighted averages. The protein concentrations of each BGS $\boldsymbol{v}'^{(l)}$ are $p_j'^{(l)} = v_j'^{(l)}/k_j$. To satisfy the reaction kinetics constraint (S2), the protein concentrations of the weighted average are $p_j = v_j/k_j = \sum_l w_l v_j'^{(l)}/k_j = \sum_l w_l p_j'^{(l)}$. As each BGS (l) fulfilled the proteome constraint (S3), $\sum_j p_j = \sum_j \sum_l w_l p_j'^{(l)} = \sum_l w_l P = P$, and thus $\boldsymbol{v}$ is a BGS. □

We can now use Theorems 1 and 2 together with results on elementary flux modes to show that any BGS can be decomposed into a weighted average of EGSs.

Theorem 3 (BGSs are weighted averages of EGSs)**.** *Any BGS $\boldsymbol{v}$ for the concentration vector $\boldsymbol{x} = [P, \boldsymbol{a}]^T$ can be decomposed into a weighted average of EGSs at $\boldsymbol{x}$.*

Proof. $\boldsymbol{v}$ is a feasible flux vector for the linearized problem defined by constraints (S5)-(S7) at constant biomass $\boldsymbol{x}$. The direction of reaction j is fixed by the sign of $k_j^{(\mathrm{eff})} = k_j(\boldsymbol{a})$, i.e., all reactions are irreversible. Under these conditions, it has been shown that $\boldsymbol{v}$ is a convex combination of elementary flux modes $\boldsymbol{v}'^{(l)}$ of the linear problem[17], i.e., $\boldsymbol{v} = \sum_l w_l' \boldsymbol{v}'^{(l)}$ with $w_l' > 0$. From Theorem 1, we know that for each of these EFMs, there exists a unique BGS $\boldsymbol{v}^{(l)} = \lambda_l \boldsymbol{v}'^{(l)}$ with $\lambda_l > 0$; according to Def. 3, this is an EGS. Thus, we can write $\boldsymbol{v} = \sum_l w_l \boldsymbol{v}^{(l)}$ as a linear combination of EGSs, with weights $w_l \equiv w_l'/\lambda_l$.

To prove that $\boldsymbol{v}$ is a weighted average of the $\boldsymbol{v}^{(l)}$, it remains to be shown that $W \equiv \sum_l w_l = 1$. According to Theorem 2, a weighted average $\boldsymbol{v}'' \equiv \sum_l \frac{w_l}{W} \boldsymbol{v}^{(l)} = \frac{1}{W}\boldsymbol{v}$ will also be a BGS. However, Theorem 1 states that there exists only one BGS in the direction of $\boldsymbol{v}$, and thus $W = 1$. □

Growth equations

In this section, we assume that the concentrations of total protein and of individual reactants, $\boldsymbol{x} \equiv [P, \boldsymbol{a}]$ are known. Mass conservation (constraint (S1)) and reaction kinetics (constraint (S2)) relate reaction fluxes to the concentration vector in two fundamentally different ways. We will now exploit this fact to eliminate the flux variables and to derive explicit expressions for $\boldsymbol{v}$, $\boldsymbol{p}$, and μ.

Note that because the concentrations $\boldsymbol{x}$ are used as input parameters in these analyses, no explicit consideration of constraints on cellular capacity, such as constraint (S4) is necessary. The given concentrations $\boldsymbol{x}$ may obey constraint (S1) or alternative capacity constraints, such as independent constraints on the capacity of cellular compartments, but these will not be used here. They will only become important when we vary $\boldsymbol{x}$ to find states of maximal growth rate in Section A..

An important requirement for the analyses below is that the active stoichiometric matrix A has full column rank, motivating the next theorem.

Theorem 4 (The active reactions of an EGS are linearly independent)**.** *Let $A \in \mathbb{R}^{(m+1)\times n}$ be the active stoichiometric matrix of an EGS. Then A has full column rank n, i.e., the columns of A are linearly independent.*

Proof. According to the definition of EGSs (Def.3), A is also the active matrix of the corresponding linearized (flux balance type) problem. It has previously been shown that the active stoichiometric matrix A of an EFM of a linear flux-balance problem has full column rank if A is formulated without an explicit "biomass" reaction (as in Def. 2)[44]. □

According to this theorem, the following theorems - which assume that A has full column rank - can in particular be applied to EGSs (and, as we will see below in Theorem 9, thus also to cBGSs with maximal growth rate).

Theorem 5 (Investment equation)**.** *Let $A \in \mathbb{R}^{(m+1)\times n}$ be an active stoichiometric matrix of a flux vector $\boldsymbol{v}$ that fulfills the mass balance constraint (S1) with concentration vector $\boldsymbol{x} = [P, \boldsymbol{a}]^T$, where A has full column rank n. Then we can split A into two submatrices $B \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{(m+1-n)\times n}$,*

$$A = \begin{bmatrix} B \\ C \end{bmatrix} \quad ,$$

such that B is a non-singular (invertible) square matrix and each row of C is a linear combination of rows of B. Let $I \equiv B^{-1}$. Let $\boldsymbol{b}$ be the subvector of reactant concentrations $\boldsymbol{a}$ that correspond to the rows of B, and let $\boldsymbol{c}$ be the subvector of the remaining reactant concentrations. Then $\boldsymbol{v}$ is given by

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

The dependent reactant concentrations $\boldsymbol{c}$ are linear combinations of the independent concentrations $[P, \boldsymbol{b}]^T$,

$$\boldsymbol{c} = D \begin{bmatrix} P \\ \boldsymbol{b} \end{bmatrix} \quad , \tag{S8}$$

with the dependence matrix $D \equiv CI$.

Proof. A may have more rows than columns ($m+1>n$). In this case, the rows for exactly n metabolites are linearly independent, as row and column rank must equal. As a consequence, the remaining $m+1-n$ metabolite concentrations are linearly dependent on the concentrations of the n independent metabolites. These dependent concentrations are not free variables, and hence they can be put aside and dealt with separately.

We decompose the linear system of equations represented by constraint (S1) into two parts, rearranging the rows of A into matrices B,C such that B contains the rows for the independent reactants. As A has full column rank, choosing linearly independent rows results in a square matrix B of full rank ($\#rows(B) = rank(B) = rank(A)$). Let $\boldsymbol{b}$ be the subvector of reactant concentrations $\boldsymbol{a}$ that correspond to the rows of B, and let $\boldsymbol{c}$ be the subvector of the remaining reactant concentrations corresponding to the rows of C. We can then split the mass balance constraint (S1) into two separate equations:

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

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

B is a square matrix of full rank, so there is always a unique inverse $I \equiv B^{-1}$. Multiplying both sides of the first equation by I from the left, we obtain the desired equation for $\boldsymbol{v}$. Inserting this result into the second equation results in the desired equation for $\boldsymbol{c}$. □

Thus, if A has full rank, then any flux vector $\boldsymbol{v}$ respecting the flux balance constraint (S1) is uniquely defined and is a linear combination of the total protein concentration P and the independent metabolite concentrations $\boldsymbol{b}$. Each entry of the inverse matrix 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. S1 for examples). In contrast to the stoichiometric matrix A, which describes local mass balances (constraint (S1)), I describes the structural allocation of reaction fluxes into the production of cellular components diluted by growth, and thus carries global, systems-level information.

B corresponds to the reduced stoichiometric matrix in Ref.[19]. 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.[19]. 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 G.**).

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 by the investment equation in Theorem 5.

We are now in a position to express the individual protein concentrations and the growth rate of a BGS as explicit functions of the concentrations $\boldsymbol{x} = [P, \boldsymbol{a}]^T$.

Theorem 6 (Individual protein concentrations as a function of the independent concentrations)**.** *Let $A \in \mathbb{R}^{(m+1)\times n}$ be an active stoichiometric matrix with full column rank n, and let $\boldsymbol{x} = [P, \boldsymbol{a}]^T$ be a concentration vector. Let $\boldsymbol{v}$ be a corresponding BGS. Let B and D be the basis and dependency matrices, respectively, as defined in Theorem 5, and let $I = B^{-1}$. Then the concentration of the protein catalyzing reaction j is*

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

Proof. As A is an active matrix, all fluxes $v_j = p_j k_j(\boldsymbol{a})$ (constraint (S2)) are non-zero. We can thus express the individual protein concentrations as $p_j = v_j / k_j(\boldsymbol{a})$. Inserting v_j from the investment equation (Theorem 5) directly leads to the above equation. □

We now insert the equations for the individual proteins into the total protein constraint (S3) to obtain an explicit expression for the growth rate.

Theorem 7 (Growth equation)**.** *Let $A \in \mathbb{R}^{(m+1)\times n}$ be an active stoichiometric matrix with full column rank n, and let $\boldsymbol{x} = [P, \boldsymbol{a}]^T$ be a concentration vector. Let $\boldsymbol{v}$ be a corresponding BGS. Let B and D be the basis and dependency matrices, respectively, as defined in Theorem 5, and let $I = B^{-1}$. Then the growth rate is*

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

if for all reactions $\frac{p_j}{\mu} = \frac{I_{jP}P + \sum_\beta I_{j\beta} b_\beta}{k_j(\boldsymbol{a})} > 0$, and no balanced growth is possible otherwise.

Proof. According to Theorem 6, the individual protein concentrations are $p_j = \mu \frac{I_{jP}P + \sum_\beta I_{j\beta} b_\beta}{k_j(\boldsymbol{a})}$. The flux v_j catalyzed by protein j must be active, and thus p_j has to be positive for all j. Substituting the expressions for p_j into the proteome constraint (S3), we obtain

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

The sum on the r.h.s. is positive, and dividing by it results in the growth equation. □

Thus, if the active matrix A of a BGS 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 growth states, as well as for all other EGSs. In this section, we did not impose any capacity constraints (such as constraint (S4)), and thus Theorems 1-7 remain valid under arbitrary capacity constraints (as long as the capacity constraints are respected by the concentration vector $\boldsymbol{x} = [P, \boldsymbol{a}]^T$).

Marginal fitness benefits and costs

In this section, we first define marginal fitness benefits and costs of concentrations. As in the previous section, the definitions make no use of the capacity constraint (S4), and thus remain valid under alternative capacity constraints. After introducing the definitions, we explore the marginal fitness benefits of cellular concentrations at optimal growth; at this point, the capacity constraint becomes central to our analysis.

Definition 4 (Marginal costs and benefits)**.** *Let $\boldsymbol{v}$ be a BGS with growth rate μ. Let $i \in \{P, \beta\}$ be an index of the concentration vector $\boldsymbol{x} = [P, \boldsymbol{b}]^T$.*

*Then the **direct marginal net benefit** of concentration x_i is defined as the relative change in growth rate due to a small change in x_i*[20]*,*

$$\eta_i^0 \equiv \frac{1}{\mu}\frac{\partial \mu}{\partial x_i} \quad .$$

*The **marginal production cost** of x_i is defined as*

$$q_i^j \equiv \frac{\mu I_{ji}}{P k_j} \quad .$$

*The **marginal kinetic benefit** of x_i is defined as*

$$u_\beta^j \equiv \frac{p_j}{P}\frac{1}{k_j}\frac{\partial k_j}{\partial b_\beta} \quad .$$

Further, we define the (total) marginal kinetic benefit of dependent reactant γ as

$$\eta_\gamma^c \equiv \sum_j u_\gamma^j \quad , \tag{S9}$$

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

$$u_\gamma^j \equiv \frac{p_j}{P}\frac{1}{k_j}\frac{\partial k_j}{\partial c_\gamma} \quad .$$

The (total) ***marginal net benefit*** *of x_i is then defined as the relative change in growth rate due to a small change in x_i, accounting for the resulting changes in the concentration of dependent metabolites c_γ,*

$$\begin{aligned}\eta_i &\equiv \frac{1}{\mu}\left(\frac{\partial\mu}{\partial x_i}+\sum_\gamma \frac{\partial\mu}{\partial c_\gamma}\frac{\partial c_\gamma}{\partial x_i}\right)\\ &= \eta_i^0+\sum_\gamma D_{\gamma i}\eta_\gamma^c \quad ,\end{aligned} \tag{S10}$$

where the second equality follows from Eq. (S8), Theorem 7, and the previous definitions.

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. (S8)). 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*** *defined as*

$$\kappa_i \equiv 1+\sum_\gamma D_{\gamma i} \quad .$$

The definition of η_i^0 accounts for the production costs of dependent reactants c_γ, as these costs are embedded in I (B is not balanced if there are dependent reactants). However, η_i^0 ignores the kinetic benefits of the dependent reactants; this is why the definition of η_i includes a separate term for their kinetic benefits but not their costs.

If I_{ji} and k_j are both positive, then the production cost 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. In contrast, the kinetic benefit u_β^j is a local effect, as it is non-zero only for reactants β directly involved in reaction j. u_β^j will generally be positive if β is a substrate of reaction j.

The marginal net benefits can be expressed as the difference between marginal benefits and costs.

Theorem 8 (Direct marginal net benefits)**.** *The direct marginal net benefits of the total protein concentration P and of independent reactant concentrations b_β ($\beta \in \{1,\dots,m\}$), respectively, are*

$$\eta_P^0 = \frac{1}{P}-\sum_j q_P^j$$

$$\eta_\beta^0 = \sum_j (u_\beta^j - q_\beta^j) \quad .$$

Proof. Taking the corresponding derivatives (see Def. (4)) of the growth equation (Theorem 7) directly leads to these equations. □

So far, we have considered BGS for a given set of active reactions (corresponding to the columns of A) and given concentrations $\boldsymbol{x} = [P,\boldsymbol{a}]^T$. Below, we will examine capacity-constrainted BGSs (cBGSs)

with maximal growth rate given the set of active reactions, optimized over all concentration vectors $\boldsymbol{x} = [P, \boldsymbol{a}]^T \in \mathbb{R}^{m+1}_{>0}$ that respect the capacity constraint (S4). As a preparation for these analyses, we first show that cBGSs at optimal growth are EGSs.

Theorem 9 (cBGSs with maximal growth rate are EGSs)**.** *Let N be a stoichiometric matrix of a general balanced growth model. Let $\boldsymbol{v}^*$ be a cBGS that maximizes the growth rate of the general problem. Then $\boldsymbol{v}^*$ is an EGS.*

Proof. Without loss of generality, we restrict $\boldsymbol{v}^*$ to its active dimensions ($v_j^* \neq 0$), with active stoichiometric matrix A. Then this reduced $\boldsymbol{v}^*$ is the optimal solution for the following non-linear optimization problem over all concentration vectors $\boldsymbol{x} \equiv [P, \boldsymbol{a}]^T \in \mathbb{R}^{m+1}_{>0}$:

$$\begin{aligned}
&\underset{\boldsymbol{x}}{\text{maximize}} \quad \mu \\
&\text{subject to:} \\
&\qquad A\boldsymbol{v} = \mu \boldsymbol{x} \\
&\qquad \forall_j \;\; v_j = p_j k_j(\boldsymbol{x}) \\
&\qquad P = \sum_j p_j \\
&\qquad \rho \geq P + \sum_\alpha a_\alpha \quad .
\end{aligned} \tag{S11}$$

Let $\boldsymbol{x}^* = [P^*, \boldsymbol{a}^*]^T$ be the concentrations and μ^* the growth rate of the optimal solution $\boldsymbol{v}^*$. Now let us consider a linearized version of this optimization problem, where me maximize the production rate v_{bio} at constant biomass composition $\boldsymbol{x}^*$ and effective kinetic constants $k_j^{(\text{eff})} \equiv k_j(\boldsymbol{a}^*)$ (see Def. 2):

$$\begin{aligned}
&\underset{\boldsymbol{x}}{\text{maximize}} \quad v_{bio} \\
&\text{subject to:} \\
&\qquad A\boldsymbol{v} = v_{bio}\boldsymbol{x}^* \\
&\qquad \forall_j \;\; v_j = p_j k_j^{(\text{eff})} \\
&\qquad P^* \geq \sum_j p_j \quad .
\end{aligned} \tag{S12}$$

We relaxed the constraint (S3) on total protein into an inequality constraint, so that Eq. (S12) describes a protein-constrained FBA problem for the active stoichiometric matrix. This is precisely the type of constrained flux balance problem analyzed in Refs.[42,43], which prove that the solutions $\boldsymbol{v}^{\text{opt}}$ to the optimization problem defined by Eq. (S12) are elementary flux modes (EFMs).

In the optimal solution to the problem defined by Eq. (S12), the protein concentration constraint will be active, that is, $P^* = \sum_j p_j$; if not, the biomass production rate v_{bio} 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. (S12) is the same as that described by Eq. (S11), except for a reduction in the dimension of the search space due to the fixed concentrations $\boldsymbol{x}^*$ (Note that the cellular capacity constraint (S4) is trivially respected in Eq. (S12) and can be ignored). Accordingly, the flux distribution $\boldsymbol{v}^*$ that maximizes the balanced growth rate μ in Eq. (S11) also maximizes the biomass production rate v_{bio} of the protein-constrained FBA problem in Eq. (S12); it is hence a representative of an EFM of the active stoichiometric matrix A with biomass $\boldsymbol{x}^*$,[42,43] and thus $\boldsymbol{v}^*$ is an EGS according to Def. 3. □

In parallel work to that presented here, de Groot *et al.* have shown that optimal solutions to balanced growth problems are elementary growth modes as defined in Ref.[18], and that the active stoichiometric matrix of elementary growth modes has full rank[18].

If instead of a single constraint on cellular capacity, multiple capacity constraints are imposed simultaneously (e.g., to describe separate constraints on different cellular compartments), then the solutions may in some cases correspond to positive linear combinations of EGSs[18,45], and the treatment below would have to be generalized. Multiple capacity constraints may play a role in the emergence of overflow metabolism in *E. coli*[46], although overflow metabolism can also arise in balanced growth models with a single capacity constraint[5].

In a cBGS with maximal growth rate for a given active stoichiometric matrix A, the cellular components will utilize the full cellular capacity ρ to saturate enzymes with their substrates. Thus, the constraint (S4) will be active, turning the inequality into an equality. The maximal balanced growth rate μ^* will thus be a function of the cellular capacity ρ. As a reference value for the total marginal net benefits of individual concentrations x_i, we now define the marginal benefit of the cellular capacity ρ (constraint (S4)). This is the first definition that makes use of the capacity constraint.

Definition 5 (Marginal benefit of the cellular capacity)**.** *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,*

$$\eta_\rho \equiv \frac{1}{\mu^*}\frac{d\mu^*}{d\rho} \quad .$$

We can now relate η_ρ to the total marginal net benefits of all concentrations. To do this, we derive necessary conditions for any optimal balanced growth state at constant cellular capacity ρ, using the

method of Lagrange multipliers. The Lagrange multipliers quantify the importance of the capacity constraint, Eq. (S4), and of the constraints for the dependent reactants, Eq. (S8), for the maximization of the objective function. The Lagrangian $\mathscr{L}$ is a function of P, $\boldsymbol{a}$, and ρ.

Theorem 10 (Balance equation)**.** *In a cBGS with maximal growth rate, the total marginal net benefit of each independent concentration $x_i \in \{P, b_\beta\}$ equals the marginal benefit of the cellular capacity ρ scaled by the capacity factor κ_i,*

$$\forall_{i\in\{P,\beta\}} \quad \eta_i = \kappa_i \eta_\rho \quad . \tag{S13}$$

Proof. We use the method of Lagrange multipliers to derive necessary conditions for any optimal cBGS at constant cellular capacity ρ. Our objective function is given by Theorem 7, which expresses the growth rate μ as an explicit function of the concentrations $\boldsymbol{x} = [P, \boldsymbol{a}]^T$. The capacity constraint (S4) will be active at maximal growth rate, i.e., it becomes an equality. The capacity constraint can then be expressed as a function g_ρ that depends on ρ and on the concentrations,

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

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 b_β,

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

We now define a Lagrangian as the sum of the objective function μ and the constraints $\boldsymbol{g}$ scaled by Lagrange multipliers λ_ρ, accounting for the capacity constraint (S4), and λ_γ, accounting for the dependence of the dependent reactants $\gamma \in \{\gamma\}$, Eq. (S8):

$$\mathscr{L} \equiv \mu + \lambda_\rho g_\rho + \sum_\gamma \lambda_\gamma g_\gamma \quad .$$

The first order necessary conditions for a constrained local maximum are that all partial derivatives of

$\mathscr{L}$ with respect to the variables P, b_β, c_γ and to the Lagrange multipliers $\lambda_\rho, \lambda_\gamma$ are zero,

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

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

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

With Theorem (8), this results in

$$\mu \ \eta_i^0 + \lambda_\rho + \sum_\gamma \lambda_\gamma D_{\gamma i} = 0 \quad . \tag{S14}$$

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

$$\frac{\partial \mathscr{L}}{\partial c_\gamma} = \frac{\partial \mu}{\partial c_\gamma} + \lambda_\rho - \lambda_\gamma = 0 \quad .$$

With Eq. (S9), we obtain

$$\lambda_\gamma = \mu \ \eta_\gamma^0 + \lambda_\rho \quad .$$

Substituting λ_γ from the last equation into Eq. (S14) gives (for $i \in \{P, \beta\}$)

$$\mu \ \eta_i^0 + \lambda_\rho + \sum_\gamma \left(\mu \eta_\gamma^c + \lambda_\rho\right) D_{\gamma i} = 0 \quad .$$

Rearranging results in

$$\begin{aligned}
0 &= \mu \ \eta_i^0 + \lambda_\rho \left(1 + \sum_\gamma D_{\gamma i}\right) + \mu \sum_\gamma D_{\gamma i} \eta_\gamma^c \\
&= \mu \ \eta_i + \lambda_\rho \kappa_i \\
&= \mu \ \eta_i - \mu \ \eta_\rho \kappa_i \quad ,
\end{aligned} \tag{S15}$$

where we used $\eta_\rho = -\lambda_\rho / \mu$, which follows directly from the envelope theorem[47]. With $\mu > 0$, we thus obtain the balance equation. □

The optimal state is perfectly balanced: the total 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 ($\forall_\gamma\, D_{\gamma i} = 0$), then the balance equation simplifies to $\eta_i = \eta_i^0 = \eta_\rho$ (Eq. (10)).

Theorem 10 states that if the dry weight density ρ would be allowed to increase by a small amount, such as 1 mg/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.

Instead of using Lagrange multipliers in the proof, one could express the total protein concentration $P = \rho - \sum_\alpha a_\alpha$ (constraint (S4)) and the dependent reactant concentrations $c_\gamma = D_{\gamma P} P + \sum_\beta D_{\gamma\beta} b_\beta$ (Eq. (S8)) in terms of ρ and of the independent reactant concentrations $\boldsymbol{b}$. Substituting the resulting expressions into the growth equation (Theorem 7) 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 as derived in the Lagrange multiplier framework, this formulation misses important insights that can be derived from the Lagrange multipliers themselves.

B. 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[48], which we here define as a relative growth rate:

$$f \equiv \frac{\mu + \delta\mu}{\mu} = 1 + \frac{\delta\mu}{\mu} \quad .$$

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

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

Note that population genetics models are frequently defined in terms of discrete generations. With generation time $T_{\text{gen}} = \ln 2/\mu$, the selection coefficient of the variant *per generation* is then[49]

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

C. Optimal ribosome protein fraction

Here we assume a very simple model for translation[26]. 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 a_T) into protein, following irreversible Michaelis-Menten kinetics:

$$k_R \equiv k_R(a_T) = k_{\text{cat}} \left(\frac{a_T}{a_T + K_{\text{m}}} \right) \tag{S16}$$

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

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.S1A happens, and the balance of net benefits of ternary complex and total protein, $\eta_T = \eta_P$ (Eq. (10)), simplifies to

$$P\, u_T^R = 1 - \frac{\mu}{k_R(a_T)} \quad .$$

Substituting the partial derivative of irreversible Michaelis-Menten kinetics (Eq. (S16)), we obtain

$$\frac{R}{a_T(1 + a_T/K_{\text{m}})} = 1 - \frac{\mu}{k_R} \quad . \tag{S17}$$

Rearranging Eq. (S16), 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},

$$a_T = \frac{K_{\text{m}}}{\frac{k_{\text{cat}} R}{v_r} - 1} \quad .$$

Substituting this into Eq. (S17) gives

$$\begin{aligned} R &= \left(1 - \frac{\mu}{k_R}\right) \left[\frac{K_{\text{m}}}{\frac{k_{\text{cat}} R}{v_r} - 1} \left(1 + \frac{1}{\frac{k_{\text{cat}} R}{v_r} - 1} \right) \right] \\ &= \left(1 - \frac{\mu}{k_R}\right) K_{\text{m}} \left[\frac{\frac{k_{\text{cat}} R}{v_r}}{\left(\frac{k_{\text{cat}} R}{v_r} - 1 \right)^2} \right] \quad . \end{aligned} \tag{S18}$$

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

$$R\, k_R = v_R = \mu P \quad .$$

Thus, substituting $\mu/k_R = R/P$ and $v_R = \mu P$ in Eq. (S18), we obtain

$$\frac{R}{P} = \left(1 - \frac{R}{P}\right) \frac{K_\mathrm{m}}{P} \left[\frac{\frac{k_\mathrm{cat} R}{\mu P}}{\left(\frac{k_\mathrm{cat} R}{\mu P} - 1\right)^2} \right] \quad .$$

This is equivalent to a quadratic equation in R/P,

$$\left(\frac{R}{P}\right)^2 + \frac{\mu}{k_\mathrm{cat}} \left(\frac{K_\mathrm{m}}{P} - 2\right) \left(\frac{R}{P}\right) + \left(\frac{\mu}{k_\mathrm{cat}}\right)^2 \left(1 - \frac{k_\mathrm{cat} K_\mathrm{m}}{\mu P}\right) = 0 \quad . \tag{S19}$$

Its two solutions are

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

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

$$k_\mathrm{cat} R = \mu P \left[1 + \frac{K_\mathrm{m}}{2P} \left(\pm \sqrt{1 + \frac{4P}{K_\mathrm{m}} \left(\frac{k_\mathrm{cat}}{\mu} - 1\right)} - 1 \right) \right] \quad . \tag{S20}$$

Because $k_\mathrm{cat} R > R\ k_R = v_R = \mu P$, the term in square brackets ($[\cdot]$) in Eq. (S20) must be > 1. Only the positive root is compatible with this condition. Thus, the ratio R/P is uniquely determined by

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

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 the final equation

$$\phi_R(\mu) = \frac{\mu r_P}{k_\mathrm{cat}} \left[1 + \frac{K_\mathrm{m}}{2P} \left(\sqrt{1 + \frac{4P}{K_\mathrm{m}} \left(\frac{k_\mathrm{cat}}{\mu} - 1\right)} - 1 \right) \right] \quad . \tag{S21}$$

The same procedure can be used to find an equation for ϕ_R that ignores the production costs. Starting from Eq. (S18) without the production cost term μ/k_R, we obtain

$$\frac{R}{P} \approx \frac{K_\mathrm{m}}{P} \left[\frac{\frac{k_\mathrm{cat} R}{\mu P}}{\left(\frac{k_\mathrm{cat} R}{\mu P} - 1\right)^2} \right] \quad ,$$

which results in a quadratic equation similar to Eq. (S19),

$$\left(\frac{R}{P}\right)^2 - 2\frac{\mu}{k_\mathrm{cat}} \frac{R}{P} + \left(\frac{\mu}{k_\mathrm{cat}}\right)^2 \left(1 - \frac{k_\mathrm{cat} K_\mathrm{m}}{\mu P}\right) \approx 0 \quad .$$

Solving for R/P gives

$$\frac{R}{P} \approx \frac{\mu}{k_{\text{cat}}} \left(1 \pm \sqrt{\frac{k_{\text{cat}} K_{\text{m}}}{\mu P}}\right) \quad . \tag{S22}$$

Again because $Rk_{\text{cat}} > \mu P$, the term in parentheses $(\cdot)$ in Eq. (S22) must be > 1, and again only the positive root is compatible with this condition. Thus, the ribosome protein fraction is uniquely determined in this approximation by

$$\phi_R \approx \frac{\mu r_P}{k_{\text{cat}}} \left(1 + \sqrt{\frac{k_{\text{cat}} K_{\text{m}}}{\mu P}}\right) \quad . \tag{S23}$$

We compared the predictions of the ribosome fraction of total protein, $\phi_R = R/P$, to quantitative proteomics data obtained by Schmidt *et al.*[32] (scaled by a factor of 0.67, as suggested by the authors, to account for a systematic error in the cell size measurements[50]). 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[51]. 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 ($r_P = 0.58$[32]), 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[32].

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.[33], with the best-fitting parameter $z = 0.124/h$.

We set the Michaelis constant of the ribosome to $K_{\text{m}} = 3 \times 10^{-6}$ mol/l, based on the diffusion limit for ternary complexes calculated in Ref.[26]. We set the ribosome's turnover number to $k_{\text{cat}} = 22$ AA/s, the highest elongation rate observed experimentally in Ref.[52]. 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.[26] 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.[32] 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 $k_{\text{cat}} = 22$ AA/s $\times$ (132.60 Da/AA)/(2,306,967 Da) $\times 3,600$s/1h = 4.55/h, and $K_{\text{m}} = 40 \times 3 \times 10^{-6}$mol/l $\times 69,167$ g/mol = 8.30 g/l.

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

Cayley *et al.*[29,35] 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 (Def. 5). The relative change in the maximal growth rate per relative change in ρ is then

$$\frac{d\ln\mu^*}{d\ln\rho} = \frac{\rho}{\mu^*}\frac{d\mu^*}{d\rho} = \rho\ \eta_\rho \tag{S24}$$

From Eq. (S13), we know that $\eta_P = \kappa_P \eta_\rho$; if there are no dependent reactants for P (i.e., $\forall_\gamma\, D_{\gamma P} = 0$), this simplifies to

$$\eta_\rho = \eta_P^0 = \frac{1}{P} - \sum_j q_P^j \quad , \tag{S25}$$

and thus

$$\frac{\rho}{\mu^*}\frac{d\mu^*}{d\rho} = \rho\ \eta_\rho = \rho\left(\frac{1}{P} - \sum_j q_P^j\right) \quad . \tag{S26}$$

The mass fraction of total protein in cell dry weight $P/\rho \approx 0.54$ has been shown to be approximately constant across growth conditions supporting intermediate to high growth rates[29]. 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. S2).

The active stoichiometric matrix A of this model and its inverse $I = A^{-1}$ are, written here with row and column labels,

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

The capacity is determined only by its two components,

$$\rho = P + a \quad ,$$

where

$$P = p_t + p_R \quad .$$

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

$$v_t = \mu(P + a) = \mu\rho \tag{S27}$$

and

$$v_R = \mu P \quad . \tag{S28}$$

From the inverse I and Eq. (3), we get

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

Combining this with Eq. (S27) and (S28) and with $\phi_R = p_R/P$ and $\phi_t = p_t/P = 1 - \phi_R$, we obtain

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

Combining this equation with Eq. (S26), we obtain

$$\begin{aligned}\rho\ \eta_\rho &= \rho\left(\frac{1}{P} - \frac{(1-\phi_R)}{\rho} - \frac{\phi_R}{P}\right) \\ &= \frac{\rho}{P} - 1 + \phi_R - \frac{\rho}{P}\phi_R \\ &= \left(\frac{\rho}{P} - 1\right)(1-\phi_R) \quad .\end{aligned} \tag{S29}$$

From Eq. (S21), 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 Ref.[35]) as $\phi_R = 0.19$. Substituting this value into Eq. (S29) together with $P/\rho = 0.54$, we estimate the relative change in the maximal growth rate per relative change in ρ as

$$\rho\ \eta_\rho = 0.69 \quad .$$

Cayley *et al.*[35] 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.[29] Errors are estimated standard deviations based on error propagation among normally distributed random variables.

Fig. S3 plots the natural logarithms of μ and ρ. Linear regression over the three available data points results in an estimated slope of 0.66, close close to our estimate of $\frac{d \ln \mu^*}{d \ln \rho} = \rho\ \eta_\rho = 0.69$.

E. An outline of possible extensions of GBA

In our development of GBA, we make several simplifying assumptions. Here, we outline some possible generalizations.

All proteins contribute to growth by acting as catalysts or transporters. This assumption can simply be removed by adding a sector of non-growth related proteins[25,27] with concentration Q to the r.h.s. of Eq. (2).

Proteins are not used as reactants. To use protein j as a reactant in reaction j', it will need an extra row in A, and its concentration p_j has to enter the concentration vector $\boldsymbol{x} \equiv [P, p_j, \boldsymbol{a}]^T$ and the kinetic function $k_{j'}(p_j, \boldsymbol{a})$. This does not affect Eq. (4). However, if p_j appears on the right hand side of Eq. (5), this equation will have to be solved for p_j before it is possible to proceed to a generalization of the growth equation.

All catalysts are proteins. We can add different catalytic RNA species as cellular components. Additionally, we may introduce reactions that combine proteins and RNA into molecular machines such as the ribosome.

A 1-to-1 correspondence between proteins and reactions. Spontaneous reactions that proceed without a catalyst have to be included in the active stoichiometric matrix A (so that I accounts for their dilution). They will need a kinetic function that relates their flux to the substrate concentrations (e.g., through mass action kinetics). However, they will not contribute to the protein sum (Eq. (2)) and hence will not directly contribute to the growth equation (6). Because in this case the flux cannot be adjusted by varying the concentration of a catalyst, only concentration vectors are feasible for which the flux through this reaction is identical when calculated based on mass conservation (through I) and on kinetics. This will reduce the dimensionality of the solution space.

In the case of isoenzymes, where both protein j and protein j' catalyze the same reaction, the optimal solution will always use the one with the more favourable kinetics at the given concentrations (e.g., protein

j if $k_j(\boldsymbol{a}) > k_{j'}(\boldsymbol{a}) > 0$).

For protein complexes, where proteins j and j' have to bind to each other before they can act as a catalyst, we can either ignore the individual proteins and include the protein complex as a cellular component in the model, or add a reaction that describes the complex formation.

Finally, if one protein (or protein complex) catalyzes reactions j and j', the substrates (and possibly products) of reaction j' will enter the kinetic function $k_j(\boldsymbol{a})$. The fluxes through both reactions are proportional to the protein concentration p. Hence, $p = v_j/k_j(\boldsymbol{a}) = v_{j'}/k_{j'}(\boldsymbol{a})$, providing an additional constraint for the fluxes $v_j, v_{j'}$. As the fluxes are unique given the concentration vector $\boldsymbol{x} = [P, \boldsymbol{a}]^T$, again not all concentration vectors $\boldsymbol{x}$ will be compatible with this condition, reducing the dimensions of the solution space of balanced growth.

Optimizing only growth. An extension of GBA 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. The dilution term $\mu\boldsymbol{x}$ in Eq. (1) would be replaced by a vector $\boldsymbol{d}(\boldsymbol{x})$ that quantifies the degradation of proteins and other molecules (with entries $d_i = z_i x_i$ and constant degradation rate z_i); an additional "output vector" $\boldsymbol{o}$ would represent the desired cellular production, with rate v_o:

$$A\boldsymbol{v} = v_o\boldsymbol{o} + \boldsymbol{d}(\boldsymbol{x}) \quad . \tag{S30}$$

The kinetics are still a function of $\boldsymbol{x}$, and we can proceed with the analysis following the same steps as for Eq. (1) to calculate $\boldsymbol{v}$, $\boldsymbol{p}$, and v_o.

F. Growth Control Analysis (GCA)

Here, we briefly explore the connection between GBA and some central concepts of *metabolic control analysis* (MCA)[16]. The results below that involve elasticity and control coefficients largely restate previous insights[22,23] in the framework of GBA. First, we rephrase the balance equation in terms of control theory.

We define the (scaled) *growth control coefficients* (GCC) as the *total* relative change in the growth rate due to a small change in the concentration x_i, *accounting for the capacity constraint*. The growth rate change is caused by two effects: the net fitness benefit of increasing x_i without considering the capacity constraint, captured by the marginal net benefits η_i; and the fitness cost of reducing the cellular capacity ρ available for all other concentrations, captured by $-\kappa_i\eta_\rho$. The GCC is then simply the sum of these two,

$$\Gamma_i^\mu = \eta_i - \kappa_i\eta_\rho \quad . \tag{S31}$$

From the balance equation, we have $\Gamma_i^\mu = 0$ at optimal growth. Growth control coefficients have been defined before for changes in growth rate due to small changes in the concentration of individual proteins, in the context of noise propagation in a model of gene expression and cellular growth[21].

We now examine the the relationship of the variables defined in GBA to the coefficients considered in MCA. The *elasticity coefficient* ε_α^j in MCA is defined as the change in the reaction rate j when varying the the substrate concentration a_α while keeping the enzyme (catalyzing protein) concentration fixed[16]. The *normalized* elasticity coefficient is thus directly related to the marginal kinetic benefit u_α^j,

$$\begin{aligned}\varepsilon_\alpha^j &\equiv \frac{a_\alpha}{v_j}\left(\frac{\partial v_j}{\partial a_\alpha}\right)_{p_j=const.} = \frac{a_\alpha}{p_j k_j} p_j \frac{\partial k_j}{\partial a_\alpha} = \frac{a_\alpha}{k_j}\frac{\partial k_j}{\partial a_\alpha} \\ &= \frac{a_\alpha}{p_j} u_\alpha^j \quad .\end{aligned}$$

Control coefficients have been defined in MCA as the change in a *response variable* y due to a change in a *state variable* x, where each $y = y(\boldsymbol{x}, \boldsymbol{\pi})$ is a function of the state variables $\boldsymbol{x}$ and the *system parameters* $\boldsymbol{\pi}$[16]. In the GBA framework, the growth rate μ, the fluxes $\boldsymbol{v}$, the protein concentrations $\boldsymbol{p}$, and the dependent concentrations $\boldsymbol{c}$ are all functions of the concentrations $\boldsymbol{x} = (P, \boldsymbol{b})$, the active matrix A, and the kinetic parameters in $\boldsymbol{k}$, Thus, $\mu, \boldsymbol{v}, \boldsymbol{p}$, and $\boldsymbol{c}$ can be seen as response variables, while the concentrations in $\boldsymbol{x}$ are state variables. In contrast to MCA, the GBA framework provides explicit functions for all response variables, and thus control coefficients can be calculated easily. The control of μ by the concentrations x_i is given by the growth control coefficient Γ_i^μ in Eq. (S31), while the control of dependent concentrations c_γ is directly determined by the dependence matrix D.

We next examine the control of fluxes v_j and protein concentrations p_j. The (scaled) *flux control coefficient* (FCC) $\Gamma_i^{v_j}$ is the relative change in v_j due to a small change in x_i (at fixed concentrations $x_{i'}$ for $i' \neq i$),

$$\Gamma_i^{v_j} \equiv \frac{1}{v_j}\frac{\partial v_j}{\partial x_i} \quad .$$

From Eq.(4), we can calculate $\partial v_j / \partial x_i$, giving

$$\begin{aligned}\Gamma_i^{v_j} &\equiv \frac{1}{v_j}\frac{\partial \mu}{\partial x_i}\frac{v_j}{\mu} + \frac{\mu}{v_j} I_{ji} = \frac{1}{\mu}\frac{\partial \mu}{\partial x_i} + \frac{\mu}{v_j} I_{ji} \\ &= \Gamma_i^\mu + \frac{P}{p_j} q_i^j \quad .\end{aligned}$$

At optimal growth, $\Gamma_i^\mu = 0$, so

$$\Gamma_i^{v_j} = \frac{P}{p_j} q_i^j \quad .$$

Thus, at optimal growth, the flux control coefficient is simply the marginal production cost incurred via reaction j, divided by the protein fraction of the catalyzing protein.

The (scaled) *protein control coefficient* (PCC) $\Gamma_i^{p_j}$ is the change in the protein fraction of protein j, p_j/P, due to a small change in the concentration x_i (at fixed concentrations $x_{i'}$ for $i' \neq i$),

$$\Gamma_i^{p_j} \equiv \frac{1}{P}\frac{\partial p_j}{\partial x_i} \quad .$$

From the kinetic constraint (S2),

$$\begin{aligned}\Gamma_i^{p_j} &= \frac{1}{P}\frac{\partial}{\partial x_i}\left(\frac{v_j}{k_j}\right) = \frac{1}{P}\left(\frac{1}{k_j}\frac{\partial v_j}{\partial x_i} - \frac{v_j}{k_j^2}\frac{\partial k_j}{\partial x_i} + \sum_\gamma D_{\gamma i}\frac{v_j}{k_j^2}\frac{\partial k_j}{\partial c_\gamma}\right) \\ &= \frac{1}{Pk_j}\frac{\partial v_j}{\partial x_i} - u_i^j - \sum_\gamma D_{\gamma i} u_\gamma^j \quad .\end{aligned}$$

Again calculating $\partial v_j/\partial x_i$ from Eq. (4), we obtain

$$\begin{aligned}\Gamma_i^{p_j} &= \frac{1}{Pk_j}\left(\frac{\partial \mu}{\partial x_i}\frac{v_j}{\mu} + \mu I_{ji}\right) - u_i^j - \sum_\gamma D_{\gamma i} u_\gamma^j = \frac{p_j}{P}\Gamma_i^\mu + q_i^j - u_i^j - \sum_\gamma D_{\gamma i} u_\gamma^j \\ &= \frac{p_j}{P}\Gamma_i^\mu - \eta_i^j \quad ,\end{aligned}$$

where we defined $\eta_i^j \equiv -q_i^j + u_i^j + \sum_\gamma D_{\gamma i} u_\gamma^j$ as the contribution of reaction (or protein) j to the marginal net benefit η_i. Summing over j, we obtain

$$\begin{aligned}\sum_j \Gamma_i^{p_j} &= \frac{\Gamma_i^\mu}{P}\sum_j p_j - \sum_j \eta_i^j \\ &= \Gamma_i^\mu - \eta_i - \kappa_i \eta_\rho \quad .\end{aligned} \tag{S32}$$

Without a capacity constraint, $\eta_\rho = 0$, and

$$\sum_j \Gamma_i^{p_j} = 0 \quad . \tag{S33}$$

Equations (S32) and (S33) can be seen as *summation theorems* that relate the GCC Γ_i^μ with the control coefficients of MCA, in a similar fashion as in[21].

At optimal growth,

$$\Gamma_i^{p_j} = -\eta_i^j \quad ,$$

and

$$\sum_j \Gamma_i^{p_j} = -\eta_i \quad .$$

Typically, reactants participate in only a small fraction of reactions, so for most combinations i, j $u_i^j = 0$ and $D_{\gamma i} u_\gamma^j = 0$; the PCC at optimal growth is then just the marginal production cost,

$$\Gamma_i^{p_j} = q_i^j \quad .$$

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

Not every reactant can be considered dependent: a reactant for which 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 $i \in \{P, \beta\}$

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

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. (9), 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., $\kappa_i = 0$, Def. 4). For this reason, the marginal net benefit of i is simply $\eta_i = 0$ (Eq. (10)).

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. (9) is replaced by a constraint on the volume of cellular dry mass per volume of cell water, ν:

$$\nu = \sigma_P P + \sum_\alpha \sigma_\alpha a_\alpha \quad ,$$

where σ_P is the specific capacity of proteins (almost constant for different proteins[53]) and σ_α is the specific capacity of reactant α, which depends on its chemical properties such as hydrophobicity and charge[54].

Supplementary Figures

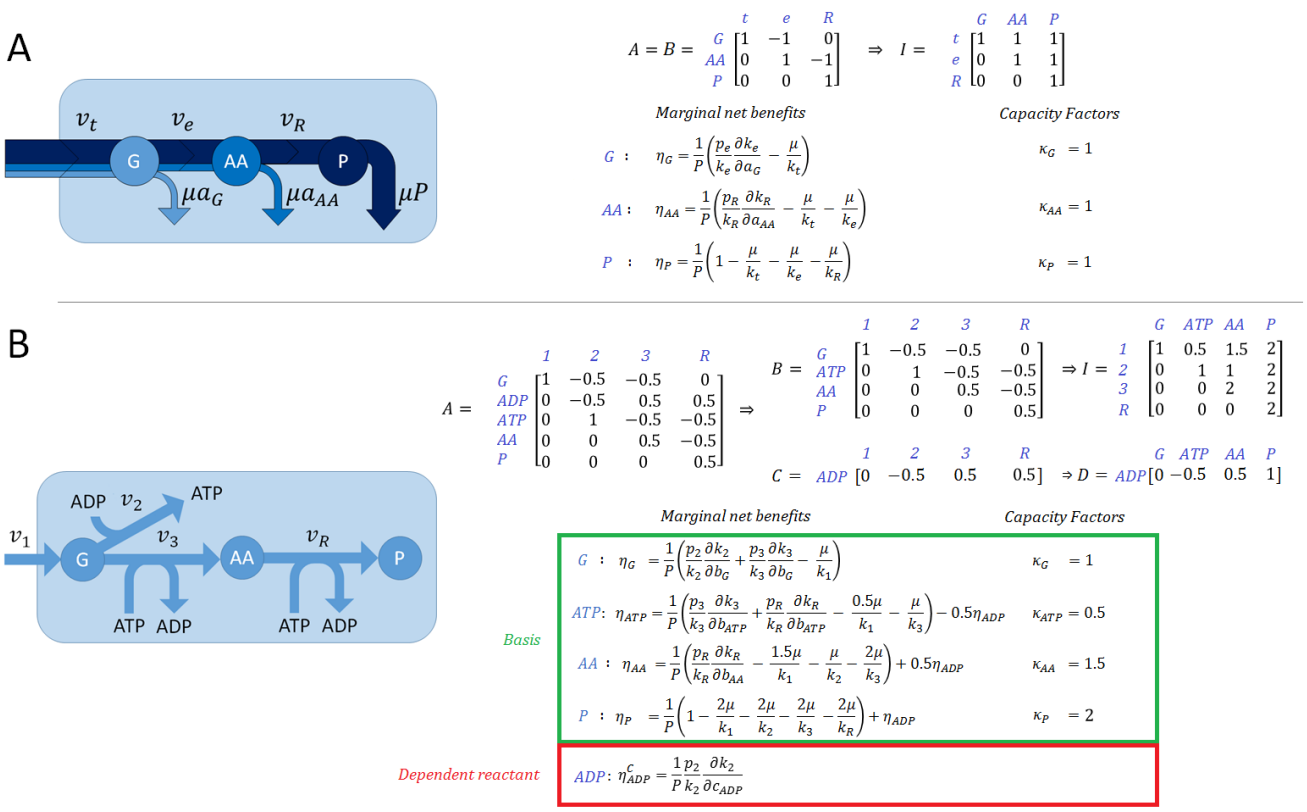

Figure S1. Examples of balanced growth models and their mathematical description, derived from the active matrix A and the kinetic functions $k_j(\boldsymbol{a})$: basis matrix B, investment matrix $I = B^{-1}$, closure matrix C, dependence matrix $D = CI$, marginal net benefits η_i, and capacity factors κ_i. (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 ($n = m + 1$). 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).

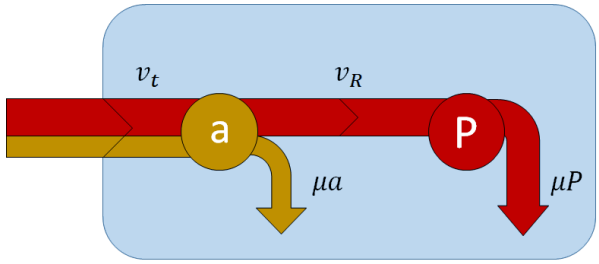

Figure S2. Minimal whole-cell model, comprising a transport reaction (with rate v_t) and the ribosome reaction (with rate v_R).

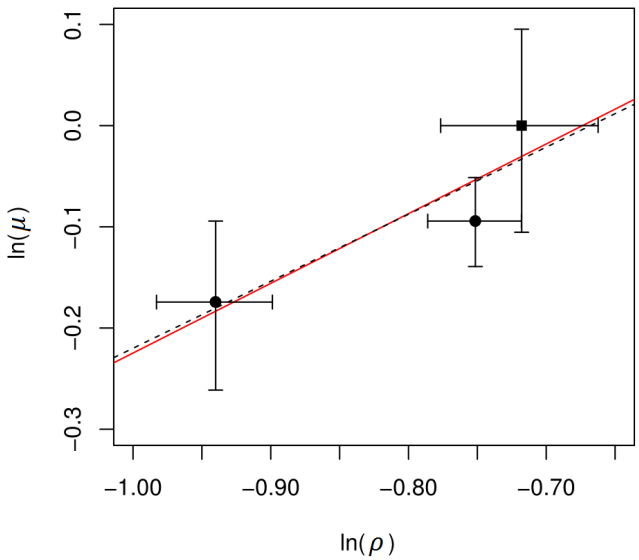

Figure S3. Dependence of the growth rate ($\ln \mu$) on dry weight per free water volume ($\ln \rho$) in *E. coli* grown at different external osmolarities[35]. The square (■) indicates the normal environmental conditions, which correspond to the maximal growth rate; dots (●) indicate growth at lower 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.

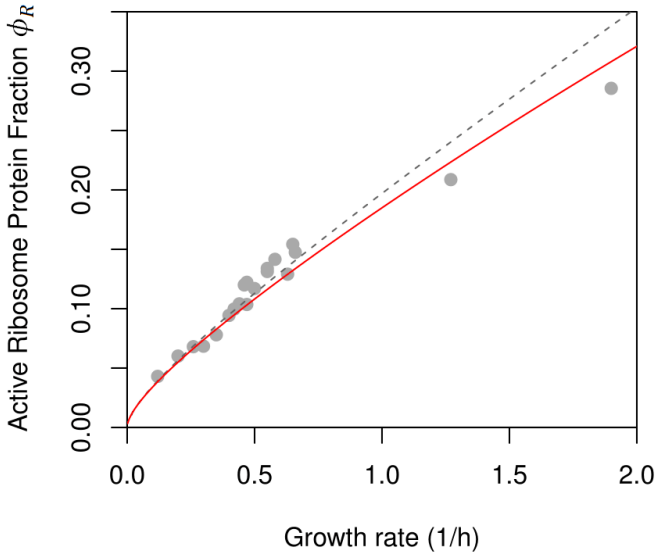

Figure S4. An approximation (dashed grey line, no free parameters; Eq. (S23)) that ignores the dilution of intermediates and hence production costs ($q_i^j \approx 0$) results in good predictions of experimentally observed *E. coli* active ribosome protein fraction[32,33] at low to intermediate growth rates (see also Ref.[34]). For comparison, we also show the full GBA prediction (red line, identical to Fig. 2).

Supplementary Tables

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.*[35], including cellular free water content $\overline{V}_{free}$ and growth rate μ across different external osmolarities, together with the respective cellular dry weight per cellular free water volume, calculated as $\rho = 1/\overline{V}_{free}$.

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
$\mathscr{L}$	Lagrangian
λ_ρ	Lagrange multiplier of the capacity constraint

Table S2. Symbols and definitions. For simplicity of notation, we also use P as an index for total protein, and ρ as an index for cellular dry weight per volume.