QUANTITATIVE PRINCIPLES OF OPTIMAL CELLULAR RESOURCE ALLOCATION

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Hugo de Morais Dourado Neto
aus Recife, Brasilien

Düsseldorf, 12. September 2019

aus dem Institut für Informatik
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Martin J. Lercher

2. Prof. Dr. Oliver Ebenhöh

Tag der mündlichen Prüfung: 19/05/2020

ERKLÄRUNG

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine Universität Düsseldorf“ erstellt worden ist.

Düsseldorf, den 12. September 2019

Hugo de Morais Dourado Neto

If you are receptive and humble, mathematics will lead you by the hand.

— Paul Dirac

ACKNOWLEDGEMENTS

I would first like to thank my supervisor Prof. Martin J. Lercher for providing me his support and the opportunity to work on inspiring projects.

I would like to thank Prof. Oliver Ebenhöh and the members of his group for a short but enjoyable stay with helpful discussions and suggestions.

I thank the German Academic Exchange Service (DAAD) and the German Research Foundation (DFG) for founding my research, including a period as part of the iGRAD-Plant project.

I'm grateful to my colleagues in the CCB group for creating a friendly environment during these years. I would like to thank Deya Alzoubi and Esther Sundermann also for all their willingness to help.

Last but not the least, I would like to thank my family: my mother Rosalia and my sister Mariana for their support capable of crossing continents, Theda for her partnership during all these years, and my son Anton for reminding me to stay curious.

CONTENTS

1 INTRODUCTION

1.1 SUMMARY

This thesis explores the mathematical implications of assuming that resource allocation in unicellular organisms has been optimized by natural selection for growth in fixed environments. In this idealized scenario, regulatory evolution takes place in a constant environment for long enough periods of time to result in "perfectly" adapted cells that do not interact to each other. In reality, cells are subject to changing environments, including shifts to different nutrients [46, 51], and are thus excepted to be under natural selection not only in steady-state conditions, but more broadly on such changing environments. Some unicellular organisms also show cooperative behaviour, and in those cases the microbial community as a whole might better be seen as an evolving unity [8]. Still, the basic theoretical framework presented here is capable of predicting important aspects of cellular resource allocation, in agreement with experimental data for *E. coli*. The corresponding mathematical theory presents an important step toward simulations of detailed, large-scale nonlinear cell models, suggests a mechanistic origin for phenomenological bacterial "growth laws" [56], and clarifies from first principles how the fitness costs and benefits of cellular components are determined by their marginal effects on protein allocation. These costs and benefits, which govern the cellular economy, are shown to relate directly to fundamental concepts in Metabolic Control Analysis.

The structure of this thesis is as follows: Chapter 1 introduces the problem of cellular resource allocation, presents a short review of the most common methods for modeling and analysis of cellular growth, states the aims of the thesis, and contains a overview of the main results in the two manuscripts that form the main body of this thesis, presented in Chapter 2. A final discussion is found in Chapter 3.

1.2 CELLULAR RESOURCE ALLOCATION

A fundamental problem in biology is what principles determine cellular resource allocation. For unicellular organisms, optimization of growth rate at steady state has been shown to explain diverse cellular properties [11], such as how metabolic networks produce biomass [12]. Optimal states of a cell are often described in terms of the costs and benefits of its components; for example, phenomenological pro-

tein costs and benefits have been employed to explain the concentration of Lac proteins at different growth rates [18].

While the costs and benefits of the concentrations of proteins and other molecule types are fundamental to our understanding of evolved biological cells, it is still unclear what exactly determines these costs and benefits. The cost of metabolic reactions is commonly equated to the production cost of the catalyzing enzymes [7, 33, 48, 59], including the required ribosomes, ATP, and carbon. Other studies indicated the importance of the costs related to the limited dry weight capacity of cellular compartments, in particular for enzymes and other macromolecules. Limiting the total protein concentration indeed improves model predictions of maximal biomass production [9, 61] and explains qualitatively the emergence of overflow metabolism [9, 44, 58].

It is commonly assumed that metabolic fluxes are predominantly determined by the concentration of the catalyzing enzymes [7, 9, 12, 58]. In contrast, modelling studies [16, 48], together with metabolite [10] and flux [52] measurements, as well as perturbation experiments [21] indicate that metabolic fluxes are determined by both enzyme and metabolite concentrations. Accordingly, while most authors consider only the costs of enzymes [7, 9, 33, 44, 48, 58, 59], the essence of cellular resource allocation can only be captured by accounting for the costs and benefits of catalysts and reactants simultaneously. The cost and benefit of reactants might then have the same origins as for catalysts; previous studies indicate that the benefit of enzymes and other proteins lies in their local effect on reaction kinetics, while the cost comes from global requirements for the maintenance of their concentrations [3, 18, 54, 57]. Thus, the cost of any cellular component can only be completely understood by a detailed modeling of its consumption of cellular resources, including all processes required for its production.

Genome-scale metabolic models have in recent years been developed for many organisms, commonly in the framework of Flux Balance Analysis (FBA) [50] and its extensions [40]. The cell modeling methods based on FBA share two common concepts: the *stoichiometric matrix* S used to express the mass balance of production and consumption of cellular components, and an objective function that is commonly set to the flux through a *biomass reaction*, which represents the balanced production of the major cellular building blocks (including nucleic acids, amino acids, and lipids) [50]. The biomass reaction is an artificial construction that is useful but comes with some limitations, which are discussed next.

1.2.1 *Biomass*

The biomass reaction has been defined in metabolic models as an artificial reaction that consumes a set of "precursor metabolites" and includes the conversion of ATP to ADP to reflect growth-related energy consumption [50]. In practice, the determination of which "precursor metabolites" are consumed in the biomass reaction, and in which proportions, is not a straightforward procedure and depends on *ad hoc* assumptions [20]. The exact biomass composition (also known as the "biomass equation" in FBA [41]) is mostly not entirely known from experiments, and measurements from closely related organisms are often used instead. It has been shown recently that this strategy is problematic, as even closely related organisms may have very different biomass compositions [38].

In principle, the biomass consists of all constituents of the cell in a given cellular state. However, this "full" biomass definition would be environment-dependent, as especially the concentrations of metabolic intermediates and co-factors can change drastically with the cellular state [23, 53]. For this reason, only a "consensus" of major cellular building blocks is generally assigned to the biomass, and the production rate of intermediate metabolites is enforced only to the extent to which it is required to support the biomass production; this strategy ignores the necessity to create a surplus of the production of intermediates to offset their dilution through cellular growth.

The difficult distinction between what should be in the biomass reaction and what not is circumvented by self-replicating models of balanced cellular growth, which demand that all components are produced in proportion to their concentrations [44]. The stoichiometric matrix of such a self-replicator model contains explicit information on how its components are produced and consumed, including the production of catalysts, which is not explicitly modeled in FBA. The assumption of balanced growth means that the rate of production of all components can be determined exactly from their concentrations. These concentrations are variables of the self-replicator problem. Thus, the fixed biomass reaction is replaced by the dilution of all cellular components through growth, and growth rate predictions do not depend on *ad hoc* assumptions on cellular composition.

Strikingly, important sectors of the protein part of biomass have been observed to change with growth rate in a simple linear fashion in *E. coli* [32, 56]. The corresponding empirical "growth laws" of protein allocation are a central piece in the current understanding of global resource allocation in bacteria [6].

1.2.2 *Growth laws*

It has long been observed that for bacteria growing on media supporting fast to moderately slow growth (e.g., 20 min to 2 hours per doubling), ribosome content increases linearly with growth rate [43]. More recently, Scott et al. [56] have further shown that *E. coli* ribosome content decreases linearly with growth rate when translation is inhibited by sublethal doses of the antibiotic chloramphenicol, and have collected these empirical correlations in a set of bacterial "growth laws". The ribosome proteome fraction ϕ_R at nutrient-limited growth depend on growth rate μ as

$$\phi_R = \phi_{R,0} + w_R \mu \tag{1}$$

where $\phi_{R,0}$ is the offset at zero growth and w_R relates to the ribosome translational efficiency. These phenomenological laws were suggested to play a practical role in theoretical biology comparable to that of earlier phenomenological laws in physics; one example is Ohm's law, which facilitated the design of electrical circuits before a microscopical understanding of electricity [56].

Hui et al. [32] have shown that other protein sectors also behave in a fashion similar to ribosomes, changing their overall concentration linearly with growth rate in response to different limitations imposed on cellular growth: catabolic section limitation (C-lim) by titrating the expression of lactose permease during growth on lactose; anabolic section limitation (A-lim) by titrating the enzyme GOGAT in the ammonia assimilation pathway; and ribosome limitation (R-lim) by sublethal amounts of an antibiotic inhibiting translation, chloramphenicol. The direction (up or down) of the growth-rate dependence of individual protein concentrations under each of the three limitations allows their grouping into 6 different protein sectors, which are related to different classes of cellular processes [32]. Analogous to the growth laws defined for the ribosome, all proteome sectors show a growth rate-independent offset, which is not explained by the optimal proteome allocation scheme presented in Ref. [32]. For the R-sector (composed by ribosomal and other translation proteins), it has been suggested that the offset is due to a fraction of non-translating ribosomes [56] or to the growth rate dependence of tRNA concentrations [35]. For the other sectors, Hui et al. [32] suggest that offsets are due to the biophysical difficulty to fully repress gene expression at zero growth rate, and the possible advantage conferred by the ability to adapt more quickly to changing environments [19, 34, 36].

Mori et al. [46] studied the possible advantage of the ribosomal offsets at famine-and-feast cycles. Based on a model for the transition kinetics from poor to rich media validated with *E. coli* data, the expected benefit is shown to be a function of the feast duration. The ribosome offset present in *E. coli* is predicted to be an optimal reserve

for feasts around 2–3h, which coincides with the feast period in the human gut microbiota [46].

The possible advantage of the over-expression of proteins in changing environments has also been explored by O'Brien et al. [51]. A fraction of "under-utilized" proteins in *E. coli* is determined as the difference between the experimental core metabolism proteome abundance [53] and the corresponding protein demand at different growth rates, estimated by using the ME-model for *E. coli* [49] (see Section 1.3.2). The benefit of under-utilized proteins on faster adaptation to improved growth conditions is estimated using ME-model simulations between the growth on galactose followed by a shift to growth on glucose. The simulations show that there is increase in the growth rate immediately after the shift, if under-utilized proteins were expressed before on the galactose growth.

However, these dynamical considerations fail to justify why the offset increases protein concentration above a simple proportionality also at high growth rates, an observation that appears inconsistent with the assumption that offsets serve as a preparation for better growth conditions. Thus, the origin of the offsets is not yet fully understood.

1.3 MODELING AND ANALYSIS OF CELLULAR REACTION NETWORKS

Next, I present a review of FBA and related constraint-based methods, together with their limitations and the advantages of the more general self-replicator models. Such self-replicator models will be the focus of the first manuscript included in Chapter 2.

1.3.1 *Flux Balance Analysis (FBA)*

Flux Balance Analysis (FBA) is currently the most widely used methodology to simulate cellular metabolism on a genome scale. FBA relies on the stoichiometry of reconstructed metabolic reactions, using linear programming to find the distribution of fluxes that maximizes a linear function of fluxes. Commonly, the objective function is set to the production rate of biomass per uptake rate (yield), which is often considered a proxy for growth rate, despite the fact that in general the maximization of molar yield is not favored by evolution [55].

In its standard application, FBA finds the maximal biomass production rate v_{bio} and its corresponding flux distribution $\boldsymbol{v}$ for a given stoichiometric matrix S, biomass reaction consuming (and in some cases producing) reactants with concentration $\mathbf{r}$, and upper ($\mathbf{u}$) and lower

(l) bounds on fluxes. Formally, this defines the linear optimization problem

$$\begin{aligned} &\underset{\boldsymbol{v}}{\text{maximize}} && v_{\mathtt{bio}} \\ &\text{subject to:} && \\ & && S\boldsymbol{v} = v_{\mathtt{bio}}\mathbf{r} \\ & && \mathbf{l} \leqslant \boldsymbol{v} \leqslant \mathbf{u} \quad , \end{aligned} \tag{2}$$

where the inequalities are defined element-wise. The lower and upper bounds on fluxes can be used to put a limit on maximal uptake rates or minimal excretion rates, or to enforce reaction directions when these are known. It is common to incorporate the biomass reaction as a column of the stoichiometric matrix with entries equal to $\mathbf{r}$ scaled by the cellular dry weight. Then $v_{\mathtt{bio}}$ is an element of the vector $\boldsymbol{v}$, and mass balance is expressed as [50] $S\boldsymbol{v} = 0$. This thesis does not use that convention in order to permit a more straightforward comparison with methods in which $\mathbf{r}$ is not fixed.

The linear optimization defined by Eq. 2 is simple to solve, but the computational ease is bought at the price of some conceptual limitations. First, the maximization of $v_{\mathtt{bio}}$ is constrained predominantly by the upper bound of the uptake reactions (e.g., the uptake rate of a limiting carbon source), so – in the absence of a growth-independent maintenance energy term – the solution corresponds to the maximal yield (ratio between uptake flux and biomass production flux), not necessarily the maximal growth rate. Furthermore, the explicit dependence of biomass production on uptake fluxes means that optimal biomass production can only be predicted if uptake fluxes are known.

Second, for realistic large-scale models, S has more columns than rows, which means there is no unique solution [50] for $\boldsymbol{v}$. Thus, one typically needs extra criteria to specify a unique prediction of the flux distribution $\boldsymbol{v}$. The most common method for this is *parsimonious FBA* (pFBA) [30], which minimizes the sum of absolute fluxes at the maximal biomass production rate; this strategy approximately reflects the expectation that the cell will minimize its resource investment into enzymes.

1.3.2 *Extensions of FBA*

Beg et al. [9] introduced a generalization of FBA that accounts for how the limitation on total macromolecular concentration limits enzyme concentration and hence metabolism, termed FBA with molecular crowding (FBAwMC). This method accounts for the (average) dependence of the fluxes on enzyme concentrations and kinetics at fixed metabolite concentrations. It assumes that the maximum mass concentration of enzymes is limited by the cytoplasmic protein density, assumed to be constant. This results in an additional constraint

for the extended FBA problem: the sum of fluxes weighted by an average *crowding coefficient* corresponds to the sum of enzyme concentrations, which cannot exceed a maximal value; this coefficient is determined by fitting growth rate predictions to experimental measurements. The flux predictions obtained with FBAwMC and related methods are closer to experimental values than those obtained from simple FBA [1, 9, 61].

Adadi et al. [1] introduced a generalization of FBAwMC termed MOMENT (MetabOlic Modeling with ENzyme kineTics), which accounts for the kinetic limitation on each *individual* reaction as $v_j \leqslant e_j k_{cat,j}$, with $k_{cat,j}$ being the catalytic constant of reaction j. Accounting for gene-to-reaction mapping, MOMENT was shown to improve FBAwMC flux predictions [1].

Goelzer et al. [25] have introduced a further generalization of FBA-type models, named Resource Balance Analysis (RBA). This method extends the cellular modelling to resource allocation outside metabolism, by accounting for the concentration of i) the necessary translation apparatus to produce proteins and ii) other macromolecules necessary for growth, such as DNA and the lipid membrane (assumed to have a constant "concentration" independent of growth rate). RBA also extends the constraint on the total concentration of macromolecules by accounting individually for limitations imposed within different cellular compartments, and extends the kinetic constraint by assuming a growth-rate (μ) dependent "apparent catalytic rate" $k_j(\mu) = a_j\mu + b_j$ for each enzyme j, where the corresponding flux is constrained by $v_j = e_j k_j(\mu)$ and the parameters a_j, b_j are obtained by fitting proteomics data to reaction fluxes. Based on this calibration, RBA has been shown to be able to predict proteome allocation in *B. subtilis* [26].

Lerman et al. [39] described an alternative way to model metabolism while accounting for the gene expression machinery, termed metabolism and macromolecular expression models (ME-Models). ME-Models generate a feasibility problem which searches for the maximal growth rate achievable with a flux distribution that requires the minimal ribosome production. Conceptually, ME-Models claim to be able to predict macromolecular composition (biomass) without requiring the assumption of a fixed biomass composition as FBA. In practice, the ME-Model for *T. maritima* presented by Lerman et al. [39] requires a "structural reaction" that comes directly from the corresponding FBA model for this organism. As in MOMENT, kinetics in ME-Models provide only an upper bound for reaction fluxes based on the corresponding k_{cat} [39].

Mori et al. [45] have introduced a variation of FBA that explicitly accounts for the growth rate dependence of proteome sectors, named Constrained Allocation Flux Balance Analysis (CAFBA). This method divides the proteome into 4 sectors: ribosome-affiliated (R-

sector), biosynthetic enzymes (E-sector), carbon intake and transport (C-sector), and core housekeeping sector, the concentration of which is independent of the growth rate (Q-sector). The corresponding proteome fractions ϕ_R, ϕ_E, ϕ_C, ϕ_Q sum up to one, and ϕ_C, ϕ_E are assumed to depend linearly on their fluxes v_C, v_E, based on experimental evidence [32, 64] and mimicking the growth law Eq.(1) for the ribosome sector [56]. With some additional assumptions, including Michaelis-Menten dependence of the C-sector flux on the external sugar concentration and again a linear dependence of each enzyme protein fraction ϕ_j on its flux v_j, a new constraint is included into the FBA problem Eq.(2)

$$w_C v_C + \sum_j w_j |v_j| + w_R \mu = \phi_{max} \quad , \tag{3}$$

where w_C is the C proteome fraction per unit of carbon influx (which depends on the external sugar concentration), w_j is the enzyme j proteome fraction per unit of flux j (absolute flux value because it does not depend on reaction direction), w_R is set to an empirical value ≈ 0.169 h [13], and $\phi_{max} = 1 - \phi_Q - \phi_{C,0} - \phi_{E,0} - \phi_{R,0}$ is the proteome fraction allocated to the growth rate dependent components of the protein sectors, which is set to 0.484 based on a previous estimation for *E. coli* [56]. The values w_j can be set in two different ways: (i) all are set equal to the same value w_E fitted to support the empirical maximal growth rate, or (ii) random values for each w_j are drawn from a distribution with mean $< w >$. The average predictions of acetate excretion and growth yield (gram of dry weight per gram of glucose) using different values for w_C to emulate carbon limitation provide a good fit to experimental values for the strain MG1655 using $< w >= 1.55 \times 10^{-3}$ gh/mmol and for the strains NCM3722 and ML308 using $< w >= 8.8 \times 10^{-4}$ gh/mmol.

All previously described methods assume the cell is in an optimal state, identified in the model by solving a linear optimization problem. However, due to the inherently non-linear nature of enzymatic rate laws, molecular biological systems are inherently nonlinear. Problems in genome-scale non-linear models, which comprise hundreds or even thousands of variables [50], are difficult or even impossible to solve numerically with current computational means. For this reason, FBA-based methods make some conceptual compromises in order to avoid formulating nonlinear optimization problems. Reaction fluxes, for example, do not account for substrate concentrations and by consequence are assumed to be a linear function of the catalyzing protein concentrations only ($v_j = e_j k_j$). The assumption that the k_j are constant across conditions, which is assumed in almost all models, may strongly affect the methods' predictive power, as it misses the importance of changes in reactant concentrations for enzyme efficiency. One alternative to the assumption of constant enzyme saturation is to use *ad hoc* rules to fit model predictions to

experimental values [9, 25, 39]; however, this phenomenological approach defeats the purpose of understanding and predicting how cell behaviour emerges from first principles.

1.3.3 *Self-replicator models*

The inherent non-linearity of biological systems can be treated exactly in simplified cellular models with only a few reactions. In an exponentially growing cell with growth rate μ, each component with concentration x needs to experience a net production rate $dx/dt = \mu x$ in order to balance its dilution by cellular volume growth. This direct connection between concentrations and growth rate makes it possible to formulate an optimization problem that directly maximizes growth rate instead of just yield. Molenaar et al. [44] proposed an optimization framework for cellular growth that accounts for the most important physicochemical constraints, assuming balanced exponential growth of a self-replicator model with up to 7 reactions.

In their mathematical description of balanced growth, Molenaar et al. [44] considered the maximization of growth rate μ under the major physicochemical constraints on cellular growth:

1. the mass balance of reactants, as in FBA; note, however, that here no particular biomass composition has to be assumed, as all reactants need to be produced in proportion to their modeled concentrations;

2. the production and mass balance of total protein used by the cell; there is a "ribosome" reaction that produces proteins at the same rate as proteins are diluted by growth;

3. the kinetics of all biochemical reactions, each of which depends on the concentrations of both the catalyst (ribosome, enzyme, or transporter) and of the reactants involved, described by non-linear kinetic rate laws;

4. an upper bound P on the total (molar) protein concentration, reflecting finite cellular capacity [1, 9, 25, 39];

5. an upper bound on the total transporter protein density on the cellular membranes, which stems from the necessity to have a minimal density of lipids to retain membrane structure [65].

The resulting optimization problem aims to find the concentrations of reactants $\mathbf{r}$ and proteins $\mathbf{p}$ (and the fluxes $\boldsymbol{\nu}$ resulting from the concentrations according to the kinetic rate laws $\mathbf{k}(\mathbf{r})$) that maximize

growth rate μ, for a given set of external reactant concentrations and a given limit on total protein concentration P,

$$
\begin{aligned}
&\underset{\mathbf{r},\mathbf{p}}{\text{maximize}} \quad \mu \\
&\text{subject to:} \\
&\qquad S\boldsymbol{\nu} = \mu\mathbf{r} \\
&\qquad \nu_R = \mu P \\
&\qquad \forall_j \;\; \nu_j = p_j k_j(\mathbf{r}) \\
&\qquad \sum_j p_j \leqslant P \\
&\qquad \sum_t p_t \leqslant r_l \\
&\qquad \mathbf{p}, \mathbf{r} \geqslant 0 \quad ;
\end{aligned}
\tag{4}
$$

here, t indexes transport proteins and l indexes lipids in the cellular membrane. For clarity of presentation, the optimization is formulated here in a slightly modified (but mathematically equivalent) manner compared to the original formulation by Molenaar et al. [44].

It is important to note two important differences between the optimization scheme (4) and the previously discussed constraint-based optimization schemes based on FBA: (i) the reactant concentrations $\mathbf{r}$ are not fixed, but are optimized for a given growth condition; (ii) concentrations are constrained to be non-negative, but there are no explicit lower and upper bounds on fluxes; limits on fluxes (including flux directions) instead arise from constraints on the concentrations of the molecules involved and from reaction kinetics, which in principle account for thermodynamics and regulation (e.g. allosteric). The first difference is particularly important: the full biomass composition varies across growth conditions and can be seen as an evolved environmental response of the cell; through changes in the reactant concentrations, it strongly influences environment-dependent kinetics.

Because this framework integrates all necessary parts of cellular physiology for self-replication (not just metabolism), it is in principle capable of predicting behaviour that emerges from the function of the cell as a whole [44]. There are, however, two main limitations of this approach: first, the optimization is assumed to be solved with numerical methods, which are capable of managing only a limited number of reactions (e.g., up to 7 reactions in Ref. [44]). Second, this approach requires quantitative knowledge of all kinetic parameters. Kinetic parameters are mostly unknown even for the best studied species; e.g., only about 10% of catalytic constants k_{cat} are known in *E. coli* [17]). Moreover, it is not clear how to best choose kinetic parameters in coarse-grained models, and so these are typically set to arbitrary values [44] or fitted to data [18, 24].

1.4 AIMS OF THIS THESIS

This thesis aims primarily to develop an analytical theory for the balanced growth of self-replicator models. Such a theory not only represents an important step toward the application of the self-replicator scheme to cellular models of any size, it also provides a deeper understanding of cell biology, by relating cellular concentrations, fluxes, and the growth rate at optimal growth through fundamental quantitative principles. Approximations to these principles will provide a rigorous basis for quantitative descriptions of optimal cellular growth in the general case of arbitrarily large systems for which kinetic parameters are not fully known.

1.5 NOTE ON THE CONTRIBUTION OF COLLABORATORS

The two manuscripts included in this thesis are the results of scientific collaborations. For the first manuscript, "An analytical theory of balanced cellular growth", I developed the GBA framework, performed all data analyses, and derived all formal results except Theorems 1-4. This manuscript has been invited for re-submission after peer review at *Nature Communications*. For the second manuscript, "Optimal catalyst and substrate concentrations in cells", I derived all formal results except Eqs.(S36-37). Together with Martin Lercher, I conceived both studies, interpreted the results, and wrote the manuscript. Xiao-Pan Hu provided the model for protein translation. Terry Hwa suggested the link to the proteome offsets in the bacterial growth laws, gave advice on data processing, and, together with Matteo Mori, helped interpret the results of the second study. Matteo Mori also helped with data processing for the second study.

Below, I use the pronoun "we" to refer to my work presented in these manuscripts.

1.6 OVERVIEW OF THE MAIN RESULTS

The first manuscript in this thesis develops an analytical basis for a general theory of cellular growth, which we termed Growth Balance Analysis (GBA). The analytical approach permits to study and predict general features of resource allocation in cell models of any size, unlike previous methods that rely on numerical optimization and are hence confined to small, coarse-grained models [44]. In the form presented in this thesis, GBA considers the simplest structural model of a complete biological cell containing a unique compartment in steady state, accounting for the concentrations of all growth-related cellular components, assumed to be constant over time. The latter assumption means that the cell increases its volume at the same rate as it produces dry weight, by increasing its water content at the same rate.

The theory incorporates a general limit on total cell dry weight, irrespective of a possible compartmentalization of cellular components. The limitation of total dry weight at fixed external osmolarity is a real, experimentally observed phenomenon [4, 14, 15].

Following the framework presented by Molenaar et al. [44], we assume that a cell is composed of two fundamental types of components: proteins that are the catalysts of reactions (ribosome, enzymes, and transporters), and reactants (metabolites, possibly RNA and DNA, and any additional component other than protein that is produced or consumed by reactions). Reactants are transported through the membrane by transporters and/or converted into other reactants by enzymes or into protein by the "ribosome" reaction. (Spontaneous reactions can be modeled by proteins with extremely high catalytic constants and extremely low Michaelis constants). This simple structural model provides a reasonable approximation to prokaryotic cells, while for eukaryotic cells an explicit consideration of cellular compartments will likely be more important.

We modify the mathematical optimization introduced in Ref. [44] by incorporating the reactants into the capacity constraint in addition to proteins. As a consequence, the total protein mass concentration P is not fixed but becomes a variable over which we optimize.

The analytical theory assumes knowledge of which reactions are active, i.e., have non-zero fluxes. Based on previous results [47, 63], we demonstrate that for any fixed set of concentrations, the corresponding set of active fluxes at maximal growth rate is an *Elementary Flux Mode* (EFM) of a corresponding linear problem, and by consequence the corresponding active stoichiometric matrix indeed has full column rank [22] (see theorems 1-4 in manuscript 1).

For a given active matrix A, kinetic functions $\mathbf{k}$ (accounting for external fixed concentrations), and cell dry weight density ρ, the general optimization problem considered in this thesis can then be formulated as:

$$
\begin{aligned}
&\underset{\mathbf{a},\mathbf{p}}{\text{maximize}} \quad \mu \\
&\text{subject to:} \\
&\qquad A\boldsymbol{\nu} = \mu \begin{bmatrix} P \\ \mathbf{a} \end{bmatrix} \\
&\qquad \forall_j \;\; \nu_j = p_j k_j(\mathbf{a}) \\
&\qquad P = \sum_j p_j \\
&\qquad P + \sum_\alpha a_\alpha = \rho \quad ,
\end{aligned}
\tag{5}
$$

where a_α are the corresponding active reactants, i.e. reactants with non-zero concentration. The mass balance of total protein and reactants is considered together, constrained by the growth rate μ and the

active matrix A (which we define such that the first row of A corresponds to the protein production by a single "ribosome" reaction, so that the resulting balance equation is simply $\nu_R = \mu P$).

This problem is reformulated in a way such that the objective function μ and the constraints become explicit functions of the concentration vector $[\mathbf{a}, P]^T$ of total protein and the reactants

$$\begin{aligned} &\underset{P,\,\mathbf{a}}{\text{maximize}} \quad \mu = \frac{P}{\sum_j \frac{I_{jP}P + \sum_\beta I_{j\beta} b_\beta}{k_j(\mathbf{a})}} \\ &\text{subject to:} \\ &\qquad P + \sum_\alpha a_\alpha = \rho \\ &\qquad \forall_\gamma \;\; c_\gamma = D_{\gamma P}P + \sum_\beta D_{\gamma\beta} b_\beta \quad ; \end{aligned} \tag{6}$$

here,

- $\mathbf{b}$ are the reactant concentrations corresponding to the linearly independent rows of A;
- $\mathbf{c}$ are reactant concentrations corresponding to the linearly dependent rows of A;
- the matrix I is the inverse of the submatrix B of linearly independent rows of A;
- the matrix D is defined as $D \equiv CI$, where C is the submatrix of A that corresponds to the linearly dependent rows.

Figure 1 shows an example of a simple balanced growth model and its mathematical description in the GBA framework.

This reformulation facilitates three different theoretical results:

1. an explicit expression for fitness costs and benefits of cellular components;
2. the analytical necessary conditions for optimal growth (using the method of Lagrange multipliers);
3. an extension of metabolic control analysis (MCA) for balanced growth models, termed Growth Control Analysis.

These three theoretical results are closely related; the framework presented here clarifies the precise mathematical connection and facilitates the integration of these different areas of theoretical cell biology.

In principle, one could simplify this problem even further by incorporating the linear constraints in Eq. (6) into the equation for μ, thereby formulating a maximization problem without constraints, which depends only on $\mathbf{b}$ and ρ. This will not be done here, as it

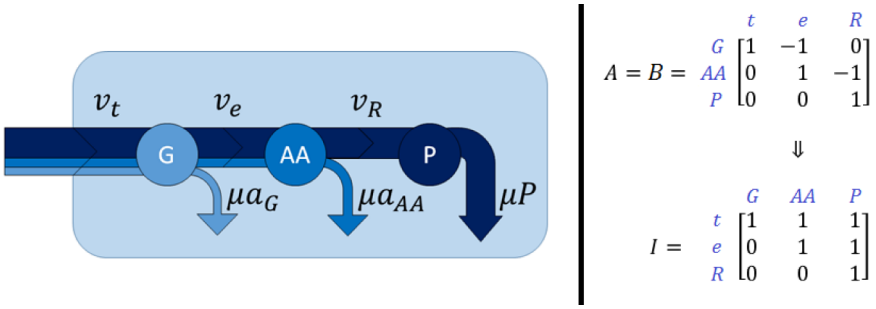

Figure 1: Example of a simple balanced growth model. Mass conservation requires that the mass flux transported into the cell is the same as that partitioned into the dilution by growth of each cellular component. The active matrix A in this case is a square matrix (as in any model with a linear reaction network), so there are no dependent reactants, and $B = A$. The mathematical description of the model is completely parametrized by the active matrix A, kinetic functions $\mathbf{k}$ (accounting for external reactant concentrations), and dry weight density ρ. The cellular state is then uniquely determined by the concentrations of total protein P and independent reactants $\mathbf{b} = \mathbf{a}$.

is biologically relevant to evaluate the individual importance of the constraints themselves; the formulation in Eq.(6) presents a more direct way to do that by using the envelope theorem [2]. In particular, we will use this formulation to quantify the influence of the capacity constraint on the maximally achievable growth rate.

The optimization problem presented in Eq.(6) can be solved analytically with the method of Lagrange multipliers. This is not true for the original problem in Eq.(5) with a general non-square matrix A; in this case, the linearly dependent rows of A imply linearly dependent gradients of the corresponding constraints (each row of A results in one constraint), and therefore the method of Lagrange multipliers cannot be applied directly.

We define the *marginal net benefit* of reactants and total protein as the relative change in growth rate caused by a small change in their concentration, ignoring the capacity constraint. We show that optimal growth requires the equality between these marginal net benefits and the marginal benefit of the capacity constraint itself, defined as the relative change in maximal growth rate due to a small increase in the cell dry mass capacity; we term these relationships *balance equations*.

The application of GBA depends on a full model parameterization of the kinetic functions; however, kinetic parameters (in particular turnover numbers k_{cat}) are in most cases not available in the literature [17]. To overcome this limitation, in practice one might explore an approximation ignoring production costs, as developed in manuscript 2.

Manuscript 2 considers the simpler optimization of dry weight density use constrained by kinetics. The corresponding first order neces-

sary conditions are identical to the balance equations with the production costs set to zero. For irreversible Michaelis-Menten kinetics, the balance equation for each reactant α equates its mass concentration with the mass concentration of unbound enzymes using α as a substrate. Strikingly, there is no dependency on the reaction fluxes $\boldsymbol{\nu}$ nor on the catalytic constants of reactions $\mathbf{k}_{cat}$, but only on the concentrations $\mathbf{p}$, a_α and the Michaelis constants K_m. The predicted optimal metabolite concentrations agree well with experimental values from *E. coli* [23, 53].

The optimal balance of catalyst and substrate concentrations also implies a unique relationship between the reaction flux and the optimal protein concentration. At higher reaction fluxes – i.e., when the flux is much higher than the predicted optimal flux at half saturation, $\nu \gg k_{cat}K_m$) – the enzyme is almost saturated with its product, and the protein concentration p is approximately equal to a linear term in ν, as expected from the empirical growth laws [56]. At lower reaction fluxes (i.e., when the flux is much lower than the predicted optimal flux at half saturation $\nu \ll k_{cat}K_m$), p depends mainly on a square root term in ν. We thus hypothesize that the combination of these two terms might explain from first principles the long observed relationship of ribosome [32, 43, 56, 57] and other catalyst concentrations [32, 56, 57] with growth rate, more recently understood in terms of the microbial growth laws [56]. In the cases where the dependence of the reaction flux on growth rate is known, there is a direct prediction for catalyst concentrations as a function of growth rate. This is the case for the ribosome and metE, the most abundant metabolic enzyme in *E. coli* growing on minimal media [53], as both fluxes must satisfy the known demand arising from growth-related protein production. The predictions for both cases assuming simple irreversible Michaelis-Menten kinetics [27, 35] are in good agreement with experimental values [53]. While the ribosome prediction results in an apparent offset, the predicted metE offset is indistinguishable from zero.

2

MANUSCRIPTS

2.1 AN ANALYTICAL THEORY OF BALANCED CELLULAR GROWTH

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

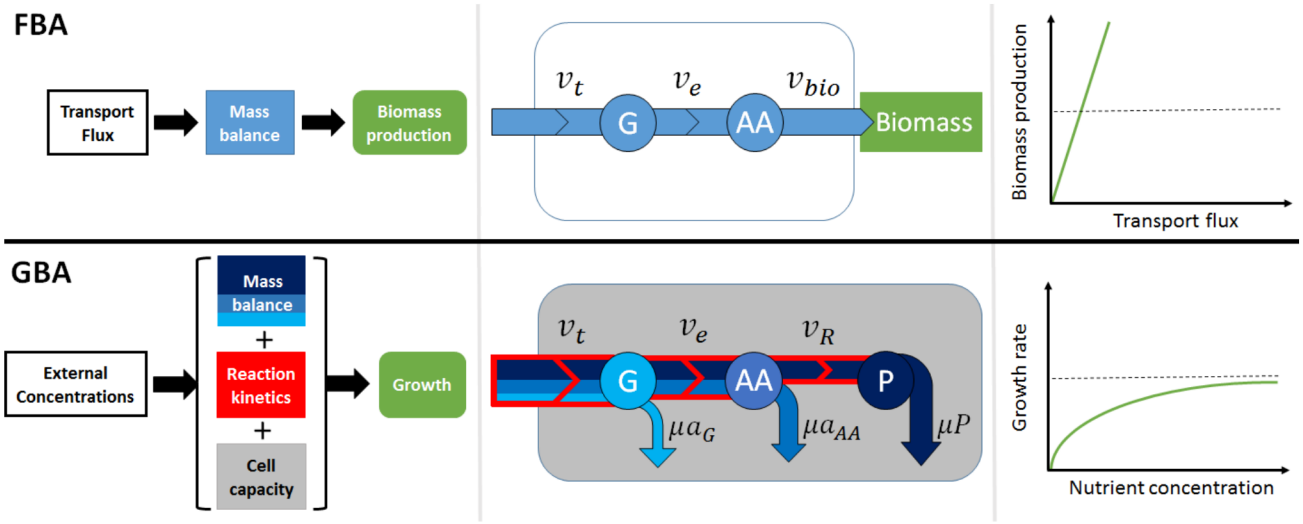

Figure 1. A schematic comparison of flux balance analysis (FBA, top) and growth balance analysis (GBA, bottom) for a simple toy model. A nutrient G is taken up through a transporter t at rate v_t and is then converted by an enzyme e with rate v_e into a precursor for protein synthesis, AA. In FBA, AA 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 $\phi_j \equiv p_j/P$ at constant μ facilitated by the increase in x_i,

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

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*:

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

(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.

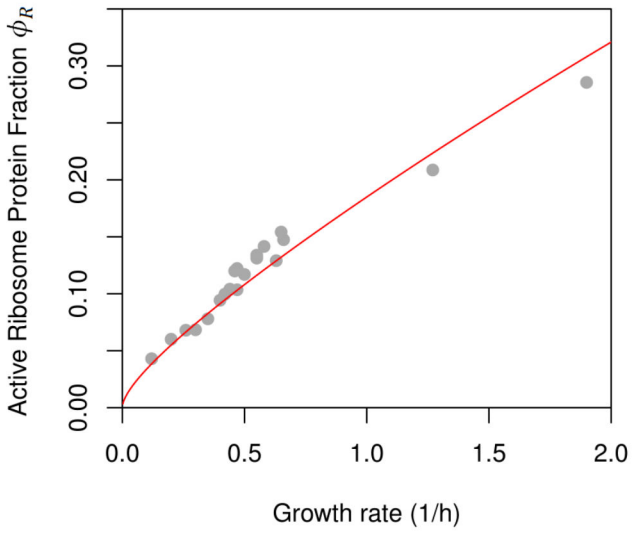

Figure 2. The GBA prediction of the active ribosomal protein fraction in *E. coli* agrees with experimental values. Comparison of GBA predictions (red line, no free parameters) and data across 20 different growth conditions[32,33] (grey dots) results in a Pearson correlation coefficient $R^2 = 0.96$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

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 $\boldsymbol{k}^{(\mathrm{eff})} \in \mathbb{R}^n_{\neq 0}$ 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:*

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

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

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

*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 $\boldsymbol{k}^{(\mathrm{eff})}$ 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*

$$\begin{aligned}\eta_P^0 &= \frac{1}{P}-\sum_j q_P^j\\ \eta_\beta^0 &= \sum_j (u_\beta^j - q_\beta^j) \quad .\end{aligned}$$

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,

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

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 [time]$^{-1}$) and Michaelis constant K_{m} (in units of [mass][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 (± 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, so in real systems Γ_i^μ might provide an objective measure of how "non-optimal" concentration x_i is. A related definition of growth control coefficient have been introduced before 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 (scaled) elasticity coefficient is thus directly related to the marginal kinetic benefit u_α^j,

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

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

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

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

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

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

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

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

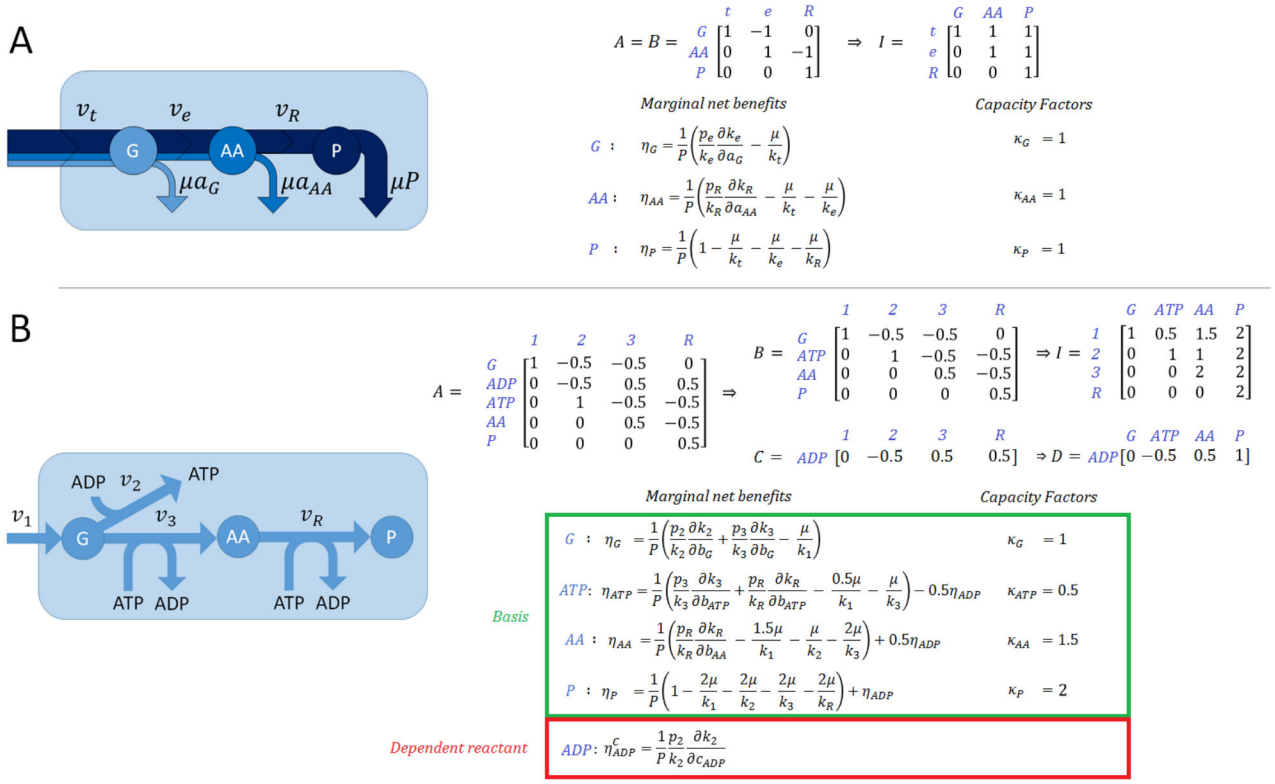

Figure S1. Examples of balanced growth models and their mathematical description, derived from the active matrix A and the kinetic functions $k_j(\boldsymbol{a})$: basis matrix B, investment matrix $I = B^{-1}$, closure matrix C, dependence matrix $D = CI$, marginal net benefits η_i, and capacity factors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

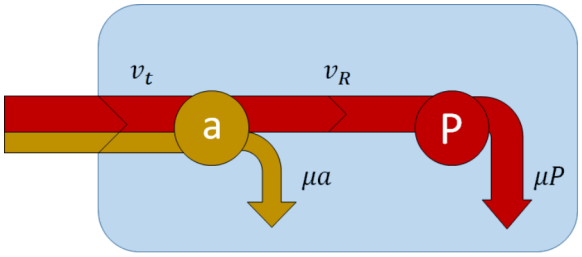

Figure S2. Minimal whole-cell model, comprising a transport reaction (with rate v_t) and the ribosome reaction (with rate v_R).

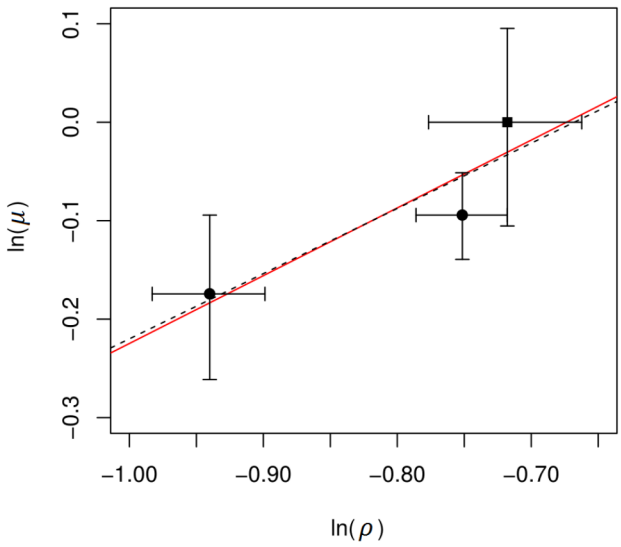

Figure S3. Dependence of the growth rate ($\ln \mu$) on dry weight per free water volume ($\ln \rho$) in *E. coli* grown at different external osmolarities[35]. The square (■) indicates the normal environmental conditions, which correspond to the maximal growth rate; dots (●) indicate growth at lower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

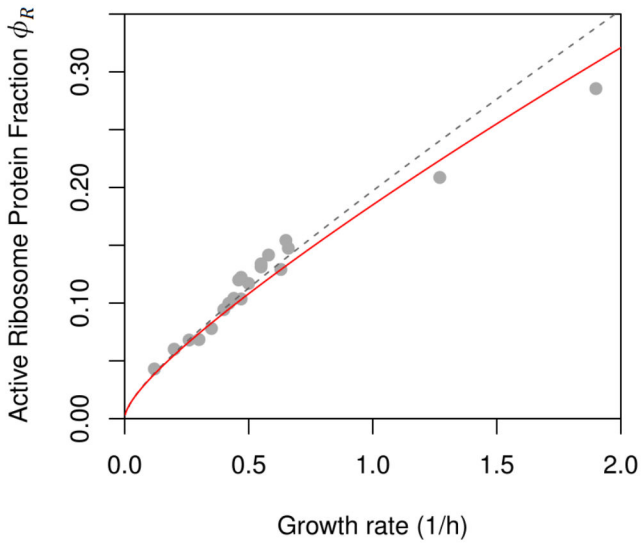

Figure S4. An approximation (dashed grey line, no free parameters; Eq. (S23)) that ignores the dilution of intermediates and hence production costs ($q_i^j \approx 0$) results in good predictions of experimentally observed *E. coli* active ribosome protein fraction[32,33] at low to intermediate growth 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.

2.2 OPTIMAL CATALYST AND SUBSTRATE CONCENTRATIONS IN CELLS

Optimal catalyst and substrate concentrations in cells

Hugo Dourado[1], Xiao-Pan Hu[1], Matteo Mori[2], Terence Hwa[2], and Martin J. Lercher[1,*]

[1]Institute for Computer Science & Department of Biology, Heinrich Heine University, 40221 Düsseldorf, Germany
[2]Department of Physics, University of California at San Diego, La Jolla, CA 92093-0374, U.S.A.
[*]to whom correspondence should be addressed: martin.lercher@hhu.de

ABSTRACT

Cells adjust their proteome for growth and survival in accordance to their environments. For exponentially growing bacterial cells, simple empirical relations have been established between the growth rate and the allocation of proteome to catabolism and biosynthesis[1,2]. These relations, arising from global constraints on protein synthesis and flux balance, have led to simple models of proteome allocation that quantitatively predict the growth and proteome composition of *E. coli* under a variety of nutrient conditions, drug treatments, and genetic perturbations[1]. The success of the proteome allocation theory is surprising, given that in its drive for simplicity, it completely neglected the metabolites which drive individual reactions in growing cells. Here, we identify the existence of an underlying global relationship between the concentrations of metabolites and enzymes in living cells, emerging from a constraint on the concentration of cellular dry mass that limits the abundance of metabolite pools. Under this constraint, maximal reaction fluxes are predicted to be achieved when the mass abundance of a metabolite equals the combined mass of the unbound enzymes consuming it. The predicted optimal scaling of enzyme saturation accounts quantitatively for the large "proteome offsets" that are observed experimentally and assumed in the existing proteome allocation models. The predictions capture the general patterns of enzyme[3] and metabolite[4] level dependencies on the growth rate found in recent studies; they agree quantitatively with observations for several specific enzymatic systems for which detailed comparisons can be made, including the ribosome. The enzyme offsets collectively account for almost half of the entire proteome, comprising the bulk of the "proteome reserve" limiting the cellular growth capacity[1,5]. Our analysis indicates that this substantial proteome investment reflects a compromise between the accumulation of enzymes and metabolites, balanced to optimally drive biochemical fluxes.

Early studies of bacterial physiology revealed that the ribosome content of a cell followed a linear relation with the growth rate when growth is varied by changing nutrient quality[6–8]. More recent quantitative studies of *E. coli* found this to be the case also for many biosynthetic proteins[1,9]. One way to express such relations is in terms of the mass fraction ϕ_j of a protein j, as

$$\phi_j = \phi_{j,0} + \frac{\mu}{\kappa_j} \quad , \tag{1}$$

where μ is the growth rate and $\phi_{j,0}, \kappa_j$ are protein-specific constants (solid red line, Fig. 1a). In contrast to the cellular resources allocated to biosynthesis, the proteome fraction allocated to many other proteins – such as those involved in cellular structures and nutrient transport – is expected to remain constant or even diminish with increasing growth rate[1,9]. Here, we focus exclusively on proteins whose abundance increases with growth rate (i.e., have $\kappa_j > 0$); these make up about 80% of protein mass for *E. coli* cells grown in glucose minimal medium, at growth rate 0.7/h (Fig. 1a)[3]. $\phi_{j,0}$ in equation (1), being a constant offset from a proportionality between protein abundance and growth rate, describes the abundance of proteins apparently maintained by the cell regardless of growth conditions. This offset is substantial: for a typical protein, $\phi_{j,0}$ is around 60% of its average protein fraction ϕ_j across different minimal media (Fig. 1b). The sum of offsets for all biosynthetic proteins amounts to 40% of the total proteome in Ref.[3] and 50% in earlier studies on other *E. coli* strains[1,5], presenting a significant load on the growth capacity of the cell. In an idealized case where all nutrients are provided and cells only need ribosomes to synthesize the cellular proteins, the maximal growth rate attainable would be $\mu_{max} = \kappa_R \phi_{R,max}$), where κ_R is the translational activity of the ribosomes and $\phi_{R,max} = 1 - \phi_0$ is the maximal fraction of the proteome allocable to the ribosomes given the total proteome offset ϕ_0[5]. Thus, a protein offset of $\phi_0 = 50\%$ would reduce the maximal ribosomal fraction and hence the maximal growth rate by half.

What physiological functions might the apparent offsets $\phi_{j,0}$ serve to justify such a high cost to steady-state growth? One often-mentioned function is their role as "proteome reserves", reducing the lag time for physiological adaptation when nutrient availability changes abruptly. Indeed, adaptation to nutrient upshift was suggested early on to rationalize the significant offset observed for the ribosomal proteins ($\phi_{R,0} \approx 5\%$)[10,11]. The beneficial effect of this reserve in fluctuating environments was experimentally demonstrated recently for ribosomal and catabolic proteins[12,13]. Similar proteome reserves may be advantageous for biosynthetic proteins during nutrient downshifts, e.g., from rich medium to minimal medium, where the demand for biosynthesis abruptly rises. However, the benefit of reserves for nutrient upshifts can in principle be realized by maintaining a finite level at slow growth only, without affecting

the abundances at fast growth; see the dashed blue line in Fig. 1a. There are also reasons to doubt the suggested role of the offsets of biosynthetic proteins as reserves for nutrient downshifts. One of the most costly biosynthetic proteins is the cobalamine-independent methionine synthase (MetE); It comprises 5.7% of the proteome during growth in glucose minimal medium[3,14], yet exhibits an almost perfect proportionality with the growth rate in minimal medium, with an offset close to zero (see below). MetE is hardly expressed in rich medium[3,14], and its accumulation is the dominant cause of growth delay for nutrient downshift from medium replete with amino acids to minimal medium. Thus, reduced transition times are likely not the primary reason these proteome offsets evolved.

We next examine the role of the offsets $\phi_{j,0}$ in equation (1) in light of basic enzymatic kinetics. For a large number of cellular reactions, the flux through a reaction can be taken as proportional to the growth rate[15], i.e., $v_j \propto \mu$. Total protein concentration is roughly constant across growth conditions[3,12], and hence the proteome fraction is approximately proportional to the enzyme concentration, $\phi_j \propto [E_j]$. For simplicity, let us consider irreversible Michaelis-Menten kinetics for reaction j. In general, the reaction flux can be written as

$$v_j = k_j([E_j] - [E_{j,free}]) \quad , \tag{2}$$

where k_j is the specific reaction rate (turnover number) and $[E_{i,free}]$ is the concentration of the unsaturated or "free" enzymes. For a reaction with a single substrate of concentration $[S]$, we have

$$[E_{j,free}] = \frac{[E_j]}{1 + [S]/K_{m,j}} \quad , \tag{3}$$

where $K_{m,j}$ is the enzyme-substrate dissociation (Michaelis) constant. Comparison of equations (1) and (2) yields $\kappa_j \propto k_j$ and $\phi_{j,0} \propto [E_{j,free}]$, suggesting the possibility that the offset $\phi_{j,0}$ might be viewed as an enzyme inefficiency caused by the unsaturated enzymes $[E_{j,free}]$. A proportional relation between ϕ_j and the growth rate μ is only obtained when the substrate concentration much exceeds $K_{m,j}$, such that $[E_{j,free}] \to 0$ and all enzymes are saturated. To examine the status of enzyme saturation more quantitatively, we note that typical $K_{m,j}$ values for cellular reactions are in the range of $10\mu M - 1mM$ (blue colored bars, Fig. 2a)[16]. Metabolomic measurements in glucose minimal medium found the concentrations of the most abundant metabolites to be of this order (red colored bars, Fig. 2a)[4], with most concentrations declining in poorer nutrient conditions (Fig. 2b). Thus, substrate availability is an important factor limiting reaction rates, as has also been observed in genetic perturbation experiments[17]. The worsening of this problem for poorer conditions (Fig. 2b) is consistent with the increasing dominance of the offset $\phi_{j,0}$ at slow growth rates according to equation (1).

Why can't the cell simply increase the metabolite concentrations to levels much exceeding $K_{m,j}$? Aside from specific biochemical reasons such as cross-reactivity and toxicity, a physiological reason can be appreciated from a simple quantitative estimate: If metabolites are kept at concentrations of 1 mM, then the set of 1000 common metabolite species would collectively amount to 10^9 molecules/cell. This would be comparable to the total number of amino acids contained in cellular proteins, which comprise over 50% of the cellular dry weight[18].

Intuitively, keeping a significant portion of the biomass in the form of metabolites instead of enzymes may be another form of inefficiency, even if the extra metabolites can keep the enzymes all saturated and work at maximal efficiency. To explore the consequence of partitioning cellular resource (biomass in this case) between the enzymes and the metabolites, we note that the cell density ρ is approximately invariant across different growth conditions, and thus the total cellular dry mass M is proportional to the cellular water content[19], V. This leads to a global constraint relating the concentrations of the enzymes $[E_j]$ and the metabolic substrates $[S_i]$:

$$\sum_j m_{E_j} \cdot [E_j] + \sum_i m_{S_i} \cdot [S_i] = \rho \quad , \tag{4}$$

where m_{E_j} and m_{S_i} are the molecular weights of enzyme $[E_j]$ and substrate $[S_i]$, respectively, and ρ is the combined cytosolic mass density of proteins and metabolites. To appreciate the effect of the biomass constraint (4) on the enzyme-metabolite partitioning, let us examine a simplified case of a single enzyme E that catalyzes the conversion of a substrate S following irreversible Michaelis-Menten kinetics. As derived mathematically in SI text 1, the maximal reaction rate according to equations (2, 3) at a given combined mass concentration of the molecules involved, $m_{E_j} \cdot [E_j] + m_{S_i} \cdot [S_i] = \rho$, is achieved when

$$m_S[S]^* = m_E[E_{free}]^* = m_E \frac{[E]^*}{1 + [S]^*/K_m} \quad . \tag{5}$$

Equation (5) describes a relation between the metabolite and enzyme concentrations at the optimal flux condition, with the only parameters being the enzyme's Michaelis constant K_m and the ratio of molecular masses between the enzyme and metabolite. In Fig. 3, we tested the predicted relation between enzyme and metabolite concentrations using the absolute abundance data generated in Refs.[3,4] for *E. coli* growing in different nutrient conditions. The prediction is seen to quantitatively capture the available data (enzyme/metabolite pairs with known K_m) without any adjustable parameters, suggesting that the endogeneous system may indeed be operating at the optimal flux condition across the range of nutrient conditions tested.

Of course, cellular metabolism involves many reactions that require more than one substrate, and conversely, many metabolites are consumed by more than one enzyme. To obtain a generalization of equation (5), we need to account for all metabolite and enzyme concentrations simultaneously. Towards this end, we make the simplifying assumption that cellular reactions are effectively irreversible. The generalized equation relates the mass devoted to a substrate S in the reference condition with the sum of free enzyme masses over all enzymes j consuming this substrate (SI Text 1.1),

$$\sum_j m_{E_j} \cdot [E_{j,free}] = m_S[S] \quad . \tag{6}$$

Inserting the substrate concentration given in Eq.(6) into the Michaelis-Menten equation, we find that optimal protein concentration has a linear-plus-square root relationship to the reaction flux,

$$[E]^* = \frac{v^*}{k_{cat}} + \sqrt{K_m \frac{m_S}{m_E} \frac{v^*}{k_{cat}}} \quad . \tag{7}$$

While the linear term gives a lower bound on the total enzyme concentration achievable at full saturation, the square root term describes the effect of incomplete saturation and is responsible for the apparent offset from the linear relation. To illustrate the predictive power of the optimality relation (7) for enzyme offsets, we consider three exemplary cases in *E. coli*. First, the enzyme agmatinase (SpeB) carries out the second step in putrescine biosynthesis; its dissociation constant for agmantine, $K_{m,SpeB} = 1150\mu M$,[20,21] is the largest known for any substrate consumed by a single enzyme. We may thus expect SpeB to be especially affected by incomplete saturation. Accordingly, a linear fit of observed SpeB abundance vs. growth rate[3] (Fig.4a) shows an offset of $[E_{SpeB}] = 0.62\mu$M (95% Confidence Interval $[0.23, 1.00]$). Expanding Equation (7) linearly around growth on glucose as a "reference condition" and assuming that SpeB is half saturated with agmantine in this condition (SI text 2), we predict an offset $[E_{SpeB,0}] = 0.95\ \mu$M ($\phi_{SpeB,0} = 0.60‰$), within the range derived from the proteome data.

Second, we mentioned that MetE has an enzyme offset indistinguishable from zero (Fig. 4b), despite being the most abundant metabolic enzyme in *E. coli* at fast growth in minimal media[3]. MetE catalyzes the last step in methionine biosynthesis. In steady state, the flux through MetE must be given by the demand of methionine in protein synthesis, i.e., the concentration of methionine stored in cellular proteins times the growth rate μ. As we thus know the exact relationship between flux and growth rate, and as kinetic parameters for MetE are available in the literature[22], Equation (7) makes quantitative predictions for the optimal MetE concentration as a function of growth rate. These predictions (red line in Fig. 4b) agree accurately with experimental observations.

Finally, the catalyst responsible for the highest proteome fraction in fast-growing *E. coli* cells is the ribosome. The ribosome can be viewed as an enzyme that converts its substrate, a specific ternary complex, into elongating peptide chains. The rate of this process can be approximated by Michaelis-Menten kinetics[10,23] and must equal the total protein dilution rate, i.e., the total protein mass concentration[3] ($\approx$ 127.4 g/L) times the growth rate μ. Thus, as for MetE, Equation (7) quantitatively predicts optimal ribosome abundance as a function of the growth rate. For intermediate to high growth rates (μ>0.3), the predictions agree accurately with experimental observations (red line in Fig. 5a). At lower growth rates (μ<0.3), observed total ribosome concentrations exceed those predicted, consistent with the existence of a reserve of de-activated ribosomes in this regime[12].

When viewed as a relationship between flux and substrate concentration, optimal resource allocation predicts that optimal substrate concentrations $[S^*]$ are proportional to the square root of the reaction fluxes, $v_i^{*1/2}$ (Eq.(S13)) . Under the expected scaling $v^* \propto \mu$ for biosynthetic proteins, we thus expect $[S^*] \propto \mu^{1/2}$, a prediction consistent with the observed scaling of metabolite concentrations[4] with growth rate (Fig. 7). We can further test this relationship quantitatively for the substrate of ribosomal elongation, the ternary complexes comprising a charged tRNA and elongation factor Tu. Comparison to experimentally determined Tu abundances[3] quantitatively confirms the predictions (red line in Fig. 5b).

In sum, the observed scaling of the abundance of enzymes and the ribosome with growth rate is highly consistent with an optimal resource distribution between catalysts and their substrates, resulting in lower saturation levels and thus increasing enzyme inefficiencies at progressively lower reaction fluxes v^*. If the concentrations of ribosome and enzymes (representing the R- and A-sectors, respectively, of coarse-grained allocation theories[1,5]) were indeed always optimal in this sense, we would expect them to drop sharply to zero at very low growth rates, deviating from the near-linear behavior generally observed. Available data provides little evidence of this expected curvature. Moreover, the optimal growth-rate dependence Eq.(7) is different for each individual reaction, and fully optimal regulation would thus require very detailed – and accordingly costly – regulatory systems. Thus, it is conceivable that natural selection only favored an approximate regulatory implementation of the predicted growth rate dependence for all but the most costly catalysts, such as the ribosome and MetE: ribosome and enzyme concentrations may scale in proportion to μ, mediated through a small number of regulatory molecules[4,24] such as ppGpp[13], while an additional level of constitutive expression (a constant offset $\phi_{j,0}$) ensures that fluxes do not become too low across commonly encountered non-rich growth media. Such an approximate implementation would also be consistent with the scatter around the optimal relationships observed when considering enzymes as

a whole (Fig. 3). Regardless of the regulatory details, we can conclude that the overexpression of enzymes over a simple proportionality to growth rate, as represented by protein offsets in the existing growth models[1,25], is rooted in the fact that internal metabolite concentrations cannot be raised to arbitrarily high levels.

Methods

E. coli data

We obtained protein concentrations of Escherichia coli strain BW25113 for 18 exponential growth conditions on minimal media[3]. For 7 of these conditions, we additionally obtained metabolite concentrations[4] for the same strain.

Metabolite concentrations in μmol/gCDW were converted to cytosolic molar concentrations based on the same conversion factor between cytosol volume and cell dry weight (2.3 ml/gCDW) used by the original authors[4]. Individual protein mass concentrations, total protein mass concentration, amino acid composition of proteins, and growth rates at exponential growth conditions were obtained from Schmidt et al. based on mass/cell and cell volumes reported by the authors at different growth rates[3]. Protein concentrations were corrected by a factor of 0.67 in cell volume as suggested by Schmidt et al. based on their recently corrected measurements of cell volume (Supplementary note 3 in Ref.[3]). The molar concentration of enzymes was determined as the minimal molar concentration among its subunits, weighted by stoichiometry; to convert this to enzyme mass concentration, we multiplied by the stoichiometry-weighted sum of molecular subunit masses. Protein subunit stoichiometries for each enzyme were obtained from Ref.[26], and assumed to equal 1 for enzymes not listed there.

Ribosome concentration was determined by calculating the geometric mean over the molar concentrations of all ribosomal proteins measured in Ref.[3]. The turnover number for translation was set to the maximal elongation rate inferred from experimental data in Ref.[18], $k_{cat} = 22$ AA/s. The ribosome's dissociation constant for its substrates, the ternary complexes, was set to an estimation based on the diffusion limit for ternary complexes[10], K_m =3 μM. As in Ref.[10], we multiplied K_m by the number of ternary complexes (n_t=40); this is equivalent to dividing the ternary complex concentration $[S_{tc}]$ in equation (3) by n_t, as the ribosome only sees one of the n_t ternary complexes at a time.

Molecular weights of the ribosome and ternary complexes were calculated from their sequences. The stoichiometry of ribosomal proteins and RNAs in the ribosome was obtained from the EcoCyc

database[27]; the stoichiometry of all components is 1 except for RplL, for which it is 4. We converted the units of the kinetic parameters from molar concentrations to mass concentrations as follows: k_{cat} =22AA/s×(132.60Da/AA)/(2,306,967Da)×3,600s/h=4.55/h (where 132.60 ± 0.09 Da is the mean amino acid molecular weight in proteins for all conditions measured in[3] and 2,306,967 Da is the ribosome molecular weight); and $K_m = 120\mu$M ×69,167 g/mol = 8.30 g/L (where 69,167 ± 1351 g/mol is the mean±SD of ternary complex molecular weights). The ribosomal fraction of total protein was calculated using the mass fraction of proteins in the ribosome[3] (=0.58) and total protein mass concentration[3] P=127.4g/L . The predicted protein fraction of elongation factor Tu was calculated using the mass fraction of Tu in the ternary complex[3] (= 43238/69167 = 0.63) and total protein concentration P.

For Figs. 2a and 3b, we used the growth in a glucose minimal medium as the reference state to define the experimentally measured "typical" substrate concentration[4] $[S_{ref}]$ and "typical" growth rate[3] μ_{ref} .

For Fig. 3, we collected a non-redundant set of Michaelis constants (K_m) of wild-type enzymes from EcoCyc[27], BRENDA[28], and UniProt[29]. All experimental values are from *E. coli*, with the exception of four metabolite-enzyme pairs where only data from other organisms are available: D-ribulose 5-phosphate–ribose-5-phosphate isomerase A (Ru5P–rpiA), 1,3-bisphospho-D-glycerate–phosphoglycerate kinase (13DGP–pgk), ADP–phosphoglycerate kinase (ADP-pgk), and glycerone phosphate–fructose bisphosphate aldolase (DHAP-fbaA); we did not consider K_m values of the extremophile Sulfolobus solfataricus, as these were obtained from measurements at 70°C. If more than one K_m was listed across the databases, we first checked if these values were mostly within the same order of magnitude (i.e., if the geometric standard deviation was ≤ 10); in this case, we used the geometric mean of all available values. Otherwise, we considered the available data for Km to be too unreliable to be included. For Figures 2a and 3b, we obtained K_m values from the dataset in Ref.[16], filtered for the organism *E. coli* and restricted to values for reaction substrates rather than products. Metabolite molecular weights were obtained from EcoCyc[27]. For SpeB's dissociation constant for agmatinase, we used the mean value from Carvajal et al.[20] (K_m= 1.1 mM) and Satishchandran et al.[21] (K_m = 1.2 mM).

Identification of dominant enzymes in *E. coli*

If the unsaturated mass concentration $m_E[E_{j,free}]^*$ of enzyme j accounts for more than half of the total protein mass utilizing a given substrate i, equation (5) approximately describes the relationship between the concentrations of the substrate and this enzyme also in the general case (SI text equation (S25)). In this case, we call j the "dominant" enzyme for i. For an automated identification of dominant enzymes,

we used the sybil and sybilSBML[30] packages in R, with the EcoCyc metabolic model for *E. coli* exported as an SBML file using Pathway Tools 19.5[31]. For each metabolite measured in Ref.[4], we first identified all reactions using it as a substrate according to the metabolic model. The gene-reaction associations given in the EcoCyc model through b-numbers were used to map the reactions to the proteins measured in Ref.[3]. For each substrate assayed in Ref.[4], we determined a dominance score (hereafter referred to simply as "dominance") for each enzyme consuming it and assayed in Ref.[3]. The dominance of an enzyme was defined as the fraction it contributes to the total mass concentration of all assayed enzymes using the substrate. An enzyme was considered "dominant" for the substrate if its dominance was >0.5, i.e., its molecules constituted more than half of the total protein mass consuming the substrate. We only attempted to assess dominance if more than half of the enzymes consuming a given substrate were assayed in Ref.[3]. We excluded membrane-bound and periplasmic enzymes based on Gene Ontology annotations[32] (GO categories 0016020 (membrane), 0005886 (plasma membrane), 0005887 (integral component of plasma membrane), 0042597 (periplasmic space), 0009279 (cell outer membrane), 0019867 (outer membrane)), as in these cases the estimated enzyme concentrations will not correspond to cytosolic concentrations. If the reaction catalysed by the dominant enzyme was reversible according to the EcoCyc model, this substrate-enzyme pair was only considered further if the flux through the reaction was measured in the corresponding direction in Ref.[4]. Cyclic AMP (cAMP) was not included in the analysis, as the major role of cAMP is not metabolic: cAMP regulates transcription through varying concentrations of cAMP-CPR; accordingly, the only enzyme using it as a substrate (cAMP phosphodiesterase) is unlikely to have a major impact on cAMP concentrations in steady-state growth.

Figures

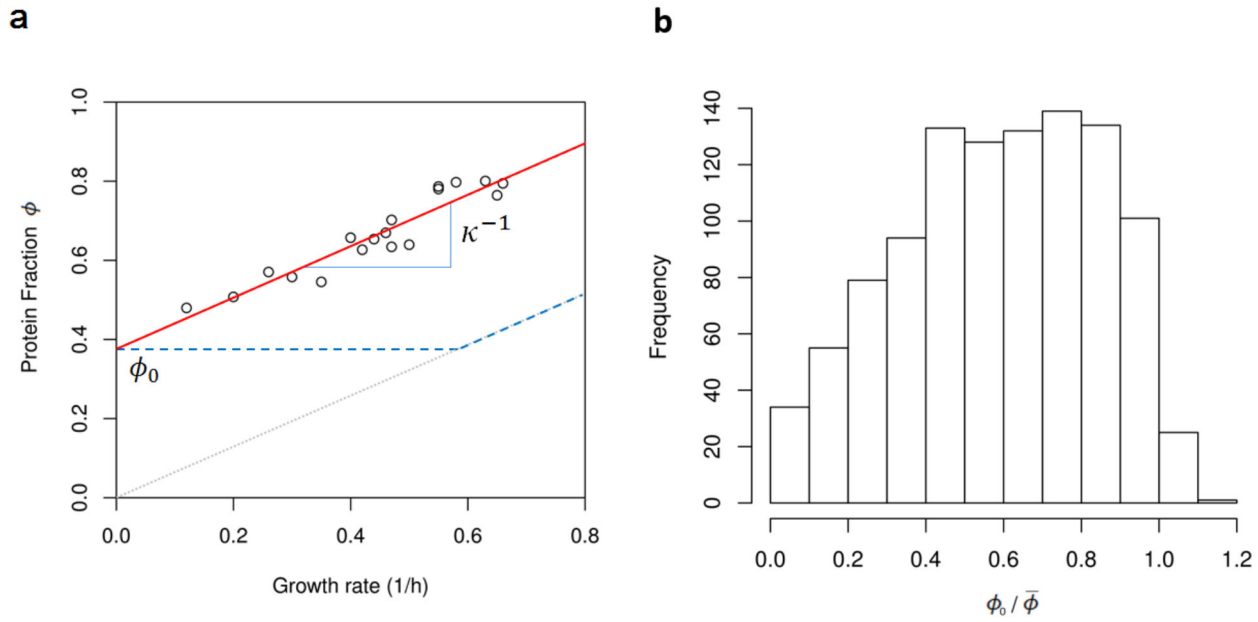

Figure 1. Protein offsets $\phi_{j,0}$ make up a substantial fraction of the proteome at low growth rates. **(a)** Comparison of the proportionality between $\phi_{j,0}$ and μ expected from the assumption that biosynthetic fluxes v^* scale in proportion to growth rate μ (dotted grey line), a protein reserve only at low growth rates (dashed blue line),and the observed constant offset across growth rates (solid red line). The points give the combined protein fraction of all proteins whose abundance increases with growth rate[3], with a combined offset $\phi_0 = 0.38$. **(b)** The distribution of the ratio between the offset $\phi_{j,0}$ observed for an individual protein and its average proteome fraction $\phi_{j,0}$ across exponential growth conditions on minimal media[3] shows a median of 0.6.

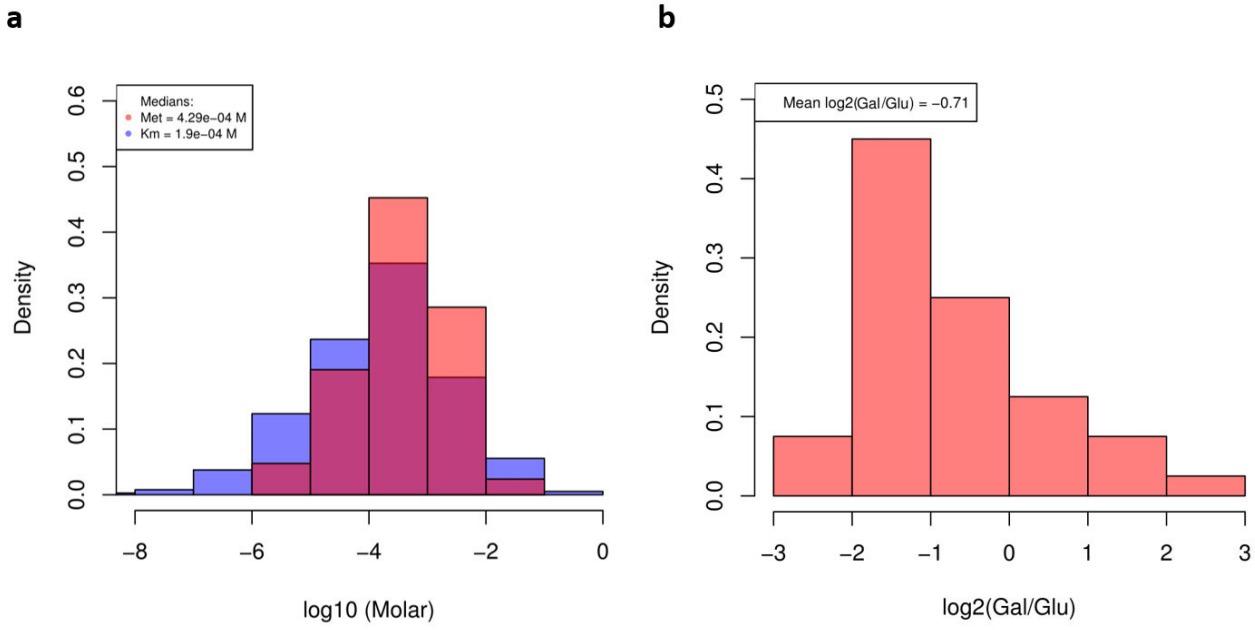

Figure 2. Substrate concentrations $[S_i]$ in a glucose reference condition are typically of the same order of magnitude as dissociation (Michaelis) constants $K_{m,i}$ and decrease in poorer nutrient conditions. **(a)** Log-scale histograms of observed metabolite concentrations $[S_i]$ (red)[4] and mean $K_{m,i}$ (blue)[16] per substrate. See Fig. 6 for the distribution of ratios $[S]/K_{m,i}$. **(b)** Log-scale histogram for the ratio of substrate concentrations between *E. coli* grown on glucose ($\mu = 0.65\ h^{-1}$) and on galactose ($\mu = 0.18\ h^{-1}$)[4], showing that metabolite concentrations in the poorer growth condition are typically about half of those in glucose.

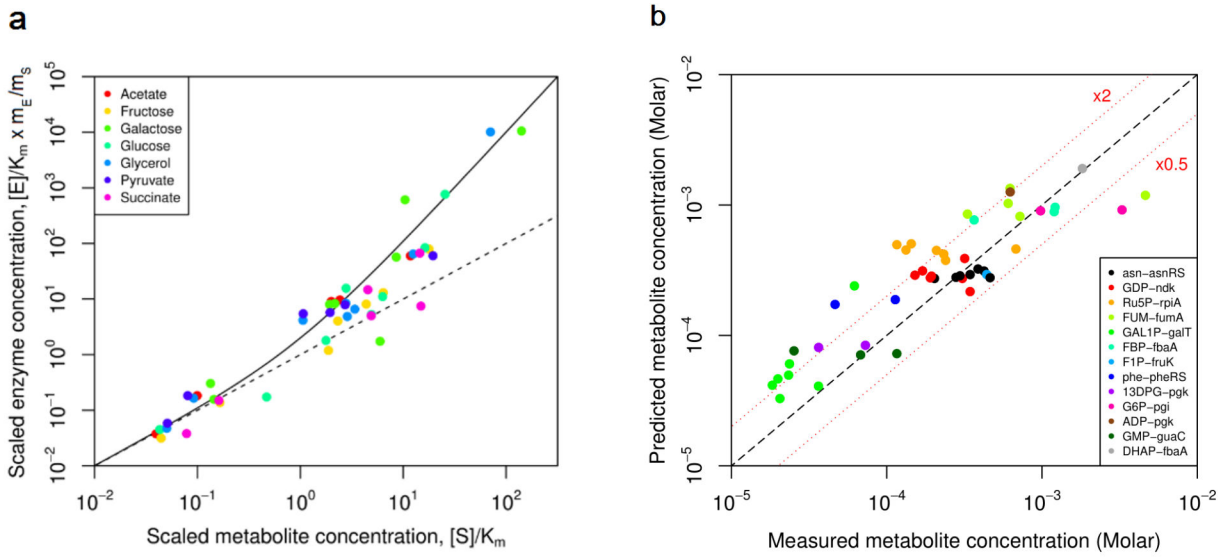

Figure 3. Experimentally observed enzyme and metabolite concentrations recover the predicted optimal scaling. **(a)** If a single enzyme j dominates the total enzyme mass consuming substrate S (SI text 1.1), we can use equation (5), rewritten for scaled enzyme concentration, $e_j^* = [E_j]^*/K_m \times m_{E_j}/m_{S_i}$ (y-axis), and scaled substrate concentration, $s^* = [S]^*/K_m$ (x-axis): this results in the dimensionless prediction $e_j^* = s^*(1+s^*)$ (solid line). The (geometric) mean fold-error of enzyme concentration predictions from observed substrate concentrations is $GMFE = 1.66$, Pearson correlation is $R^2 = 0.81$. The dashed line corresponds to identical mass concentrations of total enzyme and substrate ($e_j^* = s^*$). **(b)** Same data as in (a), but plotted as predicted versus measured metabolite concentrations, with colors indicating metabolite-enzyme pairs: asn-asnRS: L-asparagine - asparagine-tRNA ligase; GDP-ndk: guanosine-5'-diphosphate - nucleoside diphosphate kinase, Ru5P-rpiA: D-ribulose 5-phosphate - ribose-5-phosphate isomerase A, FUM-fumA: Fumarate - fumarase A, GAL1P-galT: alpha-D-galactose 1-phosphate - galactose-1-phosphate uridylyltransferase, FBP-fbaA: fructose 1,6-biphosphate - fructose bisphosphate aldolase class II, F1P-fruK: beta-D-fructofuranose 1-phosphate - 1-phosphofructokinase, phe-pheRS: L-phenylalanine - phenylalanine-tRNA ligase, 13DPG-pgk: 1,3-bisphospho-D-glycerate - phosphoglycerate kinase, G6P-pgi: beta-D-glucose 6-phosphate - phosphoglucose isomerase, ADP-pgk: adenosine-diphosphate - phosphoglycerate kinase, GMP-guaC: guanosine-5'-monophosphate - GMP reductase, DHAP-fbaA: glycerone phosphate - fructose bisphosphate aldolase class II.

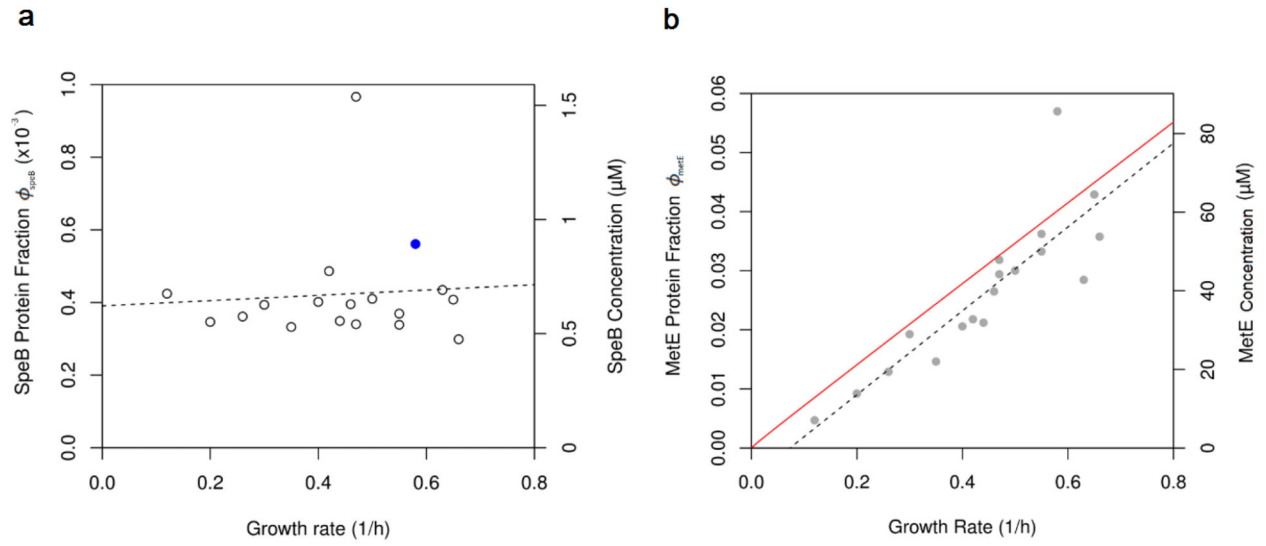

Figure 4. Observed enzyme offsets[3] for two exemplary enzymes agree with predictions from optimal resource allocation. **(a)** SpeB shows a large enzyme offset compared to its abundance in the glucose reference condition (blue dot) , with $\phi_{SpeB,0} = 0.00039$ (95% Confidence Interval $[0.00015, 0.00063]$) from linear regression (dashed line). The concentrations of SpeB's substrate, agmantine, are unknown; assuming that SpeB is half saturated in the reference condition ($[S_{ref}]^* = K_{m,SpeB}$) results in an offset prediction of $[E_{0,SpeB}] = 0.95\mu M$ (or $\phi_{SpeB,0} = 0.00060$), within the range derived from the proteome data. **(b)** metE concentration at different growth rates. A linear fit (dashed black line) results in the empirical offset $\phi_{MetE,0} = -0.0049$ (95% confidence interval: $[-0.015, +0.0049]$). The red line is the theoretical prediction from equation (7), using experimentally determined kinetic parameters (SI Text 3). The predictions are highly accurate, $GMFE = 1.29$ and $R^2 = 0.74$.

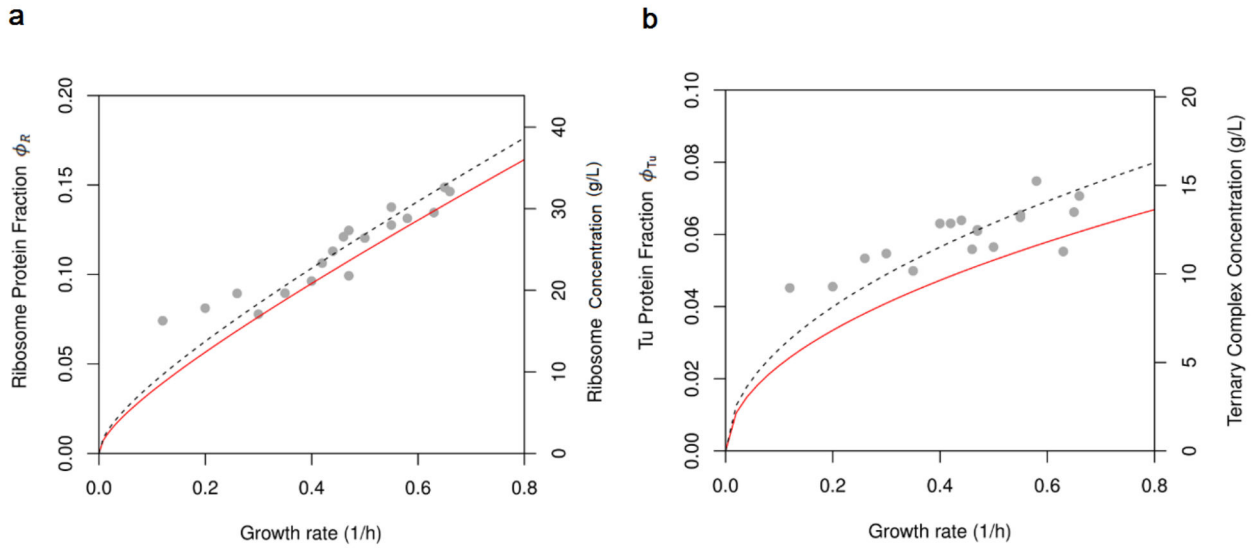

Figure 5. Optimal ribosome and Tu concentrations agree quantitatively with experimental data. **(a)** Protein fraction ϕ_R allocated to ribosomes at different growth rates[3]. A linear fit results in the empirical offset $\phi_{R,0} = 0.047$ (95% confidence interval: $[0.034, 0.060]$). The theoretical prediction from equation (7) (solid red line; SI text 4) is highly accurate, $GMFE = 1.13$ and $R^2 = 0.88$. The dashed line shows predictions based on fitted K_m (see panel b). **(b)** Ternary complex concentrations predicted from equations (S14,S42) for the ribosome, compared to experimentally observed proteome fractions of elongation factor Tu[3]. The solid red line ($GMFE = 1.24$, $R^2 = 0.64$) is calculated based on the diffusion limit for ternary complexes, which provides a lower bound for K_m; the dashed line (GMFE = 1.10, R^2 = 0.64) shows predictions based on a K_m value that is 43% larger and results from a fit to the data in this panel.

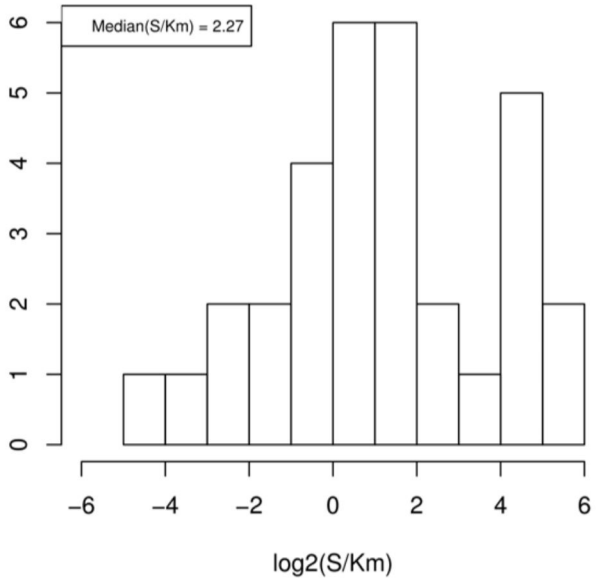

Figure 6. The ratio between absolute metabolite concentration[4] in the glucose reference condition and the geometric mean across all known *E. coli* $K_{m,i}$ values for this substrate is distributed around 1 (median: 2.27; N=32, as for 10 of the 42 metabolites assayed in Ref.[4] no $K_{m,i}$ values are available).

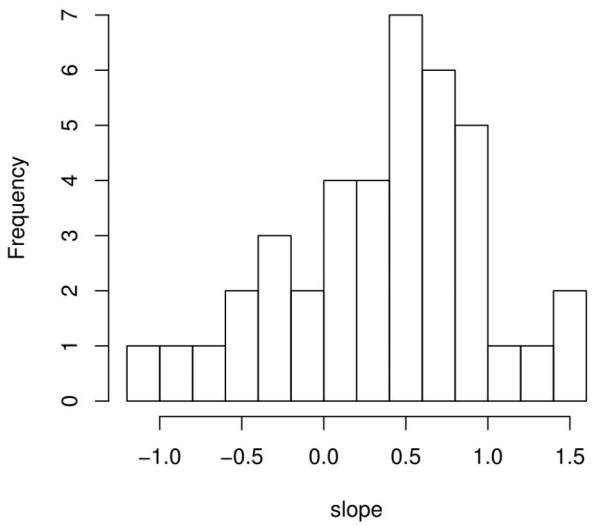

Figure 7. Substrate concentrations increase approximately with the square root of the growth rate, as predicted from optimal resource allocation. The histogram shows the distribution of slopes b in linear fits of $\log([S])$ vs. $\log(\mu)$ (power law fits corresponding to $[S] = a\mu^b$), demonstrating a peak around the expected value of 0.5; data from Ref.[4].

References

1. 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).

2. Mori, M., Hwa, T., Martin, O. C., De Martino, A. & Marinari, E. Constrained allocation flux balance analysis. *PLOS Comput. Biol.* **12**, 1–24, DOI: 10.1371/journal.pcbi.1004913 (2016).

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

4. Gerosa, L. *et al.* Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data. *Cell Syst.* **1**, 270 – 282, DOI: https://doi.org/10.1016/j.cels.2015.09.008 (2015).

5. 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).

6. Schaechter, M., MaalØe, O. & Kjeldgaard, N. O. Dependency on medium and temperature of cell size and chemical composition during balanced growth of salmonella typhimurium. *Microbiology* **19**, 592–606 (1958).

7. Neidhardt, F. C. & Curtiss, R. *Escherichia coli and Salmonella: Cellular and Molecular Biology (2 Volumes)* (ASM Press, 1996).

8. Maaløe, O. *Regulation of the Protein-Synthesizing Machinery—Ribosomes, tRNA, Factors, and So On*, 487–542 (Springer US, Boston, MA, 1979).

9. You, C. *et al.* Coordination of bacterial proteome with metabolism by cyclic amp signalling. *Nature* **500**, 301 EP – (2013). Article.

10. 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).

11. 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.

12. Mori, M., Schink, S., Erickson, D. W., Gerland, U. & Hwa, T. Quantifying the benefit of a proteome reserve in fluctuating environments. *Nat. Commun.* **8**, 1225, DOI: 10.1038/s41467-017-01242-8 (2017).

13. Erickson, D. W. *et al.* A global resource allocation strategy governs growth transition kinetics of escherichia coli. *Nature* **551**, 119 EP – (2017).

14. Li, G.-W., Burkhardt, D., Gross, C. & Weissman, J. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635, DOI: 10.1016/j.cell.2014.02.033 (2014).

15. 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).

16. 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).

17. Fendt, S.-M. *et al.* Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity. *Mol. Syst. Biol.* **6**, DOI: 10.1038/msb.2010.11 (2010).

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

19. Woldringh, C. & Nanninga, N. *Molecular Cytology of Escherichia coli*, chap. Structure of the Nucleoid and Cytoplasm in the Intact Cell (Academic Press, London, 1985).

20. Carvajal, N. *et al.* Kinetic studies and site-directed mutagenesis of escherichia coli agmatinase. a role for glu274 in binding and correct positioning of the substrate guanidinium group. *Arch. Biochem. Biophys.* **430**, 185 – 190, DOI: https://doi.org/10.1016/j.abb.2004.07.005 (2004). Highlight section: Methodologies for the measurement of the macular pigment.

21. Satishchandran, C. & Boyle, S. M. Purification and properties of agmatine ureohydrolyase, a putrescine biosynthetic enzyme in escherichia coli. *J. Bacteriol.* **165**, 843–848, DOI: 10.1128/jb.165.3.843-848.1986 (1986). https://jb.asm.org/content/165/3/843.full.pdf.

22. Whitfield, C. D., Steers, E. J. & Weissbach, H. Purification and properties of 5-methyltetrahydropteroyltriglutamate-homocysteine transmethylase. *J. Biol. Chem.* **245**, 390–401 (1970). http://www.jbc.org/content/245/2/390.full.pdf+html.

23. Wong, F., Dutta, A., Chowdhury, D. & Gunawardena, J. Structural conditions on complex networks for the michaelis–menten input–output response. *Proc. Natl. Acad. Sci.* **115**, 9738–9743, DOI: 10.1073/pnas.1808053115 (2018). https://www.pnas.org/content/115/39/9738.full.pdf.

24. 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).

25. Basan, M. Resource allocation and metabolism: the search for governing principles. *Curr. opinion microbiology* **45**, 77–83 (2018).

26. Liu, J. K. *et al.* Reconstruction and modeling protein translocation and compartmentalization in escherichia coli at the genome-scale. *BMC Syst. Biol.* **8**, 110, DOI: 10.1186/s12918-014-0110-6 (2014).

27. 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).

28. Chang, A. *et al.* BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic Acids Res.* **43**, D439–D446, DOI: 10.1093/nar/gku1068 (2014). http://oup.prod.sis.lan/nar/article-pdf/43/D1/D439/17436200/gku1068.pdf.

29. Consortium, T. U. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212, DOI: 10.1093/nar/gku989 (2014). http://oup.prod.sis.lan/nar/article-pdf/43/D1/D204/17438515/gku989.pdf.

30. Gelius-Dietrich, G., Desouki, A. A., Fritzemeier, C. J. & Lercher, M. J. sybil - efficient constraint-based modelling in r. *BMC Syst. Biol.* **7**, 125, DOI: 10.1186/1752-0509-7-125 (2013).

31. Karp, P. D. *et al.* Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. *Briefings Bioinforma.* **17**, 877–890, DOI: 10.1093/bib/bbv079 (2015). http://oup.prod.sis.lan/bib/article-pdf/17/5/877/17485679/bbv079.pdf.

32. Gene Ontology Consortium *et al.* Gene ontology consortium: Going forward. *Nucleic Acids Res.* **43**, D1049–D1056, DOI: 10.1093/nar/gku1179 (2015).

SI text

1. Detailed derivation of equations 1-7

Let us first consider the simple case of a substrate used by a single irreversible reaction. The rate v of an irreversible enzymatic reaction that converts a single substrate into a product according to a general kinetic function $k \equiv k([S], K_m, k_{cat})$ of the substrate concentration and respective kinetic constants,

$$v = [E]k\,, \tag{S1}$$

with enzyme molar concentration [*E*] and substrate molar concentration [*S*]. For irreversible Michaelis-Menten kinetics,

$$k = k_{cat}\frac{[S]}{[S]+K_m}\,, \tag{S2}$$

where k_{cat} is the turnover number and K_m is the dissociation (Michaelis) constant. The enzyme and substrate concentrations of this reaction together account for a total mass concentration *c,* measured per volume of the corresponding cellular compartment, e.g., the cytosol; *c* is a linear function of the molar concentrations [*E*] and [*S*], each multiplied with the respective molecular weights (m_E and m_S, respectively):

$$c = m_E[E] + m_S[S]\,. \tag{S3}$$

We assume that the cell is in a steady state that demands a fixed reaction rate $v > 0$. Rearranging Eq. (S1), we can express $[E]$ as a function of v and the kinetic function $k([S], K_m, k_{cat})$,

$$[E] = \frac{v}{k}\,; \tag{S4}$$

we assume $v > 0$ and thus $[S] > 0$ and $k > 0$ throughout our derivations. Substituting Eq.

(S4) into Eq. (S3), we can express the reaction's total mass concentration, *c*, as a function of the substrate concentration $[S]$ and the constants v, K_m, k_{cat}:

$$c = m_E \frac{v}{k} + m_S[S] \,. \tag{S5}$$

If *c* is minimal, a necessary condition is that the derivative of Eq. (S4) with respect to $[S]$ must be zero (at constant v):

$$\frac{dc}{d[S]} = 0 \,. \tag{S6}$$

We thus have

$$0 = \frac{dc}{d[S]} = -m_E \frac{v^*}{(k^*)^2} \frac{dk([S^*])}{d[S]} + m_S \,. \tag{S7}$$

We can simplify the further derivation if we divide all terms in Eq. (S7) by m_s and consider the ratio $a \coloneqq m_E/m_S$:

$$a \frac{v^*}{(k^*)^2} \frac{dk([S^*])}{d[S]} = 1 \,. \tag{S8}$$

Substituting the flux v using Eq.(S1):

$$a \frac{[E^*]}{k^*} \frac{dk([S^*])}{d[S]} = 1 \,. \tag{S9}$$

For irreversible Michaelis-Menten kinetics (Eq.(S2)), Eq. (S8) and (S9) result , respectively, in:

$$v^* = k_{cat} \frac{[S^*]^2}{aK_m} \,, \tag{S10}$$

$$a[E^*] = [S^*]\left(1 + \frac{[S^*]}{K_m}\right) \,. \tag{S11}$$

We note that Eq.(S11) doesn't depend on k_{cat}. Combining Eq.(S11) with Eq.(3) of the main text results in the equality between the mass concentration of substrate and free enzyme,

$$m_S[S^*] = m_E[E^*_{free}] \,. \tag{S12}$$

Both equations (S10) and (S11) can further be solved for $[S]$ to give, respectively:

$$[S^*] = \sqrt{\frac{aK_m v^*}{k_{cat}}}, \tag{S13}$$

$$[S^*] = \frac{K_m}{2}\left(\sqrt{1 + \frac{4a[E^*]}{K_m}} - 1\right) \tag{S14}$$

Substituting Eq.(S13) in Eq.(S12) and Eq.(S14) in Eq.(S11), we have, respectively:

$$v^* = k_{cat}\left([E^*] - \frac{[S^*]([E^*]; a, K_m)}{a}\right), \tag{S15}$$

$$[E^*] = \frac{v^*}{k_{cat}} + \sqrt{\frac{K_m v^*}{a k_{cat}}}, \tag{S16}$$

where $[S^*]([E^*]; a, K_m)$ is given by Eq.(S14). In both equations, we note that the second term on the right hand side is a consequence of the incomplete enzyme saturation by the metabolite.

1.1 Optimality at the systems level

Enzymatic reactions in biological cells are not isolated: the same substrate is often consumed by multiple enzymes, and the same enzyme may utilize multiple substrates. We thus need to generalize the above derivation to the systems level, considering all metabolic reactions within one cellular compartment (*e.g.*, the cytosol) simultaneously. The relevant variables can be expressed as vectors and matrices: $\vec{v}$, the vector of reaction rates v_j; $\vec{E}$, the vector of enzyme concentrations $[E_j]$ for these reactions; $\vec{k}_{cat}$, the corresponding vector of turnover numbers k_{cat_j}; $\vec{S}$, the vector of metabolite concentrations $[S_i]$; and *K*, the matrix of Michaelis constants $K_{m_{ij}}$ for metabolite *i* as substrate in reaction *j* (with $K_{m_{ij}} = 0$ if metabolite *i* is not a substrate of reaction *j*).

A non-zero rate v_j of reaction *j* can then be described using any reaction kinetics as

$$v_j = [E_j]k_j \ , \tag{S17}$$

where the effective rate $k_j = k_j(\vec{S}, K)$ is a function of the metabolite concentrations $\vec{S}$ and a set K of kinetic parameters. We assume that the cell is in a given metabolic state, *i. e.*, all reactions have a fixed rate v_j ($\vec{v}$=const). Below, we are only concerned with active reactions ($v_j > 0$), and we thus drop metabolites and enzymes involved only in non-active reactions from further consideration (*i.e.*, we assume $S_i > 0$ and $E_j > 0$ for all *i* and *j* without loss of generality).

In this metabolic state, the metabolism of a given cellular compartment accounts for a total mass concentration c_{tot}; this can be calculated as the sum of all enzyme and metabolite molar concentrations, each term multiplied by the corresponding molecular weight:

$$c_{tot} = \sum_j m_{E_j}[E_j] + \sum_i m_{S_i}[S_i] \ . \tag{S18}$$

The derivation proceeds largely as above. We can rearrange Eq. (S17) to express each enzyme concentration $[E_j]$ as a function of v_j and the vector of effective rates (which itself is a function of metabolite concentrations $\vec{S}$), as

$$[E_j] = \frac{v_j}{k_j} \ . \tag{S19}$$

It follows that for any vector of reaction rates $\vec{v}$ and any vector of non-zero metabolite concentrations $\vec{S}$, there always exists a matching vector of enzyme concentrations $\vec{E}$. Substituting Eq. (S19) into Eq. (S18), we obtain

$$c_{tot} = \sum_j m_{E_j}\frac{v_j}{k_j} + \sum_i m_{S_i}[S_i] \ , \tag{S20}$$

which is now only a function of $\vec{S}$ and the constants $\vec{v}, K, \vec{m}_E, \vec{m}_M$.

If this metabolic state has the lowest possible mass concentration (*i.e.*, c_{tot} is minimal with respect to $\vec{S}$), then all partial derivatives must vanish:

$$0 = \frac{\partial c_{tot}}{\partial S_k} = -\sum_j m_{E_j} \frac{v_j^*}{\left(k_j^*\right)^2} \frac{\partial k_j([S_k^*])}{\partial [S_k]} + m_{S_k} \tag{S21}$$

for all metabolites *k*. Dividing all terms in Eq. (S21) by m_{S_k} and rearranging, we obtain

$$\sum_j \frac{a_{kj} v_j^*}{\left(k_j^*\right)^2} \frac{\partial k_j([S_k^*])}{\partial [S_k]} = 1 \ , \tag{S22}$$

where $a_{kj} \coloneqq m_{E_j}/m_{M_k}$ is the ratio of the molecular weights of enzyme *j* and its substrate *k.* Using Eq. (S17) to re-substitute the reaction rates v_j into Eq. (S22) leads to

$$\sum_j \frac{a_{kj}[E_j^*]}{k_j^*} \frac{\partial k_j([S_k^*])}{\partial [S_k]} = 1 \ . \tag{S23}$$

If all reactions *j* follow generalized irreversible Michaelis-Menten kinetics of the form convenience kinetics [33]

$$k_j = k_{cat_j} \prod_i \left(\frac{[S_i]}{[S_i] + K_{m_{ij}}} \right), \tag{S24}$$

where the kinetic parameters K consist of turnover numbers k_{cat_j} and Michaelis constants $K_{m_{ij}}$, then Eq.(S23) results in

$$\sum_j \frac{a_{kj}[E_j^*]}{[S_k^*]\left(1 + \frac{[S_k^*]}{K_{m_{kj}}}\right)} = 1 \ , \tag{S25}$$

which only depends on the concentration and Michaelis constants of a single substrate S_k, and is independent of turnover numbers k_{cat_j}. Thus, the contribution of each individual metabolite to the total cellular cost in a maximally efficient metabolic system can be considered in isolation. Also considering generalized irreversible Michaelis-Menten kinetics, Eq.(S22) results in

$$\sum_j \frac{a_{kj} v_j^* K_{m_{kj}} \varphi_{kj}}{k_{cat_j}} = [S_k^*]^2 \ , \tag{S26}$$

where

$$\varphi_{kj}^* \coloneqq \prod_{l \neq k} \left(\frac{K_{m_{lj}}}{[S_l^*]} + 1 \right) \ ,$$

is the contribution of the other metabolites $l \neq k$ used as substrates in reaction j. Combining Eq.(S26) with Eq.(3) directly generalizes Eq.(S11) as the Eq.(6) in the main text. This equation applies to a complete metabolic system of effectively irreversible reactions following generalized Michaelis-Menten kinetics: the optimally cost-efficient concentration of each metabolite [S_k] in a given metabolic state (*i.e.*, at given reaction rates $\vec{v}$) depends only on the concentrations of the enzymes consuming it, their affinities $K_{m_{kj}}$ for the metabolite, and the enzyme/metabolite molecular weight ratios a_{kj}, but is independent of turnover numbers and reaction rates.

If one of the summands in Eq. (6) is close to 1, it will dominate this expression and we approximately recover Eq. (S11). The dominant term will usually correspond to the enzyme with the highest $a_{kj}E_j$; this is what is shown in Fig. 3 of the main text.

2. Enzyme offsets

Let us again assume that all reactions j consuming substrate i follow irreversible Michaelis-Menten kinetics in the form [33] given by Eq.(S26); we further assume that all Michaelis constants for substrate i are approximately the same, $K_{m_{ij}} \approx K_{m_i}$. This assumption is consistent with experimental observations: approximating a given Michaelis constant $K_{m_{ij}}$ by the geometric mean of $K_{m_{ij}}$, values for all other reactions j' using the same substrate results in a mean fold-error of 4.68, *i.e.*, predictions tend to be well within the right order of

magnitude. With these approximating assumptions, Eq.(S25) and Eq.(S26) result, respectively, in

$$\sum_j a_{ij}[E_j^*] = [S_i^*]\left(1 + \frac{[S_i^*]}{K_{m_i}}\right), \tag{S27}$$

$$\sqrt{\sum_j \frac{a_{ij} v_j^* K_{m_i} \varphi_{ij}^*}{k_{cat_j}}} = [S_i^*] \ , \tag{S28}$$

where

$$\varphi_{ij}^* := \prod_{l \neq i}\left(\frac{K_{m_{lj}}}{[S_l^*]} + 1\right) \ ,$$

is the contribution of the other metabolites $l \neq i$ used as substrates in reaction j. Substituting $[S_i]$ given by Eq.(S28) into Eq.(S27), we obtain

$$\sum_j a_{ij} E_j = \sum_j \frac{a_{ij} v_j^* \varphi_{ij}^*}{k_{cat_j}} + \sqrt{\sum_j \frac{a_{ij} v_j^* K_{m_i} \varphi_{ij}^*}{k_{cat_j}}} \ . \tag{S29}$$

Let us now further assume that the cell is (approximately) in the same flux mode in all conditions considered, i.e., all fluxes v_j scale linearly with the growth rate μ:

$$v_j = \beta_j \mu \tag{S30}$$

with constant β_j . Let us also assume that the φ_{ij} are approximately constant across the conditions considered. With these approximations, Eq.(S30) multiplied by the molecular weight of the substrate i describes the summed mass concentration of all enzymes consuming the substrate,

$$E_i^T = \sum_j \frac{m_{E_j} v_j \varphi_{ij}^*}{k_{cat_j}} + \sqrt{\sum_j \frac{m_{E_j} m_{S_i} K_{m_i} v_j \varphi_{ij}^*}{k_{cat_j}}} =$$

$$= x_i \mu + \sqrt{x_i m_{S_i} K_{m_i} \mu}\,. \tag{S31}$$

where $x_i \coloneqq \sum_j \frac{m_{E_j} \beta_j \varphi_{ij}^*}{k_{cat_j}}$ is constant. Expanding E_i^T around a typical growth rate μ_{ref} to first order, we obtain

$$E_i^T \approx E_i^T(\mu_{ref}) + \frac{dE_i^T}{d\mu}(\mu - \mu_{ref})$$

$$= x_i \mu_{ref} + \sqrt{x_i m_{S_i} K_{m_i} \mu_{ref}} + \left(x_i + \frac{1}{2} \frac{x_i m_{S_i} K_{m_i}}{\sqrt{x_i m_{S_i} K_{m_i} \mu_{ref}}} \right)(\mu - \mu_{ref})$$

$$= \tfrac{1}{2} \sqrt{x_i m_{S_i} K_{m_i} \mu_{ref}} + \mu \left(x_i + \tfrac{1}{2} \sqrt{\frac{x_i m_{S_i} K_{m_i}}{\mu_{ref}}} \right). \tag{S32}$$

Setting $\mu = 0$, we find the apparent offset for the summed enzyme concentration consuming the metabolite considered,

$$E_{i,0}^T = \tfrac{1}{2} \sqrt{x_i m_{S_i} K_{m_i} \mu_{ref}} = \frac{m_{S_i} [S_{i,ref}]}{2} \quad , \tag{S33}$$

with $[S_{i,ref}]$ the metabolite concentration at the typical growth rate μ_{ref}; the last equality results from inserting Eq. (S28). Under the assumption that the enzyme is half saturated, $E_{i,0}^T = \frac{m_{S_i} K_{m,i}}{2}$.

3. MetE concentration and methionine production

MetE catalyzes the only reaction consuming 5-methyltetrahydropteroyl-tri-L-glutamate, and thus Eq. (S29) simplifies to

$$[E^*_{metE}] = \frac{v^* \varphi^*_h}{k_{cat}} + \sqrt{\frac{K_m v^* \varphi^*_h}{a k_{cat}}} \quad , \tag{S34}$$

where a is the ratio between molecular weight of enzyme and substrate, and

$$\varphi^*_h := \frac{K_{m_h}}{[S^*_h]} + 1 \quad ,$$

is the contribution of homocysteine, the other metabolite used as substrate in the reaction.

At balanced growth with no protein degradation, each amino acid i built into the proteome is produced at the same rate v_i as it is diluted by growth:

$$v_i = \mu n_i \tilde{P}, \tag{S35}$$

where μ is the growth rate, n_i is the fraction of amino acid i in the proteome, and $\tilde{P}$ is the total protein concentration, which is approximately constant across conditions [1,3]. The reaction rate catalysed by MetE, which produces methionine, should thus equal the rate at which methionine is diluted in proteins (plus negligible terms for the dilution of methionine itself and for offsetting protein degradation). Substituting Eq.(S35) into Eq.(S34) gives the following expression for the concentration of MetE as a function of growth rate:

$$[E^*_{metE}](\mu) = \frac{n_{met} \tilde{P} \varphi_h}{k_{cat}} \mu + \sqrt{\frac{n_{met} \tilde{P} K_m \varphi_h}{a k_{cat}} \mu} \quad . \tag{S36}$$

For **Fig. 4b**, we assume half saturation, $\varphi_h = 2$. $n_{met} = 0.0256 \pm 0.0003$ is practically constant over all conditions explored in Schmidt *et al.* (2016). According to equation (S33), the protein offset predicted for MetE is (with $c_{Met} = n_{met}\tilde{P}$):

$$[E^*_{metE,0}] = \sqrt{\frac{K_m}{2\,k_{cat}} m_{metE} \mu_{ref} c_{Met}}\,. \tag{S37}$$

For comparison, the optimal value at the reference condition is

$$[E^*_{metE,ref}] = \frac{2\,c_{Met}}{k_{cat}}\mu_{ref} + \sqrt{\frac{2\,c_{Met}K_m}{a k_{cat}}\mu_{ref}}\quad. \tag{S38}$$

4. Translation elongation as an enzymatic reaction

We can apply the same reasoning as before to estimate the optimal balance of enzyme and substrates for the elongation phase of translation. In a simple model of elongation [10], the mRNA-bound ("active") ribosome, *R*, acts as the enzyme, while the set of ternary complexes of charged tRNAs forms the substrate; in this simple biochemical model, the kinetic constants are identical for all codons and all amino acids [10]. The following derivations will be simplified by using mass units, indicated by the use of tilde "∼".

Eq.(S16), re-written for mass concentrations instead of for molar concentrations, then predicts the optimal concentration of ribosomes as

$$[\tilde{R}^*](\tilde{v}^*) = \frac{\tilde{v}^*}{\tilde{k}_{cat}} + \sqrt{\frac{\tilde{v}^*\tilde{K}_m}{\tilde{k}_{cat}}}\quad, \tag{S39}$$

where $[\tilde{R}^*]$ is the optimal mass concentration of the active ribosomes occupied with adding amino acids to the growing peptide, $\tilde{v}$ is the corresponding elongation mass flux per volume; $\tilde{k}_{cat}$ and $\widetilde{K}_m$ are turnover number and Michaelis constant of the elongation reaction in mass units [10] (see Methods). At balanced growth with no protein degradation, each amino acid built into the proteome is produced at the same rate as it is diluted by growth:

$$\tilde{v} = \mu\tilde{P} \ , \tag{S40}$$

where $\tilde{P}$ is the total protein mass concentration, which is approximately constant across conditions [1,3]. Substituting Eq.(2) into Eq.(1) gives

$$[\tilde{R}^*](\mu) = \frac{\tilde{P}}{\tilde{k}_{cat}}\mu + \sqrt{\frac{\tilde{P}\,\widetilde{K}_m}{\tilde{k}_{cat}}\mu} \tag{S41}$$

or, in terms of the protein mass fraction of ribosomal proteins (which make up $P_R := \frac{1340000}{2306967} = 0.58$ of total ribosome mass [3], $\phi_R := P_R[\tilde{R}^*]/\tilde{P}$):

$$\phi_R(\mu) = \frac{P_R}{\tilde{k}_{cat}}\mu + P_R\sqrt{\frac{\widetilde{K}_m}{\tilde{P}\,\tilde{k}_{cat}}\mu} \ . \tag{S42}$$

A linear expansion of $[\tilde{R}^*]\,(\mu)$ around a typical growth rate μ_{ref} gives

$$[\tilde{R}^*]\,(\mu) \approx [\tilde{R}^*]\big(\mu_{ref}\big) + [\tilde{R}^*]'(\mu_{ref})(\mu - \mu_{ref}) \ , \tag{S43}$$

which can be rearranged as

$$[\tilde{R}^*]\,(\mu) \approx \tilde{R}_0 + [\tilde{R}^*]'(\mu_{ref})\mu \ , \tag{S44}$$

where $\tilde{R}_0 \coloneqq [\tilde{R}^*](\mu_{ref}) - [\tilde{R}^*]'(\mu_{ref})\mu_{ref}$ is the offset of active ribosome concentration at zero growth, and $[\tilde{R}^*]'(\mu_{ref}) = \frac{\tilde{P}}{\tilde{k}_{cat}}\left(1 + \frac{1}{2}\frac{\tilde{K}_m}{\sqrt{\frac{\tilde{P}\tilde{K}_m\mu_{ref}}{\tilde{k}_{cat}}}}\right)$. In terms of proteome fraction, the ribosome offset can be expressed as $\phi_{R,0} \coloneqq \frac{P_R\tilde{R}_0}{\tilde{P}}$.

Additional References

[33] Liebermeister W & Klipp E (2006) Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor Biol Med Model* 3:41.

3 DISCUSSION

This thesis presents the biological and mathematical foundations for an analytical theory of the balanced growth of self-replicator models, termed Growth Balance Analysis. In its strive for simplicity and generality, it intends to be as simple as possible in its assumptions, while retaining the power to predict central aspects of cellular physiology. The analytical approach permits to study and predict the resource allocation in cell models of any size, unlike previous methods that rely on numerical optimization and are hence confined to small, coarse-grained models. The necessary analytical conditions for optimal cellular growth can be understood as consequences of the main constraints on the maximal balanced growth rate [44]: (i) mass conservation, (ii) reaction kinetics, and (iii) the limited capacity for cellular dry weight, here including also small molecules. The resulting optimization scheme can be understood from the following rationale: (i) the cell growth rate equals the dry weight production rate (e.g., in units of g/h) per dry weight (e.g., in g); (ii) at growth with constant dry weight density (e.g., in units of g/L) maintained by a corresponding increase in water content, the growth rate also equals the dry weight density production (in g/L/h) per dry weight density; (iii) cell growth is constrained by reaction kinetics, which depend on the concentrations (e.g., in units of g/L) of catalysts and reactants; and (iv) the total concentration of catalysts and reactants is constrained by the dry weight density.

The constraint on total cell dry weight capacity considered here is however not the only capacity constraint on cell growth. Other constraints on the concentration of molecules depend on their size (e.g., the macromolecular crowding of proteins, RNA, and DNA [9, 15, 35, 61] and the solvent capacity for small molecules [3, 54, 62]), their localization in the cell (e.g., due to the limited membrane real estate [60, 65]), and the total number of solved molecules (for the maintenance of osmotic pressure). However, considering only the total dry weight capacity has advantages, which are discussed next.

3.1 THE CONSTRAINT ON TOTAL CELL DRY WEIGHT DENSITY

de Groot et al. [28] have recently demonstrated that more than one capacity constraint may result in a combination of Elementary Flux Modes (EFMs) at maximal metabolic rates. However, in most physiological situations only one capacity constraint will be active; a single constraint and hence a unique EFM greatly simplifies the analytical

treatment, and hence this is what is considered in the mathematical formulation of GBA presented in this thesis.

Using the dry weight capacity constraint has two main benefits: (i) it represents a coarse-grained "compromise" over all possible constraints, simplifying the formalism, and (ii) it depends on all cellular components, which is essential to connect their costs and benefits.

Metabolites are commonly assumed to be too small to contribute significantly to cell dry weight [9, 44, 48]. This appears inconsistent with experimental measurements, which show total metabolite concentration in *E. coli* to be around 300 mM [10]. Assuming an average metabolite molecular weight of $\approx$ 100 g/mol, this results in a mass density of $\approx$ 30 g/L; with a total protein mass density [32] of around 135 g/L, this value corresponds to $30/135 \approx 22\%$ of total *E. coli* protein concentration. Thus, our theoretical treatment accounts for the contribution of all cellular components directly involved in growth to cellular dry weight, including small molecules.

3.2 THE INFLUENCE OF PRODUCTION COSTS ON OPTIMAL GROWTH

The different constraints on growth have different quantitative effects on the final optimal cellular state. The mass balance and dilution of every component by growth, for example, enforces the necessary net production of each cellular component, resulting in the respective production costs. To estimate typical values for these production costs, let us consider first the balance of marginal benefits of reactants α and total protein ($\eta_\alpha = \eta_P$) in the simple scenario without dependent reactants, which results in

$$\sum_j \frac{p_j}{k_j} \frac{\partial k_j}{\partial a_\alpha} = 1 - \sum_j \frac{\mu}{k_j} \left(I_{jP} - I_{j\alpha} \right) \quad , \tag{7}$$

with the summation on the right hand side being the contribution of production costs (Eq.(10) in Manuscript 1). This contribution is negligible when its magnitude is much less than 1, in which case the previous equation can be approximated as

$$\sum_j \frac{p_j}{k_j} \frac{\partial k_j}{\partial a_\alpha} \approx 1 \quad . \tag{8}$$

The production cost contribution depends directly on the ratio μ/k_j. To estimate a typical value for this ratio *in vivo*, let us assume that k_j values are approximately equal to the average [5] $k_{cat} \approx 10\ s^{-1}$, with an average enzyme molecular weight of 100 kDa and average substrate molecular weight of 100 Da. The average k_j in mass units is then $\approx 10/s \times 3600\ s/h \times 100\ Da\ /100000\ Da = 36\ h^{-1}$. For usual growth rates of about $\mu = 1\ h^{-1}$, the estimated ratio μ/k_j is then less than 0.03. This means that if the terms $\left(I_{jP} - I_{j\alpha} \right)$ are not too large, the production cost contribution is much less than one. These

considerations indicate that Eq.(8) is a reasonable approximation for Eq.(7) in most cases. For a ribosome reaction using only one substrate, a charged tRNA "T", which is not consumed by other reactions, Eq.(7) results in

$$\frac{p_R}{k_R}\frac{\partial k_R}{\partial b_T} = 1 - \frac{\mu}{k_R} \quad , \tag{9}$$

where only the protein production cost by the ribosome does not cancel out. Considering the mass balance of protein production and consumption ($v_R = p_R k_R = \mu P$), the production cost contribution is in this case just the fraction of ribosome protein itself ($\mu/k_R = p_R/P \equiv \phi_R$), which can be estimated experimentally. Typical values for the ribosomal protein fraction are [53] $\phi_R = 0.28$ at $\mu = 1.9\ h^{-1}$, growing on LB; $\phi_R = 0.14$ at $\mu = 0.58\ h^{-1}$ on glucose in batch culture; and $\phi_R = 0.04$ at $\mu = 0.12\ h^{-1}$ on glucose in a chemostat (see Fig.2 in Manuscript 1 for the values in 20 different growth conditions). In these cases, the influence of the production costs becomes negligible at intermediate to low growth rates.

The previous considerations about the influence of production costs on optimal cellular resource allocation indicate that in general, these costs can be seen as a second order effect; this effect "fine-tunes" the concentrations that result from optimizing the utilization of the cellular capacity especially at high growth rates, and can be neglected at slow to intermediate growth. We quantified the central importance of the capacity constraint for the maximal growth rate in the GBA framework, estimating a relative change in growth rate of about 0.7 times the relative change in cellular capacity for *E. coli*. A very similar value is also suggested by the limited available data on experimental reductions in the cellular dry weight density [14]. Thus, we expect that for any given vector of reaction fluxes, the set of *in vivo* concentrations of reactants and proteins is, at least approximately, the one resulting in the optimal use of dry weight density constrained by kinetics. In this case, optimal use means two things: (i) the cell is using its full dry weight capacity ρ in order to achieve the maximal growth rate possible; and (ii) this value ρ is the minimal sum over the concentrations $\mathbf{p}, \mathbf{a}$ that results in the given flux vector, since any other combination of concentrations resulting in these fluxes but larger total dry weight density would violate the capacity limit.

3.3 THE CONSTRAINTS ON CELLULAR GROWTH AND ITS OPTIMALITY

The GBA framework is formulated for the analysis of self-replicator models where each protein catalyzes exactly one reaction. This includes, for example, models accounting for transcription through RNA polymerase (enzyme) and DNA/mRNA (reactants). In this case, the network structure and kinetics of the gene expression apparatus ex-

plicitly constrain cell growth through the corresponding kinetic functions $\mathbf{k}$ and entries in the matrix A, so the gene expression cost is included in the production costs as defined here. Translation could be modeled in a more detailed process than a single "ribosome" reaction (as explored in this thesis), resulting again in potentially more realistic predictions of optimal resource allocation due to more realistic costs and benefits of cellular components.

It is important to note that regulation of reactions at any level (transcriptional, post-transcriptional, translational, post-translational, allosteric, competitive) can be accounted for in the GBA framework through appropriate biochemical reactions and corresponding kinetic functions [42]. In this way, also the costs and benefits of regulation will become intrinsic to the costs and benefits of cellular components. More generally, the input in the optimization described in Eq.(5) is a model defined by a triple $(A, \mathbf{k}, \rho)$, and by consequence its output is expected to become more realistic as A and $\mathbf{k}$ get more realistic, i.e., as the description of how growth is constrained by network structure and reaction kinetics becomes more detailed, given that these details are sufficiently backed up by experimental data.

It is also important to note that other constraints not included in the GBA paradigm are expected to influence cellular states, and "optimality" is primarily a theoretical tool. The assumption of a fixed environment, for example, also excludes the possibility of interaction with other organisms in the growth media, such as in cooperative growth of microbial communities [8]. This scenario configures a more complex optimization that goes beyond the single cell models examined here. Even with the hypothetical knowledge of all constraints (including a realistic representation of environmental fluctuations and of the costs of cellular regulation), the theoretically optimal state is not necessarily the exact state of cells found in nature. However, the constraints accounted for here strongly influence cell growth and cannot easily be violated by living systems. Thus, there is reason to believe that the quantitative principles presented in this thesis have the potential to capture the general trends governing cellular resource allocation. In fact, the above estimates based on average values of the growth rate and of catalytic constants [5] indicate that one of the constraints accounted for here, the mass conservation, has in general a minor impact on cellular resource allocation: production costs are in general low compared to other costs and benefits. This is confirmed by experimental data from *E. coli*; concentrations of ribosome [53] (Fig.2 in Manuscript 1), enzymes [53], and metabolites [23] (Figure 3 in Manuscript 2) are well explained based only on the optimal use of dry weight capacity constrained by kinetics. This approximate approach has the benefit of relying much less on the knowledge of the (frequently unknown [17]) kinetic parameters.

3.4 GROWTH LAWS, PROTEIN OFFSETS, AND UNDER-UTILIZED PROTEINS

The growth laws of proteome allocation are an important phenomenological tool utilized to understand cellular resource allocation [6], but their mechanistic origins have not yet been completely clarified [6, 32, 46]. This thesis shows that the linear growth rate dependence of protein concentrations, and in particular their "offsets", may emerge from the simple steady-state optimization of cellular capacity utilization by catalysts and their substrates, constrained by reaction kinetics. The quantitative predictions for irreversible reactions are in good agreement with experimental values of proteins for which the growth rate dependence of flux is known quantitatively. However, many reactions have a significant backward flux and violate the irreversibility assumption; in those cases, a higher enzyme expression and consequently a larger offset is expected to compensate for the enzyme "inefficiency".

The predicted optimal protein concentrations at different growth rates also help to explain the existence of "under-utilized" proteins [51] as a consequence of (optimal) under-saturation. On the other hand, the condition-dependent expression of "un-utilized" proteins with no growth-related function in *E. coli* indicates that this organism is evolutionary adapted to changing environments, not simply to steady-state conditions [51]. Accordingly, experimental studies show that *E. coli* is capable of increasing its growth rate when it adapts to a stable environment over many generations [31].

3.5 OUTLOOK

Some important problems mentioned in this thesis remain to be explored in future work. First, the shortage of kinetic parameters greatly limits the direct application of GBA to genome-scale models, even for the simplified optimization of capacity utilization constrained by kinetics. This problem might be approached in different ways: (i) by estimating kinetic parameters *in silico*, using, for example, machine learning techniques [29]; (ii) by developing systematic methods for coarse-graining existing models to smaller networks, in a way that the necessary kinetic parameters can be systematically estimated from the incomplete set of known parameters of reactions in the original model and/or from comparison to physiological data. Small cellular models have been shown to still capture some important information about cellular behaviour, for example in terms of its proteome sectors [7, 32].

Second, the balance equations are necessary but not sufficient conditions for optimal growth. A full analytical solution of the optimization problem demands the study of second order necessary condi-

tions or of the possible convexity of the optimization problem in Eq.(6). In the case of strict convexity, the balance equation would have only one solution, and would thus be necessary and sufficient to determine the optimal growth state.

Third, the framework of GBA as presented here requires the previous knowledge of the active reaction network (i.e., the reactions carrying flux in the EFM used at maximal growth rate). The determination of this reaction set is trivial in small models with only one EFM, but becomes a major limitation for the analysis of genome-scale models. In the study of this more general problem, instead of using the Lagrange multiplier method, one needs to use the Karush–Kuhn–Tucker (KKT) conditions [37], which account for constraints on concentrations that can be active or inactive; concentrations need to be nonnegative, but some may be zero, defining the "inactive" portion of the network. Alternatively, one might use an approximation that considers a limited set of EFMs, derived, e.g., from parsimonious FBA [30] in combination with manually added constraints (e.g., to enforce overflow metabolism). Finally, the generalization of GBA to multiple capacity constraints also requires the use of KKT conditions, as some of these constraints may not be active in a given optimal state.

BIBLIOGRAPHY

[1] Roi Adadi, Benjamin Volkmer, Ron Milo, Matthias Heinemann, and Tomer Shlomi. "Prediction of Microbial Growth Rate versus Biomass Yield by a Metabolic Network with Kinetic Parameters." In: *PLOS Computational Biology* 8.7 (2012), pp. 1–9. DOI: 10.1371/journal.pcbi.1002575. URL: https://doi.org/10.1371/journal.pcbi.1002575.

[2] S. Afriat. "Theory of Maxima and the Method of Lagrange." In: *SIAM Journal on Applied Mathematics* 20.3 (1971), pp. 343–357. DOI: 10.1137/0120037. URL: https://doi.org/10.1137/0120037.

[3] Daniel E. Atkinson. "Limitation of Metabolite Concentrations and the Conservation of Solvent Capacity in the Living Cell." In: *Current Topics in Cellular Regulation* 1.C (1969), pp. 29–43. ISSN: 00702137. DOI: 10.1016/B978-0-12-152801-0.50007-9.

[4] W. W. Baldwin, Richard Myer, Nicole Powell, Erika Anderson, and Arthur L. Koch. "Buoyant density of Escherichia coli is determined solely by the osmolarity of the culture medium." In: *Archives of Microbiology* 164.2 (1995), pp. 155–157. ISSN: 1432-072X. DOI: 10.1007/s002030050248. URL: https://doi.org/10.1007/s002030050248.

[5] Arren Bar-Even, Elad Noor, Avi Flamholz, Joerg M. Buescher, and Ron Milo. "Hydrophobicity and Charge Shape Cellular Metabolite Concentrations." In: *PLOS Computational Biology* 7.10 (2011), pp. 1–7. DOI: 10.1371/journal.pcbi.1002166. URL: https://doi.org/10.1371/journal.pcbi.1002166.

[6] Markus Basan. "Resource allocation and metabolism: the search for governing principles." In: *Current opinion in microbiology* 45 (2018), pp. 77–83.

[7] Markus Basan, Sheng Hui, Hiroyuki Okano, Zhongge Zhang, Yang Shen, James R. Williamson, and Terence Hwa. "Overflow metabolism in Escherichia coli results from efficient proteome allocation." In: *Nature* 528 (2015). Article, 99 EP –. URL: https://doi.org/10.1038/nature15765.

[8] A.E. Beck, K.A. Hunt, H.C. Bernstein, and R.P. Carlson. "Interpreting and Designing Microbial Communities for Bioprocess Applications, from Components to Interactions to Emergent Properties." In: *Biotechnology for Biofuel Production and Optimization*. Elsevier, 2016, pp. 407–432. ISBN: 9780444634757. DOI: 10.1016/B978-0-444-63475-7.00015-7. URL: http://dx.doi.org/

10.1016/B978-0-444-63475-7.00015-7https://linkinghub.elsevier.com/retrieve/pii/B9780444634757000157.

[9] Q. K. Beg, A. Vazquez, J. Ernst, M. A. de Menezes, Z. Bar-Joseph, A.-L. Barabási, and Z. N. Oltvai. "Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity." In: *Proceedings of the National Academy of Sciences* 104.31 (2007), pp. 12663–12668. ISSN: 0027-8424. DOI: 10.1073/pnas.0609845104. URL: https://www.pnas.org/content/104/31/12663.

[10] Bryson D. Bennett, Elizabeth H. Kimball, Melissa Gao, Robin Osterhout, Stephen J. Van Dien, and Joshua D. Rabinowitz. "Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli." In: *Nature Chemical Biology* 5 (2009). Article, 593 EP –. URL: https://doi.org/10.1038/nchembio.186.

[11] Jan Berkhout, Frank J. Bruggeman, and Bas Teusink. "Optimality Principles in the Regulation of Metabolic Networks." In: *Metabolites* 2.3 (2012), pp. 529–552. ISSN: 2218-1989. DOI: 10.3390/metabo2030529. URL: https://www.mdpi.com/2218-1989/2/3/529.

[12] Aarash Bordbar, Jonathan M. Monk, Zachary A. King, and Bernhard O. Palsson. "Constraint-based models predict metabolic and associated cellular functions." In: *Nature Reviews Genetics* 15 (2014). Review Article, 107 EP –. URL: https://doi.org/10.1038/nrg3643.

[13] Dennis P Bremer H. "Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates." In: *EcoSal Plus* (2008). URL: http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosal.5.2.3.

[14] D. Scott Cayley, Harry J. Guttman, and M. Thomas Record. "Biophysical Characterization of Changes in Amounts and Activity of Escherichia coli Cell and Compartment Water and Turgor Pressure in Response to Osmotic Stress." In: *Biophysical Journal* 78.4 (2000), pp. 1748 –1764. ISSN: 0006-3495. DOI: https://doi.org/10.1016/S0006-3495(00)76726-9. URL: http://www.sciencedirect.com/science/article/pii/S0006349500767269.

[15] Scott Cayley, Barbara A. Lewis, Harry J. Guttman, and M.Thomas Record. "...Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity: Implications for protein-DNA interactions in vivo." In: *Journal of Molecular Biology* 222.2 (1991), pp. 281 –300. ISSN: 0022-2836. DOI: https://doi.org/10.1016/0022-2836(91)90212-O. URL: http://www.sciencedirect.com/science/article/pii/0022283691902120.

[16] Anirikh Chakrabarti, Ljubisa Miskovic, Keng Cher Soh, and Vassily Hatzimanikatis. "Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints." In: *Biotechnology Journal* 8.9 (2013), pp. 1043–1057. DOI: 10.1002/biot.201300091. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/biot.201300091.

[17] Dan Davidi, Elad Noor, Wolfram Liebermeister, Arren Bar-Even, Avi Flamholz, Katja Tummler, Uri Barenholz, Miki Goldenfeld, Tomer Shlomi, and Ron Milo. "Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements." In: *Proceedings of the National Academy of Sciences* 113.12 (2016), pp. 3401–3406. ISSN: 0027-8424. DOI: 10.1073/pnas.1514240113. arXiv: arXiv:1404.2263v1. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1514240113.

[18] Erez Dekel and Uri Alon. "Optimality and evolutionary tuning of the expression level of a protein." In: *Nature* 436.7050 (2005), pp. 588–592. ISSN: 1476-4687. DOI: 10.1038/nature03842. URL: https://doi.org/10.1038/nature03842.

[19] Patrick P. Dennis and Hans Bremer. "Differential rate of ribosomal protein synthesis in Escherichia coli B/r." In: *Journal of Molecular Biology* 84.3 (1974), pp. 407 –422. ISSN: 0022-2836. DOI: https://doi.org/10.1016/0022-2836(74)90449-5. URL: http://www.sciencedirect.com/science/article/pii/0022283674904495.

[20] Adam M Feist and Bernhard O Palsson. "The biomass objective function." In: *Current Opinion in Microbiology* 13.3 (2010). Ecology and industrial microbiology • Special section: Systems biology, pp. 344 –349. ISSN: 1369-5274. DOI: https://doi.org/10.1016/j.mib.2010.03.003. URL: http://www.sciencedirect.com/science/article/pii/S1369527410000512.

[21] Sarah-Maria Fendt, Joerg Martin Buescher, Florian Rudroff, Paola Picotti, Nicola Zamboni, and Uwe Sauer. "Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity." In: *Molecular Systems Biology* 6.1 (2010). ISSN: 1744-4292. DOI: 10.1038/msb.2010.11. URL: http://msb.embopress.org/content/6/1/356.

[22] Julien Gagneur and Steffen Klamt. "Computation of elementary modes: a unifying framework and the new binary approach." In: *BMC Bioinformatics* 5.1 (2004), p. 175. ISSN: 1471-2105. DOI: 10.1186/1471-2105-5-175. URL: https://doi.org/10.1186/1471-2105-5-175.

[23] Luca Gerosa, Bart R.B. Haverkorn van Rijsewijk, Dimitris Christodoulou, Karl Kochanowski, Thomas S.B. Schmidt, Elad Noor, and Uwe Sauer. "Pseudo-transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data." In: *Cell Systems* 1.4 (2015), pp. 270 –282. ISSN: 2405-4712. DOI: https://doi.org/10.1016/j.cels.2015.09.008. URL: http://www.sciencedirect.com/science/article/pii/S2405471215001465.

[24] Nils Giordano, Francis Mairet, Jean-Luc Gouzé, Johannes Geiselmann, and Hidde de Jong. "Dynamical Allocation of Cellular Resources as an Optimal Control Problem: Novel Insights into Microbial Growth Strategies." In: *PLOS Computational Biology* 12.3 (2016). Ed. by Oleg A Igoshin, e1004802. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004802. URL: https://project.inria.fr/reset/https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1004802.s001{\&}type=supplementaryhttp://dx.plos.org/10.1371/journal.pcbi.1004802.

[25] Anne Goelzer, Vincent Fromion, and Gérard Scorletti. "Cell Design in Bacteria As a Convex Optimization Problem." In: *Automatica* 47.6 (2011), pp. 1210–1218. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2011.02.038. URL: http://dx.doi.org/10.1016/j.automatica.2011.02.038.

[26] Anne Goelzer et al. "Quantitative prediction of genome-wide resource allocation in bacteria." In: *Metabolic Engineering* 32 (2015), pp. 232–243. ISSN: 10967176. DOI: 10.1016/j.ymben.2015.10.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S1096717615001317https://www.sciencedirect.com/science/article/pii/S1096717615001317?via{\%}3Dihub.

[27] Julio C. González, Katrina Peariso, James E. Penner-Hahn, and Rowena G. Matthews. "Cobalamin-Independent Methionine Synthase from Escherichia coli: A Zinc Metalloenzyme." In: *Biochemistry* 35.38 (1996), pp. 12228–12234. ISSN: 0006-2960. DOI: 10.1021/bi9615452. URL: https://doi.org/10.1021/bi9615452.

[28] Daan H. de Groot, Coco van Boxtel, Robert Planqué, Frank J. Bruggeman, and Bas Teusink. "The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates." In: *PLOS Computational Biology* 15.3 (2019), pp. 1–24. DOI: 10.1371/journal.pcbi.1006858. URL: https://doi.org/10.1371/journal.pcbi.1006858.

[29] David Heckmann, Colton J. Lloyd, Nathan Mih, Yuanchi Ha, Daniel C. Zielinski, Zachary B. Haiman, Abdelmoneim Amer Desouki, Martin J. Lercher, and Bernhard O. Palsson. "Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models." In: *Na-*

ture Communications 9.1 (2018), p. 5252. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07652-6. URL: http://www.nature.com/articles/s41467-018-07652-6.

[30] Hermann-Georg Holzhütter. "The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks." In: *European Journal of Biochemistry* 271.14 (2004), pp. 2905–2922. DOI: 10.1111/j.1432-1033.2004.04213.x. eprint: https://febs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1432-1033.2004.04213.x. URL: https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1432-1033.2004.04213.x.

[31] Qiang Hua, Andrew R. Joyce, Bernhard Ø. Palsson, and Stephen S. Fong. "Metabolic Characterization of Escherichia coli Strains Adapted to Growth on Lactate." In: *Applied and Environmental Microbiology* 73.14 (2007), pp. 4639–4647. ISSN: 0099-2240. DOI: 10.1128/AEM.00527-07. eprint: https://aem.asm.org/content/73/14/4639.full.pdf. URL: https://aem.asm.org/content/73/14/4639.

[32] S. Hui, J. M. Silverman, S. S. Chen, D. W. Erickson, M. Basan, J. Wang, T. Hwa, and J. R. Williamson. "Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria." In: *Molecular Systems Biology* 11.2 (2015), e784–e784. ISSN: 1744-4292. DOI: 10.15252/msb.20145697. URL: http://msb.embopress.org/cgi/doi/10.15252/msb.20145697.

[33] Moshe Kafri, Eyal Metzl-Raz, Ghil Jona, and Naama Barkai. "The Cost of Protein Production." In: *Cell Reports* 14.1 (2016), pp. 22–31. ISSN: 22111247. DOI: 10.1016/j.celrep.2015.12.015. URL: http://dx.doi.org/10.1016/j.celrep.2015.12.015https://linkinghub.elsevier.com/retrieve/pii/S221112471501428X.

[34] Niels Ole Kjeldgaard, Ole Maaloe, and Moselio Schaechter. "The transition between different physiological states during balanced growth of Salmonella typhimurium." In: *Journal of general microbiology* 19 3 (1958), pp. 607–16.

[35] Stefan Klumpp, Matthew Scott, Steen Pedersen, and Terence Hwa. "Molecular crowding limits translation and cell growth." In: *Proceedings of the National Academy of Sciences* 110.42 (2013), pp. 16754–16759. ISSN: 0027-8424. DOI: 10.1073/pnas.1310377110. URL: https://www.pnas.org/content/110/42/16754.

[36] Arthur L. Koch and Carol S. Deppe. "In vivo assay of protein synthesizing capacity of Escherichia coli from slowly growing chemostat cultures." In: *Journal of Molecular Biology* 55.3 (1971), pp. 549 –562. ISSN: 0022-2836. DOI: https://doi.org/10.1016/0022-2836(71)90336-6. URL: http://www.sciencedirect.com/science/article/pii/0022283671903366.

[37] H. W. Kuhn and A. W. Tucker. "Nonlinear Programming." In: *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (1950), pp. 481 –492.

[38] Meiyappan Lakshmanan, Sichang Long, Kok Siong Ang, Nathan Lewis, and Dong-Yup Lee. "On the impact of biomass composition in constraint-based flux analysis." In: *bioRxiv* (2019). DOI: 10.1101/652040. eprint: https://www.biorxiv.org/content/early/2019/05/28/652040.full.pdf. URL: https://www.biorxiv.org/content/early/2019/05/28/652040.

[39] Joshua A. Lerman et al. "In silico method for modelling metabolism and gene product expression at genome scale." In: *Nature Communications* 3 (2012). Article, 929 EP –. URL: https://doi.org/10.1038/ncomms1928.

[40] Nathan E. Lewis, Harish Nagarajan, and Bernhard O. Palsson. "Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods." In: *Nature Reviews Microbiology* 10.4 (2012), pp. 291–305. ISSN: 17401534. DOI: 10.1038/nrmicro2737. URL: http://dx.doi.org/10.1038/nrmicro2737.

[41] Nathan E Lewis et al. "Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models." In: *Molecular Systems Biology* 6.1 (2010), p. 390. DOI: 10.1038/msb.2010.47. eprint: https://www.embopress.org/doi/pdf/10.1038/msb.2010.47. URL: https://www.embopress.org/doi/abs/10.1038/msb.2010.47.

[42] Wolfram Liebermeister and Edda Klipp. "Bringing metabolic networks to life: convenience rate law and thermodynamic constraints." In: *Theoretical Biology and Medical Modelling* 3.1 (2006), p. 41. ISSN: 1742-4682. DOI: 10.1186/1742-4682-3-41. URL: https://doi.org/10.1186/1742-4682-3-41.

[43] O. Maaløe. "Regulation of the Protein-Synthesizing Machinery—Ribosomes, tRNA, Factors, and So On." In: *Biological Regulation and Development: Gene Expression*. Boston, MA: Springer US, 1979, pp. 487–542. ISBN: 978-1-4684-3417-0. DOI: 10.1007/978-1-4684-3417-0_12. URL: https://doi.org/10.1007/978-1-4684-3417-0_12.

[44] Douwe Molenaar, Rogier van Berlo, Dick de Ridder, and Bas Teusink. "Shifts in growth strategies reflect tradeoffs in cellular economics." In: *Molecular Systems Biology* 5.1 (2009), p. 323. DOI: 10.1038/msb.2009.82. URL: https://onlinelibrary.wiley.com/doi/abs/10.1038/msb.2009.82.

[45] Matteo Mori, Terence Hwa, Olivier C. Martin, Andrea De Martino, and Enzo Marinari. "Constrained Allocation Flux Balance Analysis." In: *PLOS Computational Biology* 12.6 (2016), pp. 1–24.

DOI: 10.1371/journal.pcbi.1004913. URL: https://doi.org/10.1371/journal.pcbi.1004913.

[46] Matteo Mori, Severin Schink, David W. Erickson, Ulrich Gerland, and Terence Hwa. "Quantifying the benefit of a proteome reserve in fluctuating environments." In: *Nature Communications* 8.1 (2017), p. 1225. ISSN: 2041-1723. DOI: 10.1038/s41467-017-01242-8. URL: https://doi.org/10.1038/s41467-017-01242-8.

[47] Stefan Müller, Georg Regensburger, and Ralf Steuer. "Enzyme allocation problems in kinetic metabolic networks: Optimal solutions are elementary flux modes." In: *Journal of Theoretical Biology* 347 (2014), pp. 182 –190. ISSN: 0022-5193. DOI: https://doi.org/10.1016/j.jtbi.2013.11.015. URL: http://www.sciencedirect.com/science/article/pii/S0022519313005420.

[48] Elad Noor, Avi Flamholz, Arren Bar-Even, Dan Davidi, Ron Milo, and Wolfram Liebermeister. "The Protein Cost of Metabolic Fluxes: Prediction from Enzymatic Rate Laws and Cost Minimization." In: *PLOS Computational Biology* 12.11 (2016), pp. 1–29. DOI: 10.1371/journal.pcbi.1005167. URL: https://doi.org/10.1371/journal.pcbi.1005167.

[49] Edward J. O'Brien, Joshua A. Lerman, Roger L. Chang, Daniel R. Hyduke, and Bernhard Palsson. "Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction." In: *Molecular Systems Biology* (2013). ISSN: 17444292. DOI: 10.1038/msb.2013.52.

[50] Jeffrey D. Orth, Ines Thiele, and Bernhard Ø Palsson. "What is flux balance analysis?" In: *Nature Biotechnology* 28 (2010), 245 EP –. URL: https://doi.org/10.1038/nbt.1614.

[51] Edward J. O'Brien, Jose Utrilla, and Bernhard O. Palsson. "Quantification and Classification of E. coli Proteome Utilization and Unused Protein Costs across Environments." In: *PLOS Computational Biology* 12.6 (2016), pp. 1–22. DOI: 10.1371/journal.pcbi.1004998. URL: https://doi.org/10.1371/journal.pcbi.1004998.

[52] Junyoung O. Park, Sara A. Rubin, Yi-Fan Xu, Daniel Amador-Noguez, Jing Fan, Tomer Shlomi, and Joshua D. Rabinowitz. "Metabolite concentrations, fluxes and free energies imply efficient enzyme usage." In: *Nature Chemical Biology* 12 (2016). Article, 482 EP –. URL: https://doi.org/10.1038/nchembio.2077.

[53] Alexander Schmidt, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer, Luciano Callipo, Kèvin Knoops, Manuel Bauer, Ruedi Aebersold, and Matthias Heinemann. "The quantitative and condition-dependent Escherichia coli proteome." In: *Nature*

Biotechnology 34 (2015), 104 EP –. URL: https://doi.org/10.1038/nbt.3418.

[54] S. Schuster, R. Schuster, and R. Heinrich. "Minimization of intermediate concentrations as a suggested optimality principle for biochemical networks." In: *Journal of Mathematical Biology* 29.5 (1991), pp. 443–455. ISSN: 1432-1416. DOI: 10.1007/BF00160471. URL: https://doi.org/10.1007/BF00160471.

[55] Stefan Schuster, Thomas Pfeiffer, and David A. Fell. "Is maximization of molar yield in metabolic networks favoured by evolution?" In: *Journal of Theoretical Biology* 252.3 (2008). In Memory of Reinhart Heinrich, pp. 497 –504. ISSN: 0022-5193. DOI: https://doi.org/10.1016/j.jtbi.2007.12.008. URL: http://www.sciencedirect.com/science/article/pii/S0022519307006340.

[56] Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. "Interdependence of cell growth and gene expression: Origins and consequences." In: *Science* 330.6007 (2010), pp. 1099–1102. ISSN: 00368075. DOI: 10.1126/science.1192588.

[57] Matthew Scott, Stefan Klumpp, Eduard M Mateescu, and Terence Hwa. "Emergence of robust growth laws from optimal regulation of ribosome synthesis." In: *Molecular Systems Biology* 10.8 (2014). ISSN: 1744-4292. DOI: 10.15252/msb.20145379. URL: http://msb.embopress.org/content/10/8/747.

[58] Tomer Shlomi, Tomer Benyamini, Eyal Gottlieb, Roded Sharan, and Eytan Ruppin. "Genome-Scale Metabolic Modeling Elucidates the Role of Proliferative Adaptation in Causing the Warburg Effect." In: *PLOS Computational Biology* 7.3 (2011), pp. 1–8. DOI: 10.1371/journal.pcbi.1002018. URL: https://doi.org/10.1371/journal.pcbi.1002018.

[59] Daniel M. Stoebel, Antony M. Dean, and Daniel E. Dykhuizen. "The Cost of Expression of Escherichia coli lac Operon Proteins Is in the Process, Not in the Products." In: *Genetics* 178.3 (2008), pp. 1653–1660. ISSN: 0016-6731. DOI: 10.1534/genetics.107.085399. URL: https://www.genetics.org/content/178/3/1653.

[60] Mariola Szenk, Ken A. Dill, and Adam M.R. de Graff. "Why Do Fast-Growing Bacteria Enter Overflow Metabolism? Testing the Membrane Real Estate Hypothesis." In: *Cell Systems* 5.2 (2017), pp. 95 –104. ISSN: 2405-4712. DOI: https://doi.org/10.1016/j.cels.2017.06.005. URL: http://www.sciencedirect.com/science/article/pii/S2405471217302338.

[61] Benjamín J Sánchez, Cheng Zhang, Avlant Nilsson, Petri-Jaan Lahtvee, Eduard J Kerkhoven, and Jens Nielsen. "Improving the phenotype predictions of a yeast genome-scale metabolic

model by incorporating enzymatic constraints." In: *Molecular Systems Biology* 13.8 (2017), p. 935. DOI: 10.15252/msb.20167411. eprint: https://www.embopress.org/doi/pdf/10.15252/msb.20167411. URL: https://www.embopress.org/doi/abs/10.15252/msb.20167411.

[62] Naama Tepper, Elad Noor, Daniel Amador-Noguez, Hulda S. Haraldsdóttir, Ron Milo, Josh Rabinowitz, Wolfram Liebermeister, and Tomer Shlomi. "Steady-State Metabolite Concentrations Reflect a Balance between Maximizing Enzyme Efficiency and Minimizing Total Metabolite Load." In: *PLOS ONE* 8.9 (2013). DOI: 10.1371/journal.pone.0075370. URL: https://doi.org/10.1371/journal.pone.0075370.

[63] Meike T. Wortel, Han Peters, Josephus Hulshof, Bas Teusink, and Frank J. Bruggeman. "Metabolic states with maximal specific rate carry flux through an elementary flux mode." In: *The FEBS Journal* 281.6 (2014), pp. 1547–1555. DOI: 10.1111/febs.12722. URL: https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/febs.12722.

[64] Conghui You, Hiroyuki Okano, Sheng Hui, Zhongge Zhang, Minsu Kim, Carl W. Gunderson, Yi-Ping Wang, Peter Lenz, Dalai Yan, and Terence Hwa. "Coordination of bacterial proteome with metabolism by cyclic AMP signalling." In: *Nature* 500 (2013). Article, 301 EP –. URL: https://doi.org/10.1038/nature12446.

[65] Kai Zhuang, Goutham N Vemuri, and Radhakrishnan Mahadevan. "Economics of membrane occupancy and respiro-fermentation." In: *Molecular Systems Biology* 7.1 (2011). ISSN: 1744-4292. DOI: 10.1038/msb.2011.34. URL: http://msb.embopress.org/content/7/1/500.