Resource allocation to cell envelopes and the scaling of bacterial growth rate

Bogi Trickovic[1*], Michael Lynch[1]

1 Center for Mechanisms of Evolution, Arizona State University, Tempe, Arizona, USA

* btrickov@asu.edu

1 Abstract

Although various empirical studies have reported a positive correlation between the specific growth rate and cell size across bacteria, it is currently unclear what causes this relationship. We conjecture that such scaling occurs because smaller cells have a larger surface-to-volume ratio and thus have to allocate a greater fraction of the total resources to the production of the cell envelope, leaving fewer resources for other biosynthetic processes. To test this theory, we developed a coarse-grained model of bacterial physiology composed of the proteome that converts nutrients into biomass, with the cell envelope acting as a resource sink. Assuming resources are partitioned to maximize the growth rate, the model yields expected scalings. Namely, the growth rate and ribosomal mass fraction scale negatively, while the mass fraction of envelope-producing enzymes scales positively with surface-to-volume. These relationships are compatible with growth measurements and quantitative proteomics data reported in the literature.

2 Author summary

It is widely known that smaller eukaryotes tend to grow faster. However, this trend does not hold in bacteria, where small-celled species grow slower. We propose that small bacteria – compared to their larger counterparts – have to invest a greater fraction of the total resources to cell envelope owing to their large surface-to-volume ratio, leaving fewer resources to internal biosynthetic processes that build the cell. By representing the cell as being composed of proteins (that convert nutrients into biomass), and cell envelope, we find that cells with large surface-to-volume ratio grow more slowly because they have to invest resources in the production of the cell envelope and the enzyme machinery that builds this structure, thus leaving fewer resources to ribosomes that replicate the cell. These predictions are corroborated by comparison with growth rate data across more than 200 bacterial species.

3 Introduction

The rate of cell growth varies across bacterial species. The inhabitant of salt marshes, *Vibrio natriegens*, divides in 10 minutes [1], whereas the causal agent of gum disease, *Treponema denticola*, takes 20 hours to divide [2]. Given the centrality of the growth rate in physiology, ecology, and evolution, efforts have been invested in understanding the causes of this variation. Previous meta-analyses have reported that the overall pattern of variation is predictable: bacterial species with larger cell volume tend to grow faster. This observation was made in both heterotrophic [3, 4] and autotrophic

bacteria [5,6]. The causes of this scaling are unclear. Eukaryotic growth rates scale negatively with body size, implying that the theories explaining this pattern – such as those invoking transport-related constraints [7,8] and limits imposed by self-shading of chlorophyll [9] – are fundamentally inadequate in accounting for the opposite scaling observed in bacteria.

Given that the surface-to-volume ratio increases with decreasing cell size, small cells will have more weight sequestered in membranes and walls than their larger counterparts. Therefore, as the cell size decreases, more resources have to be invested in the production of the cell envelope, implying that fewer resources can be invested in the biosynthetic processes that replicate the cell. It was initially suggested that this constraint might impose the limit on the smallest size that a cell can attain [10]. This idea has been further used to explain the lower limit on the size of the photoautotrophic organism and why phytoplankton growth rates appear to increase with cell volume [11,12]. Our goal is to formalize this theory and investigate whether it can explain the growth scaling in heterotrophic bacteria. We start by imagining the fastest-growing cell as a bag of self-replicating ribosomes [13]. However, the growth rate is depressed below this perfect state because the cell has to invest resources in (1) machinery that acquires and converts nutrients to fuel ribosomes, and (2) cell envelope, which is necessary for the cell to maintain the proper shape. As the cell becomes smaller, a larger fraction of the resources are diverted to envelope, and fewer resources are left for ribosomes which, in turn, means that cells grow slowly.

We formalize the afore-mentioned verbal argument in sections 4.1 and 4.2, and then use this framework to (1) obtain the simple analytical solution for maximal attainable growth rate given a particular cell size (section 4.3), and (2) explain how proteome composition ought to change across growth conditions (section 4.4). These predictions are tested using quantitative proteomic data of the model bacterium, *Escherichia coli* (section 5.1). Next, we investigate how the presence of an envelope affects the scaling of the growth rate and how these effects depend on the underlying model assumptions (section 5.2). Lastly, these cross-species scaling expectations are compared against data on bacterial growth rates, cell sizes, and proteome compositions (sections 5.4 and 5.3).

4 Materials and methods

4.1 Derivation of the steady-state growth rate

In our model, a cell is composed of two metabolite species and three protein species (Fig 1). External nutrient concentration is assumed to be constant, thus mimicking the nutrient-replete conditions when bacteria are grown in the lab. These nutrients are taken up and converted into building block b which corresponds to amino acids. Although we refer to species l as lipids, this group includes all molecules used in the construction of cell envelope, such as peptides and saccharides (including components of membrane lipoproteins and peptidoglycan). We use upper-case symbols to refer to absolute abundances, and lower-case symbols for relative abundances or concentrations. All chemical reactions obey Michaelis-Menten kinetics, and we assume that half-saturation constants (K_M) for all reactions are identical. We ultimately focus on the special case when all reactions are operating at saturation ($K_M = 0$) because this permits a simple analytical solution for the steady-state of various cellular features, and we only vary K_M to compare this limiting behavior to a more general case when cellular enzymes are not saturated. Our setup is similar to a model previously used to study the emergence of the adder model of bacterial cell division [14], except that our cell divides once a critical volume, and not critical abundance of division protein, is reached. Table 1 outlines the meaning of symbols that are used throughout the main text.

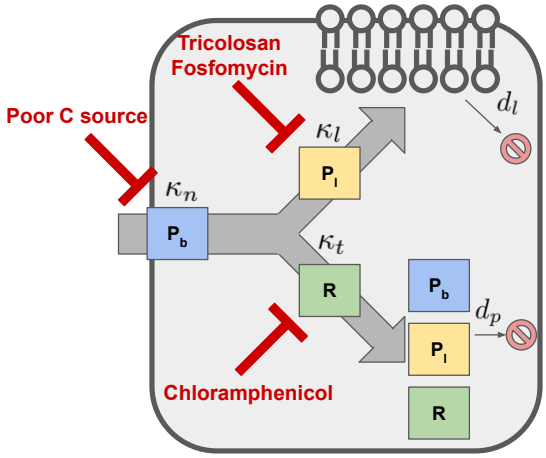

Figure 1. Model of the cell. Nutrients are taken up from the external environment via the building block producer (P_b). The produced block is then converted into lipids and other cell envelope components via the lipid-producer (P_l), and into proteins by the ribosome (R). Envelope and protein synthesis production can be inhibited by antibiotics, whereas nutrient uptake can be reduced by growing the culture on a poor carbon source. Envelope and proteins are degraded at rates d_l and d_p, respectively. Building blocks are represented as grey arrows flowing through protein machinery. Envelope components are depicted in the membrane.

The time-evolution of metabolite concentrations are:

$$\dot{b} = k_n p_b(t) - \left(k_t r(t) + k_l p_l(t)\right) \frac{b(t)}{K_M + b(t)} - \lambda(t) b(t) \tag{1a}$$

$$\dot{l} = \frac{k_l}{m_s} p_l(t) \frac{b(t)}{K_M + b(t)} - (d_l + \lambda(t)) l(t) \tag{1b}$$

The production of b occurs at rate k_n, which represents a pseudo-first-order rate constant that depends on the nutrient status of the environment inhabited by the cell. Rate constants k_n, k_l, and k_t are the turnover numbers for reactions of nutrient processing, envelope synthesis, and translation, while m_s is the size an envelope unit expressed in terms of the number of building blocks. Envelope components are eliminated from the cell by degradation at rate d_l, and by dilution due to growth at rate $\lambda(t)$. Note that the growth rate is time-dependent because it is a function of state variables (i.e., molecular abundances). We neglect the degradation of the free amino acids and focus only on the turnover of cell envelope components.

Protein species p_b, p_l, and r are produced by ribosomes which represent the autocatalytic part of the cell. All proteins have associated degradation rates d_p, with the exception of ribosomes which are reported to be remarkably stable in exponentially growing cells [15]. Sizes of the metabolic protein (P_b and P_l) and the ribosome are m_p and m_r, respectively. The time-evolution equations are identical in form to equations

Symbol	Meaning
$p_b(t)$	The relative abundance of building block-producing enzyme at time t
$p_l(t)$	The relative abundance of envelope-producing enzyme at time t
$r(t)$	The relative abundance of ribosomes at time t
$l(t)$	The relative abundance of envelope component
$P_b(t)$	The absolute abundance of building block producer
$P_l(t)$	The absolute abundance of envelope producer
$R(t)$	The absolute abundance of ribosomes
$L(t)$	The absolute abundance of envelope component
ϕ_b	The ribosome fraction allocated to translation of building block producers
ϕ_l	The ribosome fraction allocated to translation of envelope producers
ϕ_r	The ribosome fraction allocated to translation of ribosomes
Φ_B	The proteome mass fraction allocated to translation of building block producers
Φ_L	The proteome mass fraction allocated to translation of envelope producers
Φ_R	The proteome mass fraction allocated to translation of ribosomes
m_p	Length of building block producers and envelope producers in amino acids
m_r	Lenght of ribosome in amino acids
m_s	Length of the unit of envelope (lipids, peptides, and sugars) in amino acids
k_n	The turnover number of building block producers
k_l	The turnover number of envelope producers
k_t	The turnover number of ribosomes
d_p	Protein degradation rate
d_l	Envelope degradation rate
K_M	Michaelis-Menten constant for all chemical reactions
Π	Surface area-to-cytoplasmic volume ratio
β	The number of envelope units per unit of surface area
γ	The number of amino acids in proteome per unit of cytoplasmic volume
ϵ	The resource cost of the unit S/V ($m_s\beta/\gamma$)
$V(t)$	Volume of the cytoplasm
$\lambda(t)$	The specific growth rate at time t

Table 1. List of symbols used in the text and their meaning.

for metabolites:

$$\dot{p_b} = \frac{k_t}{m_p} r(t) \phi_b \frac{b(t)}{K_M + b(t)} - (d_p + \lambda(t)) p_b(t) \tag{2a}$$

$$\dot{p_l} = \frac{k_t}{m_p} r(t) \phi_l \frac{b(t)}{K_M + b(t)} - (d_p + \lambda(t)) p_l(t) \tag{2b}$$

$$\dot{r} = \frac{k_t}{m_r} r(t) \phi_r \frac{b(t)}{K_M + b(t)} - \lambda(t) r(t) \tag{2c}$$

with the only notable difference that – given the finite ribosomal pool – ribosomes have to be partitioned between different protein components, and this is denoted with ϕ_x, which represents the fraction of total ribosomal concentration that is allocated to translation of protein species x. We will assume that each molecular species contributes to volume proportional to its size in amino acids. For example, ribosomes are about 20x larger than other metabolic proteins, contributing 20x more to the cell volume. Concentrations of each species are defined as the abundance (labeled with the capital letter) divided by cytoplasmic volume V:

$$x(t) = \frac{X(t)}{V(t)}, \quad x \in \{b, l, p_b, p_l, r\}, \quad X \in \{B, L, P_b, P_l, R\} \tag{3a}$$

$$V(t) = m_r R(t) + m_p P_b(t) + m_p P_l(t) + B(t) \tag{3b}$$

The volume V corresponds to the internal volume of the cell where the chemical reactions take place, and including the envelope in the volume of the cell would imply that one can slow down chemical reactions by, say, increasing the thickness of the cell wall. As this is a non-sensical conclusion, we assume that L does not contribute to cytoplasmic volume. The cell is assumed to grow exponentially, such that $\dot{V} = \lambda(t) V(t)$. Given that both volume and concentrations are time-dependent, the left hand side of equations (1a–2c) will take the form:

$$\dot{x} = \frac{\dot{X}}{V(t)} - \frac{X(t)}{V(t)} \frac{\dot{V}}{V(t)} = \frac{\dot{X}}{V(t)} - x(t) \lambda(t) \tag{4a}$$

Substituting Eq (4a) in the system (1a–2c) and rearranging yields:

$$\dot{B} = k_n P_b(t) - \left(k_t R(t) + k_l P_l(t) \right) \frac{B(t)/V(t)}{K_M + B(t)/V(t)} \tag{5a}$$

$$\dot{L} = \frac{k_l}{m_s} P_l(t) \frac{B(t)/V(t)}{K_M + B(t)/V(t)} - d_l L(t) \tag{5b}$$

$$\dot{P_b} = \frac{k_t}{m_p} R(t) \phi_b \frac{B(t)/V(t)}{K_M + B(t)/V(t)} - d_p P_b(t) \tag{5c}$$

$$\dot{P_l} = \frac{k_t}{m_p} R(t) \phi_l \frac{B(t)/V(t)}{K_M + B(t)/V(t)} - d_p P_l(t) \tag{5d}$$

$$\dot{R} = \frac{k_t}{m_r} R(t) \phi_r \frac{B(t)/V(t)}{K_M + B(t)/V(t)} \tag{5e}$$

When cells are in a steady-state, all cellular components grow exponentially at a constant rate $\tilde{\lambda}$ which is the steady-state growth rate. We use this property to find the steady-state of the dynamical system. More precisely, we have:

$$X(t) = X(0) e^{\tilde{\lambda} t} \implies \dot{X} = \tilde{\lambda} X(0) e^{\tilde{\lambda} t} \tag{6a}$$

Note that $X(0) = k\tilde{X}$, where $\tilde{X}$ is abundance at cell division time when the cell is in the steady-state, and k is 1/2 if the abundance doubles over the life cycle. Substituting Eq 6a in the system (5a–5e) leads to:

$$\tilde{\lambda}\tilde{B} = k_n\tilde{P}_b - \left(k_t\tilde{R} + k_l\tilde{P}_l\right)\frac{\tilde{B}/\tilde{V}}{K_M + \tilde{B}/\tilde{V}} \tag{7a}$$

$$\tilde{\lambda}\tilde{L} = \frac{k_l}{m_s}\tilde{P}_l\frac{\tilde{B}/\tilde{V}}{K_M + \tilde{B}/\tilde{V}} - d_l\tilde{L} \tag{7b}$$

$$\tilde{\lambda}\tilde{P}_b = \frac{k_t}{m_p}\tilde{R}\phi_b\frac{\tilde{B}/\tilde{V}}{K_M + \tilde{B}/\tilde{V}} - d_p\tilde{P}_b \tag{7c}$$

$$\tilde{\lambda}\tilde{P}_l = \frac{k_t}{m_p}\tilde{R}\phi_l\frac{\tilde{B}/\tilde{V}}{K_M + \tilde{B}/\tilde{V}} - d_p\tilde{P}_l \tag{7d}$$

$$\tilde{\lambda}\tilde{R} = \frac{k_t}{m_r}\tilde{R}\phi_r\frac{\tilde{B}/\tilde{V}}{K_M + \tilde{B}/\tilde{V}} \tag{7e}$$

An intuitive interpretation of these equations is that the rate at which molecules are synthesized and degraded ultimately equals the rate at which they are diluted by growth. Note that $\tilde{V}$ is the steady-state cell volume which is by definition the critical volume at which cell divides. To simplify downstream expressions, let $\kappa_n = k_n/m_p$, $\kappa_t = k_t/m_r$, $\kappa_l = k_l/m_p$. These are rate constants scaled by the size of a particular enzyme, capturing how fast an enzyme is relative to its size. Solving for molecular abundances yields:

$$\tilde{B} = \left(\frac{\kappa_n\phi_b - \lambda - d_p}{\lambda} - \frac{\kappa_l\phi_l}{\kappa_t\phi_r}\right)\chi \tag{8a}$$

$$\tilde{L} = \frac{\lambda\phi_l\kappa_l/m_s}{\kappa_t\phi_r(d_l + \lambda)}\chi \tag{8b}$$

$$\tilde{P}_b = \frac{\phi_b}{m_p}\chi \tag{8c}$$

$$\tilde{P}_l = \frac{\phi_l}{m_p}\chi \tag{8d}$$

$$\tilde{R} = \frac{(\lambda + d_p)\phi_r}{m_r\lambda}\chi \tag{8e}$$

where

$$\chi = \frac{\lambda^2\kappa_t\phi_r K_M\tilde{V}}{(\lambda - \kappa_t\phi_r)(\kappa_l\phi_l\lambda + \kappa_t(\lambda + d_p - \kappa_n\phi_b)\phi_r)}$$

By definition, proteome mass fractions are:

$$\Phi_B = \frac{m_p\tilde{P}_b}{m_p\tilde{P}_b + m_p\tilde{P}_l + m_r\tilde{R}}, \quad \Phi_L = \frac{m_p\tilde{P}_l}{m_p\tilde{P}_b + m_p\tilde{P}_l + m_r\tilde{R}}, \quad \Phi_R = \frac{m_r\tilde{R}}{m_p\tilde{P}_b + m_p\tilde{P}_l + m_r\tilde{R}} \tag{9}$$

By replacing steady-state abundances with solutions in 8a–8e, we have:

$$\Phi_B = \frac{\tilde{\lambda}\phi_b}{\tilde{\lambda} + d_p\phi_r}, \quad \Phi_L = \frac{\tilde{\lambda}\phi_l}{\tilde{\lambda} + d_p\phi_r}, \quad \Phi_R = \frac{(d_p + \tilde{\lambda})\phi_r}{\tilde{\lambda} + d_p\phi_r} \tag{10}$$

In the limit when there is no protein degradation ($d_p = 0$), the proteomic mass fractions are equal to the fraction of ribosomes allocated to the synthesis of a particular protein component. This result was originally obtained in [16]. However, we are still lacking the solution for the steady-state growth rate $\tilde{\lambda}$. Given that volume increases exponentially and from the definition of the cell volume (Eq 3b), we have:

$$\lambda(t) = \dot{V}/V(t) = m_p \left[\kappa_n \frac{P_b(t)}{V(t)} - \left(\kappa_l \frac{B(t)/V(t)}{K_M + B(t)/V(t)} \frac{P_l(t)}{V(t)} + d_p \left(\frac{P_b(t)}{V(t)} + \frac{P_l(t)}{V(t)} \right) \right) \right] \tag{11a}$$

$$\tilde{\lambda} = m_p \Bigg[\underbrace{\kappa_n \frac{\tilde{P}_b}{\tilde{V}}}_{\text{aa influx}} - \Bigg(\underbrace{\kappa_l \frac{\tilde{B}/\tilde{V}}{K_M + \tilde{B}/\tilde{V}} \frac{\tilde{P}_l}{\tilde{V}}}_{\text{outflux to } L} + \underbrace{d_p \left(\frac{\tilde{P}_b}{\tilde{V}} + \frac{\tilde{P}_l}{\tilde{V}} \right)}_{\text{dissipation via degradation}} \Bigg) \Bigg] \tag{11b}$$

where the Eq 11b has been obtained by applying 6a to molecular abundances in 11a. The steady-state growth rate is equal to the rate at which the building blocks are generated from the acquired nutrients minus the building blocks that are diverted into cell envelope synthesis and thus do not contribute to cytoplasmic volume production or are dissipated through protein degradation occurring at a rate d_p. Note that diversion to lipids and other envelope constituents does not explicitly enter the growth rate because L does not contribute to the cytoplasmic volume. However, lipid production does affect the growth rate by altering the amount of available building blocks $\tilde{B}$, and by diverting proteome allocation to P_l. Finally, substituting equations 8a–8e into 11b, we retrieve a cubic polynomial:

$$a_0 + a_1\tilde{\lambda} + a_2\tilde{\lambda}^2 + a_3\tilde{\lambda}^3 = 0 \tag{12}$$

with the following coefficients:

$$a_0 = 0 \tag{13a}$$
$$a_1 = \kappa_t^2 \phi_r^2 (d_p - \kappa_n \phi_b) \tag{13b}$$
$$a_2 = \kappa_t \phi_r \Big((1 + K_M)\kappa_n \phi_b + \kappa_l \phi_l - d_p(1 + K_M(\phi_b + \phi_l)) + \kappa_t \phi_r \Big) \tag{13c}$$
$$a_3 = -(1 + K_M)\kappa_l \phi_l - \kappa_t \phi_r \tag{13d}$$

Eq12 has two non-zero roots:

$$\tilde{\lambda} = \frac{-a_2 \pm \sqrt{a_2^2 - 4a_3 a_1}}{2a_3} \tag{14}$$

In the case of the positive branch, $\tilde{\lambda} = 0$ when $\phi_r = 0$ and $\tilde{\lambda} = \kappa_t$ when $\phi_r = 1$. This is clearly incorrect, given that $\tilde{\lambda}$ should be zero both when the cell does not have ribosomes (because there is nothing to build the cell) and when the cell contains only ribosomes without any other protein components (because there is nothing to deliver building blocks to the ribosomes). On the contrary, the negative branch attains values of zero both when $\phi_r = 0$ and when $\phi_r = 1$, so we take this solution as biologically meaningful. Finally, we obtain the steady-state abundances by substituting Eq 14 in equations 8a–8e.

The analytical solution for the steady-state growth rate $\tilde{\lambda}$ (Eq 14) is validated by comparison to numerically integrating equations 5a–5e (Fig 2; upper row). We assign initial abundances and integrate the system of ODEs until the cell volume reaches the

critical division volume. Next, we reduce the initial abundances by the factor of two and restart the integration process. This emulates the process of cell division when cellular content is equipartitioned among the daughter cells. After a short out-of-equilibrium phase, the cell lineage settles into a steady-state that matches the analytical solution for $\tilde{\lambda}$ and the steady-state abundances.

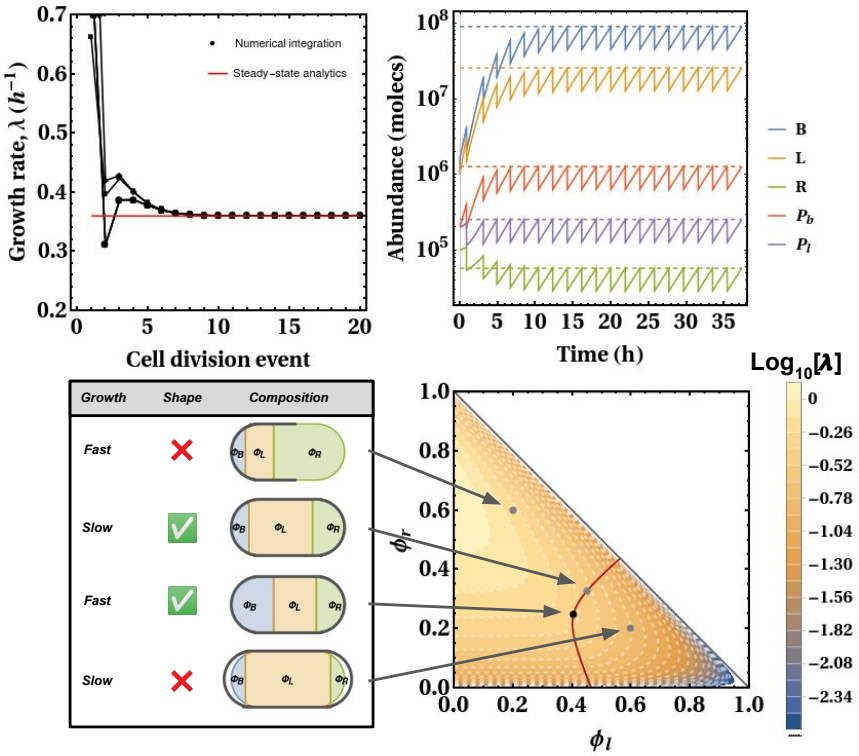

Figure 2. Comparison between the analytical and numerical solution for the cell in the steady-state. The upper left panel shows a close match of the steady-state growth rate (solid red line) relative to numerically integrated system of ODEs (black dots). Each set of points corresponds to a different set of initial molecular abundances. On the right, a similar match between analytics and numerics also exists for abundances of five molecular species; Each grey dashed line represents the steady-state analytical solution given by Eq (8a–8e). The bottom row outlines the kind of problem that a cell has to solve. Parameters (bottom figures): $\kappa_l = 4.32$ h^{-1}, $\kappa_t = 2.59$ h^{-1}, $d_p = d_l = 0$ h^{-1}, $K_M = 0.1$, $\gamma = 9.75 \times 10^8$ aa/μm^3, $\beta = 2 \times 10^6$ molecs/μm^2, $m_s = 314$ aa, $V = 3 \times 10^9$ aa/μm^3. Parameters (upper figures): $\kappa_l = 50$ h^{-1}, $\kappa_t = 1$ h^{-1}, $d_p = d_l = 0.1$ h^{-1}, $K_M = 0.01$, $V = 10^9$ aa/μm^3, $\phi_b = 0.5$, $\phi_l = 0.1$, $\phi_r = 0.4$.

Although the explicit solution for the maximal $\tilde{\lambda}$ is prohibitively difficult to obtain, we can gain some insight by examining boundary cases when the cell does not have a membrane ($\phi_l = 0$) and cellular processes are infinitely fast. When $\kappa_n \to \infty$, $\tilde{\lambda} \to \kappa_t/(1 + K_M)$, and, conversely, as $\kappa_t \to \infty$ we have $\tilde{\lambda} \to \kappa_n - d_p$. These two asymptotic results occur because the cell which instantaneously converts nutrients to building blocks will saturate the downstream translation machinery and the growth rate will be determined by the rate at which ribosomes operate. Conversely, when ribosomes are infinitely fast, the growth rate is set by how fast the building blocks are supplied to protein synthesis machinery.

4.2 Optimization problem of the cell

The steady-state growth rate $\tilde{\lambda}$ is a function of the ribosome partitioning parameters ϕ, and the cell has a task to find an optimal partitioning across different proteome components such that $\tilde{\lambda}$ is maximized. This maximization is achieved by homeostatic mechanisms that constantly take input from the environment and adjust proteome composition accordingly. Note, however, that the cell could attain maximal growth even if the partitioning parameters were hard-coded in the genome and thus cannot be actively adjusted (i.e., when homeostatic mechanisms are absent). In that case, the process of finding the peak is governed by natural selection and other evolutionary forces. Hence, our model holds regardless of whether the actual cells have regulatory mechanisms or not.

Growth rate maximization proceeds with two constraints. First, the fractions of ribosomes allocated to the production of three different proteome sectors have to sum to unity (c1). Second, the surface-to-volume ratio has to satisfy the constraint based on the shape of the cell (c2):

$$\begin{aligned} &\text{Maximize } \tilde{\lambda} \\ &\text{subject to:} \quad \phi_b + \phi_l + \phi_r = 1 \qquad \text{(c1)} \\ &\qquad\qquad \frac{\tilde{L}/\beta}{\tilde{V}/\gamma} = \Pi \qquad \text{(c2)} \end{aligned}$$

where $\Pi = 4.836(V/\gamma)^{-1/3}$ is the surface-to-volume ratio assuming the cell is a sphere. Note that V is the cytoplasmic volume, and thus the Π is the ratio of cell surface area to cytoplasmic volume. Parameters γ and β are unit conversion factors corresponding to the number of amino acids per unit of cell volume (molecs/μm^3), and the number of lipids/envelope components per unit of cell surface area (molecs/μm^2); These values are reported in Table 2. To intuitively understand the optimization problem, consider a landscape of ribosomal partitioning parameters and the resulting growth rates (Fig 2; bottom row). In the absence of geometric constraint (c2), the cell maximizes growth rate by (1) completely abolishing expression of the envelope-producing enzyme (when $\phi_l = 0$), and (2) by optimally expressing ribosomes (when $0 < \phi_r < 1$); This is because envelope-producing enzyme acts as a burden that diverts resources from other proteome components, and because the cell has to balance the production of proteins (for which high ribosomal expression is required) and the production of building blocks that fuels the translation (for which low ribosomal expression is required). These two aspects explain why the growth rate is a monotonically decreasing function of ϕ_l, and a non-monotonic function of ϕ_r.

Now suppose that the cell maximizes the growth rate, while at the same time having to maintain a surface-to-volume ratio dictated by the cell's geometry; Out of all partitioning parameters, only a subset satisfies this constraint and these parameters fall onto the red line in Fig 2. For example, a cell that has a low expression of the envelope-producing enzyme (the left-most grey point in the landscape) will have a high growth rate but won't have enough envelope to cover its cytoplasm, thus making it a non-viable option. On the other hand, a cell that has a highly expressed envelope producer (the right-most grey point) will make too much membrane relative to its volume, thus causing the cell to wrinkle. However, a cell with intermediate values of ϕ_l will produce just enough envelope to cover its cytoplasm (the intermediate grey point), and the maximal growth rate is achieved by the adjustment of partitioning parameters along this line (black point in the landscape).

4.3 Explicit solution for maximal growth under saturation kinetics

Although obtaining the analytical solution for maximal steady-state growth rate is prohibitively difficult, one can obtain this property for a special case when all enzymes are saturated, such that chemical reactions obey first-order mass-action kinetics. Indeed, most *E. coli* enzymes have K_M which is lower than their substrate's concentrations [17]. This reduces the system of Eq 7a–7e to a system of linear equations that can be readily solved for the ϕ parameters:

$$\tilde{\lambda}\tilde{L} = \frac{k_l}{m_s}\tilde{P}_l - d_l\tilde{L} \tag{16a}$$

$$\tilde{\lambda}\tilde{P}_b = \frac{k_t}{m_p}\tilde{R}\phi_b - d_p\tilde{P}_b \tag{16b}$$

$$\tilde{\lambda}\tilde{P}_l = \frac{k_t}{m_p}\tilde{R}\phi_l - d_p\tilde{P}_l \tag{16c}$$

$$\tilde{\lambda}\tilde{R} = \frac{k_t}{m_r}\tilde{R}\phi_r \tag{16d}$$

The steady-state growth is obtained from Eq 11b by setting $K_M = 0$:

$$\tilde{\lambda} = m_p\left(\kappa_n\frac{\tilde{P}_b}{\tilde{V}} - \kappa_l\frac{\tilde{P}_l}{\tilde{V}} - d_p\left(\frac{\tilde{P}_b}{\tilde{V}} + \frac{\tilde{P}_l}{\tilde{V}}\right)\right) \tag{17a}$$

Substituting Eq 17a in Eq 16a–16d:

$$\tilde{L} = \frac{\kappa_l\phi_l\kappa_t\phi_r}{m_s(\kappa_n\phi_b - \kappa_l\phi_l - d_p(\phi_b + \phi_l))(\kappa_t\phi_r + d_l)}\tilde{V} \tag{18a}$$

$$\tilde{P}_b = \frac{\kappa_t\phi_r\phi_b}{m_r(\kappa_n - d_p)\phi_b + m_r(\kappa_l + d_p)\phi_l}\tilde{V} \tag{18b}$$

$$\tilde{P}_l = \frac{\kappa_t\phi_r\phi_l}{m_r(\kappa_n - d_p)\phi_b + m_r(\kappa_l + d_p)\phi_l}\tilde{V} \tag{18c}$$

$$\tilde{R} = \frac{\phi_r(\kappa_t\phi_r + d_p)}{m_r(\kappa_n - d_p)\phi_b + m_r(\kappa_l + d_p)\phi_l}\tilde{V} \tag{18d}$$

By substituting abundances in Eq 9 for Eq 18b–18d It immediately follows that the mass fractions of protein sectors are:

$$\Phi_B = \frac{\kappa_t\phi_b}{\kappa_t + d_p}, \quad \Phi_L = \frac{\kappa_t\phi_l}{\kappa_t + d_p}, \quad \Phi_R = \frac{\kappa_t\phi_r + d_p}{\kappa_t + d_p} \tag{19}$$

Note that substitution of (18a–18d) in (17a) yields $\tilde{\lambda} = \kappa_t\phi_r$, meaning that the growth rate is simply proportional to the fraction of ribosomes that are allocated to ribosome translation. The cell, however, cannot allocate the entirety of its ribosomes to this task because this would halt the production of other necessary protein components. Therefore, to find the optimal partitioning of ribosomes that maximizes the growth rate, one has to impose additional algebraic constraints. The first algebraic constraint reflects the fact that maximal growth is achieved when the building block influx matches the outflux. If influx is higher than the outflux, then buidling blocks unnecessarily accumulate in the cell, and if the outflux is higher, then the building block pool will become completely depleted and the chemical reactions will halt. Since B is being produced and consumed at matching rates, the building block pool does not grow over

time, and retrieve the constraint by setting the left-hand side of Eq (7a) to zero and $K_M = 0$. The second expression imposes a constraint on the amount of resources that have to be diverted into cell envelope to ensure a proper cell shape:

$$m_p \kappa_n \tilde{P}_b = m_p \kappa_l \tilde{P}_l + m_r \kappa_t \tilde{R} \tag{20a}$$

$$\frac{\tilde{L}/\beta}{\tilde{V}/\gamma} = \Pi \tag{20b}$$

To obtain optimal partitioning parameters ϕ that satisfy these constraints, we substitute $\tilde{P}_b$, $\tilde{P}_l$, and $\tilde{R}$ with solutions 18a–18d. While there are three ϕ parameters, there are only two degrees of freedom given that $\phi_b = 1 - \phi_r - \phi_l$. Thus, we have a system of two linear equations in two unknowns which we solve to obtain the optimal partitioning of the proteome such that the flux of resources is balanced between catabolic and anabolic processes, and the cell has a proper shape given its size:

$$\phi_r^{\text{opt}} = \frac{\kappa_l \kappa_t (\kappa_n - d_p) - d_l (\kappa_l + \kappa_n)(\kappa_t + d_p)\epsilon\Pi}{\kappa_l \kappa_t (\kappa_n + \kappa_t) + \kappa_t (\kappa_l + \kappa_n)(\kappa_t + d_p)\epsilon\Pi} \tag{21a}$$

$$\phi_l^{\text{opt}} = \frac{(\kappa_t + d_p)(\kappa_n d_l + \kappa_t (\kappa_n + d_l - d_p))\epsilon\Pi}{\kappa_l \kappa_t (\kappa_n + \kappa_t) + \kappa_t (\kappa_l + \kappa_n)(\kappa_t + d_p)\epsilon\Pi} \tag{21b}$$

Substituting ϕ_r^{opt} and ϕ_l^{opt} in Eq 18a–18d, one retrieves molecular abundances in terms of model parameters. Substitution into Eq 19, we obtain expressions for optimal proteome mass fractions. Finally, substitution in formula for $\tilde{\lambda}$ (Eq 17a), yields a maximal steady-state growth rate $\tilde{\lambda}_{\max}$:

$$\tilde{\lambda}_{\max} = \frac{\kappa_t (\kappa_n - d_p)}{\kappa_n + \kappa_t} \frac{1}{1 + \Theta}, \tag{22a}$$

$$\Theta = \frac{(\kappa_n + \kappa_l)(\kappa_t + d_p)(\kappa_n d_l + (\kappa_n + d_l - d_p)\kappa_t)\epsilon\Pi}{(\kappa_n + \kappa_t)(\kappa_l \kappa_t (\kappa_n - d_p) - d_l (\kappa_n + \kappa_l)(\kappa_t + d_p)\epsilon\Pi)} \tag{22b}$$

Where $\epsilon = m_s \beta / \gamma$ is the resource cost of the unit S/V, or the resource investment in a unit of surface area per unit of cytoplasmic volume. For example, if $\epsilon = 2$, then each added unit of the surface area requires twice as many amino acids relative to the added unit of cytoplasmic volume.

Intuitively, the first term in Eq 22a is the steady-state growth rate in the limit of no additional resource sink, and the second term represents a deviation from maximal achievable growth due to cell envelope production. The parameter Θ is the bioenergetic cost of cell envelope production, which depends not only on the actual amount of resources that go into this cellular feature ($\epsilon\Pi$) but also on the rates of all cellular processes. Previous theoretical developments represent a special case of Eq 22a. In the limit of no investment into cell envelope ($\Pi = 0$), expressions reduce to those reported in [18], and taking this further to the case without protein degradation ($d_p = 0$) yields the result in [16].

While the explicit solution for Θ appears complicated, some insight can be gained by looking at the special case when there is no degradation ($d_l = d_p = 0$):

$$\Theta = \frac{\frac{\kappa_l + \kappa_n}{\kappa_l}}{\frac{\kappa_n + \kappa_t}{\kappa_t}} \epsilon\Pi \tag{23}$$

The part $\epsilon\Pi$ corresponds to the resource cost of producing the entire envelope structure. Parameter ϵ can be interpreted as the envelope cost of the cell with $\Pi = 1 \mu m^{-1}$. This is because γ is the total number of amino acids per unit volume and

$m_s\beta$ is the total number of amino acid equivalents required to build the unit of the cell surface. Given that Π is the surface-to-cytoplasmic volume ratio, the whole term $\epsilon\Pi$ is the cost of producing surface relative to the whole amino acid budget of the cytoplasmic volume. The fractional term in expression 23 captures the cost of producing the enzyme machinery that builds the actual envelope. One cellular process supplies the building blocks at the per unit proteome κ_n, and two other cellular processes are competing for this common pool at the per unit proteome rates of κ_l and κ_t. For instance, if κ_l is decreased, the cell maximizes growth rate by overexpressing cell envelope-producing machinery in order to compensate for the low per per unit proteome rate, thus increasing the total costs of the envelope.

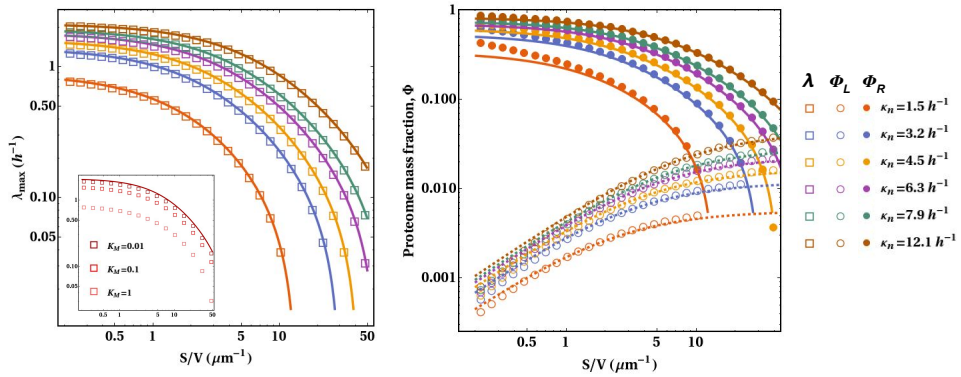

Figure 3. Maximal steady-state growth rate when all enzymes are saturated. Left panel: Growth rate scaling with the shape of the cell. Solid lines denote the analytical solution for $\tilde{\lambda}_{\max}$ (Eq 22a), while squares represent numerically maximized growth rate (Eq 14). Different lines correspond to different values of κ_n (reported in the legend on the right), while all other model parameters are identical. Right panel: Proteome composition scaling with the cell shape. Colored lines signify the optimal ribosomal (solid lines) and envelope-producer (dashed lines) mass fractions as a function of Π. Full and open circles denote the same quantities that maximize growth rate in the numerical optimization problem. Other parameters: $\kappa_l = 288$ h^{-1}, $\kappa_t = 2.59$ h^{-1}, $d_p = 0.14$ h^{-1}, $d_l = 0.17$ h^{-1}, $K_M = 0.001$, $\gamma = 9.75 \times 10^8$ aa/μm^3, $\beta = 2 \times 10^6$ lip/μm^2, $m_s = 314$ aa.

The analytical solution for $\tilde{\lambda}_{\max}$ in the saturation limit was cross-validated by comparison to the numerically maximized $\tilde{\lambda}$ with algebraic constraints of the finite ribosomal pool (c1) and geometric constraint on the cell shape (c2), as described in Section 4.2, and optimization was performed using Nelder-Mead algorithm. To ensure that the optimizer obtains the global maximum, we repeat the optimization 20 times for each set of model parameters. Each iteration starts by randomly seeding the points of the polytope. We take the highest $\tilde{\lambda}$ as the solution of the optimization problem. We generally find an excellent correspondence between analytics and numerics for both the growth rate and proteome composition (right panel in Fig 3). When enzymes in the model operate far from saturation limit (i.e., when $K_M \gg 0$), the analytics break down (see inset). Note that the large K_M has a similar effect on scaling as a reduction in nutrient quality of the media. Therefore, although the analytic solution neglects the presence of the aqueous phase – metabolites and associated water molecules – inside the cell, the incorporation of this property affects the intercept but not the overall scaling pattern.

4.4 Derivation of physiological scaling laws

The growth rate and proteome composition depend on the external environment in which the cell is reared. We obtain this dependence by modulating each of the three

rate constants κ_n, κ_t, and κ_l and deriving the response that the cell elicits. We obtain the changes in proteome composition when the nutrient quality of the media is varied in three steps. First, Eq 22a is solved for κ_n, which is now the function of the growth rate which is modulated by changing the amount of nutrients in the medium ($\tilde{\lambda}_N$). Second, this result is placed in Eq 21a and 21b to obtain optimal ribosomal allocation. Third and final, these expressions are plugged into formulae for the proteome mass fractions (Eq 10). After rearrangement:

$$\Phi_R = \frac{d_p}{d_p + \kappa_t} + \frac{1}{\kappa_t + d_p}\tilde{\lambda}_N \tag{24a}$$

$$\Phi_L = \frac{d_l \epsilon \Pi}{\kappa_l} + \frac{\epsilon \Pi}{\kappa_l}\tilde{\lambda}_N \tag{24b}$$

Growth rate can also be altered either by changing the concentration of translation inhibitor ($\tilde{\lambda}_T$). Thus, by solving Eq 22a for κ_t and following the same procedure outlined above, we get the optimal partitioning when the growth rate is varied by translation-inhibiting antibiotic:

$$\Phi_R = 1 - \frac{d_l(\kappa_n + \kappa_l)\epsilon\Pi}{(\kappa_n - d_p)\kappa_l} - \frac{\kappa_l + (\kappa_n + \kappa_l)\epsilon\Pi}{\kappa_l(\kappa_n - d_p)}\tilde{\lambda}_T \tag{25a}$$

$$\Phi_L = \frac{d_l \epsilon \Pi}{\kappa_l} + \frac{\epsilon \Pi}{\kappa_l}\tilde{\lambda}_T \tag{25b}$$

Finally, when the growth is perturbed via changes in concentration of envelope synthesis inhibitor ($\tilde{\lambda}_L$), we solve Eq 22a for κ_l and obtain:

$$\Phi_R = \frac{d_p}{d_p + \kappa_t} + \frac{1}{\kappa_t + d_p}\tilde{\lambda}_L \tag{26a}$$

$$\Phi_L = \frac{(\kappa_n - d_p)\kappa_t - d_l(\kappa_t + d_p)\epsilon\Pi}{\kappa_n(\kappa_t + d_p)} - \frac{\kappa_n + \kappa_t(\kappa_t + d_p)\epsilon\Pi}{\kappa_n(\kappa_t + d_p)}\tilde{\lambda}_L \tag{26b}$$

Thus, the mass fractions are a linear function of the growth rate, when the latter is perturbed either by changing the carbon source in the media ($\tilde{\lambda}_\mathrm{N}$), or by altering the concentration of antibiotics ($\tilde{\lambda}_\mathrm{L}$ or $\tilde{\lambda}_\mathrm{T}$). In the limit of no degradation, Eq 24a and 25a reduce to $\tilde{\lambda}_\mathrm{N}/\kappa_t$ and $1 - \tilde{\lambda}_\mathrm{T}/\kappa_n$, respectively, which are similar to the growth laws first reported in [16] with the exception that we did not account for minimal and maximal possible ribosomal content. One can intuitively interpret the intercept and the slope in Fig 4 the following way.

The intercept quantifies the basal expression of each proteomic component required to replenish constantly degraded components. As the exponential growth rate tends to zero, the ribosomal expression is completely abolished because protein synthesis is not needed in a non-dividing cell. However, when degradation is present, ribosomes are necessary even when the cell is not dividing in order for protein synthesis to balance the protein degradation. Hence, the non-zero intercept in Eq 24a. The same idea was recently proposed in [19], and presents the alternative to the idea that the excess ribosomes at slow growth represent the reserve that allows cells to quickly tune their growth to cycles of feast and famine [20]. A similar explanation holds for intercept in the Eq 24b. Absent lipid degradation, a non-dividing cell abolishes the expression of the envelope-producing enzyme because envelope synthesis only occurs during growth. However, if the envelope is being degraded, then even a non-dividing cell has to allocate some of its resources to an enzyme that will re-synthesize the degraded components. In general, the higher the degradation rates, the higher the intercept because more biosynthetic enzymes are needed to replenish a greater amount of degraded components.

The slope is the rate at which the mass fraction of a particular enzyme changes as the growth rate changes, and it is inversely related to the rate of that enzyme. The smaller the rate constant of an enzyme, the more limiting this step is in the process of growth. As the growth conditions improve, the cell has to allocate more of its resources to alleviate this limitation compared to the case when the enzyme is much faster. Take, for instance, changes in ribosomal content when growth is modulated by the nutrient quality of the media (Eq 24a): if κ_t is small, then modulating growth by rearing the cells with better carbon source requires a faster increase in ribosomal mass fraction relative to the case when protein synthesis rate is large. A similar interpretation holds for other growth laws: if the step is slow, then the cell has to allocate resources to it faster to meet the same increase in growth rate. There are generally three types of growth laws, illustrated in Fig 4: proteomic changes when nutrients are varied (Eq 24a, 24b), the translation rate is varied (Eq 25b, 25a), and the envelope synthesis rate is varied (Eq 26b, 26a).

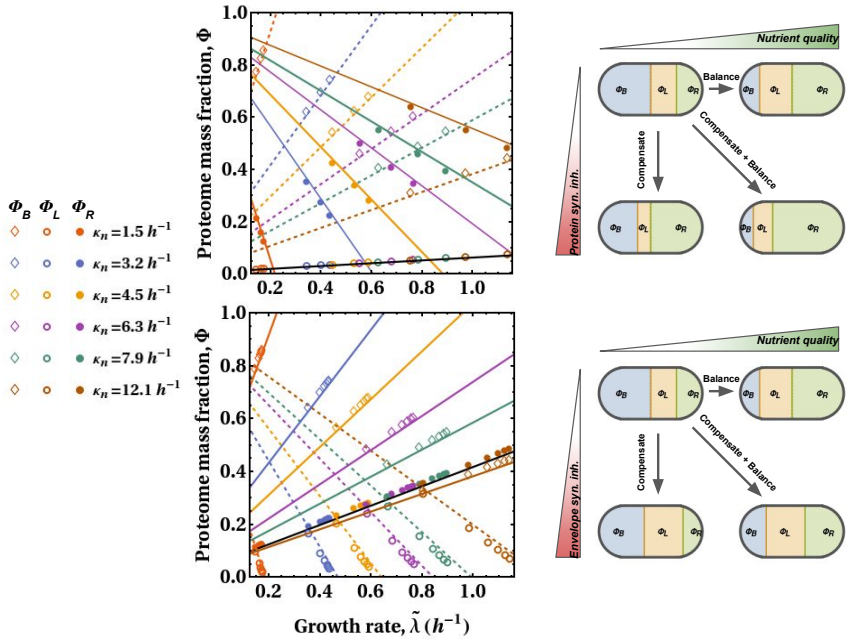

Figure 4. Changes in proteome composition across growth conditions. Panels on the left represent quantitative changes in all three proteome sectors, while the cartoons on the right give graphical interpretation of those changes. Upper row: Proteomic changes when nutrient quality and protein synthesis rate are modulated. Diamond and circle symbols represent numerically-computed optimal proteome mass fractions, and lines denote corresponding analytic solutions. Black line – Φ_L (Eq 25b); Colored solid lines – Φ_R (Eq 25a); Colored dashed lines – Φ_B (the rest of the proteome mass). Lower row: Proteomic changes when nutrient quality and envelope synthesis rate are modulated. Black line – Φ_R (Eq 26a); Colored dashed lines – Φ_L (Eq 26b); Colored solid lines – Φ_B (the rest of the proteome). Colors correspond to different κ_n values reported in the legend.

Consider optimal proteomic changes when cells are reared with translation inhibitor under poor nutrient conditions (e.g., blue points and lines in upper panel). Increasing the concentration of protein synthesis inhibitor increases the fraction of ribosomes that are inhibited, so the cell compensates for this inhibition by upregulating the ribosomes (note the negative scaling between Φ_R and $\tilde{\lambda}$). Given the finite proteome, this up-regulation is coupled to the down-regulation of envelope producer and building block producer. Now imagine that the cells were shifted to a richer media or a better carbon

source (e.g., brown points and lines in upper panel). Two kinds of changes occur. First, the increased nutrient influx is balanced by an increase in the outflux of building blocks into the biomass by upregulating the ribosomal mass fraction. Second, the same amount of envelope has to be produced in a shorter time. Given that κ_l is fixed, this new goal can only be reached by increasing the Φ_L. In short, rearing cells under better nutrient conditions leads to an increase in the expression of ribosomes and envelope producers.

An identical explanation holds when proteomic changes are induced by growing cells in the presence of envelope synthesis inhibitor (bottom panel). In this case, however, compensation for inhibition is achieved by overexpressing envelope producers (note the negative scaling between Φ_L and $\tilde{\lambda}_{\max}$), thus causing down-regulation of the two other proteome components. Similar to the previous up-shift experiment, cells reared under better nutrient conditions balance the increased nutrient influx by increasing the outflux of building blocks into the proteins (via up-regulation of ribosomes), and into the envelope (via up-regulation of envelope producer). Finally, note that there is a good agreement between numerically computed optimal proteomic mass fractions and analytical solutions under saturation kinetics (marker symbols and lines in Fig 4, respectively).

4.5 Data collection and normalization procedures

Growth rate and cell size data were obtained from the literature. Because most of the species were reared in nutrient-rich medium, often containing yeast extract, under optimal temperature, pH, and salinity, we assumed that the reported values correspond to the maximal growth that a species can attain. Given that different species are grown in disparate temperatures, we normalized growth rates to 20°C using Q_{10}-correction with Q_{10} coefficient of 2.5, to exclude the potential confounding effect of temperature on growth scaling. The linear dimensions of the cell were also obtained from the literature, and were used to calculate cell volume and surface. Each cell is classified into one of the three categories based on its shape: spheres, rods, and helices.

Because we are interested in whether variation in growth rate can be solely explained in terms of variation in cell size and shape, we work with chemoorganoheterotrophic species. This ensures that variation in growth is not caused by differences in the type of metabolism that species have. However, even chemoorganoheterotrophs may generate energy in a variety of ways. One of the biggest differences is between oxidative phosphorylation and substrate phosphorylation (i.e., fermentation), the latter having lower energetic content. Hence, one can argue that obligate fermenters ought to grow slower owing to a slower rate of energy extraction from nutrients. In our model, this would translate to an organism having a lower κ_n. If the growth rate of *E. coli* under anaerobic conditions is 63% of the growth rate under aerobic ones [21], one can assume that κ_n of anaerobes is also 63% of κ_n of aerobes. Formally, let f be the growth rate as the function of nutritional capacity κ_n (Eq 22a). Then the growth rate of an anaerobe ($\tilde{\lambda}^*_{\max}$) is going to be a function of the anaerobic nutritional capacity (κ^*_n). Let anaerobic κ_n be $\alpha = 0.63$ of aerobic κ_n:

$$\tilde{\lambda}^*_{\max} = f(\kappa^*_n) = f(\alpha\,\kappa_n) = f(\kappa_n)/a \tag{27}$$

where a is conversion factor that depends on the model parameters and is given by:

$$a = \frac{(\kappa_l(d_p-\kappa_n)\kappa_t + d_l\epsilon\Pi(\kappa_l+\kappa_n)(d_p+\kappa_t))(\kappa_l(\alpha\kappa_n+\kappa_t)+\epsilon\Pi(\kappa_l+\alpha\kappa_n)(d_p+\kappa_t))}{(\kappa_l(\kappa_n+\kappa_t)+\epsilon\Pi(\kappa_l+\kappa_n)(d_p+\kappa_t))(\kappa_l(d_p-\alpha\kappa_n)\kappa_t + d_l\epsilon\Pi(\kappa_l+\alpha\kappa_n)(d_p+\kappa_t)))} \tag{28}$$

Therefore, to compare aerobes and anaerobes on the same footing, the growth rate

of anaerobes obtained from the literature should be multiplied by the factor of a; The term a can be biologically interpreted as the fold-increase in growth rate that anaerobes would experience if they were to switch to oxidative phosphorylation as the means of generating energy. More precisely, a is obtained by taking the ratio of aerobic (with nutritional capacity being κ_n) to anaerobic growth rate (with nutritional capacity being $0.63\kappa_n$). Each organism was determined whether it is an obligate fermenter by investigating its citrate cycle in KEGG PATHWAY database. A bacterium is deemed a fermenter if it does not have a total of 12 steps listed in the database, starting and ending with oxaloacetate. For species without biochemical annotation, we extracted information on their energy metabolism from the literature. Lastly, when we could not find data on the type of energy-generation pathways, we simply assumed that those species respire.

4.6 Proteomic data analysis

We use proteomic data to 1) infer the rate constants via physiological scaling laws, and 2) test the predictions on the scaling of the proteomic composition with the surface-to-volume ratio of the cell.

As outlined in the section 4.4, one needs to co-measure the growth rate across different conditions and the proteome mass fraction allocated to three sectors to infer capacities from the slope of regressions (Eq 24a–26b). To this end, we use two types of studies: Those that directly quantified absolute abundances of each protein [22–27], and those that indirectly measured ribosomal mass fraction from the total RNA-to-total protein ratio [16, 28–31]. We use [25] as a primary source because it has the highest coverage of the *E. coli* proteome (in excess of 90% of the proteome is detected). To estimate mass fractions belonging to nutrient-processing, lipid-producing, or ribosomal protein, we classified the total mass (in units fg/cell) of each reported protein in one of the three possible groups based on its designation in KEGG BRITE database [32]. If a protein's assigned function in the BRITE database had keywords "fatty acid biosynthesis", "lipopolysaccharide biosynthesis", "peptidoglycan biosynthesis", we classified such protein as contributing to Φ_L: If protein's designation contained the keyword "ribosomal protein", we classified it as contributing to Φ_R. All other proteins were grouped in the Φ_B fraction. After binning, we calculate each mass fraction by dividing the total mass of proteins per cell in that class by the total mass of proteome per cell. Ribosomal mass fractions from indirect sources are obtained by multiplying the total RNA/total protein ratio with a conversion constant as reported in [31]. Some studies reported relative protein abundances, and we converted these into relative mass fractions by multiplying each entry with the molecular mass of the given protein.

To test the prediction on scaling of proteome composition with S/V, we collected quantified proteomes of bacterial species with S/V ranging from less than 5 μm^{-1} in *Lactococcus lactis* to more than 50 μm^{-1} in *Spiroplasma poulsonii*. In addition to proteomic studies, we also collected data for ribosome abundance and converted it to proteome mass fractions in the following way. First, the total mass of ribosomes was calculated as:

$$M_{\text{ribo}} = \frac{N_{\text{ribo}} \times m_r \times m_{aa}}{N_{\text{Avg}}} \tag{29}$$

where $m_{aa} = 110$ g/mol is the molar mass of an average amino acid, $N_{\text{Avg}} = 6.022 \times 10^{23}$ molecules/mol, $m_r = 7336$ amino acids is the length of the protein component of a single ribosome (assumed to be constant across bacteria) and N_{ribo} is the ribosome abundance per cell. To calculate the mass of the total proteome, we first use the empirical scaling law (Eq 30a from prokaryote data in Fig 7.1 reported in

Chapter 7 of [33]) that allows calculation of dry cell weight (in grams) from the cell volume V (in units of μm^3). Then, knowing that the proteome of most bacterial cells accounts for 54% of the total dry weight (Fig 7.3 reported in Chapter 7 of [33]), the ribosomal mass fraction is calculated in three steps from cell volume:

$$M_{\text{cell}} = 4.95 \times 10^{-13} V_{\text{cell}}^{0.928} \tag{30a}$$

$$M_{\text{prot}} = 0.54 M_{\text{cell}} \tag{30b}$$

$$\Phi_R = M_{\text{ribo}}/M_{\text{prot}} \tag{30c}$$

For cross-species data, we used proteomes from PaxDB [34], a recent compendium of proteomes across 100 species [35], a collection of ribosomal abundances from [36] and [37], and a number of additional quantitative proteomic studies not reported in above-mentioned databases: *Borrelia burgdorferi* [38], *Treponema pallidum* [39], *Polynucleobacter asymbioticus* [40], *Mesoplasma florum* [41], and *Spiroplasma poulsonii* [42]. In total, we have quantified Φ_R and Φ_L for 40 and 30 bacterial species, respectively.

4.7 Parameterization of the model

Eq 22a has six parameters (κ_n, κ_l, κ_t, ϵ, d_l, d_p) and one variable (Π). Our goal was to constrain the rates of chemical reactions to the values occurring in *Escherichia coli*, and then ask how the growth rate would scale if the variable Π was altered. That is, we are interested in understanding how the growth rate would scale if all bacterial species were biochemically identical to an *E. coli* cell, and only differed in cell size.

We estimated the resource cost of the envelope in a cell with a unit S/V, ϵ, by assuming that resources investment is equal to the mass M of a structure. Given that our model consists of the whole envelope (lipids, peptides, and sugar attachments) with mass M_{env} and proteins of mass M_{prot}, and that envelope and proteins account for roughly 30% and 55% of total cell mass M_T [43], and $\Pi = 5$ (Table 2):

$$\epsilon\Pi = \frac{m_s \beta S}{\gamma V} = \frac{M_{\text{env}}}{M_{\text{prot}}} = \frac{M_{\text{env}}/M_T}{M_{\text{prot}}/M_T} = \frac{0.30}{0.55} = 0.55 \implies \epsilon = 0.11 \mu m \tag{31}$$

This is a crude estimate because it does not include the direct costs of envelope production and proteome component (i.e., resources needed to convert one chemical compound into another one), but these usually account for a small fraction of the total costs [44], and small changes in ϵ do not affect our conclusion.

The degradation rates are estimated separately from the literature. We assumed that the ribosomal degradation rate is zero, and we justify this assertion by three observations. First, ribosomal rRNA is remarkably stable in the exponential and stationary phases, and the degradation occurs only when the culture transitions between these two growth stages [15, 45]. Second, almost all ribosomal proteins have degradation rates close to zero [46]. Third, ribosomal proteins are mutually exchangeable when damaged, meaning that replacement is favored over degradation and re-synthesis [47]. The protein degradation rate was set to 0.05 per hour [48]. We also tried out estimates for various protein classes [49] but found little variation in the scaling pattern (Fig S1.1). Envelope degradation rate d_l is difficult to estimate exactly due to the diverse components that make up this structure, but was set to 1 per hour [50]. This study reports a hypermetric scaling of the peptidoglycan turnover and the growth rate ($d_l = 0.7 \times (\text{Log}[2]/\tau)^{1.38}$). Assuming that an *E. coli* cell divides in 30 minutes, we have $d_l \simeq 1$. Unfortunately, this is a very crude estimate based on the peptidoglycan degradation rate in *B. subtilis*, which constitutes a large part of its envelope. Peptidoglycan degradation rates are somewhat lower in *E. coli* [51] but on the

other hand, that species has an outer membrane containing polysaccharides, and we do not know how fast this component is degraded.

The rate constants of chemical reactions (κ_t, κ_l, κ_n) were inferred from *Escherichia coli* proteomic data across different growth conditions, leveraging growth laws derived in section 4.4. We ignore the intercept of regressions (Eq 24a–26b) and infer parameters strictly from the slopes of these relationships. The slope of the regression of the growth rate on the mass fraction of ribosomal proteins allows one to compute κ_t from Eq 24a. Similarly, the value of κ_l can be computed from the slope of the regression of the growth rate on the mass fraction of envelope-producing enzymes and Eq 24b. Our data for growth rates across different bacterial species is normalized to 20°C using Q_{10} correction with the coefficient of 2.5. Hence, we also normalized the inferred rate constants using the same method. Estimates of rate constants for *E. coli* are reported in Table 2.

Symbol	Definition	Value	Units	Source
κ_n(M63+gly)	Nutritional capacity in M63 with glycerol	0.476 ± 0.026	h^{-1}	[16], ∘
κ_n(M63+glc)	Nutritional capacity in M63 with glucose	0.979 ± 0.112	h^{-1}	[16], ∘
κ_n(cAA+gly)	Nutritional capacity in cAA and glycerol	1.34 ± 0.211	h^{-1}	[16], ∘
κ_n(cAA+glc)	Nutritional capacity in cAA and glucose	1.84 ± 0.401	h^{-1}	[16], ∘
κ_n(RDM+gly)	Nutritional capacity in RDM with glycerol	2.285 ± 0.619	h^{-1}	[16], ∘
κ_n(RDM+glc)	Nutritional capacity in RDM with glucose	3.360 ± 1.344	h^{-1}	[16], ∘
κ_l	The envelope synthesis rate	39.603 ± 8.288	h^{-1}	[25], ∘
κ_t	The protein synthesis rate	2.579 ± 0.175	h^{-1}	[25], ∘
ϵ	The cost of cell per unit of S/V	0.11	μm	∗
Π	Surface-to-volume ratio	5	μm^{-1}	†
d_p	Protein degradation rate	0.011	h^{-1}	[48]
d_l	Cell envelope degradation rate	0.211	h^{-1}	[50]

Table 2. Parameterization of the model. Standard errors were obtained using error propagation (see Section S1.2 in S1 Appendix). † Surface-to-volume ratio represents the average across different dimensions of *E. coli* in our dataset. ∗ Calculated in the text. All rate parameters are Q_{10}-corrected as described in text. ∘ Rate constants estimated from the noted study.

Because κ_n captures both the intrinsic efficiency of metabolism to convert nutrients into building blocks and the nutrient state of the external environment, there will be one κ_n for every medium that *E. coli* is reared in. Intuitively, one would expect minimal media to have lower κ_n than rich media, as the latter contains better quality nutrients and thus leads to more amino acids being generated per unit time per molecule P_b. The parameter κ_n is calculated from the slopes of ribosomal mass fraction across growth conditions with varying concentrations of translation inhibitor, after plugging in values of κ_l, d_p, and $\epsilon\Pi$ in Eq 25a. The values are computed for the following media: M63 with glycerol ($\kappa_n = 0.476 \pm 0.026\, h^{-1}$), M63 with glucose ($\kappa_n = 0.979 \pm 0.112\, h^{-1}$), casamino acids with glycerol ($\kappa_n = 1.34 \pm 0.211\, h^{-1}$), casamino acids with glucose ($\kappa_n = 1.84 \pm 0.401\, h^{-1}$), rich defined media with glycerol ($\kappa_n = 2.285 \pm 0.619\, h^{-1}$), and rich defined media with glucose ($\kappa_n = 3.360 \pm 1.344\, h^{-1}$).

One can intuitively explain the estimated rate constants by using logic outlined in [52]. It takes roughly 10^9 glucose molecules to produce the entire carbon skeleton of an *E. coli* cell. Given that each amino acid is worth about as much ATP as a single glucose molecule, one could say that the cell requires 10^9 amino acid-equivalents for its

construction. Therefore, with roughly 3×10^6 copies of metabolic proteins in the cell, each building block producer making $k_n = \kappa_n m_p = 2 \times 325$ amino acids per hour, it will take ~ 30 minutes for the metabolic proteins to generate enough amino acids to replace the cell. This is also the experimentally measured cell division time of an *E. coli* cell grown under favorable conditions. For the protein synthesis rate κ_t, note that each ribosome converts $\kappa_t \times m_r$ amino acids into proteins (where the length of the ribosome m_r is 7336 amino acids). This means that the inferred translation rate k_t in our model is about 25 amino acids per ribosome per second, which is close (albeit slightly larger) to empirical values for an *E. coli* cell [53]. Length of metabolic protein was taken to be the median length of an *E. coli* protein (325 aa from [54]), and ribosome length is set to the total number of amino acids in ribosomal proteins (7336 aa from [29]).

5 Results

The model of the cell developed in section 4 purports to explain two phenomena: (1) the optimal partitioning of the proteome when the same bacterial species is reared under different growth conditions; and (2) the scaling of the growth rate and proteome composition across bacteria of different shapes and sizes. Both sets of predictions are tested in the ensuing sections. Applying the developed theory to data on growth and proteome composition of the model organism *Escherichia coli*, we first show that our model yields a good qualitative description of physiological responses to changes in the environment. We then use this correspondence between theory and data to estimate the parameter values that yield a good quantitative fit as well (section 5.1).

By parameterizing our model with values obtained from *E. coli*, one can address how the growth rate is supposed to scale with the shape and size of the cell (sections 5.2 and 5.3). This implicitly assumes that all bacterial species are biochemically identical to *E. coli*, such that the rates of all chemical reactions are the same. Lastly, in an attempt to simplify the expectation and remove dependence on parameters that are difficult to estimate, we look at the special case when degradation is absent, and the envelope is costly enough such that $\tilde{\lambda}_{\max}$ is inversely proportional to Π. This expectation is tested in section 5.4.

5.1 Proteome reallocation across growth conditions

The internal homeostatic mechanisms allow the cell to allocate the proteome to different cellular tasks such that the growth rate is maximized. While *E. coli* cells achieve this via alarmone ppGpp [55], our model is agnostic of the exact mechanism and simply assumes that such resource-tuning strategy exists. One can express the proteome mass fraction of a particular component as a linear function of growth rate when the latter is altered via changes in growth conditions (derived in section 4.4). Given that there are three cellular processes in our representation of the cell (building block production, cell envelope synthesis, and protein synthesis), we ask how the proteome composition changes when each process is perturbed. More precisely, we are interested in how the proteome composition changes as the growth rate is modulated by changing the values of κ_n, κ_l, or κ_t.

The first expectation is that the ribosomal mass fraction of the proteome Φ_R scales with the growth rate across different nutrient conditions $\tilde{\lambda}_N$ as:

$$\Phi_R = \Phi_R^{\{N,\min\}} + \frac{1}{\kappa_t + d_p}\tilde{\lambda}_N \tag{32}$$

The cell re-balances building block supply and biosynthesis demand by allocating proteome to the limiting process, which is, in this case, protein synthesis. The

proteomic data qualitatively corroborate this expectation (Fig 5, upper left panel). Data from different studies appear to show a small amount of variation, but the overall trend is strong. The second expectation is that the envelope-producer fraction increases with the growth rate under nutrient perturbation:

$$\Phi_L = \Phi_L^{\{N,\min\}} + \frac{\epsilon\Pi}{\kappa_l}\tilde{\lambda}_N \tag{33}$$

The slope captures the fact that faster growth implies that a cell has to produce its envelope faster. Given that each envelope-producer operates at a constant rate, the only way this constraint can be met is by allocating more enzymes to this task. While the proteomic data on this proteome component is less accurate compared to ribosomal proteins (probably owing to low abundance of envelope-producers), the predicted positive scaling still occurs in the two most comprehensive proteomic studies (orange and blue points in left panels of Fig 5).

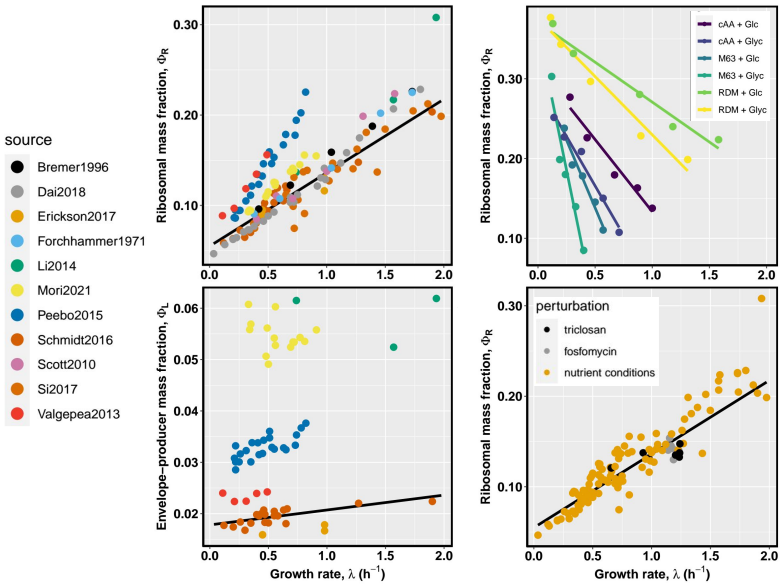

Figure 5. Comparison of empirical proteome composition changes with theoretical expectations. On the left, upper and lower panels represent Φ_R and Φ_L across various studies reported in the legend. Upper right panel is data on Φ_R under chloramphenicol treatment obtained from [16]. Each line corresponds to a different media as outlined in the corner legend. Lower right panel contrasts changes in Φ_R across nutrient conditions (orange points) with Φ_R changes under envelope producer inhibitors (black and grey dots), with data from [30]. Lines signify the ordinary least squares fit to the data from [16] (upper right panel), and [25] all other panels.

The third expectation is concerned with scaling of ribosomal mass fraction under translation inhibition:

$$\Phi_R = \Phi_R^{\{T,\max\}} - \frac{\kappa_l + (\kappa_n + \kappa_l)\epsilon\Pi}{\kappa_l(\kappa_n - d_p)}\tilde{\lambda}_T \tag{34}$$

As the concentration of a translation-inhibiting drug is increased, so is the fraction of inhibited ribosomes. The cell compensates for this inhibition by overexpressing ribosomes, which causes a negative scaling between Φ_R and $\tilde{\lambda}_T$. We see a good fit to data (upper right panel in Fig 5), and similar relationships can be obtained by using the alternative source of data from [30] (data not shown).

The fourth prediction is that the response of Φ_R to nutrient quality perturbation is the same as the response to envelope-synthesis inhibition. This follows from the fact that the relationship

$$\Phi_R = \Phi_R^{\{L,\min\}} + \frac{1}{\kappa_t + d_p}\tilde{\lambda}_L \tag{35}$$

is identical to Eq 32. Thus, the prediction is that points from two types of perturbation experiments ought to fall onto the same line. Indeed, the treatment of *E. coli* culture with fosfomycin (peptidoglycan synthesis inhibitor) and triclosan (fatty acid synthesis inhibitor) reduces both the growth rate and the ribosomal mass fraction (lower-right panel in Fig 5). One can intuitively understand this behavior by noting that the cell responds to the envelope-synthesis inhibition by overexpressing envelope-producer to compensate for the poisoning of the enzymes, which reduces the resources available for ribosomal proteins thus causing the reduction in their expression.

5.2 The growth-impeding effects of cell envelope

We are ultimately interested in how the growth rate ($\tilde{\lambda}_{\max}$) scales with the cell size across bacterial species with different shapes. We start by investigating the cell without degradation processes. In the limit of saturation kinetics, our model yields a closed-form solution for the maximal attainable growth given a particular cell size and shape. For simplicity, let us consider a case when degradation is absent ($d_p = 0$, $d_l = 0$), so that the Eq 22a simplifies to:

$$\tilde{\lambda}_{\max} = \frac{\kappa_t \kappa_n}{\kappa_n + \kappa_t}\frac{1}{1+\Theta}, \quad \Theta = \frac{1/\kappa_n + 1/\kappa_l}{1/\kappa_n + 1/\kappa_t}\epsilon\Pi \tag{36a}$$

This expression has a simple, intuitive interpretation. The first term corresponds to Monod's law [56], meaning that the growth rate is a hyperbolic function of the nutrient quality (κ_n) and protein synthesis rate (κ_t). As more nutrients are supplied (i.e., κ_n increased), the production of building blocks becomes less limiting than the downstream step of incorporating those components into the biomass. In the limit of infinite resource concentration ($\kappa_n \to \infty$), the rate of growth will equal the rate at which those resources are converted into the proteins by ribosomes ($\tilde{\lambda}_{\max} = \kappa_t$). Likewise, in the limit of infinitely fast translation ($\kappa_t \to \infty$) that instantaneously convert amino acids into biomass, the rate of growth will be identical to the rate at which nutrients are assimilated and supplied to ribosomes ($\tilde{\lambda}_{\max} = \kappa_n$).

The second term in Eq 36a captures the impediment of the growth rate caused by the production of the cell envelope, which acts as a resource sink. Here, two important insights emerge. First, the impediment term is a monotonically increasing function of Π, meaning that either smaller or elongated cells ought to have low growth rates. For example, two cells with identical cell volumes should have different growth rates if one is round and the other is elongated. Second, the growth rate impediment depends not only on the bioenergetic costs of the envelope – set by $\epsilon\Pi$ term but also on the rates of cellular reactions. The term Θ can be intuitively interpreted as the time required to produce a unit of surface area relative to the time needed for the production of a unit of proteome. Because κ is the rate of the biochemical step, $1/\kappa$ is the mean time to completion of that step. If the rate of the envelope-producing enzyme is low such that the time to build an envelope is large, then the cell has to reallocate more of its proteome toward envelope producers.

Therefore, the envelope has two kinds of growth-impeding effects, or costs: (1) structure-related, via allocation of resources to the envelope itself; and (2) machinery-related, via allocation of resources to enzymes that build the envelope. Both

of these features divert resources away from processes that replicate the cell and thus impede the growth. The costs are computed by comparing the growth rates with and without a particular feature. This is simply the definition of a selection coefficient in reproduction occurs in the continuous-time [57]. For instance, the total growth impediment caused by envelope is:

$$s_{\mathrm{T}} = \lim_{\Pi \to 0} \tilde{\lambda}_{\max} - \tilde{\lambda}_{\max} = \frac{\kappa_n \kappa_t}{\kappa_n + \kappa_t} \frac{\Theta}{1+\Theta} \tag{37}$$

The structure-related cost is obtained by comparing growth rates of the cell lacking both structure and machinery ($\Pi \to 0$) to the one with structure and without machinery ($\kappa_l \to \infty$). Intuitively, the proteome fraction allocated to envelope producers tends to zero when the envelope synthesis rate is infinitely fast. Likewise, the machinery-related cost is retrieved by contrasting the growth rates of the cell with structure and without machinery ($\kappa_l \to \infty$), with the one that has both of these components (wild type growth rate). In the first case, the sole cause of the difference in growth rates is the presence of the structure, while in the second case, it is the presence of machinery:

$$s_{\mathrm{S}} = \lim_{\Pi \to 0} \tilde{\lambda}_{\max} - \lim_{\kappa_l \to \infty} \tilde{\lambda}_{\max} = \frac{\kappa_n \kappa_t}{\kappa_n + \kappa_t} \frac{\Theta}{1+\Theta+\kappa_n/\kappa_l} \tag{38}$$

$$s_{\mathrm{M}} = \lim_{\kappa_l \to \infty} \tilde{\lambda}_{\max} - \tilde{\lambda}_{\max} = \frac{\kappa_n \kappa_t}{\kappa_n + \kappa_t} \frac{\Theta}{1+\Theta} \frac{\kappa_n/\kappa_l}{1+\Theta+\kappa_n/\kappa_l} \tag{39}$$

Three interesting consequences emerge from expressions 37–39. First, the machinery costs decrease as envelope synthesis rate κ_l increases (because $\partial s_{\mathrm{M}}/\partial \kappa_l < 0$). The faster the envelope producer is, the fewer proteins are needed to achieve the same net rate. Second, the structure cost depends not only on the resources needed for the envelope as a structure ($\epsilon\Pi$) but also on nutrient processing and protein synthesis rates. Envelope synthesis competes with protein synthesis for the common building block pool, so even instantaneous production of the envelope diverts resources away from protein synthesis and thus slows down the growth. Third, the total cost increases with Π (because $\partial s_{\mathrm{T}}/\partial \Pi > 0$), and structure costs eventually come to dominate the total costs while machinery cost approaches zero (as $\Pi \to \infty$, $s_{\mathrm{S}} \to s_{\mathrm{T}}$ and $s_{\mathrm{M}} \to 0$).

The inclusion of protein degradation in the model does not change the previously-reached conclusions significantly. Letting $d_l = 0$ and $d_p > 0$, in Eq 17a yields:

$$\tilde{\lambda}_{\max} = \frac{\kappa_t(\kappa_n - d_p)}{\kappa_n + \kappa_t} \frac{1}{1+\Theta}, \quad \Theta = \frac{\frac{\kappa_n+\kappa_l}{\kappa_l}}{\frac{\kappa_n+\kappa_t}{\kappa_t+d_p}} \epsilon\Pi \tag{40}$$

Relative to the no-degradation model, the growth rate is further impaired in two ways (solid and dashed lines in Fig 6). First, the maximal attainable growth rate in the absence of envelope is lowered because resources are now partly dissipated via protein degradation (purple lines). Second, the overall growth scaling is reduced because nutrient- and envelope-producing enzymes have to be overexpressed relative to the base scenario without degradation (orange lines) to compensate for lowered flux caused by the continual degradation of machinery. Finally, note that $\kappa_n > d_p$ for the growth rate to be positive. If this condition is not met, the cell dissipates resources faster than it assimilates them, ultimately leading to complete destruction. Despite these quantitative differences, one still expects the growth rate to eventually approach inverse scaling with Π, as in the no-degradation case.

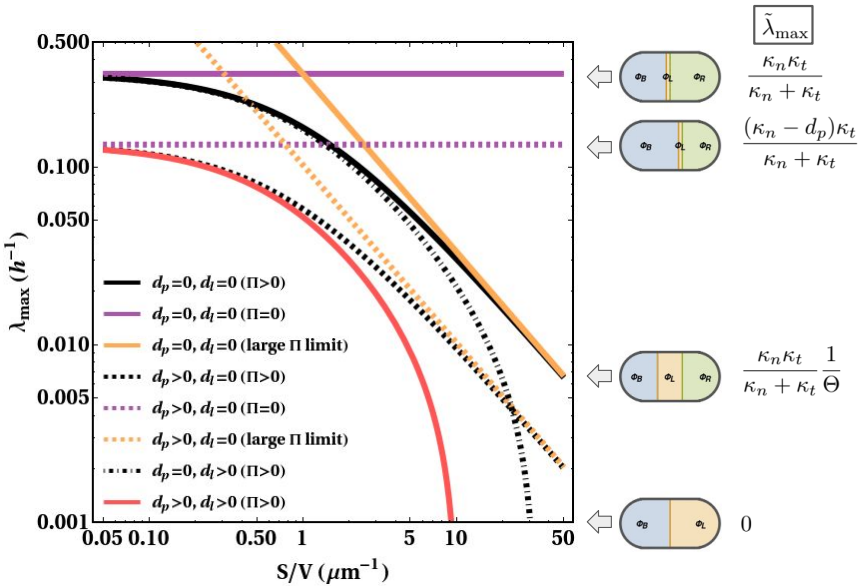

Figure 6. Analytics for the growth rate scaling. The general model is denoted with red dot-dashed curve. Special cases with absent or only one type of degradation are represented by black curves (see inset legend for details). Orange (purple) lines signify the growth rate scaling in the limit of large (small) S/V: Solid lines correspond to the case without protein degradation, while the dashed lines represent the case without the envelope degradation. Side cartoon outlines proteome composition and respective growth rates for solid purple, dashed purple, solid orange, and red solid lines. Parameters: $\kappa_l = \kappa_t = \epsilon = 1$, $\kappa_n = 0.5$, $d_l = 0.01$, $d_p = 0.3$.

In the case of only envelope degradation ($d_l > 0$, $d_p = 0$), the expected scaling changes fundamentally (black dot-dashed line in Fig 6):

$$\tilde{\lambda}_{\max} = \frac{\kappa_t\kappa_n}{\kappa_n+\kappa_t}\frac{1}{1+\Theta}, \quad \Theta = \frac{(\kappa_n+\kappa_l)(\kappa_n\kappa_t + d_l(\kappa_n+\kappa_t))\epsilon\Pi}{(\kappa_n+\kappa_t)(\kappa_n\kappa_l - d_l(\kappa_n+\kappa_l)\epsilon\Pi)} \tag{41}$$

Two consequences are immediately clear. First, the asymptotic growth as cells become large ($\Pi \to 0$) is the same as in the no-degradation model. An infinitely large cell (i.e., $\Pi \to 0$) still has to re-cycle proteins, but it does not have to re-cycle envelope because S/V approaches zero, implying that the asymptotic growth rate is larger in the case of an envelope- than in protein degradation case. Second, for every piece of an envelope that is added, envelope producers have to be overexpressed relative to the no-degradation case to meet an ever-increasing degradation demand. More formally from Eq 41, $\partial\Theta/\partial\Pi > 0$ (i.e., costs increase with Π) and $\partial^2\Theta/\partial\Pi^2 > 0$, (i.e., costs increase increasingly fast). This behavior ultimately leads to a critical surface-to-volume ratio Π_{crit} where the entire proteome is devoted to building, and re-cycling envelope and no resources are left for ribosomes. We find this point by setting $\tilde{\lambda}_{\max} = 0$ in Eq 41 and solving for Π:

$$\Pi_{\text{crit}} = \frac{\kappa_l\kappa_n}{(\kappa_l+\kappa_n)d_l\epsilon} \tag{42}$$

To prove that our interpretation is correct, set $d_p = 0$ in the expression for optimal fraction of ribosomes allocated to ribosome translation (Eq 21a), then substitute Π_{crit}, and lastly simplify to zero. The value of Π_{crit} increases with nutrient quality κ_n and envelope synthesis rate κ_l because faster rates mean that fewer resources have to be allocated to the respective proteome component to achieve the same total flux.

Moreover, it decreases with envelope degradation rate d_l and costs ϵ as this requires heavier processing and synthesis machinery investment, thus shifting the critical point to smaller cells. Finally, allowing for both envelope and lipid degradation (Eq 22a) only exacerbates the envelope burden described here (red solid line).

5.3 Scaling behavior in a fully specified model

The theory is tested by comparing the expected to the observed scaling relationships using growth, cell shape, and proteomic data across heterotrophic bacteria. This requires specifying model parameters. Because we are dealing with a somewhat metabolically homogeneous set of species (i.e., all are heterotrophs), we assume that all the cellular reaction rates are identical, and the only feature that varies across bacteria is Π. Furthermore, most of the maximal growth rates in our dataset were measured in the media containing yeast extract, so we assume that κ_n is fixed across species as well. The prerequisite for these assumptions to hold is that bacteria's type of metabolism does not correlate with cell size and shape. That is, the variation in metabolism and external environment might affect dispersion around the scaling but does not affect the scaling itself.

The model is determined by six parameters: rate constants of three types of enzymes (κ_n, κ_l, κ_t), degradation rates (d_p, d_l), and resource cost of the unit S/V (ϵ). Degradation rates of lipids, peptidoglycan, and proteins are set to values determined in pulse-chase experiments obtained from the literature. We take the energetic cost of the envelope, ϵ, from our earlier calculation (Eq 31). Rate constants are not only related to turnover numbers of respective enzyme classes but also depend on other cellular properties that are not explicitly accounted for in the model, such as the nutrient status of the environment, the number of enzymes belonging to a particular sector, the topology of a particular biochemical pathway, the amount of nucleic acid in the cell, and so on. We resolve this problem by estimating these values from the empirical relationships we derived in section 5.1.

The protein synthesis rate, κ_t, can be estimated from the slope of Eq 32 if the degradation rate d_p is known. The envelope synthesis rate, κ_l, is retrieved from the slope in Eq 35, if the envelope cost $\epsilon\Pi$ is known. And finally, the nutrient processing rate κ_n is inferred from slopes of changes in Φ_R as the growth rate is perturbed by the addition of translation-inhibiting drug chloramphenicol (Eq 34), after plugging in all of the previously-estimated parameters. Because κ_n also depends on the nutrient quality, the values were inferred for various media ranging from nutrient-depleted to nutrient-replete. All rate constants were estimated by fitting physiological scaling relationships to *E. coli* proteomic data. We first classified each protein into one of the three sectors based on its function (for details, see section 4.6), and then pooled their individual masses to compute the total mass fractions Φ_R and Φ_L. The surface-to-volume ratio was calculated from the linear dimensions of the cell. Cytoplasmic volume was calculated by subtracting $2\times$ thickness of the bacterial envelope, where the latter was taken to be 30nm [58].

First, curves parameterized with κ_n from minimal (grey) to rich media (dark blue) capture most of the variation in growth rate and proteome composition data (Fig 7). Hence, in principle, one could attribute the scatter along the y-axis to differences in efficiencies in which bacteria take up and convert the nutrients into building blocks. Second, the analytic solutions qualitatively explain the absence of a simple -1 slope that one would intuitively expect in the limit of very large $\epsilon\Pi$. For small S/V, the envelope mainly burdens the cell by production, whereas for large S/V region these costs rapidly increase due to additional envelope degradation processes. Third, Φ_L is expected to have a slightly positive and Φ_R a negative scaling. The prediction is not unambiguously corroborated by the data. A large amount of variation in proteomic

data around the expectation might reflect growth and temperature differences which we do not know how to correct for. Also, Φ_L is a low abundance proteome sector as it accounts for no more than a few percent of total proteome mass. These two effects might jointly make the data unreliable.

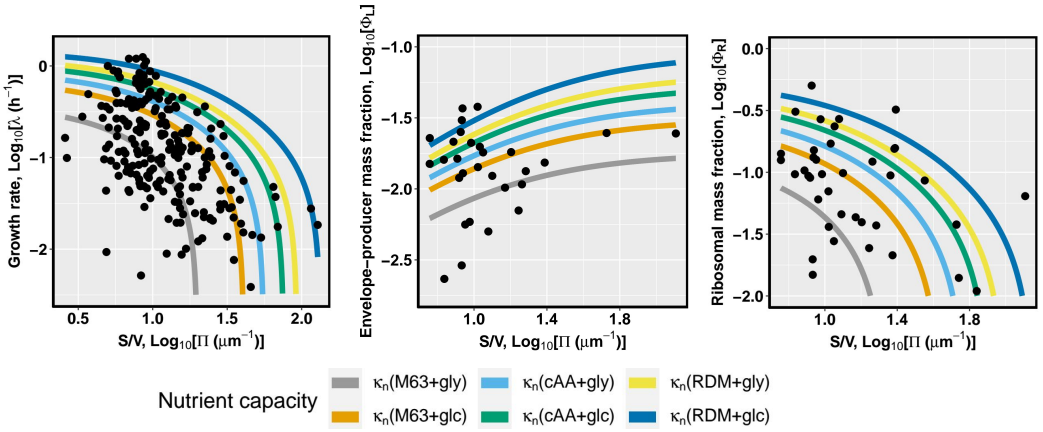

Figure 7. The comparison of expected and observed scaling relationships. Upper row: scaling of growth rate with S/V; Colored curves denote expected theoretical scaling given by Eq 36a and parameterized with κ_n estimated from various media. Lower row: scaling of ribosomal and envelope-producer mass fractions. Colored curves are theoretical expectation (Eq 22a) parameterized using values in Table 2.

A large variation in growth rate data across bacteria might also stem from variation in morphology (e.g., motility structures), metabolism (e.g., aerobic or anaerobic capabilities), or environment (e.g., from oceanic sediments to digestive tract of mammals). Ideally, one would want to compare the growth rates across species with the same type of metabolism and the same environment but differ in shape or size. This way, all of the morphological, metabolic, and environmental factors might be controlled, and the entirety of variation in growth can be attributed to variation in size. To this end, we analyzed the scaling of the growth rate with S/V in Lenski's long-term evolution experiment in *Escherichia coli* and a more recent experimental evolution of *Mycoplasma mycoides*.

First, we used S/V in combination with the relative fitness data for four LTEE time points spanning 50,000 generations [59]. Given that the relative fitness is the ratio of Malthusian parameters of evolved to ancestral line determined from the head-to-head competition (see [60]), we converted this metric to the growth rate by multiplying it with the growth rate of the ancestral strain (see $V_{\max}$ in [61]). However, because the competition experiment lasts for one day, competing strains enter into stationary phase after a few hours, so the competitive advantage of the evolved line might reflect not only differences in the growth rates but also the ability to survive in the stationary phase. Thus, we sought to cross-validate our approach by including data from [62], which reported the growth rates and cell volumes across LTEE. Cell volumes were converted to S/V using the empirical scaling $S \sim 2\pi V^{2/3}$ reported to hold across many bacterial species [63]. Second, we collected relative fitness and cell size of wild-type *Mycoplasma mycoides* and *M. mycoides* whose genome has been minimized by removal of non-essential genes [64].

Comparing LTEE data with our analytical model (Fig 8), we find that the observed scaling falls very near to expected scaling for *E. coli* grown in the media with similar composition (M63) as that used in LTEE minimal media with glycerol (Davis broth). Note that both indirect estimates (red points) and direct measurements (green points) of growth rate fall onto the same line. A similar trend is observed in the experimental

evolution of *M. mycoides*. The growth rate of the ancestors was not reported, so we used the value from our dataset to convert relative fitness into the growth rate. The blue line denotes the expected scaling for media with casamino acids and glycerol.

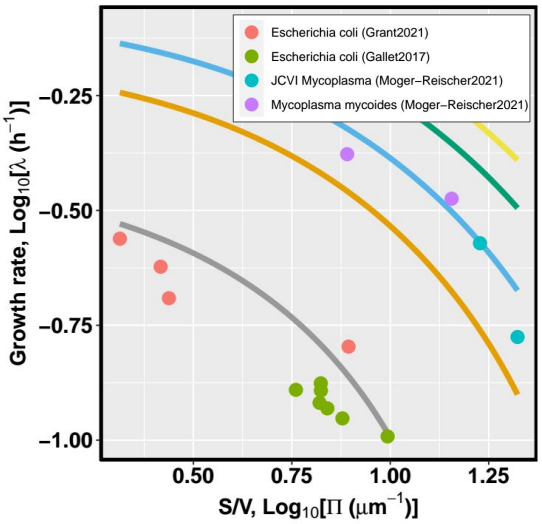

Figure 8. Scaling of growth rate with cell shape in experimental evolution setup. Red and green points – indirect estimate from relative fitness data and direct measurement of the growth rate in LTEE; Blue and purple points – Indirect growth rate estimate the from relative fitness data in experimental evolution of *Mycoplasma mycoides* wild-type and engineered cell with the minimal genome. Lines are the analytical solutions for different κ_n as in previous figures.

5.4 Scaling in the limit of large envelope costs and no degradation

Given the uncertainty of model parameters (especially envelope degradation), we attempted to simplify the theoretical expectation by focusing on the simplest possible case when degradation is absent, and the costs of the envelope are sufficiently high that Eq 22a can be approximated as $\tilde{\lambda}_{\max} \propto \Pi^{-1}$, Eq 21a as $\Phi_R \propto \Pi^{-1}$, and Eq 21b as $\Phi_L \propto \Pi^0$. While this scenario might not be the most realistic, the benefit is that it does not require any further specification of model parameters because the prediction is that the slope of regression $\mathrm{Log}_{10}(\tilde{\lambda}_{\max})$~$\mathrm{Log}_{10}(\Pi)$ and $\mathrm{Log}_{10}(\Phi_R)$~$\mathrm{Log}_{10}(\Pi)$ should be -1, and the slope of $\mathrm{Log}_{10}(\Phi_L)$~$\mathrm{Log}_{10}(\Pi)$ should be zero. Assuming that the shape is constant across bacteria, similar scalings can be obtained with cell volumes as the independent variable. One expects $\mathrm{Log}_{10}(\tilde{\lambda}_{\max})$~0.33 $\mathrm{Log}_{10}(V)$, $\mathrm{Log}_{10}(\Phi_R)$~0.33 $\mathrm{Log}_{10}(V)$, and Φ_L retains independence as with S/V. We use these expectations as the null hypothesis for the slopes in the regression analysis. In addition to scaling expectations, we also wish to test whether surface-to-volume is a better predictor than the cell volume. To this end, we performed OLS regression with either S/V or V as independent, and the growth rate, or proteomic mass fractions as the dependent variable.

	$\tilde{\lambda}_{\max}$	Φ_R	Φ_L
S/V_{cyt}	-0.984 ± 0.109 (-1)	-0.496 ± 0.190 (-1)	0.186 ± 0.226 (0)
S/V_{tot}	-1.244 ± 0.149 (-1)	-0.93 ± 0.26 (-1)	0.129 ± 0.371 (0)
V_{cyt}	0.08 ± 0.059 (0.33)	0.035 ± 0.109 (0.33)	-0.046 ± 0.139 (0)
V_{tot}	0.015 ± 0.06 (0.33)	0.103 ± 0.124 (0.33)	0.003 ± 0.152 (0)

Table 3. Observed and expected scaling exponent of the growth and proteome composition. Dependent variables are denoted in the first row, while independent variables are listed in the first column. Each table cell contains the observed slope with standard error, while the expected scaling exponent is reported in the parentheses. Blue color highlights the cases where the observed slope is not significantly different from the expected (i.e., where $p > 0.05$). Sample size: $n(\tilde{\lambda}_{\max}) = 229$, $n(\Phi_R) = 40$, $n(\Phi_L) = 30$.

Five insights emerge. First, S/V_{cyt} is a moderate predictor of growth (adjusted $R^2 = 0.26$), while cell volume appears neither to match the expected 0.33 scaling nor to show any correlation with the growth rate. Second, the observed scaling of growth with S/V is not significanly different from -1. Third, the scaling of proteomic mass fractions with S/V largely meets the expected slopes of -1, for Φ_R, and 0, for Φ_L. Fourth, both S/V_{cyt} and S/V_{tot} perform equally well, with S/V_{cyt} being a slightly better predictor of growth than S/V_{tot} (adjusted R^2 of 0.26 vs. 0.23) and the scaling exponent being within one standard error of the estimated -1 in the case of the former. Fifth, Φ_R scales isometrically with S/V_{tot} but not with S/V_{cyt}.

We cross-validated these results by analyzing a recently-published study [65], which contains data on linear dimensions of the cell and the minimal doubling time. Although we do not find any correlation between cell volume and the growth rate, we find a very weak negative relationship (adjusted $R^2 = 0.08$) between S/V and the growth rate (Fig S1.3) with a scaling exponent of -0.97 which is not significantly different from -1 ($p = 0.92$). Given that bacterial species with large S/V (i.e., small coccal or thin helical cells) are frequently parasitic, it is possible that the pattern is mainly driven by large-S/V species growing slower because of their natural habitat and not because trade-offs in investments. If that is the case, then excluding parasitic species from the dataset would eliminate negative scaling. To control for this confounding factor, we separated data into free-living and host-associated species and then performed the regression analysis (see Section S1.4 in S1 Appendix). We find a very weak negative relationship between S/V and the growth rate in both of these datasets and pooled data.

Thus far, we have assumed that all species have an envelope that is ∼30 nm thick, like that of an *E. coli* [58]. However, larger bacterial species may have thicker envelopes which require higher resource investments. If resource investments stay the same across the size range, then the variation in growth cannot be explained by the invariant S/V. To account for this possibility, we collected data on envelope thicknesses across 45 species and used these values to compute Π (see Section S1.3 in S1 Appendix). We find that $\tilde{\lambda}_{\max}$ scales with Π with exponent -1.09 (Fig S1.2) which is no significantly different from -1 ($p = 0.62$). Similarly, $\tilde{\lambda}_{\max}$ scales with V_{cyt} with exponent 0.44 (Table S1.2) which is not significantly different from 0.33 ($p = 0.42$). Therefore, the inclusion of envelope thickness data does not alter the conclusions reached by using fixed thickness.

Three conclusions are reached from the preceding analysis. First, Π is a moderate predictor of $\tilde{\lambda}_{\max}$. It accounts for roughly a quarter of growth rate variation in the whole dataset. Second, S/V_{cyt} is slightly better predictor of growth rate than S/V_{tot}. Third, the proteome composition qualitatively – but not quantitatively – fits the expected pattern. The ribosomal mass fraction of the proteome scales negatively with Π, albeit with a slope which is shallower than the expected -1, and the envelope

producer mass fraction appears independent of Π, which is in accordance with theory.

6 Discussion

Motivated by the observation that larger bacteria tend to grow faster, we proposed a theory that explains this relationship in terms of a trade-off between investment into surface features and biosynthetic machinery that builds the cell. By formalizing this verbal statement into a quantitative model, we obtained predictions that we sought to test with published data. Three conclusions are reached. First, the model recapitulates the previous physiological responses ($\Phi_R \sim \tilde{\lambda}_N$, and $\Phi_R \sim \tilde{\lambda}_T$), and uncovers new ones ($\Phi_L \sim \tilde{\lambda}_N$, and $\Phi_R \sim \tilde{\lambda}_L$). Second, it reveals that the natural variable that governs the scaling of the growth rate with bacterium size is the ratio of the cell surface area to volume and not the cell volume itself. And third, the model correctly predicts negative scaling of ribosomal content with S/V. On the other hand, the purported mild positive scaling of Φ_L is not unambiguously corroborated by proteomic data.

We attempted to control for factors that might affect the scaling. By cross-validating with an independently published dataset, we find that our conclusions are not the artifact of the dataset. By separately regressing data on host-associated and free-living species, we excluded the possibility that the overall scaling is caused by small-celled species being dominated by organisms selected for slow growth to avoid excessively damaging their host. By including data on envelope thickness when calculating S/V_{cyt}, we excluded the possibility that the envelope becomes thicker with cell size, thus increasing the relative investment into the envelope as cells become larger [66, 67]. By comparing growth rates across the experimental evolution study where fast growth is selected, we controlled for other factors – such as the environment that bacteria inhabit – that might cause the growth rate scaling. While our analysis was restricted to one particular theory, others have proposed alternative explanations for these phenomena. We now discuss those possibilities.

It was proposed that the larger bacteria have larger genomes and a greater gene repertoire, which allows them to metabolize a more diverse set of nutrients compared to their smaller counterparts [3]. There are a few problems with this idea. First, it is difficult to test this hypothesis because it does not offer a quantitative prediction on how genome length or gene number scaling should translate to the scaling of the growth rate. For example, should one expect the growth rate to scale with an identical exponent as gene number? If so, their data refutes the theory, given that the growth rate scales with the power of 0.73 and gene number and genome size with the power of 0.35. On the other hand, if the growth rate scales with the square of the gene number, then their explanation seems more plausible. Second, even if the genome hypothesis is true, it begs the question what causes genome size (or gene number) to scale with the power of 0.35 with cell mass instead of some other value. Third, while this hypothesis might in principle explain the interspecific growth scaling, it cannot explain scaling of growth in experimental evolution as all data points correspond to the same species with the identical gene repertoire.

Another alternative hypothesis is that larger cells are growing faster because their cytoplasm is less crowded, thus alleviating constraints from the internal diffusion of macromolecules [62]. According to this hypothesis, growth is proportional to the abundance of metabolic proteins, implying that faster growth can be only achieved by increasing their abundance. However, increased abundance of effectors can slow down intercellular diffusion and thus reduce the encounter rates between cellular components. To solve this problem, an increase in the total mass of these effectors has to be followed by the increase of volume to offset this kind of effect. On the contrary, volume and mass scale proportionaly in our data set (slope of 0.93 ± 0.084 with adjusted $R^2 = 0.88$ is not

significantly different from unity, $p = 0.412$), meaning that larger cells are not less dense. Furthermore, if the selection optimizes the density of cytoplasm such that the rates of cellular reactions are maximized [68, 69], one would expect cell density to remain constant over the course of experimental evolution. However, given that the mass-to-volume ratio decreases, it is not clear why this would not increase the mean time required for two proteins to collide with one another and, thus, decrease the rate of cellular reactions.

The theory developed here conflicts with three previously published conclusions. First, contrary to the assumption that the selection coefficient (i.e., the evolutionary cost) of a particular cellular feature is directly proprotional to the fraction of total resources invested in the production of the trait [4, 70], here we show that this is not necessarily true, and that it can also depend on the rates of cellular reactions (Eq 37–39). This is because the costs of the trait also have to include the costs of machinery that builds the trait (in this case, the envelope-producing enzyme), and the amount of enzymes produced depends on how fast it operates relative to other proteome components: If the enzyme is slow, then a large fraction of the proteome has to be allocated to it to achieve an appropriate flux. Second, although a recently-published analysis of Madin2020 dataset [71] concluded that there is no correlation between cell diameter and growth rate across prokaryotes, our analysis reveals that growth rate correlates with S/V, at least when one focuses on bacteria for which both linear dimensions are available.

Third, it was argued that the observed scaling of various bacterial features necessarily imposes the upper limit on how large a bacterium can be. According to this view, the ribosomes have to replicate themselves, and all other proteins within the cell doubling time. So, if ribosomes, other proteins, and doubling time scale with different exponents, it is possible for a cell to reach a particular volume at which it does not have enough ribosomes to replicate the entire proteome within the doubling time inferred from a specific power function – the event deemed the "ribosome catastrophe" [72]. However, given that many functions can fit the same data, it is unclear whether fitting pure power functions is adequate. Indeed, our model can explain the scaling of both proteome and growth rate without any fundamental cell size limit. Of course, there might be many reasons why bacteria cannot evolve extreme cell sizes, but inferring this limit from the scaling laws within extant bacterial species may be a questionable approach.

Our model also offers a causal explanation for the correlation between genomic features associated with translation and growth rate. For example, faster-growing bacteria tend to have a greater number of rRNA operons [73]. According to our theory, bacteria grow slowly because they cannot allocate enough resources to ribosomes, given that other surface-related constraints have to be satisfied. Suppose the high copy number of rRNA genes is caused by the need to meet the high demand for ribosomes. In that case, species with large S/V will have no selective advantage in possessing additional gene copies, meaning that they will be purged by selection to reduce the costs of replicating the added DNA. The same explanation holds true for tRNA genes. Unfortunately, our theory does not yield the precise expectation for the scaling of genomic features, so this hypothesis cannot be tested at the moment. Nonetheless, we find a negative log-log scaling between S/V of the species and its rRNA ($p < 10^{-9}$, adjusted $R^2 = 0.171$) and tRNA ($p < 10^{-13}$, adjusted $R^2 = 0.249$) gene number (see Section S1.5 in S1 Appendix for details).

The model presented here is not universal, as it cannot account for the scaling of growth across the entire Tree of Life. We would expect the growth rate to monotonically increase as S/V decreases, but we know that this is not true because eukaryotes with much smaller S/V have growth rates that decrease with size [74]. Even

large bacteria, like *Metabacterium polyspora* with a volume of 480 μm^3, have doubling times measured in days [75]. While the selection for fast growth in bacteria leads to the reduction in S/V, the opposite is true for the fungus *Kluyveromyces marxianus* where propagation in auxostat leads to an increase in S/V [76], which contradicts our theory. Therefore, the scaling law proposed here breaks at some point, as other constraints become more dominant. Although advances in integrating multiple physical constraints acting at once have been made [77], the theory that mechanistically unifies them and derives expected scaling laws remains to be developed.

7 Data availability

All notebooks, data, and code used to derive the theoretical results and plot all the figures is available in Supporting information and at `https://github.com/BogiTrick/growth_scaling_envelope`.

8 Supporting information

S1 Appendix. Robustness, corrections, and cross-validation. Contains the sensitivity of the scaling to variation in protein degradation rates, correction for cell envelope thickness, and cross-validation of the scaling trends with independent dataset. The method for obtaining errors of inferred parameters is described.

S1 File. Notebooks and scripts. Contains *Mathematica* notebooks for reproduction of entire theoretical derivation, spreadsheets with raw data, and R scripts used for data processing.

9 Acknowledgements

We are thankful to Paul Schavemaker, Jon Harrison, Jesse Taylor, and Kerry Geiler-Samerotte for the critical reading of the manuscript, and to Nkrumah Grant for sharing the cell size data in the long-term experimental evolution of *E. coli*.

10 Funding

This work was supported by the Moore–Simons Project on the Origin of the Eukaryotic Cell, Simons Foundation 735927, `https://doi.org/10.46714/735927`; US Department of the Army, MURI award W911NF-14-1-0411, 2014-2019, Innovation in Prokaryotic Evolution (co-PI, with Pat Foster, Jay Lennon, Jake McKinlay, and Allan Drummond); National Institutes of Health, R35-GM122566-01, 2017-2022, Causes and Population-genetic Consequences of Molecular Variation; National Science Foundation, DBI-2119963, 2021-2026, BII: Mechanisms of Cellular Evolution (with W. Frasch, K. Geiler-Samerotte, K. Hu, J. Wideman).

11 Competing interests

The authors declare that they have no competing interests.

References

1. Eagon RG. *Pseudomonas natriegens*, a marine bacterium with a generation time of less than 10 minutes. J Bacteriol. 1962;83(4):736–737.

2. Tanno-Nakanishi M, Kikuchi Y, Kokubu E, Yamada S, Ishihara K. *Treponema denticola* transcriptional profiles in serum-restricted conditions. FEMS Microbiol Lett. 2002;365(16):fny171.

3. DeLong JP, Okie JG, Moses ME, Sibly RM, Brown JH. Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. Proc Natl Acad Sci U S A. 2010;107(29):12941–12945.

4. Lynch M, Marinov GK. The bioenergetic costs of a gene. Proc Natl Acad Sci USA. 2015;112:15690–15695.

5. Chen B, Liu H. Relationships between phytoplankton growth and cell size in surface oceans: Interactive effects of temperature, nutrients, and grazing. Limnol Oceanogr. 2010;55(3):965–972.

6. Marañón et al E. Unimodal size scaling of phytoplankton growth and the size dependence of nutrient uptake and use. Ecol Lett. 2013;16(3):371–379.

7. West GB, Woodruff WH, Brown JH. Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. Proc Natl Acad Sci U S A. 2002;99(Suppl. 1):2473–2478.

8. Banavar JR, Damuth J, Maritan A, Rinaldo A. Supply–demand balance and metabolic scaling. Proc Natl Acad Sci USA. 2002;99(16):10506–10509.

9. Finkel ZV. Light absorption and size scaling of light-limited metabolism in marine diatoms. Limnol Oceanogr. 2001;46(1):86–94.

10. Pirie NW. On being the right size. Annu Rev Microbiol. 1973;27(1):119–132.

11. Raven JA. Why are there no picoplanktonic O_2 evolvers with volumes less than 10^{-19} m^3? J Plankt Res. 1962;16(5):565–580.

12. Raven J. The twelfth Tansley lecture. Small is beautiful: the picophytoplankton. Funct Ecol. 1998;12(4):503–513.

13. Molenaar D, Van Berlo R, De Ridder D, Teusink B. Shifts in growth strategies reflect tradeoffs in cellular economics. Mol Syst Biol. 2009;5(1):1099–1102.

14. Bertaux F, von Kügelgen J, Marguerat S, Shahrezaei V. A bacterial size law revealed by a coarse-grained model of cell physiology. PLoS Comp Biol. 2020;16(9):1–21.

15. Piir K, Paier A, Liiv A, Tenson T, Maiväli Ü. Ribosome degradation in growing bacteria. EMBO Rep. 2011;12(5):458–462.

16. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. Interdependence of cell growth and gene expression: origins and consequences. Science. 2010;330(6007):1099–1102.

17. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. Nat Chem Biol. 2009;5(8):593.

18. Pandey PP, Jain S. Analytic derivation of bacterial growth laws from a simple model of intracellular chemical dynamics. Theory Biosci. 2016;135(3):pp.121–130.

19. Calabrese L, Grilli J, Osella M, Kempes CP, Lagomarsino MC, Ciandrini L. Role of protein degradation in growth laws. bioRxiv. 2020;.

20. Mori M, Schink S, Erickson DW, Gerland U, Hwa T. Quantifying the benefit of a proteome reserve in fluctuating environments. Nat Commun. 2017;8(1):1–8.

21. Monk JM, Koza A, Campodonico MA, Machado D, Seoane JM, Palsson BO, et al. Multi–omics quantification of species variation of *Escherichia coli* links molecular features with strain phenotypes. Cell Syst. 2009;3(3):238–251.

22. Valgepea K, Adamberg K, Seiman A, Vilu R. Escherichia coli achieves faster growth by increasing catalytic and translation rates of proteins. Mol Biosyst. 2013;9(9):2344–2358.

23. Li GW, Burkhardt D, Gross C, Weissman JS. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell. 2014;157(3):624–635.

24. Peebo K, Valgepea K, Maser A, Nahku R, Adamberg K, Vilu R. Proteome reallocation in Escherichia coli with increasing specific growth rate. Mol BioSyst. 2015;11(4):1184–1193.

25. Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, et al. The quantitative and condition-dependent Escherichia coli proteome. Nat Biotechnol. 2016;34(1):104–110.

26. Erickson DW, Schink SJ, Patsalo V, Williamson JR, Gerland U, Hwa T. A global resource allocation strategy governs growth transition kinetics of Escherichia coli. Nature. 2017;551(7678):119–123.

27. Mori M, Zhang Z, Banaei-Esfahani A, Lalanne JB, Okano H, Collins BC, et al. From coarse to fine: the absolute *Escherichia coli* proteome under diverse growth conditions. Mol Syst Biol. 2021;17(5):e9536.

28. Forchhammer J, Lindahl L. Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of Escherichia coli 15. J Mol Bio. 1971;55(3):563–568.

29. Bremer HDPP, Dennis PP. Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli* and *Salmonella*: cellular and molecular biology. 1996;2(2):1553–69.

30. Si F, Li D, Cox SE, Sauls JT, Azizi O, Sou C, et al. Invariance of initiation mass and predictability of cell size in *Escherichia coli*. Curr Biol. 2017;27(9):pp.1278–1287.

31. Dai X, Zhu M, Warren M, Balakrishnan R, Patsalo V, Okano H, et al. Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth. Nat Microbiol. 2016;2(2):1–9.

32. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28(1):27–30.

33. Lynch M. The origins of cellular architecture. Oxford University Press; in press.

34. Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, Hengartner MO, et al. PaxDb, a database of protein abundance averages across all three domains of life. Mol Cell Proteomics. 2012;11(8):492–500.

35. Müller JB, Geyer PE, Colaço AR, Treit PV, Strauss MT, Oroshi M, et al. The proteome landscape of the kingdoms of life. Nature. 2020;582(7813):592–596.

36. Lynch M, Marinov GK. Membranes, energetics, and evolution across the prokaryote-eukaryote divide. Elife. 2017;6:e20437.

37. Zhao X, Schwartz CL, Pierson J, Giovannoni SJ, McIntosh JR, Nicastro D. Three-dimensional structure of the ultraoligotrophic marine bacterium "*Candidatus Pelagibacter ubique*". Appl Environ Microbiol. 2017;83(3):e02807–16.

38. Angel TE, Luft BJ, Yang X, Nicora CD, Camp II DG, Jacobs JM, et al. Proteome analysis of *Borrelia burgdorferi* response to environmental change. PloS one. 2010;5(11):e13800.

39. Osbak KK, Houston S, Lithgow KV, Meehan CJ, Strouhal M, Šmajs D, et al. Characterizing the syphilis-causing *Treponema pallidum ssp. pallidum* proteome using complementary mass spectrometry. PLoS Negl Trop Dis. 2016;10(9):e0004988.

40. Srivastava A, Murugaiyan J, Garcia JA, De Corte D, Hoetzinger M, Eravci M, et al. Combined methylome, transcriptome and proteome analyses document rapid acclimatization of a bacterium to environmental changes. Front Microbiol. 2020;11:2197.

41. Matteau D, Lachance JC, Grenier F, Gauthier S, Daubenspeck JM, Dybvig K, et al. Integrative characterization of the near-minimal bacterium *Mesoplasma florum*. Mol Sys Bio. 2020;16(12):e9844.

42. Masson F, Rommelaere S, Marra A, Schüpfer F, Lemaitre B. Dual proteomics of *Drosophila melanogaster* hemolymph infected with the heritable endosymbiont *Spiroplasma poulsonii*. Plos One. 2021;16(14):e0250524.

43. Stouthamer AH. A theoretical study on the amount of ATP required for synthesis of microbial cell material. Antonie van Leeuwenhoek. 1973;39(1):545–565.

44. Lynch M, Trickovic B. A theoretical framework for evolutionary cell biology. J Mol Bio. 2020; p. 1861–1879.

45. Marvaldi J, Pichon J, Delaage M, Marchis-Mouren G. Individual ribosomal protein pool size and turnover rate in *Escherichia coli*. J Mol Bio. 1974;84(1):83–96.

46. Chen SS, Williamson JR. Characterization of the ribosome biogenesis landscape in E. coli using quantitative mass spectrometry. J Mol Bio. 2011;425(4):767–779.

47. Pulk A, Liiv A, Peil L, Maiväli Ü, Nierhaus K, Remme J. Ribosome reactivation by replacement of damaged proteins. Mol Microbio. 2010;75(4):801–814.

48. Nath K, Koch AL. Protein degradation in *Escherichia coli*: I. Measurement of rapidly and slowly decaying components. J Biol Chem. 1970;245(11):2889–2900.

49. Maurizi M. Proteases and protein degradation in *Escherichia coli*. Experientia. 1992;48(2):178–201.

50. Cheung HY, Vitković L, Freese E. Rates of peptidoglycan turnover and cell growth of *Bacillus subtilis* are correlated. J Bacteriol. 1983;156(3):1099–1106.

51. Goodell EW, Schwarz U. Release of cell wall peptides into culture medium by exponentially growing *Escherichia coli*. J Bacteriol. 1983;162(1):391–397.

52. Mahmoudabadi G, Phillips R, Lynch M, Milo R. Defining the Energetic Costs of Cellular Structures. bioRxiv. 2019; p. p.666040.

53. Milo R, Phillips R. Cell biology by the numbers. Garland Science, New York, NY; 2015.

54. Zhang J. Protein-length distributions for the three domains of life. Trends Genet. 2000;16(3):107–9.

55. Wu C, Balakrishnan R, Mori M, Manzanarez G, Zhang Z, Hwa T. Cellular perception of growth rate and the mechanistic origin of bacterial growth laws. bioRxiv. 2021;.

56. Monod J. The growth of bacterial cultures. Annu Rev Microbiol. 1949;3(1):371–394.

57. Chevin LM. On measuring selection in experimental evolution. Biol Lett. 2011;7(2):210–213.

58. Matias VR, Al-Amoudi aDJ A, Beveridge TJ. Cryo-transmission electron microscopy of frozen-hydrated sections of *Escherichia coli* and *Pseudomonas aeruginosa*. J Bacteriol. 2003;185(20):6112–6118.

59. Grant NA, Magid AA, Franklin J, Dufour YS, Lenski R. Changes in Cell Size and Shape During 50,000 Generations of Experimental Evolution with *Escherichia coli*. J Bacteriol. 2021;203(10):e00469–20.

60. Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. Am Nat. 1991;138(6):1315–1341.

61. Vasi F, Travisano M, Lenski RE. Long-term experimental evolution in Escherichia coli. II. Changes in life-history traits during adaptation to a seasonal environment. Am Nat. 1994;144(3):432–456.

62. Gallet R, Violle C, Fromin N, Jabbour-Zahab R, Enquist BJ, Lenormand T. The evolution of bacterial cell size: the internal diffusion-constraint hypothesis. ISME J. 2017;11(7):1559–1568.

63. Ojkic N, Serbanescu D, Banerjee S. Surface-to-volume scaling and aspect ratio preservation in rod-shaped bacteria. eLife. 2019;8:e47033.

64. Moger-Reischer RZ, Glass JI, Wise K, Bittencour D, Sun L, Lynch M, et al. Evolution of a minimal cell. bioRxiv. 2021;.

65. Madin et al JS. A synthesis of bacterial and archaeal phenotypic trait data. Sci Data. 2020;7(1):1–8.

66. Raven JA. The energetics of freshwater algae; energy requirements for biosynthesis and volume regulation. New Phytol. 1982;92(1):1–20.

67. Koch AL. What size should a bacterium be? A question of scale. Annu Rev Microbiol. 1996;50:317–348.

68. Dill K, Ghosh K, Schmit J. Physical limits of cells and proteomes. Proc Natl Acad Sci U S A. 2011;108:17876–17882.

69. Soh S, Banaszak M, Kandere-Grzybowska K, Grzybowski BA. Why are cells microscopic: A transport-time perspective. J Phys Chem Lett. 2001;4(6):861–865.

70. Ilker E, Hinczewski M. Modeling the growth of organisms validates a general relation between metabolic costs and natural selection. Phys Rev Lett. 2019;122(23):p.238101.

71. Westoby M, Nielsen DA, Gillings MR, Litchman E, Madin JS, Paulsen IT, et al. Cell size, genome size, and maximum growth rate are near-independent dimensions of ecological variation across bacteria and archaea. Ecol Evol. 2021;11(9):3956–3976.

72. Kempes CP, Wang L, Amend JP, Doyle J, Hoehler T. Evolutionary tradeoffs in cellular composition across diverse bacteria. ISME J. 2016;10(9):2145.

73. Vieira-Silva S, Rocha EP. The systemic imprint of growth and its uses in ecological (meta) genomics. PLoS Genet. 2021;6(1):e1000808.

74. Hatton IA, Dobson AP, Storch D, Galbraith ED, Loreau M. Linking scaling laws across eukaryotes. Proc Natl Acad Sci U S A. 2019;116(43):21616–21622.

75. Stünkel S, Alves J, Kunstýř I. Characterization of two "*Metabacterium*" sp. from the gut of rodents. Folia microbiologica. 1993;38(3):171–175.

76. Groeneveld P, Stouthamer AH, V WH. Super life–how and why 'cell selection' leads to the fastest-growing eukaryote. FEBS J. 2009;276(1):254–70.

77. Kempes CP, Koehl MAR, West GB. The scales that limit: the physical boundaries of evolution. Front Ecol Evol. 2019;7:242.