

Fundamental limits on the rate of bacterial growth

Nathan M. Belliveau^{†, 1}, Griffin Chure^{†, 2}, Christina L. Hueschen³, Hernan G. Garcia⁴, Jane Kondev⁵, Daniel S. Fisher⁶, Julie A. Theriot^{1, 7, *}, Rob Phillips^{8, 9, *}

¹Department of Biology, University of Washington, Seattle, WA, USA; ²Department of Applied Physics, California Institute of Technology, Pasadena, CA, USA; ³Department of Chemical Engineering, Stanford University, Stanford, CA, USA; ⁴Department of Molecular Cell Biology and Department of Physics, University of California Berkeley, Berkeley, CA, USA; ⁵Department of Physics, Brandeis University, Waltham, MA, USA; ⁶Department of Applied Physics, Stanford University, Stanford, CA, USA; ⁷Allen Institute for Cell Science, Seattle, WA, USA; ⁸Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA; ⁹Department of Physics, California Institute of Technology, Pasadena, CA, USA; *Co-corresponding authors. Address correspondence to phillips@pboc.caltech.edu and jtheriot@uw.edu; [†]These authors contributed equally to this work

Abstract Recent years have seen an experimental deluge interrogating the relationship between bacterial growth rate, cell size, and protein content, quantifying the abundance of proteins across growth conditions with unprecedented resolution. However, we still lack a rigorous understanding of what sets the scale of these quantities and when protein abundances should (or should not) depend on growth rate. Here, we seek to quantitatively understand this relationship across a collection of *Escherichia coli* proteomic data covering ≈ 4000 proteins and 36 growth rates. We estimate the basic requirements for steady-state growth by considering key processes in nutrient transport, energy generation, cell envelope biogenesis, and the central dogma. From these estimates, ribosome biogenesis emerges as a primary determinant of growth rate. We expand on this assessment by exploring a model of proteomic regulation as a function of the nutrient supply, revealing a mechanism that ties cell size and growth rate to ribosomal content.

Introduction

The observed range of bacterial growth rates is enormously diverse. In natural environments, some microbial organisms may double only once per year (?) while in comfortable laboratory conditions, growth can be rapid with several divisions per hour (?). This six order-of-magnitude difference in time scales of growth encompasses different microbial species and lifestyles, yet even for a single species such as *Escherichia coli*, the growth rate can be modulated over a comparably large scale by tuning the type and amount of nutrients in the growth medium (?). This remarkable plasticity in growth rate illustrates the intimate relationship between environmental conditions and the rates at which cells convert nutrients into new cellular material – a relationship that has remained a major topic of inquiry in bacterial physiology for over a century (?).

A key discovery of the past 70 years was the identification of bacterial "growth laws" (?); empirical relationships that relate the bacterial growth rate to the protein and RNA composition of the intracellular milieu in a number of different species. Over the past decade, a flurry of work (??????) have examined these growth laws at a quantitative level, developing a series of phenomenological models from which the growth laws naturally emerge. In parallel, a "molecular revolution" in biology has yielded an increasingly refined molecular census of the cell, particularly for bacteria such as the microbial workhorse *E. coli* (?). In light of the now expansive trove of quantitative biological data, we can revisit several of the evergreen questions about bacterial growth and physiology that were originally raised by microbiologists in the middle of the 20th century. Specifically, what biological processes are the primary determinants for how quickly bacterial cells can grow and reproduce? Why do cells modulate the absolute numbers and relative ratios of their molecular constituents as a function of changes in growth rate or nutrient availability?

In this paper, we begin by considering these two questions from two distinct angles. First, as a result of an array of high-quality

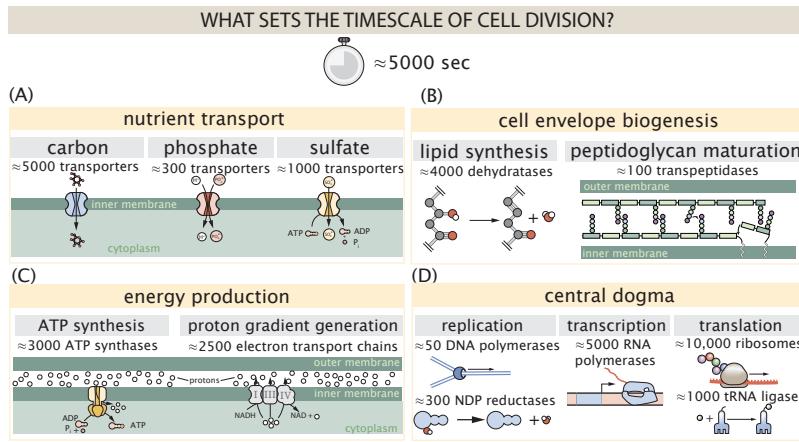


Figure 1. Transport and synthesis processes necessary for cell division. We consider an array of processes necessary for a cell to double its molecular components, broadly grouped into four classes. These categories are (A) nutrient transport across the cell membrane, (B) cell envelope biogenesis, (C) energy production (namely, ATP synthesis), and (D) processes associated with the central dogma. Numbers shown are the approximate number of complexes of each type observed at a growth rate of 0.5 hr^{-1} , or a cell doubling time of $\approx 5000 \text{ s}$.

proteome-wide measurements of *E. coli* under diverse growth conditions, we have generated a census that allows us to explore how the number of key molecular players change as a function of growth rate. Here, we have assembled a singular data set of protein copy numbers using measurements collected over the past decade via mass spectrometry (???) or ribosomal profiling (?) of the composition of the *E. coli* proteome across 36 unique growth rates. Due to notable changes in cell size and cellular composition as a function of growth rate (??), substantial care was taken to ensure consistency on a per cellular basis, and detailed analysis and additional discussion is provided in Appendix Section "??". Second, by compiling molecular turnover rate measurements for many of the fundamental processes associated with bacterial growth, we make quantitative estimates of a handful of key cellular processes (schematized in ??) to determine whether our current understanding of the dynamics of these processes are sufficient to explain the magnitude of the observed protein copy numbers across conditions (see ?? describing the philosophy behind this approach). The census, combined with these estimates, provide a window into the question of whether the rates of central processes such as energy generation or DNA synthesis vary systematically as a function of cell growth rate by altering protein copy number.

Throughout our estimates, we consider an archetypal growth rate of $\approx 0.5 \text{ hr}^{-1}$ corresponding to a doubling time of ≈ 5000 seconds, as the data sets examined here heavily sample this growth regime. While we formulate point estimates for the protein abundances at this division time, we also consider how these values will vary at other growth rates due to changes in cell size, surface area, and chromosome copy number (??). For the majority of the processes considered, we find that the protein copy numbers appear tuned for the task of cell doubling across a continuum of growth rates. Thus, our understanding of the kinetics of various biological processes is sufficient to quantitatively explain the observed abundances of these proteins.

From these estimates, it emerges that translation, particularly the synthesis of ribosomal proteins is a plausible candidate that limits the rate of cellular growth in *E. coli*. We reach this conclusion by considering that ribosome synthesis is 1) a rate limiting step for the *fastest* bacterial division, and 2) the main determinant of bacterial growth rate across nutrient conditions associated with moderate to fast growth rates. In addition, a strict dependence between the maximal growth rate and ribosomal mass fraction coincides with the regime where the growth laws appear most valid (??). This enables us to suggest that the long-observed correlation between growth rate and cell size (??) can be simply attributed to the increased absolute number of ribosomes per cell under conditions supporting extremely rapid growth. To better understand how the observed alterations in absolute protein abundances, and in particular, changes in ribosome copy number, influence growth rate across different nutrient conditions we consider a minimal model of cellular growth. Our conclusions from these analyses provide important insight into how *E. coli* regulates growth across conditions of differing nutrient availability and identifies fundamental constraints in bacterial growth more broadly.

Nutrient Transport

We begin by considering the critical transport processes diagrammed in ??(A). In order to build new cellular mass, the molecular and elemental building blocks must be scavenged from the environment in different forms. Carbon, for example, is acquired via the transport of carbohydrates and sugar alcohols with some carbon sources receiving preferential treatment in their consumption (?). Phosphorus, sulfur, and nitrogen, on the other hand, are harvested primarily in the forms of inorganic salts, namely phosphate,

Box 1. The Rules of Engagement for Order-Of-Magnitude Estimates

This work relies heavily on "back-of-the-envelope" estimates to understand the growth-rate dependent abundances of molecular complexes. This moniker arises from the limitation that any estimate should be able to fit on the back of a postage envelope. As such, we must draw a set of rules governing our precision and sources of key values.

The rule of "one, few, and ten". The philosophy behind order-of-magnitude estimates is to provide an estimate of the appropriate scale, not a prediction with many significant digits (?). We therefore define three different scales of precision in making estimates. The scale of "one" is reserved for values that range between 1 and 2. For example, If a particular process has been experimentally measured to transport 1.87 protons for a process to occur, we approximate this process to require 2 protons per event. The scale of "few" is reserved for values ranging between 3 and 7. For example, we will often use Avogadro's number to compute the number of molecules in a cell given a concentration and a volume. Rather than using Avogadro's number as 6.02214×10^{23} , we will approximate it as 5×10^{23} . Finally, the scale of "ten" is reserved for values which we know within an order of magnitude. If a particular protein complex is present at 883 copies per cell, we say that it is present in approximately 10^3 copies per cell. These different scales will be used to arrive at simple estimates that report the expected scale of the observed data. Therefore, the estimates presented here should not be viewed as hard-and-fast predictions of precise copy numbers, but as approximate lower (or upper) bounds for the number of complexes that may be needed to satisfy some cellular requirement. Furthermore, we use equality symbols (=) sparingly and frequently defer to approximation (\approx) or scaling (~) symbols when reporting an estimate. When \approx is used, we are implicitly stating that we are confident in this estimate within a factor of a few. When a scaling symbol ~ is used, we are stating that we are confident in our estimate to within an order of magnitude.

The BioNumbers Database as a source for values. In making our estimates, we often require approximate values for key cellular properties, such as the elemental composition of the cell, the average dry mass, or approximate rates of synthesis. We rely heavily on the BioNumbers Database (bionumbers.hms.harvard.edu, ?) as a repository for such information. Every value we draw from this database has an associated BioNumbers ID number, abbreviated as BNID, and we provide this reference in grey-boxes in each figure.

Uncertainty in the data sets and the accuracy of an estimate. The data sets presented in this work are the products of careful experimentation with the aim to report, to the best of their ability, the absolute copy numbers of proteins in the cell. These data, collected over the span of a few years, come from different labs and use different internal standards, controls, and even techniques (discussed further in Appendix ??). As a result, there is notable disagreement in the measured copy numbers for some complexes across data sets. In assessing whether our estimates could explain the observed scales and growth-rate dependencies, we also considered the degree of variation between the different data sets. For example, say a particular estimate undercuts the observed data by an order of magnitude. If all data sets agree within a factor of a few of each other, we revisit our estimate and consider what we may have missed. However, if the data sets themselves disagree by an order of magnitude, we determine that our estimate is appropriate given the variation in the data.

Point versus continuum estimates For each estimate performed in this work, we begin with a simple order-of-magnitude estimate for the abundance of the complex in question at an archetypal growth rate of around 0.5 hr^{-1} , followed by a more refined estimate across a continuum of growth rates from around 0.05 to 2.0 hr^{-1} . The former estimate, always outlined in the associated figure and indicated as a translucent brown point in the corresponding plots, will rely on making coarse-grained approximations of cell mass, cell volume, and/or typical surface areas. The continuum estimates, displayed as a grey curve on the various plots, relax these assertions and incorporate empirical findings from the literature of how cell masses, volumes, and surface areas scale with the cellular growth rate. Thus, it is possible for the point estimate at a growth rate of 0.5 hr^{-1} may not perfectly agree with the continuum estimate at 0.5 hr^{-1} . We emphasize that both the point and continuum estimates are not hard predictions for the complex abundance, but rather reflect an order-of-magnitude estimates. While the proteomic measurements will often not fall directly on the curve corresponding to the continuum estimate, we do not view this as a failure of the approach. Rather, we gauge the accuracy of our estimates by examining whether the 1) the point and continuum estimates are within the same order of magnitude as the observations and 2) whether the continuum estimate qualitatively captures the growth-rate dependence observed in the data. We direct the reader to the Appendix section entitled "Extending Estimates to a Continuum of Growth Rates" for a more in-depth discussion of point versus continuum estimates.

sulfate, and ammonium/ammonia (??????). All of these compounds have different membrane permeabilities (?) and most require some energetic investment either via ATP hydrolysis or through the proton electrochemical gradient to bring the material across the hydrophobic cell membrane.

The elemental composition of *E. coli* has received much quantitative attention over the past half century (????), providing us with a starting point for estimating how many atoms of each element must be scavenged from the environment. A synthesis of these studies presents an approximate dry mass composition of $\approx 50\%$ carbon (BNID: 100649; see ?? for explanation of BNID references), $\approx 15\%$ nitrogen (BNID: 106666), $\approx 3\%$ phosphorus (BNID: 100653), and 1% sulfur (BNID: 100655) with the remainder being attributable to oxygen, hydrogen, and various transition metals. We use this stoichiometric breakdown to estimate the abundance and growth rate dependence of a variety of transporters responsible for carbon uptake, and provide more extensive investigation of the other critical elements – phosphorus, sulfur, and nitrogen – in the Appendix Section "??".

Using ≈ 0.3 pg as the typical *E. coli* dry mass at a growth rate of $\approx 0.5 \text{ hr}^{-1}$ (BNID: 103904), coupled with an approximation that $\approx 50\%$ of this mass is carbon, we estimate that $\sim 10^{10}$ carbon atoms must be brought into the cell in order to double all of the carbon-containing molecules (??(A), top). Typical laboratory growth conditions provide carbon as a single class of sugar (such as glucose, galactose, or xylose) often transported across the cell membrane by a transporter complex specific to that particular sugar. One such mechanism of transport is via the PTS system, which is a highly modular system capable of transporting a diverse range of sugars with high specificity (?). The glucose-specific component of this system transports ≈ 200 glucose molecules (≈ 1200 carbon atoms) per second per transporter (BNID: 114686). Making the assumption that this is a typical sugar transport rate for the PTS system, coupled with the need to transport $\sim 10^{10}$ carbon atoms, we then expect on the order of ≈ 2000 transporters must be expressed per cell in order to bring in enough carbon atoms (??(A), top).

We find, however, that the experimental measurements exceed this by several fold (??(A), bottom), implying that the cell is capable of transporting more carbon atoms than strictly needed for biosynthesis. We can also abstract this calculation to consider any particular growth rate given knowledge of the cell density and volume as a function of growth rate and direct the reader to the Appendix Section "??" for more information. This abstraction, shown as a grey line in ??(A), reveals an excess of transporters even at faster growth rates. This contrasts with our observations for uptake of phosphorus and sulfur, which turn out to align well with our expectations across different growth conditions (??–?? and discussed further in the Appendix Section "??").

It is important to note that so far we have neglected any specifics of the regulation of the carbon transport system. Using the diverse array of growth conditions available in the data, we can explore how individual carbon transport systems depend on specific carbon availability. In ??(B), we show the total number of carbohydrate transporters specific to different carbon sources. A striking observation, shown in the top-left plot of ??(B), is the constancy in the expression of the glucose-specific transport systems, an observation that stands in contrast with other species of transporters. Additionally, we note that the total number of glucose-specific transporters is tightly distributed at $\approx 10^4$ per cell, the approximate number of transporters needed to sustain rapid growth of several divisions per hour. This illustrates that *E. coli* maintains a substantial number of complexes present for transporting glucose regardless of growth condition, which is known to be the preferential carbon source (??).

Many metabolic operons are regulated with dual-input logic gates that are only expressed when glucose concentrations are low and the concentration of other carbon sources are elevated (?????). Points colored in red in ??(B) (labeled by red text-boxes) correspond to growth conditions in which the specific carbon source (glycerol, xylose, or fructose) is present as the sole source of carbon. The grey lines in ??(B) show the estimated number of transporters needed at each growth rate to satisfy the cellular carbon requirement, adjusted for the specific carbon source in terms of number of carbon atoms per molecule and the rate of transport for the particular transporter species. These plots show that, even in the absence of the particular carbon source, expression of the transporters is maintained on the order of $\sim 10^2$ per cell. The low but non-zero abundances may reflect the specific regulatory logic involved, requiring that cells are able to transport some minimal amount of an alternative carbon source in order to induce expression of these alternative carbon-source systems when needed.

Limits on Transporter Expression

If acquisition of nutrients was a limiting process in cell division under the typical growth conditions explored here, the growth rate could be theoretically increased simply by expressing more transporters, but is this feasible at a physiological level? A way to approach this question is to compute the amount of space in the bacterial membrane that could be occupied by nutrient transporters. Considering a rule-of-thumb for the surface area of *E. coli* of about $5 \mu\text{m}^2$ (BNID: 101792), we expect an areal density for 2000 transporters to be approximately a few hundred transporters per μm^2 . For a typical transporter occupying about 50 nm^2 , this amounts to about only $\approx 1\%$ of the total inner membrane surface area (?). Bacterial cell membranes typically have densities of 10^5 proteins/ μm^2 (?), with roughly 60 % of the surface area occupied by protein (BNID: 100078), implying that the cell could easily

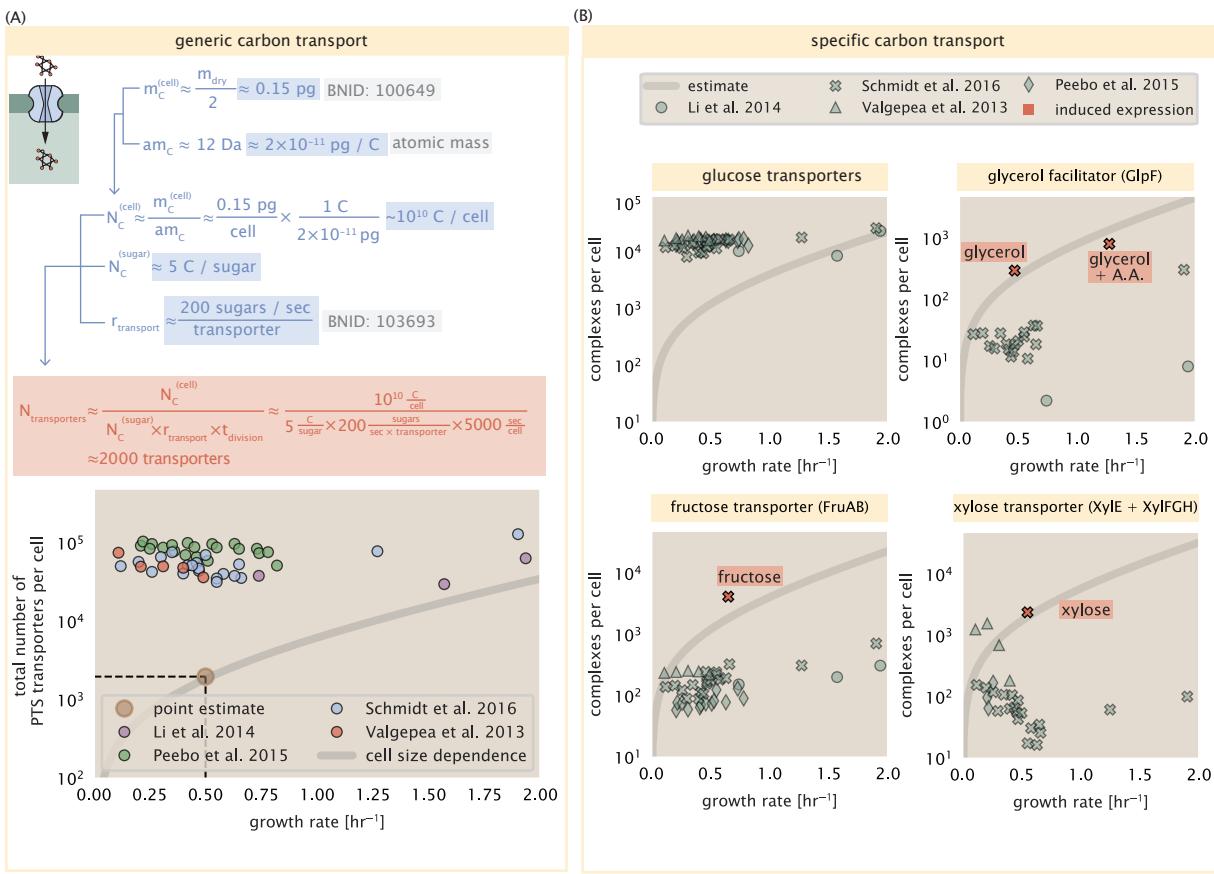


Figure 2. The abundance of carbon transport systems across growth rates. (A) A simple estimate for the minimum number of generic carbohydrate transport systems (top) assumes $\sim 10^{10}$ C are needed to complete division, each transported sugar contains ≈ 5 C, and each transporter conducts sugar molecules at a rate of ≈ 200 per second. Bottom plot shows the estimated number of transporters needed at a growth rate of ≈ 0.5 per hr (light-brown point and dashed lines). Colored points correspond to the mean number of complexes involved in carbohydrate import (complexes annotated with the Gene Ontology terms GO:0009401 and GO:0098704) for different growth conditions across different published datasets. (B) The abundance of various specific carbon transport systems plotted as a function of the population growth rate. The rates of substrate transport differ between these transporter species. To compute the continuum growth rate estimate (grey line), we used the following transport rates for each transporter species: 200 glucose· s⁻¹ (BNID: 103693), 2000 glycerol· s⁻¹ (?), 200 fructose· s⁻¹ (assumed to be similar to PtsI, BNID: 103693), and 50 xylose· s⁻¹ (assumed to be comparable to LacY, BNID: 103159). Red points and highlighted text indicate conditions in which the only source of carbon in the growth medium induces expression of the transport system. Grey lines in (A) and (B) represents the estimated number of transporters per cell at a continuum of growth rates.

Figure 2–Figure supplement 1. Estimates and observed abundances of phosphate and sulfate transporters.

accommodate more transporters. There are, however, additional constraints on the space that can be devoted to nutrient uptake due to occupancy by proteins involved in processes like cell wall synthesis and energy production, and we will consider this further in the coming sections.

Cell Envelope Biogenesis

In contrast to nutrient transporters, which support the synthesis of biomolecules throughout the cell and therefore need to scale with the cell size, here we must consider the synthesis of components that will need to scale with the surface area of the cell. *E. coli* is a rod-shaped bacterium with a remarkably robust length-to-width aspect ratio of $\approx 4:1$ (??). Assuming this surface area is approximately the same between the inner and outer membranes of *E. coli*, and the fact that each membrane is itself a lipid bilayer, our rule-of-thumb of $5 \mu\text{m}^2$ per surface suggests a total membrane surface area of $\approx 20\mu\text{m}^2$ (see the Appendix Section "??" for a description of the calculation of cell surface area as a function of cell size). In this section, we will estimate the number of key protein complexes needed to synthesize the lipids as well as the complexes involved in assembling the peptidoglycan scaffold that make up the cell envelope.

Lipid Synthesis

The dense packing of the membrane with proteins means that the cell membranes are not composed entirely of lipid molecules, with only $\approx 40\%$ of the membrane area occupied by lipids (BNID: 100078). Using a rule-of-thumb of 0.5 nm^2 as the surface area of the typical lipid (BNID: 106993), we can estimate $\sim 2 \times 10^7$ lipids per cell, which is in close agreement with experimental measurements (BNID: 100071, 102996).

The membranes of *E. coli* are composed of a variety of different lipids, each of which are unique in their structures and biosynthetic pathways (?). Recently, a combination of stochastic kinetic modeling (?) and *in vitro* kinetic measurements (??) has revealed remarkably slow steps in the fatty acid synthesis pathways which may serve as the rate limiting reactions for making new membrane phospholipids in *E. coli*. One such step is the removal of hydroxyl groups from the fatty-acid chain by ACP dehydratase that leads to the formation of carbon-carbon double bonds. This reaction, catalyzed by proteins FabZ and FabA (?), have been estimated to have kinetic turnover rates of ≈ 1 dehydration per second per enzyme (?). Thus, given this rate and the need to synthesize $\approx 2 \times 10^7$ lipids over 5000 seconds, one can estimate that a typical cell requires ≈ 4000 ACP dehydratases. This is in reasonable agreement with the experimentally observed copy numbers of FabZ and FabA (??(A)). Furthermore, we can extend this estimate to account for the change in membrane surface area as a function of the growth rate (grey line in ??(A)), which captures the observed growth rate dependent expression of these two enzymes.

Peptidoglycan Synthesis

The exquisite control of bacteria over their cell shape is due primarily to a stiff, several nanometer thick meshwork of polymerized disaccharides that makes up the cell wall termed the peptidoglycan. The formation of the peptidoglycan is an intricate process involving many macromolecular players (??), whose coordinated action maintains cell shape and integrity even in the face of large-scale perturbations (??). The peptidoglycan alone comprises $\approx 3\%$ of the cellular dry mass (BNID: 1019360), making it the most massive molecule in *E. coli*. The polymerized unit of the peptidoglycan is a N-acetylglucosamine and N-acetylmuramic acid disaccharide, of which the former is functionalized with a short pentapeptide. With a mass of ≈ 1000 Da, this unit, which we refer to as a murein monomer, is polymerized to form long strands in the periplasm which are then attached to each other via their peptide linkers. Together, these quantities provide an estimate of $\approx 5 \times 10^6$ murein monomers per cell.

The crosslinking of the pentapeptides between adjacent glycan strands is responsible for maintaining the structural integrity of the cell wall and, in principle, each murein monomer can be involved in such a crosslink. In some microbes, such as in gram-positive bacterium *Staphylococcus aureus*, the extent of crosslinking can be large with $> 90\%$ of pentapeptides forming a connection between glycan strands. In *E. coli*, however, a much smaller proportion ($\approx 20\%$) of the peptides are crosslinked, resulting in a weaker and more porous cell wall (??). The formation of these crosslinks occurs primarily during the polymerization of the murein monomers and is facilitated by a family of enzymes called transpeptidases. The four primary transpeptidases of *E. coli* have only recently been quantitatively characterized *in vivo*, via liquid chromatography mass spectrometry, which revealed a notably slow kinetic turnover rate of ≈ 2 crosslinking reactions formed per second per enzyme (?).

Assembling these quantities permits us to make an estimate that on the order of ≈ 100 transpeptidases per cell are needed for complete maturation of the peptidoglycan, given a division time of ≈ 5000 seconds; a value that is comparable to experimental observations [??(B)]. Expanding this estimate to account for the changing mass of the peptidoglycan as a function of growth rate [grey line in ??(B)] predicts an order-of-magnitude increase in the abundance of the transpeptidases when the grow rate is increased

CELL ENVELOPE BIOSYNTHESIS

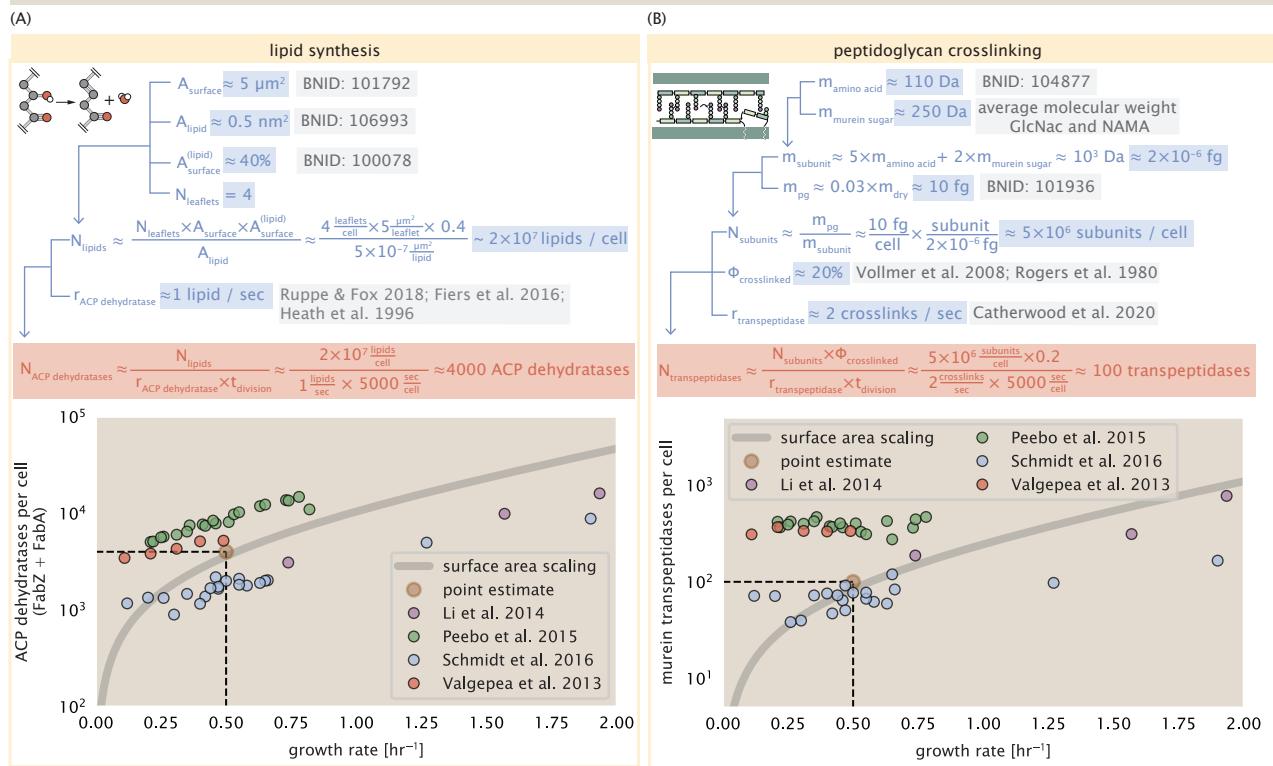


Figure 3. (A) Top panel shows an estimation for the number of ACP dehydratases necessary to form functional phospholipids, which is assumed to be a rate-limiting step on lipid synthesis. The rate of ACP dehydratases was inferred from experimental measurements via a stochastic kinetic model described in ?. Bottom panel shows the experimentally observed complex copy numbers using the stoichiometries $[\text{FabA}]_2$ and $[\text{FabZ}]_2$. (B) An estimate for the number of peptidoglycan transpeptidases needed to complete maturation of the peptidoglycan. The mass of the murein monomer was estimated by approximating each amino acid in the pentapeptide chain as having a mass of 110 Da and each sugar in the disaccharide having a mass of ≈ 250 Da. The *in vivo* rate of transpeptidation in *E. coli* was taken from recent analysis by ?. The bottom panel shows experimental measurements of the transpeptidase complexes in *E. coli* following the stoichiometries $[\text{MrcA}]_2$, $[\text{MrcB}]_2$, $[\text{MrdA}]_1$, and $[\text{MrdB}]_1$. Grey curves in each plot show the estimated number of complexes needed to satisfy the synthesis requirements scaled by the surface area as a function of growth rate.

by a factor of four. However, it is difficult to compare this to the observed complex abundances as systematic disagreements between the different data obfuscates any significant dependence on growth rate.

Limits on Cell Wall Biogenesis

While the processes we have considered represent only a small portion of proteins devoted to cell envelope biogenesis, we find it unlikely that they limit cellular growth in general. The relative amount of mass required for lipid and peptidoglycan components decrease at faster growth rates due to a decrease in their surface area to volume (S/V) ratio (?). Furthermore, despite the slow catalytic rate of FabZ and FabA in lipid synthesis, experimental data and recent computational modeling has shown that the rate of fatty-acid synthesis can be drastically increased by increasing the concentration of FabZ (??). With a proteome size of $\approx 3 \times 10^6$ proteins, a hypothetical 10-fold increase in expression from 4000 to 40,000 ACP dehydratases would result in a paltry $\approx 1\%$ increase in the size of the proteome. In the context of peptidoglycan synthesis, we note that our estimate considers only the transpeptidase enzymes that are involved in lateral and longitudinal elongation of the peptidoglycan. This neglects the presence of other transpeptidases that are present in the periplasm and also involved in remodeling and maturation of the peptidoglycan. It is therefore possible that if this was setting the speed limit for cell division, the simple expression of more transpeptidases may be sufficient to maintain the structural integrity of the cell wall.

Energy Production

Cells consume and generate energy predominantly in the form of nucleoside triphosphates (NTPs) in order to grow. The high-energy phosphodiester bonds of (primarily) ATP power a variety of cellular processes that drive biological systems away from thermodynamic equilibrium. We now turn to the synthesis of ATP as a potential process that may limit growth, which also requires us to consider the maintenance of the electrochemical proton gradient which powers it.

ATP Synthesis

Hydrolysis of the terminal phosphodiester bond of ATP into ADP (or alternatively GTP and GDP) and an inorganic phosphate provides the thermodynamic driving force in a wide array of biochemical reactions. One such reaction is the formation of peptide bonds during translation, which requires ≈ 2 ATPs for the charging of an amino acid to the tRNA and ≈ 2 GTPs for the formation of each peptide bond. Assuming the ATP costs associated with error correction and post-translational modifications of proteins are negligible, we can make the approximation that each peptide bond has a net cost of ≈ 4 ATP (BNID: 101442). Formation of GTP from ATP is achieved via the action of nucleoside diphosphate kinase, which catalyzes this reaction without an energy investment (?) and therefore consider all NTP requirements of the cell to be functionally equivalent to being exclusively ATP. In total, the energetic costs of peptide bond formation consume $\approx 80\%$ of the cells ATP budget [BNID: 107782; 106158; 101637; 111918, ??]. The pool of ATP is produced by the F₁-F₀ ATP synthase – a membrane-bound rotary motor which under ideal conditions can yield ≈ 300 ATP per second [BNID: 114701; ?].

To estimate the total number of ATP equivalents consumed during a cell cycle, we will make the approximation that there are $\approx 3 \times 10^6$ proteins per cell with an average protein length of ≈ 300 peptide bonds (BNID: 115702; 108986; 104877). Taking these values together, coupled with an estimate of ≈ 4 ATP equivalents per peptide bond, we find that the typical *E. coli* cell consumes $\sim 5 \times 10^9$ ATP per cell cycle on protein synthesis alone. Assuming that each ATP synthases operates at its maximal speed (300 ATP per second per synthase), ≈ 3000 ATP synthases are needed to keep up with the energy demands of the cell. This estimate is comparable with the experimental observations, shown in ??(A). We note that this estimate assumes all ATP is synthesized via ATP synthase and neglects synthesis via fermentative metabolism. This simplification may explain why at the fastest growth rates ($\approx 2 \text{ hr}^{-1}$), our continuum estimate predicts more synthase than is experimentally observed (gray line in ??). At rapid growth rates, *E. coli* enters a type of overflow metabolism where non-respiratory routes for ATP synthesis become pronounced (??).

Generating the Proton Electrochemical Gradient

In order to produce ATP, the F₁-F₀ ATP synthase itself must consume energy. Rather than burning through its own product (and violating thermodynamics), this intricate macromolecular machine has evolved to exploit the electrochemical potential established across the inner membrane through cellular respiration. This electrochemical gradient is manifest by the pumping of protons into the intermembrane space via the electron transport chains as they reduce NADH. In *E. coli*, this potential difference is ≈ -200 mV (BNID: 102120). A simple estimate of the inner membrane as a capacitor with a working voltage of -200 mV reveals that $\approx 2 \times 10^4$ protons must be present in the intermembrane space. However, each rotation of an ATP synthase shuttles ≈ 4 protons into the cytosol (BNID: 103390). With a few thousand ATP synthases producing ATP at their maximal rate, the potential difference would be rapidly abolished in a few milliseconds if it were not being actively maintained.

The electrochemistry of the electron transport complexes of *E. coli* have been the subject of intense biochemical and biophysical study (????). A recent work (?) examined the respiratory capacity of the *E. coli* electron transport complexes using structural and biochemical data, revealing that each electron transport chain rapidly pumps protons into the intermembrane space at a rate of ≈ 1500 protons per second (BNID: 114704; 114687). Using our estimate of the number of ATP synthases required per cell [??(A)], coupled with these recent measurements, we estimate that ≈ 3000 electron transport complexes would be necessary to facilitate the $\sim 5 \times 10^6$ protons per second diet of the cellular ATP synthases. This estimate is in agreement with the number of complexes identified in the proteomic datasets [plot in ??(B)]. This suggests that every ATP synthase must be accompanied by ≈ 1 functional electron transport chain.

Limits on Biosynthesis in a Crowded Membrane

Our estimates thus far have focused on biochemistry at the periphery of the cell. Since surface area and volume do not scale identically as cell size changes, in order to better understand the physical constraints on transport and energy production it is necessary to consider the consequence of a changing S/V ratio, which will decrease at faster growth rates. Here we use our analysis of ATP production to consider this constraint.

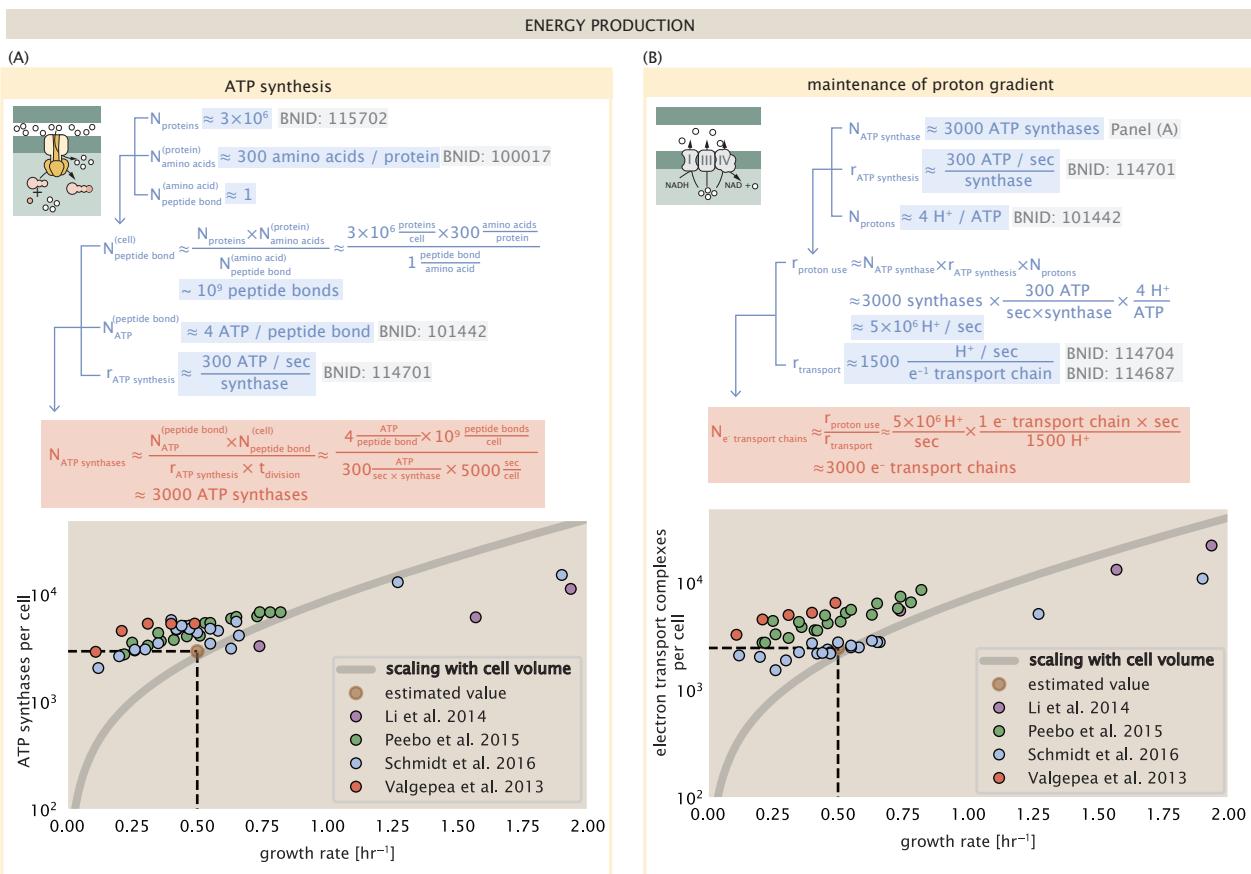


Figure 4. The abundance of F_1 - F_0 ATP synthases and electron transport chain complexes as a function of growth rate. (A) Estimate of the number of F_1 - F_0 ATP synthase complexes needed to accommodate peptide bond formation and other NTP dependent processes. Points in plot correspond to the mean number of complete F_1 - F_0 ATP synthase complexes that can be formed given proteomic measurements and the subunit stoichiometry $[AtpE]_{10}[AtpF]_2[AtpB][AtpC][AtpH][AtpA]_3[AtpG][AtpD]_3$. (B) Estimate of the number of electron transport chain complexes needed to maintain a membrane potential of -200 mV given estimate of number of F_1 - F_0 ATP synthases from (A). Points in plot correspond to the average number of complexes identified as being involved in aerobic respiration by the Gene Ontology identifier GO:0019646 that could be formed given proteomic observations. These complexes include cytochromes *bd1* ($[CydA][CydB][CydX][CydH]$), *bdII* ($[AppC][AppB]$), *bo₃*, $[CyoD][CyoA][CyoB][CyoC]$ and NADH:quinone oxioreduclease ($[NuqA][NuqH][NuqJ][NuqK][NuqL][NuqM][NuqO][NuqN][NuqB][NuqC][NuqE][NuqF][NuqG][NuqI]$) and II ($[Ndh]$). Grey lines in both (A) and (B) correspond to the estimate procedure described, but applied to a continuum of growth rates. We direct the reader to the Supporting Information for a more thorough description of this approach.

In our estimate of ATP production above we found that a cell demands about 5×10^9 ATP per cell cycle or 10^6 ATP/s. With a cell volume of roughly 1 fL (BNID: 100004), this corresponds to about 2×10^{10} ATP per fL of cell volume, in line with previous estimates (??). In ??(A) we plot this ATP demand as a function of the S/V ratio in green, where we have considered a range of cell shapes from spherical to rod-shaped with an aspect ratio (length/width) equal to 4. In order to consider the maximum ATP that could be produced, we consider the amount of ATP that can be generated by a membrane filled with ATP synthase and electron transport complexes and a maximal production rate of about 3 ATP / ($\text{nm}^2 \cdot \text{s}$) (?). This is shown in blue in ??(A), which shows that at least for the growth rates observed (right column in plot), the energy demand is roughly an order of magnitude less. Interestingly, ? found that ATP production by respiration is less efficient than by fermentation on a per membrane area basis, due to the additional proteins of the electron transport chain. This suggests that, even under anaerobic growth, cells will have sufficient membrane space for ATP production.

The analysis highlights that there will indeed be a maximum attainable cell size due to a diminishing capacity to provide resources as the cell increases in size. The maximum energy production in ??(A), however, does represent a somewhat unachievable limit since the inner membrane must also include other proteins such as those we've considered for nutrient transport and cell wall biogenesis. To better understand the overall proteomic makeup of the inner membrane, we therefore used Gene Ontology (GO) annotations (??) to identify all proteins embedded or peripheral to the inner membrane (GO term: 0005886). Those associated but not membrane-bound include proteins like MreB and FtsZ and must nonetheless be considered as a vital component occupying space on the membrane. In ??(B), we find that the total protein mass per μm^2 is nearly constant across growth rates. Interestingly, when we consider the distribution of proteins grouped by their Clusters of Orthologous Groups (COG) (?), the relative abundance of each category is relatively constant across growth rates. This suggests that no one process (energy production, nutrient uptake, etc.) is dominating even at fast growth rates [??(C)].

Processes of the Central Dogma

Up to this point, we have considered a variety of transport and biosynthetic processes that are critical to acquiring and generating new cell mass. While there are of course many other metabolic processes we could consider, we now turn our focus to some of the most important processes which *must* be undertaken irrespective of the growth conditions – those of the central dogma.

DNA Replication

Most bacteria (including *E. coli*) harbor a single, circular chromosome and can have extra-chromosomal plasmids up to ~ 100 kbp in length. While we consider the starting material dNTPs in ??-?? and discussed further in the Appendix Section "??", here we focus our quantitative thinking on the chromosome of *E. coli*, which harbors ≈ 5000 genes and $\approx 5 \times 10^6$ base pairs.

To successfully divide and produce viable progeny, this chromosome must be faithfully replicated and segregated into each nascent cell. Replication is initiated at a single region of the chromosome termed the *oriC* locus where a pair of replisomes, each consisting of two DNA polymerase III, begin their high-fidelity replication of the genome in opposite directions (?). *In vitro* measurements have shown that DNA Polymerase III copies DNA at a rate of ≈ 600 nucleotides per second (BNID: 104120). Therefore, to replicate a single chromosome, two replisomes moving at their maximal rate would copy the entire genome in ≈ 4000 s. Thus, with a division time of 5000 seconds, there is sufficient time for a pair of replisomes complexes to replicate the entire genome.

In rapidly growing cultures, bacteria like *E. coli* can initiate as many as 10 - 12 replication forks at a given time (??), we expect only a few DNA polymerases (≈ 10) are needed. However, as shown in ??, DNA polymerase III is nearly an order of magnitude more abundant. This discrepancy can be understood by considering its binding constant to DNA. *In vitro* characterization has quantified the K_D of DNA polymerase III holoenzyme to single-stranded and double-stranded DNA to be 50 and 200 nM, respectively (?). The right-hand plot in ?? shows that the concentration of DNA polymerase III across all data sets is within this range. Thus, its copy number appears to vary such that its concentration is approximately equal to the dissociation constant to the DNA. While the processes regulating the initiation of DNA replication are complex and involve more than just the holoenzyme, these data indicate that the kinetics of replication rather than the explicit copy number of the DNA polymerase III holoenzyme is the more relevant feature of DNA replication to consider. In light of this, the data in ?? suggests that for bacteria like *E. coli*, DNA replication does not represent a rate-limiting step in cell division. However, it is worth noting that for bacterium like *C. crescentus* whose chromosomal replication is initiated only once per cell cycle (?), the time to double their chromosome indeed represents an upper limit to their growth rate.

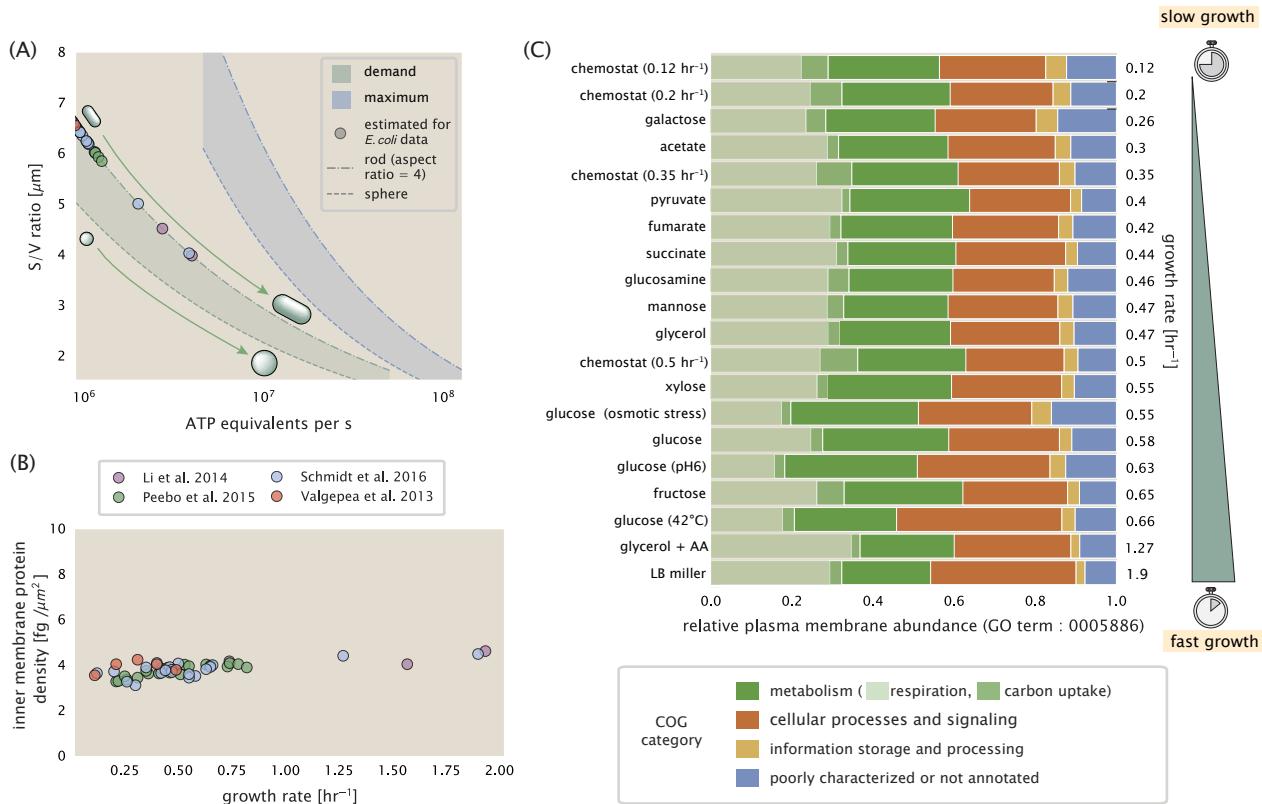


Figure 5. Influence of cell size and surface area to volume (S/V) ratio on ATP production and inner membrane composition. (A) Scaling of ATP demand and maximum ATP production through respiration as a function of S/V ratio. Cell volumes of 0.5 fL to 50 fL were considered, with the dashed (- -) line corresponding to a sphere and the dash-dot line (- -.) reflecting a rod-shaped bacterium like *E. coli* with a typical aspect ratio (length / width) of 4 (?). The ATP demand is calculated as 10⁶ ATP/(μm³ s), while the maximum ATP production rate is taken to be 3 ATP / (nm²·s) (?), with calculations of *E. coli* volume and surface area detailed in Appendix ???. In this calculation, 50% of the bacterial inner membrane is assumed to be protein, with the remainder lipid. (B) Total protein mass per μm² calculated for proteins with inner membrane annotation (GO term: 0005886). (C) Relative protein abundances by mass based on COG annotation. Metabolic proteins are further separated into respiration (F_1 - F_0 ATP synthase, NADH dehydrogenase I, succinate:quinone oxidoreductase, cytochrome bo₃ ubiquinol oxidase, cytochrome bd-I ubiquinol oxidase) and carbohydrate transport (GO term: GO:0008643). Note that the elongation factor EF-Tu can also associate with the inner membrane, but was excluded in this analysis due to its high relative abundance (roughly identical to the summed protein shown in part (B)).

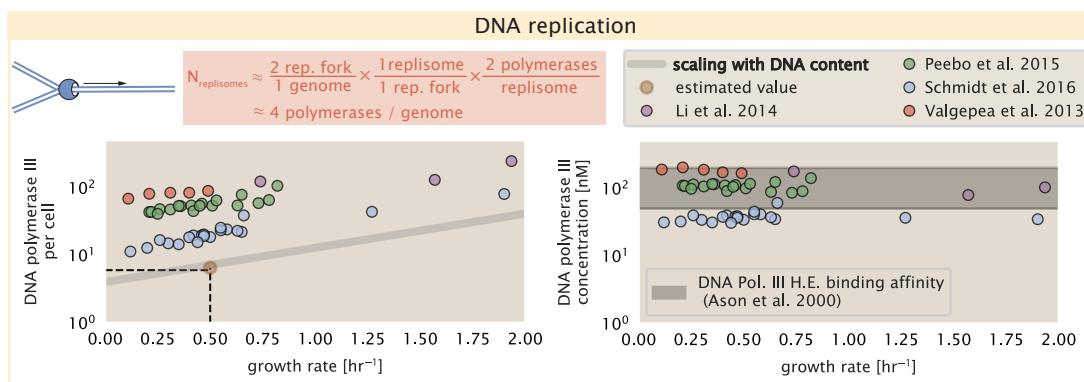


Figure 6. Complex abundance estimates for dNTP synthesis and DNA replication. An estimate for the minimum number of DNA polymerase holoenzyme complexes needed to facilitate replication of a single genome. Points in the left-hand plot correspond to the total number of DNA polymerase III holoenzyme complexes ([DnaE]₃[DnaQ]₃[HolE]₃[DnaX]₅[HolB][HolA][DnaN]₄[HolC]₄[HolD]₄) per cell. Right-hand plot shows the effective concentration of DNA polymerase III holoenzyme (See Appendix ?? for calculation of cell size). Grey lines in left-hand panel show the estimated number of complexes needed as a function of growth, the details of which are described in the Supplemental Information.

Figure 6–Figure supplement 1. Estimate and observations of the abundance of ribonucleotide reductase, a key component in dNTP synthesis.

RNA Synthesis

We now turn our attention to the next stage of the central dogma – the transcription of DNA to form RNA. We consider three major groupings of RNA, namely the RNA associated with ribosomes (rRNA), the RNA encoding the amino-acid sequence of proteins (mRNA), and the RNA which links codon sequence to amino-acid identity during translation (tRNA).

rRNA serves as the catalytic and structural component of the ribosome, comprising approximately 2/3 of the total ribosomal mass, and is decorated with ≈ 50 ribosomal proteins. Each ribosome contains three rRNA molecules of lengths 120, 1542, and 2904 nucleotides (BNID: 108093), meaning each ribosome contains ≈ 4500 nucleotides overall. *In vivo* measurements of the kinetics of rRNA transcription have revealed that RNA polymerases are loaded onto the promoter of an rRNA gene at a rate of ≈ 1 per second (BNID: 111997, 102362). If RNA polymerases are constantly loaded at this rate, then we can assume that ≈ 1 functional rRNA unit is synthesized per second per rRNA operon. While *E. coli* possesses 7 of these operons per chromosome, the fact that chromosome replication can be parallelized means that the average dosage of rRNA genes can be substantially higher (up to ≈ 70 copies) at fast growth rates (?). At a growth rate of $\approx 0.5 \text{ hr}^{-1}$, however, the average cell has ≈ 1 copy of its chromosome and therefore approximately ≈ 7 copies of the rRNA operons, producing ≈ 7 rRNA units per second. With a 5000 second division time, this means the cell is able to generate around 3×10^4 functional rRNA units, comparable within an order of magnitude to the number of ribosomes per cell.

How many RNA polymerases are then needed to constantly transcribe the required rRNA? If one polymerase is loaded per second, and the transcription rate is ≈ 40 nucleotides per second (BNID: 101094), then the typical spacing between polymerases will be ≈ 40 nucleotides. However, we must note that the polymerase itself has a footprint of ≈ 40 nucleotides (BNID: 107873), meaning that one could expect to find one RNA polymerase per 80 nucleotide stretch of an rRNA gene. With a total length of ≈ 4500 nucleotides per operon and 7 operons per cell, the number of RNA polymerases transcribing rRNA at any given time is then ≈ 500 per cell.

As outlined in ??, and discussed further the Appendix Section "??", synthesis of mRNA and tRNA together require on the order of ≈ 400 RNAP. Thus, in total, one would expect the typical cell to require ≈ 1000 RNAP to satisfy its transcriptional demands. As is revealed in ??(B), this estimate is about an order of magnitude below the observed number of RNA polymerase complexes per cell ($\approx 5000 - 7000$). The difference between the estimated number of RNA polymerase needed for transcription and these observations, however, are consistent with literature revealing that $\approx 80\%$ of RNA polymerases in *E. coli* are not transcriptionally active (?), with the majority non-specifically bound to DNA.

Our estimates also neglect other mechanistic features of transcription and transcriptional initiation more broadly. For example, we acknowledge that some fraction of the RNAP pool is nonspecifically bound to DNA during its search for promoters from which to begin transcription. Furthermore, we ignore the obstacles that RNA polymerase and DNA polymerase present to each other as they move along the DNA (?). Additionally, while they represent the core machinery for transcription, RNA polymerase is not sufficient to initiate transcription. Initiation of transcription is dependent on the presence of σ -factors, protein cofactors that bind directly to the polymerase (?) and aid in promoter recognition. In ??–??, we show that the predicted RNA polymerase copy number indeed is more comparable with the abundance of σ -70 (RpoD), the primary sigma factor in *E. coli*. There therefore remains more to be investigated as to what sets the observed abundance of RNA polymerase in these proteomic data sets. However, we conclude that the observed RNA polymerase abundances are generally in excess of what appears to be needed for growth, suggesting that the synthesis of RNA polymerase themselves are not particularly limiting.

Protein Synthesis

We conclude our dialogue between back-of-the-envelope estimates and comparison with the proteomic data by examining the final process in the central dogma – translation. In doing so, we will begin with an estimate of the number of ribosomes needed to replicate the cellular proteome. While the rate at which ribosomes translate is well known to be dependent on the growth rate (?), a phenomenon we consider later in this work we begin by making the approximation that translation occurs at a modest rate of ≈ 15 amino acids per second per ribosome (BNID: 100233). Under this approximation and our previous estimate of 10^9 peptide bonds per cell at a growth rate of 0.5 hr^{-1} , we can easily arrive at an estimate of $\approx 10^4$ ribosomes needed per cell to replicate the entire protein mass ??(A), top). This point estimate, as well as the corresponding estimate across a continuum of growth rates, proves to be notably comparable to the experimental observations, shown in the bottom panel of ??(A). While the ribosome is responsible for the formation of peptide bonds, we do not diminish the importance of charging tRNAs with their appropriate amino acid, a process which occurs with remarkable fidelity. In the Appendix and in ??–??, we consider the process of ligating tRNAs to their corresponding amino acid and again find notable accord between the data and our quantitative expectations.

Having completed our circuit through key processes of cellular growth outlined in ??, we can now take stock of our understanding

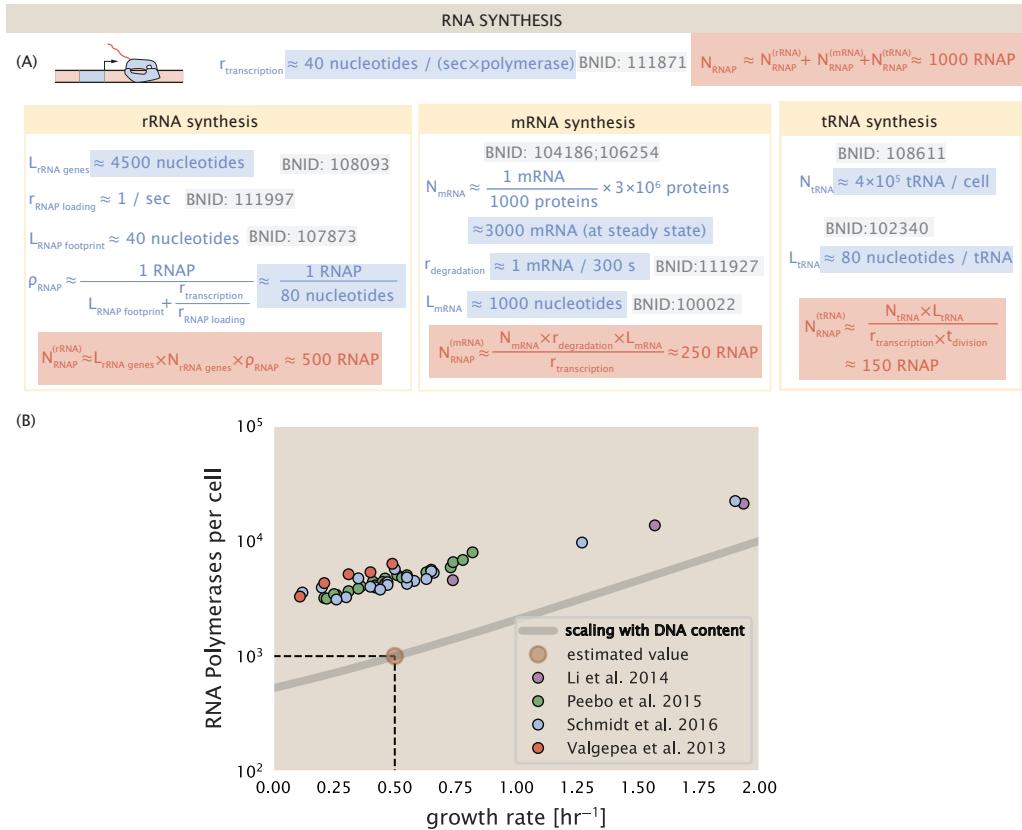


Figure 7. Estimation of the RNA polymerase demand and comparison with experimental data. (A) Estimations for the number of RNA polymerase needed to synthesize sufficient quantities of rRNA, mRNA, and tRNA from left to right, respectively. (B) The RNA polymerase core enzyme copy number as a function of growth rate. Colored points correspond to the average number RNA polymerase core enzymes that could be formed given a subunit stoichiometry of $[RpoA]_2[RpoC][RpoB]$.

Figure 7-Figure supplement 1. Abundance and growth rate dependence of σ -70.

of the observed growth rate dependence and abundances of various protein complexes. We note that, broadly speaking, these simple estimates have been reasonably successful in quantitatively describing the observations in the proteomic data, suggesting that the proteome is tuned in composition and absolute abundance to match the growth rate requirements without any one process representing a singular bottleneck or rate limiting step in division. However, in our effort to identify key limitations on growth, there are two notable observations that we wish to emphasize.

The first is a recurring theme throughout the estimates investigated here, which is that any inherent biochemical rate limitation can be overcome by expressing more proteins. We can view this as a parallelization of each biosynthesis task, which helps explain why bacteria tend to increase their protein content (and cell size) as growth rate increases (?). The second, and ultimately the most significant in defining the cellular growth rate, is that the synthesis of ribosomal proteins presents a special case where parallelization is *not* possible and thereby imposes a limit on the fastest possible growth rate. Each ribosome has ≈ 7500 amino acids across all of its protein components which must be strung together as peptide bonds through the action of another ribosome. Once again using a modest elongation rate of ≈ 15 amino acids per second, we arrive at an estimate of ≈ 500 seconds or ≈ 7 minutes to replicate a single ribosome. This limit, as remarked upon by others (?), serves as a hard theoretical boundary for how quickly a bacterium like *E. coli* can replicate. As each ribosome would therefore need to copy itself, this 7 minute speed limit is independent of the number of ribosomes per cell [??(B)], yet assumes that the only proteins that need to be replicated for division to occur are ribosomal proteins, a regime not met in biological reality. This poses an optimization problem for the cell – how are the translational demands of the entire proteome met without investing resources in the production of an excess of ribosomes?

This question, more frequently presented as a question of optimal resource allocation, has been the target of an extensive dialogue between experiment and theory over the past decade. In a now seminal work, ? present an elegant treatment of resource allocation through partitioning of the proteome into sectors – one of which being ribosome-associated proteins whose relative size ultimately defines the total cellular growth rate. In more recent years, this view has been more thoroughly dissected experimentally (?????) and together have led to a paradigm-shift in how we think of cellular physiology at the proteomic-level. However, the quantitative description of these observations is often couched in terms of phenomenological constants and effective parameters with the key observable features of expression often computed in relative, rather than absolute, abundances. Furthermore, these approaches often exclude or integrate away effects of cell size and chromosome content, which we have found through our estimates to have important connections to the observed cellular growth rate.

In the closing sections of this work, we explore how ribosomal content, total protein abundance, and chromosomal replication are intertwined in their control over the cellular growth rate. To do so, we take a more careful view of ribosome abundance, increasing the sophistication of our analysis by exchanging our order-of-magnitude estimates for a minimal mathematical model of growth rate control. This is defined by parameters with tangible connections to the biological processes underlying cellular growth and protein synthesis. Using this model, we interrogate how the size of the ribosome pool and its corresponding translational capacity enable cells to maintain a balance between the supply of amino acids via metabolism and catabolism and their consumption through the peptide bond formation required for growth.

Maximum Growth Rate is Determined by the Ribosomal Mass Fraction

The 7 minute speed limit shown in ??(B) assumes all proteins in the cell are ribosomes. In order to connect this to the experimental data (and physiological reality more broadly), we first need to relax this assumption and determine a translation-limited growth rate. Here, we will assume that the cell is composed of N_{pep} peptide bonds and R ribosomes, whose precise values will depend on the growth rate λ . The protein subunits of each ribosomal protein sum to a total of ≈ 7500 amino acids as noted earlier, which we denote by L_R . With an average mass of an amino acid of $m_{\text{AA}} \approx 110$ Da (BNID: 104877), the total ribosomal mass fraction Φ_R is given by

$$\Phi_R = \frac{m_{\text{ribosomes}}}{m_{\text{proteome}}} \approx \frac{m_{\text{AA}} \times R \times L_R}{m_{\text{AA}} \times N_{\text{pep}}} = \frac{R \times L_R}{N_{\text{pep}}}. \quad (1)$$

For exponentially growing cells (?), the rate of cellular growth will be related to the rate of protein synthesis via

$$\lambda N_{\text{pep}} = r_t \times R \times f_a, \quad (2)$$

where r_t is the translation rate. Here, we've introduced a multiplicative factor f_a which represents the fraction of the ribosomes that are actively translating. This term allows us to account for immature or non-functional ribosomes or active sequestration of ribosomes through the action of the secondary messenger alarmone (p)ppGpp in poorer nutrient conditions (?).

Combining ?? and ?? results in an expression for a translation-limited growth rate, which is given by

$$\lambda_{\text{translation-limited}} = \frac{r_t \times \Phi_R \times f_a}{L_R}. \quad (3)$$

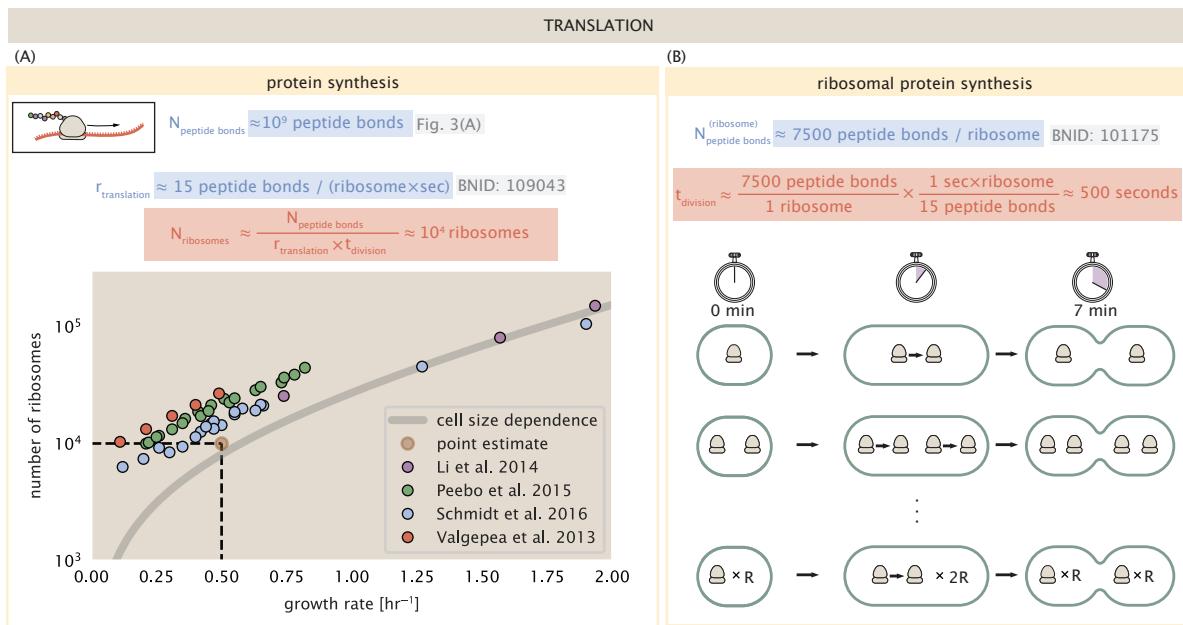


Figure 8. Estimation of the required number of ribosomes and the speed limit for bacterial replication. (A) Estimation of the number of ribosomes required to synthesize 10^9 peptide bonds with an elongation rate of 15 peptide bonds per second. The average abundance of ribosomes is plotted as a function of growth rate. Our estimated values are shown for a growth rate of 0.5 hr^{-1} . Grey lines correspond to the estimated complex abundance calculated at different growth rates. (B) Estimation for the time to replicate a ribosome. This rate is independent of the number of ribosomes R and instead is limited by the time required to double an individual ribosome.

Figure 8—Figure supplement 1. Estimate and observed abundance and growth rate dependence of tRNA ligases.

This result, derived in a similar manner by others (??), reflects mass-balance under steady state growth and has long provided a rationalization of the apparent linear increase in *E. coli*'s ribosomal content as a function of growth rate (????). The left-hand panel of ??(A) shows this growth rate plotted as a function of the ribosomal mass fraction. In the regime where all ribosomes are active ($f_a = 1$) and the entire proteome is composed of ribosomal proteins ($\Phi_R = 1$), indeed, we arrive at the maximum theoretical growth rate of r_t/L_R , and $\approx 7 \text{ min}$ for *E. coli*.

Connecting ?? to the proteomic data serving as the centerpiece of our work, however, requires knowledge of f_a at each growth rate as proteomic measurements only provide a measure of Φ_R . Recently, ? determined f_a as a function of the growth rate (??(A), right-hand panel, inset), revealing that $f_a \approx 1$ at growth rates above 0.75 hr^{-1} and $f_a < 1$ as the growth rate slows. Using these data, we inferred the approximate active fraction (see the Appendix Section "??") at each growth rate and used this to compute $\Phi_R \times f_a$ (??(A), colored points in right-hand panel). In general, these data skirt the translation-limited growth rate determined using ??, where we have taken r_t to be the maximal elongation rate of 17 amino acids per second measured by ?. There is a notable discrepancy between the data collected in ?? and that collected from ?. When compared to other measurements (non-proteomic based) of the active ribosome mass fraction (??(B), grey points in right-hand panel), the data from ? and ? are notably different, suggesting a systematic error in these data. These additional measurements come from a number of recent studies and are determined from measurements of total RNA to total protein mass ratios (??-??).

Together, these results illustrate that the growth rates observed across the amalgamated data sets are close to the translation-limited growth rate determined through ribosomal activity, at least for the data reported in ? and ?. While this is a useful framework to consider how the relative abundance of ribosomes (compared to all other proteins) defines the growth rate, it is worth noting that as growth rate increases, so does the cell size and therefore so will the total proteomic mass (?). With a handle on how elongation rate and the total number of peptide bonds per proteome is related to the growth rate, we now expand this description to account for the increasing cell size and ribosome copy number at faster growth rates, enabling us to identify a potential bottleneck in the synthesis of rRNA.

rRNA Synthesis Presents a Potential Bottleneck During Rapid Growth

Even under idealized experimental conditions, *E. coli* rarely exhibits growth rates above 2 hr^{-1} (?), which is still well-below the synthesis rate of a single ribosome, and below the maximum growth rates reported for several other bacteria (?). While we have

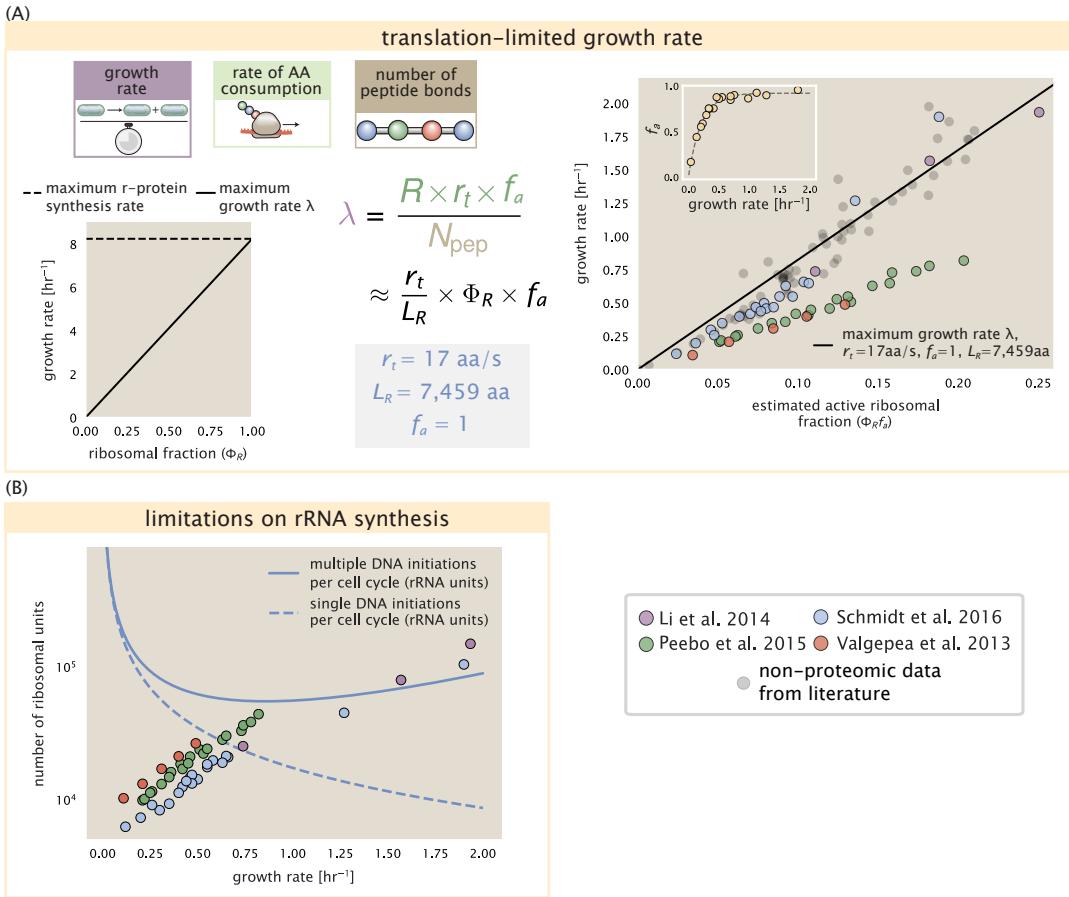


Figure 9. Translation-limited growth rate. (A) left: Translation-limited growth as a function of the ribosomal fraction. The solid line is calculated for an elongation rate of 17 aa per second. The dashed line corresponds to the maximum rate of ribosomal protein synthesis ($\approx 7 \text{ min}$). right: Translation-limited growth rate as a function of the actively translating ribosomal fraction. The actively translating ribosomal fraction is calculated using the estimated values of f_a from ? (shown in inset; see Appendix ?? for additional detail). Gray data points show additional measurements from literature and considered further in the supplemental figure part (A). (B) Maximum number of rRNA units that can be synthesized as a function of growth rate. Solid curve corresponding to the rRNA copy number is calculated by multiplying the number of rRNA operons by the estimated number of (# ori) at each growth rate. The quantity (# ori) was calculated using Equation 4 and the measurements from ?. The dashed line shows the maximal number of functional rRNA units produced from a single chromosomal initiation per cell cycle.

Figure 9–Figure supplement 1. Comparison of $\Phi_R f_a$ with literature and estimation of (# ori).

considered potential limits imposed by translation of ribosomal *proteins*, here we consider potential limiting regimes specific to the synthesis of rRNA.

Due to multiple initiations of chromosomal replication per cell doubling, the effective number of rRNA operons increases with growth rate and will do so in proportion to the average number of origins per cell, $\langle \# \text{ ori} \rangle$. This later parameter is set by how often replication must be initiated in order to keep up with cell doubling times τ , whose time may be shorter than the cell cycle time τ_{cyc} (referring to the time from replication initiation to cell division) (??). This is quantified by

$$\langle \# \text{ ori} \rangle = 2^{\tau_{\text{cyc}}/\tau} = 2^{\tau_{\text{cyc}} A / \log(2)}, \quad (4)$$

where the doubling time $s_0\tau$ is related to the growth rate by $\tau_{\text{au}} = \log(2)/\lambda$. We used the experimental measurements of τ_{cyc} (the timescale of chromosome replication and cell division) and τ (the timescale of a cell doubling) from ? (??-??(B)) to calculate $\langle \# \text{ ori} \rangle$ with ?? as a function of growth rates. For growth rates above about 0.5 hr^{-1} , t_{cyc} is approximately constant at about 70 minutes, implying that $\langle \# \text{ ori} \rangle$ will grow exponentially with growth rates beyond 0.5 hr^{-1} . As the rRNA operons are predominantly located close to the origin of replication (BNID: 100352), we make the simplifying assumption that that the number of rRNA operons will be directly proportional to $\langle \# \text{ ori} \rangle$.

Returning to our rule-of-thumb of one functional rRNA unit is produced per second per transcribing operon, we estimate the maximum number of ribosomes that could be made as a function of growth rate (??(B), blue curve). Although we expect this estimate to significantly overestimate rRNA abundance at slower growth rates ($\lambda < 0.5 \text{ hr}^{-1}$), this provides a useful reference alongside the proteomic measurements, particularly in the regime of fast growth. For growth rates above about 1 hr^{-1} , for example, we find that cells will need to transcribe rRNA near their maximal rate. As a counter example, if *E. coli* did not initiate multiple rounds of replication, but managed to replicate their chromosome within the requisite time limit, they would be unable to make enough rRNA for the observed number of ribosomes (dashed blue curve in ??(C)). The convergence between the maximum rRNA production and measured ribosome copy number suggests rRNA synthesis may begin to present a bottleneck at the fastest growth rates due to the still-limited copies of rRNA genes.

Rapid Growth Requires *E. coli* to Increase Both Cell Size and Ribosomal Mass Fraction

In the right-hand side of ??(B), we also find that above about 0.75 hr^{-1} , the growth rate is determined solely by the ribosomal mass fraction Φ_R , since f_a is close to 1, and r_t is near its maximal rate (?). While Φ_R will need to increase in order for cells to grow faster, the fractional dependence in ?? gives little insight into how this scaling is actually achieved by the cell.

It is now well-documented that *E. coli* cells add a constant volume per origin of replication, which is robust to a remarkable array of cellular perturbations (?). Given the proteomic measurements featured in this work, we find that the ribosome copy number also scales in proportion to $\langle \# \text{ ori} \rangle$ (??(A)). However, an increase in ribosome abundance alone is not necessarily sufficient to increase growth rate and we also need to consider how Φ_R varies with $\langle \# \text{ ori} \rangle$. Importantly, as shown in ??(B), we find that the deviations in protein expression with $\langle \# \text{ ori} \rangle$ are largely restricted to regions of ribosomal protein genes ??(B). Here we have calculated the position-dependent protein expression across the chromosome by a running Gaussian average of protein copy number (20 kbp st. dev. averaging window) based on each gene's transcriptional start site. These were median-subtracted to account for the change in total protein abundance with $\langle \# \text{ ori} \rangle$. This result suggests that Φ_R is also being tuned in proportion to $\langle \# \text{ ori} \rangle$ under nutrient-limited growth, and in particular, it is through this additional dependence on Φ_R , combined with the exponential increase in $\langle \# \text{ ori} \rangle$, that *E. coli* exhibits an exponential increase in cell size with growth rate.

A Minimal Model of Nutrient-Mediated Growth Rate Control

While the preceeding subsections highlight a dominant role for ribosomes in setting the growth rate, our analysis on the whole emphasizes how the total proteomic content also changes in response to variable growth conditions and growth rate. In this final section we use a minimal model of growth rate control to better understand how this interconnection between ribosomal abundance and total protein abundance influences the observed growth rate.

Here we propose that cells modulate their protein abundance in direct response to the availability of nutrients in their environment. As noted earlier, bacteria can modulate ribosomal activity through the secondary messenger molecules like (p)ppGpp in poorer nutrient conditions (??(C) - inset; ?). Importantly, these secondary messengers also cause global changes in transcriptional and translational activity (??). In *E. coli*, amino acid starvation leads to the accumulation of de-acetylated tRNAs at the ribosome's A-site and a strong increase in (p)ppGpp synthesis activity by the enzyme RelA (?). Along with this, there is increasing evidence that (p)ppGpp also acts to inhibit the initiation of DNA replication (?), providing a potential mechanism for cells to lower $\langle \# \text{ ori} \rangle$ and maintain a smaller cell size in poorer nutrient conditions (?).

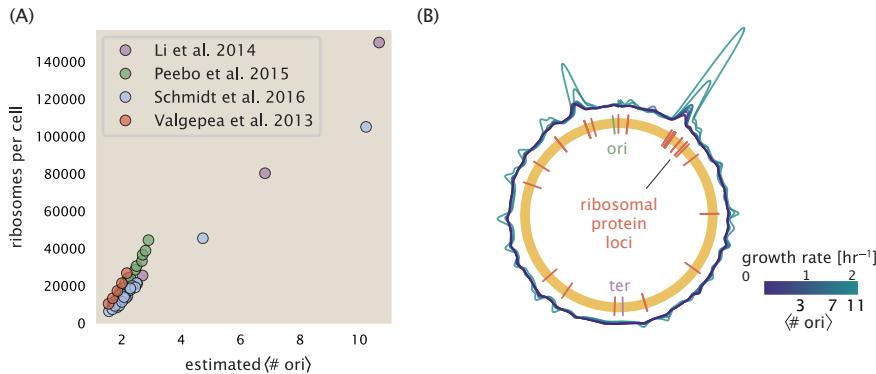


Figure 10. Cells increase both absolute ribosome abundance and Φ_R with (# ori). (A) Plot of the ribosome copy number estimated from the proteomic data against the estimated (# ori) (see Appendix ?? for additional details). (B) A running Gaussian average (20 kbp st. dev.) of protein copy number is calculated for each growth condition considered by (?) based on each gene's transcriptional start site. Since total protein abundance increases with growth rate, protein copy numbers are median-subtracted to allow comparison between growth conditions. (# ori) are estimated using the data in (A) and Equation ??.

To consider this quantitatively, we assume that cells modulate their proteome (N_{pep} , R , Φ_R) to better maximize their rate of peptide elongation r_t . The elongation rate r_t will depend on how quickly the ribosomes can match codons with their correct aminoacyl tRNA, along with the subsequent steps of peptide bond formation and translocation. This ultimately depends on the cellular concentration of amino acids, which we treat as a single effective species, $[AA]_{\text{eff}}$. In our model, we determine the the rate of peptide elongation r_t and achievable growth rate as simply depending on the supply of amino acids (and, therefore, also aminoacyl tRNAs), through a parameter r_{AA} in units of AA per second, and the rate of amino acid consumption by protein synthesis ($r_t \times R \times f_a$). This is shown schematically in ??(A) and derived in the Appendix Section "??". Given our observation that protein synthesis and energy production are not limiting, we assume that other molecular players required by ribosomes such as elongation factors and GTP are available in sufficient abundance.

In ??(B), we illustrate how the elongation rate will depend on the ribosomal copy number. Here, we have considered an arbitrarily chosen $r_{AA} = 5 \times 10^6 \text{ AA} \cdot s^{-1} \cdot \mu\text{m}^{-3}$ and $f_a = 1$ for a unit cell volume $V = 1\text{fL}$ (we provide the interactive figure ??-?? which allows the user to explore different regimes of this parameter space). At low ribosome copy numbers, the observed elongation rate is dependent primarily on $[AA]_{\text{eff}}$ through r_{AA} [as $r_t^{\max} \times R \times f_a \ll r_{AA}$, point (1) in ??(B)]. As the ribosome copy number is increased such that the amino acid supply rate and consumption rate are nearly equal [point (2) in ??(B)], the observed elongation rate begins to decrease sharply. When the ribosome copy number is increased even further, consumption at the maximum elongation rate exceeds the supply rate, yielding a significantly reduced elongation rate [point (3) in ??(B)]. While the elongation rate will always be dominated by the amino acid supply rate at sufficiently low ribosome copy numbers, the elongation rate at larger ribosome abundances can be increased by tuning f_a such that not all ribosomes are elongating, reducing their total consumption rate.

Optimal Ribosomal Content and Cell Size Depend on Nutrient Availability and Metabolic Capacity

To relate elongation rate to growth rate, we constrain the set of parameters based on our available proteomic measurements; namely, we restrict the values of R , N_{pep} , and cell size V to those associated with the amalgamated proteomic data (described in the Appendix Section "??"). We then consider how changes in the nutrient conditions, through the parameter r_{AA} , influence the maximum growth rate as determined by ???. ??(C) shows how the growth rate depends on the rate of amino acid supply r_{AA} as a function of the cellular ribosome copy number. A feature immediately apparent is the presence of a maximal growth rate whose dependence on R (and consequently, the cell size) increases with increasing r_{AA} . Importantly, however, there is an optimum set of R , N_{pep} , and V that are strictly dependent on the value of r_{AA} . This shows that increasing the ribosomal concentration beyond the cell's metabolic capacity will have the adverse consequence of depleting the supply of amino acids $[AA]_{\text{eff}}$ and a concomitant decrease in the elongation rate r_t (??(B)) and growth rate.

Also of note is the growth rate trends observed at low values of r_{AA} [purple and blue lines in ??(C)], representative of growth in nutrient-poor media. In these conditions, there no longer exists a peak in growth, at least within the range of physiologically-relevant ribosome copy numbers. This is a regime, associated with slower growth rates, where cells limit their pool of actively translating ribosomes by decreasing f_a (?), which we find would help maintain the pool of available amino acids $[AA]_{\text{eff}}$ and increase the achievable elongation rate.

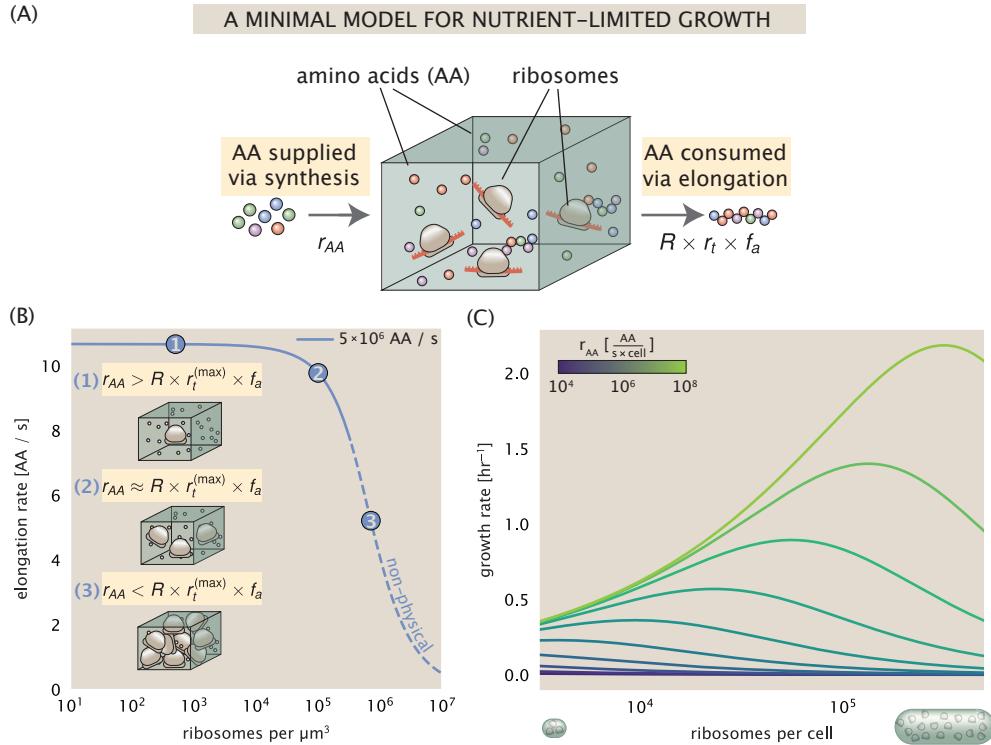


Figure 11. A minimal model of growth rate control under nutrient limitation. (A) We consider a unit volume of cellular material composed of amino acids (colored spheres) provided at a supply rate r_{AA} . These amino acids are polymerized by a pool of ribosomes (brown blobs) at a rate $r_t \times R \times f_a$, where r_t is the elongation rate, R is the ribosome copy number in the unit volume, and f_a is the fraction of those ribosomes actively translating. (B) The observed elongation rate is plotted as a function of the number of ribosomes. The three points correspond to three regimes of ribosome copy numbers and are shown schematically on the left-hand side. The region of the curve shown as dashed lines represents a non-physical copy number, but is shown for illustrative purposes. This curve was generated using an amino acid supply rate of $5 \times 10^6 \text{ AA / s}$, a maximal elongation rate of 17.1 AA / s , $f_a = 1$, and a unit cell volume of $V = 1 \text{ fL}$. See Appendix ?? for additional model details. (C) The cellular growth rate is plotted as a function of total cellular ribosome copy number for different cellular amino acid supply rates, with blue and green curves corresponding to low and high supply rates, respectively. As the ribosome copy number is increased, so too is the cell size and total protein abundance N_{pep} . We direct the reader to the Supplemental Information for discussion on the inference of the relationship between cell size, number of peptide bonds, and ribosome copy number.

Figure 11–Figure supplement 1. An interactive figure for exploration of the model parameter space.

Discussion

Continued experimental and technological improvements have led to a treasure trove of quantitative biological data (??????), and an ever advancing molecular view and mechanistic understanding of the constituents that support bacterial growth (??????). In this work we have compiled what we believe to be the state-of-the-art knowledge on proteomic copy number across a broad range of growth conditions in *E. coli*. We have made this data accessible through a [GitHub repository](#), and an [interactive figure](#) that allows exploration of specific protein and protein complex copy numbers. Through a series of order-of-magnitude estimates that traverse key steps in the bacterial cell cycle, this proteomic data has been a resource to guide our understanding of two key questions: what biological processes limit the absolute speed limit of bacterial growth, and how do cells alter their molecular constituents as a function of changes in growth rate or nutrient availability? While not exhaustive, our series of estimates provide insight on the scales of macromolecular complex abundance across four classes of cellular processes – the transport of nutrients, the production of energy, the synthesis of the membrane and cell wall, and the numerous steps of the central dogma.

In general, the copy numbers of the complexes involved in these processes were in reasonable agreement with our order-of-magnitude estimates. Since many of these estimates represent soft lower-bound quantities, this suggests that cells do not express proteins grossly in excess of what is needed for a particular growth rate. Several exceptions, however, also highlight the dichotomy between a proteome that appears to "optimize" expression according to growth rate and one that must be able to quickly adapt to environments of different nutritional quality. Take, for example, the expression of carbon transporters. Shown in ??(B), we find that cells always express a similar number of glucose transporters irrespective of growth condition. At the same time, it is interesting to note that many of the alternative carbon transporters are still expressed in low but non-zero numbers ($\approx 10\text{-}100$ copies per cell) across growth conditions. This may relate to the regulatory configuration for many of these operons, which require the presence of a metabolite signal in order for alternative carbon utilization operons to be induced (??). Furthermore, upon induction, these transporters are expressed and present in abundances in close agreement with a simple estimate.

Of the processes illustrated in ??, we arrive at a ribosome-centric view of cellular growth rate control. This is in some sense unsurprising given the long-held observation that *E. coli* and many other organisms vary their ribosomal abundance as a function of growth conditions and growth rate (??). However, through our dialogue with the proteomic data, two additional key points emerge. The first relates to our question of what process sets the absolute speed limit of bacterial growth. While a cell can parallelize many of its processes simply by increasing the abundance of specific proteins or firing multiple rounds of DNA replication, this is not so for synthesis of ribosomes [??(A)]. The translation time for each ribosome [≈ 7 min, ?] places an inherent limit on the growth rate that can only be surpassed if the cell were to increase their polypeptide elongation rate, or if they could reduce the total protein and rRNA mass of the ribosome. The second point relates to the long-observed correlations between growth rate and cell size (??), and between growth rate and ribosomal mass fraction. While both trends have sparked tremendous curiosity and driven substantial amounts of research in their own regards, these relationships are themselves intertwined. In particular, it is the need for cells to increase their absolute number of ribosomes under conditions of rapid growth that require cells to also grow in size. Further experiments are needed to test the validity of this hypothesis. In particular, we believe that the change in growth rate in response to translation-inhibitory drugs (such as chloramphenicol) could be quantitatively predicted, given one had precision measurement of the relevant parameters, including the fraction of actively translating ribosomes f_a and changes in the metabolic capacity of the cell (i.e. the rate that amino acids can be made available) for a particular growth condition.

While the generation of new ribosomes plays a dominant role in growth rate control, there exist other physical limits to the function of cellular processes. One of the key motivations for considering energy production was the physical constraints on total volume and surface area as cells vary their size (??). As *E. coli* get larger at faster growth rates, an additional constraint begins to arise in energy production and nutrient uptake due to the relative decrease in total surface area, where ATP is predominantly produced (?). Specifically, the cell interior requires an amount of energy that scales cubically with cell size, but the available surface area only grows quadratically [??(A)]. While this threshold does not appear to be met for *E. coli* cells growing at 2 hr^{-1} or less, it highlights an additional constraint on growth given the apparent need to increase cell size in order to grow faster. This limit is relevant even to eukaryotic organisms, whose mitochondria exhibit convoluted membrane structures that nevertheless remain bacteria-sized organelles (?). In the context of bacterial growth and energy production more generally, we have mainly limited our analysis to the aerobic growth conditions associated with the proteomic data and further consideration will be needed for anaerobic growth.¹

This work is by no means meant to be a complete dissection of bacterial growth rate control, and there are many aspects of the bacterial proteome and growth that we neglected to consider. For example, other recent work (???) has explored how the proteome is structured and how that structure depends on growth rate. In the work of ?, the authors coarse-grained the proteome into six discrete categories being related to either translation, catabolism, anabolism, and others related to signaling and core metabolism.

The relative mass fraction of the proteome occupied by each sector could be modulated by external application of drugs or simply by changing the nutritional content of the medium. While we have explored how the quantities of individual complexes are related to cell growth, we acknowledge that higher-order interactions between groups of complexes or metabolic networks at a systems-level may reveal additional insights into how these growth-rate dependences are mechanistically achieved. Furthermore, while we anticipate the conclusions summarized here are applicable to a wide collection of bacteria with similar lifestyles as *E. coli*, other bacteria and archaea may have evolved other strategies that were not considered. Further experiments with the level of rigor now possible in *E. coli* will need to be performed in a variety of microbial organisms to learn more about how regulation of proteomic composition and growth rate control has evolved over the past 3.5 billion years.

Methods

Data Analysis and Availability

All proteomic measurements come from the experimental work of ??? (mass spectrometry) and ? (ribosomal profiling). Data curation and analysis was done programmatically in Python, and compiled data and analysis files are accessible through a [GitHub repository](#) (DOI:XXX) associated with this paper as well as on the associated [paper website](#). Additionally, we provide two interactive figures that allow for [rapid exploration of the compiled data sets](#) as well as [exploration of the parameter space of the minimal model](#).

Acknowledgements

We thank Matthias Heinemann, Alexander Schmidt, and Gene-Wei Li for additional input regarding their data. We also thank members of the Phillips, Theriot, Kondek, Garcia labs, as well as Ron Milo and Terry Hwa for useful discussions. We thank Suzannah M. Beeler, Avi Flamholz, Soichi Hirokawa, and Manuel Razo-Mejia for reading and providing comments on drafts of this manuscript. R.P. is supported by La Fondation Pierre-Gilles de Gennes, the Rosen Center at Caltech, and the NIH 1R35 GM118043 (MIRA). J.A.T. is supported by the Howard Hughes Medical Institute, and NIH Grant R37-AI036929. N.M.B is a HHMI Fellow of The Jane Coffin Childs Memorial Fund. H.G.G. is supported by the Burroughs Wellcome Fund Career Award at the Scientific Interface, the Sloan Research Foundation, the Human Frontiers Science Program, the Searle Scholars Program, the Shurl & Kay Curci Foundation, the Hellman Foundation, the NIH Director's New Innovator Award (DP2 OD024541-01), and an NSF CAREER Award (1652236). D.S.F. is supported by an NSF award (PHY-1607606).

Competing Interests

The authors declare no competing interests.

main_figs/fig2-S1_phospho_sulfo_transport.pdf

Figure 2-Figure supplement 1. (A) Estimate for the number of PitA phosphate transport systems needed to maintain a 3% phosphorus *E. coli* dry mass. Points in plot correspond to the total number of PitA transporters per cell. (B) Estimate of the number of CysUWA complexes necessary to maintain a 1% sulfur *E. coli* dry mass. Points in plot correspond to average number of CysUWA transporter complexes that can be formed given the transporter stoichiometry [CysA]₂[CysU][CysW][Sbp/CysP]. Grey line in (A) and (B) represents the estimated number of transporters per cell at a continuum of growth rates.

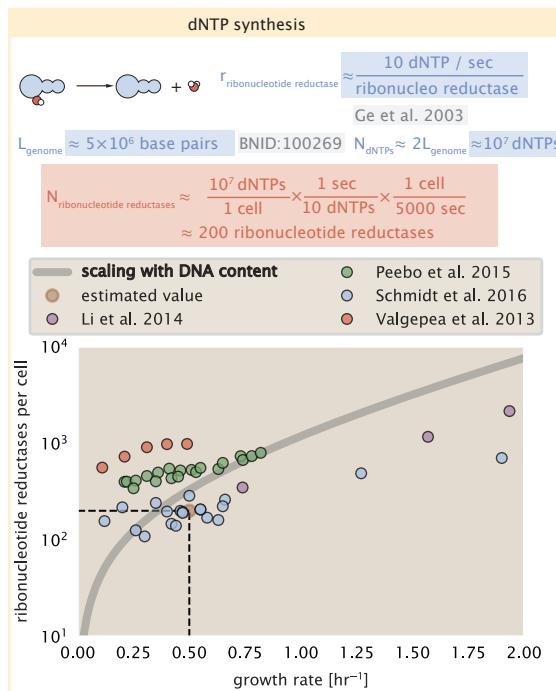


Figure 6-Figure supplement 1. Estimate of the number of ribonucleotide reductase enzymes needed to facilitate the synthesis of $\approx 10^7$ dNTPs over the course of a 5000 second generation time. Points in the plot correspond to the total number of ribonucleotide reductase I ($[\text{NrdA}]_2[\text{NrdB}]_2$) and ribonucleotide reductase II ($[\text{NrdE}]_2[\text{NrdF}]_2$) complexes. Grey lines in top panel show the estimated number of complexes needed as a function of growth, the details of which are described in the Appendix.

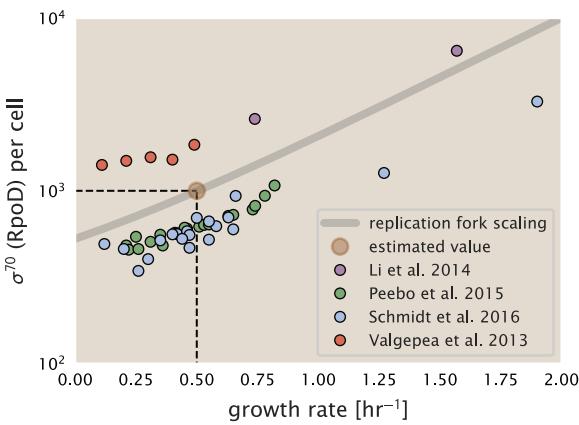


Figure 7-Figure supplement 1. The abundance of σ^{70} as a function of growth rate. Estimated value for the number of RNAP is shown as a translucent brown point and grey line.

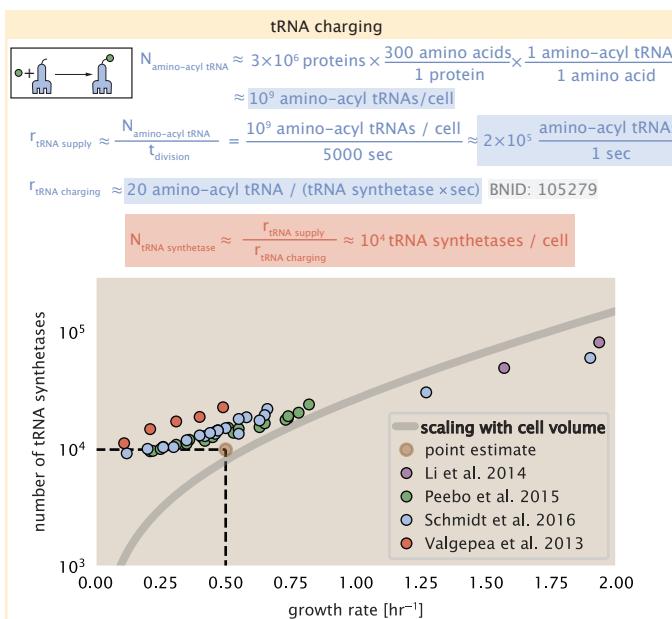


Figure 8-Figure supplement 1. Estimation for the number of tRNA synthetases that will supply the required amino acid demand. The sum of all tRNA synthetases copy numbers are plotted as a function of growth rate ([ArgS], [CysS], [GlnS], [GltX], [IleS], [LeuS], [ValS], [AlaS]₂, [AsnS]₂, [AspS]₂, [TyrS]₂, [TrpS]₂, [ThrS]₂, [SerS]₂, [ProS]₂, [PheS]₂[PheT]₂, [MetG]₂, [lysS]₂, [HisS]₂, [GlyS]₂[GlyQ]₂).

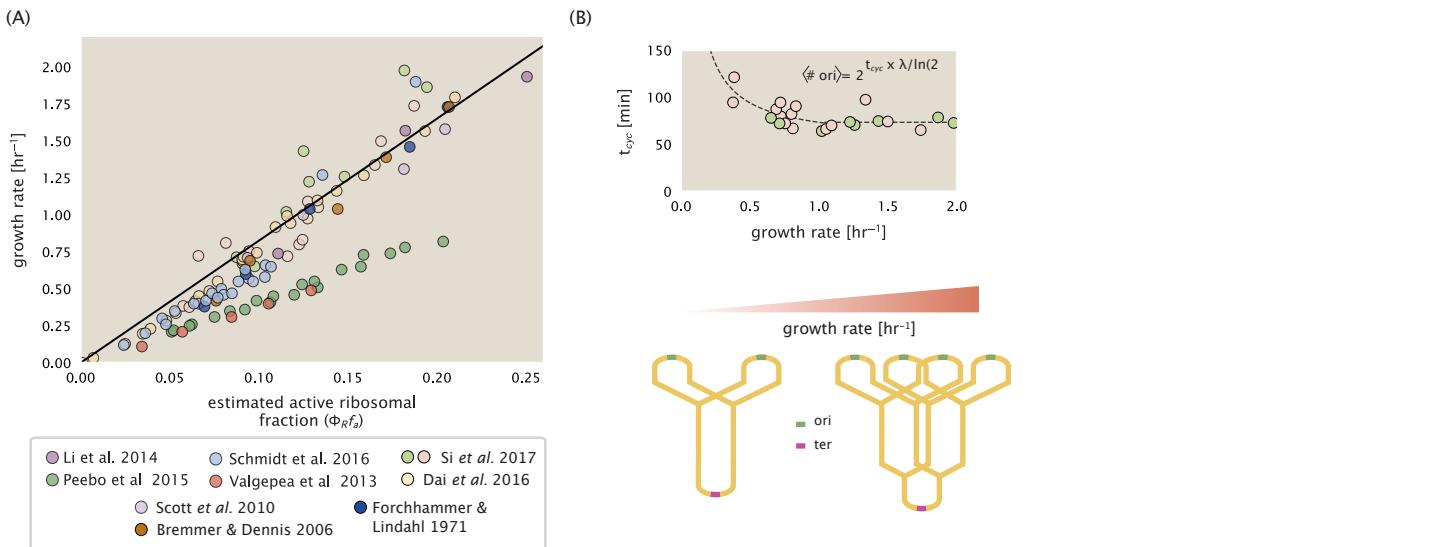


Figure 9—Figure supplement 1. (A) Actively translating ribosomal fraction versus growth rate. The actively translating ribosomal fraction is calculated using the estimated values of f_a from ? (shown in inset; see the Appendix Section "???" for additional detail). Additional measurements in addition to the proteomic measurements are based on measurements of cellular RNA to protein ratio, with $\Phi_R \approx$ the cellular RNA to protein ratio divided by 2.1 (?). (B) Experimental measurements of the cell doubling time τ and cell cycle time t_{cyc} from Si et al. (2017). Dashed line shows fit to the data, which were used to estimate $\langle \# \text{ ori} \rangle$. t_{cyc} was assumed to vary in proportion to τ for doubling times greater than 40 minutes, and reach a minimum value of 73 minutes. See Appendix ?? for additional details exact estimation of rRNA copy number. Red data points correspond to measurements in strain MG1655, while light green points are for strain NCM3722. Schematic shows the expected increase in replication forks (or number of ori regions) as *E. coli* cells grow faster.

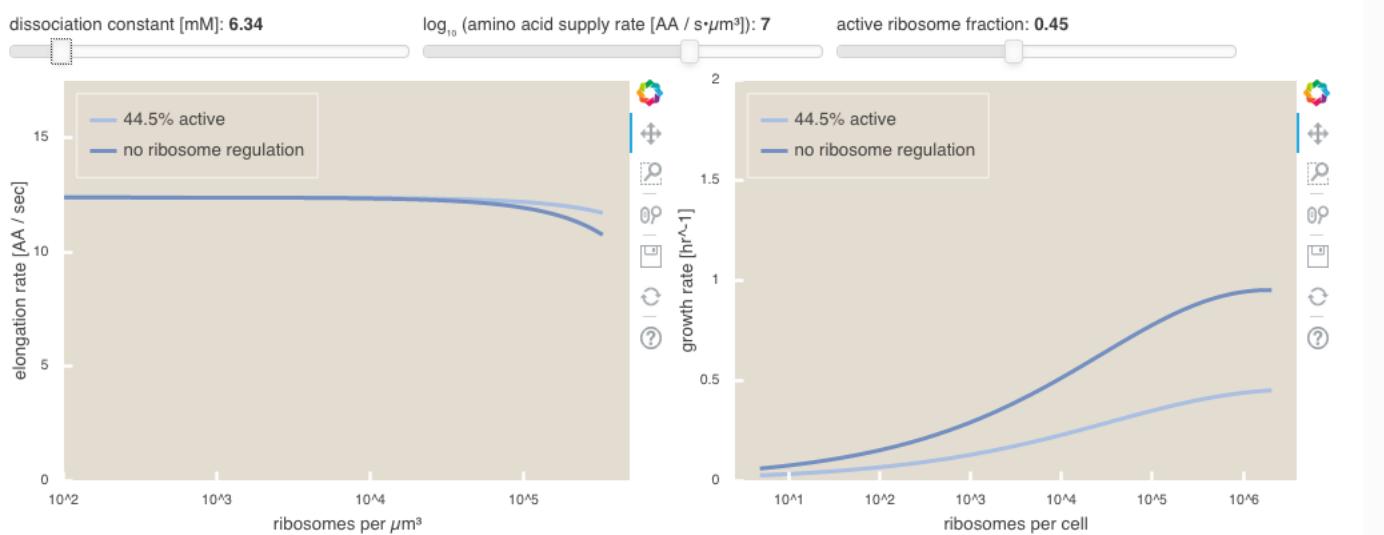


Figure 11—Figure supplement 1. An interactive version of parts (B) and (C) of ?? which permit the user to modulate the rate of amino acid supply, the dissociation constant of amino acids to the ribosome, and the fraction of the ribosome pool that is actively translating. This interactive figure, and the code used to generate it, is available on the [paper website](#).