

Fundamental limits on the rate of bacterial cell division

³ Nathan M. Belliveau^{†, 1}, Griffin Chure^{†, 2}, Christina L. Hueschen³, Hernan G. Garcia⁴, Jane
⁴ Kondev⁵, Daniel S. Fisher⁶, Julie A. Theriot^{1, 7, *}, Rob Phillips^{8, 9, *}

⁵ ¹Department of Biology, University of Washington, Seattle, WA, USA; ²Department of Applied Physics,
⁶ California Institute of Technology, Pasadena, CA, USA; ³Department of Chemical Engineering,
⁷ Stanford University, Stanford, CA, USA; ⁴Department of Molecular Cell Biology and Department of
⁸ Physics, University of California Berkeley, Berkeley, CA, USA; ⁵Department of Physics, Brandeis
⁹ University, Waltham, MA, USA; ⁶Department of Applied Physics, Stanford University, Stanford, CA,
¹⁰ USA; ⁷Allen Institute for Cell Science, Seattle, WA, USA; ⁸Division of Biology and Biological Engineering,
¹¹ California Institute of Technology, Pasadena, CA, USA; ⁹Department of Physics, California Institute of
¹² Technology, Pasadena, CA, USA; *Co-corresponding authors. Address correspondence to
¹³ phillips@pboc.caltech.edu and jtheriot@uw.edu; [†]These authors contributed equally to this work

¹⁴

¹⁵ **Abstract** Recent years have seen a deluge of experiments dissecting the relationship between bacterial
¹⁶ growth rate, cell size, and protein content, quantifying the abundance of proteins across growth conditions with
¹⁷ unprecedented resolution. However, we still lack a rigorous understanding of what sets the scale of these
¹⁸ quantities and when protein abundances should (or should not) depend on growth rate. Here, we seek to
¹⁹ quantitatively understand this relationship across a collection of *Escherichia coli* proteomic data sets covering \approx
²⁰ 4000 proteins and 36 growth rates. We estimate the basic requirements for steady-state growth by considering
²¹ key processes in nutrient transport, energy generation, cell envelope biogenesis, and the central dogma. From
²² these estimates, ribosome biogenesis emerges as a primary determinant of growth rate. We expand on this
²³ assessment by exploring a model of ribosomal regulation as a function of the nutrient supply, revealing a
²⁴ mechanism that ties cell size and growth rate to ribosomal content.

²⁵

²⁶ Introduction

²⁷ The observed range of bacterial growth rates is enormously diverse. In natural environments, some microbial or-
²⁸ ganisms may double only once per year (?) while in comfortable laboratory conditions, growth can be rapid with
²⁹ several divisions per hour (?). This six order-of-magnitude difference in time scales encompasses different micro-
³⁰ bial species and lifestyles, yet even for a single species such as *Escherichia coli*, the growth rate can be modulated
³¹ over a comparably large scale by tuning the type and amount of nutrients in the growth medium. This remarkable
³² flexibility in growth rate illustrates the intimate relationship between environmental conditions and the rates at
³³ which cells convert nutrients into new cellular material – a relationship that has remained a major topic of inquiry
³⁴ in bacterial physiology for over a century (?).

³⁵ Jacques Monod once remarked that “the study of the growth of bacterial cultures does not constitute a special-
³⁶ ized subject or branch of research, it is the basic method of Microbiology.” Those words ring as true today as they
³⁷ did when they were written 70 years ago (?) with the quantitative power of this “method” recently undergoing
³⁸ renaissance. Many of the key questions addressed by the pioneering efforts in the middle of the last century can
³⁹ be revisited by examining them through the lens of the increasingly refined molecular census that is available for
⁴⁰ bacteria such as the microbial workhorse *E. coli*.

⁴¹ Several of the evergreen questions about bacterial growth and physiology that were originally raised by micro-

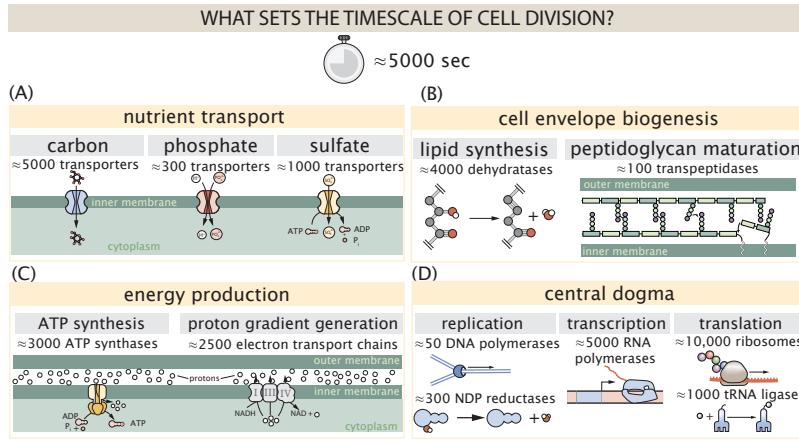


Figure 1. Transport and synthesis processes necessary for cell division. We consider an array of processes necessary for a cell to double its molecular components, broadly grouped into four classes. These categories are (A) nutrient transport across the cell membrane, (B) cell envelope biogenesis, (C) 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 $\approx 5000 \text{ s}$.

42 biologists in the middle of the 20th century can now be reframed in light of this newly available data. For example,
 43 what biological processes are the primary determinants for how quickly bacterial cells can grow and reproduce?
 44 How do cells modulate the absolute numbers and relative ratios of their molecular constituents as a function of
 45 changes in growth rate or nutrient availability? In this paper, we begin by considering these two questions from
 46 two distinct angles. First, as a result of an array of high-quality proteome-wide measurements of *E. coli* under
 47 diverse growth conditions, we have a census that allows us to explore how the number of key molecular players
 48 change as a function of growth rate. Here, we have assembled a singular data set using measurements collected
 49 over the past decade via mass spectrometry (???) or ribosomal profiling (?) of the composition of the *E. coli* pro-
 50 teome across 36 unique growth rates (see Appendix ?? for further discussion of the data). Second, by compiling
 51 molecular turnover rate measurements for many of the fundamental processes associated with bacterial growth,
 52 we make quantitative estimates of key cellular processes (schematized in ??) to determine whether our current
 53 understanding of the dynamics of these processes are sufficient to explain the magnitude of the observed protein
 54 copy numbers across conditions. The census, combined with these estimates, provide a window into the question
 55 of whether the rates of central processes such as energy generation or DNA synthesis are regulated systematically
 56 as a function of cell growth rate by altering protein copy number.

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

64 From these estimates, it emerges that translation, particularly the synthesis of ribosomal proteins, is a plausi-
 65 ble candidate that limits the rate of cell division. We reach this conclusion by considering that ribosome synthe-
 66 sis is 1) a rate limiting step for the *fastest* bacterial division, and 2) a major determinant of bacterial growth across the
 67 nutrient conditions we have considered under steady state, exponential growth. This enables us to suggest that
 68 the long-observed correlation between growth rate and cell size (??) can be simply attributed to the increased ab-
 69 solute number of ribosomes per cell under conditions supporting extremely rapid growth. To better understand
 70 how the observed alterations in absolute protein abundances, and in particular, changes in ribosome copy num-
 71 ber, influence growth rate across different nutrient conditions we consider a minimal model of cellular growth.
 72 Our conclusions from these analyses provide important insight into how *E. coli* regulates growth across conditions
 73 of differing nutrient availability and identifies fundamental constraints in bacterial growth more broadly.

78

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

76

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.

77

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.

78

Furthermore, we use equality symbols (=) sparingly and frequently defer to approximation (\approx) or scaling (\sim) 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 \sim is used, we are stating that we are confident in our estimate to within an order of magnitude.

79

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.

80

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.

113 **Nutrient Transport**

114 We begin by considering the critical transport processes diagrammed in ??(A). In order to build new cellular mass,
115 the molecular and elemental building blocks must be scavenged from the environment in different forms. Car-
116 bon, for example, is acquired via the transport of carbohydrates and sugar alcohols with some carbon sources
117 receiving preferential treatment in their consumption (?). Phosphorus, sulfur, and nitrogen, on the other hand, are
118 harvested primarily in the forms of inorganic salts, namely phosphate, sulfate, and ammonia (??????). All of these
119 compounds have different membrane permeabilities (?) and most require some energetic investment either via
120 ATP hydrolysis or through the proton electrochemical gradient to bring the material across the hydrophobic cell
121 membrane.

122 The elemental composition of *E. coli* has received much quantitative attention over the past half century (????),
123 providing us with a starting point for estimating how many atoms of each element must be scavenged from the
124 environment. A synthesis of these studies presents an approximate dry mass composition of $\approx 50\%$ carbon (BNID:
125 100649, see ??), $\approx 15\%$ nitrogen (BNID: 106666), $\approx 3\%$ phosphorus (BNID: 100653), and 1% sulfur (BNID: 100655)
126 with remainder being attributable to oxygen, hydrogen, and various transition metals. We use this stoichiometric
127 breakdown to estimate the abundance and growth rate dependence of a variety of transporters responsible for
128 carbon uptake, and provide more extensive investigation of the other critical elements – phosphorus, sulfur, and
129 nitrogen – in the Appendix ??.

130 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
131 approximation that $\approx 50\%$ of this mass is carbon, we estimate that $\sim 10^{10}$ carbon atoms must be brought into the
132 cell in order to double all of the carbon-containing molecules ??(A, top). Typical laboratory growth conditions
133 provide carbon as a single class of sugar (such as glucose, galactose, or xylose) often transported cross the cell
134 membrane by a transporter complex specific to that particular sugar. One such mechanism of transport is via
135 the PTS system which is a highly modular system capable of transporting a diverse range of sugars with high
136 specificity (?). The glucose-specific component of this system transports ≈ 200 glucose molecules (≈ 1200 carbon
137 atoms) per second per transporter (BNID: 114686). Making the assumption that this is a typical sugar transport
138 rate for the PTS system, coupled with the need to transport $\sim 10^{10}$ carbon atoms, we then expect on the order of
139 ≈ 1000 transporters must be expressed per cell in order to bring in enough carbon atoms ??(A, top).

140 However, we find this estimate to be exceeded by several fold by experimental measurements ??(A, bottom),
141 implying that the cell is capable of transporting more carbon atoms than strictly needed for biosynthesis. While
142 we estimate ≈ 1000 transporters are needed with a 5000 second division time, we can abstract this calculation to
143 consider any particular growth rate given knowledge of the cell density and volume as a function of growth rate
144 and direct the reader to the Appendix ?? for more information. This abstraction, shown as a grey line in ??(A),
145 reveals an excess of transporters even at faster growth rates. This contrasts with our observations for uptake
146 of phosphorus and sulfur, which align well with our expectations across different growth conditions (??-?? and
147 discussed further in Appendix ??).

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

157 Many metabolic operons are regulated with dual-input logic gates that are only expressed when glucose con-
158 centrations are low and the concentration of other carbon sources are elevated (?????). Points colored in red in
159 ??(B) (labeled by red text-boxes) correspond to growth conditions in which the specific carbon source (glycerol,
160 xylose, or fructose) is present as the sole source of carbon. The grey lines in ??(B) show the estimated number
161 of transporters needed at each growth rate to satisfy the cellular carbon requirement, adjusted for the specific
162 carbon source in terms of number of carbon atoms per molecule and the rate of transport for the particular

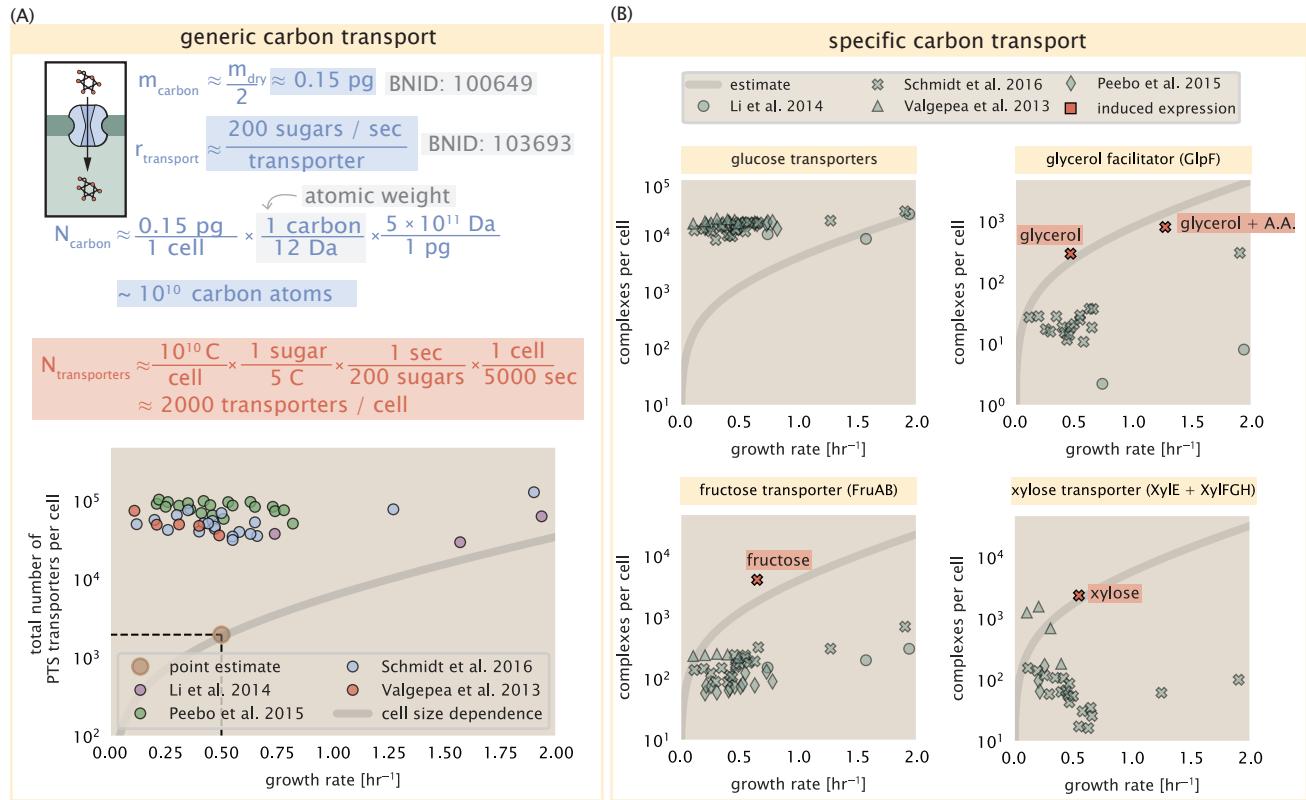


Figure 2. The abundance of carbon transport systems across growth rates. (A) A simple estimate for the minimum number of generic carbohydrate transport systems (top) assumes $\sim 10^{10}$ C are needed to complete division, each transported sugar contains ≈ 5 C, and each transporter conducts sugar molecules 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.

163 transporter species. These plots show that, in the absence of the particular carbon source, expression of the
164 transporters is maintained on the order of $\sim 10^2$ per cell. The low but non-zero abundances may reflect the spe-
165 cific regulatory logic involved, requiring that cells are able to transport some minimal amount of an alternative
166 carbon source in order to induce expression of these alternative carbon-source systems.

167 **Limits on Transporter Expression**

168 If acquisition of nutrients was a limiting process in cell division under the typical growth conditions explored here,
169 the growth rate could be theoretically increased simply by expressing more transporters, but is this feasible at a
170 physiological level? A way to approach this question is to compute the amount of space in the bacterial membrane
171 that could be occupied by nutrient transporters. Considering a rule-of-thumb for the surface area of *E. coli* of about
172 $5 \mu\text{m}^2$ (BNID: 101792), we expect an areal density for 1000 transporters to be approximately 200 transporters/ μm^2 .
173 For a typical transporter occupying about 50 nm^2 , this amounts to about only $\approx 1\%$ of the total inner membrane
174 area (?). Additionally, bacterial cell membranes typically have densities of 10^5 proteins/ μm^2 (?), implying that the
175 cell could accommodate more membrane and this places additional limitations on cell size and surface area that
176 we will consider further in the coming sections.

177 **Cell Envelope Biogenesis**

178 In contrast to nutrient transporters, which support the synthesis of biomolecules throughout the cell and therefore
179 need to scale with the cell size, here we must consider the synthesis of components that will need to scale with
180 the surface area of the cell. *E. coli* is a rod-shaped bacterium with a remarkably robust length-to-width aspect
181 ratio of $\approx 4:1$ (?). At modest growth rates, the total cell surface area is $\approx 5 \mu\text{m}^2$ (BNID: 101792). Assuming this
182 surface area is approximately the same between the inner and outer membranes of *E. coli*, and the fact that each
183 membrane is itself a lipid bilayer, cells have a the total membrane surface area of $\approx 20 \mu\text{m}^2$ (see Appendix ?? for
184 a description of the calculation of cell surface area as a function of cell size). In this section, we will estimate the
185 number of protein complexes needed to produce this membrane surface area as well as the complexes involved
186 in assembling the peptidoglycan scaffold it encapsulates.

187 **Lipid Synthesis**

188 The dense packing of the membrane with proteins means that the cell membranes are not composed entirely
189 of lipid molecules, with only $\approx 40\%$ of the membrane area is occupied by lipids (BNID: 100078). Using a rule-of-
190 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,
191 which is in close agreement with experimental measurements (BNID: 100071, 102996).

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

203 **Peptidoglycan Synthesis**

204 Bacterial cells demonstrate exquisite control over their cell shape. This is primarily due to the cell wall, a stiff, sev-
205 eral nanometer thick meshwork of polymerized disaccharides. The formation of the peptidoglycan is an intricate
206 process involving many macromolecular players (?), whose coordinated action maintains cell shape and integrity
207 even in the face of large-scale perturbations (?). The peptidoglycan alone comprises $\approx 3\%$ of the cellular dry mass
208 (BNID: 1019360, making it the most massive molecule in *E. coli*. The polymerized unit of the peptidoglycan is a
209 N-acetylglucosamine and N-acetylmuramic acid disaccharide, of which the former is functionalized with a short

210 pentapeptide. With a mass of \approx 1000 Da, this unit, which we refer to as a murein monomer, it is polymerized
211 to form long strands in the periplasm which are then attached to each other via their peptide linkers. Together,
212 these quantities provide an estimate of $\approx 5 \times 10^6$ murein monomers per cell.

213 The crosslinking of the pentapeptides between adjacent glycan strands is responsible for maintaining the structural
214 integrity of the cell wall and, in principle, each murein monomer can be involved in such a crosslink. In some
215 microbes, such as in gram-positive bacterium *Staphylococcus aureus*, the extent of crosslinking can be large with
216 > 90% of pentapeptides forming a connection between glycan strands. In *E. coli*, however, a much smaller proportion
217 (\approx 20%) of the peptides are crosslinked, resulting in a weaker and more porous cell wall ??. The formation of
218 these crosslinks occurs primarily during the polymerization of the murein monomers and is facilitated by a family
219 of enzymes called transpeptidases. The four primary transpeptidases of *E. coli* have only recently been quantitatively
220 characterized *in vivo* via liquid chromatography mass spectrometry which revealed a notably slow kinetic
221 turnover rate of \approx 2 crosslinking reactions formed per second per enzyme (?).

222 Assembling these quantities permits us to make an estimate that on the order of \approx 100 transpeptidases per
223 cell are needed for complete maturation of the peptidoglycan, given a division time of \approx 5000 seconds; a value
224 that is comparable to experimental observations (??(B)). Expanding this estimate to account for the changing
225 mass of the peptidoglycan as a function of growth rate (grey line in ??(B)) also qualitatively captures the observed
226 dependence in the data, though systematic disagreements between the different data sets makes the comparison
227 more difficult.

228 Limits on Cell Wall Biogenesis

229 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
230 and peptidoglycan components decrease at faster growth rates due to a decrease in their surface area to volume
231 (S/V) ratio (?). Furthermore, despite the slow catalytic rate of FabZ and FabA in lipid synthesis, experimental data
232 and recent computational modeling has shown that the rate of fatty-acid synthesis can be drastically increased
233 by increasing the concentration of FabZ (??). With a proteome size of $\approx 3 \times 10^6$ proteins, a hypothetical 10-fold increase
234 in expression from 4000 to 40,000 ACP dehydratases would result in a paltry \approx 1% increase in the size of the
235 proteome. In the context of peptidoglycan synthesis, we note that our estimate considers only the transpeptidase
236 enzymes that are involved lateral and longitudinal elongation of the peptidoglycan. This neglects the presence of
237 other transpeptidases that are present in the periplasm and also involved in remodeling and maturation of the
238 peptidoglycan. It is therefore possible that if this was setting the speed limit for cell division, the simple expression
239 of more transpeptidases may be sufficient to maintain the structural integrity of the cell wall.
240

241 Energy Production

242 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
243 biological systems away from thermodynamic equilibrium. We next turn to the synthesis of ATP as a potential
244 process that may limit growth, which also requires us to consider the maintenance of the electrochemical proton
245 gradient which powers it.
246

247 ATP Synthesis

248 Hydrolysis of the terminal phosphodiester bond of ATP into ADP (or alternatively GTP and GDP) and an inorganic
249 phosphate provides the thermodynamic driving force in a wide array of biochemical reactions. One such reaction
250 is the formation of peptide bonds during translation, which requires \approx 2 ATPs for the charging of an amino
251 acid to the tRNA and \approx 2 GTP for the formation of each peptide bond. Assuming the ATP costs associated with
252 error correction and post-translational modifications of proteins are negligible, we can make the approximation
253 that each peptide bond has a net cost of \approx 4 ATP (BNID: 101442, ?). Formation of GTP from ATP is achieved via
254 the action of nucleoside diphosphate kinase, which catalyzes this reaction without an energy investment (?) and
255 therefore consider all NTP requirements of the cell to be functionally equivalent to being exclusively ATP. In total,
256 the energetic costs of peptide bond formation consume \approx 80% of the cells ATP budget (BNID: 107782; 106158;

CELL ENVELOPE BIOSYNTHESIS

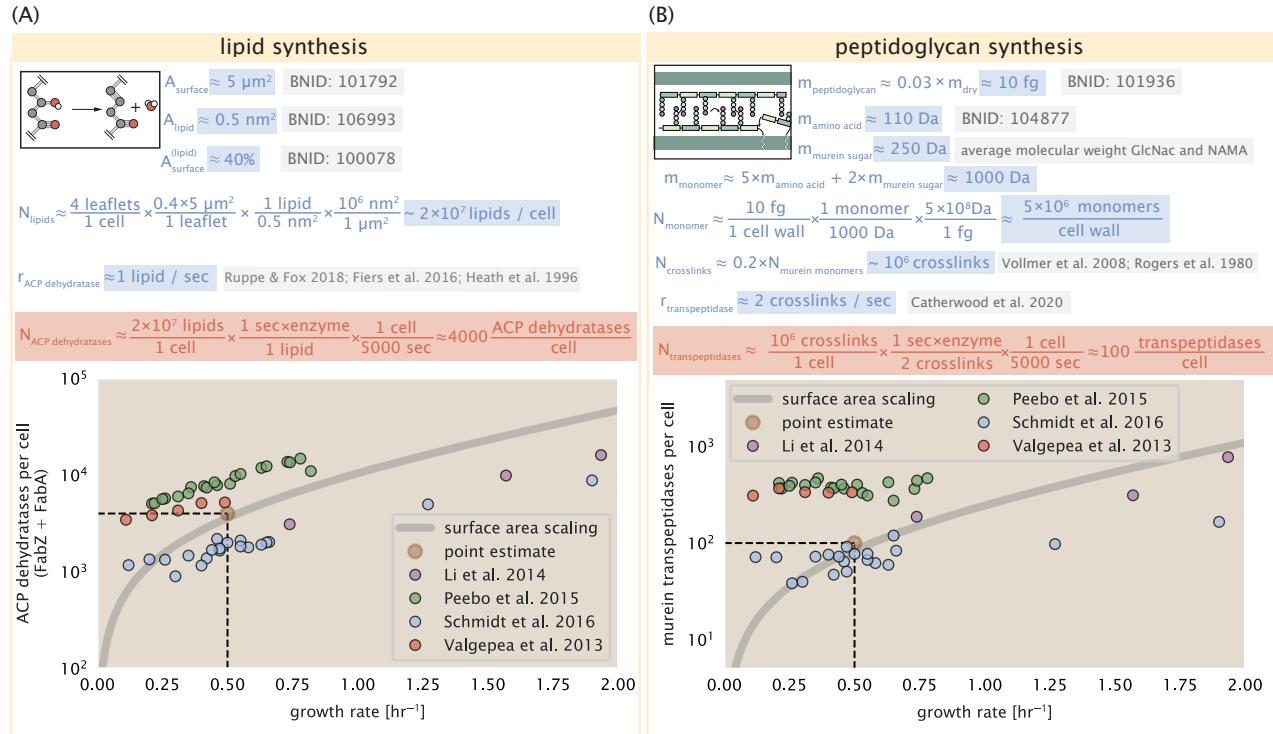


Figure 3. (A) Top panel shows an estimation for the number of ACP dehydratases necessary to form functional phospholipids, which is assumed to be a rate-limiting step on lipid synthesis. The rate of ACP dehydratases was inferred from experimental measurements via a stochastic kinetic model described 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 $r_{\text{E. coli}}$ 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.

257 101637; 111918, ??). The pool of ATP is produced by the F₁-F₀ ATP synthase – a membrane-bound rotary motor
258 which under ideal conditions can yield \approx 300 ATP per second (BNID: 114701; ?).

259 To estimate the total number of ATP equivalents consumed during a cell cycle, we will make the approximation
260 that there are $\approx 3 \times 10^6$ proteins per cell with an average protein length of \approx 300 peptide bonds (BNID: 115702;
261 108986; 104877). Taking these values together, coupled with an estimate of \approx 4 ATP equivalents per peptide bond,
262 we find that the typical *E. coli* cell consumes $\sim 5 \times 10^9$ ATP per cell cycle on protein synthesis alone. Assuming that
263 each ATP synthases operates at its maximal speed (300 ATP per second per synthase), \approx 3000 ATP synthases
264 are needed to keep up with the energy demands of the cell. This estimate is comparable with the experimental
265 observations, shown in ?? (A). We note that this estimate assumes all ATP is synthesized via ATP synthase and
266 neglects synthesis via fermentative metabolism. This assumption may explain why at the fastest growth rates (\approx
267 2 hr⁻¹), our continuum estimate predicts more synthase than is experimentally observed (gray line in ??). At rapid
268 growth rates, *E. coli* enters a type of overflow metabolism where fermentative metabolism becomes pronounced
269 (?).

270 Generating the Proton Electrochemical Gradient

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

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

289 Limits on Biosynthesis in a Crowded Membrane

290 Our estimates thus far have focused on biochemistry at the periphery of the cell and have generally been con-
291 cordant with the abundances predicted by our estimates. However, as surface area and volume do not scale
292 identically, it is necessary to consider the physical limits for transport and energy production given the S/V ratio,
293 which as we've noted will decrease at faster growth rates.

294 In our estimate of ATP production above we found that a cell demands about 5×10^9 ATP per cell cycle or 10^6
295 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,
296 in line with previous estimates (?). In ?? (A) we plot this ATP demand as a function of the S/V ratio in green, where
297 we have considered a range of cell shapes from spherical to rod-shaped with an aspect ratio (length/width) equal
298 to 4. In order to consider the maximum ATP that could be produced, we consider the amount of ATP that can be
299 generated by a membrane filled with ATP synthase and electron transport complexes, which provides a maximal
300 production of about 3 ATP / (nm²·s) (?). This is shown in blue in ??(A), which shows that at least for the growth
301 rates observed (right column in plot), the energy demand is roughly an order of magnitude less. Interestingly, ?
302 also found that ATP production by respiration is less efficient than by fermentation per membrane area occupied
303 due to the additional proteins of the electron transport chain. This suggests that, even under anaerobic growth,
304 there will be sufficient membrane space for ATP production.

305 The analysis highlights the diminishing capacity to provide resources as the cell increases in size. However, the

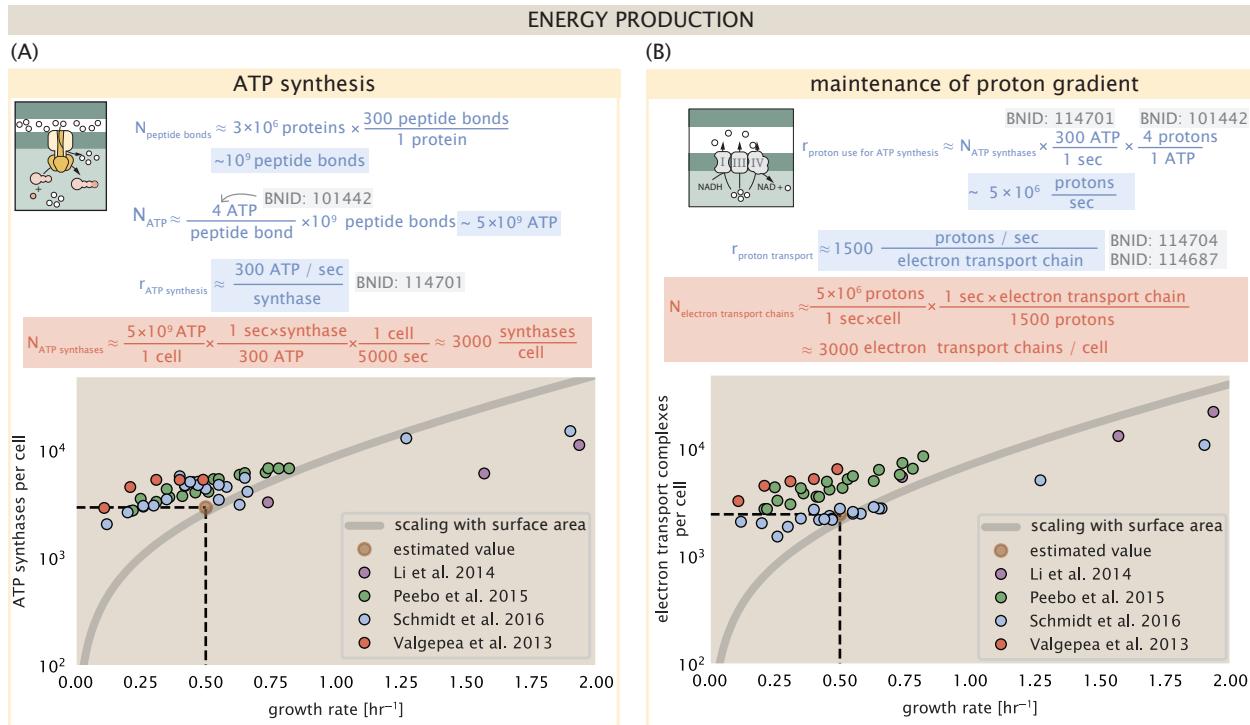


Figure 4. The abundance of F₁-F₀ ATP synthases and electron transport chain complexes as a function of growth rate.

(A) Estimate of the number of F₁-F₀ ATP synthase complexes needed to accommodate peptide bond formation and other NTP dependent processes. Points in plot correspond to the mean number of complete F₁-F₀ 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 F₁-F₀ 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 oxioreductase I ([NuoA][NuoH][NuoJ][NuoK][NuoL][NuoM][NuoN][NuoB][NuoC][NuoE][NuoF][NuoG][NuoI]) 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.

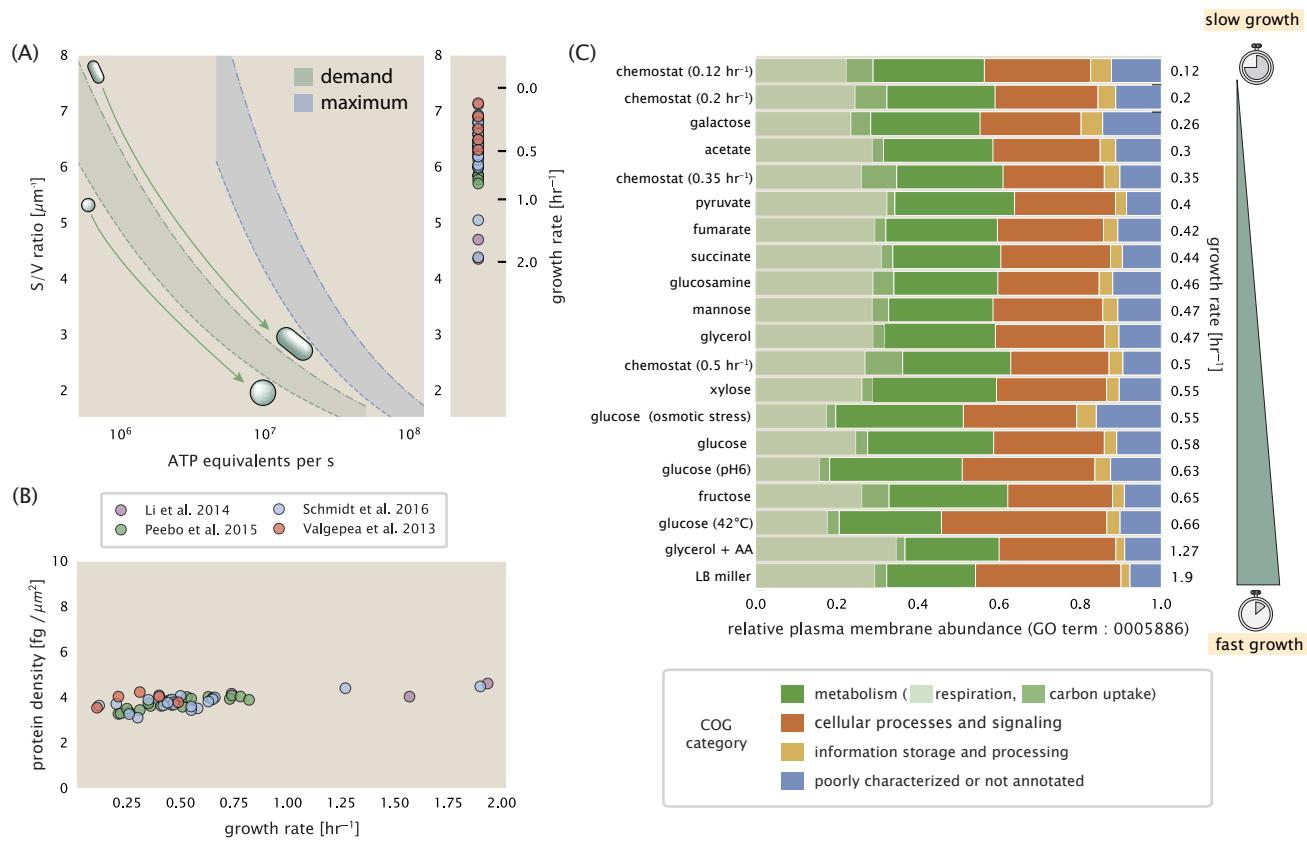


Figure 5. Influence of cell size and S/V ratio on ATP production and inner membrane composition. (A) Scaling of ATP demand and maximum ATP production as a function of S/V ratio. Cell volumes of 0.5 fL to 50 fL were considered, with the dashed (—) line corresponding to a sphere and the dash-dot line (---) reflecting a rod-shaped bacterium like *E. coli* with a typical aspect ratio (length / width) of 4 (?). Approximately 50% of the bacterial inner membrane is assumed to be protein, with the remainder lipid. The right plot shows the measured growth rates, and their estimated S/V ratio using growth rate dependent size measurements from ? (See Appendix ?? on calculation). (B) Total protein mass per μm^2 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)).

maximum energy production in ??(A) does represent a somewhat unachievable limit since the inner membrane must also include other proteins including those required for lipid and membrane synthesis. 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 for those in metabolism (including ATP synthesis via respiration) is also relatively constant across growth rates, suggesting that no one process (energy production, nutrient uptake, etc.) is particularly 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 central processes which *must* be undertaken irrespective of the growth

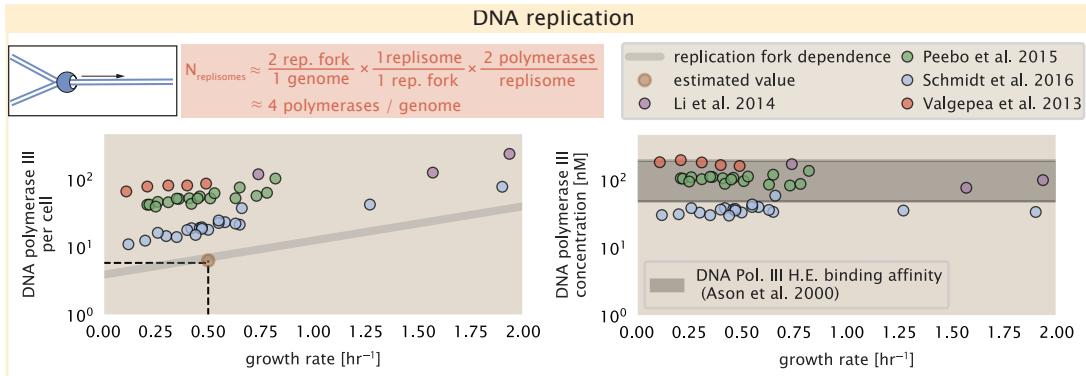


Figure 6. Complex abundance estimates for dNTP synthesis and DNA replication. An estimate for the minimum number of DNA polymerase holoenzyme complexes needed to facilitate replication of a single genome. Points in the left-hand plot correspond to the total number of DNA polymerase III holoenzyme complexes ($[DnaE]_3[DnaQ]_3[HolE]_3[DnaX]_5[HolB][HolA][DnaN]_4[HolC]_4[Hold]_4$) 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.

320 conditions – those of the central dogma.

321 DNA Replication

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

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

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

346 RNA Synthesis

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

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

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

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

375 Our estimates also neglect other mechanistic features of transcription and transcriptional initiation more
376 broadly. For example, we acknowledge that some fraction of the RNAP pool is nonspecifically bound to DNA
377 during its search for promoters from which to begin transcription. Furthermore, we ignore the obstacles that RNA
378 polymerase and DNA polymerase present to each other as they move along the DNA (?). Finally, we neglect the
379 fact that RNA polymerase also require σ -factors for promoter recognition and transcription initiation (?).

380 While they are the machinery for transcription, RNA polymerase is not sufficient to initiate transcription. Promoter
381 recognition and initiation of transcription is dependent on the presence of σ -factors, protein cofactors which
382 bind directly to the polymerase (?). In ??–??, we show that the predicted RNA polymerase copy number indeed is
383 more comparable with the abundance of σ -70 (RpoD), the primary sigma factor in *E. coli*. There therefore remains
384 more to be investigated as to what sets the observed abundance of RNA polymerase in these proteomic data sets.
385 However, we conclude that the observed excess in abundance for RNA polymerase abundances are generally
386 in excess of what appears to be needed for growth, suggesting that the abundance of RNA polymerase itself is
387 not particularly limiting.

388 Protein Synthesis

389 The last process of the central dogma – the translation of RNA into protein – is the final subject in our dialogue
390 between back-of-the-envelope estimates and comparison with proteomic data. Here we consider the number of
391 ribosomes needed to replicate the cellular proteome. While the rate at which ribosomes translates is well known
392 to have a growth rate dependence (?), here we make the approximation that translation occurs at a rate of \approx 15
393 amino acids per second per ribosome (BNID: 100233). Under this approximation and assuming a division time of
394 5000 seconds, we can easily arrive at an estimate of \approx 10^4 ribosomes needed per cell to replicate the entire protein
395 mass (??(A), top). This point estimate and the corresponding estimate across a continuum of growth rates proves

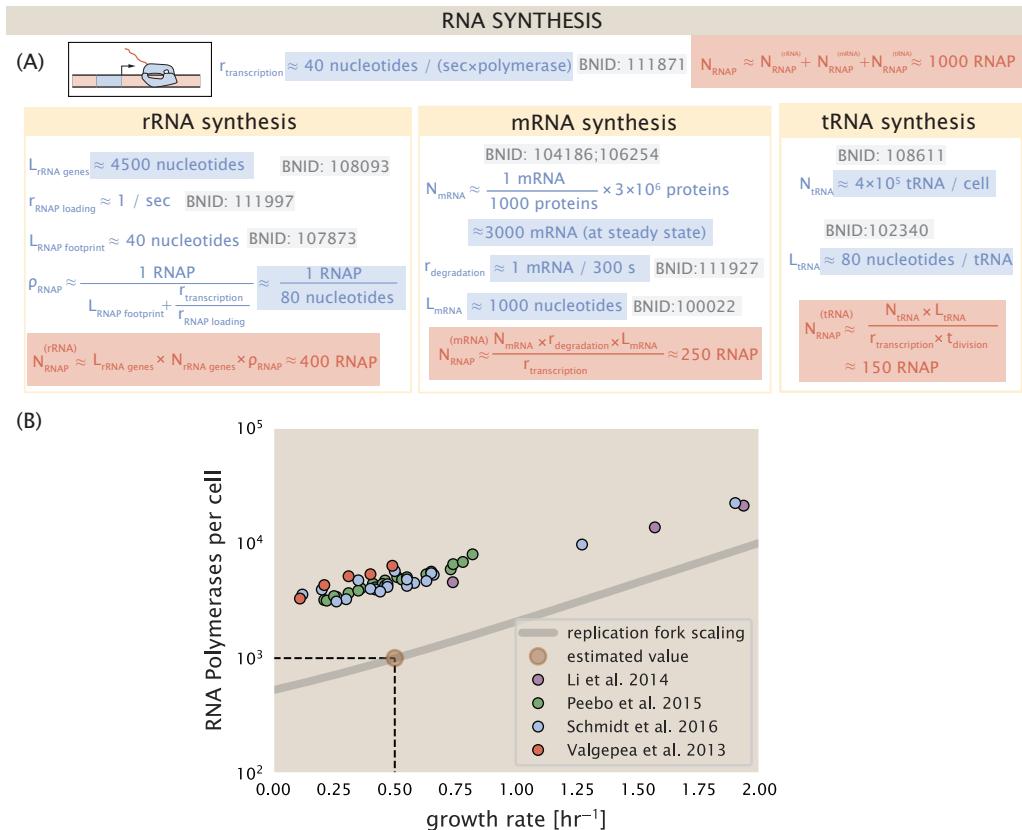


Figure 7. Estimation of the RNA polymerase demand and comparison with experimental data. (A) Estimations for the number of RNA polymerase needed to synthesize sufficient quantities of rRNA, mRNA, and tRNA from left to right, respectively.(B) The RNA polymerase core enzyme copy number as a function of 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.

396 to be notably comparable to the experimental observations, shown in ??(A). While the ribosome is responsible for
397 the formation of peptide bonds, we do not diminish the importance of the charging of tRNAs with the appropriate
398 amino acid, a process which occurs with remarkable accuracy. In the Appendix and in ??-??, we consider the process
399 of ligating tRNAs to their corresponding amino acid and again find notable comparability with the data.

400 Having completed our circuit through key processes of cellular growth outlined in ??, we can now take stock
401 on our understanding of the observed growth rate dependence and abundances of various protein complexes.
402 We note that in general, these simple estimates have been reasonably successful in quantitatively describing the
403 observations in the proteomic data, suggesting that the proteome is tuned in composition and absolute abun-
404 dance to match the growth rate requirements, without any one process representing a particular bottleneck or
405 rate limiting step in division. However, in our effort to identify key limitations on growth, there are two notable
406 observations that we wish to emphasize.

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

420 This question, more frequently presented as a question of optimal resource allocation, has been the target of
421 an extensive dialogue between experiment and theory over the past decade. In a now seminal work, ? presents
422 an elegant treatment of resource allocation through partitioning of the proteome into sectors – one of which being
423 ribosome-associated proteins – the relative sizes of which ultimately define the total cellular growth rate. In
424 more recent years, this view has been more thoroughly dissected experimentally (?????) and together represent
425 a paradigm-shift in how we think of cellular physiology at the proteome-level. However, the quantitative descrip-
426 tion of these observations is often couched in terms of phenomenological constants and effective parameters with
427 the key observable features of expression being computed in relative, rather than absolute, abundances. Further-
428 more, these approaches often exclude or integrate away effects of cell size and chromosome content, which we
429 have illustrated in our estimates to have deep connections to the observed cellular growth rate.

430 In the closing sections of this work, we explore how ribosome content, cell size, and chromosome content are
431 intertwined in their control over the cellular growth rate. To do so, we take a more detailed view of ribosome abun-
432 dance, exchanging our order-of-magnitude estimates for a minimal mathematical model of growth rate control.
433 This is defined by parameters with tangible connections to the biological processes underlying cellular growth
434 and protein synthesis. Using this model, we interrogate how the size of the ribosome pool and its corresponding
435 activity is regulated to balance the supply of amino acids via metabolism and catabolism with the consumption of
436 amino acids through peptide bond formation required for growth.

437 **Maximum Growth Rate is Determined by the Ribosomal Mass Fraction**

438 The 7 minute speed limit shown in ??(B) assumes all necessary proteins in the cell are strictly ribosomal. Here, we
439 relax this assumption to derive a translation-limited growth rate in order to make a comparison to the experimen-
440 tal data.

441 We will assume at a given growth rate λ , the cell is composed of N_{pep} peptide bonds and R ribosomes. Each
442 ribosome, though mostly rRNA, has $L_R \approx 7500$ amino acids. With an average mass of an amino acid of $m_{\text{AA}} \approx$
443 110 Da (BNID: 104877). The total mass fraction of the proteome associated with ribosomal proteins Φ_R can be
444 computed as

$$\Phi_R =$$

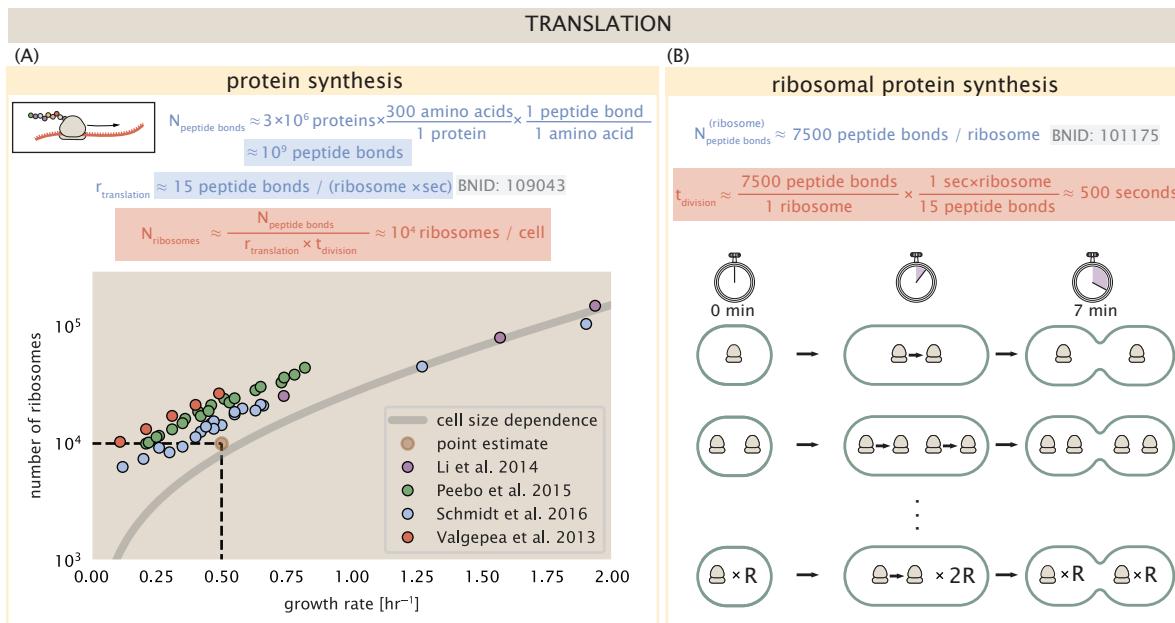


Figure 8. Estimation of the required number of ribosomes and the speed limit for bacterial replication. (A) Estimation of the number of ribosomes required to synthesize 10^9 peptide bonds with an elongation rate of 15 peptide bonds per second. The average abundance of ribosomes is plotted as a function 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.

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

446 In ??(C) we find that above about 0.75 hr^{-1} , the growth rate is determined by the ribosomal mass fraction Φ_R ,
 447 since f_a is close to 1, and r_i is near its maximal rate [cite and refer to figure/ supplemental]. While Φ_R will need to
 448 increase in order for cells to grow faster, the fractional dependence in ?? gives little insight into how this is actually
 449 achieved in the cell and we consider this further here.

450 It is now well-documented that *E. coli* cells add a constant volume per origin of replication, which is robust
 451 to a remarkable array of cellular perturbations (?). We find that ribosomal copy number also scales its ribosome
 452 copy number in proportion to $\langle \# \text{ ori} \rangle$??(A). Importantly, however, it will only be due to an increase in Φ_R at these
 453 moderate to fast growth rates that cells can achieve an increase in their growth rate. Indeed, we find that the
 454 deviations in protein expression with $\langle \# \text{ ori} \rangle$ are largely restricted to regions of ribosomal protein genes ??(B). Here
 455 we have calculated the position-dependent protein expression across the chromosome by a running Gaussian
 456 average of protein copy number (20 kbp st. dev. averaging window) based on each gene's transcriptional start
 457 site. These were median-subtracted to account for the change in total protein abundance with $\langle \# \text{ ori} \rangle$. This result
 458 suggests that Φ_R is also being tuned in proportion to $\langle \# \text{ ori} \rangle$ under nutrient-limited growth, and in particular, it
 459 is through this additional dependence on Φ_R that *E. coli* exhibits an exponential increase in cell size with growth
 460 rate.

461 A Minimal Model of Nutrient-Mediated Growth Rate Control

462 While the preceding subsections highlight a dominant role for ribosomes in setting the growth rate, our analysis
 463 on the whole emphasizes that the total proteomic content must also change in response to variable growth con-
 464 ditions and growth rate. In this final section we use a minimal model of growth rate control to better understand
 465 how the interconnection between ribosomal abundance and total protein influences the observed growth rate.
 466 Here we propose that cells modulate their protein abundance in direct response to the availability of nutrients

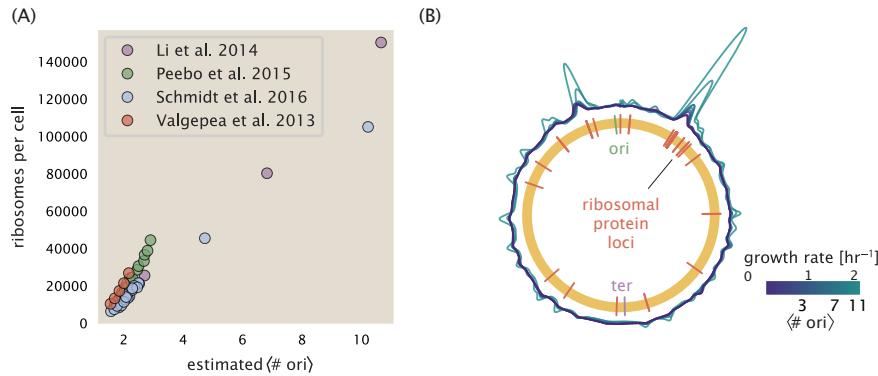


Figure 9. Cells increase both absolute ribosome abundance and Φ_R with $(\# \text{ori})$. (A) Plot of the ribosome copy number estimated from the proteomic data against the estimated $(\# \text{ori})$ (see Appendix ?? for additional details). (B) A running Gaussian average (20 kbp st. dev.) of protein 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. $(\# \text{ori})$ are estimated using the data in (A) and Equation ??.

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-acylated 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 to lower $(\# \text{ori})$ and maintain a smaller cell size in poorer growth conditions (?).

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_e . The elongation rate r_e will depend on how quickly the ribosomes can match codons with their correct amino-acyl tRNA, along with the subsequent steps of peptide bond formation and translocation. This ultimately depends on the cellular concentration 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_e and achievable growth rate as simply depending on the supply of amino acids (and, therefore, also amino-acyl tRNAs), through a parameter r_{AA} in units of AA per second, and the rate of amino acid consumption by protein synthesis ($r_e \times R \times f_a$). This is shown schematically in ??(A) and derived in Appendix ???. Given our observation that protein synthesis and energy production are not limiting, we assume that other molecular players required by ribosomes like 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 $r_{AA} = 5 \times 10^6 \text{ AA} \cdot \text{s}^{-1} \cdot \mu\text{m}^{-3}$ and $f_a = 1$ for a unit cell volume $V = 1\text{fL}$. At low ribosome copy numbers, the observed elongation rate is dependent primarily on $[AA]_{\text{eff}}$ through r_{AA} [as $r_e^{\max} \times R \times f_a << r_{AA}$, 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 the 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 to those associated with the amalgamated proteomic data (described in Appendix ??). 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 observed growth

(A)

A MINIMAL MODEL FOR NUTRIENT-LIMITED GROWTH

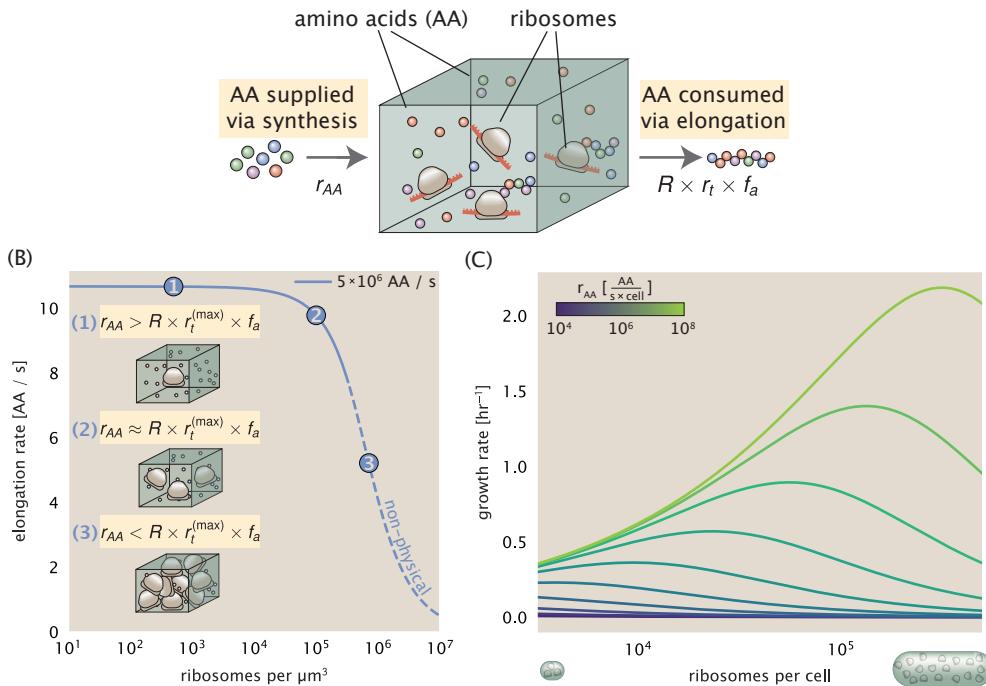


Figure 10. A minimal model of growth rate control under nutrient limitation. (A) We consider a unit volume of cellular material composed of amino acids (colored spheres) provided at a supply rate r_{AA} . These amino acids are polymerized by a pool of ribosomes (brown blobs) at a rate $r_t \times R \times f_a$, 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 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 the parameters $r_{AA} = 5 \times 10^6 \text{ AA} / \text{s}$, $r_t^{(\text{max})} = 17.1 \text{ AA} / \text{s}$, $f_a = 1$, 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.

499 rate depends on the rate of amino acid supply r_{AA} as a function of the cellular ribosome copy number. A feature
500 immediately apparent is the presence of a maximal growth rate whose dependence on R (and consequently, the
501 cell size) increases with increasing r_{AA} . Importantly, however, there is an optimum set of R , N_{pep} , and V that are
502 strictly dependent on the value of r_{AA} . Increasing the ribosomal concentration beyond the cell's metabolic capacity
503 has the adverse consequence of depleting the supply of amino acids and a concomitant decrease in the elongation
504 rate r_f [??(B)].

505 Also of note is the growth rate profiles shown for low amino acid supply rates [purple and blue lines in ??(C)],
506 representing growth in nutrient-poor media. In these conditions, there no longer exists a peak in growth, at least in
507 the range of physiologically-relevant ribosome copy numbers. Instead, cells limit their pool of actively translating
508 ribosomes by decreasing f_a (?), which would help maintain the pool of available amino acids [$AA]_{eff}$ and increase
509 the achievable elongation rate. This observation is in agreement with the central premise of the cellular resource
510 allocation principle proposed by ??? and ?.

511 Discussion

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

524 In general, the copy numbers of the complexes involved in these processes were reasonable agreement with
525 our order-of-magnitude estimates. Since many of these estimates represent soft lower-bound quantities, this
526 suggests that cells do not express proteins grossly in excess of what is needed for a particular growth rate. Several
527 exceptions, however, also highlight the dichotomy between a proteome that appears to "optimize" expression
528 according to growth rate and one that must be able to quickly adapt to environments of different nutritional
529 quality. Take, for example, the expression of carbon transporters. Shown in ??(B), we find that cells always express
530 a similar number of glucose transporters irrespective of growth condition. At the same time, it is interesting to
531 note that many of the alternative carbon transporters are still expressed in low but non-zero numbers (≈ 10 -
532 100 copies per cell) across growth conditions. This may relate to the regulatory configuration for many of these
533 operons, which require the presence of a metabolite signal in order for alternative carbon utilization operons to
534 be induced (?). Furthermore, upon induction, these transporters are expressed and present in abundances in
535 close agreement with a simple estimate.

536 Of the processes illustrated in ??, we arrive at a ribosome-centric view of cellular growth rate control. This
537 is in some sense unsurprising given the long-held observation that *E. coli* and many other organisms vary their
538 ribosomal abundance as a function of growth conditions and growth rate (?). However, through our dialogue
539 with the proteomic data, two additional key points emerge. The first relates to our question of what process sets
540 the absolute speed limit of bacterial growth. While a cell can parallelize many of its processes simply by increas-
541 ing the abundance of specific proteins or firing multiple rounds of DNA replication, this is not so for synthesis of
542 ribosomes (?)(A)). The translation time for each ribosome [≈ 7 min, ?] places an inherent limit on the growth rate
543 that can only be surpassed if the cell were to increase their polypeptide elongation rate, or if they could reduce
544 the total protein and rRNA mass of the ribosome. The second point relates to the long-observed correlations
545 between growth rate and cell size (?), and between growth rate and ribosomal mass fraction. While both trends
546 have sparked tremendous curiosity and driven substantial amounts of research in their own regards, these re-
547 lationships are themselves intertwined. In particular, it is the need for cells to increase their absolute number
548 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 parameter r_{AA} in our minimal model) 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 (??). While *E. coli* get larger as it expresses more ribosomes, an additional constraint begins to arise in energy production due to a 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 in cell size 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 bacteria growth and energy production more generally, we have limited our analysis to the aerobic growth conditions associated with the proteomic data and further consideration will be needed for anaerobic growth.

This work is by no means meant to be an exhaustive 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. An interactive figure that allows exploration of specific protein and protein complex copy numbers is available at [link].

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, Kondev, and Garcia labs for useful discussions. 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.

Competing Interests

The authors declare no competing interests.

Appendix for: Fundamental limits on the rate of bacterial cell division

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

608 **Contents**

Table 1. Overview of proteomic data sets.

Author	Method	Reported Quantity
Taniguchi <i>et al.</i> (2010)	YFP-fusion, cell fluorescence	fg/copies per cell
Valgepea <i>et al.</i> (2012)	mass spectrometry	fg/copies per cell
Peebo <i>et al.</i> (2014)	mass spectrometry	fg/copies per fl
Li <i>et al.</i> (2014)	ribosomal profiling	fg/copies per cell ^a
Soufi <i>et al.</i> (2015)	mass spectrometry	fg/copies per cell
Schmidt <i>et al.</i> (2016)	mass spectrometry	fg/copies per cell ^b

a. The reported values assume that the proteins are long-lived compared to the generation time but are unable to account for post-translational modifications that may alter absolute protein abundances.

b. This mass spectrometry approach differs substantially from the others since in addition to the relative proteome-wide abundance measurements, the authors performed absolute quantification of 41 proteins across all growth conditions (see Section ?? for more details on this).

609 Experimental Details Behind Proteomic Data

610 Here we provide a brief summary of the experiments behind each proteomic data set. The purpose of this section
611 is to identify how the authors arrived at absolute protein abundances. In the following section (Section ??) we will
612 then provide a summary of the final protein abundance measurements that were used throughout the main text.
613 Table ?? provides an overview of the publications we considered. These are predominately mass spectrometry-
614 based, with the exception of the work from ? which used ribosomal profiling, and the fluorescence-based counting
615 done in ?.

616 Fluorescence based measurements

617 In the work of ?, the authors used a chromosomal YFP fusion library where individual strains have a specific gene
618 tagged with a YFP-coding sequence. 1018 of their 1400 attempted strains were used in the work. A fluorescence
619 microscope was used to collect cellular YFP intensities across all these strains. Through automated image analy-
620 sis, the authors normalized intensity measurements by cell size to account for the change in size and expression
621 variability across the cell cycle. Following correction of YFP intensities for cellular autofluorescence, final absolute
622 protein levels were determined by a calibration curve with single-molecule fluorescence intensities. This calibra-
623 tion experiment was performed separately using a purified YFP solution.

624 Ribosomal profiling measurements

625 The work of ? takes a sequencing based approach to estimate protein abundance. Ribosomal profiling, which
626 refers to the deep sequencing of ribosome-protected mRNA fragments, can provide a quantitative measurement
627 of the protein synthesis rate. As long as the protein life-time is long relative to the cell doubling time, it is possible to
628 estimate absolute protein copy numbers. The absolute protein synthesis rate has units of proteins per generation,
629 and for stable proteins will also correspond to the protein copy number per cell.

630 In the experiments, ribosome-protected mRNA is extracted from cell lysate and selected on a denaturing poly-
631 acrylamide gel for deep sequencing (15–45 nt long fragments collected and sequenced by using an Illumina HiSeq
632 2000 in ?). Counts of ribosome footprints from the sequencing data were then corrected empirically for position-
633 dependent biases in ribosomal density across each gene, as well as dependencies on specific sequences including
634 the Shine-Dalgarno sequence. These data-corrected ribosome densities represent relative protein synthesis rates.
635 Absolute protein synthesis rates are obtained by multiplying the relative rates by the total cellular protein per cell.
636 The total protein per unit volume was determined with the Lowry method to quantify total protein, calibrated
637 against bovine serum albumin (BSA). By counting colony-forming units following serial dilution of their cell cul-
638 tures, they then calculated the total protein per cell.

639 **Mass spectrometry measurements**

640 Perhaps not surprisingly, the data is predominantly mass spectrometry based. This is largely due to tremendous
641 improvements in the sensitivity of mass spectrometers, as well as improvements in sample preparation and data
642 analysis pipelines. It is now a relatively routine task to extract protein from a cell and quantify the majority of
643 proteins present by shotgun proteomics. In general, this involves lysing cells, enzymatically digesting the proteins
644 into short peptide fragments, and then introducing them into the mass spectrometer (e.g. with liquid chromatog-
645 raphy and electrospray ionization), which itself can have multiple rounds of detection and further fragmentation
646 of the peptides.

647 Most quantitative experiments rely on labeling protein with stable isotopes, which allow multiple samples to
648 be measured together by the mass spectrometer. By measuring samples of known total protein abundance sim-
649 taneously (i.e. one sample of interest, and one reference), it is possible to determine relative protein abundances.
650 Absolute protein abundances can be estimated following the same approach used above for ribosomal profiling,
651 which is to multiply each relative abundance measurement by the total cellular protein per cell. This is the ap-
652 proach taken by ? and ?, with relative protein abundances determined based on the relative peptide intensities
653 (label free quantification 'LFQ' intensities). For the data of ?, total protein per cell was determined by measuring
654 total protein by the Lowry method, and counting colony-forming units following serial dilution. For the data from
655 ?, the authors did not determine cell quantities and instead report the cellular protein abundances in protein per
656 unit volume by assuming a mass density of 1.1 g/ml, with a 30% dry mass fraction.

657 An alternative way to arrive at absolute protein abundances is to dope in synthetic peptide fragments of known
658 abundance. These can serve as a direct way to calibrate mass spectrometry signal intensities to absolute mass.
659 This is the approach taken by ?. In addition to a set of shotgun proteomic measurements to determine proteome-
660 wide relative abundances, the authors also performed absolute quantification of 41 proteins covering over four
661 orders of magnitude in cellular abundance. Here, a synthetic peptide was generated for each of the proteins,
662 doped into each protein sample, and used these to determine absolute protein abundances of the 41 proteins.
663 These absolute measurements, determined for every growth condition, were then used as a calibration curve to
664 convert proteomic-wide relative abundances into absolute protein abundance per cell. A more extensive discus-
665 sion of the ? data set can be found in Section ??.

666 **Summary of Proteomic Data**

667 In the work of the main text we only used the data from ??. As shown in ??(A), the reported total protein abun-
668 dances in the work of ? and ? differed quite substantially from the other work. For the work of ? this is in part
669 due to a lower coverage in total proteomic mass quantified, though we also noticed that most proteins appear
670 undercounted when compared to the other data.

671 ??(B) summarizes the total protein mass for each data point in our final compiled data set. We note that
672 protein abundances were all scaled so they followed a common growth rate-dependent change in total protein
673 mass. While our inclination initially was to leave reported copy numbers untouched, a notable discrepancy in the
674 scaling total protein per cell between ? and the other data sets forced us to dig deeper into those measurements
675 (compare ? and ? data in ??(A)). The particular trend in ? appears to be due to assumptions of cell size and we
676 provide a more extensive discussion and analysis of that data set in section ???. As a compromise, and in an effort
677 to treat all data equally, we instead scaled all protein abundance values to a data-driven estimate of total protein
678 per cell. Here we used cell size measurements from ??, and an estimate of total protein content through expected
679 dry mass. Total protein per cell was estimated using available data on total DNA, RNA, and protein from ??, which
680 account for the majority of dry mass in the cell. We consider these details in sections ?? and ?? that follows.

681 Lastly, in ?? we show the total proteomic coverage and overlap of proteins quantified across each data set.
682 Here we have used an UpSet diagram (?) to compare the data. Overall, the overlap in quantified proteins is quite
683 high, with 1157 proteins quantified across all data sets. The sequencing based approach of ? has substantially
684 higher coverage compared to the mass spectrometry data sets (3394 genes versus the 2041 genes quantified in
685 the work of ?). However, in terms of total protein mass, the data from ??? each quantify roughly equivalent total
686 protein mass. An exception to this is in the data from ?, where we find that the total protein quantified in ? is
687 90-95 % of the total protein mass (when using the data from ? as a reference).

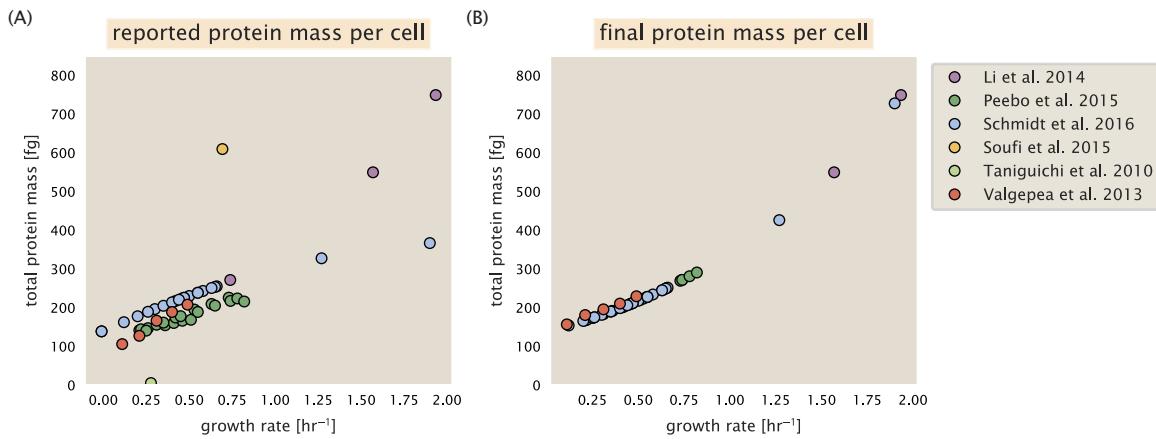


Figure 11. Summary of the growth-rate dependent total protein abundance for each data set. (A) Total protein abundance per cell as originally reported in the data sets of ??????. Note that the data from ? only reported protein abundances per unit volume and total protein per cell was found by multiplying these by the growth-rate dependent cell size as determined by ?. (B) Adjusted total protein abundances across the proteomic data sets are summarized. Protein abundances were adjusted so that all data shared a common set of growth-rate dependent total protein per cell and cellular protein concentration following the cell size expectations of ? (see section on ?? for further details).

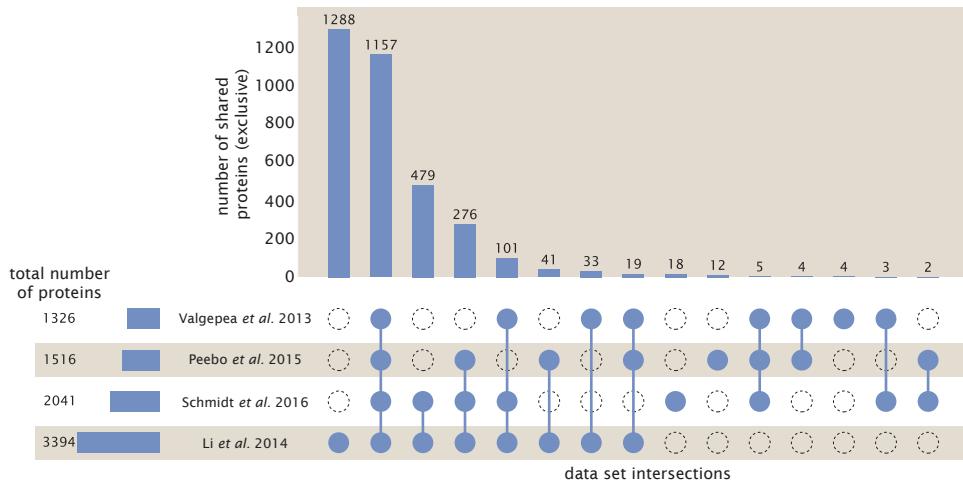


Figure 12. Comparison of proteomic coverage across different data sets. An UpSet diagram (?) summarizes the total number of protein coding genes whose protein abundance was reported in the data sets of ?????. The total number of genes reported in each individual data set are noted on the bottom left, while their overlap is summarized in the bar chart. Each column here refers to the intersection of a set of data sets identified by blue circles.

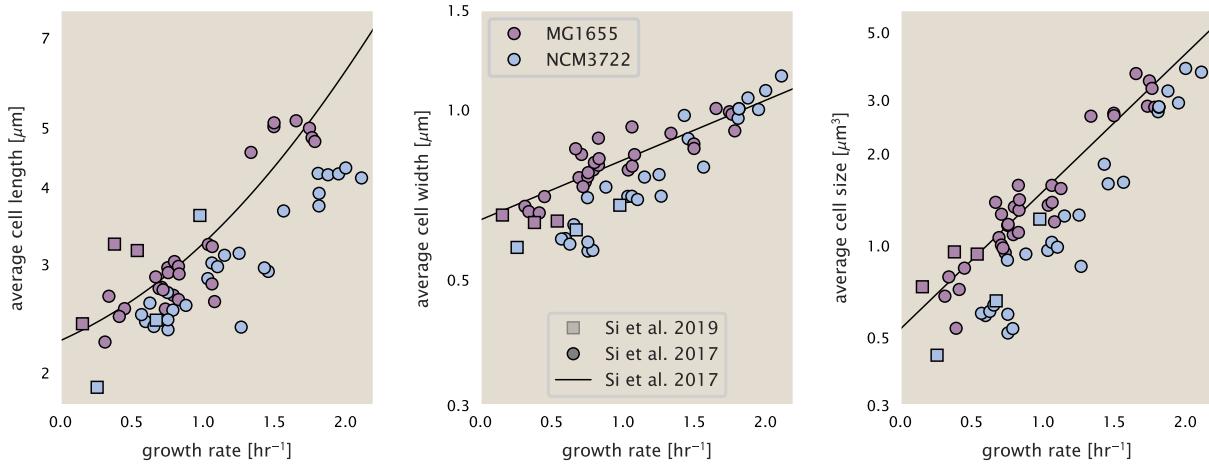


Figure 13. Summary of size measurements from Si et al. 2017, 2019. Cell lengths and widths were measured from cell contours obtained from phase contrast images, and refer to the long and short axis respectively. (A) Cell lengths and (B) cell widths show the mean measurements reported (they report 140-300 images and 5,000-30,000 for each set of samples; which likely means about 1,000-5,000 measurements per mean value reported here since they considered about 6 conditions at a time). Fits were made to the MG1655 strain data; length: $0.5 e^{1.09 \cdot \lambda} + 1.76 \mu\text{m}$, width: $0.64 e^{0.24 \cdot \lambda} \mu\text{m}$. (C) Cell size, V , was calculated as cylinders with two hemispherical ends (Equation ??). The MG1655 strain data gave a best fit of $0.533 e^{1.037 \cdot \lambda} \mu\text{m}^3$.

688 Estimation of Cell Size and Surface Area

689 Since most of the proteomic data sets lack cell size (i.e. volume) measurements, we chose instead to use a common
 690 estimate of size for any analysis requiring cell size or surface area. Since each of the data sets used either K-12
 691 MG1655 or its derivative, BW25113 (from the lab of Barry L. Wanner; the parent strain of the Keio collection (??)),
 692 we fit the MG1655 cell size data from the supplemental material of ?? using the optimize.curve_fit function from
 693 the Scipy python package (?).

694 The average size measurements from each of their experiments are shown in Figure ??, with cell length and
 695 width shown in (A) and (B), respectively. The length data was well described by the exponential function $0.5 e^{1.09 \cdot \lambda}$
 696 + $1.76 \mu\text{m}$, while the width data was well described by $0.64 e^{0.24 \cdot \lambda} \mu\text{m}$. In order to estimate cell size we take the cell
 697 as a cylinders with two hemispherical ends (??). Specifically, cell size is estimated from,
 698

$$V = \pi \cdot r^2 \cdot (l - 2r/3), \quad (1)$$

699 where r is half the cell width. A best fit to the data is described by $0.533 e^{1.037 \cdot \lambda} \mu\text{m}^3$. Calculation of the cell surface
 700 area is given by,

$$S = \eta \cdot \pi \left(\frac{\eta \cdot \pi}{4} - \frac{\pi}{12} \right)^{-2/3} V^{2/3}, \quad (1)$$

702 where η is the aspect ratio ($\eta = l/w$) (?).

703 Estimation of Total Protein Content per Cell

704 In order to estimate total protein per cell for a particular growth rate, we begin by estimating the cell size from the
 705 fit shown in Figure ??(C) ($0.533 e^{1.037 \cdot \lambda} \mu\text{m}^3$). We then estimate the total protein content from the total dry mass of
 706 the cell. Here we begin by noting that for almost the entire range of growth rates considered here, protein, DNA,
 707 and RNA were reported to account for at least 90 % of the dry mass (?). The authors also found that the total dry
 708 mass concentration was roughly constant across growth conditions. Under such a scenario, we can calculate the
 709 total dry mass concentration for protein, DNA, and RNA, which is given by $1.1 \text{ g/ml} \times 30 \% \times 90 \% \text{ or about } [M_p] =$
 710 300 fg per fl . Multiplying this by our prediction of cell size gives the total dry mass per cell.

711 However, even if dry mass concentration is relatively constant across growth conditions, it is not obvious how
 712 protein concentration might vary due to the substantial increase in rRNA at faster growth rates (?). This is a well-
 713 documented result that arises from an increase in ribosomal abundance at faster growth rates (?). To proceed

714 therefore rely on experimental measurements of total DNA content per cell that also come from Basan *et al.*, and
715 RNA to protein ratios that were measured in Dai *et al.* (and cover the entire range of growth conditions considered
716 here). These are reproduced in Figure ??(A) and (B), respectively.

717 Assuming that the protein, DNA, and RNA account for 90 % of the total dry mass, the protein mass can then de-
718 termined by first subtracting the experimentally measured DNA mass, and then using the experimental estimate
719 of the RNA to protein ratio. The total protein per cell is will be related to the summed RNA and protein mass by,
720

$$M_P = \frac{[M_P + M_{RNA}]}{1 + (RP_{ratio})}. \quad (1)$$

721 (RP_{ratio} refers to the RNA to protein ratio as measured by Dai *et al.*. In Figure ??(C) we plot the estimated cellular
722 concentrations for protein, DNA, and RNA from these calculations, and in Figure ??(D) we plot their total expected
723 mass per cell. This later quantity is the growth rate-dependent total protein mass that was used to estimate total
724 protein abundance across all data sets (and summarized in ??(B)).

725 **Estimating Volume and Number of Amino Acids from Ribosome Copy Number**

726 Towards the end of the main text, we examine a coarse-grained model of nutrient-limited growth. A key point
727 in our analysis was to consider how elongation rate r_i and growth rate λ vary with respect to the experimentally
728 observed changes in cell size, total number of peptide bonds per cell N_{pep} , and ribosomal content. In order to
729 do maintain parameters in line with the experimental data, but otherwise allow us to explore the model, we
730 performed a phenomenological fit of N_{pep} and V as a function of the measured ribosomal copy number R . As has
731 been described in the preceding sections of this supplement, we estimate cell volume for each growth condition
732 using the size measurements from ??, and N_{pep} is approximated by taking the total protein mass and dividing this
733 number by the average mass of an amino acid, 110 Da (BNID: 104877).

734 Given the exponential scaling of V and N_{pep} with growth rate, we performed a linear regression of the log trans-
735 form of these parameters as a function of the log transform of the ribosome copy number. Using optimization by
736 minimization, we estimated the best-fit values of the intercept and slope for each regression. ?? shows the result
737 of each regression as a dashed line.

738 **Additional Considerations of Schmidt *et al.* Data Set**

739 While the data set from ? remains a heroic effort that our labs continue to return to as a resource, there were
740 steps taken in their calculation of protein copy number that we felt needed further consideration. In particular,
741 the authors made an assumption of constant cellular protein concentration across all growth conditions and
742 used measurements of cell volume that appear inconsistent with an expected exponential scaling of cell size
743 with growth rate that is well-documented in *E. coli* (??).

744 We begin by looking at their cell volume measurements, which are shown in blue in Figure ???. As a comparison,
745 we also plot cell sizes reported in three other recent papers: measurements from Taheri-Araghi *et al.* and Si *et al.*
746 come from the lab of Suckjoon Jun, while those from Basan *et al.* come from the lab of Terence Hwa. Each set of
747 measurements used microscopy and cell segmentation to determine the length and width, and then calculated
748 cell size by treating the cell is a cylinder with two hemispherical ends, as we considered in the previous section.
749 While there is notable discrepancy between the two research groups, which are both using strain NCM3722, Basan
750 *et al.* found that this came specifically from uncertainty in determining the cell width. This is prone to inaccuracy
751 given the small cell size and optical resolution limits (further described in their supplemental text). Perhaps the
752 more concerning point is that while each of these alternative measurements show an exponential increase in
753 cell size at faster growth rates, the measurements used by Schmidt *et al.* appear to plateau. This resulted in an
754 analogous trend in their final reported total cellular protein per cell as shown in Figure ?? (purple data points), and
755 is in disagreement with other measurements of total protein at these growth rates (?).

756 Since it is not obvious how measurements of cell size influenced their reported protein abundances, in the
757 following subsections we begin by considering this calculation. We then consider three different approaches to
758 estimate the growth-rate dependent total protein mass to compare with those values reported from ?. The results
759 of this are summarized in ??(B), with the original values from both ? and ? shown in ??(A) for reference. For most
760 growth conditions, we find that total protein per cell is still in reasonable agreement. However, for the fastest

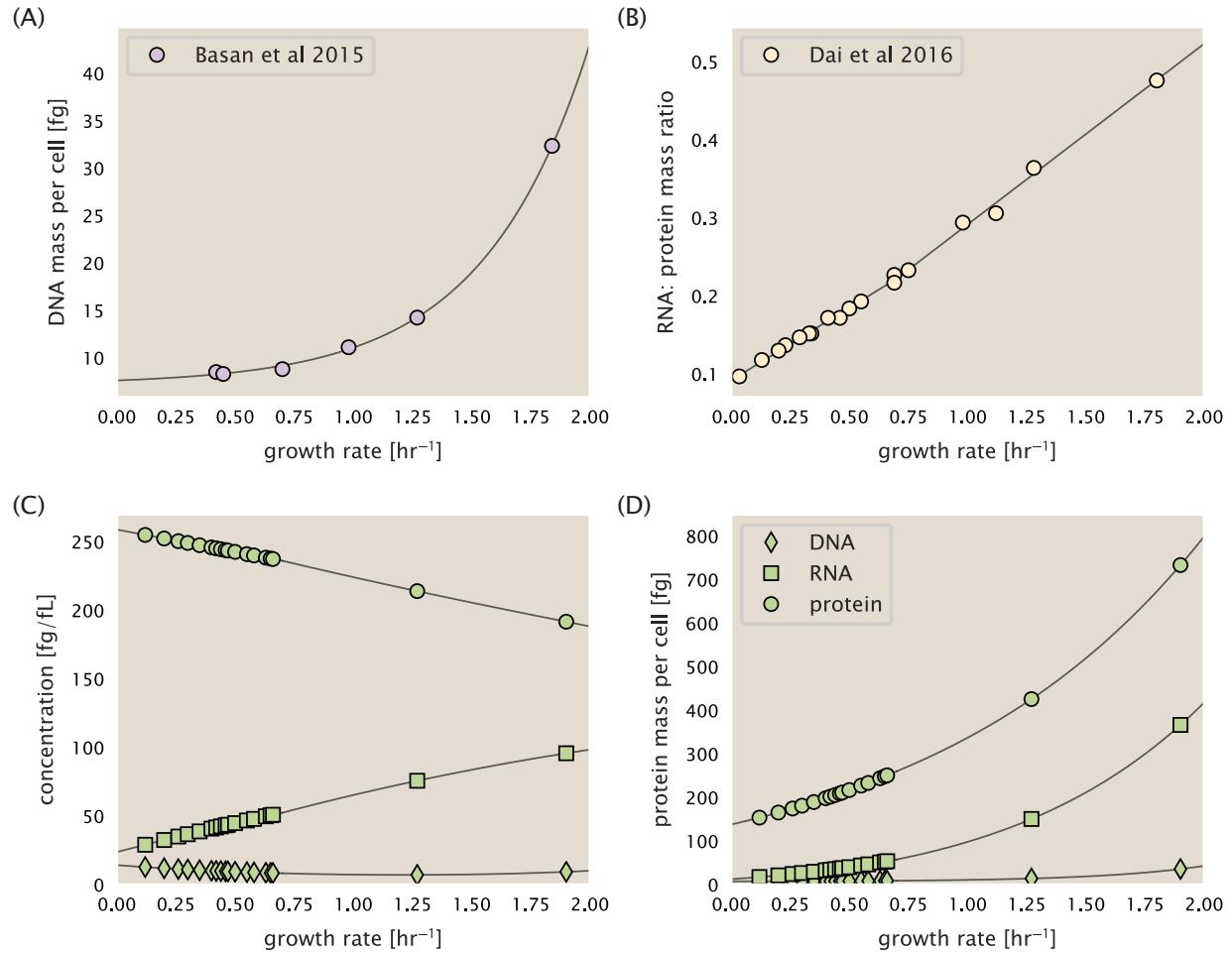


Figure 14. Empirical estimate of cellular protein, DNA, and RNA as a function of growth rate. (A) Measured DNA mass per cell as a function of growth rate, reproduced from Basan *et al.* 2015. The data was fit to an exponential curve (DNA mass in fg per cell is given by $0.42 e^{2.23 \cdot \lambda} + 7.2$ fg per cell, where λ is the growth rate in hr⁻¹). (B) RNA to protein measurements as a function of growth rate. The data was fit to two lines: for growth rates below 0.7 hr⁻¹, the RNA/protein ratio is $0.18 \cdot \lambda + 0.093$, while for growth rates faster than 0.7 hr⁻¹ the RNA/protein ratio is given by $0.25 \cdot \lambda + 0.035$. For (A) and (B) cells are grown under varying levels of nutrient limitation, with cells grown in minimal media with different carbon sources for the slowest growth conditions, and rich-defined media for fast growth rates. (C) Predictions of cellular protein, DNA, and RNA concentration. (D) Total cellular mass predicted for protein, DNA, and RNA using the cell size predictions from Si *et al.*. Symbols (diamond: DNA, square: RNA, circle: protein) show estimated values of mass concentration and mass per cell for the specific growth rates in ?.

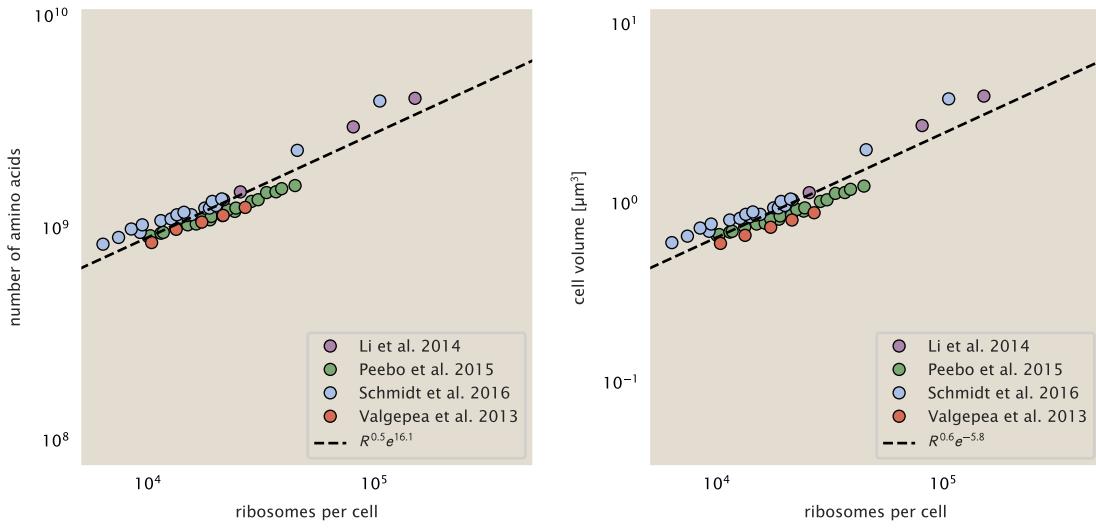


Figure 15. Phenomenological regression of cell volume and number of amino acids per cell as a function of the ribosome copy number. Colored points correspond to the measured value (or calculated value in the case of the cell volume) with colors denoting different data sets. The dashed black line shows the result of the fit, with the functional form of the equation given in the legend with R representing the ribosome copy number.

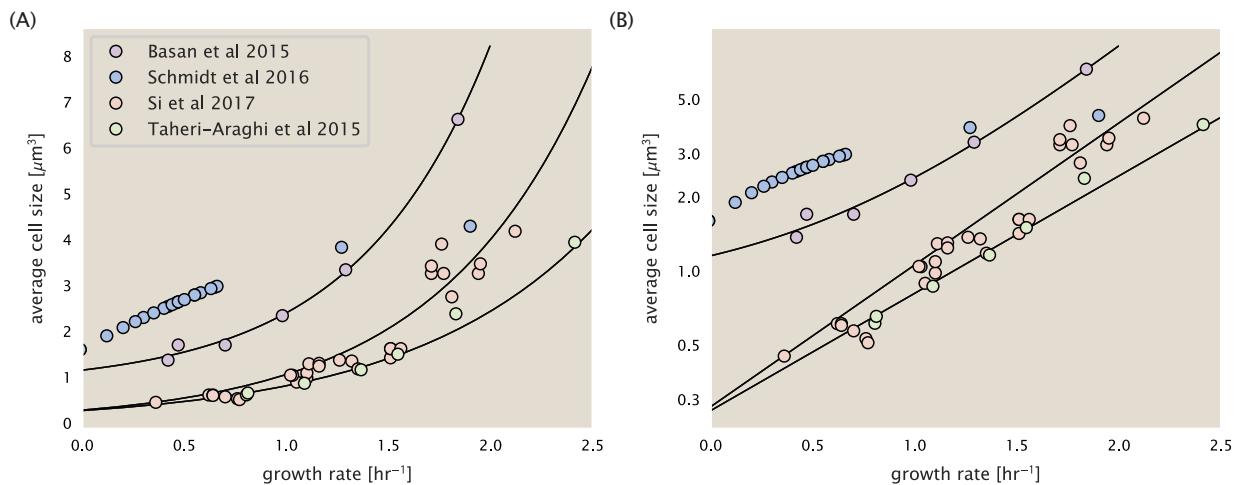


Figure 16. Measurements of cell size as a function of growth rate. (A) Plot of the reported cell sizes from several recent papers. The data in blue come from Volkmer and Heinemann, 2011 (?) and were used in the work of Schmidt *et al.*. Data from the lab of Terence Hwa are shown in purple (?), while the two data sets shown in green and red come from the lab of Suckjoon Jun (??). (B) Same as in (A) but with the data plotted on a logarithmic y-axis to highlight the exponential scaling that is expected for *E. coli*.

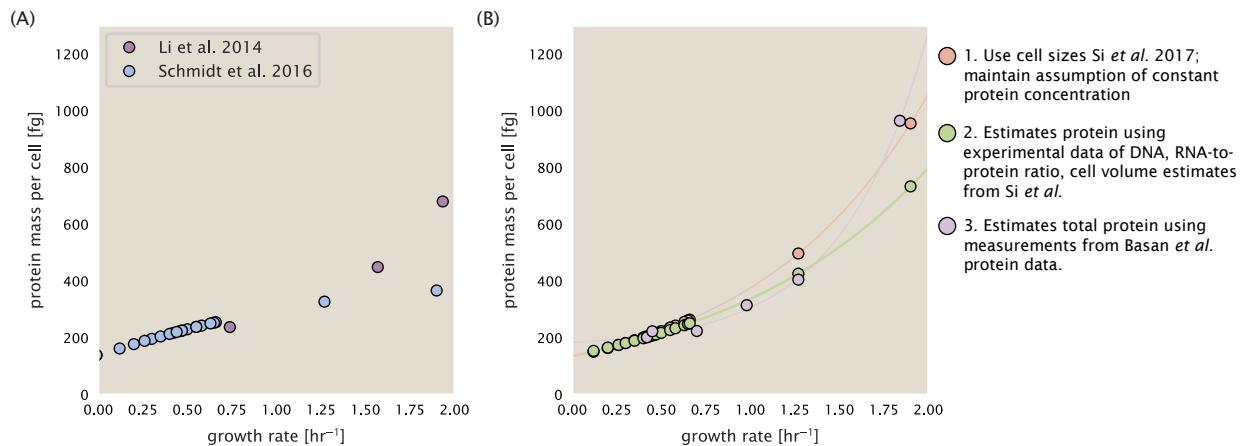


Figure 17. Alternative estimates of total cellular protein for the growth conditions considered in Schmidt *et al.* (A) The original protein mass from Schmidt *et al.* and Li *et al.* are shown in purple and blue, respectively. (B) Three alternative estimates of total protein per cell. 1. light red: Rescaling of total protein mass assuming a growth rate independent protein concentration and cell volumes estimated from Si *et al.* 2017. 2. light green: Rescaling of total protein mass using estimates of growth rate-dependent protein concentrations and cell volumes estimated from Si *et al.* 2017. Total protein per cell is calculated by assuming a 1.1 g/ml cellular mass density, 30% dry mass, with 90% of the dry mass corresponding to DNA, RNA, and protein (?). See ?? for details on calculation. 3. light purple: Rescaling of total protein mass using the experimental measurements from Basan *et al.* 2015.

761 growth conditions, with glycerol + supplemented amino acids, and LB media, all estimates are substantially higher
 762 than those originally reported. This is the main reason why we chose to readjusted protein abundance as shown
 763 in ??(B) (with the calculation described in section ??).

764 Effect of cell volume on reported absolute protein abundances

765 As noted in section ??, the authors calculated proteome-wide protein abundances by first determining absolute
 766 abundances of 41 pre-selected proteins, which relied on adding synthetic heavy reference peptides into their
 767 protein samples at known abundance. This absolute quantitation was performed in replicate for each growth
 768 condition. Separately, the authors also performed a more conventional mass spectrometry measurement for
 769 samples from each growth condition, which attempted to maximize the number of quantified proteins but only
 770 provided relative abundances based on peptide intensities. Finally, using their 41 proteins with absolute abun-
 771 dances already determined, they then created calibration curves with which to relate their relative intensity to
 772 absolute protein abundance for each growth condition. This allowed them to estimate absolute protein abun-
 773 dance for all proteins detected in their proteome-wide data set. Combined with their flow cytometry cell counts,
 774 they were then able to determine absolute abundance of each protein detected on a per cell basis.

775 While this approach provided absolute abundances, another necessary step to arrive at total cellular protein
 776 was to account for any protein loss during their various protein extraction steps. Here the authors attempted
 777 to determine total protein separately using a BCA protein assay. In personal communications, it was noted that
 778 determining reasonable total protein abundances by BCA across their array of growth conditions was particularly
 779 troublesome. Instead, they noted confidence in their total protein measurements for cells grown in M9 minimal
 780 media + glucose and used this as a reference point with which to estimate the total protein for all other growth
 781 conditions.

782 For cells grown in M9 minimal media + glucose an average total mass of $M_p = 240$ fg per cell was measured.
 783 Using their reported cell volume, reported as $V_{orig} = 2.84$ fl, a cellular protein concentration of $[M_p]_{orig} = M_p/V_{orig} =$
 784 85 fg/fl. Now, taking the assumption that cellular protein concentration is relatively independent of growth rate,
 785 they could then estimate the total protein mass for all other growth conditions from,

786

$$M_{P,i} = [M_p]_{orig} \cdot V_i \quad (1)$$

787 where M_{P_i} represents the total protein mass per cell and V_i is the cell volume for each growth condition i as measured in Volkmer and Heinemann, 2011. Here the thinking is that the values of M_{P_i} reflects the total cellular protein for growth condition i , where any discrepancy from their absolute protein abundance is assumed to be due to protein loss during sample preparation. The protein abundances from their absolute abundance measurements noted above were therefore scaled to their estimates and are shown in Figure ?? (purple data points).

792 If we instead consider the cell volumes predicted in the work of Si *et al.*, we again need to take growth in M9
793 minimal media + glucose as a reference with known total mass, but we can follow a similar approach to estimate
794 total protein mass for all other growth conditions. Letting $V_{Si_glu} = 0.6 \text{ fL}$ be the predicted cell volume, the cellular
795 protein concentration becomes $[M_P]_{Si} = M_P / V_{Si_glu} = 400 \text{ fg/fL}$. The new total protein mass per cell can then be
796 calculated from,

797

$$M'_{P,i} = [M_P]_{Si} \cdot V_{Si,i} \quad (1)$$

798 where $M'_{P,i}$ is the new protein mass prediction, and $V_{Si,i}$ refers to the new volume prediction for each condition i ,
799 These are shown as red data points in Figure ??(B).

800 Relaxing assumption of constant protein concentration across growth conditions

801 We next relax the assumption that cellular protein concentration is constant and instead, attempt to estimate it
802 using experimental data. Here we use the estimation of total protein mass per cell detailed in section ?? for all
803 data points in the ? data set. The green data points in ??(B) show this prediction, and this represents the approach
804 used to estimate total protein per cell for all data sets.

805 Comparison with total protein measurements from Basan *et al.* 2015.

806 One of the challenges in our estimates in the preceding sections is the need to estimate protein concentration
807 and cell volumes. These are inherently difficult to accurately due to the small size of *E. coli*. Indeed, for all the
808 additional measurements of cell volume included in Figure ??, no measurements were performed for cells growing
809 at rates below 0.5 hr^{-1} . It therefore remains to be determined whether our extrapolated cell volume estimates
810 are appropriate, with the possibility that the logarithmic scaling of cell size might break down for slower growth.

811 In our last approach we therefore attempt to estimate total protein using experimental data that required no
812 estimates of concentration or cell volume. Specifically, in the work of Basan *et al*, the authors measured total
813 protein per cell for a broad range of growth rates (reproduced in Figure ??). These were determined by first
814 measuring bulk protein from cell lysate, measured by the colorimetric Biuret method (?), and then abundance
815 per cell was calculated from cell counts from either plating cells or a Coulter counter. While it is unclear why
816 Schmidt *et al.* was unable to take a similar approach, the results from Basan *et al* appear more consistent with
817 our expectation that cell mass will increase exponentially with faster growth rates. In addition, although they do
818 not consider growth rates below about 0.5 hr^{-1} , it is interesting to note that the protein mass per cell appears to
819 plateau to a minimum value at slow growth. In contrast, our estimates using cell volume so far have predicted
820 that total protein mass should continue to decrease slightly for slower growing cells. By fitting this data to an
821 exponential function dependent on growth rate, we could then estimate the total protein per cell for each growth
822 condition considered by ?. These are plotted as red data points in ??(B).

823 Calculation of Complex Abundance

824 All protein data quantified the abundance of individual proteins per cell. However, this work requires estimates
825 on the abundance of individual protein *complexes*, rather than the copy number of individual proteins. In this
826 section, we outline the approach we used to annotate proteins as being part of a macromolecular complex and
827 how we computed their absolute abundances per cell.

828 Protein complexes, and proteins individually, often have a variety of names, both longform and shorthand. As
829 individual proteins can have a variety of different synonyms, we sought to ensure that each protein annotated in
830 the data sets used the same synonym. To do use, we relied heavily on the EcoCyc Genomic Database (?). Each
831 protein in available data sets included an annotation of one of the gene name synonyms as well as an accession
832 ID – either a UniProt or Blattner "b-number". We programmatically matched up individual accession IDs between

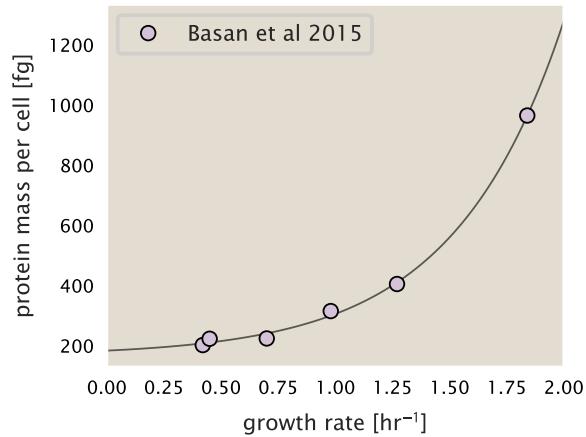


Figure 18. Total cellular protein reported in Basan et al. 2015. Measured protein mass as a function of growth rate as reproduced from Basan *et al.* 2015, with cells grown under different levels of nutrient limitation. The data was fit to an exponential curve where protein mass in fg per cell is given by $14.65 e^{2.180 \cdot \lambda} + 172$ fg per cell, where λ is the growth rate in hr⁻¹.

$$N_{\text{RNA polymerase}}(\lambda) \approx L_{\text{operon}} D_{\text{rRNA}} \rho_{\text{RNA polymerase}}, \quad (1)$$

where L_{operon} is the total length of an rRNA operon (≈ 4500 bp) and $\rho_{\text{RNA polymerase}}$ is packing density of RNA polymerase on a given operon, taken to be 1 RNA polymerase per 80 nucleotides.

Calculation of active ribosomal fraction.

In the main text we used the active ribosomal fraction f_a that was reported in the work of ? to estimate the active ribosomal mass fraction $\Phi_R \times f_a$ across growth conditions. We lacked any specific model to consider how f_a should vary with growth rate, and instead find that the data is well-approximated by fitting to an exponential curve ($f_a = -0.889 e^{4.6 \cdot \lambda} + 0.922$; dashed line in inset of ??(C)). We use this function to estimate f_a for each of the data points shown in ??(C).

Estimation of $\langle \# \text{ori} \rangle / \langle \# \text{ter} \rangle$ and $\langle \# \text{ori} \rangle$.

E. coli shows robust scaling of cell size with the average the number of origins $\langle \# \text{ori} \rangle$ per cell (?). Since protein makes up a majority of the cell's dry mass, the change in cell size is also a reflection of the changes in proteomic composition and total abundance across growth conditions. Given the potential constraints on rRNA synthesis and changes in ribosomal copy number with $\langle \# \text{ori} \rangle$, it becomes important to also consider how protein copy numbers vary with the state of chromosomal replication. This is particularly true when trying to make sense of the changes in ribosomal fraction and growth-rate dependent changes in proteomic composition at a mechanistic level. As considered in the main text, it is becoming increasingly apparent that regulation through the secondary messengers (p)ppGpp may act to limit DNA replication and also reduce ribosomal activity in poorer nutrient conditions. In this context, both $\langle \# \text{ori} \rangle$, as well as the $\langle \# \text{ori} \rangle / \langle \# \text{ter} \rangle$ ratio become important parameters to consider and keep tract of. An increase in $\langle \# \text{ori} \rangle / \langle \# \text{ter} \rangle$ ratio in particular, causes a relatively higher gene dosage in rRNA and r-protein genes due to skew in genes near the origin, where the majority of these are located

In the main text we estimated the change in $\langle \# \text{ori} \rangle$ with growth rate using the nutrient-limited wild-type cell data from ?. We consider their measurements of DNA replication time (t_C , 'C' period of cell division), total cell cycle time (t_{cyc} , 'C' + 'D' period of cell division), and doubling time τ from wild-type *E. coli* growing across a range of growth conditions. Here we show how we estimate this parameter, as well as the $\langle \# \text{ori} \rangle / \langle \# \text{ter} \rangle$ ratio from their data. We begin by considering $\langle \# \text{ori} \rangle$. If the cell cycle time takes longer than the time of cell division, the cell will need to initiate DNA replication more often than its rate of division, $2^{\lambda t} = 2^{\ln(2) \cdot t / \tau}$ to maintain steady state growth. Cells will need to do this in proportion to the ratio $\lambda_{\text{cyc}} / \lambda = t_{\text{cyc}} / \tau$, and the number of origins per cell (on average) is then given by $2^{t_{\text{cyc}} / \tau}$. The average number of termini will in contrast depend on the lag time between DNA replication and cell division, t_D , with $\langle \# \text{ori} \rangle / \langle \# \text{ter} \rangle$ ratio = $2^{t_{\text{cyc}} / \tau - t_D / \tau} = 2^{t_C / \tau}$.

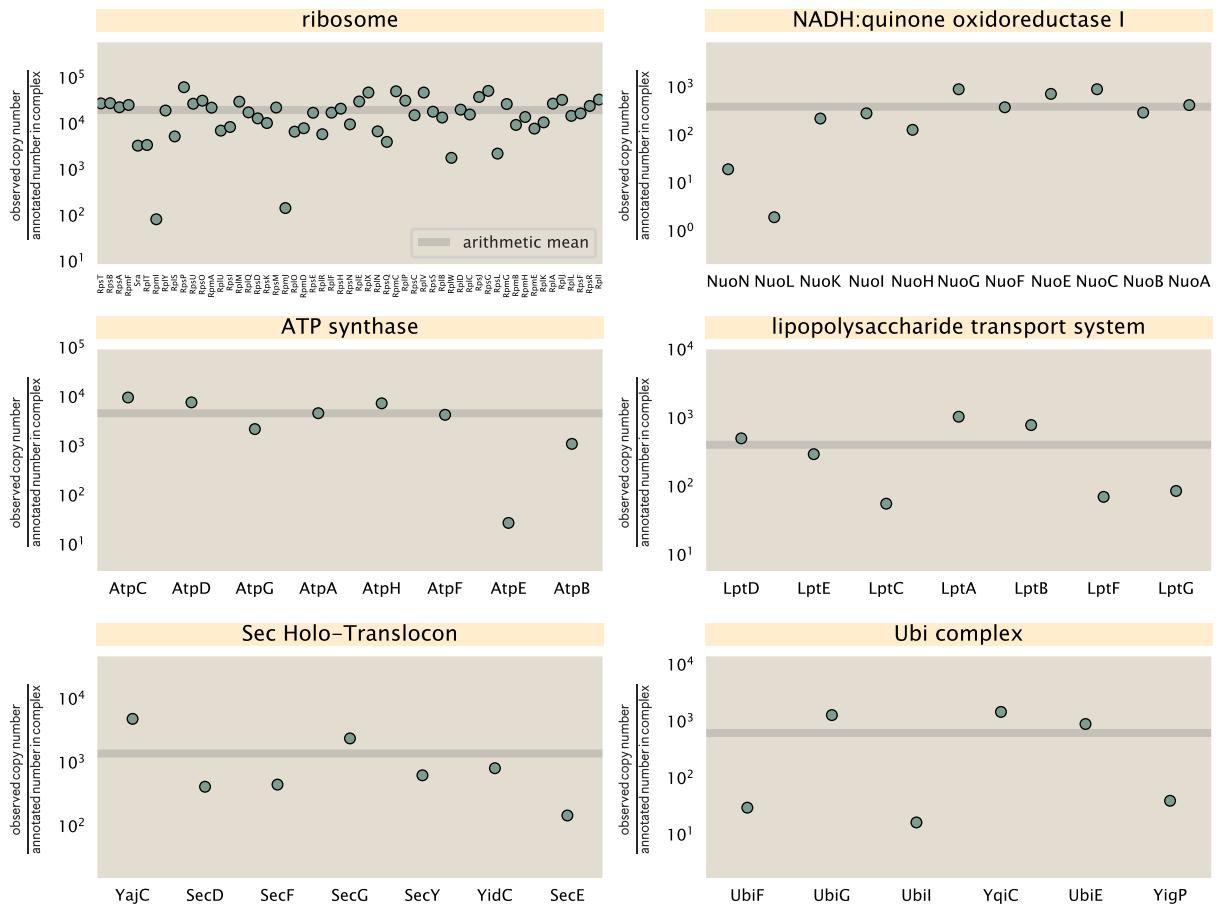


Figure 19. Calculation of the mean complex abundance from measurements of single subunits. Six of the largest complexes (by number of subunits) in *E. coli*. Points correspond to the maximum number of complexes that can be formed given measurement of that individual protein. Solid grey line corresponds to the arithmetic mean across all subunits. These data correspond to measurements from ? in a glucose-supplemented minimal growth medium.

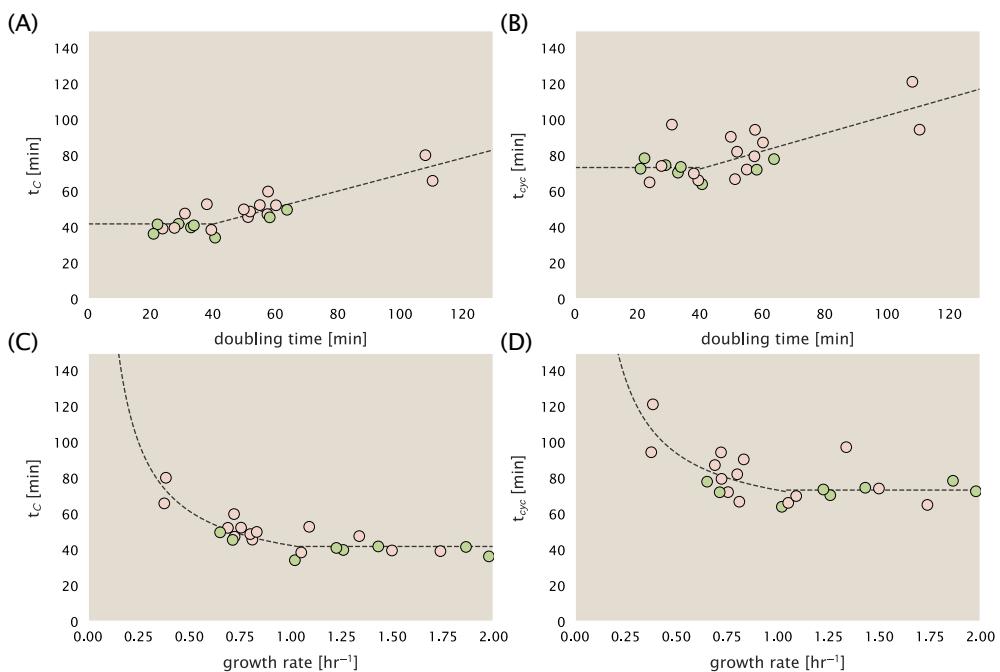


Figure 20. Estimation of $\langle \#ori \rangle / \langle \#ter \rangle$ and $\langle \#ori \rangle$ using data from Si *et al.* (2017). (A) and (B) plot the reported t_C and t_{cyc} as a function of cell doubling time τ , respectively. The dashed lines show a piecewise fit to the data. For short doubling times (rich media), t_C and t_{cyc} are assumed constant ($t_C = 42$ minutes, $t_{cyc} = 73$ minutes). At the transition, taken to occur at 40 minutes, the dashed line corresponds to an assumed proportional increase in each parameter as a function of the doubling time ($t_C = 0.46 \tau + 23.3$ minutes, $t_{cyc} = 0.50 \tau + 52.7$ minutes). (C) and (D) plot the same data as in (A) and (B), but as a function of growth rate, given by $\lambda = \ln(2)/\tau$.

973 In Figure ??(A) and (B) we plot the measured t_C and t_{cyc} values versus the doubling time from ?. The authors
974 estimated t_C by marker frequency analysis using qPCR, while $t_{cyc} = t_C + t_D$ were inferred from t_C and τ . In the
975 plots we see that both t_C and t_{cyc} reach a minimum at around 40 and 75 minutes, respectively. For a C period
976 of 40 minutes, this would correspond to a maximum rate of elongation of about 1,000 bp/sec. Since we lacked
977 a specific model to describe how each of these parameters vary with growth condition, we assumed that they
978 were linearly dependent on the doubling time. For each parameter, t_C and t_{cyc} , we split them up into two domains
979 corresponding to poorer nutrient conditions and rich nutrient conditions (cut off at $\tau \approx 40$ minutes where chro-
980 mosomal replication becomes nearly constant). The fit lines are shown as solid black lines. In Figure ??(C) and (D)
981 we also show t_C and t_{cyc} as a function of growth rate λ along with our piecewise linear fits, which match the plots
982 in the main text.

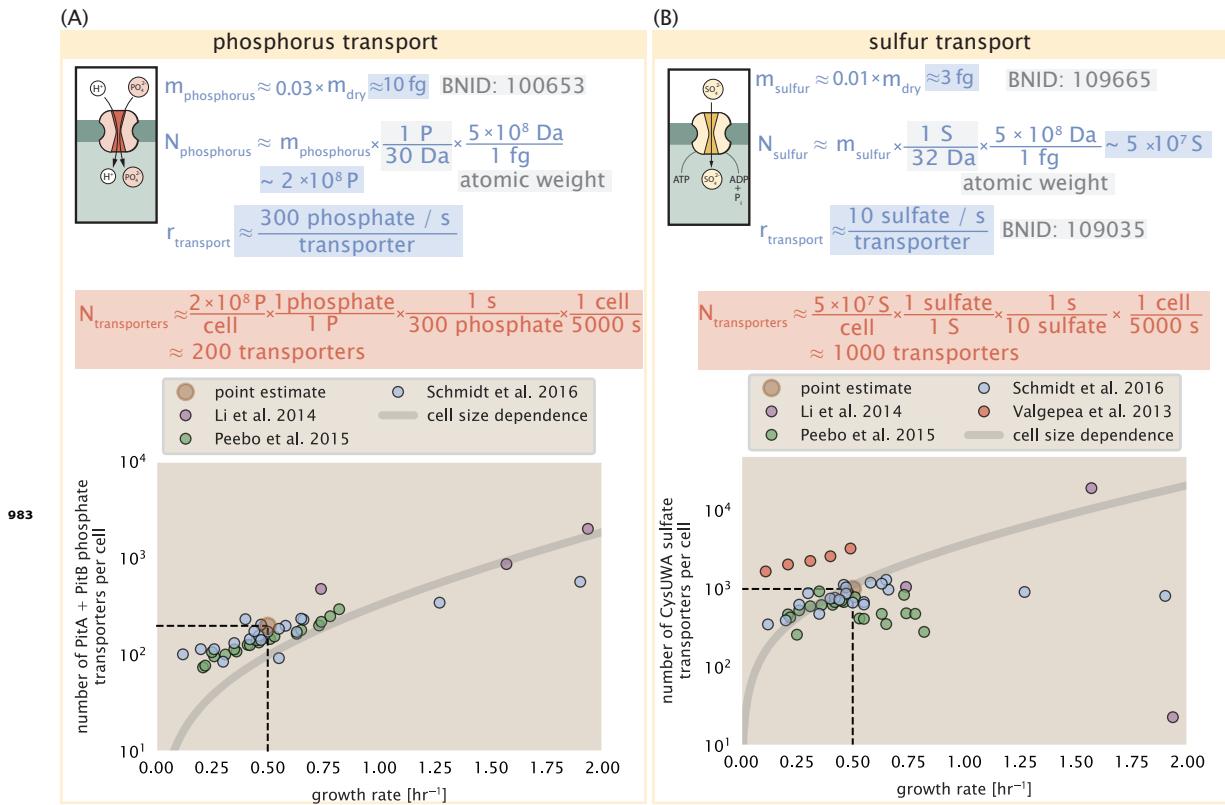


Figure 2-Figure supplement 1. (A) Estimate for the number of PitA phosphate transport systems needed to maintain a 3% phosphorus *E. coli* dry mass. Points in plot correspond to the total number of PitA transporters per cell. (B) Estimate of the number of CysUWA complexes necessary to maintain a 1% sulfur *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.

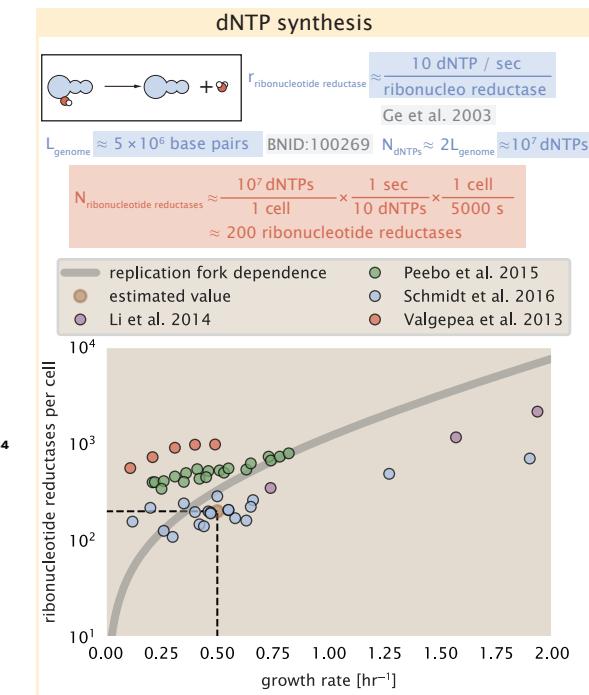


Figure 6–Figure supplement 1. Estimate of the number of ribonucleotide reductase enzymes needed to facilitate the synthesis of $\approx 10^7$ dNTPs over the course of a 5000 second generation time. Points in the plot correspond to the total number of ribonucleotide reductase I ($[\text{NrdA}]_2[\text{NrdB}]_2$) 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.

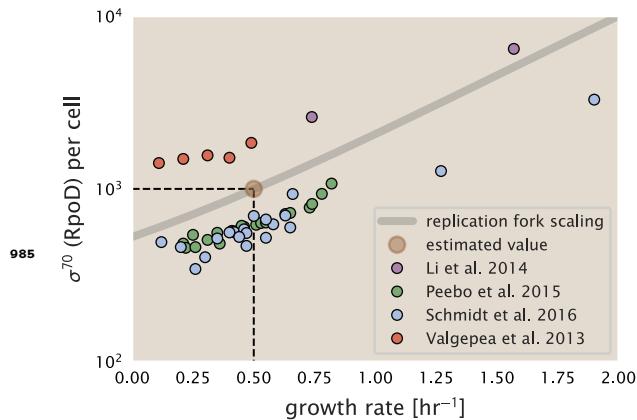


Figure 7–Figure supplement 1. The abundance of σ^{70} as a function of growth rate. Estimated value for the number of RNAP is shown as a translucent brown point and grey line.

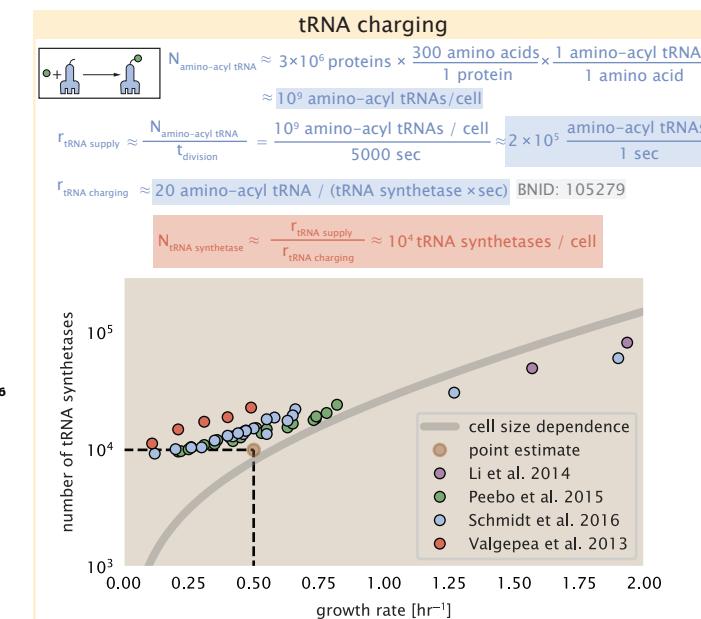


Figure 8-Figure supplement 1. Estimation for the number of tRNA synthetases that will supply the required amino acid demand. The sum of all tRNA synthetases copy numbers are plotted as a function of growth rate ([ArgS], [CysS], [GlnS], [GltX], [IleS], [LeuS], [ValS], [AlaS]₂, [AsnS]₂, [AspS]₂, [TyrS]₂,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