

# Fundamental limits on the rate of bacterial cell division

<sup>3</sup> **Nathan M. Belliveau<sup>†, 1</sup>, Griffin Chure<sup>†, 2, 3</sup>, Christina L. Hueschen<sup>4</sup>, Hernan G.  
<sup>4</sup> Garcia<sup>5</sup>, Jane Kondev<sup>6</sup>, Daniel S. Fisher<sup>7</sup>, Julie A. Theriot<sup>1, 8</sup>, Rob Phillips<sup>2, 9,\*</sup>**

\*For correspondence:

<sup>†</sup>These authors contributed equally to this work

<sup>5</sup> <sup>1</sup>Department of Biology, University of Washington, Seattle, WA, USA; <sup>2</sup>Division of  
<sup>6</sup> Biology and Biological Engineering, California Institute of Technology, Pasadena, CA,  
<sup>7</sup> USA; <sup>3</sup>Department of Applied Physics, California Institute of Technology, Pasadena, CA,  
<sup>8</sup> USA; <sup>4</sup>Department of Chemical Engineering, Stanford University, Stanford, CA, USA;  
<sup>9</sup> <sup>5</sup>Department of Molecular Cell Biology and Department of Physics, University of  
<sup>10</sup> California Berkeley, Berkeley, CA, USA; <sup>6</sup>Department of Physics, Brandeis University,  
<sup>11</sup> Waltham, MA, USA; <sup>7</sup>Department of Applied Physics, Stanford University, Stanford, CA,  
<sup>12</sup> USA; <sup>8</sup>Allen Institute for Cell Science, Seattle, WA, USA; <sup>9</sup>Department of Physics,  
<sup>13</sup> California Institute of Technology, Pasadena, CA, USA; \*Contributed equally

<sup>14</sup> 

---

<sup>15</sup> **Abstract** This will be written next (promise).

<sup>16</sup> 

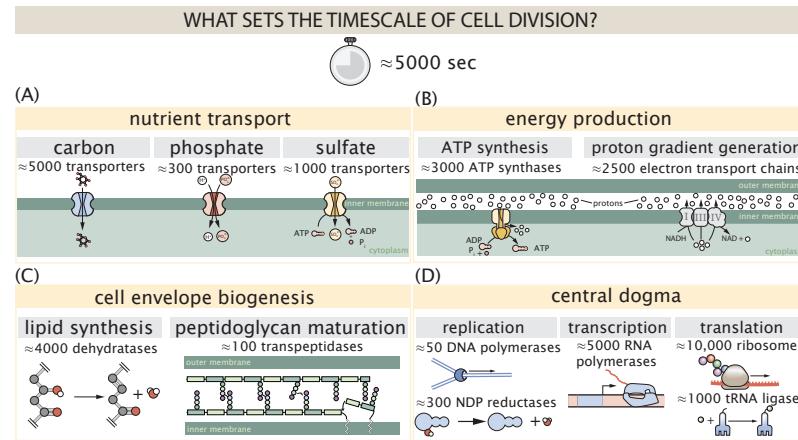
---

## <sup>17</sup> Introduction

<sup>18</sup> The observed range of bacterial growth rates is enormously diverse. In natural environments,  
<sup>19</sup> some microbial organisms might double only once per year (?) while in comfortable laboratory  
<sup>20</sup> conditions, growth can be rapid with several divisions per hour (?). This six order-of-magnitude  
<sup>21</sup> difference in time scales encompasses different microbial species and lifestyles, yet even for a sin-  
<sup>22</sup> gle species such as *E. coli*, the growth rate can be modulated over a similar scale by tuning the  
<sup>23</sup> type and amount of nutrients in the growth medium. This remarkable flexibility in growth rate  
<sup>24</sup> illustrates the intimate relationship between environmental conditions and the rates at which cells  
<sup>25</sup> convert nutrients into new cellular material – a relationship that has remained a major topic of  
<sup>26</sup> inquiry in bacterial physiology for over a century (?).

<sup>27</sup> As was noted by Jacques Monod, “the study of the growth of bacterial cultures does not con-  
<sup>28</sup> stitute a specialized subject or branch of research, it is the basic method of Microbiology.” Those  
<sup>29</sup> words ring as true today as they did when they were written 70 years ago (?). Indeed, the study  
<sup>30</sup> of bacterial growth has undergone a renaissance. Many of the key questions addressed by the  
<sup>31</sup> pioneering efforts in the middle of the last century can be revisited by examining them through  
<sup>32</sup> the lens of the increasingly refined molecular census that is available for bacteria such as the mi-  
<sup>33</sup> crobial workhorse *Escherichia coli*. In this work, we explore an amalgamation of recent proteomic  
<sup>34</sup> data sets to explore fundamental limits of bacterial growth.

<sup>35</sup> Several of the evergreen questions about bacterial growth that were originally raised by micro-  
<sup>36</sup> biologists in the middle of the 20th century can now be reframed in light of this newly available  
<sup>37</sup> data. For example, what biological processes set the absolute speed limit for how fast bacterial  
<sup>38</sup> cells can grow and reproduce? How do cells alter the absolute numbers and relative ratios of  
<sup>39</sup> their molecular constituents as a function of changes in growth rate or nutrient availability? In  
<sup>40</sup> this paper, we address these two questions from two distinct angles. First, as a result of an array  
<sup>41</sup> of high-quality proteome-wide measurements of the *E. coli* proteome under myriad growth condi-



**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) energy production (namely, ATP synthesis), (C) cell envelope biogenesis, 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  $0.5 \text{ hr}^{-1}$ , or a cell doubling time of  $\approx 5000 \text{ s}$ .

42 tions, we have a census that allows us to explore how the number of key molecular players change  
 43 as a function of growth rate. This census provides a window into the question of whether the rates  
 44 of central processes such as energy generation or DNA synthesis are regulated systematically as  
 45 a function of cell growth rate by altering protein copy number in individual cells. Second, by com-  
 46 piling molecular turnover rate measurements for many of the fundamental processes associated  
 47 with bacterial growth, we can make quantitative estimates to determine whether the observed pro-  
 48 tein copy numbers under varying conditions appear to be in excess of what would be minimally  
 49 required to support cell growth at the observed rates.

50 In this paper, we make a series of order-of-magnitude estimates for the copy numbers and  
 51 growth rate dependent expression of a variety of different processes, schematized in ??, informed  
 52 by the collection of proteomic data sets. We use these estimates to explore which, if any, of the  
 53 hypothesis illustrated n ?? may act as molecular bottlenecks that limit bacterial growth. Specifically,  
 54 we leverage a combination of *E. coli* proteomic data sets collected over the past decade using either  
 55 mass spectrometry (???) or ribosomal profiling (?) across 31 unique growth conditions. Through-  
 56 out, our estimates we consider a modest growth rate of  $\approx 0.5 \text{ hr}^{-1}$  corresponding to a doubl-  
 57 ing time of  $\approx 5000$  seconds, as the the data sets heavily sample this regime. While we formulate point  
 58 estimates for the complex abundances at this division time, we consider how these values will vary  
 59 at other growth rates due to changes in cell size, surface area, and chromosome copy number (?).

60 Broadly, we find that for the majority of these estimates the protein copy numbers appear  
 61 well-tuned for the task of cell doubling at a given growth rate. From our analysis, it emerges that  
 62 translation, particularly of ribosomal proteins, is the most plausible candidate for a molecular bot-  
 63 tleneck. We reach this conclusion by considering that translation is 1) a rate limiting step for the  
 64 *fastest* bacterial division, and 2) a major determinant of bacterial growth across the nutrient condi-  
 65 tions we have considered under steady state, exponential growth. This enables us to suggest that  
 66 the long-observed correlation between growth rate and cell size (??) can be simply attributed to  
 67 the increased absolute number of ribosomes per cell under conditions supporting extremely rapid  
 68 growth, a hypothesis which we formally mathematize and explore.

### 113 Uptake of Nutrients

114 We begin our series of estimates by considering the critical transport processes diagrammed in  
 115 ??(A). In order to build new cellular mass, the molecular and elemental building blocks must be scav-

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

71 This work relies heavily on so-called "back-of-the-envelope" estimates to understand the abundances and growth-rate dependences of a variety of molecular complexes. This moniker  
 72 arises from the limitation that any estimate should be able to fit on the back of a postage  
 73 envelope. Therefore, we must draw a set of rules governing our precision and sources of key  
 74 values.  
 75

76 **The rule of "one, few, and ten".** The philosophy behind order-of-magnitude estimates is to  
 77 provide a estimate of the appropriate scale, not a prediction with infinite accuracy. As such,  
 78 we define three different scales of precision in making estimates. The scale of "one" is re-  
 79 served for values that range between 1 and 2. For example, If a particular process has been  
 80 experimentally measured to transport 1.87 protons for a process to occur, we approximate  
 81 this process to require 2 protons per event. The scale of "few" is reserved for values ranging  
 82 between 3 and 5. For example, we will often use Avogadro's number to compute the number  
 83 of molecules in a cell given a concentration and a volume. Rather than using Avogadro's num-  
 84 ber as  $6.02214 \times 10^{23}$ , we will approximate it as  $5 \times 10^{23}$ . Finally, the scale of "ten" is reserved  
 85 for values which we know within an order of magnitude. If a particular protein complex is  
 86 present at 883 copies per cell, we say that it is present in approximately  $10^3$  copies per cell.  
 87 These different scales will be used to arrive at simple estimates that report the expected scale  
 88 of the observed data. Therefore, the estimates presented here should not be viewed as hard-  
 89 and-fast predictions of precise copy numbers, but as approximate lower (or upper) bounds  
 90 for the number of complexes that may be needed to satisfy some cellular requirement.  
 91 Furthermore, we use equality symbols (=) sparingly and frequently defer to approximation ( $\approx$ )  
 92 or scaling ( $\sim$ ) symbols when reporting an estimate. When  $\approx$  is used, we are implicitly stating  
 93 that we are confident in this estimate within a factor of a few. When a scaling symbol  $\sim$  is  
 94 used, we are stating that we are confident in our estimate to within an order of magnitude.

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

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

engaged from the environment in different forms. Carbon, for example, is acquired via the transport of carbohydrates and sugar alcohols with some carbon sources receiving preferential treatment in their consumption (?). Phosphorus, sulfur, and nitrogen, on the other hand, are harvested primarily in the forms of inorganic salts, namely phosphate, sulfate, and ammonia (??????). All of these compounds have different permeabilities across the cell membrane ? and most require some energetic investment either via ATP hydrolysis or through the proton electrochemical gradient to bring the material across the hydrophobic cell membrane. Given the diversity of biological transport mechanisms and the vast number of inputs needed to build a cell, we begin by considering transport of some of the most important cellular ingredients: carbon, nitrogen, oxygen, hydrogen, phosphorus, and sulfur.

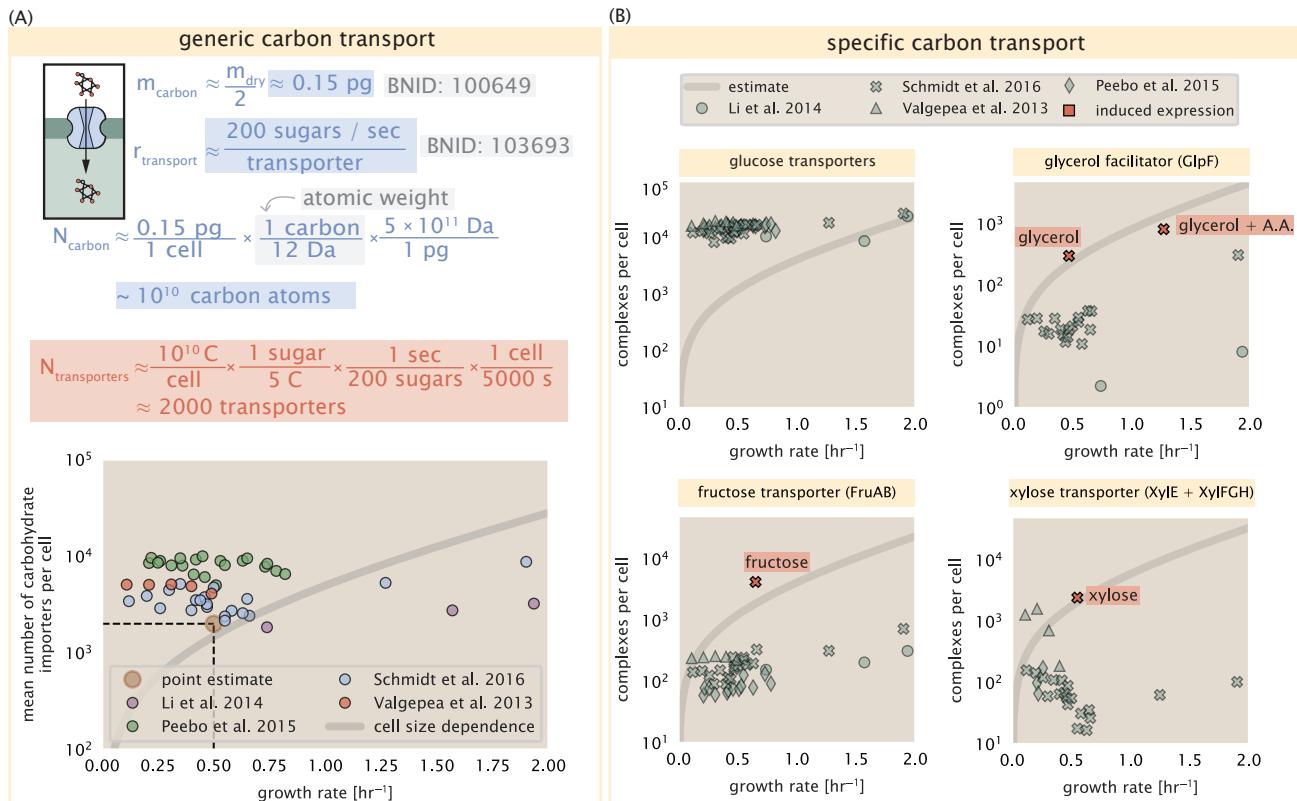
The elemental composition of *E. coli* has received much quantitative attention over the past half century (????), providing us with a starting point for estimating the copy numbers of various transporters. While there is some variability in the exact elemental percentages (with different uncertainties), we can estimate that the dry mass of a typical *E. coli* cell is  $\approx$  45% carbon (BioNumber ID: 100649, see ??),  $\approx$  15% nitrogen (BNID: 106666),  $\approx$  3% phosphorus (BNID: 100653), and 1% sulfur (BNID: 100655). In the coming paragraphs, we will engage in a dialogue between back-of-the-envelope estimates for the numbers of transporters needed to facilitate these chemical stoichiometries and the experimental proteomic measurements of the biological reality. Such an approach provides the opportunity to test if our biological knowledge is sufficient to understand the scale at which these complexes are produced. At the end of this section, we discuss physical limits as to the number of transporters that can be present, and comment on the plausibility of this process acting as a molecular bottleneck.

### 138 **Nitrogen Transport**

We must first address which elemental sources must require proteinaceous transport, meaning that the cell cannot acquire appreciable amounts simply via diffusion across the membrane. The permeability of the lipid membrane to a large number of solutes has been extensively characterized over the past century. Large, polar molecular species (such as various sugar molecules, sulfate, and phosphate) have low permeabilities while small, non-polar compounds (such as oxygen, carbon dioxide, and ammonia) can readily diffuse across the membrane. Ammonia, a primary source of nitrogen in typical laboratory conditions, has a permeability on par with water ( $\sim 10^5$  nm/s, BNID:110824). In particularly nitrogen-poor conditions, *E. coli* expresses a transporter (AmtB) which appears to aid in nitrogen assimilation, though the mechanism and kinetic details of transport are still a matter of debate (??). Beyond ammonia, another plentiful source of nitrogen come in the form of glutamate, which has its own complex metabolism and scavenging pathways. However, nitrogen is plentiful in the growth conditions examined in this work, permitting us to neglect nitrogen transport as a potential rate limiting process in cell division in typical experimental conditions.

### 153 **Carbon Transport**

We begin with the most abundant element in *E. coli* by mass, carbon. Using  $\approx 0.3$  pg as the typical *E. coli* dry mass (BNID: 103904), we estimate that  $\sim 10^{10}$  carbon atoms must be brought into the cell in order to double all of the carbon-containing molecules (??(A, top)). Typical laboratory growth conditions, such as those explored in the aforementioned proteomic data sets, provide carbon as a single class of sugar such as glucose, galactose, or xylose to name a few. *E. coli* has evolved myriad mechanisms by which these sugars can be transported across the cell membrane. One such mechanism of transport is via the PTS system which is a highly modular system capable of transporting a diverse range of sugars (?). The glucose-specific component of this system transports  $\approx 200$  glucose molecules per second per transporter (BNID: 114686). Making the assumption that this is a typical sugar transport rate, coupled with the need to transport  $\sim 10^{10}$  carbon atoms, we arrive at the conclusion that on the order of 1,000 transporters must be expressed in order to bring



**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  $\approx 5$  C, and each transporter conducts sugar molecules at a rate of  $\approx 200$  per second. Bottom plot shows the estimated number of transporters needed at a growth rate of  $\approx 0.5$  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 to compute the continuum growth rate estimate (grey line) were 200 glucose·s<sup>-1</sup> (BNID: 103693), 2000 glycerol·s<sup>-1</sup> (?), 200 fructose·s<sup>-1</sup> (assumed to be similar to PtsL, BNID: 103693), and 50 xylose·s<sup>-1</sup> (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 line in (A) and (B) represents the estimated number of transporters per cell at a continuum of growth rates.

in enough carbon atoms to divide in 5000 s, diagrammed in the top panel of ??(A). This estimate, along with the observed average number of the PTS system carbohydrate transporters present in the proteomic data sets (????), is shown in ??(A). While we estimate 1500 transporters are needed with a 5000 s division time, we can abstract this calculation to consider any particular growth rate given knowledge of the cell density and volume as a function of growth rate and direct the reader to the Supplemental Information for more information. As revealed in ??(A), experimental measurements exceed the estimate by several fold, illustrating that transport of carbon into the cell is not rate limiting for cell division. Abstracting this point estimate at 5000 s to a continuum of growth rates (grey line in ??(A)) reveals an excess of transporters at other growth rates, though in rapid growth regimes, the abundance is below our simple estimate.

The estimate presented in ??(A) neglects any specifics of the regulation of the carbon transport system and presents a view of how many carbohydrate transporters are present on average. Using the diverse array of growth conditions explored in the proteomic data sets, we can explore how individual carbon transport systems depend on the population growth rate. In ??(B), we show the

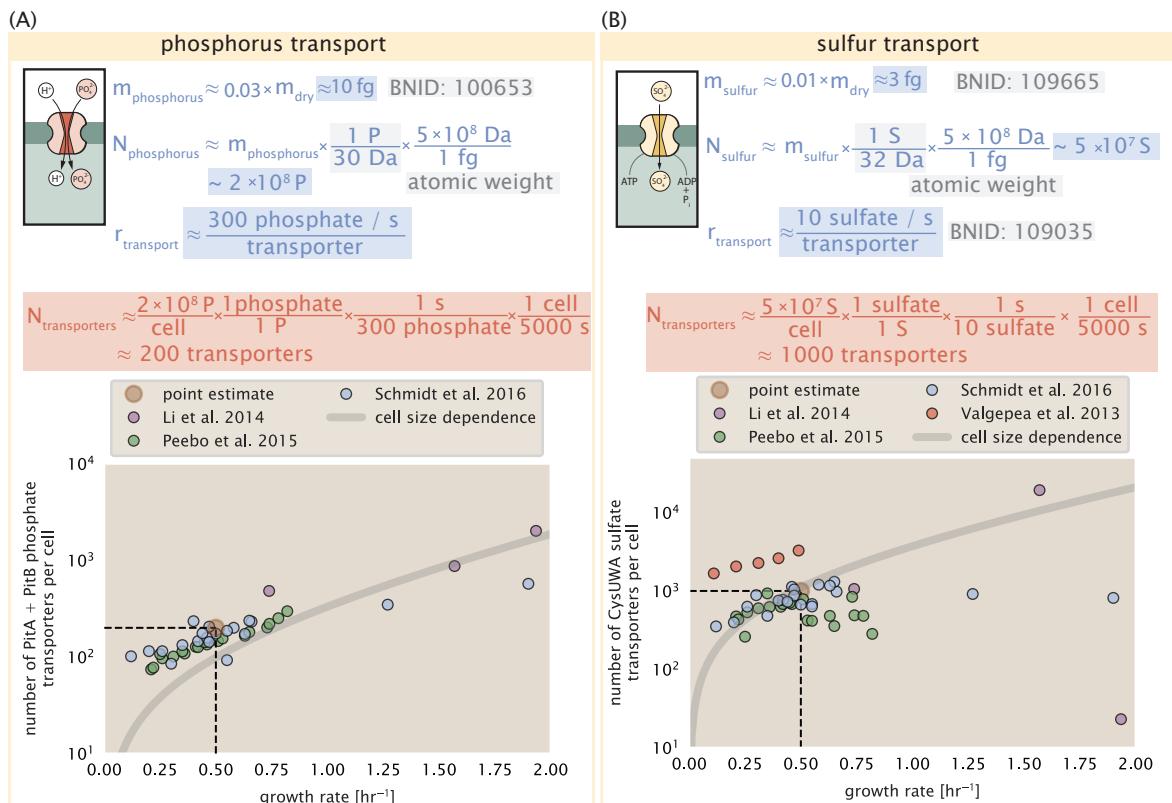
179 total number of carbohydrate transporters specific to different carbon sources. A striking obser-  
 180 vation, shown in the top-left plot of ??(B), is the constancy in the expression of the glucose-specific  
 181 transport systems. Additionally, we note that the total number of glucose-specific transporters is  
 182 tightly distributed at  $\approx 10^4$  per cell, the approximate number of transporters needed to sustain  
 183 rapid growth of several divisions per hour. This illustrates that *E. coli* maintains a substantial num-  
 184 ber of complexes present for transporting glucose regardless of growth rate, which is known to be  
 185 the preferential carbon source (???).

186 It is now understood that a large number of metabolic operons are regulated with dual-input  
 187 logic gates that are only expressed when glucose concentrations are low (mediated by cyclic-AMP  
 188 receptor protein CRP) and the concentration of other carbon sources are elevated (??). A famed  
 189 example of such dual-input regulatory logic is in the regulation of the *lac* operon which is only  
 190 natively activated in the absence of glucose and the presence of allolactose, an intermediate in  
 191 lactose metabolism (?), though we now know of many other such examples (???). This illustrates  
 192 that once glucose is depleted from the environment, cells have a means to dramatically increase  
 193 the abundance of the specific transporter needed to digest the next sugar that is present. Several  
 194 examples of induced expression of specific carbon-source transporters are shown in ??(B). Points  
 195 colored in red (labeled by red text-boxes) correspond to growth conditions in which the specific  
 196 carbon source (glycerol, xylose, or fructose) is present. These plots show that, in the absence of the  
 197 particular carbon source, expression of the transporters is maintained on the order of  $\sim 10^2$  per cell.  
 198 However, when the transport substrate is present, expression is induced and the transporters be-  
 199 come highly-expressed. The grey lines in ??(B) show the estimated number of transporters needed  
 200 at each growth rate to satisfy the cellular carbon requirement. It is notable that in all cases, the  
 201 magnitude of induced expression (shown in red) falls close to the estimate, illustrating the ability  
 202 of the cell to tune expression in response to changing environments. Together, this generic esti-  
 203 mation and the specific examples of induced expression suggest that transport of carbon across  
 204 the cell membrane, while critical for growth, is not the rate-limiting step of cell division.

## 205 **Phosphorus and Sulfur Transport**

206 We now turn our attention towards other essential elements, namely phosphorus and sulfur. Phos-  
 207 phorus is critical to the cellular energy economy in the form of high-energy phosphodiester bonds  
 208 making up DNA, RNA, and the NTP energy pool as well as playing a critical role in the post-translational  
 209 modification of proteins and defining the polar-heads of lipids. In total, phosphorus makes up  
 210  $\approx 3\%$  of the cellular dry mass which in typical experimental conditions is in the form of inorganic  
 211 phosphate. The cell membrane has remarkably low permeability to this highly-charged and critical  
 212 molecule, therefore requiring the expression of active transport systems. In *E. coli*, the proton elec-  
 213 trochemical gradient across the inner membrane is leveraged to transport inorganic phosphate  
 214 into the cell (?). Proton-solute symporters are widespread in *E. coli* (??) and can have rapid trans-  
 215 port rates of 50 to 100 molecules per second for sugars and other solutes (BNID: 103159; 111777).  
 216 As a more extreme example, the proton transporters in the F<sub>1</sub>-F<sub>0</sub> ATP synthase, which use the pro-  
 217 ton electrochemical gradient for rotational motion, can shuttle protons across the membrane at  
 218 a rate of  $\approx 1000$  per second (BNID: 104890; 103390). In *E. coli* the PitA phosphate transport sys-  
 219 tem has been shown to be very tightly coupled with the proton electrochemical gradient with a  
 220 1:1 proton:phosphate stoichiometric ratio (??). Taking the geometric mean of the aforementioned  
 221 estimates gives a plausible rate of phosphate transport on the order of 300 per second. Illustrated  
 222 in ??(A), we can estimate that  $\approx 200$  phosphate transporters are necessary to maintain an  $\approx 3\%$  dry  
 223 mass with a 5000 s division time. This estimate is consistent with observation when we examine  
 224 the observed copy numbers of PitA in proteomic data sets (plot in ??(A)). While our estimate is very  
 225 much in line with the observed numbers, we emphasize that this is likely a slight overestimate of  
 226 the number of transporters needed as there are other phosphorous scavenging systems, such as  
 227 the ATP-dependent phosphate transporter Pst system which we have neglected.

228 Satisfied that there are a sufficient number of phosphate transporters present in the cell, we



**Figure 3. Estimates and measurements of phosphate and sulfate transport systems as a function of growth rate.** (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]<sub>2</sub>[CysU][CysW][Sbp/CysP]. Grey line in (A) and (B) represents the estimated number of transporters per cell at a continuum of growth rates.

now turn to sulfur transport as another potentially rate limiting process. Similar to phosphate, sulfate is highly-charged and not particularly membrane permeable, requiring active transport. While there exists a H<sup>+</sup>/sulfate symporter in *E. coli*, it is in relatively low abundance and is not well characterized (?). Sulfate is predominantly acquired via the ATP-dependent ABC transporter CysUWA system which also plays an important role in selenium transport (??). While specific kinetic details of this transport system are not readily available, generic ATP transport systems in prokaryotes transport on the order of 1 to 10 molecules per second (BNID: 109035). Combining this generic transport rate, measurement of sulfur comprising 1% of dry mass, and a 5000 second division time yields an estimate of  $\approx 1000$  CysUWA complexes per cell (??(B)). Once again, this estimate is in notable agreement with proteomic data sets, suggesting that there are sufficient transporters present to acquire the necessary sulfur. In a similar spirit of our estimate of phosphorus transport, we emphasize that this is likely an overestimate of the number of necessary transporters as we have neglected other sulfur scavenging systems that are in lower abundance.

#### 242 Limits on Transporter Expression

So which, if any, of these processes may be rate limiting for growth? As suggested by ?? (B), induced expression can lead to an order-of-magnitude (or more) increase in the amount of transporters needed to facilitate transport. Thus, if acquisition of nutrients was the limiting state in cell division, could expression simply be increased to accommodate faster growth? A way to approach this question is to compute the amount of space in the bacterial membrane that could be occupied

248 by nutrient transporters. Considering a rule-of-thumb for the surface area of *E. coli* of about 5  
 249  $\mu\text{m}^2$  (BNID: 101792), we expect an areal density for 1000 transporters to be approximately 200  
 250 transporters/  $\mu\text{m}^2$ . For a typical transporter occupying about 50  $\text{nm}^2/\text{dimer}$ , this amounts to about  
 251 only 1 percent of the total inner membrane (?). In addition, bacterial cell membranes typically have  
 252 densities of  $10^5$  proteins/ $\mu\text{m}^2$  (?), implying that the cell could accommodate more transporters of a  
 253 variety of species if it were rate limiting. As we will see in the next section, however, occupancy of  
 254 the membrane can impose other limits on the rate of energy production.

## 255 Energy Production

256 While the transport of nutrients is required to build new cell mass, the metabolic pathways both  
 257 consume and generate energy in the form of NTPs. The high-energy phosphodiester bonds of  
 258 (primarily) ATP power a variety of cellular processes that drive biological systems away from ther-  
 259 modynamic equilibrium. The next set of processes we hypothesize might control the rate of cell  
 260 division considers the energy budget of a dividing cell in terms of the synthesis of ATP from ADP  
 261 and inorganic phosphate as well as maintenance of the electrochemical proton gradient which  
 262 powers it.

## 263 ATP Synthesis

264 Hydrolysis of the terminal phosphodiester bond of ATP forming ADP and an inorganic phosphate is  
 265 a kinetic driving force in a wide array of biochemical reactions. One such reaction is the formation  
 266 of peptide bonds during translation which requires  $\approx 2$  ATPs for the charging of an amino acid  
 267 to the tRNA and  $\approx 2$  ATP equivalents for the formation of the peptide bond between amino acids.  
 268 Considering the ATP costs associated with error correction and post-translational modifications of  
 269 proteins, we can make the approximation that each peptide bond has a net cost of  $\approx 4$  ATP (BNID:  
 270 101442, ?). In total, the energetic costs of peptide bond formation consume  $\approx 80\%$  of the cells ATP  
 271 budget (BNID: 107782; 106158; 101637; 111918, ???). The pool of ATP is produced by the  $F_1$ - $F_0$  ATP  
 272 synthase – a membrane-bound rotary motor which under ideal conditions can yield  $\approx 300$  ATP per  
 273 second (BNID: 114701; ??).

274 To estimate the total number of ATP equivalents consumed during a cell cycle, we will make  
 275 the approximation that there are  $\approx 3 \times 10^6$  proteins per cell with an average protein length of  $\approx$   
 276 300 peptide bonds (BNID: 115702; 108986; 104877, ?). Taking these values together, we estimate  
 277 that the typical *E. coli* cell consumes  $\approx 5 \times 10^9$  ATP per cell cycle on protein synthesis alone and  
 278  $\approx 6 \times 10^9$  ATP in total. Assuming that the ATP synthases are operating at their fastest possible  
 279 rate,  $\approx 3000$  ATP synthases are needed to keep up with the energy demands of the cell. This  
 280 estimate and a comparison with the data are shown in ?? (A). Despite our assumption of maximal  
 281 ATP production rate per synthase and approximation of all NTP consuming reactions being the  
 282 same as ATP, we find that an estimate of a few thousand complete synthases per cell to agree  
 283 well with the experimental data. Much as we did for the estimates of transporter copy number  
 284 in the previous section, we can generalize this estimate to consider a continuum of growth rates  
 285 rather than a point estimate of 5000 s, indicated by the gray lines in ??, and find that this approach  
 286 adequately describes the observed growth rate dependence.

287 If the direct production of ATP was a rate limiting step for growth, could the cell simply express  
 288 more ATP synthase complexes? This requires us to consider several features of cellular physiology,  
 289 namely the physical space on the inner membrane as well as the ability to maintain the proton  
 290 chemical gradient leveraged by the synthase to drive ATP production out of equilibrium.

## 291 Generating the Proton Electrochemical Gradient

292 In order to produce ATP, the  $F_1$ - $F_0$  ATP synthase itself must consume energy. Rather than burning  
 293 through its own product, this intricate macromolecular machine has evolved to exploit the elec-  
 294 trochemical potential established across the inner membrane through cellular respiration. This  
 295 electrochemical gradient is manifest by the pumping of protons into the intermembrane space via

296 the electron transport chains as they reduce NADH. In *E. coli*, this potential difference is  $\approx -200$  mV  
 297 (BNID: 102120, ?). A simple estimate of the inner membrane as a capacitor with a working voltage  
 298 of  $-200$  mV (as performed in the Supplemental Information) reveals that  $\approx 2 \times 10^4$  protons must be  
 299 present in the intermembrane space.

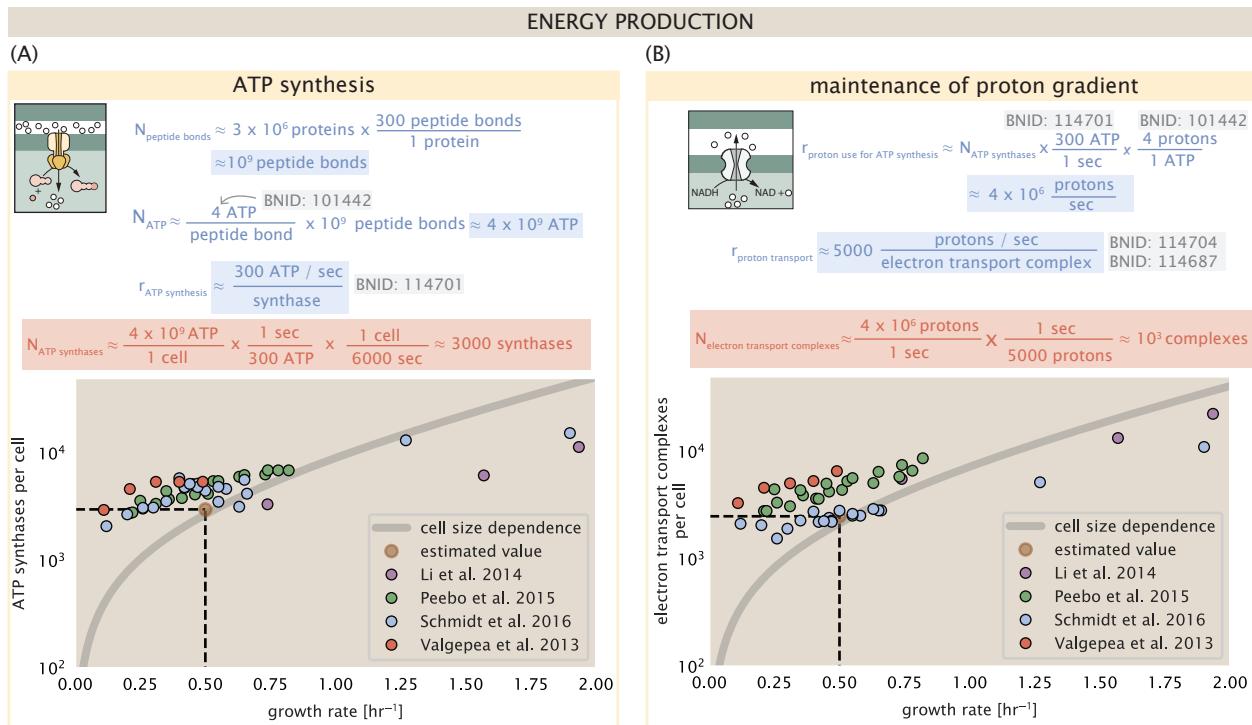
300 However, the constant rotation of the ATP synthases would rapidly abolish this potential difference if it were not being actively maintained. To undergo a complete rotation (and produce a  
 301 single ATP), the F<sub>1</sub>-F<sub>0</sub> ATP synthase must shuttle  $\approx 4$  protons across the membrane into the cytosol  
 302 (BNID: 103390, ?). With  $\approx 3000$  ATP synthases each generating 300 ATP per second, the  $2 \times 10^4$   
 303 protons establishing the 200 mV potential would be consumed in only a few milliseconds. This  
 304 brings us to our next estimate: how many electron transport complexes are needed to support  
 305 the consumption rate of the ATP synthases?

306 The electrochemistry of the electron transport complexes of *E. coli* have been the subject of  
 307 intense biochemical and biophysical study over the past half century (????). A recent work (?) ex-  
 308 amined the respiratory capacity of the *E. coli* electron transport complexes using structural and  
 309 biochemical data, revealing that each electron transport chain rapidly pumps protons into the in-  
 310 termembrane space at a rate of  $\approx 1500$  protons per second (BIND: 114704; 114687, ?). Using our  
 311 estimate of the number of ATP synthases required per cell (??(A)), coupled with these recent mea-  
 312 surements, we estimate that  $\approx 1000$  electron transport complexes would be necessary to facilitate  
 313 the  $\approx 4 \times 10^6$  protons per second diet of the cellular ATP synthases. This estimate (along with a  
 314 generalization to the entire range of observed growth rates) is in agreement with the number of  
 315 complexes identified in the proteomic datasets (plot in ??(B)). This suggests that every ATP syn-  
 316 thase must be accompanied by  $\approx 1$  functional electron transport chain. Again, if this were a rate  
 317 limiting process for bacterial growth, one must conclude that it is not possible for the cell to simply  
 318 increase the production of both the number of electron transport chain complexes as well as ATP  
 319 synthases. As both of these components only function bound to the inner membrane, we now  
 320 turn our attention towards the available space in the membrane as well as surface-area-to-volume  
 321 constraints.

### 322 Energy Production in a Crowded Membrane.

323 For each protein considered so far, the data shows that in general their numbers increase with  
 324 growth rate. This is in part a consequence of the increase in cell length and width at that is com-  
 325 mon to many rod-shaped bacteria at faster growth rates (??). For the particular case of *E. coli*, the  
 326 total cellular protein and cell size increase logarithmically with growth rate (??). Recall however  
 327 that each transport process, as well as the ATP production via respiration, is performed at the bac-  
 328 terial membrane. This means that their maximum productivity can only increase in proportion to  
 329 the cell's surface area divided by the cell doubling time. This difference in scaling would vary in  
 330 proportion to the surface area-to-volume (S/V) ratio. Earlier we found that there was more than  
 331 sufficient membrane real estate for carbon intake in our earlier estimate. However, since the total  
 332 number of ATP synthases and electron chain transport complexes both exhibit a clear increase in  
 333 copy number with growth rate, it was important to also consider the consequences of this S/V ratio  
 334 scaling in more detail.

335 In our estimate of ATP production above we found that a cell demands about  $6 \times 10^9$  ATP per  
 336 cell cycle or  $10^6$  ATP/s. With a cell volume of roughly 1 fL, this corresponds to about  $2 \times 10^{10}$  ATP  
 337 per fL of cell volume, in line with previous estimates (??). In ?? (A) we plot this ATP demand as a  
 338 function of the S/V ratio in green, where we have considered a range of cell shapes from spheri-  
 339 cal to rod-shaped with an aspect ratio (length/width) equal to 4 (See appendix for calculations of  
 340 cell volume and surface area). In order to consider the maximum power that could be produced,  
 341 we consider the amount of ATP that can generated by a membrane filled with ATP synthase and  
 342 electron transport complexes, which provides a maximal production of about 3 ATP / (nm<sup>2</sup>·s) (?).  
 343 This is shown in blue in ??(A), which shows that at least for the growth rates observed, the energy  
 344 demand is roughly an order of magnitude less. Interestingly, ? also found that ATP production



**Figure 4. The abundance of F<sub>1</sub>-F<sub>0</sub> ATP synthases and electron transport chain complexes as a function of growth rate.** (A) Estimate of the number of F<sub>1</sub>-F<sub>0</sub> ATP synthase complexes needed to accommodate peptide bond formation and other NTP dependent processes. Points in plot correspond to the mean number of complete F<sub>1</sub>-F<sub>0</sub> ATP synthase complexes that can be formed given proteomic measurements and the subunit stoichiometry [AtpE]<sub>10</sub>[AtpF]<sub>2</sub>[AtpB][AtpC][AtpH][AtpA]<sub>3</sub>[AtpG][AtpD]<sub>3</sub>. (B) Estimate of the number of electron transport chain complexes needed to maintain a membrane potential of -200 mV given estimate of number of F<sub>1</sub>-F<sub>0</sub> 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.

**346** by respiration is less efficient than by fermentation per membrane area occupied due to the addi-  
**347** tional proteins of the electron transport chain. This suggests that, even under anaerobic growth,  
**348** there will be sufficient membrane space for ATP production in general.

**349** While the analysis in ??(A) serves to highlight the diminishing capacity to provide resources to  
**350** grow if the cell increases in size (and its S/V decreases), maximum energy production represents a  
**351** somewhat unachievable limit since the inner membrane must also include other proteins including  
**352** those required for lipid and membrane synthesis. We used the proteomic data to look at the  
**353** distribution of proteins on the inner membrane, relying on the Gene Ontology (GO) annotations (??)  
**354** to identify all proteins embedded or peripheral to the inner membrane (GO term: 0005886). Those  
**355** associated but not membrane-bound include proteins like MreB and FtsZ, that traverse the inner  
**356** membrane by treadmilling and must nonetheless be considered as a vital component occupying  
**357** space on the membrane. In ?? (B), we find that the total protein mass per  $\mu\text{m}^2$  is surprisingly  
**358** constant across growth rates. Interestingly, when we consider the distribution of proteins grouped  
**359** by their Clusters of Orthologous Groups (COG) (?), the relative abundance for those in metabolism  
**360** (including ATP synthesis via respiration) is also relatively constant across growth rates, suggesting  
**361** that many other membrane associated proteins also increase in similar proportions to proteins  
**362** devoted to energy production ?? (C).

### **363** **Synthesis of the Cell Envelope**

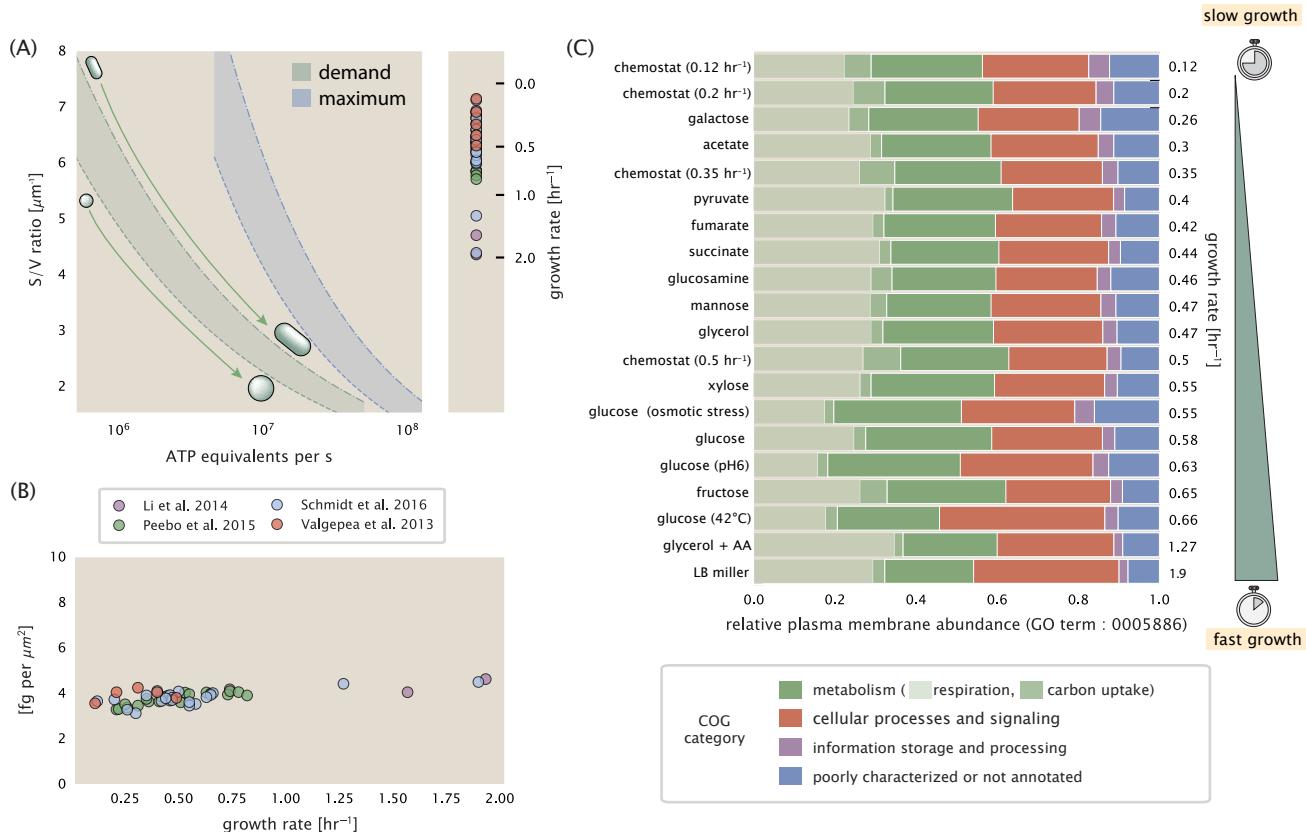
**364** The subjects of our estimates thus far have been localized to the periphery of the cell, embedded  
**365** within the hydrophobic lipid bilayer of the inner membrane. As outlined in ??, cells could in principle  
**366** increase the expression of the membrane-bound ATP synthases and electron transport chains  
**367** to support a larger energy budget across a wide range of cell volumes and membrane surface areas.  
**368** This ability, however, is contingent on the ability of the cell to expand the surface area of the  
**369** cell by synthesizing new lipids and peptidoglycan for the cell wall. In this next class of estimates,  
**370** we will turn our focus to these processes and consider the copy numbers of the relevant enzymes.

### **371** **Lipid Synthesis**

**372** The cell envelopes of gram negative bacteria (such as *E. coli*) are composed of inner and outer  
**373** phospholipid bilayer membranes separated by a  $\approx 10$  nm periplasmic space (BNID: 100016, ?). As  
**374** mentioned in our discussion of the surface area to volume constraints on energy production, *E. coli*  
**375** is a rod-shaped bacterium with a 4:1 length-to-width aspect ratio. At modest growth rates, such as  
**376** our stopwatch of 5000 s, the total cell surface area is  $\approx 5 \mu\text{m}^2$  (BNID: 101792, ?). As there are two  
**377** membranes, each of which composed of two lipid leaflets, the total membrane area is  $\approx 20 \mu\text{m}^2$ , a  
**378** remarkable value compared to the  $\approx 2 \mu\text{m}$  length of the cell.

**379** While this represents the total area of the membrane, this does not mean that it is composed  
**380** entirely of lipid molecules. Rather, the dense packing of the membrane with proteins means that  
**381** only  $\approx 40$  % of the membrane area is occupied by lipids (BNID: 100078, ?). Using a rule-of-thumb  
**382** of  $0.5 \text{ nm}^2$  as the surface area of the typical lipid (BNID: 106993, ?), we arrive at an estimate of  
**383**  $\approx 2 \times 10^7$  lipids per cell, an estimate in close agreement with experimental measurements (BNID:  
**384** 100071, 102996; ?).

**385** The membranes of *E. coli* are composed of a variety of different lipids, each of which are unique  
**386** in their structures and biosynthetic pathways (?). With such diversity in biosynthesis, it becomes  
**387** difficult to identify which step(s) may be the rate-limiting, an objective further complicated by the  
**388** sparsity of *in vivo* kinetic data. Recently, a combination of stochastic kinetic modeling (?) and *in*  
**389** *vitro* kinetic measurements (??) have revealed remarkably slow steps in the fatty acid synthesis  
**390** pathways which may serve as the rate limiting reactions. One such step is the removal of hydroxyl  
**391** groups from the fatty-acid chain by ACP dehydratase that leads to the formation of carbon-carbon  
**392** double bonds. This reaction, catalyzed by proteins FabZ and FabA in *E. coli* (?), have been estimated  
**393** to have kinetic turnover rates of  $\approx 1$  dehydration per second per enzyme (?). Combined with this  
**394** rate, our previous estimates for the number of lipids to be formed, and a 5000 second division



**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 Supplemental XX on calculation). (B) Total protein mass per  $\mu\text{m}^2$  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 bo<sub>3</sub> 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)).

395 yields an estimate that the cell requires  $\approx$  4000 ACP dehydratases. This is in reasonable agreement  
 396 with the experimentally observed copy numbers of FabZ and FabA (??(A)). Furthermore, we can  
 397 extend this estimate to account for the change in membrane surface area as a function of the  
 398 growth rate (grey line in ??(A)), which captures the observed growth rate dependent expression of  
 399 these two enzymes.

400 Despite the slow catalytic rate of FabZ and FabA, we argue that the generation of fatty acids  
 401 is not a bottleneck in cell division and is not the key process responsible for setting the bacterial  
 402 growth rate. Experimental evidence has shown that the rate of fatty-acid synthesis can be dras-  
 403 tically increased *in vitro* by increasing the concentration of FabZ ?. Stochastic simulations of the  
 404 complete fatty acid synthesis pathway of *E. coli* further supports this experimental observation ?.  
 405 Thus, if this step was the determining factor in cell division, increasing growth rate could be as  
 406 simple as increasing the number of ACP dehydratases per cell. With a proteome size of  $\approx 3 \times 10^6$   
 407 proteins, a hypothetical increase in expression from 4000 to 40,000 ACP dehydratases would re-  
 408 sult in a  $\approx 1\%$  increase in the size of the proteome. As many other proteins are in much larger  
 409 abundance than 4000 per cell (as we will see in the coming sections), it is unlikely that expression  
 410 of ACP dehydratases couldn't be increased to facilitate faster growth.

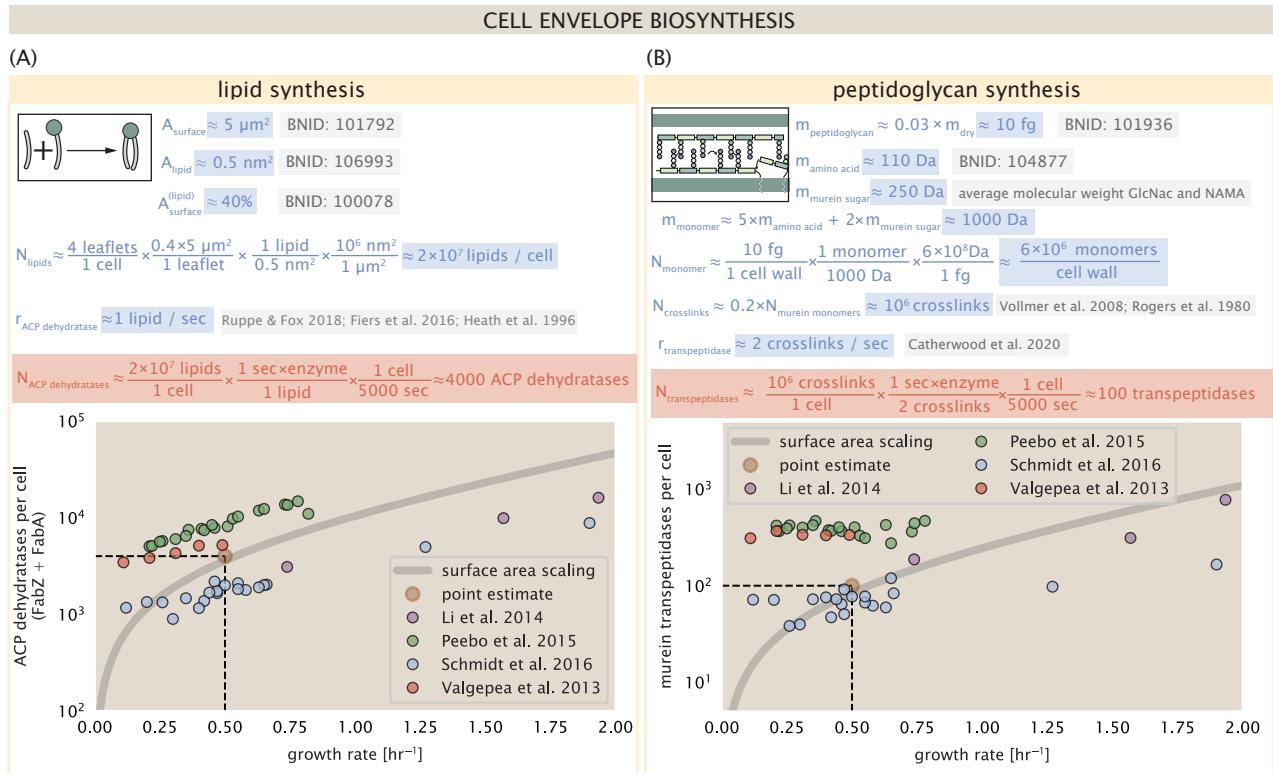
### 411 Peptidoglycan Synthesis

412 While variation in cell size can vary substantially across growth conditions, bacterial cells demon-  
 413 strate exquisite control over their cell shape. This is primarily due to the cell wall, a stiff meshwork  
 414 of polymerized disaccharides interspersed with short peptide crosslinks termed the peptidogly-  
 415 can. The cell wall is also a vital structural component that counteracts turgor pressure. In *E. coli*, this  
 416 enormous peptidoglycan molecule is a few nanometers thick and resides within the periplasmic  
 417 space between the inner and outer membrane. The formation of the peptidoglycan is an intricate  
 418 process, involving the bacterial actin homolog MreB (?) along with a variety of membrane-bound  
 419 and periplasmic enzymes (?). The coordinated action of these components result in a highly-robust  
 420 polymerized meshwork that maintains cell shape even in the face of large-scale perturbations and  
 421 can restore rod-shaped morphology even after digestion of the peptidoglycan (??).

422 In glucose-supported steady-state growth, the peptidoglycan alone comprises  $\approx 3\%$  of the cellular  
 423 dry mass (BNID: 101936, ?), making it the most massive molecule in *E. coli*. The polymerized unit  
 424 of the peptidoglycan is a N-acetylglucosamine and N-acetylmuramic acid disaccharide, of which the  
 425 former is functionalized with a short pentapeptide. With a mass of  $\approx 1000$  Da, this unit, which we  
 426 refer to as a murein monomer, is polymerized to form long strands in the periplasm which are then  
 427 attached to each other via their peptide linkers. Using the aforementioned measurement that  $\approx$   
 428 3% of the dry mass is peptidoglycan, it can be estimated that the peptidoglycan is composed of  $\approx$   
 429  $6 \times 10^6$  murein monomers.

430 During growth, peptidoglycan is constantly being broken down to allow insertion of new murein  
 431 monomers and cellular expansion. In order to maintain structural integrity these monomers must  
 432 be crosslinked into the expanding cell wall, potentially limiting how quickly new material can be  
 433 added and we consider this process as a possible rate-limiting step. In principle, each one of these  
 434 murein monomers can be crosslinked to another glycan strand via the pentapeptide. In some  
 435 species, such as in gram-positive bacterium *Staphylococcus aureus*, the extent of crosslinking can be  
 436 large with > 90% of pentapeptides forming a connection between glycan strands. In *E. coli*, however,  
 437 a much smaller proportion ( $\approx 20\%$ ) of the peptides are crosslinked, resulting in a weaker and more  
 438 porous cell wall ?. The formation of these crosslinks primarily occur during the polymerization of  
 439 the murein monomers and is facilitated by a family of enzymes called transpeptidases. In *E. coli*,  
 440 there are four primary transpeptidases that are involved in lateral and longitudinal extension of  
 441 the peptidoglycan. These transpeptidases have only recently been quantitatively characterized *in*  
 442 *vivo* via liquid chromatography mass spectrometry (?), which revealed a kinetic turnover rate of  
 443  $\approx 1 - 2$  crosslinking reactions formed per second per enzyme.

444 Pulling these measurements together permits us to make an estimate that on the order of  $\approx$



**Figure 6. Estimation of the key components involved in cell envelope biosynthesis.** (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  $\approx 250$  Da. The *in vivo* rate of transpeptidation in *E. coli* was taken from recent analysis by ?. The bottom panel shows experimental measurements of the transpeptidase complexes in *E. coli* following the stoichiometries  $[\text{MrcA}]_2$ ,  $[\text{MrcB}]_2$ ,  $[\text{MrdA}]_1$ , and  $[\text{MrdB}]_1$ . Grey curves in each plot show the estimated number of complexes needed to satisfy the synthesis requirements scaled by the surface area as a function of growth rate. We direct the reader to the supplemental information for a more detailed discussion of this estimate.

445 100 transpeptidases are needed for complete maturation of the peptidoglycan, given a division  
 446 time of  $\approx$  5000 seconds, a value that is closely aligned with the experimental observations (??(B)).  
 447 Expanding this estimate to account for the changing volume of the peptidoglycan as a function of  
 448 growth rates (grey line in ??(B)) also qualitatively captures the observed dependence in the data,  
 449 though systematic disagreements between the different data sets makes the comparison more  
 450 difficult.

451 Much as in the case of fatty acid synthesis, we find it unlikely that the formation of peptidoglycan  
 452 is a rate limiting step in bacterial cell division. The estimate we have presented considered only the  
 453 transpeptidase enzymes that are involved lateral and longitudinal elongation of the peptidoglycan  
 454 (proteins MrdA, MrdB, MrcA, and MrcB). This neglects the presence of other transpeptidases that  
 455 are present in the periplasm and also involved in remodeling and maturation of the peptidoglycan.  
 456 It is therefore possible that if this was setting the speed limit for cell division, the simple expression  
 457 of more transpeptidases may be sufficient to maintain the structural integrity of the cell wall.

### 458 Function of the Central Dogma

459 Up to this point, we have considered a variety of transport and biosynthetic processes that are  
 460 critical to acquiring and generating new cell mass. While there are of course many other metabolic  
 461 processes we could consider and perform estimates of (such as the components of fermentative  
 462 versus aerobic respiration), we now turn our focus to some of the most central processes which  
 463 must be undertaken irrespective of the growth conditions – the processes of the central dogma.

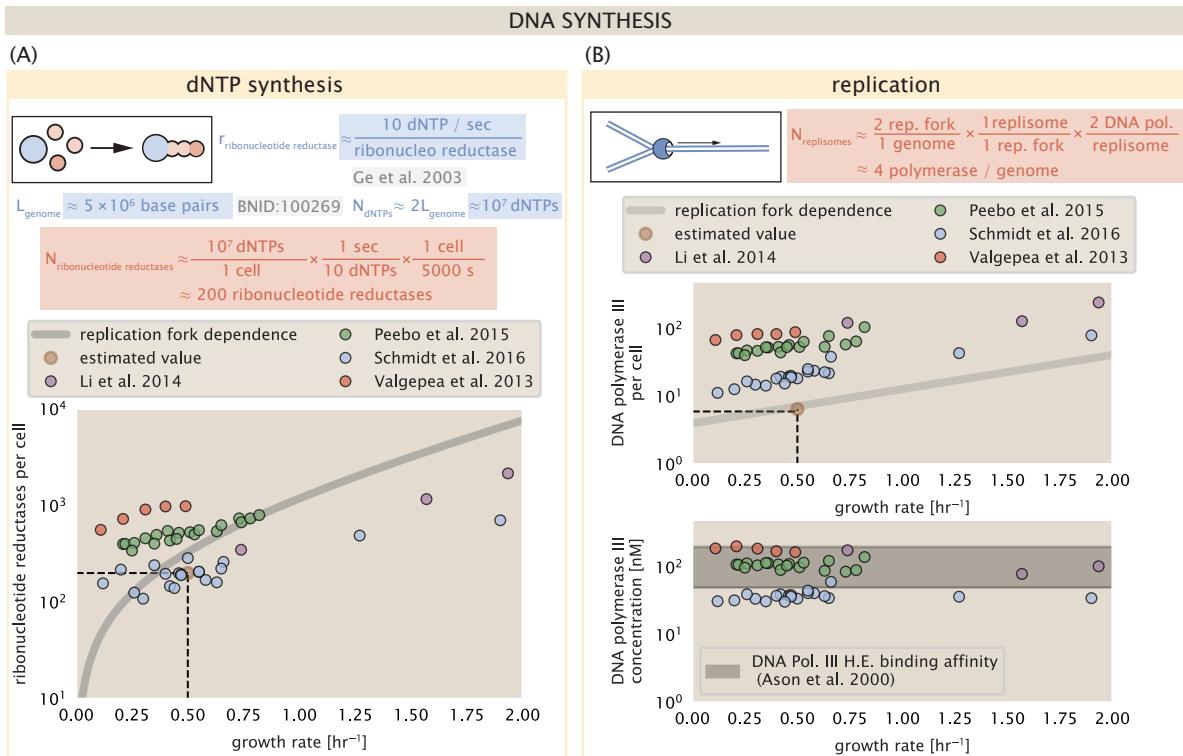
### 464 DNA

465 Most bacteria (including *E. coli*) harbor a single, circular chromosome and can have extra-chromosomal  
 466 plasmids up to  $\sim$  100 kbp in length. We will focus our quantitative thinking solely on the chromo-  
 467 some of *E. coli* which harbors  $\approx$  5000 genes and  $\approx$   $5 \times 10^6$  base pairs. To successfully divide and  
 468 produce viable progeny, this chromosome must be faithfully replicated and segregated into each  
 469 nascent cell. We again rely on the near century of literature in molecular biology to provide some  
 470 insight on the rates and mechanics of the replicative feat as well as the production of the required  
 471 starting materials, dNTPs.

### 472 dNTP synthesis

473 We begin our exploration DNA replication by examining the production of the deoxyribonucleotide  
 474 triphosphates (dNTPs). The four major dNTPs (dATP, dTTP, dCTP, and dGTP) are synthesized *de*  
 475 *novo* in separate pathways, requiring different building blocks. However, a critical step present  
 476 in all dNTP synthesis pathways is the conversion from ribonucleotide to deoxyribonucleotide via  
 477 the removal of the 3' hydroxyl group of the ribose ring (?). This reaction is mediated by a class of  
 478 enzymes termed ribonucleotide reductases, of which *E. coli* possesses two aerobically active com-  
 479 plexes (termed I and II) and a single anaerobically active enzyme. Due to their peculiar formation  
 480 of a radical intermediate, these enzymes have received much biochemical, kinetic, and structural  
 481 characterization. One such work (?) performed a detailed *in vitro* measurement of the steady-state  
 482 kinetic rates of these complexes, revealing a turnover rate of  $\approx$  10 dNTP per second.

483 Since this reaction is central to the synthesis of all dNTPs, it is reasonable to consider the abun-  
 484 dance of these complexes is a measure of the total dNTP production in *E. coli*. Illustrated schemati-  
 485 cally in ?? (A), we consider the fact that to replicate the cell's genome, on the order of  $\approx$   $10^7$  dNTPs  
 486 must be synthesized. Assuming a production rate of 10 per second per ribonucleotide reductase  
 487 complex and a cell division time of 5000 seconds, we arrive at an estimate of  $\approx$  200 complexes  
 488 needed per cell. As shown in the bottom panel of ?? (A), this estimate agrees with the experimen-  
 489 tal measurements of these complexes abundances within  $\approx$  1/2 an order of magnitude. Extension  
 490 of this estimate across a continuum of growth rate, including the fact that multiple chromosomes  
 491 can be replicated at a given time, is shown as a grey transparent line in ??(A). Similarly to our point



**Figure 7. Complex abundance estimates for dNTP synthesis and DNA replication.** (A) 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. (B) An estimate for the minimum number of DNA polymerase holoenzyme complexes needed to facilitate replication of a single genome. Points in the top plot correspond to the total number of DNA polymerase III holoenzyme complexes ( $[\text{DnaE}]_3[\text{DnaQ}]_3[\text{HolE}]_3[\text{DnaX}]_5[\text{HolB}]_5[\text{HolA}]_5[\text{DnaN}]_4[\text{HolC}]_4[\text{HolD}]_4$ ) per cell. Bottom plot shows the effective concentration of DNA polymerase III holoenzyme given cell volumes calculated from the growth rate as in ? (See Supplemental Information Section 4). Grey lines in (A) and top panel of (B) show the estimated number of complexes needed as a function of growth, the details of which are described in the Supplemental Information.

492 estimate, this refinement agrees well with the data, accurately describing both the magnitude of  
493 the complex abundance and the dependence on growth rate.

494 Recent work has revealed that during replication, the ribonucleotide reductase complexes co-  
495 alesce to form discrete foci colocalized with the DNA replisome complex (?). This is particularly  
496 pronounced in conditions where growth is slow, indicating that spatial organization and regula-  
497 tion of the activity of the complexes plays an important role.

#### 498 DNA Replication

499 We now turn our focus to the integration of these dNTP building blocks into the replicated chromo-  
500 some strand via the DNA polymerase. Replication is initiated at a single region of the chromosome  
501 termed the *oriC* locus at which a pair of DNA polymerases bind and begin their high-fidelity repli-  
502 cation of the genome in opposite directions. Assuming equivalence between the two replication  
503 forks, this means that the two DNA polymerase complexes (termed replisomes) meet at the mid-  
504 way point of the circular chromosome termed the *ter* locus. The kinetics of the five types of DNA  
505 polymerases (I – V) have been intensely studied, revealing that DNA polymerase III performs the  
506 high fidelity processive replication of the genome with the other "accessory" polymerases playing  
507 auxiliary roles (?). *In vitro* measurements have shown that DNA Polymerase III copies DNA at a rate  
508 of  $\approx 600$  nucleotides per second (BNID: 104120, ?). Therefore, to replicate a single chromosome,

509 two replisomes (containing two DNA polymerase III each) moving at their maximal rate would copy  
 510 the entire genome in  $\approx$  4000 s. Thus, with a division time of 5000 s (our "typical" growth rate for  
 511 the purposes of this work), there is sufficient time for a pair of replisomes complexes to replicate  
 512 the entire genome. However, this estimate implies that 4000 s would be the upper-limit time scale  
 513 for bacterial division which is at odds with the familiar 20 minute (1200 s) doubling time of *E. coli*  
 514 in rich medium.

515 It is well known that *E. coli* can parallelize its DNA replication such that multiple chromosomes  
 516 are being replicated at once, with as many as 10 - 12 replication forks at a given time (??). Thus,  
 517 even in rapidly growing cultures, we expect only a few polymerases ( $\approx$  10) are needed to replicate  
 518 the chromosome per cell doubling. However, as shown in ??(B), DNA polymerase III is nearly an  
 519 nearly an order of magnitude more abundant. This discrepancy can be understood by considering  
 520 its binding constant to DNA. DNA polymerase III is highly processive, facilitated by a strong affinity  
 521 of the complex to the DNA. *In vitro* biochemical characterization has quantified the  $K_D$  of DNA  
 522 polymerase III holoenzyme to single-stranded and double-stranded DNA to be 50 and 200 nM,  
 523 respectively (?). The bottom plot in ?? (B) shows that the concentration of the DNA polymerase  
 524 III across all data sets and growth conditions is within this range. Thus, while the copy number  
 525 of the DNA polymerase III is in excess of the strict number required to replicate the genome, its  
 526 copy number appears to vary such that its concentration is approximately equal to the dissociation  
 527 constant to the DNA. While the processes regulating the initiation of DNA replication are complex  
 528 and involve more than just the holoenzyme, these data indicate that the kinetics of replication  
 529 rather than the explicit copy number of the DNA polymerase III holoenzyme is the more relevant  
 530 feature of DNA replication to consider. In light of this, the data in ??(B) suggests that for bacteria  
 531 like *E. coli*, DNA replication does no that represent a rate-limiting step in cell division. However,  
 532 it is worth noting that for bacterium like *C. crescentus* whose chromosomal replication is initiated  
 533 only once per cell cycle (?), the time to double their chromosome likely represents an upper limit  
 534 to their growth rate.

### 535 RNA Synthesis

536 With the machinery governing the replication of the genome accounted for, we now turn our attention  
 537 to the next stage of the central dogma – the transcription of DNA to form RNA. We primarily  
 538 consider three major groupings of RNA, namely the RNA associated with ribosomes (rRNA), the  
 539 RNA encoding the amino-acid sequence of proteins (mRNA), and the RNA which links codon sequence  
 540 to amino-acid identity during translation (tRNA). Despite the varied function of these RNA species,  
 541 they share a commonality in that they are transcribed from DNA via the action of RNA polymerase.  
 542 In the coming paragraphs, we will consider the synthesis of RNA as a rate limiting step in bacterial division by estimating how many RNA polymerases must be present to synthesize  
 543 all necessary rRNA, mRNA, and tRNA.  
 544

#### 545 rRNA

546 We begin with an estimation of the number of RNA polymerases needed to synthesize the rRNA  
 547 that serve as catalytic and structural elements of the ribosome. Each ribosome contains three rRNA  
 548 molecules of lengths 120, 1542, and 2904 nucleotides (BNID: 108093, ?), meaning each ribosome  
 549 contains  $\approx$  4500 nucleotides. As the *E. coli* RNA polymerase transcribes DNA to RNA at a rate of  $\approx$   
 550 40 nucleotides per second (BNID: 101904, ?), it takes a single RNA polymerase  $\approx$  100 s to synthesize  
 551 the RNA needed to form a single functional ribosome. Therefore, in a 5000 s division time, a single  
 552 RNA polymerase transcribing rRNA at a time would result in only  $\approx$  50 functional ribosomal rRNA  
 553 units – far below the observed number of  $\approx$  10<sup>4</sup> ribosomes per cell.

554 Of course, there can be more than one RNA polymerase transcribing the rRNA genes at any  
 555 given time. To elucidate the *maximum* number of rRNA units that can be synthesized given a single  
 556 copy of each rRNA gene, we will consider a hypothesis in which the rRNA operon is completely tiled  
 557 with RNA polymerase. *In vivo* measurements of the kinetics of rRNA transcription have revealed

558 that RNA polymerases are loaded onto the promoter of an rRNA gene at a rate of  $\approx 1$  per second  
 559 (BNID: 111997; 102362, ?). If RNA polymerases are being constantly loaded on to the rRNA genes  
 560 at this rate, then we can assume that  $\approx 1$  functional rRNA unit is synthesized per second. With a  
 561 5000 second division time, this hypothesis leads to a maximal value of 5000 functional rRNA units,  
 562 still undershooting the observed number of  $10^4$  ribosomes per cell.

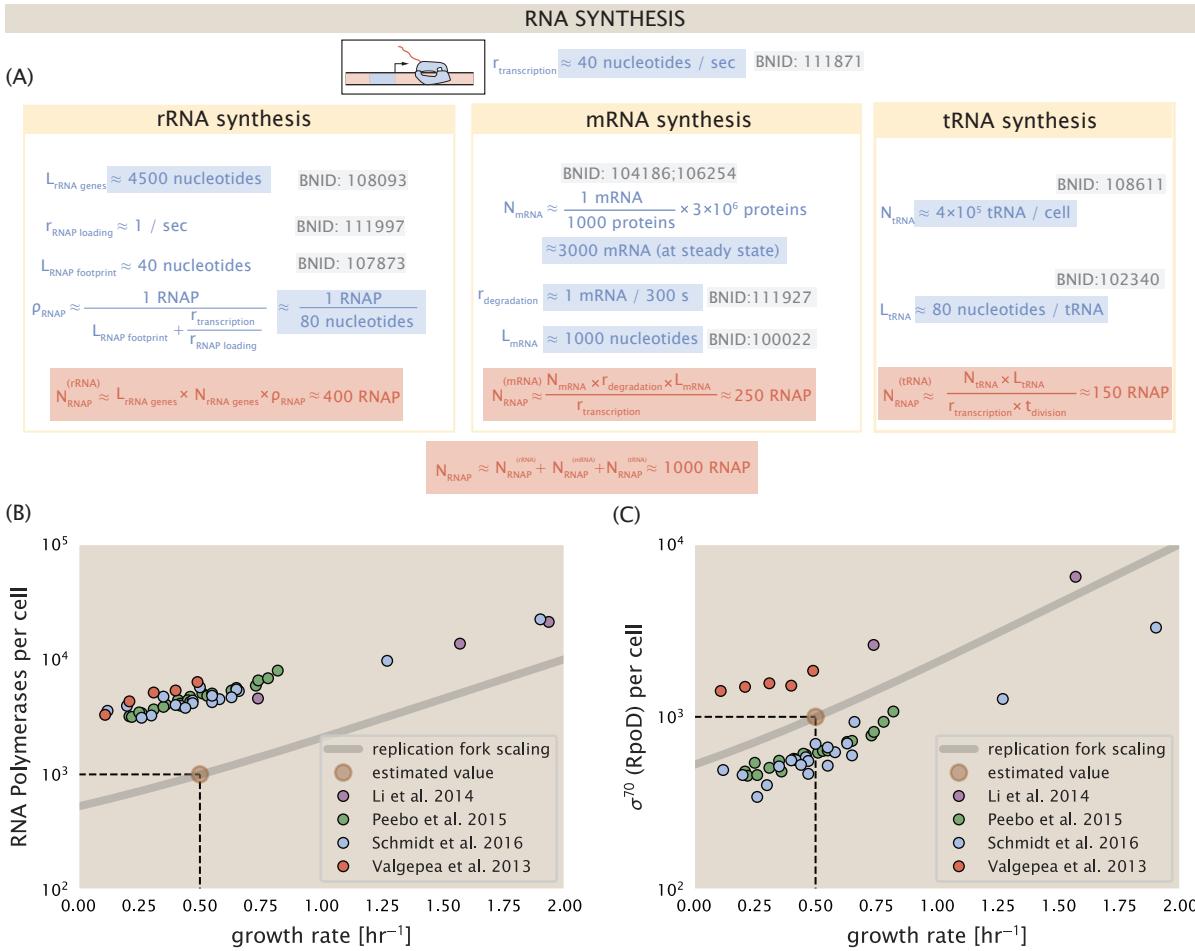
563 *E. coli*, like many other bacterium, have evolved a clever mechanism to surpass this kinetic limit  
 564 for the rate of rRNA production. Rather than having only one copy of each rRNA gene, *E. coli* has  
 565 seven copies of the operon (BIND: 100352, ?) four of which are localized directly adjacent to the  
 566 origin of replication (?). As fast growth also implies an increased gene dosage due to parallelized  
 567 chromosomal replication, the total number of rRNA genes can be on the order of  $\approx 10 - 70$  copies  
 568 at moderate to fast growth rates (?). Using our standard time scale of a 5000 second division time,  
 569 we can make the lower-bound estimate that the typical cell will have 7 copies of the rRNA operon.  
 570 Synthesizing one functional rRNA unit per second per rRNA operon, a total of  $4 \times 10^4$  rRNA units  
 571 can be synthesized, comfortably above the observed number of ribosomes per cell.

572 How many RNA polymerases are then needed to constantly transcribe 7 copies of the rRNA  
 573 genes? We approach this estimate by considering the maximum number of RNA polymerases tiled  
 574 along the rRNA genes with a loading rate of 1 per second and a transcription rate of 40 nucleotides  
 575 per second. Considering that a RNA polymerase has a physical footprint of approximately 40 nu-  
 576 cleotides (BNID: 107873, ?), we can expect  $\approx 1$  RNA polymerase per 80 nucleotides. With a total  
 577 length of  $\approx 4500$  nucleotides per operon and 7 operons per cell, the maximum number of RNA  
 578 polymerases that can be transcribing rRNA at any given time is  $\approx 400$ . As we will see in the coming  
 579 sections, the synthesis of rRNA is the dominant requirement of the RNA polymerase pool.

#### 580 mRNA

581 To form a functional protein, all protein coding genes must first be transcribed from DNA to form an  
 582 mRNA molecule. While each protein requires an mRNA blueprint, many copies of the protein can  
 583 be synthesized from a single mRNA. Factors such as strength of the ribosomal binding site, mRNA  
 584 stability, and rare codon usage frequency dictate the number of proteins that can be made from  
 585 a single mRNA, with yields ranging from  $10^1$  to  $10^4$  (BNID: 104186; 100196; 106254, ?). Computing  
 586 the geometric mean of this range yields  $\approx 1000$  proteins synthesized per mRNA, a value that agrees  
 587 with experimental measurements of the number of proteins per cell ( $\approx 3 \times 10^6$ , BNID: 100088, ?)  
 588 and total number of mRNA per cell ( $\approx 3 \times 10^3$ , BNID: 100064, ?).

589 This estimation captures the *steady-state* mRNA copy number, meaning that at any given time,  
 590 there will exist approximately 3000 unique mRNA molecules. To determine the *total* number of  
 591 mRNA that need to be synthesized over the cell's lifetime, we must consider degradation of the  
 592 mRNA. In most bacteria, mRNAs are rather unstable with life times on the order of several minutes  
 593 (BNID: 104324; 106253; 111927; 111998, ?). For convenience, we assume that the typical mRNA in  
 594 our cell of interest has a typical lifetime of  $\approx 300$  seconds. Using this value, we can determine  
 595 the total mRNA production rate to maintain a steady-state copy number of 3000 mRNA per cell.  
 596 While we direct the reader to the appendix for a more detailed discussion of mRNA transcriptional  
 597 dynamics, we state here that the total mRNA production rate must be on the order of  $\approx 15$  mRNA  
 598 per second. In *E. coli*, the average protein is  $\approx 300$  amino acids in length (BNID: 108986, ?), meaning  
 599 that the corresponding mRNA is  $\approx 900$  nucleotides which we will further approximate as  $\approx 1000$   
 600 nucleotides to account for the non-protein coding regions on the 5' and 3' ends. This means that  
 601 the cell must have enough RNA polymerase molecules about to sustain a transcription rate of  
 602  $\approx 1.5 \times 10^4$  nucleotides per second. Knowing that a single RNA polymerase polymerizes RNA at a  
 603 clip of 40 nucleotides per second, we arrive at a comfortable estimate of  $\approx 250$  RNA polymerase  
 604 complexes needed to satisfy the mRNA demands of the cell. It is worth noting that this number is  
 605 approximately half of that required to synthesize enough rRNA, as we saw in the previous section.  
 606 We find this to be a striking result as these 250 RNA polymerase molecules are responsible for the  
 607 transcription of the  $\approx 4000$  protein coding genes that are not ribosome associated.



**Figure 8. 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. Bionumber Identifiers (BNIDs) are provided for key quantities used in the estimates. (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  $[RpoA]_2[RpoC][RpoB]$ . (C) The abundance of  $\sigma^{70}$  as a function of growth rate. Estimated value for the number of RNAP is shown in (B) and (C) as a translucent brown point at a growth rate of  $0.5 \text{ hr}^{-1}$ .

### 608 tRNA

609 The final class of RNA molecules worthy of quantitative consideration are the tRNAs that are used  
 610 during translation to map codon sequence on mRNA to specific amino acids. Unlike mRNA or rRNA,  
 611 each individual tRNA is remarkably short, ranging from 70 to 95 nucleotides each (BNID: 109645;  
 612 102340, ?). What they lack in length, they make up for in abundance, with reported values ranging  
 613 from  $\approx 6 \times 10^4$  (BNID: 105280, ?) to  $\approx 4 \times 10^5$  (BNID: 108611). To test tRNA synthesis as a possible  
 614 growth-rate limiting stage, we will err towards a higher abundance of  $\approx 4 \times 10^5$  per cell. Combining  
 615 the abundance and tRNA length measurements, we make the estimate that  $\approx 5 \times 10^7$  nucleotides  
 616 are sequestered in tRNA per cell. Unlike mRNA, tRNA is remarkably stable with typical lifetimes *in*  
 617 *vivo* on the order of  $\approx 48$  hours (??) – well beyond the timescale of division. Once again using our  
 618 rule-of-thumb for the rate of transcription to be 40 nucleotides per second and assuming a division  
 619 time of  $\approx 5000$  seconds, we arrive at an estimate of  $\approx 150$  RNA polymerases to synthesize enough  
 620 tRNA. This requirement pales in comparison to the number of polymerases needed to generate  
 621 the rRNA and mRNA pools and can be neglected as a significant transcriptional burden.

**622 RNA Polymerase and  $\sigma$ -factor Abundance**

623 These estimates, summarized in ??(A), reveal that synthesis of rRNA and mRNA are the dominant  
 624 RNA species synthesized by RNA polymerase, suggesting the need for  $\approx 700$  RNA polymerases  
 625 per cell. As is revealed in ??(B), this estimate is about an order of magnitude below the observed  
 626 number of RNA polymerase complexes per cell ( $\approx 5000 - 7000$ ). The disagreement between the  
 627 estimated number of RNA polymerases and these observations are at least consistent with recent  
 628 literature revealing that  $\approx 80\%$  of RNA polymerases in *E. coli* are not transcriptionally active (?).  
 629 Our estimate ignores the possibility that some fraction is only nonspecifically bound to DNA, as  
 630 well as the obstacles that RNA polymerase and DNA polymerase present for each other as they  
 631 move along the DNA (?).

632 In addition, it is also vital to consider the role of  $\sigma$ -factors which help RNA polymerase iden-  
 633 tify and bind to transcriptional start sites (?). Here we consider  $\sigma^{70}$  (RpoD) which is the dominant  
 634 "general-purpose"  $\sigma$ -factor in *E. coli*. While initially thought of as being solely involved in transcrip-  
 635 tional initiation, the past two decades of single-molecule work has revealed a more multipurpose  
 636 role for  $\sigma^{70}$  including facilitating transcriptional elongation (?????). ??(B) is suggestive of such a role  
 637 as the number of  $\sigma^{70}$  proteins per cell is in close agreement with our estimate of the number of  
 638 transcriptional complexes needed.

639 These estimates provide insight as to the observed magnitude of both RNA polymerase and the  
 640  $\sigma$ -70 factor. As we have done in the previous sections, and described in the supplemental informa-  
 641 tion, we can generalize these estimates across a wide range of growth rates (grey line in ??(B)).  
 642 While there remains some disagreement in the magnitude of the copy number, this estimate ap-  
 643 pears to very adequately describe the growth rate dependence of these complexes. Furthermore,  
 644 these findings illustrate that transcription cannot be the rate limiting step in bacterial division. ??(B)  
 645 (A) reveals that the availability of RNA polymerase is not a limiting factor for cell division as the cell  
 646 always has an apparent  $\sim 10$ -fold excess than needed. Furthermore, if more transcriptional activity  
 647 was needed to satisfy the cellular requirements, more  $\sigma^{70}$ -factors could be expressed to utilize a  
 648 larger fraction of the RNA polymerase pool.

**649 Translation and Ribosomal Synthesis**

650 Lastly, we turn our attention to the process of synthesizing new proteins, translation. This process  
 651 stands as a good candidate for potentially limiting growth since the synthesis of new proteins relies  
 652 on the generation of ribosomes, themselves proteinaceous molecules. As we will see in the coming  
 653 sections of this work, this poses a "chicken-or-the-egg" problem where the synthesis of ribosomes  
 654 requires ribosomes in the first place.

655 We will begin our exploration of protein translation in the same spirit as we have in previous sec-  
 656 tions – we will draw order-of-magnitude estimates based on our intuition and available literature,  
 657 and then compare these estimates to the observed data. In doing so, we will estimate both the  
 658 absolute number of ribosomes necessary for replication of the proteome as well as the synthesis  
 659 of amino-acyl tRNAs. From there we consider the limitations on ribosomal synthesis in light of our  
 660 estimates on both the synthesis of ribosomal proteins and our earlier results on rRNA synthesis.

**661 tRNA Synthetases**

662 We begin by first estimating the number of tRNA synthetases in *E. coli* needed to convert free amino-  
 663 acids to polypeptide chains. At a doubling time of  $\approx 5000$  s, *E. coli* has roughly  $3 \times 10^6$  proteins per  
 664 cell (BNID: 115702; ?). Assuming that the typical protein is on the order of  $\approx 300$  amino acids in  
 665 length (BNID: 100017; ?), we can estimate that a total of  $\approx 10^9$  amino acids are stitched together by  
 666 peptide bonds.

667 How many tRNAs are needed to facilitate this remarkable number of amino acid delivery events  
 668 to the translating ribosomes? It is important to note that tRNAs are recycled after they've passed  
 669 through the ribosome and can be recharged with a new amino acid, ready for another round of  
 670 peptide bond formation. While some *in vitro* data exists on the turnover of tRNA in *E. coli* for

<sup>671</sup> different amino acids, we can make a reasonable estimate by comparing the number of amino  
<sup>672</sup> acids to be polymerized to cell division time. Using our stopwatch of 5000 s and  $10^9$  amino acids,  
<sup>673</sup> we arrive at a requirement of  $\approx 2 \times 10^5$  tRNA molecules to be consumed by the ribosome per  
<sup>674</sup> second.

<sup>675</sup> There are many processes which go into synthesizing a tRNA and ligating it with the appropriate  
<sup>676</sup> amino acids. As we discussed previously, there appear to be more than enough RNA polymerases  
<sup>677</sup> per cell to synthesize the needed pool of tRNAs. Without considering the many ways in which  
<sup>678</sup> amino acids can be scavenged or synthesized *de novo*, we can explore ligation as a potential  
<sup>679</sup> rate limiting step. The enzymes which link the correct amino acid to the tRNA, known as tRNA  
<sup>680</sup> synthetases or tRNA ligases, are incredible in their proofreading of substrates with the incorrect  
<sup>681</sup> amino acid being ligated once out of every  $10^4$  to  $10^5$  times (BNID: 103469, ?). This is due in part  
<sup>682</sup> to the consumption of energy as well as a multi-step pathway to ligation. While the rate at which  
<sup>683</sup> tRNA is ligated is highly dependent on the identity of the amino acid, it is reasonable to state that  
<sup>684</sup> the typical tRNA synthetase has charging rate of  $\approx 20$  AA per tRNA synthetase per second (BNID:  
<sup>685</sup> 105279, ?).

<sup>686</sup> We can make an assumption that aminio-acyl tRNAs are in steady-state where they are pro-  
<sup>687</sup> duced at the same rate they are consumed, meaning that  $2 \times 10^5$  tRNAs must be charged per second.  
<sup>688</sup> Combining these estimates together, as shown schematically in ??(A), yields an estimate of  $\approx 10^4$   
<sup>689</sup> tRNA synthetases per cell with a division time of 5000 s. This point estimate is in very close agree-  
<sup>690</sup> ment with the observed number of synthetases (the sum of all 20 tRNA synthetases in *E. coli*). This  
<sup>691</sup> estimation strategy seems to adequately describe the observed growth rate dependence of the  
<sup>692</sup> tRNA synthetase copy number (shown as the grey line in ??(B)), suggesting that the copy number  
<sup>693</sup> scales with the cell volume.

<sup>694</sup> In total, the estimated and observed  $\approx 10^4$  tRNA synthetases occupy only a meager fraction of  
<sup>695</sup> the total cell proteome, around 0.5% by abundance. It is reasonable to assume that if tRNA charg-  
<sup>696</sup> ing was a rate limiting process, cells would be able to increase their growth rate by devoting more  
<sup>697</sup> cellular resources to making more tRNA synthases. As the synthesis of tRNAs and the correspond-  
<sup>698</sup> ing charging can be highly parallelized, we can argue that tRNA charging is not a rate limiting step  
<sup>699</sup> in cell division, at least for the growth conditions explored in this work.

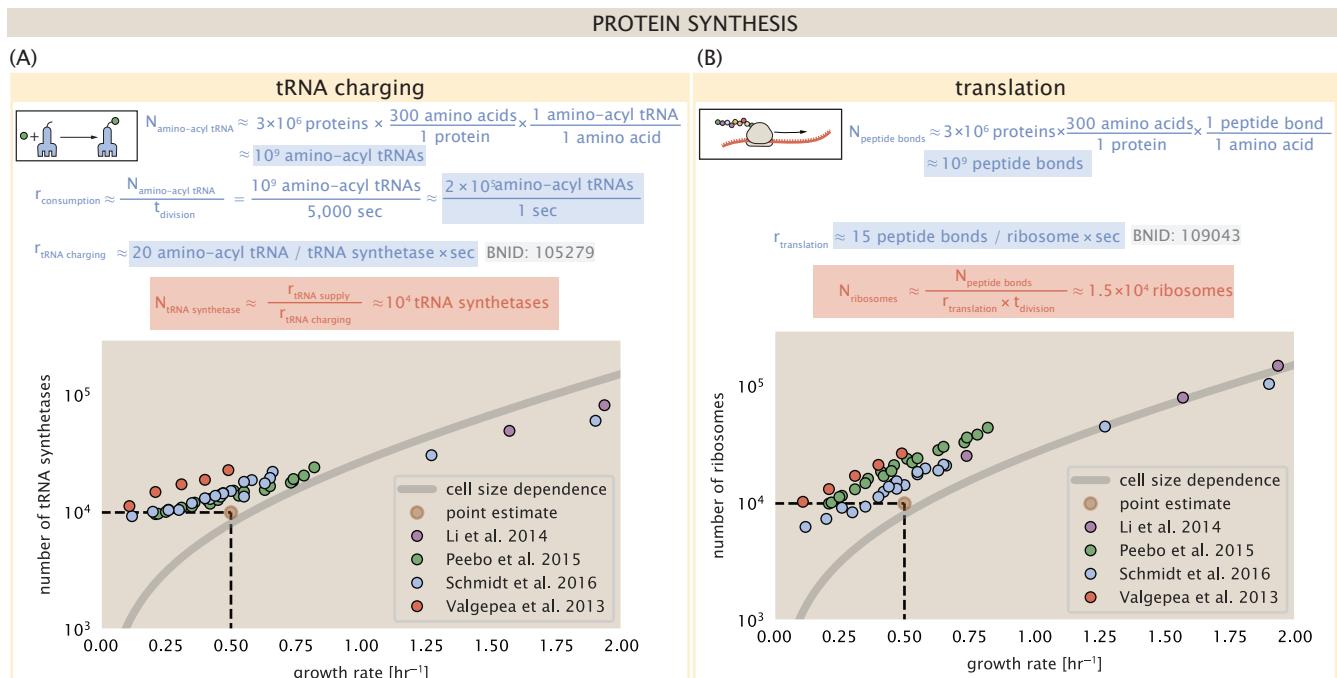
#### <sup>700</sup> Protein Synthesis

<sup>701</sup> With the number of tRNA synthetases accounted for, we now consider the abundance of the pro-  
<sup>702</sup> tein synthesis machines themselves, ribosomes. Ribosomes are enormous protein/rRNA com-  
<sup>703</sup> plexes that facilitate the peptide bond formation between amino acids in the correct sequence  
<sup>704</sup> as defined by the coding mRNA. Before we examine the synthesis of the ribosome proteins and  
<sup>705</sup> the limits that may place on the observed bacterial growth rates, let's consider replication of the  
<sup>706</sup> cellular proteome.

<sup>707</sup> As described in the previous section, an *E. coli* cell consisting of  $\approx 3 \times 10^6$  proteins will have  
<sup>708</sup> on the order  $\approx 10^9$  peptide bonds per proteome. While the rate at which ribosomes translates is  
<sup>709</sup> well known to have a growth rate dependence ? and is a topic which we discuss in detail in the  
<sup>710</sup> coming sections. However, for the purposes of our order-of-magnitude estimate, we can make  
<sup>711</sup> the approximation that translation occurs at a rate of  $\approx 15$  amino acids per second per ribosome  
<sup>712</sup> (BNID: 100233, ?). Under this approximation and assuming a division time of 5000 s, we can arrive  
<sup>713</sup> at an estimate of  $\approx 10^4$  ribosomes are needed to replicate the cellular proteome, shown in ??(B).  
<sup>714</sup> This point estimate, while glossing over important details such as chromosome copy number and  
<sup>715</sup> growth-rate dependent translation rates, proves to be notably accurate when compared to the  
<sup>716</sup> experimental observations (??(B)).

#### <sup>717</sup> Translation and Ribosomal Synthesis as a Growth-Rate Limiting Step

<sup>718</sup> Thus far, the general back-of-the-envelope estimates have been reasonably successful in explain-  
<sup>719</sup> ing what sets the scale of absolute protein copy number as well as their observed dependence



**Figure 9. Estimation of the required tRNA synthetases and ribosomes.** (A) 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], [Glx], [IleS], [LeuS], [ValS], [AlaS]<sub>2</sub>, [AsnS]<sub>2</sub>, [AspS]<sub>2</sub>, [TyrS]<sub>2</sub>, [TrpS]<sub>2</sub>, [ThrS]<sub>2</sub>, [SerS]<sub>2</sub>, [ProS]<sub>2</sub>, [PheS]<sub>2</sub>[PheT]<sub>2</sub>, [MetG]<sub>2</sub>, [LysS]<sub>2</sub>, [HisS]<sub>2</sub>, [GlyS]<sub>2</sub>[GlyQ]<sub>2</sub>). (B) Estimation of the number of ribosomes required to synthesize 10<sup>9</sup> 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<sup>-1</sup>. Grey lines correspond to the estimated complex abundance calculated at different growth rates. See Supplemental Information XX for a more detail description of this calculation.

on the cellular growth rate. A recurring theme that has arisen is the ability of cells to parallelize their biosynthesis tasks. For example, while DNA replication speed-limit is  $\approx 40$  minutes to replicate a genome, cells can divide faster than this by initiating more than one round of replication per doubling. The process of protein synthesis overall doesn't appear to be rate-limiting, since for example, cells are able to induce the expression of additional enzymes to grow on alternative carbon sources. However, as we will see, the synthesis of ribosomal proteins presents a special case where parallelization is not possible (??(A)). Thus, it is plausible that translation may be a key factor in determining the cellular growth rate.

To gain some intuition into how translation can set the speed of bacterial growth, we again consider the total number of peptide bonds that must be synthesized, which we denote as  $N_{\text{pep}}$ . With cells growing exponentially in time (?), we can compute the number of amino acids to be polymerized as

$$N_{\text{pep}}\lambda = r_t R f_a, \quad (1)$$

where  $\lambda$  is the cell growth rate in  $s^{-1}$ ,  $r_t$  is the maximum elongation rate in  $\text{AA}\cdot s^{-1}$ , and  $R$  is the average ribosome copy number per cell. The addition factor  $f_a$  refers to the fraction of actively translating ribosomes, and allows us to account for the possibility of nonfunctional, immature ribosomes or active sequestration of ribosomes mediated by the secondary-messenger molecule alarmones, guanosine pentaphosphate [(p)ppGpp] at slow growth (??). Knowing the number of peptide bonds to be formed permits us to compute the translation-limited growth rate as

$$\lambda_{\text{translation-limited}} = \frac{r_t R f_a}{N_{\text{pep}}}. \quad (2)$$

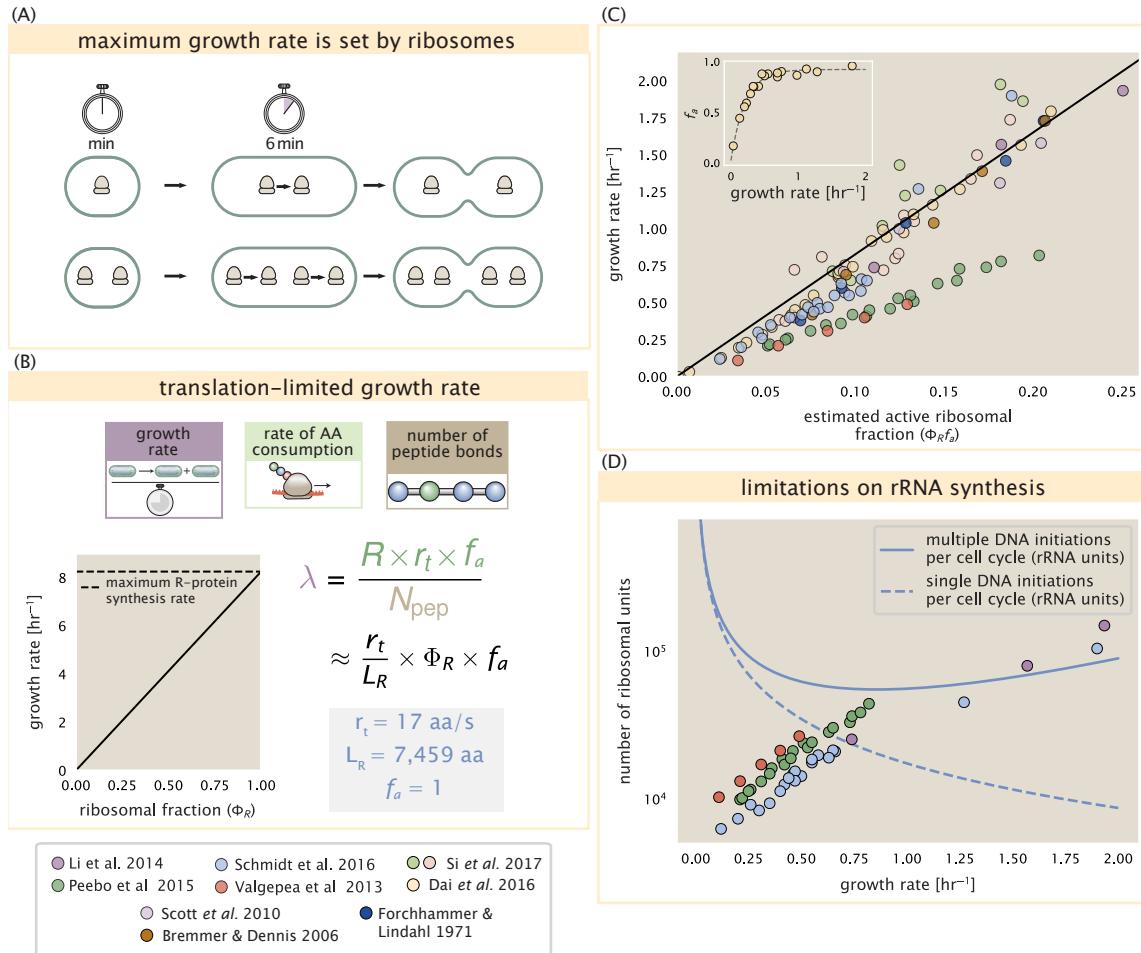
Alternatively, since  $N_{\text{pep}}$  is related to the total protein mass through the molecular weight of each protein, we can also consider the growth rate in terms of the fraction of the total proteome mass dedicated to ribosomal proteins. By making the approximation that an average amino acid has a molecular weight of 110 Da (BNID: 104877, ?), the total protein mass  $m_{\text{protein}}$  is related to  $N_{\text{AA}}$  by  $(m_{\text{protein}}/110 \text{ Da}) \times N_A$ , where  $N_A$  is Avogadro's number. Similarly,  $R$  is related to the ribosomal protein mass by  $R \approx (m_R/800 \text{ Da}) \times N_A$ , where 800 Da reflects the summed molecular weight of all ribosomal subunits. This allows us to approximate  $R/N_{\text{pep}} \approx \Phi_R/L_R$ , where  $\Phi_R$  is the ribosomal mass fraction  $m_{\text{protein}}/m_R$ , and  $L_R$  the ratio of 800 kDa / 110 Da per amino acid or, alternatively, the total length in amino acids that make up a ribosome. The translation-limited growth rate can then be written in the form

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

This is plotted as a function of ribosomal fraction  $\Phi_R$  in ??(B), where we take  $L_R = 7459$  AA, corresponding to the length in amino acids for all ribosomal subunits of the 50S and 30S complex (BNID: 101175, (?)), and  $f_a = 1$ .

The growth rate defined by ?? reflects mass-balance under steady-state growth and has long provided a rationalization of the apparent linear increase in *E. coli*'s ribosomal content as a function of growth rate (??). Here we see that there will be a maximum rate when  $\Phi_R = 1$ , which for a translation rate of 17 amino acids per second, gives us  $\lambda \approx 8 \text{ hr}^{-1}$ , or a doubling time just under 6 minutes (??(B), dashed line). Interestingly, this limit is independent of the absolute number of ribosomes and is simply given by the time to translate an entire ribosome,  $L_R/r_t$ . As shown in ??(A), we can reconcile this with the observation that in order to double the average number of ribosomes, each ribosome must produce a second ribosome and this process cannot be parallelized. Unless protein synthesis can increase, or cells can trim their total ribosomal protein mass, this must represent an absolute speed limit for cell doubling.

In recent work from ?, the authors made independent measurements of  $r_t$ ,  $\Phi_R$  (via RNA-to-protein ratios, and directly by mass spectrometry), and growth rate, enabling inference of the active fraction  $f_a$  across the entire range of growth rates considered here. In ??(?) we use this measurement of  $f_a$  to estimate the active fraction of ribosomal protein across the proteomic data sets.



**Figure 10. Translation-limited growth rate.** (A) An inherent maximum growth rate is set by the time to synthesize all ribosomal subunits. This growth rate is given by  $r_t/L_R$ , where  $r_t$  is the elongation rate and  $L_R$  is the total number of amino acids that make up the entire ribosomal complex. This rate is independent of the number of ribosomes; instead limited serial requirement for each ribosome to double itself. (B) Translation-limited growth as a function of the ribosomal fraction. The solid line is calculated for an elongation rate of 17 aa per second. The dashed line corresponds to the maximum rate of ribosomal protein (R-protein) synthesis considered in part (A). (C) Actively translating ribosomal fraction versus growth rate. The actively translating ribosomal fraction is calculated using the estimated values of  $f_a$  from ? (shown in inset; see Supplemental Section XX for additional detail). (D) Maximum number of rRNA units that can be synthesized as a function of growth rate. Solid curve corresponds to the rRNA copy number is calculated by multiplying the number of rRNA operons by the estimated number of  $\langle \# \text{ ori} \rangle$  at each growth rate. The quantity  $\langle \# \text{ ori} \rangle$  was calculated using Equation 4 and the measurements from ? that are plotted in ??(A). The dashed line shows the maximal number of functional rRNA units produced from a single chromosomal initiation per cell cycle.

765 Importantly, we find that from the perspective of actively translating ribosomes, cells are effectively  
 766 skirting the limit in growth rate that is set by ??, as nutrient conditions vary.

767 Earlier, we considered rRNA synthesis (see Section **RNA Synthesis**), finding that, when the rRNA  
 768 operons are maximally loaded with RNA polymerase, the cell can produce  $\approx 1$  functional rRNA unit  
 769 per second per operon. In ??(C), we show the maximum number of ribosomes that could be made  
 770 as a function of growth rate given this rRNA production rule-of-thumb. While each *E. coli* genome  
 771 has 7 copies of the rRNA operon (BNID: 107866, ?), parallelization of DNA synthesis by firing multi-  
 772 ple rounds of replication at a time can drastically increase the effective number of rRNA operons. The blue  
 773 curve in ??(C), we assume that the effective number of rRNA operons increases in proportion to  
 774 the number of origins of replication  $\langle \# \text{ ori} \rangle$  (solid blue line; with the calculation of  $\langle \# \text{ ori} \rangle$  described  
 775 in the next section). Although we expect this value to drastically overestimate rRNA abundance at  
 776 slower growth rates ( $\lambda < 0.5 \text{ hr}^{-1}$ ), it provides a useful reference when considered along with the  
 777 proteomic measurements that are also plotted. For growth rates above about  $1 \text{ hr}^{-1}$ , we find that  
 778 cells will need to transcribe rRNA near their maximal rate. The dashed blue curve in ??(C) shows  
 779 the maximal number of functional rRNA units that could be synthesized from a single genome (igno-  
 780 ring the chromosome replication speed limit of  $\approx 40$  minutes per genome). The convergence  
 781 between the maximum rRNA production with parallelization and the experimentally measured ri-  
 782 bosome copy number (points in ??(C)), suggests rRNA synthesis may begin to present a bottleneck  
 783 in cell division at the fastest growth rates. While this strain of *E. coli* is rarely reported to grow  
 784 faster than  $2 \text{ hr}^{-1}$ , other bacteria with more copies of rRNA genes have been found that surpass  
 785 this growth rate (??).

### 786 **Relationship Between Cell Size and Growth Rate**

787 The relationship between cell size and growth rate has long been of interest in the study of bacterial  
 788 physiology, particularly following the now six decade-old observation that cell volume appears to  
 789 increase exponentially with growth rate; known as Schaechter's growth law (??). However, the  
 790 mechanism that governs this relationship, and even the question of whether the change in average  
 791 cell size is truly exponential, has remained under debate (?). Given the importance of cell size in  
 792 determining the total protein mass that must be doubled (as well as in setting other parameters  
 793 like the surface-area-to-volume ratio), we examine the influence size may have in setting the scales  
 794 of protein abundance and growth dependence observed in the proteomic datasets.

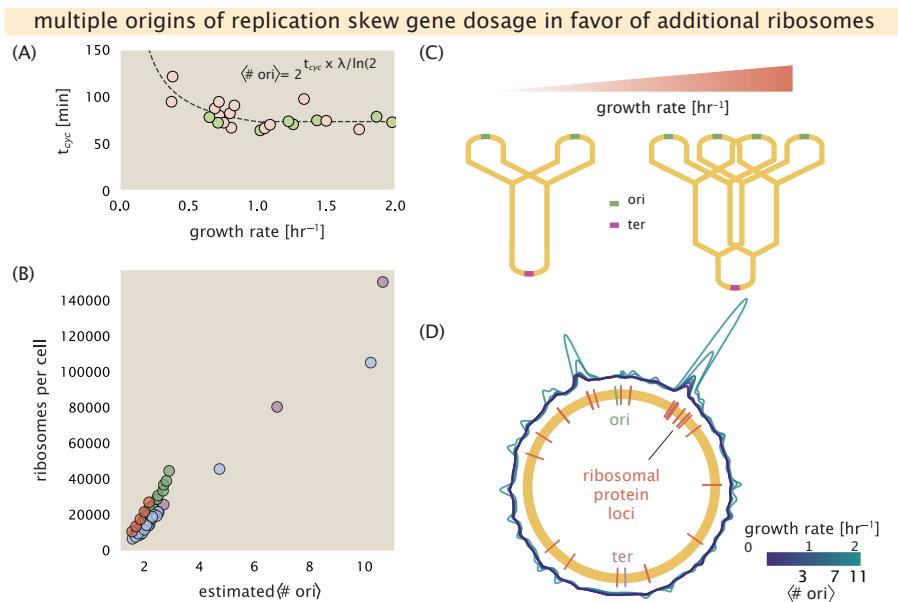
795 As shown in ??(B), cells grow at a near-maximal rate dictated by their total ribosomal mass  
 796 fraction  $\Phi_R$ , at least at moderate growth rates above  $0.5 \text{ hr}^{-1}$ , suggesting that growth rate could  
 797 increase simply by making more ribosomes and increasing  $\Phi_R$ . In reality, however, large swaths of  
 798 the proteome increase in absolute protein abundance as cells grow faster (Supplemental Figure X),  
 799 and the ability to add additional ribosomes is likely constrained by other factors including crowding  
 800 due to their large size (??). Rather, it is well-documented that *E. coli* cells add a constant volume  
 801 per origin of replication (termed a "unit cell" or "initiation mass"), which is robust to a remarkable  
 802 array of cellular perturbations (?). To consider this in the context of the proteomic data, we used  
 803 the measurements from ? for wild-type *E. coli* cells grown in different nutrient conditions (??(A))  
 804 to estimate the average number of origins per cell  $\langle \# \text{ ori} \rangle$  across the data. Indeed, we find an  
 805 approximately linear correlation between ribosome copy number and  $\langle \# \text{ ori} \rangle$  (??(B)).

806 The average number of origins  $\langle \# \text{ ori} \rangle$  is set by how often replication must be initiated per cell  
 807 doubling under steady-state growth. This can be quantified as

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

808 where  $\tau_{\text{cyc}}$  is the cell cycle time (referring to the time from replication initiation to cell division), and  
 809  $\tau$  is the cell doubling time. For a constant cell cycle time, observed at growth rates above about  $0.5$   
 810  $\text{hr}^{-1}$  (?), ?? states that  $\langle \# \text{ ori} \rangle$  will increase exponentially with the growth rate.

811 Why does *E. coli* add a constant volume per  $\langle \# \text{ ori} \rangle$ ? To gain insight on this phenomenological  
 812 discovery and how it pertains to growth, we must consider how the protome size and composi-



**Figure 11. Multiple replication initiations bias protein synthesis in favor of more ribosomes.** (A)

Experimental data from Si *et al.* (2017). Dashed line shows fit to the data, which were used to estimate  $\langle \# \text{ori} \rangle$ .  $t_{\text{cyc}}$  was assumed to vary in proportion to  $\tau$  for doubling times great than 40 minutes, and then reach a minimum value of [fill in] minutes below this (see Supplemental Appendix X for additional details). Red data points correspond to measurements in strain MG1655, while light green points are for strain NCM3722. (B) Plot of the ribosome copy number estimated from the proteomic data against the estimated  $\langle \# \text{ori} \rangle$ . (C) Schematic shows the expected increase in replication forks (or number of ori regions) as *E. coli* cells grow faster. (D) A running Gaussian average (20 kbp st. dev.) of protein copy number is calculated for each growth condition considered by (?). Since total protein abundance increases with growth rate, protein copy numbers are median-subtracted to allow comparison between growth conditions.  $\langle \# \text{ori} \rangle$  are estimated using the data in (A) and Equation ???. [still looking into how best to use this type of analysis]

tion changes with respect to growth rate. In ??(D), we consider the position-dependent protein expression across the chromosome for each of the growth conditions from ?. Here, we calculated a running Gaussian average of protein copy number (20 kbp st. dev. averaging window) based on each gene's transcriptional start site, which were then median-subtracted to account for the differences in total protein abundance with each growth condition. Importantly, we find that the major deviations in protein copy number are largely restricted to regions of ribosomal protein genes, with substantially higher deviations observed for cells with high  $\langle \# \text{ori} \rangle$  (teal), as compared to those with low  $\langle \# \text{ori} \rangle$  (purple). This is particularly apparent for genes closer to the origin, where the majority of ribosomal proteins are found. This suggests that in addition to the linear scaling between protein abundance and  $\langle \# \text{ori} \rangle$ , the relative ribosomal abundance is tuned in proportion to  $\langle \# \text{ori} \rangle$ . Since growth rate depends specifically on the ribosomal fraction  $\Phi_R$ , this result suggests that cells are changing their size as a way to tune  $\Phi_R$  to match the available nutrient conditions.

### 825 Alarmone-Mediated Regulation Controls the Rate of Protein Synthesis

As we have seen, cell size, total proteomic content, and the number of ribosomes are all interconnected and influence the achievable growth rate. The drastic change in these parameters across different growth conditions also suggests that cells are tuning them to better match their biosynthetic capacity to the specific environment. Take, as another illustration of this, the recent experimental work by ?. In one set of experiments the authors considered growth in cells whose primary glucose transport system was disrupted ( $\Delta ptsG$ ). Unsurprisingly, the growth rate was reduced, and was measured at about two-fold slower than their wild-type line. This change, however, was not simply the result of now-limiting carbon uptake. Instead, cells accommodated this perturbation by

834 also reducing their ribosomal mass fraction by a factor of two, which is still in line with ?? under  
 835 translation-limited growth. In this final, we explore the interconnection between cell size, ribo-  
 836 some content, and growth rate by formulating a minimal model of growth rate control. We use it  
 837 to quantitatively show how tuning these parameters help cells maximize their growth rate.

838 To react to changes in nutrient conditions, bacteria rely on the synthesis or degradation of  
 839 secondary-messenger molecules like (p)ppGpp, which cause global changes in transcriptional and  
 840 translational activity. In *E. coli*, amino acid starvation causes the accumulation of de-acylated tRNAs  
 841 at the ribosome's A-site and leads to a strong increase in (p)ppGpp synthesis activity by the enzyme  
 842 RelA (?). Cells also accumulate (p)ppGpp during steady-state growth in poorer growth conditions,  
 843 which leads to a decrease in the fraction of actively translating ribosomes,  $f_a$  (with  $f_a \approx 0.5$  at a  
 844 growth rate of  $\approx 0.3 \text{ hr}^{-1}$ ).

845 Furthermore, (p)ppGpp can inhibit the initiation of DNA replication by mediating a change in  
 846 transcriptional activity and the supercoiling state of the origin of replication (?). These observations  
 847 all raise the possibility that it is through (p)ppGpp that cells mediate the growth-rate dependent  
 848 changes in  $\langle \# \text{ ori} \rangle$ , cell size, and ribosomal abundance and activity (??). Indeed, recent work in a  
 849 (p)ppGpp deficient strain of *E. coli* found that cells exhibited a high ratio of  $\langle \# \text{ ori} \rangle$  to  $\langle \# \text{ ter} \rangle$ , and  
 850 cell sizes that were more consistent with a fast growth state where (p)ppGpp levels are normally  
 851 low (?).

### 852 Ribosomal Elongation Rate and Cellular Growth Rate are Linked by Amino Acid 853 Scarcity

854 To better understand how cells maximize their growth rate across growth conditions, we consider  
 855 a mode of regulation in which the rate of peptide elongation  $r_t$  depends only on the availability of  
 856 amino acids (and, therefore, also amino-acyl tRNAs). It is through the elongation rate  $r_t$  that we  
 857 assume cells adjust their ribosomal content ( $R, \Phi_R$ ) according to nutrient availability. As the rate of  
 858 amino acid supply, denote by  $r_{AA}$ , decreases, the cell can tune the rate of amino acid consumption  
 859 (mathematized as  $r_t \times R \times f_a$ ) to remain in steady-state growth, shown schematically in ??(A). Under  
 860 this model, other molecular players required for translation like elongation factors and GTP are  
 861 considered in sufficient abundance, which appear to be valid assumptions given our analysis of  
 862 the proteomic data and energy production thus far.

863 For simplicity, we consider all amino acids as a single species with an effective cellular con-  
 864 centration  $[AA]_{\text{eff}}$ . The rate of elongation  $r_t$  will depend on how quickly the ribosomes can match  
 865 codons with their correct amino-acyl tRNA, along with the subsequent steps of peptide bond for-  
 866 mation and translocation. We therefore coarse-grain the steps of elongation to two time-scales,  
 867 1) the time required to find and bind each correct amino-acyl tRNA, and 2) the remaining steps in  
 868 peptide elongation that will not depend on the amino acid availability. The time to translate each  
 869 codon is given by the inverse of the elongation rate  $r_t$ , which can be written as,

$$\frac{1}{r_t} = \frac{1}{k_{on}\alpha[AA]_{\text{eff}}} + \frac{1}{r_t^{\max}}. \quad (5)$$

870 where we have assumed that the rate of binding by amino-acyl tRNA  $k_{on}$  is proportional to  $[AA]_{\text{eff}}$   
 871 by a constant  $\alpha$ . The second term on the right-hand side reflects our assumption that other steps  
 872 in peptide elongation are not rate-limiting, with a maximum elongation rate  $r_t^{\max}$  of about 17 amino  
 873 acids per second ?. This can be stated more succinctly in terms of an effective dissociation constant,

$$K_D = \frac{r_t^{\max}}{\alpha k_{on}}, \quad (6)$$

874 where the elongation rate  $r_t$  is now given by

$$r_t = \frac{r_t^{\max}}{1 + K_D/[AA]_{\text{eff}}}. \quad (7)$$

875 Under steady-state growth, the amino acid concentration is constant ( $\frac{d[AA]_{\text{eff}}}{dt} = 0$ ), meaning that  
 876 synthesis and consumption are matched. The effective amino acid concentration  $[AA]_{\text{eff}}$  will relate  
 877 to the rate of amino acid synthesis (or import, for rich media) and/or tRNA charging, as  $r_{AA}$ , and  
 878 the rate of consumption,  $r_t \times R \times f_a$  by,

$$\int_0^t \frac{d[AA]_{\text{eff}}}{dt} dt = \int_0^t ([r_{AA}] - [r_t \times R \times f_a]) dt, \quad (8)$$

879 where the time from 0 to  $t$  is an arbitrary length of time, and the square brackets indicate concen-  
 880 trations per unit time. Integrating ?? yields.

$$[AA]_{\text{eff}} = t([r_{AA}] - [r_t \times R \times f_a]). \quad (9)$$

881 Alternatively, we can state this in terms of absolute ribosome copy number  $R$  by considering a  
 882 unit volume  $V$ ,

$$[AA]_{\text{eff}} = \frac{t(r_{AA} - r_t \times R \times f_a)}{V}, \quad (10)$$

883 where  $r_{AA}$  is in units of AA per unit time and  $r_t$  is in units of AA per unit time per ribosome. With  
 884 an expression for  $[AA]_{\text{eff}}$  in hand, we can now solve ?? for  $r_t$ , which is a quadratic function with a  
 885 physically-meaningful positive root of

$$r_t = \frac{-t(r_{AA} + r_t^{(\max)} R f_a) - K_D V \pm \sqrt{(r_{AA} t + r_t^{(\max)} R f_a t + K_D V)^2 - 4(R f_a t)(r_t^{(\max)} r_{AA} t)}}{-2R f_a t}. \quad (11)$$

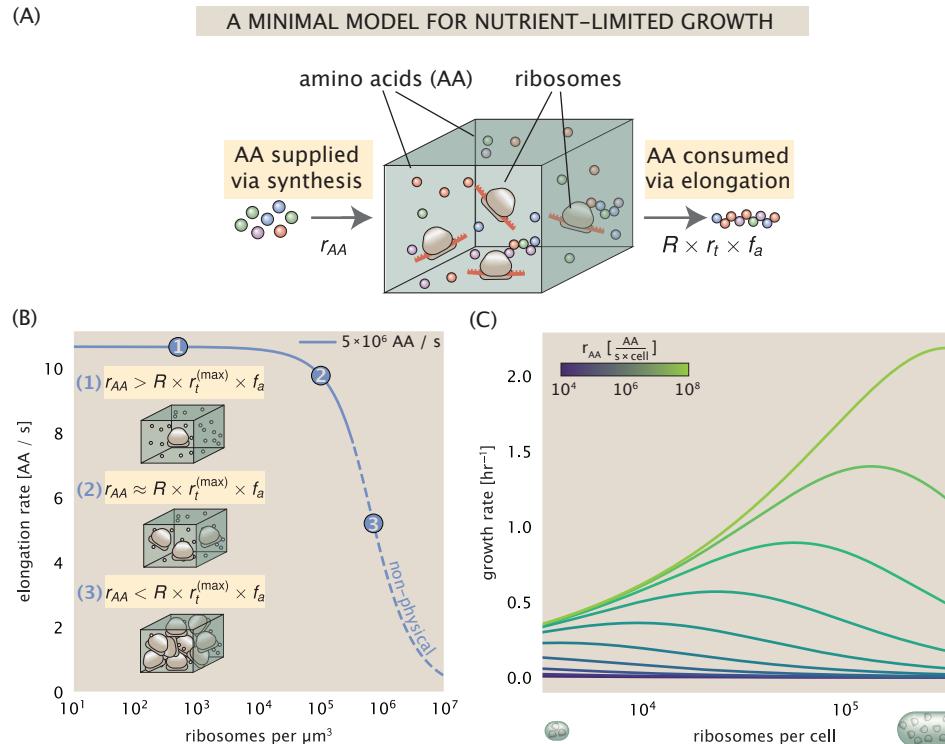
886 In ??(B), we illustrate how the elongation rate depends on the ribosomal copy number. Here,  
 887 we have considered a unit volume  $V = 1\mu\text{m}^3$ , a unit time  $t = 1\text{ s}$ , a  $K_D = 5\text{ mM}$  (inferred from ?),  
 888  $f_a = 1$ , and an arbitrarily chosen  $r_{AA} = 5 \times 10^6 \text{ AA} \cdot \text{s}^{-1} \cdot \mu\text{m}^{-3}$ . At low ribosome copy numbers, the  
 889 observed elongation rate is dependent primarily on the ratio of  $K_D/V r_{AA}$  [as  $r_t^{(\max)} \times R \times f_a \ll r_{AA}$ ,  
 890 point (1) in ??(B)]. As the ribosome copy number is increased such that the amino acid supply rate  
 891 and consumption rate are nearly equal [point (2) in ??(B)], the observed elongation rate begins to  
 892 decrease sharply. When the ribosome copy number is increased even further, consumption at the  
 893 maximum elongation rate exceeds the supply rate, yielding a significantly reduced elongation rate  
 894 [point (3) in ??(B)]. While the elongation rate will always be dominated by the amino acid supply rate  
 895 at sufficiently low ribosome copy numbers, the elongation rate at larger ribosome abundances can  
 896 be increased by tuning  $f_a$  such that not all ribosomes are elongating, reducing the total consump-  
 897 tion rate.

898 It is important to note that thus far, this model quantifies only the relationship between amino  
 899 acid supply and consumption as a function of the ribosome copy number and states nothing about  
 900 the cellular growth rate. With a sense of how elongation rate is tied to amino acid availability, we  
 901 now turn to how this relates to the cellular growth rate.

902 Optimal Growth Rate, Ribosomal Content, and Cell Size Depend on Nutrient Availability  
 903 and Metabolic Capacity.

904 To relate the elongation rate to growth rate, we constrain the set of parameters based on measured  
 905 proteomic changes; namely, we will restrict the values of  $R$ ,  $N_{\text{pep}}$ , and  $V$  to those associated with  
 906 the amalgamated proteomic data. We will then consider how changes in the nutrient conditions,  
 907 through the parameter  $r_{AA}$ , influence the maximum growth rate.

908 Earlier, we considered ribosome biosynthesis as the growth-rate determining cellular process  
 909 in ?? by stating that the cellular growth rate  $\lambda$  was related to the ribosome abundance, elongation  
 910 rate, active ribosome fraction, and the total number of peptide bonds to be formed,  $N_{\text{pep}}$ . We  
 911 return to this limit in light of our expression for a condition-dependent elongation rate  $r_t$  given by  
 912 ?. ?. ?. ??(C) shows how the observed growth rate depends on the rate of amino acid supply  $r_{AA}$  as a  
 913 function of the cellular ribosome copy number. A feature immediately apparent is the presence  
 914 of a maximal growth rate whose dependence on  $R$  (and consequently, the cell volume) increases



**Figure 12. A minimal model for regulation of growth rate 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 ribosome copy numbers in a unit volume  $\mu\text{m}^3$ . 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  $r_{AA} = 5 \times 10^6 \text{ AA / s}$ ,  $K_D = 5 \text{ mM}$ , and  $r_t^{(\max)} = 17.1 \text{ AA / s}$ . (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 volume and total protein abundance. We direct the reader to the Supplemental Information for discussion on the inference of the relationship between cell volume, number of peptide bonds, and ribosome copy number.

915 with increasing  $r_{AA}$ . Importantly, however, there is an optimum set of  $R$ ,  $N_{pep}$ , and  $V$  that are strictly  
916 dependent on the value of  $r_{AA}$ . Increasing the ribosomal concentration beyond the cell's metabolic  
917 capacity has the adverse consequence of depleting the supply of amino acids and a concomitant  
918 decrease in the elongation rate  $r_e$  [??(B)].

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