

# Fundamental limits on the rate of bacterial cell division

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

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

<sup>14</sup>

<sup>15</sup> **Abstract** Recent years have seen a deluge of experiments dissecting the relationship between bacterial  
<sup>16</sup> growth rate, cell size, and protein content, quantifying the abundance of proteins across growth conditions with  
<sup>17</sup> unprecedented resolution. However, we still lack a rigorous understanding of what sets the scale of these  
<sup>18</sup> quantities and when protein abundances should (or should not) depend on growth rate. Here, we seek to  
<sup>19</sup> quantitatively understand this relationship across a collection of *Escherichia coli* proteomic data sets covering  $\approx$   
<sup>20</sup> 4000 proteins and 31 growth conditions. We estimate the basic requirements for steady-state growth by  
<sup>21</sup> considering key processes in nutrient transport, energy generation, cell envelope biogenesis, and the central  
<sup>22</sup> dogma, from which ribosome biogenesis emerges as a primary determinant of growth rate. We conclude by  
<sup>23</sup> exploring a model of ribosomal regulation as a function of the nutrient supply, revealing a mechanism that ties  
<sup>24</sup> cell size and growth rate to ribosomal content.

<sup>25</sup>

## <sup>26</sup> Introduction

<sup>27</sup> The observed range of bacterial growth rates is enormously diverse. In natural environments, some microbial or-  
<sup>28</sup> ganisms might double only once per year (*Mikucki et al., 2009*) while in comfortable laboratory conditions, growth  
<sup>29</sup> can be rapid with several divisions per hour (*Schaechter et al., 1958*). This six order-of-magnitude difference in  
<sup>30</sup> time scales encompasses different microbial species and lifestyles, yet even for a single species such as *Escherichia*  
<sup>31</sup> *coli*, the growth rate can be modulated over a similarly large scale by tuning the type and amount of nutrients in  
<sup>32</sup> the growth medium. This remarkable flexibility in growth rate illustrates the intimate relationship between envi-  
<sup>33</sup> ronmental conditions and the rates at which cells convert nutrients into new cellular material – a relationship that  
<sup>34</sup> has remained a major topic of inquiry in bacterial physiology for over a century (*Jun et al., 2018*).

<sup>35</sup> As noted by Jacques Monod, “the study of the growth of bacterial cultures does not constitute a specialized  
<sup>36</sup> subject or branch of research, it is the basic method of Microbiology.” Those words ring as true today as they did  
<sup>37</sup> when they were written 70 years ago (*Monod, 1949*) and the study of bacterial growth has recently undergone a  
<sup>38</sup> renaissance. Many of the key questions addressed by the pioneering efforts in the middle of the last century can  
<sup>39</sup> be revisited by examining them through the lens of the increasingly refined molecular census that is available for  
<sup>40</sup> bacteria such as the microbial workhorse *E. coli*. In this work, we explore an amalgamation of recent proteomic  
<sup>41</sup> data sets to explore fundamental limits of bacterial growth.

42 Several of the evergreen questions about bacterial growth that were originally raised by microbiologists in the  
43 middle of the 20th century can now be reframed in light of this newly available data. For example, what biological  
44 processes are the primary determinants for quickly bacterial cells can grow and reproduce? How do cells alter the  
45 absolute numbers and relative ratios of their molecular constituents as a function of changes in growth rate or  
46 nutrient availability? In this paper, we address these two questions from two distinct angles. First, as a result of an  
47 array of high-quality proteome-wide measurements of *E. coli* under diverse growth conditions, we have a census  
48 that allows us to explore how the number of key molecular players change as a function of growth rate. Here,  
49 we have compiled a combination of data sets collected over the past decade using either mass spectrometry  
50 (*Schmidt et al., 2016; Peebo et al., 2015; Valgepea et al., 2013*) or ribosomal profiling (*Li et al., 2014*) of *E. coli*  
51 across 36 unique growth rates (see Appendix Experimental Details Behind Proteomic Data for further discussion  
52 of these data sets). Second, by compiling molecular turnover rate measurements for many of the fundamental  
53 processes associated with bacterial growth, we make quantitative estimates of key cellular processes (schematized  
54 in *Figure 1*) to determine whether our current understanding of the dynamics of these processes are sufficient  
55 to explain the magnitude of the observed protein copy numbers across conditions. The census, combined with  
56 these estimates, provide a window into the question of whether the rates of central processes such as energy  
57 generation or DNA synthesis are regulated systematically as a function of cell growth rate by altering protein copy  
58 number in individual cells.

59 Throughout this work, we consider an archetypal growth rate of  $\approx 0.5 \text{ hr}^{-1}$  corresponding to a doubling time  
60 of  $\approx 5000$  seconds, as the the data sets heavily sample this regime. While we formulate point estimates for the  
61 complex abundances at this division time, we consider how these values will vary at other growth rates due to  
62 changes in cell volume, surface area, and chromosome copy number (*Taheri-Araghi et al., 2015*). Broadly, we find  
63 that for the majority of these estimates, the protein copy numbers appear tuned for the task of cell doubling at a  
64 given growth rate. Thus, our understanding of the kinetics of myriad biological processes is sufficient to quantita-  
65 tively explain the observed abundances of these proteins. Furthermore, the comparison between estimate and  
66 observation reveals at a quantitative level what key phenotypic variable (such as volume, surface area, and/or  
67 chromosome content) determines the observed dependence on growth rate.

68 From these estimates, it emerges that translation, particularly of ribosomal proteins, is a plausible candidate  
69 for a molecular bottleneck for cell division. We reach this conclusion by considering that translation is 1) a rate  
70 limiting step for the *fastest* bacterial division, and 2) a major determinant of bacterial growth across the nutrient  
71 conditions we have considered under steady state, exponential growth. We mathematically consider the effects  
72 of ribosome concentration, peptide chain elongation rate, and the regulation of ribosome activity through the  
73 small molecule alarmone ppGpp. Exploration of this model is suggestive that the long-observed correlation be-  
74 tween growth rate and cell size (*Schaechter et al., 1958; Si et al., 2017*) can be simply attributed to the increased  
75 absolute number of ribosomes per cell under conditions supporting extremely rapid growth. Our conclusions  
76 from this exploration present testable hypotheses for how the observed growth rate should depend on ribosome  
77 abundance.

## 117 Nutrient Transport

118 We begin by considering the critical transport processes diagrammed in *Figure 1(A)*. In order to build new cellular  
119 mass, the molecular and elemental building blocks must be scavenged from the environment in different forms.  
120 Carbon, for example, is acquired via the transport of carbohydrates and sugar alcohols with some carbon sources  
121 receiving preferential treatment in their consumption (*Monod, 1947*). Phosphorus, sulfur, and nitrogen, on the  
122 other hand, are harvested primarily in the forms of inorganic salts, namely phosphate, sulfate, and ammonia (*Jun  
123 et al., 2018; Assentoft et al., 2016; Stasi et al., 2019; Antonenko et al., 1997; Rosenberg et al., 1977; Willsky et al.,  
124 1973*). All of these compounds have different permeabilities across the cell membrane (*Phillips, 2018*) and most  
125 require some energetic investment either via ATP hydrolysis or through the proton electrochemical gradient to  
126 bring the material across the hydrophobic cell membrane.

127 The elemental composition of *E. coli* has received much quantitative attention over the past half century (*Nei-  
128 dhardt et al., 1991; Taymaz-Nikerel et al., 2010; Heldal et al., 1985; Bauer and Ziv, 1976*), providing us with a  
129 starting point for estimating how many atoms of each element must be scavenged from the environment. A syn-  
130 thesis of these studies presents an approximate dry mass composition of  $\approx 45\%$  carbon (BioNumber ID: 100649,

79

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

80

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

81

**The rule of "one, few, and ten".** The philosophy behind order-of-magnitude estimates is to provide an estimate of the appropriate scale, not a prediction with infinite accuracy. As such, we define three different scales of precision in making estimates. The scale of "one" is reserved for values that range between 1 and 2. For example, if a particular process has been experimentally measured to transport 1.87 protons for a process to occur, we approximate this process to require 2 protons per event. The scale of "few" is reserved for values ranging between 3 and 5. For example, we will often use Avogadro's number to compute the number of molecules in a cell given a concentration and a volume. Rather than using Avogadro's number as  $6.02214 \times 10^{23}$ , we will approximate it as  $5 \times 10^{23}$ . Finally, the scale of "ten" is reserved for values which we know within an order of magnitude. If a particular protein complex is present at 883 copies per cell, we say that it is present in approximately  $10^3$  copies per cell. These different scales will be used to arrive at simple estimates that report the expected scale of the observed data. Therefore, the estimates presented here should not be viewed as hard-and-fast predictions of precise copy numbers, but as approximate lower (or upper) bounds for the number of complexes that may be needed to satisfy some cellular requirement.

82

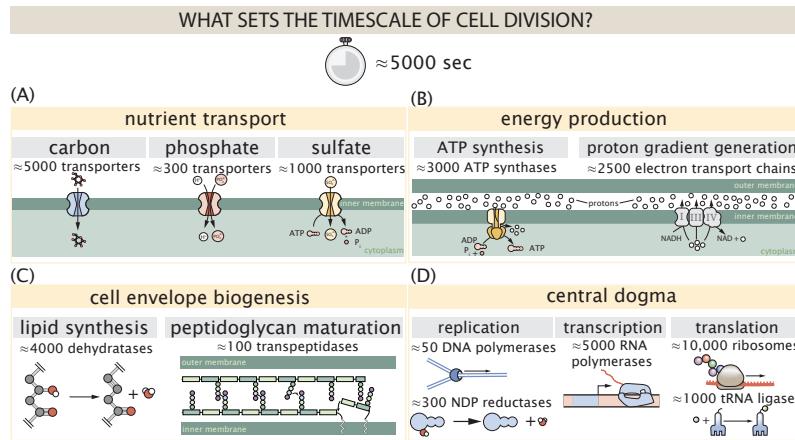
Furthermore, we use equality symbols (=) sparingly and frequently defer to approximation ( $\approx$ ) or scaling ( $\sim$ ) symbols when reporting an estimate. When  $\approx$  is used, we are implicitly stating that we are confident in this estimate within a factor of a few. When a scaling symbol  $\sim$  is used, we are stating that we are confident in our estimate to within an order of magnitude.

83

**The BioNumbers Database as a source for values.** In making our estimates, we often require approximate values for key cellular properties, such as the elemental composition of the cell, the average dry mass, or approximate rates of synthesis. We rely heavily on the BioNumbers Database ([Milo et al., 2010](#)) as a repository for such information. Every value we draw from this database has an associated BioNumbers ID number, abbreviated as BNID, and we provide this reference in grey-boxes in each figure.

84

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



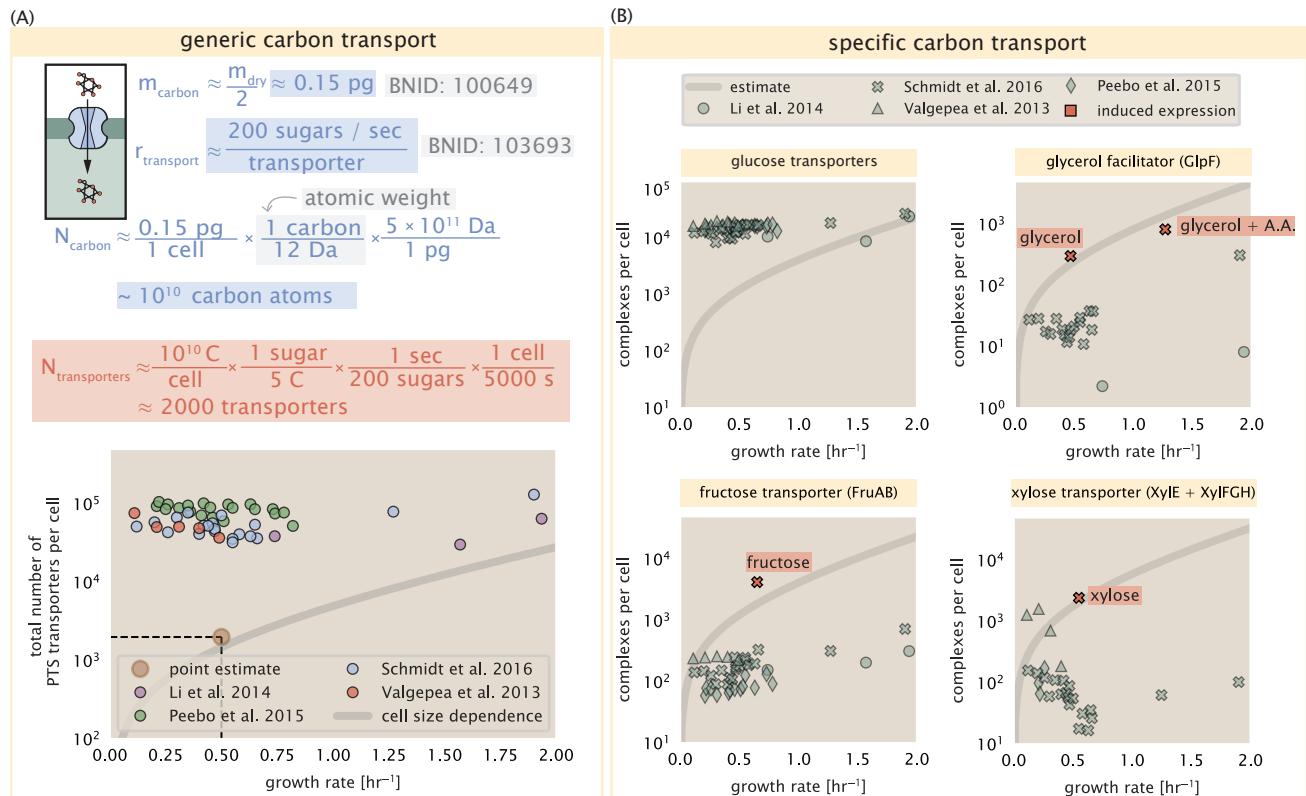
**Figure 1. Transport and synthesis processes necessary for cell division.** We consider an array of processes necessary for a cell to double its molecular components, broadly grouped into four classes. These categories are (A) nutrient transport across the cell membrane, (B) energy production (namely,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

131 see **Box 1**),  $\approx 15\%$  nitrogen (BNID: 106666),  $\approx 3\%$  phosphorus (BNID: 100653), and  $1\%$  sulfur (BNID: 100655) with  
 132 remainder being attributable to oxygen, hydrogen, and various metals. We use this stoichiometric breakdown to  
 133 estimate the abundance and growth rate dependence of a variety of transporters responsible for carbon uptake,  
 134 and provide more extensive investigation of the other critical elements – phosphorus, sulfur, and nitrogen – in the  
 135 Appendix ??.

136 Using  $\approx 0.3 \text{ pg}$  as the typical *E. coli* dry mass (BNID: 103904) at a growth rate of  $\approx 0.5 \text{ hr}^{-1}$  coupled with an  
 137 approximation that  $\approx 50\%$  of this is carbon, we estimate that  $\sim 10^{10}$  carbon atoms must be brought into the cell  
 138 in order to double all of the carbon-containing molecules (**Figure 2(A, top)**). Typical laboratory growth conditions  
 139 provide carbon as a single class of sugar (such as glucose, galactose, or xylose) for which *E. coli* has evolved myriad  
 140 mechanisms by which they can be transported across the cell membrane. One such mechanism of transport is  
 141 via the PTS system which is a highly modular system capable of transporting a diverse range of sugars (**Escalante**  
 142 **et al., 2012**). The glucose-specific component of this system transports  $\approx 200$  glucose molecules ( $\approx 1200$  carbon  
 143 atoms) per second per transporter (BNID: 114686). Making the assumption that this is a typical sugar transport  
 144 rate, coupled with the need to transport  $\sim 10^{10}$  carbon atoms, we then expect on the order of  $\approx 1000$  transporters  
 145 must be expressed in order to bring in enough carbon atoms, diagrammed in the top panel of **Figure 2(A)**.

146 As revealed in **Figure 2(A)**, experimental measurements exceed the estimate by several fold, implying that  
 147 the cell is capable of transporting more carbon atoms than needed. While we estimate  $\approx 1000$  transporters are  
 148 needed with a  $5000 \text{ s}$  division time, we can abstract this calculation to consider any particular growth rate given  
 149 knowledge of the cell density and volume as a function of growth rate and direct the reader to the Appendix  
 150 Extending Estimates to a Continuum of Growth Rates for more information. This abstraction, shown as a grey line  
 151 in **Figure 2(A)**, reveals an excess of transporters even at faster growth rates. This contrasts with our observations  
 152 for uptake of phosphorus and sulfur, which align well with our expectations across different growth conditions  
 153 (**Figure 2–Figure Supplement 1** and discussed further in Appendix ??)

154 It is important to note, however, that this estimate neglects any specifics of the regulation of the carbon trans-  
 155 port system and the data shows how many carbohydrate transporters are present on average. Using the diverse  
 156 array of growth conditions available in the data, we also explore how individual carbon transport systems depend  
 157 on specific carbon availability. In **Figure 2(B)**, we show the total number of carbohydrate transporters specific to  
 158 different carbon sources. A striking observation, shown in the top-left plot of **Figure 2(B)**, is the constancy in the  
 159 expression of the glucose-specific transport systems, an observation that stands in contrast with other species of  
 160 transporters. Additionally, we note that the total number of glucose-specific transporters is tightly distributed at  
 161  $\approx 10^4$  per cell, the approximate number of transporters needed to sustain rapid growth of several divisions per  
 162 hour. This illustrates that *E. coli* maintains a substantial number of complexes present for transporting glucose



**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· $s^{-1}$  (BNID: 103693), 2000 glycerol· $s^{-1}$  (**Lu et al., 2003**), 200 fructose· $s^{-1}$  (assumed to be similar to PtsL, BNID: 103693), and 50 xylose· $s^{-1}$  (assumed to be comparable to LacY, BNID: 103159). Red points and highlighted text indicate conditions in which the only source of carbon in the growth medium induces expression of the transport system. Grey line in (A) and (B) represents the estimated number of transporters per cell at a continuum of growth rates.

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

163 regardless of growth condition, which is known to be the preferential carbon source (*Monod, 1947; Liu et al., 2005; Aidelberg et al., 2014*).

165 Many metabolic operons are regulated with dual-input logic gates that are only expressed when glucose concentrations are low (mediated by cyclic-AMP receptor protein CRP) and the concentration of other carbon sources are elevated (*Gama-Castro et al., 2016; Zhang et al., 2014; Gama-Castro et al., 2016; Belliveau et al., 2018; Ireland et al., 2020*). Points colored in red in *Figure 2(B)* (labeled by red text-boxes) correspond to growth conditions in which the specific carbon source (glycerol, xylose, or fructose) is present. The grey lines in *Figure 2(B)* show the estimated number of transporters needed at each growth rate to satisfy the cellular carbon requirement, adjusted for the specific carbon source. These plots show that, in the absence of the particular carbon source, expression of the transporters is maintained on the order of  $\sim 10^2$  per cell. The low but non-zero abundances may reflect the specific regulatory logic involved, requiring that cells are able to transport some minimal amount of an alternative carbon source in order to induce expression of these alternative carbon-source systems.

175 If acquisition of nutrients was limiting process in cell division under the typical growth conditions explored here, 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 by nutrient transporters. 176 Considering a rule-of-thumb for the surface area of *E. coli* of about  $5 \mu\text{m}^2$  (BNID: 101792), we expect an areal density for 1000 transporters to be approximately 200 transporters/ $\mu\text{m}^2$ . For a typical transporter occupying 177 about  $50 \text{ nm}^2/\text{dimer}$ , this amounts to about only  $\approx 1\%$  of the total inner membrane area (*Szenk et al., 2017*). 178 Additionally, bacterial cell membranes typically have densities of  $10^5$  proteins/ $\mu\text{m}^2$  (*Phillips, 2018*), implying that 179 the cell could accommodate more transporters if any one species was limiting.

## 183 Cell Envelope Biogenesis

184 The carbon transported across the membrane is incorporated into nearly every biological molecule synthesized by the cell and, as we will see, many of these synthesis requirements scale with the cell volume. One unique exception 185 is in the synthesis of components of the cell envelope – namely the synthesis of lipids and peptidoglycan – which 186 define the cell surface area. Given our conclusion that expression of the carbon transporters could be increased to 187 avert rate limiting transport, we will now explore some of the key processes dedicated to building this membrane 188 real estate.

## 190 Lipid Synthesis

191 The cell envelopes of gram negative bacteria (such as *E. coli*) are composed of inner and outer phospholipid bilayer 192 membranes separated by a  $\approx 10 \text{ nm}$  periplasmic space (BNID: 100016). *E. coli* is a rod-shaped bacterium with a 193 remarkably robust length-to-width aspect ratio of  $\approx 4:1$  (*Harris and Theriot, 2018*). At modest growth rates, such 194 as our stopwatch of 5000 s, the total cell surface area is  $\approx 5 \mu\text{m}^2$  (BNID: 101792, *Milo et al. (2010)*). Assuming this 195 surface area is approximately the same between the inner and outer membranes of *E. coli*, and the fact that each 196 membrane is itself a lipid bilayer, the total membrane surface area is  $\approx 20 \mu\text{m}^2$ , a remarkable value compared to 197 the  $\approx 2 \mu\text{m}$  length of the cell.

198 While this represents the total area of the membrane, this does not mean that it is composed entirely of lipid 199 molecules. Rather, the dense packing of the membrane with proteins means that only  $\approx 40\%$  of the membrane 200 area is occupied by lipids (BNID: 100078). Using a rule-of-thumb of  $0.5 \text{ nm}^2$  as the surface area of the typical 201 lipid (BNID: 106993), we arrive at an estimate of  $\sim 2 \times 10^7$  lipids per cell, an estimate in close agreement with 202 experimental measurements (BNID: 100071, 102996).

203 The membranes of *E. coli* are composed of a variety of different lipids, each of which are unique in their structures and biosynthetic pathways (*Sohlenkamp and Geiger, 2016*). Recently, a combination of stochastic kinetic 204 modeling (*Ruppe and Fox, 2018*) and *in vitro* kinetic measurements (*Ranganathan et al., 2012; Yu et al., 2011*) 205 have revealed remarkably slow steps in the fatty acid synthesis pathways which may serve as the rate limiting 206 reactions. One such step is the removal of hydroxyl groups from the fatty-acid chain by ACP dehydratase that 207 leads to the formation of carbon-carbon double bonds. This reaction, catalyzed by proteins FabZ and FabA in *E. 208 coli* (*Yu et al., 2011*), have been estimated to have kinetic turnover rates of  $\approx 1$  dehydration per second per enzyme 209 (*Ruppe and Fox, 2018*). Thus, given this rate and the need to synthesize  $\approx 2 \times 10^7$  in a  $\approx 5000$  second time window, 210 one can estimate that a typical cell requires  $\approx 4000$  ACP dehydratases. This is in reasonable agreement with the 211

212 experimentally observed copy numbers of FabZ and FabA (*Figure 3(A)*). Furthermore, we can extend this estimate  
213 to account for the change in membrane surface area as a function of the growth rate (grey line in *Figure 3(A)*),  
214 which captures the observed growth rate dependent expression of these two enzymes.

215 Despite the slow catalytic rate of FabZ and FabA, we find it unlikely that the generation of fatty acids would be a  
216 bottleneck in cell division and is not the key process responsible for setting the bacterial growth rate. Experimental  
217 data and recent computational model has shown that the rate of fatty-acid synthesis can be drastically increased  
218 by increasing the concentration of FabZ *Yu et al. (2011); Ruppe and Fox (2018)*. With a proteome size of  $\approx 3 \times 10^6$   
219 proteins, a hypothetical 10-fold increase in expression from 4000 to 40,000 ACP dehydratases would result in a  $\approx$   
220 1% increase in the size of the proteome. As many other proteins are in much larger abundance than 4000 per cell  
221 (as we will see in the coming sections), it is unlikely that expression of ACP dehydratases couldn't be increased to  
222 speed up lipid synthesis and thereby facilitate faster growth.

### 223 Peptidoglycan Synthesis

224 While variation in cell size can vary substantially across growth conditions, bacterial cells demonstrate exquisite  
225 control over their cell shape. This is primarily due to the cell wall, a stiff, several nanometer thick meshwork of poly-  
226 merized discaccharides. The formation of the peptidoglycan is an intricate process involving many macromolecular  
227 players (*Shi et al., 2018; Morgenstein et al., 2015*) whose coordinated action maintains cell shape even in the face  
228 of large-scale perturbations, and can restore rod-shaped morphology even after digestion of the peptidoglycan  
229 (*Harris and Theriot, 2018; Shi et al., 2018*).

230 At our archetypal growth rate, the peptidoglycan alone comprises  $\approx 3\%$  of the cellular dry mass (BNID: 101936),  
231 making it the most massive molecule in *E. coli*. The polymerized unit of the peptidoglycan is a N-acetylglucosamine  
232 and N-acetylmuramic acid disaccharide, of which the former is functionalized with a short pentapeptide. With a  
233 mass of  $\approx 1000$  Da, this unit, which we refer to as a murein monomer, is polymerized to form long strands in the  
234 periplasm which are then attached to each other via their peptide linkers. Together, these quantities provide an  
235 estimate of  $\approx 5 \times 10^6$  murein monomers per cell.

236 The crosslinking of the pentapeptides between adjacent glycan strands is responsible for maintaining the struc-  
237 tural integrity of the cell wall and, in principle, each murein monomer can be involved in such a crosslink. In some  
238 microbes, such as in gram-positive bacterium *Staphylococcus aureus*, the extent of crosslinking can be large with  
239  $> 90\%$  of pentapeptides forming a connection between glycan strands. In *E. coli*, however, a much smaller propor-  
240 tion ( $\approx 20\%$ ) of the peptides are crosslinked, resulting in a weaker and more porous cell wall *Vollmer et al. (2008)*;  
241 *Rogers et al. (1980)*. The formation of these crosslinks occurs primarily during the polymerization of the murein  
242 monomers and is facilitated by a family of enzymes called transpeptidases. The four primary transpeptidases of  
243 *E. coli* have only recently been quantitatively characterized *in vivo* via liquid chromatography mass spectrometry  
244 (*Catherwood et al., 2020*), which revealed a notably slow kinetic turnover rate of  $\approx 2$  crosslinking reactions formed  
245 per second per enzyme.

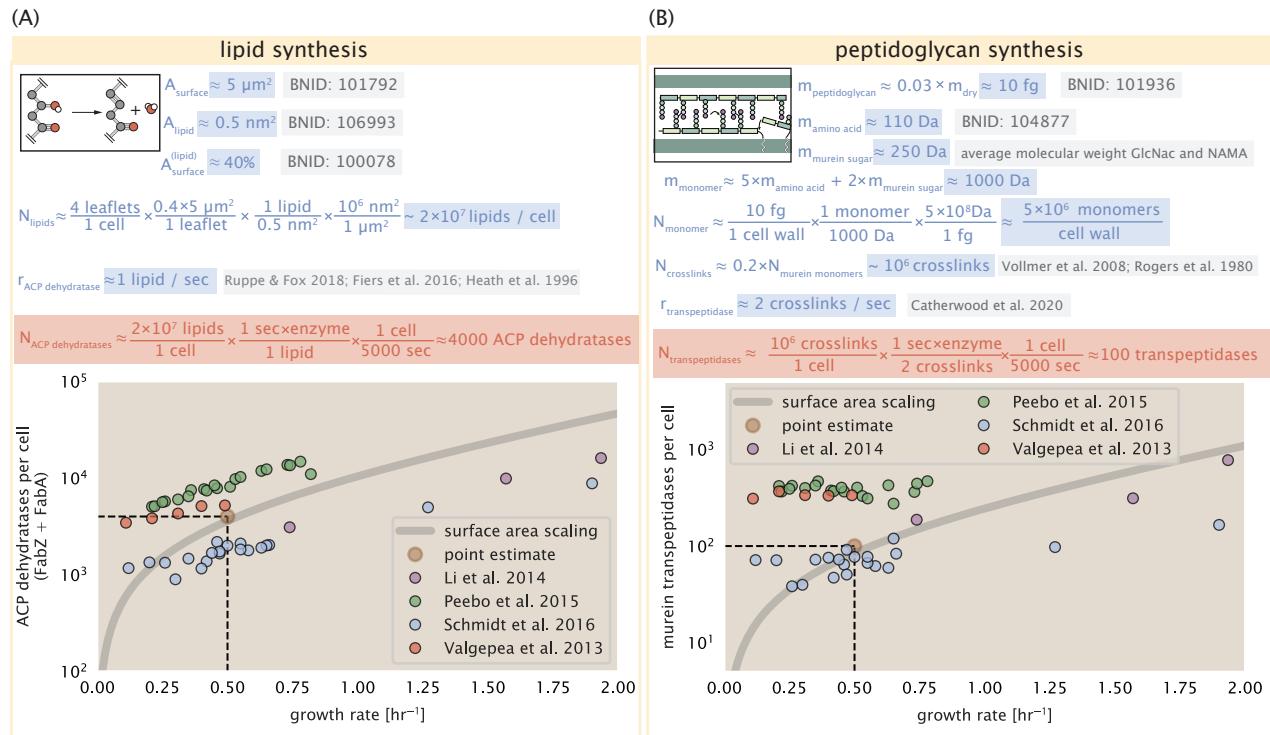
246 Assembling these quantities together permits us to make an estimate that on the order of  $\approx 100$  transpepti-  
247 dases per cell are needed for complete maturation of the peptidoglycan, given a division time of  $\approx 5000$  seconds,  
248 a value that is comparable to experimental observations (*Figure 3(B)*). Expanding this estimate to account for the  
249 changing volume of the peptidoglycan as a function of growth rate (grey line in *Figure 3(B)*) also qualitatively cap-  
250 tures the observed dependence in the data, though systematic disagreements between the different data sets  
251 makes the comparison more difficult.

252 Much as in the case of fatty acid synthesis, we find it unlikely that the formation of peptidoglycan is a process  
253 which defines the rate of bacterial cell division. The estimate we have presented considered only the transpepti-  
254 dase enzymes that are involved lateral and longitudinal elongation of the peptidoglycan. This neglects the pres-  
255 ence of other transpeptidases that are present in the periplasm and also involved in remodeling and maturation  
256 of the peptidoglycan. It is therefore possible that if this was setting the speed limit for cell division, the simple  
257 expression of more transpeptidases may be sufficient to maintain the structural integrity of the cell wall.

### 258 Energy Production

259 While the uptake of nutrients provides essential resources, some of which are utilized to build the lipid scaffold  
260 housing the transporters, the variety of metabolic pathways must then consume and generate energy in the form

## CELL ENVELOPE BIOSYNTHESIS



**Figure 3.** (A) Top panel shows an estimation for the number of ACP dehydratases necessary to form functional phospholipids, which is assumed to be a rate-limiting step on lipid synthesis. The rate of ACP dehydratases was inferred from experimental measurements via a stochastic kinetic model described in [Ruppe and Fox \(2018\)](#). Bottom panel shows the experimentally observed complex copy numbers using the stoichiometries  $[\text{FabA}]_2$  and  $[\text{FabZ}]_2$ . (B) An estimate for the number of peptidoglycan transpeptidases needed to complete maturation of the peptidoglycan. The mass of the murein monomer was estimated by approximating each amino acid in the pentapeptide chain as having a mass of 110 Da and each sugar in the disaccharide having a mass of 250 Da. The *in vivo* rate of transpeptidation in *E. coli* was taken from recent analysis by [Catherwood et al. \(2020\)](#). 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.

261 of NTPs to build new cell mass. The high-energy phosphodiester bonds of (primarily) ATP power a variety of  
262 cellular processes that drive biological systems away from thermodynamic equilibrium. The next set of processes  
263 we consider as a potential bottleneck in growth is the synthesis of ATP from ADP and inorganic phosphate as well  
264 as maintenance of the electrochemical proton gradient which powers it.

## 265 **ATP Synthesis**

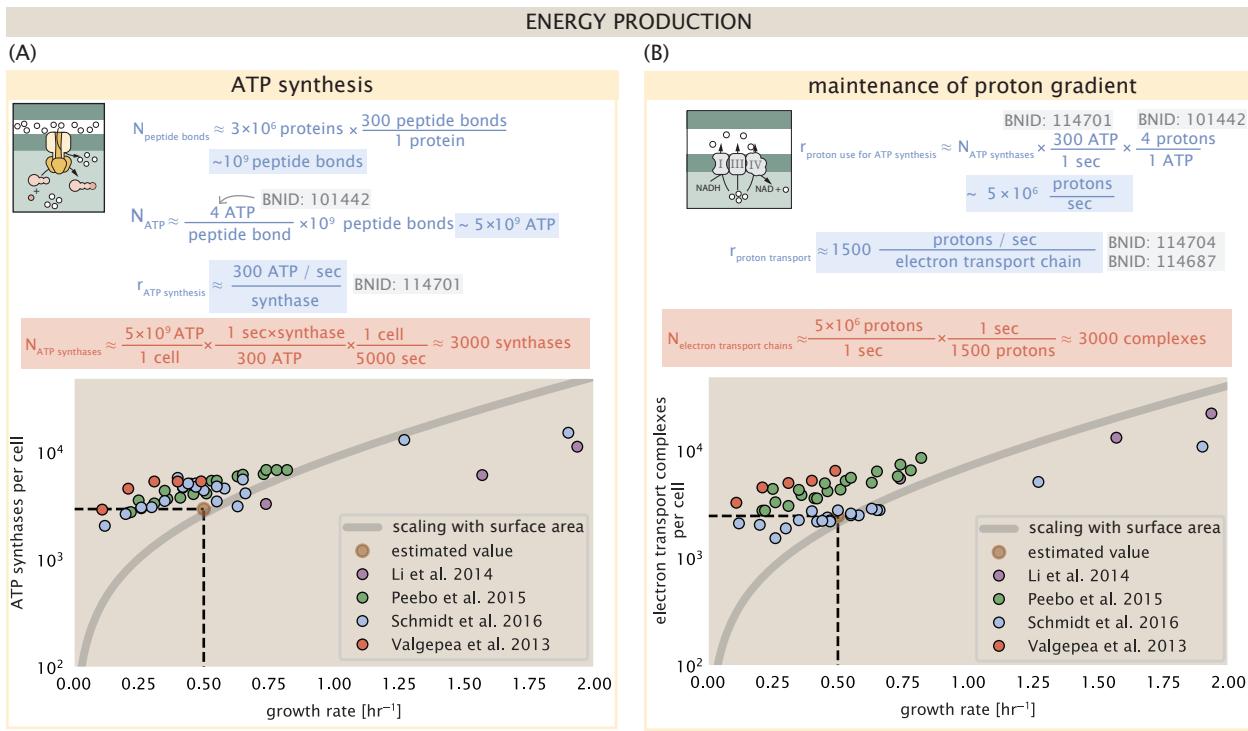
266 Hydrolysis of the terminal phosphodiester bond of ATP into ADP (or alternatively GTP and GDP) and an inorganic  
267 phosphate provides the kinetic driving force in a wide array of biochemical reactions. One such reaction is the for-  
268 mation of peptide bonds during translation, which requires  $\approx 2$  ATPs for the charging of an amino acid to the tRNA  
269 and  $\approx 2$  GTP for the formation of each peptide bond. Assuming the ATP costs associated with error correction  
270 and post-translational modifications of proteins are negligible, we can make the approximation that each peptide  
271 bond has a net cost of  $\approx 4$  ATP (BNID: 101442, *Milo et al. (2010)*). Formation of GTP from ATP is achieved via the  
272 action of nucleoside diphosphate kinase, which catalyzes this reaction without an energy investment (*Lascu and*  
273 *Gonin, 2000*) and therefore consider all NTP requirements of the cell to be functionally equivalent to being exclu-  
274 sively ATP. In total, the energetic costs of peptide bond formation consume  $\approx 80\%$  of the cells ATP budget (BNID:  
275 107782; 106158; 101637; 111918, *Lynch and Marinov (2015); Stouthamer (1973)*). The pool of ATP is produced by  
276 the  $F_1$ - $F_0$  ATP synthase – a membrane-bound rotary motor which under ideal conditions can yield  $\approx 300$  ATP per  
277 second (BNID: 114701; *Weber and Senior (2003)*).

278 To estimate the total number of ATP equivalents consumed during a cell cycle, we will make the approximation  
279 that there are  $\approx 3 \times 10^6$  proteins per cell with an average protein length of  $\approx 300$  peptide bonds (BNID: 115702;  
280 108986; 104877). Taking these values together, we find that the typical *E. coli* cell consumes  $\sim 5 \times 10^9$  ATP per cell  
281 cycle on protein synthesis alone. Assuming that each ATP synthases operates at its maximal speed (300 ATP per  
282 second per synthase),  $\approx 3000$  ATP synthases are needed to keep up with the energy demands of the cell. This  
283 estimate and a is comparable with the experimental observations, shown in *Figure 4* (A). Much as we did for the  
284 estimates of transporter copy numbers, we can generalize this estimate across different growth rates (indicated by  
285 the gray line in *Figure 4*), which appears to capture the growth rate dependence of ATP synthase abundance. We  
286 note that this estimate assumes all ATP is synthesized via ATP synthase and neglects synthesis via fermentative  
287 metabolism. This assumption may explain why at the fastest growth rates ( $\approx 2 \text{ hr}^{-1}$ ), our continuum estimate  
288 predicts more synthase than is experimentally observed. Recent work has shown that at rapid growth rates, *E.*  
289 *coli* enters a stage of overflow metabolism where fermentative metabolism becomes pronounced *Szenk et al.*  
290 (*2017*).

## 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 through its own  
293 product (and violating thermodynamics), this intricate macromolecular machine has evolved to exploit the elec-  
294 trochemical potential established across the inner membrane through cellular respiration. This electrochemical  
295 gradient is manifest by the pumping of protons into the intermembrane space via the electron transport chains  
296 as they reduce NADH. In *E. coli*, this potential difference is  $\approx -200$  mV (BNID: 102120). A simple estimate of the  
297 inner membrane as a capacitor with a working voltage of -200 mV reveals that  $\approx 2 \times 10^4$  protons must be present  
298 in the intermembrane space. However, the constant rotation of the ATP synthases would abolish this potential  
299 difference in a few milliseconds if it were not being actively maintained.

300 The electrochemistry of the electron transport complexes of *E. coli* have been the subject of intense biochemical  
301 and biophysical study (*Ingledew and Poole, 1984; Khademian and Imlay, 2017; Cox et al., 1970; Henkel et al., 2014*).  
302 A recent work (*Szenk et al., 2017*) examined the respiratory capacity of the *E. coli* electron transport complexes  
303 using structural and biochemical data, revealing that each electron transport chain rapidly pumps protons into  
304 the intermembrane space at a rate of  $\approx 1500$  protons per second (BNID: 114704; 114687). Using our estimate  
305 of the number of ATP synthases required per cell [*Figure 4(A)*], coupled with these recent measurements, we  
306 estimate that  $\approx 1000$  electron transport complexes would be necessary to facilitate the  $\sim 5 \times 10^6$  protons per  
307 second diet of the cellular ATP synthases. This estimate is in agreement with the number of complexes identified  
308 in the proteomic datasets (plot in *Figure 4(B)*). This suggests that every ATP synthase must be accompanied by  $\approx$   
309 1 functional electron transport chain. Again, to consider whether energy production may become a rate-limiting



**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]), *bo<sub>3</sub>*,([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.

310 step in growth, we need to consider whether cells could simply increase their abundances in the cell membrane,  
 311 and we turn our attention towards the available space in the membrane next.

### 312 Biosynthesis in a Crowded Membrane

313 Our estimates thus far have focused on biochemistry at the periphery of the cell and have generally been con-  
 314 cordant with the abundances expected from simple estimates. Furthermore, we have been able to describe the  
 315 growth-rate dependences considering increasing cell volume and/or surface area. However, as surface area and  
 316 cell volume do not scale identically, it is worth considering what physical limits may be imposed for transport or  
 317 energy production given the ratio of surface area to volume (S/V), which decreases with increasing growth rate.

318 In our estimate of ATP production above we found that a cell demands about  $5 \times 10^9$  ATP per cell cycle or  $10^6$   
 319 ATP/s. With a cell volume of roughly 1 fL (BNID: 100004), this corresponds to about  $2 \times 10^{10}$  ATP per fL of cell  
 320 volume, in line with previous estimates (*Stouthamer and Bettenhausen, 1977; Szenk et al., 2017*). In *Figure 5(A)* we plot this ATP demand as a function of the S/V ratio in green, where we have considered a range of cell shapes  
 321 from spherical to rod-shaped with an aspect ratio (length/width) equal to 4. In order to consider the maximum  
 322 ATP that could be produced, we consider the amount of ATP that can be generated by a membrane filled with ATP  
 323 synthase and electron transport complexes, which provides a maximal production of about 3 ATP / (nm<sup>2</sup>·s) (*Szenk*  
 324 *et al., 2017*). This is shown in blue in *Figure 5(A)*, which shows that at least for the growth rates observed (right  
 325 column in plot), the energy demand is roughly an order of magnitude less. Interestingly, *Szenk et al. (2017)* also

327 found that ATP production by respiration is less efficient than by fermentation per membrane area occupied due  
328 to the additional proteins of the electron transport chain. This suggests that, even under anaerobic growth, there  
329 will be sufficient membrane space for ATP production.

330 The analysis highlights the diminishing capacity to provide resources as the cell increases in size. However,  
331 the maximum energy production in *Figure 5(A)* does represent a somewhat unachievable limit since the inner  
332 membrane must also include other proteins including those required for lipid and membrane synthesis. To bet-  
333 ter understand the overall proteomic makeup of the inner membrane, we therefore used Gene Ontology (GO)  
334 annotations (*Ashburner et al., 2000; The Gene Ontology Consortium, 2018*) to identify all proteins embedded  
335 or peripheral to the inner membrane (GO term: 0005886). Those associated but not membrane-bound include  
336 proteins like MreB and FtsZ and must nonetheless be considered as a vital component occupying space on the  
337 membrane. In *Figure 5(B)*, we find that the total protein mass per  $\mu\text{m}^2$  is nearly constant across growth rates. In-  
338 terestingly, when we consider the distribution of proteins grouped by their Clusters of Orthologous Groups (COG)  
339 (*Tatusov et al., 2000*), the relative abundance for those in metabolism (including ATP synthesis via respiration) is  
340 also relatively constant across growth rates, suggesting that no one process (energy production, nutrient uptake,  
341 etc.) is particularly dominating even at fast growth rates *Figure 5(C)*.

## 342 Processes of the Central Dogma

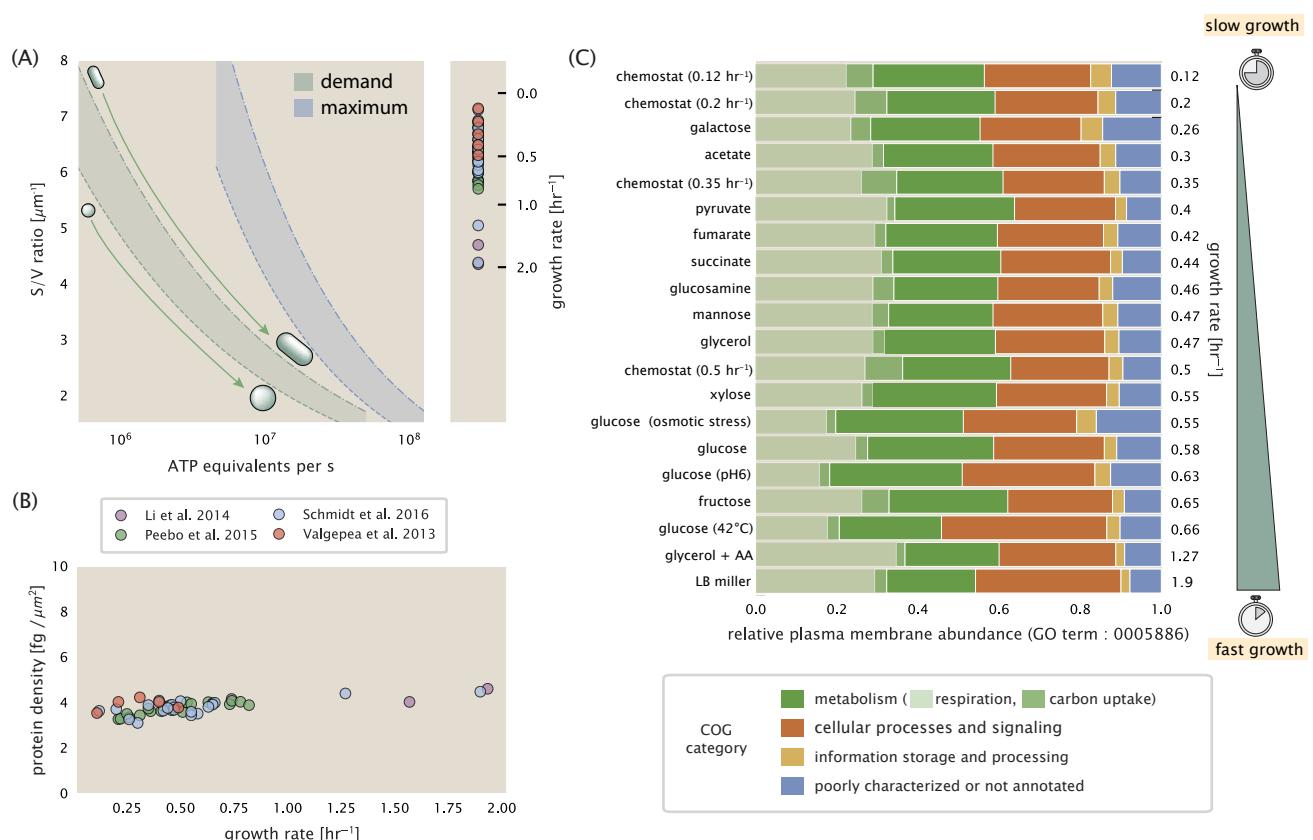
343 Up to this point, we have considered a variety of transport and biosynthetic processes that are critical to acquiring  
344 and generating new cell mass. While there are of course many other metabolic processes we could consider and  
345 perform estimates of, we now turn our focus to some of the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

### 347 DNA Replication

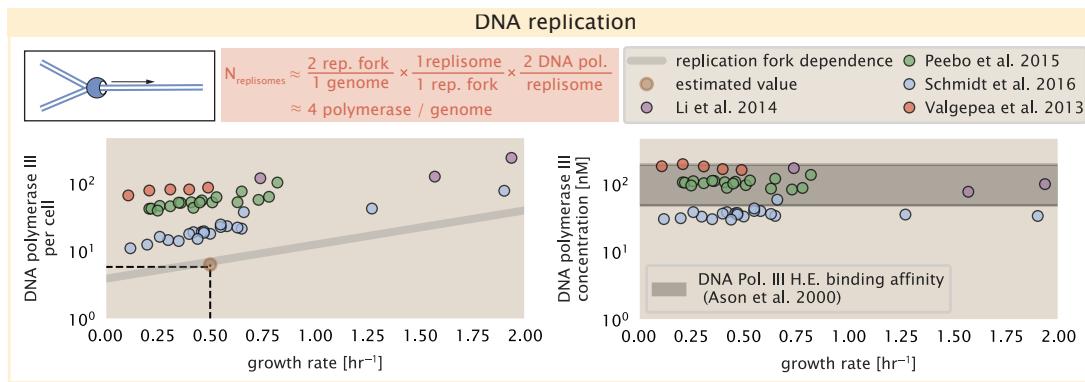
348 Most bacteria (including *E. coli*) harbor a single, circular chromosome and can have extra-chromosomal plasmids  
349 up to  $\sim 100$  kbp in length. We will focus our quantitative thinking solely on the chromosome of *E. coli* which harbors  
350  $\approx 5000$  genes and  $\approx 5 \times 10^6$  base pairs. To successfully divide and produce viable progeny, this chromosome must  
351 be faithfully replicated and segregated into each nascent cell. We again rely on the near century of literature  
352 in molecular biology to provide some insight on the rates and mechanics of this replicative feat, as well as the  
353 production of the required starting materials dNTPs (*Figure 6–Figure Supplement 1*, and discussed in Appendix  
354 ??).

355 Replication is initiated at a single region of the chromosome termed the *oriC* locus at which a pair of DNA poly-  
356 merases bind and begin their high-fidelity replication of the genome in opposite directions. Assuming equivalence  
357 between the two replication forks, this means that the two DNA polymerase complexes (termed replisomes) meet  
358 at the midway point of the circular chromosome termed the *ter* locus. The kinetics of the five types of DNA poly-  
359 merases (I – V) have been intensely studied, revealing that DNA polymerase III performs the high fidelity processive  
360 replication of the genome with the other "accessory" polymerases playing auxiliary roles (*Fijalkowska et al., 2012*).  
361 *In vitro* measurements have shown that DNA Polymerase III copies DNA at a rate of  $\approx 600$  nucleotides per second  
362 (BNID: 104120). Therefore, to replicate a single chromosome, two replisomes (containing two DNA polymerase III  
363 each) moving at their maximal rate would copy the entire genome in  $\approx 4000$  s. Thus, with a division time of 5000 s  
364 (our "typical" growth rate for the purposes of this work), there is sufficient time for a pair of replisomes complexes  
365 to replicate the entire genome. However, this estimate implies that 4000 s would be the upper-limit time scale for  
366 bacterial division which is at odds with the familiar 20 minute (1200 s) doubling time of *E. coli* in rich medium.

367 Even in rapidly growing cultures, where bacteria like *E. coli* parallelize its DNA replication with as many as 10 -  
368 12 replication forks at a given time (*Bremer and Dennis, 2008; Si et al., 2017*), we expect only a few polymerases  
369 ( $\approx 10$ ) are needed. However, as shown in *Figure 6* DNA polymerase III is nearly an order of magnitude more  
370 abundant. This discrepancy can be understood by considering its binding constant to DNA. DNA polymerase III is  
371 highly processive, facilitated by a strong affinity of the complex to the DNA. *In vitro* biochemical characterization  
372 has quantified the  $K_D$  of DNA polymerase III holoenzyme to single-stranded and double-stranded DNA to be  
373 50 and 200 nM, respectively (*Ason et al., 2000*). The right-hand plot in *Figure 6* shows that the concentration  
374 of the DNA polymerase III across all data sets and growth conditions is within this range. Thus, while the copy  
375 number of the DNA polymerase III is in excess of the strict number required to replicate the genome, its copy



**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 (Shi et al., 2018). 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 Si et al. (2017) (See Appendix Estimation of Cell Size and Surface Area 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  $bo_3$  ubiquinol oxidase, cytochrome bd-I ubiquinol oxidase) and carbohydrate transport (GO term: GO:0008643). Note that the elongation factor EF-Tu can also associate with the inner membrane, but was excluded in this analysis due to its high relative abundance (roughly identical to the summed protein shown in part (B)).



**Figure 6. Complex abundance estimates for dNTP synthesis and DNA replication.** An estimate for the minimum number of DNA polymerase holoenzyme complexes needed to facilitate replication of a single genome. Points in the left-hand plot correspond to the total number of DNA polymerase III holoenzyme complexes ( $[DnaE]_3[DnaQ]_3[HolE]_3[DnaX]_5[HolB][HolA][DnaN]_4[HolC]_4[Hold]_4$ ) per cell. Right-hand plot shows the effective concentration of DNA polymerase III holoenzyme (See Appendix Estimation of Cell Size and Surface Area for calculate of cell volumes). Grey lines in left-hand panel show the estimated number of complexes needed as a function of growth, the details of which are described in the Supplemental Information.

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

number appears to vary such that its concentration is approximately equal to the dissociation constant to the DNA. While the processes regulating the initiation of DNA replication are complex and involve more than just the holoenzyme, these data indicate that the kinetics of replication rather than the explicit copy number of the DNA polymerase III holoenzyme is the more relevant feature of DNA replication to consider. In light of this, the data in **Figure 6** suggests that for bacteria like *E. coli*, DNA replication does not represent a rate-limiting step in cell division. However, it is worth noting that for bacterium like *C. crescentus* whose chromosomal replication is initiated only once per cell cycle (*Jensen et al., 2001*), the time to double their chromosome indeed represents an upper limit to their growth rate.

### RNA Synthesis

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

rRNA serves as the catalytic and structural framework along with myriad ribosomal proteins as part of a complete ribosomal complex. Each ribosome contains three rRNA molecules of lengths 120, 1542, and 2904 nucleotides (BNID: 108093), meaning each ribosome contains  $\approx 4500$  nucleotides overall. As the *E. coli* RNA polymerase transcribes DNA to RNA at a rate of  $\approx 40$  nucleotides per second (BNID: 101904), it takes a single RNA polymerase  $\approx 100$  s to synthesize the RNA needed to form a single functional ribosome. Rather than rely on an estimate for the number of ribosomes needed per cell (an estimate we consider in depth towards the end of this work), we will aim to elucidate the *maximum* number of rRNA units that can be synthesized given the 7 copies of the rRNA genes present on the *E. coli* chromosome (BNID: 100352). To do so, we will imagine that each rRNA operon is completely tiled with RNA polymerase. *In vivo* measurements of the kinetics of rRNA transcription have revealed that RNA polymerases are loaded onto the promoter of an rRNA gene at a rate of  $\approx 1$  per second (BNID: 111997, 102362). If RNA polymerases are constantly loaded at this rate, then we can assume that  $\approx 1$  functional rRNA unit is synthesized per second per rRNA operon. While *E. coli* possesses 7 of these operons per chromosome (5 of which are located closely to the origin of replication), the fact that chromosome replication can be parallelized means that the average dosage of rRNA genes can be in the range of 10 to 70 copies per cell at fast growth rates. At a growth rate of  $\approx 0.5 \text{ hr}^{-1}$ , the average cell has  $\approx 1$  copy of its chromosome and therefore approximately  $\approx 7$  copies of the rRNA operons, therefore producing  $\approx 7$  rRNA units per second. At fast growth rates, this can be

405 further increased as multiple copies of the chromosome will be present per cell, yielding large rRNA gene dosages  
406 between 10 and 70 copies per cell. With a 5000 second division time, this means the cell is able to generate around  
407  $3 \times 10^4$  functional rRNA units, comparable within an order of magnitude to the number of ribosomes per cell.

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

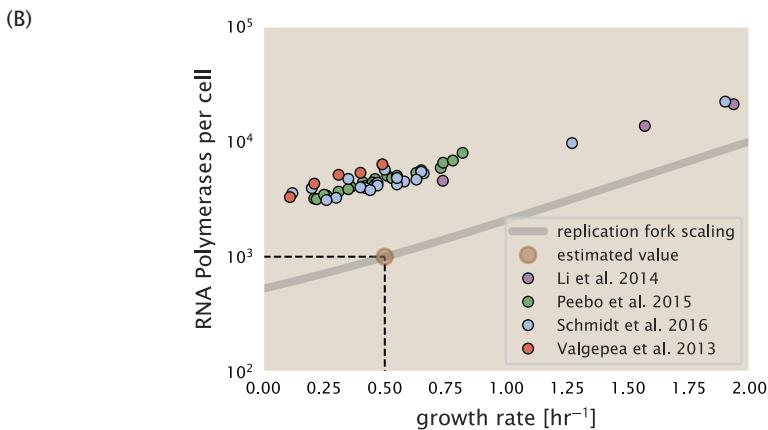
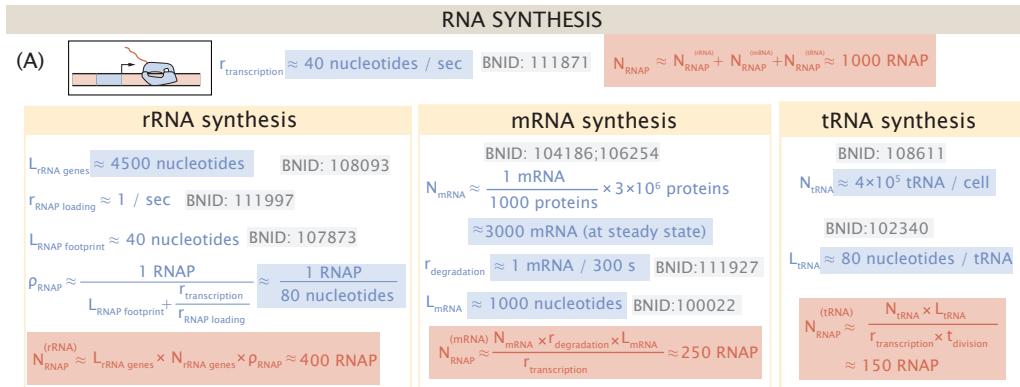
414 The synthesis of rRNA demands the majority share of the required RNA polymerase per cell. As outlined in  
415 **Figure 7**, and discussed further the Appendix ??, synthesis of mRNA and tRNA together require on the order of  
416  $\approx 400$  RNAP. Thus, in total, one would expect the typical cell to require  $\approx 1000$  RNAP to satisfy its transcriptional  
417 demands. As is revealed in **Figure 7(B)**, this estimate is about an order of magnitude below the observed number  
418 of RNA polymerase complexes per cell ( $\approx 5000 - 7000$ ). The difference between the estimated number of RNA  
419 polymerase needed for transcription and these observations are consistent with recent literature revealing  
420 that  $\approx 80\%$  of RNA polymerases in *E. coli* are not transcriptionally active (Patrick et al., 2015).

421 Our estimate neglects other mechanistic features of transcription and transcriptional initiation more broadly.  
422 For example, we acknowledge that some fraction of the RNAP pool is nonspecifically bound to DNA searching for  
423 a promoter from which to begin transcription. Furthermore, we ignore the obstacles that RNA polymerase and  
424 DNA polymerase present for each other as they move along the DNA (Finkelstein and Greene, 2013). Finally, we  
425 neglect the fact that, while they are the machinery for transcription, RNA polymerase is not sufficient to initiate  
426 transcription. Promoter recognition and initiation of transcription is dependent on the presence of  $\sigma$ -factors,  
427 proteinaceous cofactors which bind directly to the polymerase (Browning and Busby, 2016). In **Figure 7–Figure**  
428 **Supplement 1**, we show that the predicted RNA polymerase copy number is more comparable with the abundance  
429 of  $\sigma$ -70 (RpoD), the primary sigma factor in *E. coli*. There remains more to be investigated as to what sets the  
430 observed abundance of RNA polymerase in these proteomic data sets.

### 431 Protein Synthesis

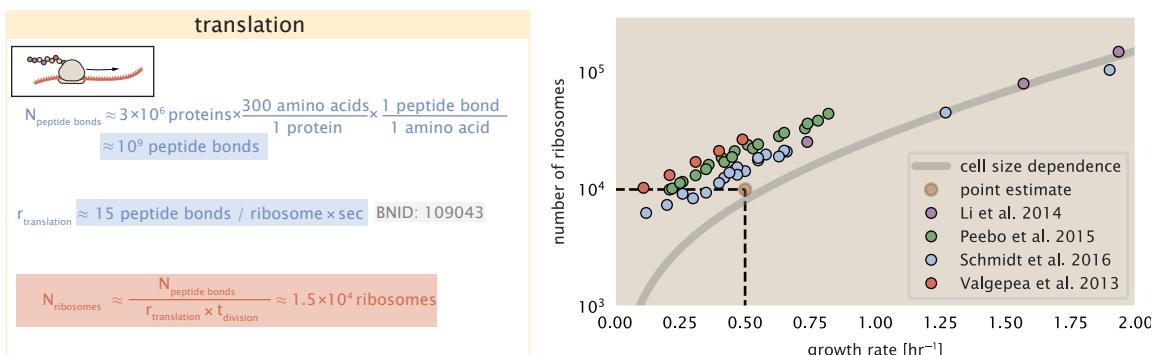
432 The final process of the central dogma – the translation of RNA into function protein – becomes the next and  
433 final target for our back-of-the-envelope estimates. To do so, we consider the abundance of ribosomes needed  
434 to replicate the cellular proteome. While the rate at which ribosomes translates is well known to have a growth  
435 rate dependence Dai et al. (2018) and we consider this more carefully in the sections that follow, here we make  
436 the approximation that translation occurs at a rate of  $\approx 15$  amino acids per second per ribosome (BNID: 100233).  
437 Under this approximation and assuming a division time of 5000 s, we can arrive at an estimate of  $\approx 10^4$  ribosomes  
438 are needed to replicate the cellular proteome, shown in **Figure 8**. This point estimate and the corresponding  
439 estimate across a continuum of growth rates proves to be notably comparable to the experimental observations,  
440 shown in **Figure 8(B)**. While the ribosome is responsible for the formation of peptide bonds, we do not diminish  
441 the importance of the charging of tRNAs with the appropriate amino acid, a process with occurs with remarkable  
442 accuracy. In the Appendix and in **Figure 8–Figure Supplement 1**, we consider the process of ligating tRNAs to their  
443 corresponding amino acid and again find notable comparability with the data.

444 To this point, the general back-of-the-envelope estimates have been reasonably successful in predicting the  
445 scale of absolute protein copy number as well as their observed dependence on the cellular growth rate. A recurring  
446 theme across these varied biological processes is the ability of cells to parallelize tasks through the expression  
447 of additional proteins. Even when that is not possible, like in chromosomal replication which requires a minimum  
448 of  $\approx 40$  minutes, *E. coli* and many other bacteria surpass this limit by initiating additional rounds of replication  
449 per doubling as we have noted. However, the synthesis of ribosomal proteins presents a special case where parallelization  
450 is not possible and must be doubled in quantity on average with every cell division (**Figure 9(A)**). In  
451 the remaining sections of this work, we will consider in detail how this limit leads to the observed relationship  
452 between growth rate and cell volume.



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

**Figure 7-Figure supplement 1.** Abundance and growth rate dependence of  $\sigma$ -70.



**Figure 8. Estimation of the required number of ribosomes.** Estimation of the number of ribosomes required to synthesize  $10^9$  peptide bonds with an elongation rate of 15 peptide bonds per second. The average abundance of ribosomes is plotted as a function of growth rate. Our estimated values are shown for a growth rate of  $0.5 \text{ hr}^{-1}$ . Grey lines correspond to the estimated complex abundance calculated at different growth rates.

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

453 **Translation and Ribosomal Synthesis as a Growth-Rate Limiting Step**

454 To gain some intuition into how translation and ribosomal synthesis may limit bacterial growth, we again consider  
455 the total number of peptide bonds that must be synthesized, which we denote as  $N_{\text{pep}}$ . With cells growing expo-  
456 nentially in time (*Godin et al., 2010*), the rate of cellular growth will be related to the rate of protein synthesis by

457

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

458 where  $\lambda$  is the cell growth rate in  $\text{s}^{-1}$ ,  $r_t$  is the maximum elongation rate in  $\text{AA}\cdot\text{s}^{-1}$ , and  $R$  is the average ribosome  
459 copy number per cell. The addition factor  $f_a$  refers to the fraction of actively translating ribosomes, and allows us to  
460 account for the possibility of nonfunctional, immature ribosomes or active sequestration of ribosomes, mediated  
461 by the secondary-messenger molecule alarmones, guanosine pentaphosphate [(p)ppGpp] at slow growth (*Dennis*  
462 *et al., 2004; Dai et al., 2016*). Knowing the number of peptide bonds formed per cell permits us to compute the  
463 translation-limited growth rate as

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

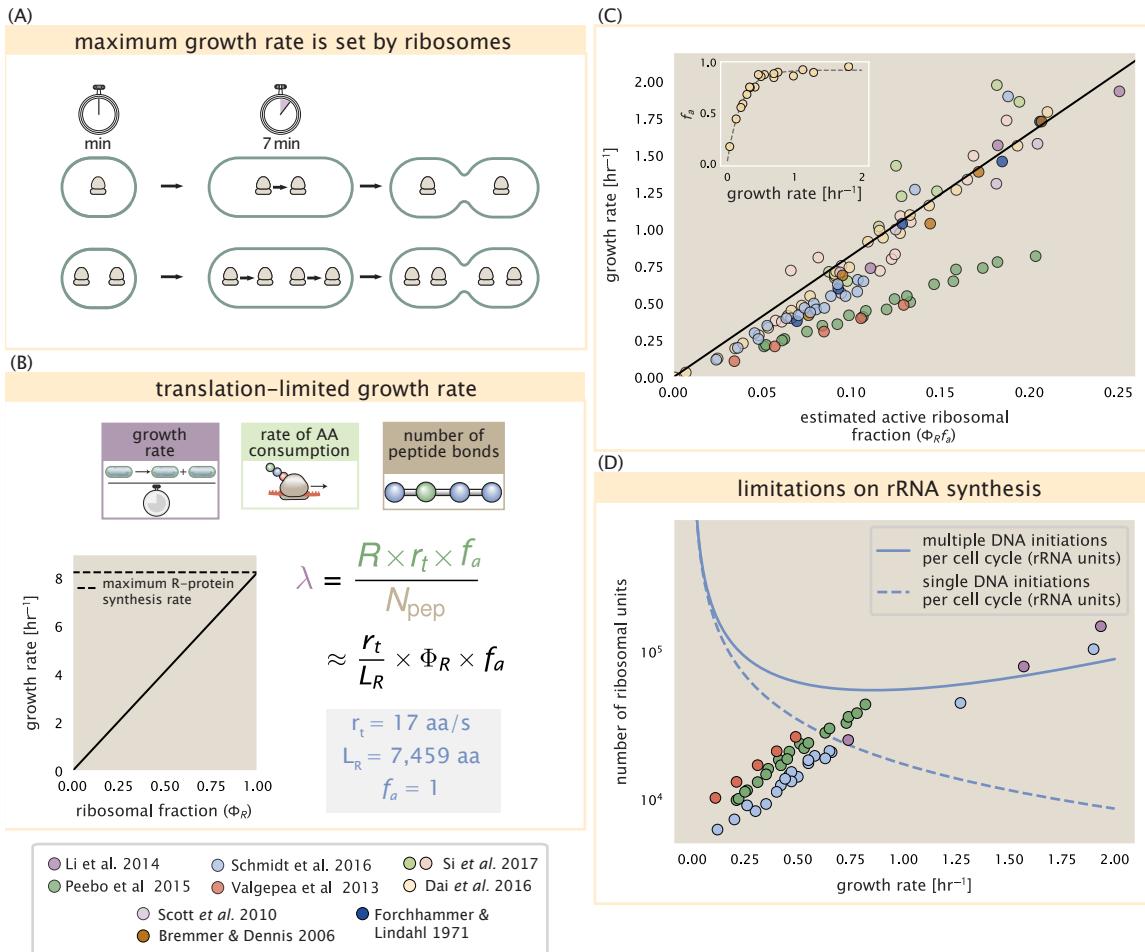
464 Alternatively, since  $N_{\text{pep}}$  is related to the total protein mass through the molecular weight of each protein, we  
465 can also consider the growth rate in terms of the fraction of the total proteome mass dedicated to ribosomal  
466 proteins. By making the approximation that an average amino acid has a molecular weight of 110 Da (BNID:  
467 104877), the total protein mass  $m_{\text{protein}}$  is related to  $N_{\text{pep}}$  by  $(m_{\text{protein}}/110 \text{ Da}) \times N_A$ , where  $N_A$  is Avogadro's number.  
468 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  
469 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  
470 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  
471 length in amino acids that make up a ribosome. The translation-limited growth rate can then be written in the  
472 form

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

473 This is plotted as a function of ribosomal fraction  $\Phi_R$  in *Figure 9(B)*, where we take  $L_R \approx 7500 \text{ AA}$ , corresponding  
474 to the length in amino acids for all ribosomal subunits of the 50S and 30S complex (BNID: 101175), and  $f_a =$   
475 1. In *Figure 9(C)* we use the recent measurements of  $f_a$  from *Dai et al. (2016)* to estimate the active fraction of  
476 ribosomal protein across the proteomic data sets and number of other recent measurements. We see that cells  
477 are consistently skirting the limit in growth rate set by *Equation 3* as nutrient conditions vary.

478 The growth rate defined by *Equation 3* reflects mass-balance under steady-state growth and has long provided  
479 a rationalization of the apparent linear increase in *E. coli*'s ribosomal content as a function of growth rate (*Maaløe,*  
480 *1979; Scott et al., 2010*). The maximum rate, when  $\Phi_R = 1$ , could only be achieved if a cell contained only ribo-  
481 somes. This corresponds to the synthesis time of all ribosomal subunits,  $L_R/r_t \approx 7 \text{ minutes}$  (*Dill et al., 2011*) and  
482 interestingly, is independent of the absolute number of ribosomes. This is because, in order to double the cell's  
483 ribosomal mass, each ribosome must produce a second ribosome; a process which cannot be parallelized. Unless  
484 elongation rate increased, or cells could trim their total ribosomal protein mass, this dependency limits both the  
485 maximum growth rate (when  $\Phi_R = 1$ ), and also the achievable growth rate under more realistic values of  $\Phi_R$ .

486 *E. coli* rarely exhibits growth rates above  $2 \text{ hr}^{-1}$  (*Bremer and Dennis, 2008; Roller et al., 2016*) and in *Figure 9(C)*  
487 we consider ribosomal generation from the perspective of rRNA synthesis. Here we use our rule-of-thumb of  
488 1 functional rRNA unit per second per operon and estimate the maximum number of ribosomes that could be  
489 made as a function of growth rate (blue curve). Although we expect this estimate to drastically overestimate  
490 rRNA abundance at slower growth rates ( $\lambda < 0.5 \text{ hr}^{-1}$ ), it provides a useful reference alongside the proteomic  
491 measurements. For growth rates above about  $1 \text{ hr}^{-1}$ , we find that cells will need to transcribe rRNA near their  
492 maximal rate. As a counter example, if *E. coli* did not initiate multiple rounds of replication, they would be unable  
493 to make enough rRNA for the observed number of ribosomes (dashed blue curve in *Figure 9(C)*). The convergence  
494 between the maximum rRNA production and measured ribosome copy number suggests rRNA synthesis may  
495 begin to present a bottleneck at the fastest growth rates due to the limited copies of rRNA genes.



**Figure 9. 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 [Dai et al. \(2016\)](#) (shown in inset; see Appendix Calculation of active ribosomal fraction 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 (# ori) at each growth rate. The quantity (# ori) was calculated using Equation 4 and the measurements from [Si et al. \(2017\)](#) that are plotted in [Figure 10\(A\)](#). The dashed line shows the maximal number of functional rRNA units produced from a single chromosomal initiation per cell cycle.

## 496 Relationship Between Cell Size and Growth Rate

497 The relationship between cell size and growth rate has long been of interest in the study of bacterial physiology,  
498 particularly following the now six decade-old observation that cell volume appears to increase exponentially with  
499 growth rate; known as Schaechter's growth law (*Schaechter et al., 1958; Taheri-Araghi et al., 2015*). However, the  
500 mechanism that governs this relationship, and even the question of whether the change in average cell size is truly  
501 exponential, has remained under debate (*Harris and Theriot, 2018*). Here we examine the influence of ribosomal  
502 content and total protein abundance on cell size.

503 Cells grow at a near-maximal rate dictated by their total ribosomal mass fraction  $\Phi_R$ , at least at moderate  
504 growth rates above  $0.5 \text{ hr}^{-1}$  (where  $f_a$  is close to 1, and  $r_i$  is near its maximal rate). Here, growth rate can be  
505 increased only by increasing  $\Phi_R$ , though the simple addition of more ribosomes is likely constrained by aspects  
506 physical constrains like macromolecular crowding (*Delarue et al., 2018; Soler-Bistué et al., 2020*). As *E. coli* grows  
507 faster, large swaths of its proteome increase in absolute abundance. It is now well-documented that *E. coli* cells  
508 add a constant volume per origin of replication, which is robust to a remarkable array of cellular perturbations  
509 (*Si et al., 2017*). To consider this dependency in the context of the proteomic data, we used measurements from  
510 *Si et al. (2017)* (*Figure 10(A)*) to estimate the average number of origins per cell ( $\langle \# \text{ ori} \rangle$ ) at different growth rates.  
511  $\langle \# \text{ ori} \rangle$  is set by how often replication must be initiated per cell doubling under steady-state growth. This can be  
512 quantified as

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

513 where  $t_{\text{cyc}}$  is the cell cycle time (referring to the time from replication initiation to cell division), and  $\tau$  is the cell dou-  
514 bling time. For ribosomal synthesis, we find an approximately linear correlation between ribosome copy number  
515 and  $\langle \# \text{ ori} \rangle$  (*Figure 10(B)*).

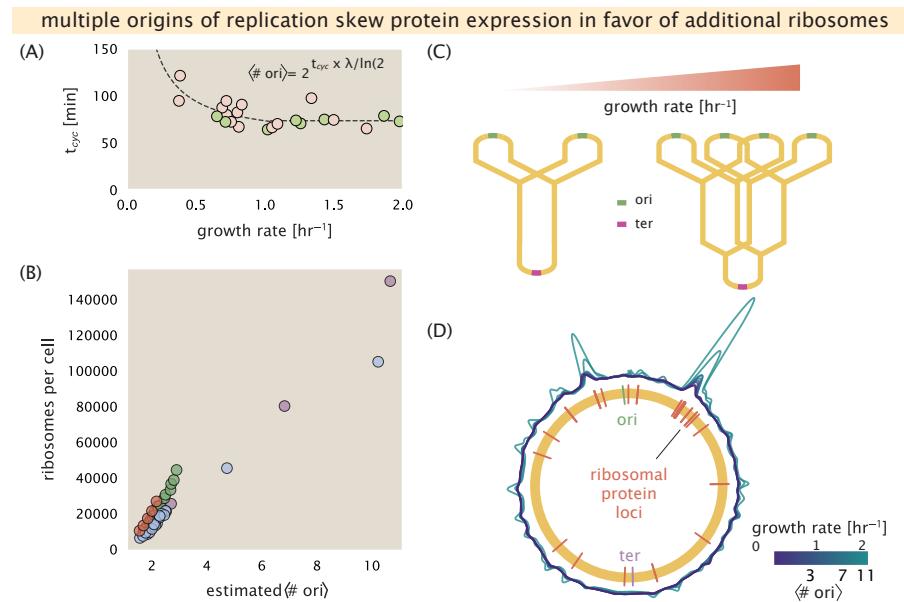
516 For a constant cell cycle time, observed at growth rates above about  $0.5 \text{ hr}^{-1}$  (*Helmstetter and Cooper, 1968*),  
517 **Equation 4** states that  $\langle \# \text{ ori} \rangle$  will need to increase exponentially with the growth rate. While this says nothing of  
518 the observed scaling with cell size, the additional dependency on ribosomal content, which increases with  $\langle \# \text{ ori} \rangle$ ,  
519 provides a link. In *Figure 10(D)*, we consider the position-dependent protein expression across the chromosome by  
520 calculating a running Gaussian average of protein copy number (20 kbp st. dev. averaging window) based on each  
521 gene's transcriptional start site, which were then median-subtracted to account for the differences in total protein  
522 abundance. Importantly, major deviations in protein copy number are largely restricted to regions of ribosomal  
523 protein genes. This suggests that the relative ribosomal abundance  $\Phi_R$  is also being tuned in proportion to  $\langle \# \text{ ori} \rangle$ ,  
524 with the exponential relationship between cell size and growth rate following from how *E. coli* varies its number  
525 of ribosomes per cell.

## 526 Nutrient-Mediated Regulation of Proteomic Composition and Growth Rate

527 To react to changes in nutrient conditions, bacteria rely on the synthesis or degradation of secondary-messenger  
528 molecules like (p)ppGpp, which cause global changes in transcriptional and translational activity. In *E. coli*, amino  
529 acid starvation causes the accumulation of de-acylated tRNAs at the ribosome's A-site and leads to a strong in-  
530 crease in (p)ppGpp synthesis activity by the enzyme RelA (*Hauryliuk et al., 2015*). Cells also accumulate (p)ppGpp  
531 during steady-state growth in poorer growth conditions, which leads to a decrease in the fraction of actively trans-  
532 lating ribosomes,  $f_a$  (with  $f_a \approx 0.5$  at a growth rate of  $\approx 0.3 \text{ hr}^{-1}$ ; *Figure 9(C)* - inset).

533 Furthermore, (p)ppGpp can inhibit the initiation of DNA replication by mediating a change in transcriptional  
534 activity and the supercoiling state of the origin of replication (*Kraemer et al., 2019*). These observations all raise  
535 the possibility that it is through (p)ppGpp that cells mediate the growth-rate dependent changes in  $\langle \# \text{ ori} \rangle$ , cell size,  
536 and ribosomal abundance and activity (*Zhu and Dai, 2019; Büke et al., 2020*). Indeed, recent work in a (p)ppGpp  
537 deficient strain of *E. coli* found that cells exhibited a high ratio in the number of origins of replication to the number  
538 of termini, and cell sizes that were more consistent with a fast growth state where (p)ppGpp levels are normally  
539 low (*Fernández-Coll et al., 2020*).

540 As we have seen, cell size, total proteomic content, and the number of ribosomes are all interconnected and  
541 influence the achievable growth rate. The drastic changes observed in cell size, proteomic composition, and ribo-  
542 some abundance across different growth conditions suggests a hypothesis that each parameter is being tuned  
543 to better match the cell's biosynthetic capacity to the specific environment. In this final section, we explore the



**Figure 10. Cells increase absolute ribosome abundance with  $\langle \# \text{ori} \rangle$ .** (A) Experimental data from Si *et al.* (2017). Dashed line shows fit to the data, which were used to estimate  $\langle \# \text{ori} \rangle$ .  $t_{cyc}$  was assumed to vary in proportion to  $\tau$  for doubling times greater than 40 minutes, and reach a minimum value of 73 minutes (see Appendix Estimation of  $\langle \# \text{ori} \rangle / \langle \# \text{ter} \rangle$  and  $\langle \# \text{ori} \rangle$  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 (Schmidt *et al.*, 2016). 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 4.

544 interconnection between cell size, ribosome content, and growth rate by formulating a minimal model of growth  
 545 rate control. We use it to quantitatively show how tuning these parameters help cells maximize their growth rate  
 546 for a particular environment.

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

548 Here we consider a mode of regulation in which the rate of peptide elongation  $r_t$  depends only on the availability  
 549 of amino acids (and, therefore, also amino-acyl tRNAs). It is through the elongation rate  $r_t$  that we assume cells  
 550 adjust their ribosomal content ( $R, \Phi_R$ ) according to nutrient availability and for simplicity, do not explicitly model  
 551 changes in (# ori) or regulation by (p)ppGpp.

552 The rate of elongation  $r_t$  will depend on how quickly the ribosomes can match codons with their correct amino-  
 553 acyl tRNA, along with the subsequent steps of peptide bond formation and translocation. We therefore coarse-  
 554 grain the steps of elongation to two time-scales, 1) the time required to find and bind each correct amino-acyl  
 555 tRNA, and 2) the remaining steps in peptide elongation that will not depend on the amino acid availability. The  
 556 availability of amino acids will depend on their cellular concentration, which we treat as a single effective species,  
 557  $[AA]_{\text{eff}}$ . Under this model, other molecular players required for translation like elongation factors and GTP are  
 558 considered in sufficient abundance, which appear to be valid assumptions given our analysis of the proteomic  
 559 data and energy production thus far. The time to translate each codon is given by the inverse of the elongation  
 560 rate  $r_t$ , which can be written as,

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

561 where we have assumed that the rate of binding by amino-acyl tRNA  $k_{on}$  is proportional to  $[AA]_{\text{eff}}$  by a constant  
 562  $\alpha$ . The second term on the right-hand side reflects our assumption that other steps in peptide elongation are  
 563 not rate-limiting, with a maximum elongation rate  $r_t^{\max}$  of about 17 amino acids per second **Dai et al. (2016)**. As  
 564 the rate of amino acid supply, denote by  $r_{AA}$ , varies with changing nutrient conditions, the cell can maximize  
 565 the rate of protein synthesis by tuning the rate of amino acid consumption ( $r_t \times R \times f_a$ ), shown schematically in  
 566 **Figure 11(A)**. This can be stated more succinctly in terms of an effective dissociation constant,  $K_D = r_t^{\max}/\alpha k_{on}$ ,  
 567 where the elongation rate  $r_t$  is now given by

$$r_t = \frac{r_t^{\max}}{1 + K_D/[AA]_{\text{eff}}}. \quad (6)$$

568 Under steady-state growth, the amino acid concentration is constant ( $\frac{d[AA]_{\text{eff}}}{dt} = 0$ ) and will relate to the rate  
 569 of amino acid synthesis (or import, for rich media) and/or tRNA charging, as  $r_{AA}$ , and the rate of consumption,  
 570  $r_t \times R \times f_a$ . We calculate  $[AA]_{\text{eff}}$  by,

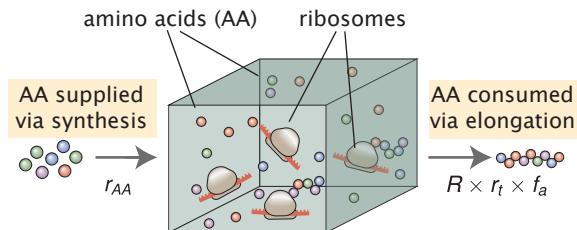
$$[AA]_{\text{eff}} = \frac{t(r_{AA} - r_t \times R \times f_a)}{V}, \quad (7)$$

571 which allows us to then solve for  $r_t$  explicitly (further described in Appendix ??). Here  $r_{AA}$  is in units of AA per unit  
 572 time, and  $V$  reflects the volume of the cell over a time period  $t$ .

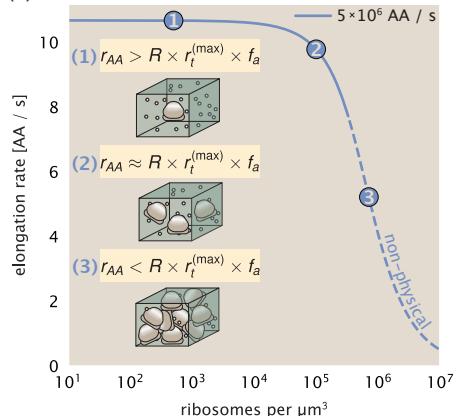
573 In **Figure 11(B)**, we illustrate how the elongation rate depends on the ribosomal copy number. Here, we have  
 574 considered a unit volume  $V = 1\mu\text{m}^3$ , a unit time  $t = 1\text{ s}$ , a  $K_D = 5\text{ mM}$  (inferred from **Bennett et al. (2009)**),  
 575  $f_a = 1$ , and an arbitrarily chosen  $r_{AA} = 5 \times 10^6 \text{ AA} \cdot \text{s}^{-1} \cdot \mu\text{m}^{-3}$ . At low ribosome copy numbers, the observed  
 576 elongation rate is dependent primarily on the ratio of  $K_D/Vr_{AA}$  [as  $r_t^{\max} \times R \times f_a \ll r_{AA}$ , point (1) in **Figure 11(B)**].  
 577 As the ribosome copy number is increased such that the amino acid supply rate and consumption rate are nearly  
 578 equal [point (2) in **Figure 11(B)**], the observed elongation rate begins to decrease sharply. When the ribosome  
 579 copy number is increased even further, consumption at the maximum elongation rate exceeds the supply rate,  
 580 yielding a significantly reduced elongation rate [point (3) in **Figure 11B**]. While the elongation rate will always be  
 581 dominated by the amino acid supply rate at sufficiently low ribosome copy numbers, the elongation rate at larger  
 582 ribosome abundances can be increased by tuning  $f_a$  such that not all ribosomes are elongating, reducing the total  
 583 consumption rate.

(A)

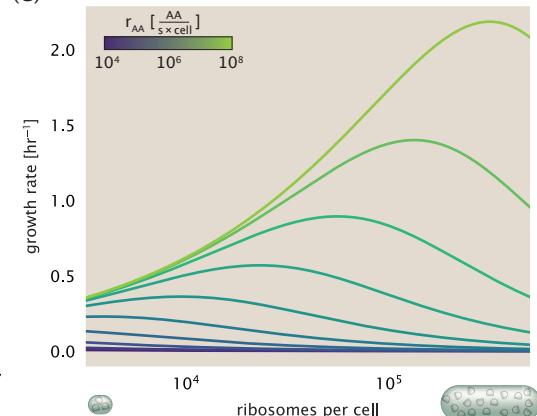
## A MINIMAL MODEL FOR NUTRIENT-LIMITED GROWTH



(B)



(C)



**Figure 11. A minimal model of growth rate control under nutrient limitation.** (A) We consider a unit volume of cellular material composed of amino acids (colored spheres) provided at a supply rate  $r_{AA}$ . These amino acids are polymerized by a pool of ribosomes (brown blobs) at a rate  $r_t \times R \times f_a$ , where  $r_t$  is the elongation rate,  $R$  is the ribosome copy number in the unit volume, and  $f_a$  is the fraction of those ribosomes actively translating. (B) The observed elongation rate is plotted as a function of ribosomes 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.

584 **Optimal Growth Rate, Ribosomal Content, and Cell Size Depend on Nutrient Availability and  
585 Metabolic Capacity.**

586 To relate elongation rate to growth rate, we constrain the set of parameters based on our available proteomic mea-  
587 surements; namely, we restrict the values of  $R$ ,  $N_{pep}$ , and  $V$  to those associated with the amalgamated proteomic  
588 data (described in Appendix Estimation of Total Protein Content per Cell). We then consider how changes in the  
589 nutrient conditions, through the parameter  $r_{AA}$ , influence the maximum growth rate as determined by **Equation 2**.  
590 **Figure 11(C)** shows how the observed growth rate depends on the rate of amino acid supply  $r_{AA}$  as a function of  
591 the cellular ribosome copy number. A feature immediately apparent is the presence of a maximal growth rate  
592 whose dependence on  $R$  (and consequently, the cell size) increases with increasing  $r_{AA}$ . Importantly, however,  
593 there is an optimum set of  $R$ ,  $N_{pep}$ , and  $V$  that are strictly dependent on the value of  $r_{AA}$ . Increasing the ribosomal  
594 concentration beyond the cell's metabolic capacity has the adverse consequence of depleting the supply of amino  
595 acids and a concomitant decrease in the elongation rate  $r_e$  [**Figure 11(B)**].

596 Also of note is the growth rate profiles shown for low amino acid supply rates [purple and blue lines in **Fig-  
597 ure 11(C)**], representing growth in nutrient-poor media. In these conditions, there no longer exists a peak in  
598 growth, at least in the range of physiologically-relevant ribosome copy numbers. Instead, cells limit their pool of  
599 actively translating ribosomes by decreasing  $f_a$  ([Dai et al., 2016](#)), which would help maintain the pool of available  
600 amino acids  $[AA]_{eff}$  and increase the achievable elongation rate. This observation is in agreement with the central  
601 premise of the cellular resource allocation principle proposed by [Scott et al. \(2010\)](#); [Klumpp et al. \(2009\)](#); [Klumpp  
602 and Hwa \(2014\)](#) and [Hui et al. \(2015\)](#).

603 **Discussion**

604 Continued experimental and technological improvements have led to a treasure trove of quantitative biological  
605 data ([Hui et al., 2015](#); [Schmidt et al., 2016](#); [Si et al., 2017](#); [Gallagher et al., 2020](#); [Peebo et al., 2015](#); [Valgepea et al.,  
606 2013](#)), and an ever advancing molecular view and mechanistic understanding of the constituents that support  
607 bacterial growth ([Taheri-Araghi et al., 2015](#); [Morgenstein et al., 2015](#); [Si et al., 2019](#); [Karr et al., 2012](#); [Kostinski and  
608 Reveni, 2020](#)). In this work we have compiled what we believe to be the state-of-the-art knowledge on proteomic  
609 copy number across a broad range of growth conditions in *E. coli*. We have made this data accessible through a  
610 [GitHub repository](#), and an interactive figure that allows exploration of specific protein and protein complex copy  
611 numbers. Through a series of order-of-magnitude estimates that traverse key steps in the bacterial cell cycle, this  
612 proteomic data has been a resource to guide our understanding of two key questions: what biological processes  
613 limit the absolute speed limit of bacterial growth, and how do cells alter their molecular constituents as a function  
614 of changes in growth rate or nutrient availability? While not exhaustive, our series of estimates provide insight  
615 on the scales of macromolecular complex abundance across four classes of cellular processes – the transport of  
616 nutrients, the production of energy, the synthesis of the membrane and cell wall, and the numerous steps of the  
617 central dogma.

618 In general, the copy numbers of the complexes involved in these processes were reasonable agreement with  
619 our order-of-magnitude estimates. Since many of these estimates represent soft lower-bound quantities, this  
620 suggests that cells do not express proteins grossly in excess of what is needed for a particular growth rate. Several  
621 exceptions, however, also highlight the dichotomy between a proteome that appears to "optimize" expression  
622 according to growth rate and one that must be able to quickly adapt to environments of different nutritional quality.  
623 Take, for example, the expression of carbon transporters. Shown in **Figure 2(B)**, we find that cells always express  
624 a similar number of glucose transporters irrespective of growth condition. At the same time, it is interesting to  
625 note that many of the alternative carbon transporters are still expressed in low but non-zero numbers ( $\approx 10$ -  
626 100 copies per cell) across growth conditions. This may relate to the regulatory configuration for many of these  
627 operons, which require the presence of a metabolite signal in order for alternative carbon utilization operons to  
628 be induced ([Monod, 1949](#); [Laxhuber et al., 2020](#)). Furthermore, upon induction, these transporters are expressed  
629 and present in abundances in close agreement with a simple estimate.

630 Of the processes illustrated in **Figure 1**, we arrive at a ribosome-centric view of cellular growth rate control.  
631 This is in some sense unsurprising given the long-held observation that *E. coli* and many other organisms vary  
632 their ribosomal abundance as a function of growth conditions and growth rate [Scott et al. \(2010\)](#); [Metzl-Raz et al.](#)

(2017). However, through our dialogue with the proteomic data, two additional key points emerge. The first relates to our question of what process sets the absolute speed limit of bacterial growth. While a cell can parallelize many of its processes simply by increasing the abundance of specific proteins or firing multiple rounds of DNA replication, this is not so for synthesis of ribosomes (Figure 9(A)). The translation time for each ribosome [ $\approx$  6 min, Dill et al. (2011)] places an inherent limit on the growth rate that can only be surpassed if the cell were to increase their polypeptide elongation rate, or if they could reduce the total protein and rRNA mass of the ribosome. The second point relates to the long-observed correlations between growth rate and cell size (Schaechter et al., 1958; Si et al., 2017), and between growth rate and ribosomal mass fraction. While both trends have sparked tremendous curiosity and driven substantial amounts of research in their own regards, these relationships are themselves intertwined. In particular, it is the need for cells to increase their absolute number of ribosomes under conditions of rapid growth that require cells to also grow in size. Further experiments are needed to test the validity of this hypothesis. In particular, we believe that the change in growth rate in response to translation-inhibitory drugs (such as chloramphenicol) could be quantitatively predicted, given one had precision measurement of the relevant parameters, including the fraction of actively translating ribosomes  $f_a$  and changes in the metabolic capacity of the cell (i.e. the parameter  $r_{AA}$  in our minimal model) for a particular growth condition.

While the generation of new ribosomes plays a dominant role in growth rate control, there exist other physical limits to the function of cellular processes. One of the key motivations for considering energy production was the physical constraints on total volume and surface area as cells vary their size (Harris and Theriot, 2018; Ojikic et al., 2019). While *E. coli* get larger as it expresses more ribosomes, an additional constraint begins to arise in energy production due to a relative decrease in total surface area where ATP is predominantly produced (Szenk et al., 2017). Specifically, the cell interior requires an amount of energy that scales cubically with cell size, but the available surface area only grows quadratically (Figure 5(A)). While this threshold does not appear to be met for *E. coli* cells growing at 2 hr<sup>-1</sup> or less, it highlights an additional constraint on growth given the apparent need to increase in cell size to grow faster. This limit is relevant even to eukaryotic organisms, whose mitochondria exhibit convoluted membrane structures that nevertheless remain bacteria-sized organelles (Guo et al., 2018). In the context of bacteria growth and energy production more generally, we have limited our analysis to the aerobic growth conditions associated with the proteomic data and further consideration will be needed for anaerobic growth.

This work is by no means meant to be an exhaustive dissection of bacterial growth rate control, and there are many aspects of the bacterial proteome and growth that we neglected to consider. For example, other recent work (Liebermeister et al., 2014; Hui et al., 2015; Schmidt et al., 2016) has explored how the proteome is structured and how that structure depends on growth rate. In the work of Hui et al. (2015), the authors coarse-grained the proteome into six discrete categories being related to either translation, catabolism, anabolism, and others related to signaling and core metabolism. The relative mass fraction of the proteome occupied by each sector could be modulated by external application of drugs or simply by changing the nutritional content of the medium. While we have explored how the quantities of individual complexes are related to cell growth, we acknowledge that higher-order interactions between groups of complexes or metabolic networks at a systems-level may reveal additional insights into how these growth-rate dependences are mechanistically achieved. Furthermore, while we anticipate the conclusions summarized here are applicable to a wide collection of bacteria with similar lifestyles as *E. coli*, other bacteria and archaea may have evolved other strategies that were not considered. Further experiments with the level of rigor now possible in *E. coli* will need to be performed in a variety of microbial organisms to learn more about how regulation of proteomic composition and growth rate control has evolved over the past 3.5 billion years.

## Methods

### Data Analysis and Availability

All proteomic measurements come from the experimental work of Schmidt et al. (2016); Peebo et al. (2015); Valgepea et al. (2013) (mass spectrometry) and Li et al. (2014) (ribosomal profiling). Data curation and analysis was done programmatically in Python, and compiled data and analysis files are accessible through a [GitHub repository] (DOI:XXX) associated with this paper as well as on the associated [paper website](#). An interactive figure that

<sup>682</sup> allows exploration of specific protein and protein complex copy numbers is available at [link].

<sup>683</sup> **Acknowledgements**

<sup>684</sup> We thank Matthias Heinemann, Alexander Schmidt, and Gene-Wei Li for additional input regarding their data. We  
<sup>685</sup> also thank members of the Phillips, Theriot, Kondek, and Garcia labs for useful discussions. R.P. is supported by  
<sup>686</sup> La Fondation Pierre-Gilles de Gennes, the Rosen Center at Caltech, and the NIH 1R35 GM118043 (MIRA). J.A.T. is  
<sup>687</sup> supported by the Howard Hughes Medical Institute, and NIH Grant R37-AI036929. N.M.B is a HHMI Fellow of The  
<sup>688</sup> Jane Coffin Childs Memorial Fund.

<sup>689</sup> **Competing Interests**

<sup>690</sup> The authors declare no competing interests.

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

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

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

<b>704</b>	<b>Contents</b>	
<b>705</b>	<b>Introduction</b>	<b>1</b>
<b>706</b>	<b>Nutrient Transport</b>	<b>2</b>
<b>707</b>	<b>Cell Envelope Biogenesis</b>	<b>6</b>
<b>708</b>	Lipid Synthesis . . . . .	6
<b>709</b>	Peptidoglycan Synthesis . . . . .	7
<b>710</b>	<b>Energy Production</b>	<b>7</b>
<b>711</b>	ATP Synthesis . . . . .	9
<b>712</b>	Generating the Proton Electrochemical Gradient . . . . .	9
<b>713</b>	<b>Biosynthesis in a Crowded Membrane</b>	<b>10</b>
<b>714</b>	<b>Processes of the Central Dogma</b>	<b>11</b>
<b>715</b>	DNA Replication . . . . .	11
<b>716</b>	RNA Synthesis . . . . .	13
<b>717</b>	Protein Synthesis . . . . .	14
<b>718</b>	<b>Translation and Ribosomal Synthesis as a Growth-Rate Limiting Step</b>	<b>16</b>
<b>719</b>	Relationship Between Cell Size and Growth Rate . . . . .	18
<b>720</b>	Nutrient-Mediated Regulation of Proteomic Composition and Growth Rate . . . . .	18
<b>721</b>	Ribosomal Elongation Rate and Cellular Growth Rate are Linked by Amino Acid Scarcity . . . . .	20
<b>722</b>	Optimal Growth Rate, Ribosomal Content, and Cell Size Depend on Nutrient Availability and Metabolic Capacity. . . . .	22
<b>724</b>	<b>Discussion</b>	<b>22</b>
<b>725</b>	<b>Methods</b>	<b>23</b>
<b>726</b>	Data Analysis and Availability . . . . .	23
<b>727</b>	<b>Acknowledgements</b>	<b>24</b>
<b>728</b>	<b>Competing Interests</b>	<b>24</b>
<b>729</b>	<b>Experimental Details Behind Proteomic Data</b>	<b>28</b>
<b>730</b>	Fluorescence based measurements . . . . .	28
<b>731</b>	Ribosomal profiling measurements . . . . .	28
<b>732</b>	Mass spectrometry measurements . . . . .	29
<b>733</b>	<b>Summary of Proteomic Data</b>	<b>29</b>
<b>734</b>	<b>Estimation of Cell Size and Surface Area</b>	<b>30</b>
<b>735</b>	<b>Estimation of Total Protein Content per Cell</b>	<b>32</b>
<b>736</b>	Estimating Volume and Number of Amino Acids from Ribosome Copy Number . . . . .	32
<b>737</b>	<b>Additional Considerations of Schmidt <i>et al.</i> Data Set</b>	<b>32</b>
<b>738</b>	Effect of cell volume on reported absolute protein abundances . . . . .	34
<b>739</b>	Relaxing assumption of constant protein concentration across growth conditions . . . . .	36
<b>740</b>	Comparison with total protein measurements from Basan <i>et al.</i> 2015. . . . .	36
<b>741</b>	<b>Calculation of Complex Abundance</b>	<b>37</b>

<b>742</b>	<b>Extending Estimates to a Continuum of Growth Rates</b>	<b>39</b>
743	Estimation of the total cell mass . . . . .	39
744	Complex Abundance Scaling With Cell Volume . . . . .	39
745	A Relation for Complex Abundance Scaling With Surface Area . . . . .	40
746	Number of Lipids . . . . .	40
747	Number of Murein Monomers . . . . .	40
748	Complex Abundance Scaling With Number of Origins, and rRNA Synthesis . . . . .	41
<b>749</b>	<b>Calculation of active ribosomal fraction.</b>	<b>41</b>
<b>750</b>	<b>Estimation of <math>\langle \#ori \rangle / \langle \#ter \rangle</math> and <math>\langle \#ori \rangle</math>.</b>	<b>41</b>

**Table 1.** Overview of proteomic data sets.

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

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

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

## 751 Experimental Details Behind Proteomic Data

752 Here we provide a brief summary of the experiments behind each proteomic data set. The purpose of this section  
753 is to identify how the authors arrived at absolute protein abundances. In the following section (Section Summary  
754 of Proteomic Data) we will then provide a summary of the final protein abundance measurements that were  
755 used throughout the main text. Table 1 provides an overview of the publications we considered. These are pre-  
756 dominately mass spectrometry-based, with the exception of the work from Li *et al.* (2014) which used ribosomal  
757 profiling, and the fluorescence-based counting done in Taniguchi *et al.* (2010).

## 758 Fluorescence based measurements

759 In the work of Taniguchi *et al.* (2010), the authors used a chromosomal YFP fusion library where individual strains  
760 have a specific gene tagged with a YFP-coding sequence. 1018 of their 1400 attempted strains were used in the  
761 work. A fluorescence microscope was used to collect cellular YFP intensities across all these strains. Through au-  
762 tomated image analysis, the authors normalized intensity measurements by cell size to account for the change  
763 in size and expression variability across the cell cycle. Following correction of YFP intensities for cellular autoflu-  
764 orescence, final absolute protein levels were determined by a calibration curve with single-molecule fluorescence  
765 intensities. This calibration experiment was performed separately using a purified YFP solution.

## 766 Ribosomal profiling measurements

767 The work of Li *et al.* (2014) takes a sequencing based approach to estimate protein abundance. Ribosomal pro-  
768 filing, which refers to the deep sequencing of ribosome-protected mRNA fragments, can provide a quantitative  
769 measurement of the protein synthesis rate. As long as the protein life-time is long relative to the cell doubling  
770 time, it is possible to estimate absolute protein copy numbers. The absolute protein synthesis rate has units of  
771 proteins per generation, and for stable proteins will also correspond to the protein copy number per cell.

772 In the experiments, ribosome-protected mRNA is extracted from cell lysate and selected on a denaturing poly-  
773 acrylamide gel for deep sequencing (15–45 nt long fragments collected and sequenced by using an Illumina HiSeq  
774 2000 in Li *et al.* (2014)). Counts of ribosome footprints from the sequencing data were then corrected empiri-  
775 cally for position-dependent biases in ribosomal density across each gene, as well as dependencies on specific  
776 sequences including the Shine-Dalgarno sequence. These data-corrected ribosome densities represent relative  
777 protein synthesis rates. Absolute protein synthesis rates are obtained by multiplying the relative rates by the total  
778 cellular protein per cell. The total protein per unit volume was determined with the Lowry method to quantify  
779 total protein, calibrated against bovine serum albumin (BSA). By counting colony-forming units following serial  
780 dilution of their cell cultures, they then calculated the total protein per cell.

781 **Mass spectrometry measurements**

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

789 Most quantitative experiments rely on labeling protein with stable isotopes, which allow multiple samples to  
790 be measured together by the mass spectrometer. By measuring samples of known total protein abundance simul-  
791 taneously (i.e. one sample of interest, and one reference), it is possible to determine relative protein abundances.  
792 Absolute protein abundances can be estimated following the same approach used above for ribosomal profil-  
793 ing, which is to multiply each relative abundance measurement by the total cellular protein per cell. This is the  
794 approach taken by *Valgepea et al. (2013)* and *Peebo et al. (2015)*, with relative protein abundances determined  
795 based on the relative peptide intensities (label free quantification 'LFQ' intensities). For the data of *Valgepea et al.*  
796 (*2013*), total protein per cell was determined by measuring total protein by the Lowry method, and counting colony-  
797 forming units following serial dilution. For the data from *Peebo et al. (2015)*, the authors did not determine cell  
798 quantities and instead report the cellular protein abundances in protein per unit volume by assuming a mass  
799 density of 1.1 g/ml, with a 30% dry mass fraction.

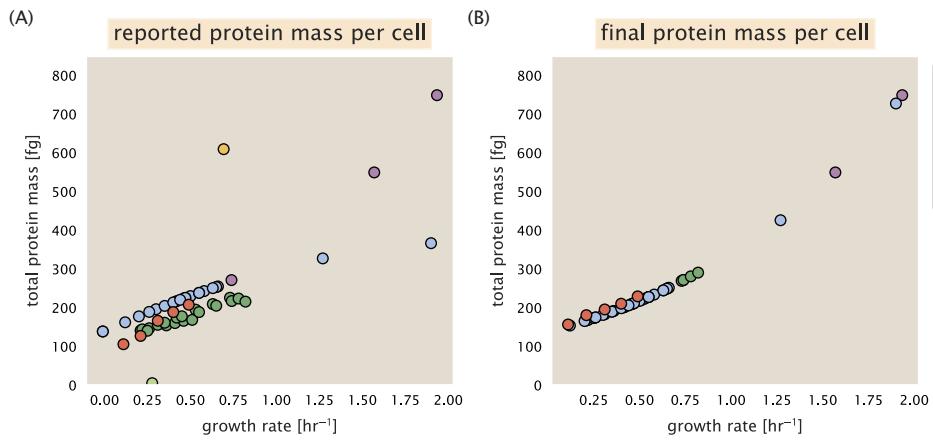
800 An alternative way to arrive at absolute protein abundances is to dope in synthetic peptide fragments of known  
801 abundance. These can serve as a direct way to calibrate mass spectrometry signal intensities to absolute mass.  
802 This is the approach taken by *Schmidt et al. (2016)*. In addition to a set of shotgun proteomic measurements to  
803 determine proteome-wide relative abundances, the authors also performed absolute quantification of 41 proteins  
804 covering over four orders of magnitude in cellular abundance. Here, a synthetic peptide was generated for each of  
805 the proteins, doped into each protein sample, and used these to determine absolute protein abundances of the 41  
806 proteins. These absolute measurements, determined for every growth condition, were then used as a calibration  
807 curve to convert proteomic-wide relative abundances into absolute protein abundance per cell. A more extensive  
808 discussion of the *Schmidt et al. (2016)* data set can be found in Section Additional Considerations of Schmidt *et al.*  
809 Data Set.

810 **Summary of Proteomic Data**

811 In the work of the main text we only used the data from *Valgepea et al. (2013)*; *Li et al. (2014)*; *Peebo et al. (2015)*;  
812 *Schmidt et al. (2016)*. As shown in *Figure 12(A)*, the reported total protein abundances in the work of *Taniguchi*  
813 *et al. (2010)* and *Soufi et al. (2015)* differed quite substantially from the other work. For the work of *Taniguchi*  
814 *et al. (2010)* this is in part due to a lower coverage in total proteomic mass quantified, though we also noticed that  
815 most proteins appear undercounted when compared to the other data.

816 *Figure 12(B)* summarizes the total protein mass for each data point in our final compiled data set. We note that  
817 protein abundances were all scaled so they followed a common growth rate-dependent change in total protein  
818 mass. While our inclination initially was to leave reported copy numbers untouched, a notable discrepancy in the  
819 scaling total protein per cell between *Schmidt et al. (2016)* and the other data sets forced us to dig deeper into  
820 those measurements (compare *Schmidt et al. (2016)* and *Li et al. (2014)* data in *Figure 12(A)*). The particular trend  
821 in *Schmidt et al. (2016)* appears to be due to assumptions of cell size and we provide a more extensive discussion  
822 and analysis of that data set in section Additional Considerations of Schmidt *et al.* Data Set. As a compromise, and  
823 in an effort to treat all data equally, we instead scaled all protein abundance values to a data-driven estimate of  
824 total protein per cell. Here we used cell size measurements from *Si et al. (2017, 2019)*, and an estimate of total  
825 protein content through expected dry mass. Total protein per cell was estimated using available data on total  
826 DNA, RNA, and protein from *Basan et al. (2015)*; *Dai et al. (2016)*, which account for the majority of dry mass in the  
827 cell. We consider these details in sections Estimation of Cell Size and Surface Area and Estimation of Total Protein  
828 Content per Cell that follows.

829 Lastly, in *Figure 13* we show the total proteomic coverage and overlap of proteins quantified across each data



**Figure 12. Summary of the growth-rate dependent total protein abundance for each data set.** (A) Total protein abundance per cell as original reported in the data sets of *Taniguchi et al. (2010)*; *Valgepea et al. (2013)*; *Li et al. (2014)*; *Soufi et al. (2015)*; *Peebo et al. (2015)*; *Schmidt et al. (2016)*. Note that the data from *Peebo et al. (2015)* only reported protein abundances per unit volume and total protein per cell was found by multiplying these by the growth-rate dependent cell size as determined by *Si et al. (2017)*. (B) Adjusted total protein abundances across the proteomic data sets are summarized. Protein abundances were adjusted so that all data shared a common set of growth-rate dependent total protein per cell and cellular protein concentration following the cell size expectations of *Si et al. (2017)* (see section on Estimation of Cell Size and Surface Area for further details).

set. Here we have used an UpSet diagram (*Lex et al., 2014*) to compare the data. Overall, the overlap in quantified proteins is quite high, with 1157 proteins quantified across all data sets. The sequencing based approach of *Li et al. (2014)* has substantially higher coverage compared to the mass spectrometry data sets (3394 genes versus the 2041 genes quantified in the work of *Schmidt et al. (2016)*). However, in terms of total protein mass, the data from *Li et al. (2014)*; *Schmidt et al. (2016)*; *Peebo et al. (2015)* each quantify roughly equivalent total protein mass. An exception to this is in the data from *Valgepea et al. (2013)*, where we find that the total protein quantified in *Valgepea et al. (2013)* is 90-95 % of the total protein mass (when using the data from *Schmidt et al. (2016)* as a reference).

### Estimation of Cell Size and Surface Area

Since most of the proteomic data sets lack cell size (i.e. volume) measurements, we chose instead to use a common estimate of size for any analysis requiring cell size or surface area. Since each of the data sets used either K-12 MG1655 or its derivative, BW25113 (from the lab of Barry L. Wanner; the parent strain of the Keio collection (*Datsenko and Wanner, 2000*; *Baba et al., 2006*)), we fit the MG1655 cell size data from the supplemental material of *Si et al. (2017, 2019)* using the `optimize.curve_fit` function from the Scipy python package (*Virtanen et al., 2020*).

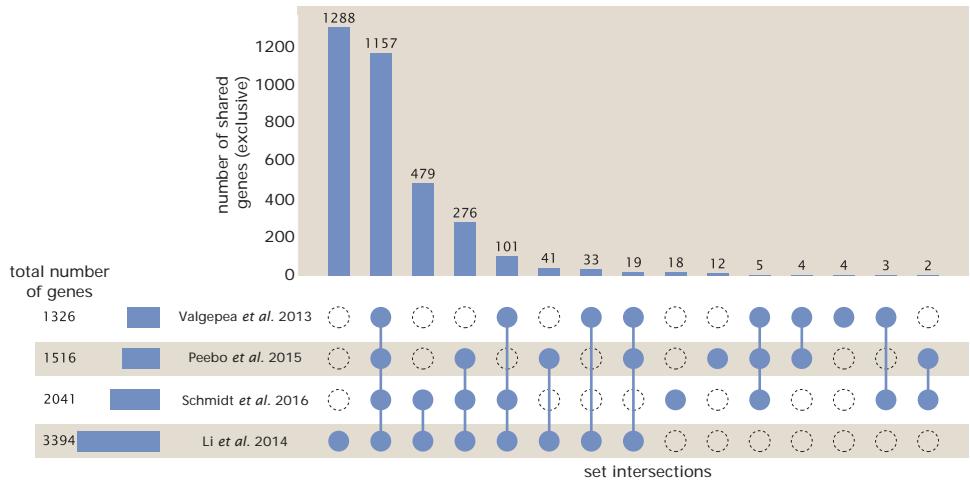
The average size measurements from each of their experiments are shown in Figure *Figure 14*, with cell length and width shown in (A) and (B), respectively. The length data was well described by the exponential function  $0.5 e^{1.09 \cdot \lambda} + 1.76 \mu\text{m}$ , while the width data was well described by  $0.64 e^{0.24 \cdot \lambda} \mu\text{m}$ . In order to estimate cell size we take the cell as a cylinders with two hemispherical ends (*Si et al., 2017*; *Basan et al., 2015*). Specifically, cell size is estimated from,

$$V = \pi \cdot r^2 \cdot (l - 2r/3), \quad (8)$$

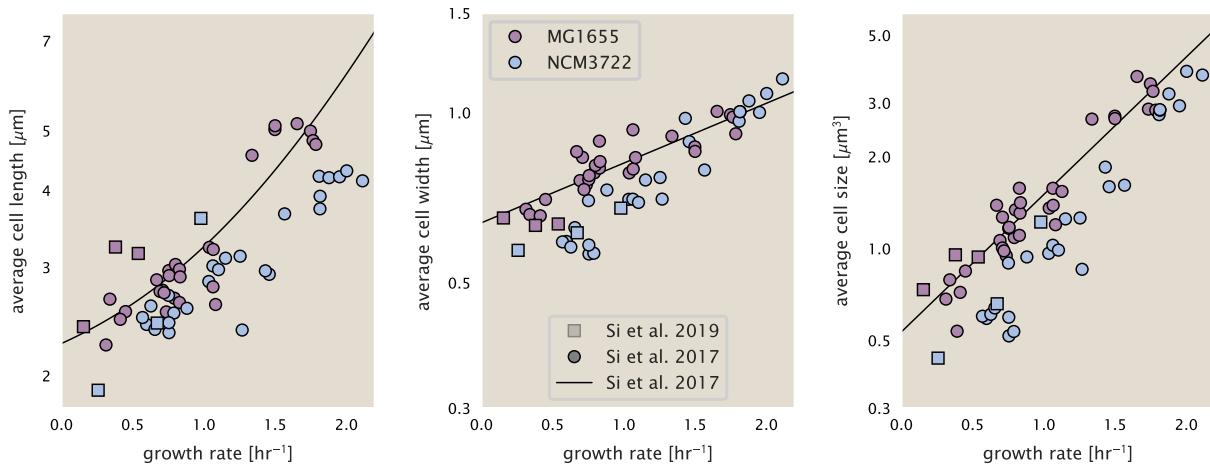
where  $r$  is half the cell width. A best fit to the data is described by  $0.533 e^{1.037 \cdot \lambda} \mu\text{m}^3$ . Calculation of the cell surface area is given by,

$$S = \eta \cdot \pi \left( \frac{\eta \cdot \pi}{4} - \frac{\pi}{12} \right)^{-2/3} V^{2/3}, \quad (9)$$

where  $\eta$  is the aspect ratio ( $\eta = l/w$ ) (*Ojkic et al., 2019*).



**Figure 13. Comparison of proteomic coverage across different data sets.** An UpSet diagram (*Lex et al., 2014*) summarizes the total number of protein coding genes whose protein abundance was reported in the data sets of *Valgepea et al. (2013); Li et al. (2014); Schmidt et al. (2016); Peebo et al. (2015)*. The total number of genes reported in each individual data set are noted on the bottom left, while their overlap is summarized in the bar chart. Each column here refers to the intersection of a set of data sets identified by blue circles.



**Figure 14. Summary of size measurements from Si et al. 2017, 2019.** Cell lengths and widths were measured from cell contours obtained from phase contrast images, and refer to the long and short axis respectively. (A) Cell lengths and (B) cell widths show the mean measurements reported (they report 140-300 images and 5,000-30,000 for each set of samples; which likely means about 1,000-5,000 measurements per mean value reported here since they considered about 6 conditions at a time). Fits were made to the MG1655 strain data; length:  $0.5 e^{1.09 \cdot \lambda} + 1.76 \mu\text{m}$ , width:  $0.64 e^{0.24 \cdot \lambda} \mu\text{m}$ . (C) Cell size,  $V$ , was calculated as cylinders with two hemispherical ends (Equation 8). The MG1655 strain data gave a best fit of  $0.533 e^{1.037 \cdot \lambda} \mu\text{m}^3$ .

## 852 Estimation of Total Protein Content per Cell

853 In order to estimate total protein per cell for a particular growth rate, we begin by estimating the cell size from  
854 the fit shown in Figure **Figure 14(C)** ( $0.533 e^{1.037 \cdot \lambda} \mu\text{m}^3$ ). We then estimate the total protein content from the total  
855 dry mass of the cell. Here we begin by noting that for almost the entire range of growth rates considered here,  
856 protein, DNA, and RNA were reported to account for at least 90 % of the dry mass (**Basan et al. (2015)**). The authors  
857 also found that the total dry mass concentration was roughly constant across growth conditions. Under such a  
858 scenario, we can calculate the total dry mass concentration for protein, DNA, and RNA, which is given by 1.1 g/ml  
859  $\times 30\% \times 90\%$  or about  $[M_p] = 300 \text{ fg per fl}$ . Multiplying this by our prediction of cell size gives the total dry mass  
860 per cell.

861 However, even if dry mass concentration is relatively constant across growth conditions, it is not obvious how  
862 protein concentration might vary due to the substantial increase in rRNA at faster growth rates (**Dai et al. (2016)**).  
863 This is a well-documented result that arises from an increase in ribosomal abundance at faster growth rates (**Scott**  
864 **et al. (2010)**). To proceed therefore rely on experimental measurements of total DNA content per cell that also  
865 come from Basan *et al.*, and RNA to protein ratios that were measured in Dai *et al.* (and cover the entire range of  
866 growth conditions considered here). These are reproduced in Figure **Figure 15(A)** and (B), respectively.

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

$$870 M_p = \frac{[M_p + M_{RNA}]}{1 + (RP_{ratio})}. \quad (10)$$

871 ( $RP_{ratio}$  refers to the RNA to protein ratio as measured by Dai *et al.*). In Figure **Figure 15(C)** we plot the estimated  
872 cellular concentrations for protein, DNA, and RNA from these calculations, and in Figure **Figure 15(D)** we plot their  
873 total expected mass per cell. This later quantity is the growth rate-dependent total protein mass that was used to  
874 estimate total protein abundance across all data sets (and summarized in **Figure 12(B)**).

## 874 Estimating Volume and Number of Amino Acids from Ribosome Copy Number

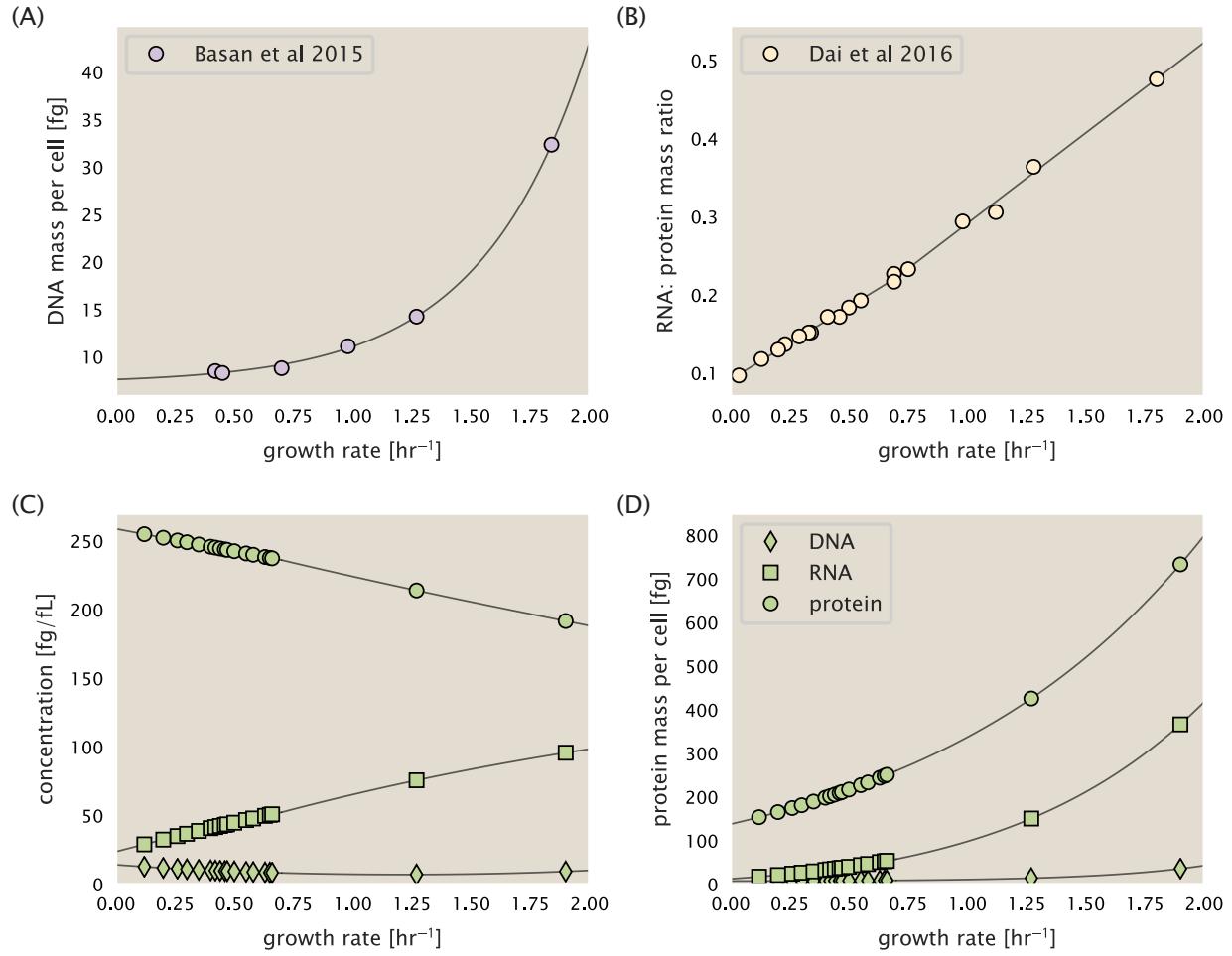
875 Towards the end of the main text, we examine a coarse-grained model of nutrient-limited growth. A key point  
876 in our analysis was to consider how elongation rate  $r_e$  and growth rate  $\lambda$  vary with respect to the experimentally  
877 observed changes in cell size, total number of peptide bonds per cell  $N_{pep}$ , and ribosomal content. In order to  
878 do maintain parameters in line with the experimental data, but otherwise allow us to explore the model, we  
879 performed a phenomenological fit of  $N_{pep}$  and  $V$  as a function of the measured ribosomal copy number  $R$ . As has  
880 been described in the preceding sections of this supplement, we estimate cell volume for each growth condition  
881 using the size measurements from **Si et al. (2017, 2019)**, and  $N_{pep}$  is approximated by taking the total protein mass  
882 and dividing this number by the average mass of an amino acid, 110 Da (BNID: 104877).

883 Given the exponential scaling of  $V$  and  $N_{pep}$  with growth rate, we performed a linear regression of the log trans-  
884 form of these parameters as a function of the log transform of the ribosome copy number. Using optimization  
885 by minimization, we estimated the best-fit values of the intercept and slope for each regression. **Figure 16** shows  
886 the result of each regression as a dashed line.

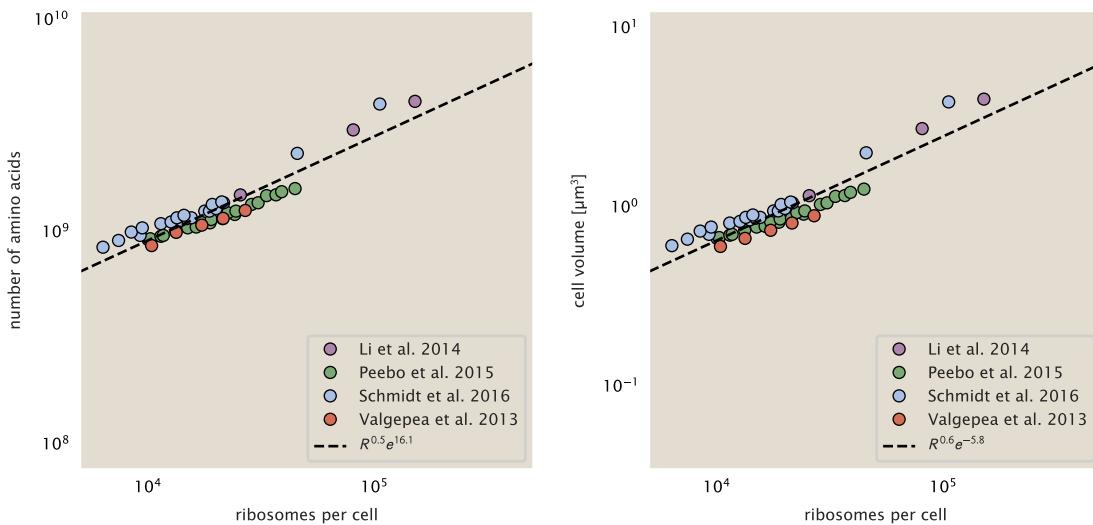
## 887 Additional Considerations of Schmidt *et al.* Data Set

888 While the data set from **Schmidt et al. (2016)** remains a heroic effort that our labs continue to return to as a  
889 resource, there were steps taken in their calculation of protein copy number that we felt needed further consider-  
890 ation. In particular, the authors made an assumption of constant cellular protein concentration across all growth  
891 conditions and used measurements of cell volume that appear inconsistent with an expected exponential scaling  
892 of cell size with growth rate that is well-documented in *E. coli* (**Schaechter et al. (1958); Taheri-Araghi et al. (2015);**  
893 **Si et al. (2017)**).

894 We begin by looking at their cell volume measurements, which are shown in blue in Figure **Figure 17**. As a  
895 comparison, we also plot cell sizes reported in three other recent papers: measurements from Taheri-Araghi *et*  
896 *al.* and Si *et al.* come from the lab of Suckjoon Jun, while those from Basan *et al.* come from the lab of Terence  
897 Hwa. Each set of measurements used microscopy and cell segmentation to determine the length and width, and



**Figure 15. Empirical estimate of cellular protein, DNA, and RNA as a function of growth rate.** (A) Measured DNA mass per cell as a function of growth rate, reproduced from Basan *et al.* 2015. The data was fit to an exponential curve (DNA mass in fg per cell is given by  $0.42 e^{2.23 \cdot \lambda} + 7.2$  fg per cell, where  $\lambda$  is the growth rate in hr<sup>-1</sup>). (B) RNA to protein measurements as a function of growth rate. The data was fit to two lines: for growth rates below 0.7 hr<sup>-1</sup>, the RNA/protein ratio is  $0.18 \cdot \lambda + 0.093$ , while for growth rates faster than 0.7 hr<sup>-1</sup> the RNA/protein ratio is given by  $0.25 \cdot \lambda + 0.035$ . For (A) and (B) cells are grown under varying levels of nutrient limitation, with cells grown in minimal media with different carbon sources for the slowest growth conditions, and rich-defined media for fast growth rates. (C) Predictions of cellular protein, DNA, and RNA concentration. (D) Total cellular mass predicted for protein, DNA, and RNA using the cell size predictions from Si *et al.*. Symbols (diamond: DNA, square: RNA, circle: protein) show estimated values of mass concentration and mass per cell for the specific growth rates in Schmidt *et al.* (2016).



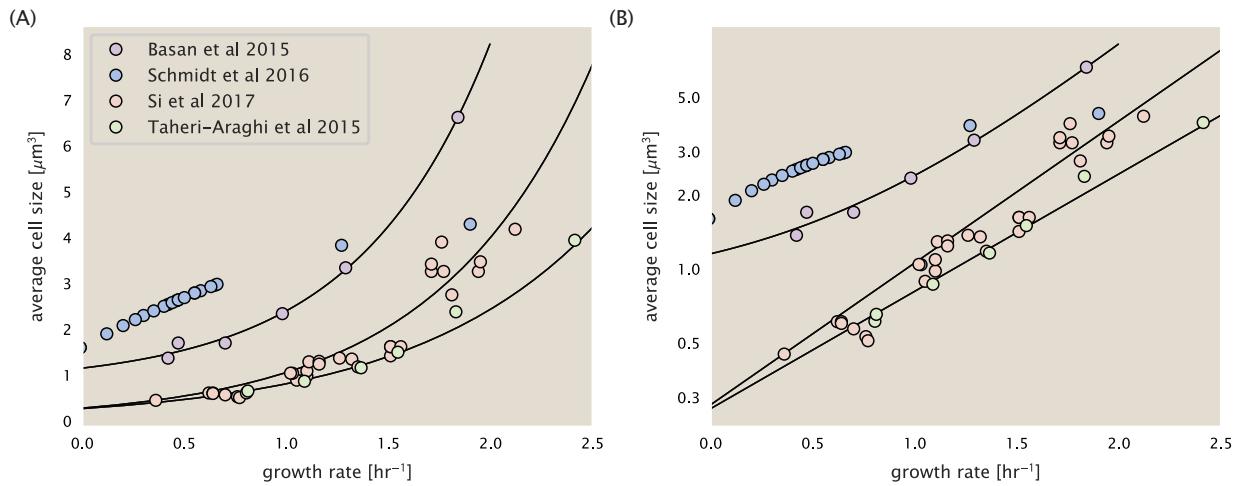
**Figure 16. Phenomenological regression of cell volume and number of amino acids per cell as a function of the ribosome copy number.** Colored points correspond to the measured value (or calculated value in the case of the cell volume) with colors denoting different data sets. The dashed black line shows the result of the fit, with the functional form of the equation given in the legend with  $R$  representing the ribosome copy number.

then calculated cell size by treating the cell as a cylinder with two hemispherical ends, as we considered in the previous section. While there is notable discrepancy between the two research groups, which are both using strain NCM3722, Basan *et al.* found that this came specifically from uncertainty in determining the cell width. This is prone to inaccuracy given the small cell size and optical resolution limits (further described in their supplemental text). Perhaps the more concerning point is that while each of these alternative measurements show an exponential increase in cell size at faster growth rates, the measurements used by Schmidt *et al.* appear to plateau. This resulted in an analogous trend in their final reported total cellular protein per cell as shown in Figure [Figure 18](#) (purple data points), and is in disagreement with other measurements of total protein at these growth rates ([Basan et al., 2015](#)).

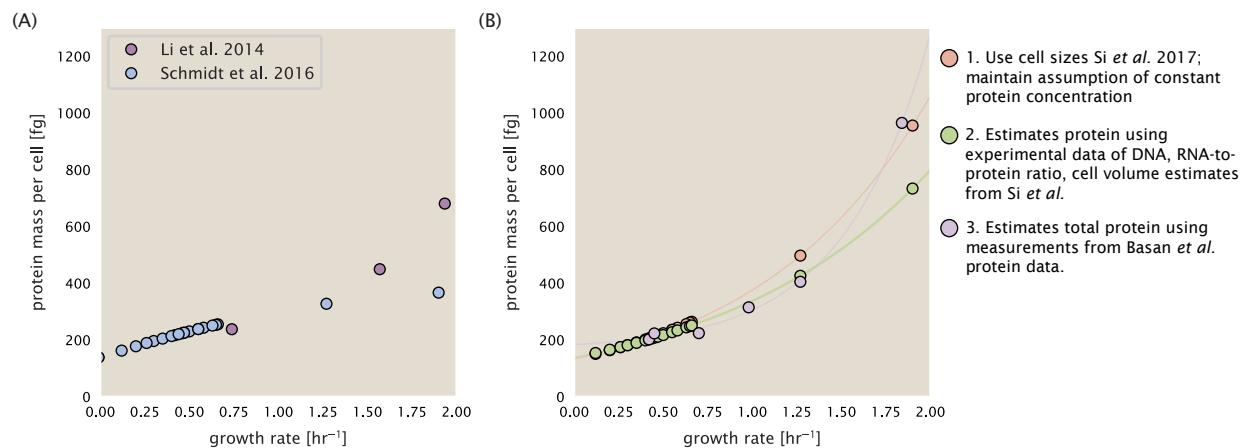
Since it is not obvious how measurements of cell size influenced their reported protein abundances, in the following subsections we begin by considering this calculation. We then consider three different approaches to estimate the growth-rate dependent total protein mass to compare with those values reported from [Schmidt et al. \(2016\)](#). The results of this are summarized in [Figure 17\(B\)](#), with the original values from both [Schmidt et al. \(2016\)](#) and [Li et al. \(2014\)](#) shown in [Figure 17\(A\)](#) for reference. For most growth conditions, we find that total protein per cell is still in reasonable agreement. However, for the fastest growth conditions, with glycerol + supplemented amino acids, and LB media, all estimates are substantially higher than those originally reported. This is the main reason why we chose to readjusted protein abundance as shown in [Figure 12\(B\)](#) (with the calculation described in section [Estimation of Total Protein Content per Cell](#)).

### Effect of cell volume on reported absolute protein abundances

As noted in section [Experimental Details Behind Proteomic Data](#), the authors calculated proteome-wide protein abundances by first determining absolute abundances of 41 pre-selected proteins, which relied on adding synthetic heavy reference peptides into their protein samples at known abundance. This absolute quantitation was performed in replicate for each growth condition. Separately, the authors also performed a more conventional mass spectrometry measurement for samples from each growth condition, which attempted to maximize the number of quantified proteins but only provided relative abundances based on peptide intensities. Finally, using their 41 proteins with absolute abundances already determined, they then created calibration curves with which to relate their relative intensity to absolute protein abundance for each growth condition. This allowed them to estimate absolute protein abundance for all proteins detected in their proteome-wide data set. Combined with



**Figure 17. Measurements of cell size as a function of growth rate.** (A) Plot of the reported cell sizes from several recent papers. The data in blue come from Volkmer and Heinemann, 2011 ([Volkmer and Heinemann \(2011\)](#)) and were used in the work of Schmidt *et al.*. Data from the lab of Terence Hwa are shown in purple ([Basan \*et al.\* \(2015\)](#)), while the two data sets shown in green and red come from the lab of Suckjoon Jun ([Taheri-Araghi \*et al.\* \(2015\); Si \*et al.\* \(2017\)](#)). (B) Same as in (A) but with the data plotted on a logarithmic y-axis to highlight the exponential scaling that is expected for *E. coli*.



**Figure 18. Alternative estimates of total cellular protein for the growth conditions considered in Schmidt *et al.*** (A) The original protein mass from Schmidt *et al.* and Li *et al.* are shown in purple and blue, respectively. (B) Three alternative estimates of total protein per cell. 1. *light red*: Rescaling of total protein mass assuming a growth rate independent protein concentration and cell volumes estimated from Si *et al.* 2017. 2. *light green*: Rescaling of total protein mass using estimates of growth rate-dependent protein concentrations and cell volumes estimated from Si *et al.* 2017. Total protein per cell is calculated by assuming a 1.1 g/ml cellular mass density, 30% dry mass, with 90% of the dry mass corresponding to DNA, RNA, and protein ([Basan \*et al.\*, 2015](#)). See [Estimation of Total Protein Content per Cell](#) for details on calculation. 3. *light purple*: Rescaling of total protein mass using the experimental measurements from Basan *et al.* 2015.

their flow cytometry cell counts, they were then able to determine absolute abundance of each protein detected on a per cell basis.

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

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

$$M_{P\_i} = [M_p]_{orig} \cdot V_i \quad (11)$$

where  $M_{P_i}$  represents the total protein mass per cell and  $V_i$  is the cell volume for each growth condition  $i$  as measured in Volkmer and Heinemann, 2011. Here the thinking is that the values of  $M_{P_i}$  reflects the total cellular protein for growth condition  $i$ , where any discrepancy from their absolute protein abundance is assumed to be due to protein loss during sample preparation. The protein abundances from their absolute abundance measurements noted above were therefore scaled to their estimates and are shown in Figure **Figure 18** (purple data points).

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

$$M'_{P\_i} = [M_p]_{Si} \cdot V_{Si\_i} \quad (12)$$

where  $M'_{P_i}$  is the new protein mass prediction, and  $V_{Si_i}$  refers to the new volume prediction for each condition  $i$ . These are shown as red data points in Figure **Figure 18(B)**.

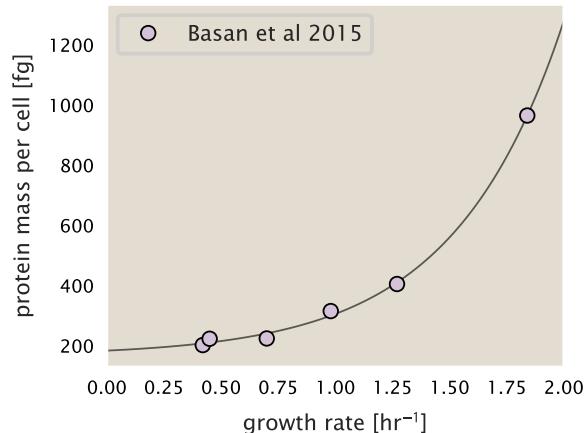
### Relaxing assumption of constant protein concentration across growth conditions

We next relax the assumption that cellular protein concentration is constant and instead, attempt to estimate it using experimental data. Here we use the estimation of total protein mass per cell detailed in section Estimation of Total Protein Content per Cell for all data points in the *Schmidt et al. (2016)* data set. The green data points in **Figure 18(B)** show this prediction, and this represents the approach used to estimate total protein per cell for all data sets.

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

One of the challenges in our estimates in the preceding sections is the need to estimate protein concentration and cell volumes. These are inherently difficult to accurately due to the small size of *E. coli*. Indeed, for all the additional measurements of cell volume included in Figure **Figure 17**, no measurements were performed for cells growing at rates below  $0.5\text{ hr}^{-1}$ . It therefore remains to be determined whether our extrapolated cell volume estimates are appropriate, with the possibility that the logarithmic scaling of cell size might break down for slower growth.

In our last approach we therefore attempt to estimate total protein using experimental data that required no estimates of concentration or cell volume. Specifically, in the work of Basan *et al.*, the authors measured total protein per cell for a broad range of growth rates (reproduced in Figure **Figure 19**). These were determined by first measuring bulk protein from cell lysate, measured by the colorimetric Biuret method (*You et al. (2013)*), and then abundance per cell was calculated from cell counts from either plating cells or a Coulter counter. While it is unclear why Schmidt *et al.* was unable to take a similar approach, the results from Basan *et al* appear more



**Figure 19. Total cellular protein reported in Basan et al. 2015.** Measured protein mass as a function of growth rate as reproduced from Basan *et al.* 2015, with cells grown under different levels of nutrient limitation. The data was fit to an exponential curve where protein mass in fg per cell is given by  $14.65 e^{2.180 \cdot \lambda} + 172$  fg per cell, where  $\lambda$  is the growth rate in  $hr^{-1}$ .

consistent with our expectation that cell mass will increase exponentially with faster growth rates. In addition, although they do not consider growth rates below about  $0.5\ hr^{-1}$ , it is interesting to note that the protein mass per cell appears to plateau to a minimum value at slow growth. In contrast, our estimates using cell volume so far have predicted that total protein mass should continue to decrease slightly for slower growing cells. By fitting this data to an exponential function dependent on growth rate, we could then estimate the total protein per cell for each growth condition considered by Schmidt *et al.* (2016). These are plotted as red data points in Figure 18(B).

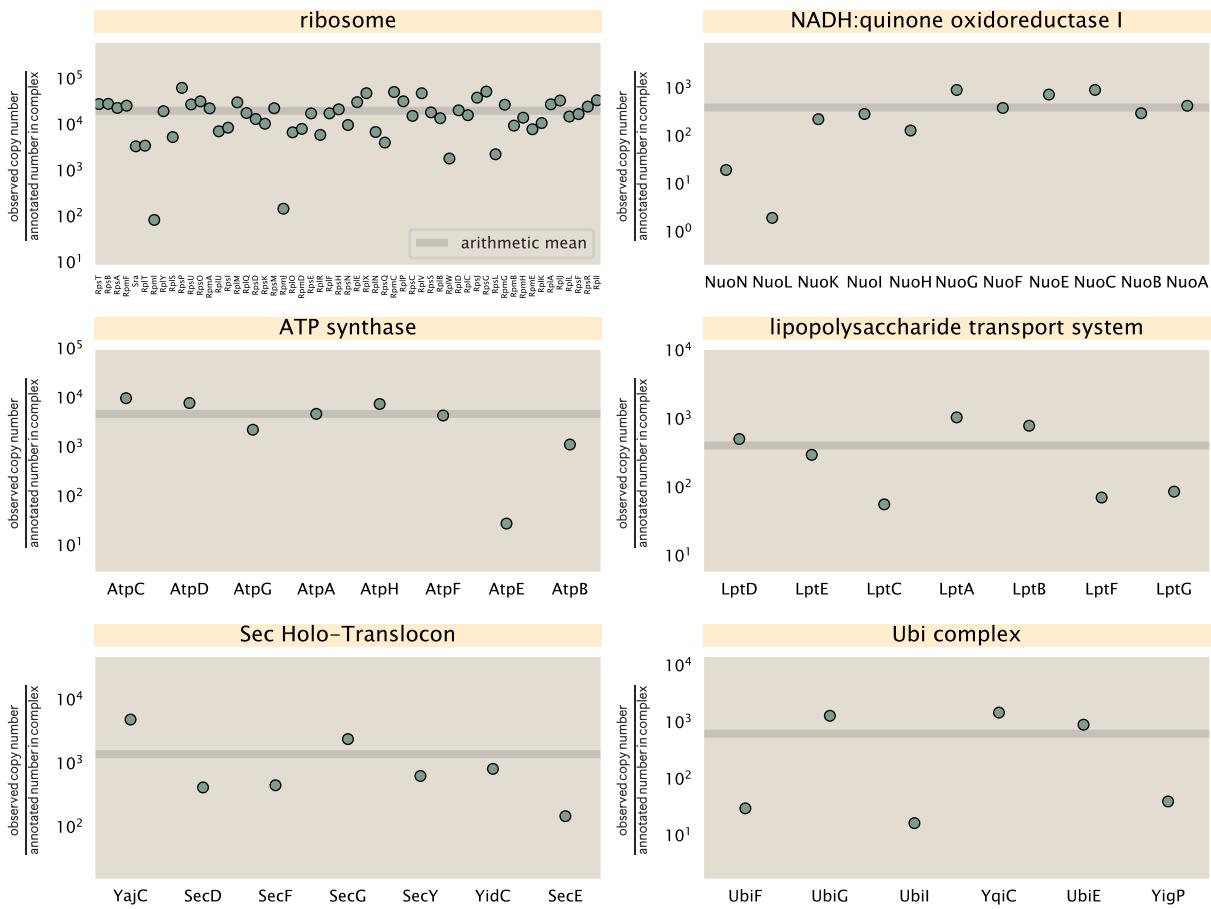
### Calculation of Complex Abundance

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

Protein complexes, and proteins individually, often have a variety of names, both longform and shorthand. As individual proteins can have a variety of different synonyms, we sought to ensure that each protein annotated in the data sets used the same synonym. To do use, we relied heavily on the EcoCyc Genomic Database (Keseler *et al.*, 2017). Each protein in available data sets included an annotation of one of the gene name synonyms as well as an accession ID – either a UniProt or Blattner “b-number”. We programmatically matched up individual accession IDs between the proteins in different data sets. In cases where accession IDs matched but the gene names were different, we manually verified that the gene product was the same between the datasets and chose a single synonym. All code used in the data cleaning and unification procedures can be found on the associated GitHub repository (DOI:XXX) associated with this paper as well as on the associated paper website.

With each protein conforming to a single identification scheme, we then needed to identify the molecular complexes each protein was a member of. Additionally, we needed to identify how many copies of each protein were present in each complex (i.e. the subunit copy number) and compute the estimated abundance complex that accounted for fluctuations in subunit stoichiometry. To map proteins to complexes, we accessed the EcoCyc *E. coli* database Keseler *et al.* (2017) using PathwayTools version 23.0 Karp *et al.* (2019). With a license for PathWay Tools, we mapped each unique protein to its annotated complexes via the BioCyc Python package. As we mapped each protein with *all* of its complex annotations, there was redundancy in the dataset. For example, ribosomal protein L20 (RplT) is annotated to be a component of the 50S ribosome (EcoCyc complex CPLX-03962) as well as a component of the mature 70S ribosome (EcoCyc complex CPLX-03964).

In addition to the annotated complex, we collected information on the stoichiometry of each macromolecular complex. For a complex with  $N_{subunits}$  protein species, for each protein subunit  $i$  we first calculate the number of



**Figure 20. Calculation of the mean complex abundance from measurements of single subunits.** Six of the largest complexes (by number of subunits) in *E. coli*. Points correspond to the maximum number of complexes that can be formed given measurement of that individual protein. Solid grey line corresponds to the arithmetic mean across all subunits. These data correspond to measurements from Schmidt et al. (2016) in a glucose-supplemented minimal growth medium.

1001 complexes that could be formed given the measured protein copy numbers per cell,

$$N_{\text{complex}}(\text{subunit } i) = \frac{P_{\text{subunit } i}^{(\text{measured})}}{m_{\text{subunit } i}}. \quad (13)$$

1002 Here,  $P_{\text{subunit } i}^{(\text{measured})}$  refers to the measured protein copy number of species  $i$ , and  $m$  refers to the number of monomers  
 1003 present for that protein in the complex. For example, the 70S mature ribosome complex has 55 protein components,  
 1004 all of which are present in a single copy except L4 (RplL), which is present in 4 copies ( $m = 4$ ). For each  
 1005 ribosomal protein, we then calculate the maximum number of complexes that could be formed using Equation 13.  
 1006 This example, along with example from 5 other macromolecular complexes, can be seen in Figure 20.

1007 It is important to note that measurement noise, efficiency of protein extraction, and differences in protein  
 1008 stability will mean that the precise value of each calculation will be different for each component of a given complex.  
 1009 Thus, to report the total complex abundance, we use the arithmetic mean of across all subunits in the complex,

$$\langle N_{\text{complex}} \rangle = \frac{1}{N_{\text{subunits}}} \sum_i^{N_{\text{subunits}}} \frac{P_i^{(\text{measured})}}{m_{\text{subunit } i}}. \quad (14)$$

1010 in Figure 20, we show this mean value as a grey line for a variety of different complexes. Additionally, we have  
 1011 built an interactive figure accessible on the [paper website](#) where the validity of this approach can be examined  
 1012 for any complex with more than two subunits (thus, excluding monomers and dimers).

1013 **Extending Estimates to a Continuum of Growth Rates**

1014 In the main text, we considered a standard stopwatch of 5000 s to estimate the abundance of the various protein  
1015 complexes considered. In addition to point estimates, we also showed the estimate as a function of growth rate  
1016 as transparent grey curves. In this section, we elaborate on this continuum estimate, giving examples of estimates  
1017 that scale with either cell volume, cell surface area, or number of origins of replication.

1018 **Estimation of the total cell mass**

1019 For many of the processes estimated in the main text we relied on a cellular dry mass of  $\approx 300$  fg from which we  
1020 computed elemental and protein fractions using knowledge of fractional composition of the dry mass. At modest  
1021 growth rates, such as the 5000 s doubling time used in the main text, this is a reasonable number to use as the  
1022 typical cell mass is  $\approx 1$  pg and *E. coli* cells can approximated as 70% water by volume. However, as we have shown  
1023 in the preceding sections, the cell size is highly dependent on the growth rate. This means that a dry mass of 300  
1024 fg cannot be used reliably across all growth rates.

1025 Rather, using the phenomenological description of cell volume scaling exponentially with growth rate, and  
1026 using a rule-of-thumb of a cell buoyant density of  $\approx 1.1$  pg / fL (BNID: 103875), we can calculate the cell dry mass  
1027 across a range of physiological growth rates as

$$m_{\text{cell}} \approx \rho V(\lambda) \approx \rho a e^{\lambda * b} \quad (15)$$

1028 where  $a$  and  $b$  are constants with units of  $\mu\text{m}^3$  and hr, respectively. The value of these constants can be estimated  
1029 from the careful volume measurements performed by *Si et al. (2017,?)*, as considered in Appendix Estimation of  
1030 Cell Size and Surface Area earlier.

1031 **Complex Abundance Scaling With Cell Volume**

1032 Several of the estimates performed in the main text are implicitly dependent on the cell volume. This includes  
1033 processes such as ATP utilization and, most prominently, the transport of nutrients, whose demand will be pro-  
1034 portional to the volume of the cell. Of the latter, we estimated the number of transporters that would be needed  
1035 to shuttle enough carbon, phosphorus, and sulfur across the membrane to build new cell mass. To do so, we  
1036 used elemental composition measurements combined with a 300 fg cell dry mass to make the point estimate. As  
1037 we now have a means to estimate the total cell mass as a function of volume, we can generalize these estimates  
1038 across growth rates.

1039 Rather than discussing the particular details of each transport system, we will derive this scaling expression in  
1040 very general terms. Consider that we wish to estimate the number of transporters for some substance  $X$ , which  
1041 has been measured to be made up some fraction of the dry mass,  $\theta_X$ . If we assume that, irrespective of growth  
1042 rate, the cell dry mass is relatively constant (*Basan et al., 2015*) and  $\approx 30\%$  of the total cell mass, we can state that  
1043 the total mass of substance  $X$  as a function of growth rate is

$$m_X \approx 0.3 \times \rho V(\lambda) \theta_X, \quad (16)$$

1044 where we have used  $\rho V(\lambda)$  as an estimate of the total cell mass, defined in *Equation 15*. To convert this to the  
1045 number of units  $N_X$  of substance  $X$  in the cell, we can use the formula weight  $w_X$  of a single unit of  $X$  in conjunction  
1046 with *Equation 16*,

$$N_X \approx \frac{m_X}{w_X}. \quad (17)$$

1047 To estimate the number of transporters needed, we make the approximation that loss of units of  $X$  via diffusion  
1048 through porins or due to the permeability of the membrane is negligible and that a single transporter complex  
1049 can transport substance  $X$  at a rate  $r_X$ . As this rate  $r_X$  is in units of  $X$  per time per transporter, we must provide  
1050 a time window over which the transport process can occur. This is related to the cell doubling time  $\tau$ , which can  
1051 be calculated from the the growth rate  $\lambda$  as  $\tau = \log(2)/\lambda$ . Putting everything together, we arrive at a generalized  
1052 transport scaling relation of

$$N_{\text{transporters}}(\lambda) = \frac{0.3 \times \rho V(\lambda) \theta_X}{w_X r_X \tau}. \quad (18)$$

1053 This function is used to draw the continuum estimates for the number of transporters seen in Figures 2 and  
 1054 3 as transparent grey curves. Occasionally, this continuum scaling relationship will not precisely agree with the  
 1055 point estimate outlined in the main text. This is due to the choice of  $\approx 300$  fg total dry mass per cell for the point  
 1056 estimate, whereas we considered more precise values of cell mass in the continuum estimate. We note, however,  
 1057 that both this scaling relation and the point estimates are meant to describe the order-of-magnitude observed,  
 1058 and not the predict the exact values of the abundances.

1059 **Equation 18** is a very general relation for processes where the cell volume is the "natural variable" of the  
 1060 problem. This means that, as the cell increases in volume, the requirements for substance  $X$  also scale with  
 1061 volume rather than scaling with surface area, for example. So long as the rate of the process, the fraction of the  
 1062 dry mass attributable to the substance, and the formula mass of the substance is known, **Equation 18** can be used  
 1063 to compute the number of complexes needed. For example, to compute the number of ATP synthases per cell,  
 1064 **Equation 18** can be slightly modified to the form

$$N_{\text{ATP synthases}}(\lambda) = \frac{0.3 \times \rho V(\lambda) \theta_{\text{protein}} N_{\text{ATP}}}{w_{AA} r_{\text{ATP}} \tau}, \quad (19)$$

1065 where we have included the term  $N_{\text{ATP}}$  to account for the number of ATP equivalents needed per amino acid for  
 1066 translation ( $\approx 4$ , BNID: 114971), and  $w_{AA}$  is the average mass of an amino acid. The grey curves in Figure 4 o the  
 1067 main text were made using this type of expression.

## 1068 A Relation for Complex Abundance Scaling With Surface Area

1069 In our estimation for the number of complexes needed for lipid synthesis and peptidoglycan maturation, we used  
 1070 a particular estimate for the cell surface area ( $\approx 5 \mu\text{m}$ , BNID: 101792) and the fraction of dry mass attributable to  
 1071 peptidoglycan ( $\approx 3\%$ , BNID: 101936). Both of these values come from glucose-fed *E. coli* in balance growth. As we  
 1072 are interested in describing the scaling as a function of the growth rate, we must also consider how these values  
 1073 scale with cell surface area, which is the natural variable for these types of processes. In the coming paragraphs,  
 1074 we highlight how we incorporate a condition-dependent surface area into our calculation of the number of lipids  
 1075 and murein monomers that need to be synthesized and crosslinked, respectively.

### 1076 Number of Lipids

1077 To compute the number of lipids as a function of growth rate, we make the assumption that some features, such as  
 1078 the surface area of a single lipid ( $A_{\text{lipid}} \approx 0.5 \text{ nm}^2$ , BNID: 106993) and the total fraction of the membrane composed  
 1079 of lipids ( $\approx 40\%$ , BNID: 100078) are independent of the growth rate. Using these approximations combined with  
 1080 **Equation 9**, and recognizing that each membrane is composed of two leaflets, we can compute the number of  
 1081 lipids as a function of growth rate as

$$N_{\text{lipids}}(\lambda) \approx \frac{4 \text{ leaflets} \times 0.4 \times \eta \pi \left( \frac{\eta \pi}{4} - \frac{\pi}{12} \right)^{-2/3} V(\lambda)^{2/3}}{A_{\text{lipid}}} \quad (20)$$

1082 where  $\eta$  is the length-to-width aspect ratio and  $V$  is the cell volume.

### 1083 Number of Murein Monomers

1084 In calculation of the number of transpeptidases needed for maturation of the peptidoglycan, we used an empirical  
 1085 measurement that  $\approx 3\%$  of the dry mass is attributable to peptidoglycan and that a single murien monomer is  
 1086  $m_{\text{murein}} \approx 1000 \text{ Da}$ . While the latter is independent of growth rate, the former is not. As the peptidoglycan exists as  
 1087 a thin shell with a width of  $w \approx 10 \text{ nm}$  encapsulating the cell, one would expect the number of murein monomers  
 1088 scales with the surface area of this shell. In a similar spirit to our calculation of the number of lipids, the total  
 1089 number of murein monomers as a function of growth rate can be calculated as

$$N_{\text{murein monomers}}(\lambda) \approx \frac{\rho_{\text{pg}} w \eta \pi \left( \frac{\eta \pi}{4} - \frac{\pi}{12} \right)^{-2/3} V(\lambda)^{2/3}}{m_{\text{murein}}}, \quad (21)$$

1090 where  $\rho_{\text{pg}}$  is the density of peptidoglycan.

**1091 Complex Abundance Scaling With Number of Origins, and rRNA Synthesis**

**1092** While the majority of our estimates hinge on the total cell volume or surface area, processes related to the central  
**1093** dogma, namely DNA replication and synthesis of rRNA, depend on the number of chromosomes present in the  
**1094** cell. As discussed in the main text, the ability of *E. coli* to parallelize the replication of its chromosome by having  
**1095** multiple active origins of replication is critical to synthesize enough rRNA, especially at fast growth rates. Derived  
**1096** in *Si et al. (2017)* and reproduced in the main text and Appendix Estimation of  $\langle \#ori \rangle / \langle \#ter \rangle$  and  $\langle \#ori \rangle$  below, the  
**1097** average number of origins of replication at a given growth rate can be calculated as

$$\langle \#ori \rangle \approx 2^{t_{\text{cyc}} \lambda / \ln 2} \quad (22)$$

**1098** where  $t_{\text{cyc}}$  is the total time of replication and division. We can make the approximation that  $t_{\text{cyc}} \approx 70$  min, which is  
**1099** the time it takes two replisomes to copy an entire chromosome.

**1100** In the case of rRNA synthesis, the majority of the rRNA operons are surrounding the origin of replication. Thus,  
**1101** at a given growth rate  $\lambda$ , the average dosage of rRNA operons per cell  $D_{\text{rRNA}}$  is

$$D_{\text{rRNA}}(\lambda) \approx N_{\text{rRNA operons}} \times 2^{t_{\text{cyc}} \lambda / \ln 2}. \quad (23)$$

**1102** This makes the approximation that *all* rRNA operons are localized around the origin. In reality, the operons  
**1103** are some distance away from the origin, making **Equation 23** an approximation (*Dennis et al., 2004*).

**1104** In the main text, we stated that at a growth rate of  $0.5 \text{ hr}^{-1}$ , there is  $\approx 1$  chromosome per cell. While a fair  
**1105** approximation, **Equation 22** illustrates that is not precisely true, even at slow growth rates. In estimating the  
**1106** number of RNA polymerases as a function of growth rate, we consider that regardless of the number of rRNA  
**1107** operons, they are all sufficiently loaded with RNA polymerase such that each operon produces one rRNA per  
**1108** second. Thus, the total number of RNA polymerase as a function of the growth rate can be calculated as

$$N_{\text{RNA polymerase}}(\lambda) \approx L_{\text{operon}} D_{\text{rRNA}} \rho_{\text{RNA polymerase}}, \quad (24)$$

**1109** where  $L_{\text{operon}}$  is the total length of an rRNA operon ( $\approx 4500$  bp) and  $\rho_{\text{RNA polymerase}}$  is packing density of RNA poly-  
**1110** merase on a given operon, taken to be 1 RNA polymerase per 80 nucleotides.

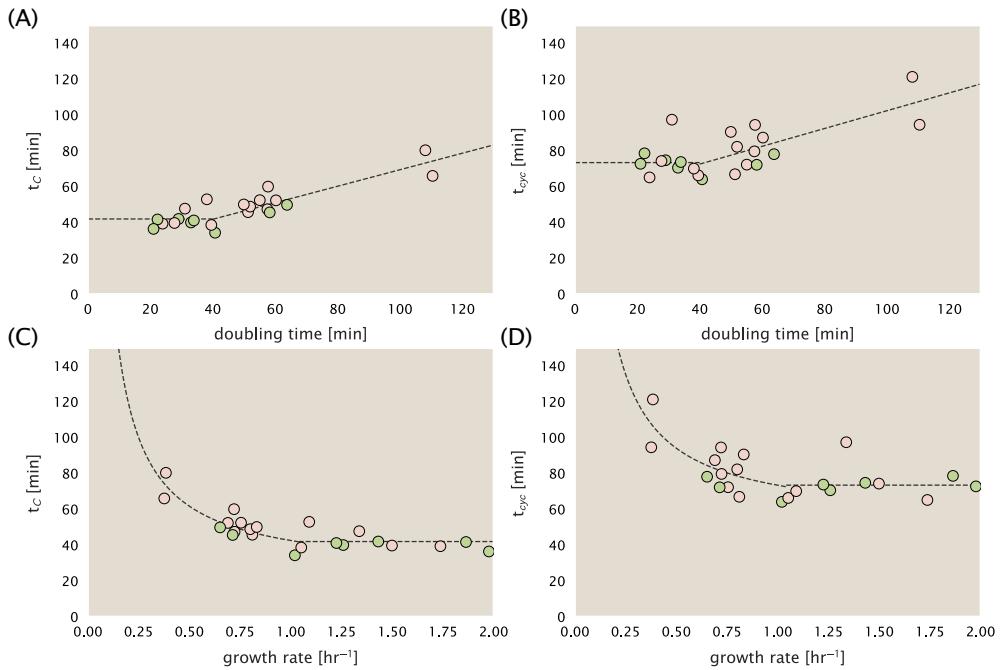
**1111 Calculation of active ribosomal fraction.**

**1112** In the main text we used the active ribosomal fraction  $f_a$  that was reported in the work of *Dai et al. (2016)* to  
**1113** estimate the active ribosomal mass fraction  $\Phi_R \times f_a$  across growth conditions. We lacked any specific model to  
**1114** consider how  $f_a$  should vary with growth rate, and instead find that the data is well-approximated by fitting to an  
**1115** exponential curve ( $f_a = -0.889 e^{4.6 \cdot \lambda} + 0.922$ ; dashed line in inset of **Figure 9(C)**). We use this function to estimate  $f_a$   
**1116** for each of the data points shown in **Figure 9(C)**.

**1117 Estimation of  $\langle \#ori \rangle / \langle \#ter \rangle$  and  $\langle \#ori \rangle$ .**

**1118** *E. coli* shows robust scaling of cell size with the average the number of origins  $\langle \#ori \rangle$  per cell (*Si et al., 2017*). Since  
**1119** protein makes up a majority of the cell's dry mass, the change in cell size is also a reflection of the changes in  
**1120** proteomic composition and total abundance across growth conditions. Given the potential constraints on rRNA  
**1121** synthesis and changes in ribosomal copy number with  $\langle \#ori \rangle$ , it becomes important to also consider how pro-  
**1122** tein copy numbers vary with the state of chromosomal replication. This is particularly true when trying to make  
**1123** sense of the changes in ribosomal fraction and growth-rate dependent changes in proteomic composition at a  
**1124** mechanistic level. As considered in the main text, it is becoming increasingly apparent that regulation through  
**1125** the secondary messengers (p)ppGpp may act to limit DNA replication and also reduce ribosomal activity in poorer  
**1126** nutrient conditions. In this context, both  $\langle \#ori \rangle$ , as well as the  $\langle \#ori \rangle / \langle \#ter \rangle$  ratio become important parameters  
**1127** to consider and keep tract of. An increase in  $\langle \#ori \rangle / \langle \#ter \rangle$  ratio in particular, causes a relatively higher gene  
**1128** dosage in rRNA and r-protein genes due to skew in genes near the origin, where the majority of these are located

**1129** In the main text we estimated the change in  $\langle \#ori \rangle$  with growth rate using the nutrient-limited wild-type cell  
**1130** data from *Si et al. (2017)*. We consider their measurements of DNA replication time ( $t_C$ , 'C' period of cell division),  
**1131** total cell cycle time ( $t_{\text{cyc}}$ , 'C' + 'D' period of cell division), and doubling time  $\tau$  from wild-type *E. coli* growing across  
**1132** a range of growth conditions. Here we show how we estimate this parameter, as well as the  $\langle \#ori \rangle / \langle \#ter \rangle$  ratio



**Figure 21. Estimation of  $\langle \#ori \rangle / \langle \# ter \rangle$  and  $\langle \#ori \rangle$  using data from Si et al. (2017).** (A) and (B) plot the reported  $t_C$  and  $t_{cyc}$  as a function of cell doubling time  $\tau$ , respectively. The dashed lines show a piecewise fit to the data. For short doubling times (rich media),  $t_C$  and  $t_{cyc}$  are assumed constant ( $t_C = 42$  minutes,  $t_{cyc} = 73$  minutes). At the transition, taken to occur at 40 minutes, the dashed line corresponds to an assumed proportional increase in each parameter as a function of the doubling time ( $t_C = 0.46\tau + 23.3$  minutes,  $t_{cyc} = 0.50\tau + 52.7$  minutes). (C) and (D) plot the same data as in (A) and (B), but as a function of growth rate, given by  $\lambda = \ln(2)/\tau$ .

from their data. We begin by considering  $\langle \#ori \rangle$ . If the cell cycle time takes longer than the time of cell division, the cell will need to initiate DNA replication more often than its rate of division,  $2^{\lambda t} = 2^{\ln(2)\cdot t/\tau}$  to maintain steady-state growth. Cells will need to do this in proportion to the ratio  $\lambda_{cyc}/\lambda = t_{cyc}/\tau$ , and the number of origins per cell (on average) is then given by  $2^{t_{cyc}/\tau}$ . The average number of termini will in contrast depend on the lag time between DNA replication and cell division,  $t_D$ , with  $\langle \#ori \rangle / \langle \# ter \rangle$  ratio =  $2^{t_{cyc}/\tau - t_D/\tau} = 2^{t_C/\tau}$ .

In Figure 21(A) and (B) we plot the measured  $t_C$  and  $t_{cyc}$  values versus the doubling time from Si et al. (2017). The authors estimated  $t_C$  by marker frequency analysis using qPCR, while  $t_{cyc} = t_C + t_D$  were inferred from  $t_C$  and  $\tau$ . In the plots we see that both  $t_C$  and  $t_{cyc}$  reach a minimum at around 40 and 75 minutes, respectively. For a C period of 40 minutes, this would correspond to a maximum rate of elongation of about 1,000 bp/sec. Since we lacked a specific model to describe how each of these parameters vary with growth condition, we assumed that they were linearly dependent on the doubling time. For each parameter,  $t_C$  and  $t_{cyc}$ , we split them up into two domains corresponding to poorer nutrient conditions and rich nutrient conditions (cut off at  $\tau \approx 40$  minutes where chromosomal replication becomes nearly constant). The fit lines are shown as solid black lines. In Figure 21(C) and (D) we also show  $t_C$  and  $t_{cyc}$  as a function of growth rate  $\lambda$  along with our piecewise linear fits, which match the plots in the main text.

1148 **References**

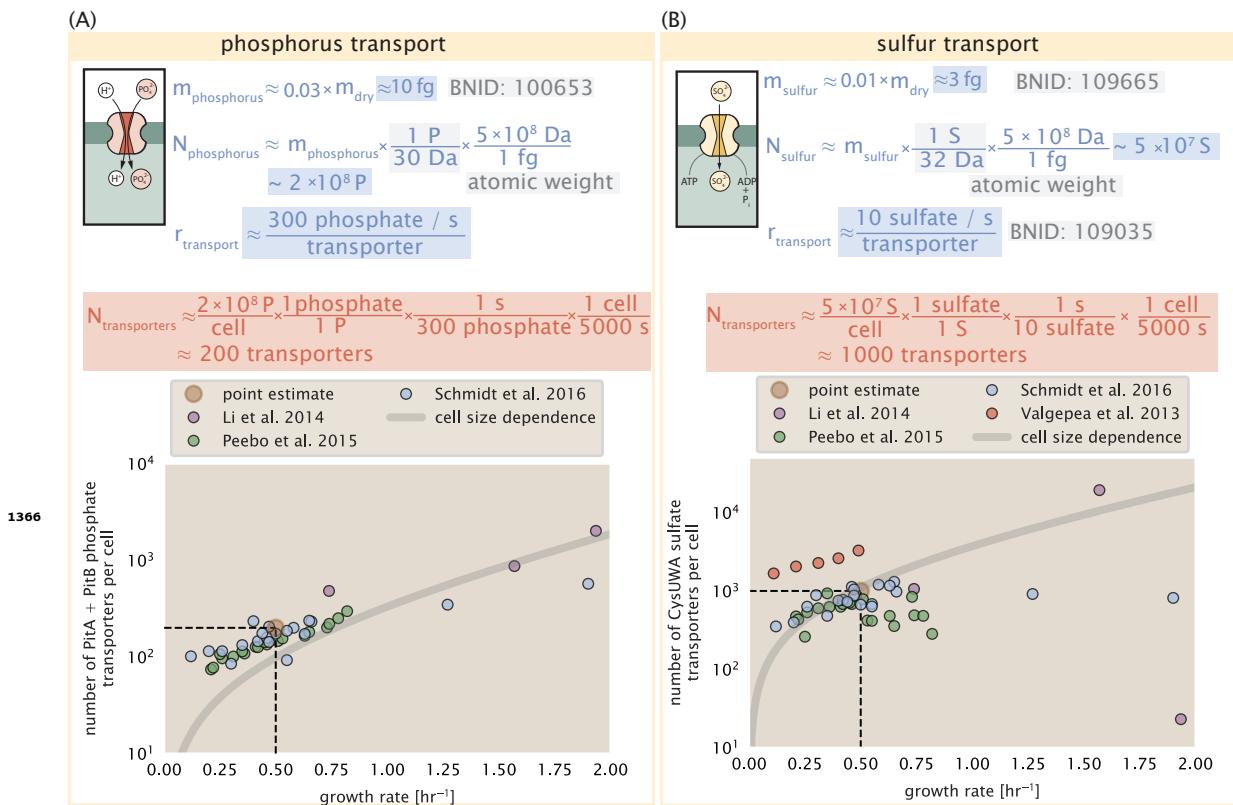
- 1149 Aidelberg, G., Towbin, B. D., Rothschild, D., Dekel, E., Bren, A., and Alon, U. (2014). Hierarchy of non-glucose sugars in *Escherichia*  
1150 *coli*. *BMC Systems Biology*, 8(1):133.
- 1151 Antonenko, Y. N., Pohl, P., and Denisov, G. A. (1997). Permeation of ammonia across bilayer lipid membranes studied by ammo-  
1152 nium ion selective microelectrodes. *Biophysical Journal*, 72(5):2187–2195.
- 1153 Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris,  
1154 M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock,  
1155 G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29.
- 1156 Ason, B., Bertram, J. G., Hingorani, M. M., Beechem, J. M., O'Donnell, M., Goodman, M. F., and Bloom, L. B. (2000). A Model  
1157 for *Escherichia coli* DNA Polymerase III Holoenzyme Assembly at Primer/Template Ends: DNA Triggers A Change In Binding  
1158 Specificity of the  $\gamma$  Complex Clamp Loader. *Journal of Biological Chemistry*, 275(4):3006–3015.
- 1159 Assentoft, M., Kaptan, S., Schneider, H.-P., Deitmer, J. W., de Groot, B. L., and MacAulay, N. (2016). Aquaporin 4 as a NH<sub>3</sub> Channel.  
1160 *Journal of Biological Chemistry*, 291(36):19184–19195.
- 1161 Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006).  
1162 Construction of *Escherichia coli*K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*,  
1163 2(1):2460.
- 1164 Basan, M., Zhu, M., Dai, X., Warren, M., Sévin, D., Wang, Y.-P., and Hwa, T. (2015). Inflating bacterial cells by increased protein  
1165 synthesis. *Molecular Systems Biology*, 11(10):836.
- 1166 Bauer, S. and Ziv, E. (1976). Dense growth of aerobic bacteria in a bench-scale fermentor. *Biotechnology and Bioengineering*,  
1167 18(1):81–94. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/bit.260180107>.
- 1168 Belliveau, N. M., Barnes, S. L., Ireland, W. T., Jones, D. L., Sweredoski, M. J., Moradian, A., Hess, S., Kinney, J. B., and Phillips, R.  
1169 (2018). Systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria. *Proceedings  
1170 of the National Academy of Sciences*, 115(21):E4796–E4805.
- 1171 Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009). Absolute metabolite concentra-  
1172 tions and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chemical Biology*, 5(8):593–599.
- 1173 Bremer, H. and Dennis, P. P. (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Expon-  
1174 ential Growth Rates. *EcoSal Plus*, 3(1).
- 1175 Browning, D. F. and Busby, S. J. W. (2016). Local and global regulation of transcription initiation in bacteria. *Nature Reviews  
1176 Microbiology*, 14(10):638–650.
- 1177 Büke, F., Grilli, J., Lagomarsino, M. C., Bokinsky, G., and Tans, S. (2020). ppGpp is a bacterial cell size regulator. *bioRxiv*,  
1178 266:2020.06.16.154187.
- 1179 Catherwood, A. C., Lloyd, A. J., Tod, J. A., Chauhan, S., Slade, S. E., Walkowiak, G. P., Galley, N. F., Punekar, A. S., Smart, K., Rea, D.,  
1180 Evans, N. D., Chappell, M. J., Roper, D. I., and Dowson, C. G. (2020). Substrate and Stereochemical Control of Peptidoglycan  
1181 Cross-Linking by Transpeptidation by *Escherichia coli* PBP1B. *Journal of the American Chemical Society*, 142(11):5034–5048.
- 1182 Cox, G. B., Newton, N. A., Gibson, F., Snoswell, A. M., and Hamilton, J. A. (1970). The function of ubiquinone in *Escherichia coli*.  
1183 *Biochemical Journal*, 117(3):551–562.
- 1184 Dai, X., Zhu, M., Warren, M., Balakrishnan, R., Okano, H., Williamson, J. R., Fredrick, K., and Hwa, T. (2018). Slowdown of Transla-  
1185 tional Elongation in *Escherichia coli* under Hyperosmotic Stress. - PubMed - NCBI. *mBio*, 9(1):281.
- 1186 Dai, X., Zhu, M., Warren, M., Balakrishnan, R., Patsalo, V., Okano, H., Williamson, J. R., Fredrick, K., Wang, Y.-P., and Hwa, T.  
1187 (2016). Reduction of translating ribosomes enables *Escherichia coli* to maintain elongation rates during slow growth. *Nature  
1188 Microbiology*, 2(2):16231.
- 1189 Datsenko, K. A. and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products.  
1190 *Proceedings of the National Academy of Sciences*, 97(12):6640–6645.
- 1191 Delarue, M., Brittingham, G. P., Pfeffer, S., Surovtsev, I. V., Pinglay, S., Kennedy, K. J., Schaffer, M., Gutierrez, J. I., Sang, D.,  
1192 Poterewicz, G., Chung, J. K., Plitzko, J. M., Groves, J. T., Jacobs-Wagner, C., Engel, B. D., and Holt, L. J. (2018). mTORC1 Con-  
1193 trols Phase Separation and the Biophysical Properties of the Cytoplasm by Tuning Crowding. *Cell*, 174(2):338–349.e20.
- 1194 Dennis, P. P., Ehrenberg, M., and Bremer, H. (2004). Control of rRNA Synthesis in *Escherichia coli*: a Systems Biology Approach.  
1195 *Microbiology and Molecular Biology Reviews*, 68(4):639–668.

- 1196 Dill, K. A., Ghosh, K., and Schmit, J. D. (2011). Physical limits of cells and proteomes. *Proceedings of the National Academy of Sciences*, 108(44):17876–17882.
- 1198 Escalante, A., Salinas Cervantes, A., Gosset, G., and Bolívar, F. (2012). Current knowledge of the *Escherichia coli* phosphoenolpyruvate–carbohydrate phosphotransferase system: Peculiarities of regulation and impact on growth and product formation. *Applied Microbiology and Biotechnology*, 94(6):1483–1494.
- 1201 Fernández-Coll, L., Maciąg-Dorszynska, M., Tailor, K., Vadia, S., Levin, P. A., Szalewska-Palasz, A., Cashel, M., and Dunny, G. M. (2020). The Absence of (p)ppGpp Renders Initiation of *Escherichia coli* Chromosomal DNA Synthesis Independent of Growth Rates. *mBio*, 11(2):45.
- 1204 Fijalkowska, I. J., Schaaper, R. M., and Jonczyk, P. (2012). DNA replication fidelity in *Escherichia coli*: A multi-DNA polymerase affair. *FEMS Microbiology Reviews*, 36(6):1105–1121.
- 1206 Finkelstein, I. J. and Greene, E. C. (2013). Molecular Traffic Jams on DNA. *Annual Review of Biophysics*, 42(1):241–263.
- 1207 Gallagher, L. A., Bailey, J., and Manoil, C. (2020). Ranking essential bacterial processes by speed of mutant death. *Proceedings of the National Academy of Sciences*.
- 1209 Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeida, D., Muñiz-Rascado, L., García-Sotelo, J. S., Alquicira-Hernández, K., Martínez-Flores, I., Pannier, L., Castro-Mondragón, J. A., Medina-Rivera, A., Solano-Lira, H., Bonavides-Martínez, C., Pérez-Rueda, E., Alquicira-Hernández, S., Porrón-Sotelo, L., López-Fuentes, A., Hernández-Koutoucheva, A., Moral-Chávez, V. D., Rinaldi, F., and Collado-Vides, J. (2016). RegulonDB version 9.0: High-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Research*, 44(D1):D133–D143.
- 1214 Godin, M., Delgado, F. F., Son, S., Grover, W. H., Bryan, A. K., Tzur, A., Jorgensen, P., Payer, K., Grossman, A. D., Kirschner, M. W., and Manalis, S. R. (2010). Using buoyant mass to measure the growth of single cells. *Nature Methods*, 7(5):387–390.
- 1216 Guo, Y., Li, D., Zhang, S., Yang, Y., Liu, J.-J., Wang, X., Liu, C., Milkie, D. E., Moore, R. P., Tulu, U. S., Kiehart, D. P., Hu, J., Lippincott-Schwartz, J., Betzig, E., and Li, D. (2018). Visualizing Intracellular Organelle and Cytoskeletal Interactions at Nanoscale Resolution on Millisecond Timescales. *Cell*, 175(5):1430–1442.e17.
- 1219 Harris, L. K. and Theriot, J. A. (2018). Surface Area to Volume Ratio: A Natural Variable for Bacterial Morphogenesis. *Trends in microbiology*, 26(10):815–832.
- 1221 Hauryliuk, V., Atkinson, G. C., Murakami, K. S., Tenson, T., and Gerdes, K. (2015). Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nature Reviews Microbiology*, 13(5):298–309.
- 1223 Heldal, M., Norland, S., and Tumyr, O. (1985). X-ray microanalytic method for measurement of dry matter and elemental content of individual bacteria. *Applied and Environmental Microbiology*, 50(5):1251–1257.
- 1225 Helmstetter, C. E. and Cooper, S. (1968). DNA synthesis during the division cycle of rapidly growing *Escherichia coli* Br. *Journal of Molecular Biology*, 31(3):507–518.
- 1227 Henkel, S. G., Beek, A. T., Steinsiek, S., Stagge, S., Bettenbrock, K., de Mattos, M. J. T., Sauter, T., Sawodny, O., and Ederer, M. (2014). Basic Regulatory Principles of *Escherichia coli*'s Electron Transport Chain for Varying Oxygen Conditions. *PLoS ONE*, 9(9):e107640.
- 1230 Hui, S., Silverman, J. M., Chen, S. S., Erickson, D. W., Basan, M., Wang, J., Hwa, T., and Williamson, J. R. (2015). Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular Systems Biology*, 11(2):e784–e784.
- 1233 Ingledew, W. J. and Poole, R. K. (1984). The respiratory chains of *Escherichia coli*. *Microbiological Reviews*, 48(3):222–271.
- 1234 Ireland, W. T., Beeler, S. M., Flores-Bautista, E., Belliveau, N. M., Sweredoski, M. J., Moradian, A., Kinney, J. B., and Phillips, R. (2020). Deciphering the regulatory genome of *Escherichia coli*, one hundred promoters at a time. *bioRxiv*.
- 1236 Jensen, R. B., Wang, S. C., and Shapiro, L. (2001). A moving DNA replication factory in *Caulobacter crescentus*. *The EMBO journal*, 20(17):4952–4963.
- 1238 Jun, S., Si, F., Pugatch, R., and Scott, M. (2018). Fundamental principles in bacterial physiology - history, recent progress, and the future with focus on cell size control: A review. *Reports on Progress in Physics*, 81(5):056601.
- 1240 Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S. M., and Subhraveti, P. (2019). The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, 20(4):1085–1093.

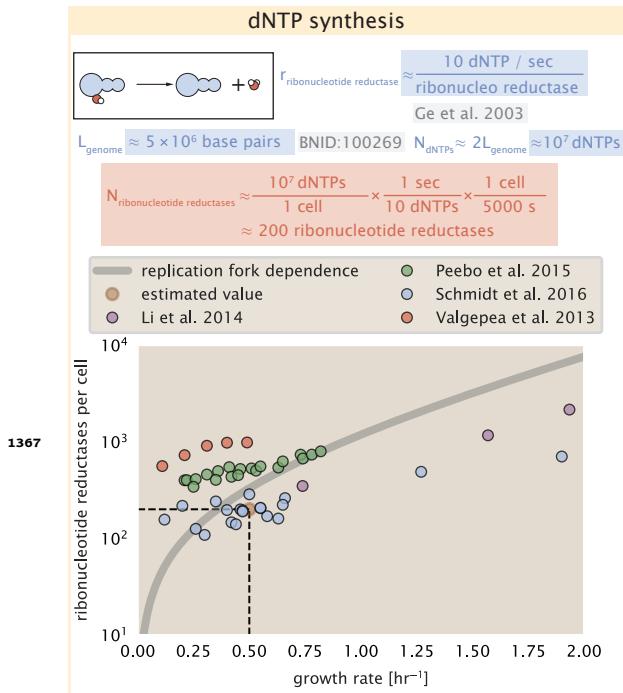
- 1243 Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Assad-Garcia, N., Glass, J. I., and Covert, M. W.  
1244 (2012). A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell*, 150(2):389–401.
- 1245 Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S.,  
1246 Kothari, A., Krummenacker, M., Latendresse, M., Muñiz-Rascado, L., Ong, Q., Paley, S., Peralta-Gil, M., Subhraveti, P., Velázquez-  
1247 Ramírez, D. A., Weaver, D., Collado-Vides, J., Paulsen, I., and Karp, P. D. (2017). The EcoCyc database: reflecting new knowledge  
1248 about Escherichia coliK-12. *Nucleic Acids Research*, 45(D1):D543–D550.
- 1249 Khademian, M. and Imlay, J. A. (2017). *Escherichia coli* cytochrome c peroxidase is a respiratory oxidase that enables the use of  
1250 hydrogen peroxide as a terminal electron acceptor. *Proceedings of the National Academy of Sciences*, 114(33):E6922–E6931.
- 1251 Klumpp, S. and Hwa, T. (2014). Bacterial growth: Global effects on gene expression, growth feedback and proteome partition.  
1252 *Current Opinion in Biotechnology*, 28:96–102.
- 1253 Klumpp, S., Zhang, Z., and Hwa, T. (2009). Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell*,  
1254 139(7):1366–1375.
- 1255 Kostinski, S. and Reuveni, S. (2020). Ribosome Composition Maximizes Cellular Growth Rates in *E. coli*. *Physical Review Letters*,  
1256 125(2):028103.
- 1257 Kraemer, J. A., Sanderlin, A. G., and Laub, M. T. (2019). The Stringent Response Inhibits DNA Replication Initiation in *E. coli* by  
1258 Modulating Supercoiling of oriC. *mBio*, 10(4):822.
- 1259 Lascu, I. and Gonin, P. (2000). The Catalytic Mechanism of Nucleoside Diphosphate Kinases. *Journal of Bioenergetics and Biomem-  
1260 branes*, 32(3):237–246.
- 1261 Laxhuber, K. S., Morrison, M. J., Chure, G., Belliveau, N. M., Strandkvist, C., Naughton, K. L., and Phillips, R. (2020). Theoretical  
1262 investigation of a genetic switch for metabolic adaptation. *PLOS ONE*, 15(5):e0226453.
- 1263 Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: visualization of intersecting sets. *IEEE Transactions  
1264 on Visualization and Computer Graphics*, 20(12):1983–1992.
- 1265 Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles  
1266 underlying allocation of cellular resources. *Cell*, 157(3):624–635.
- 1267 Liebermeister, W., Noor, E., Flamholz, A., Davidi, D., Bernhardt, J., and Milo, R. (2014). Visual account of protein investment in  
1268 cellular functions. *Proceedings of the National Academy of Sciences*, 111(23):8488–8493.
- 1269 Liu, M., Durfee, T., Cabrera, J. E., Zhao, K., Jin, D. J., and Blattner, F. R. (2005). Global Transcriptional Programs Reveal a Carbon  
1270 Source Foraging Strategy by *Escherichia coli*. *Journal of Biological Chemistry*, 280(16):15921–15927.
- 1271 Lu, D., Grayson, P., and Schulten, K. (2003). Glycerol Conductance and Physical Asymmetry of the *Escherichia coli* Glycerol Facili-  
1272 tator GlpF. *Biophysical Journal*, 85(5):2977–2987.
- 1273 Lynch, M. and Marinov, G. K. (2015). The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences*,  
1274 112(51):15690–15695.
- 1275 Maaløe, O. (1979). *Regulation of the Protein-Synthesizing Machinery—Ribosomes, tRNA, Factors, and So On*. Gene Expression.  
Springer.
- 1277 Metzl-Raz, E., Kafri, M., Yaakov, G., Soifer, I., Gurvich, Y., and Barkai, N. (2017). Principles of cellular resource allocation revealed  
1278 by condition-dependent proteome profiling. *eLife*, 6:e03528.
- 1279 Mikucki, J. A., Pearson, A., Johnston, D. T., Turchyn, A. V., Farquhar, J., Schrag, D. P., Anbar, A. D., Priscu, J. C., and Lee, P. A. (2009).  
1280 A Contemporary Microbially Maintained Subglacial Ferrous "Ocean". *Science*, 324(5925):397–400.
- 1281 Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). BioNumbers—the database of key numbers in molecular  
1282 and cell biology. *Nucleic Acids Research*, 38(suppl\_1):D750–D753.
- 1283 Monod, J. (1947). The phenomenon of enzymatic adaptation and its bearings on problems of genetics and cellular differentiation.  
1284 *Growth Symposium*, 9:223–289.
- 1285 Monod, J. (1949). The Growth of Bacterial Cultures. *Annual Review of Microbiology*, 3(1):371–394.
- 1286 Morgenstein, R. M., Bratton, B. P., Nguyen, J. P., Ouzounov, N., Shaevitz, J. W., and Gitai, Z. (2015). RodZ links MreB to cell wall  
1287 synthesis to mediate MreB rotation and robust morphogenesis. *Proceedings of the National Academy of Sciences*, 112(40):12510–  
1288 12515.

- 1289 Neidhardt, F. C., Ingraham, J., and Schaechter, M. (1991). *Physiology of the Bacterial Cell - A Molecular Approach*, volume 1. Elsevier.
- 1290 Ojikic, N., Serbanescu, D., and Banerjee, S. (2019). Surface-to-volume scaling and aspect ratio preservation in rod-shaped bacteria. *eLife*, 8:642.
- 1292 Patrick, M., Dennis, P. P., Ehrenberg, M., and Bremer, H. (2015). Free RNA polymerase in *Escherichia coli*. *Biochimie*, 119:80–91.
- 1293 Peebo, K., Valgepea, K., Maser, A., Nahku, R., Adamberg, K., and Vilu, R. (2015). Proteome reallocation in *Escherichia coli* with increasing specific growth rate. *Molecular BioSystems*, 11(4):1184–1193.
- 1295 Phillips, R. (2018). Membranes by the Numbers. In *Physics of Biological Membranes*, pages 73–105. Springer, Cham, Cham.
- 1296 Ranganathan, S., Tee, T. W., Chowdhury, A., Zomorodi, A. R., Yoon, J. M., Fu, Y., Shanks, J. V., and Maranas, C. D. (2012). An integrated computational and experimental study for overproducing fatty acids in *Escherichia coli*. *Metabolic Engineering*, 14(6):687–704.
- 1299 Rogers, H., Perkins, H., and Ward, J. (1980). *Microbial Cell Walls and Membranes*. Chapman and Hall, London.
- 1300 Roller, B. R. K., Stoddard, S. F., and Schmidt, T. M. (2016). Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nature microbiology*, 1(11):1–7.
- 1302 Rosenberg, H., Gerdes, R. G., and Chegwidden, K. (1977). Two systems for the uptake of phosphate in *Escherichia coli*. *Journal of Bacteriology*, 131(2):505–511.
- 1304 Ruppe, A. and Fox, J. M. (2018). Analysis of Interdependent Kinetic Controls of Fatty Acid Synthases. *ACS Catalysis*, 8(12):11722–11734.
- 1306 Schaechter, M., MaalØe, O., and Kjeldgaard, N. O. (1958). Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of *Salmonella typhimurium*. *Microbiology*, 19(3):592–606.
- 1308 Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., and Heinemann, M. (2016). The quantitative and condition-dependent *Escherichia coli* proteome. *Nature Biotechnology*, 34(1):104–110.
- 1310 Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330(6007):1099–1102.
- 1312 Shi, H., Bratton, B. P., Gitai, Z., and Huang, K. C. (2018). How to Build a Bacterial Cell: MreB as the Foreman of *E. coli* Construction. *Cell*, 172(6):1294–1305.
- 1314 Si, F., Le Treut, G., Sauls, J. T., Vadia, S., Levin, P. A., and Jun, S. (2019). Mechanistic Origin of Cell-Size Control and Homeostasis in Bacteria. *Current Biology*, 29(11):1760–1770.e7.
- 1316 Si, F., Li, D., Cox, S. E., Sauls, J. T., Azizi, O., Sou, C., Schwartz, A. B., Erickstad, M. J., Jun, Y., Li, X., and Jun, S. (2017). Invariance of Initiation Mass and Predictability of Cell Size in *Escherichia coli*. *Current Biology*, 27(9):1278–1287.
- 1318 Sohlenkamp, C. and Geiger, O. (2016). Bacterial membrane lipids: Diversity in structures and pathways. *FEMS Microbiology Reviews*, 40(1):133–159.
- 1320 Soler-Bistué, A., Aguilar-Pierlé, S., García-Garcé, M., Val, M.-E., Sismeiro, O., Varet, H., Sieira, R., Krin, E., Skovgaard, O., Comerci, D. J., Eduardo P. C. Rocha, and Mazel, D. (2020). Macromolecular crowding links ribosomal protein gene dosage to growth rate in *Vibrio cholerae*. *BMC Biology*, 18(1):1–18.
- 1323 Soufi, B., Krug, K., Harst, A., and Macek, B. (2015). Characterization of the *E. coli* proteome and its modifications during growth and ethanol stress. *Frontiers in Microbiology*, 6:198.
- 1325 Stasi, R., Neves, H. I., and Spira, B. (2019). Phosphate uptake by the phosphonate transport system PhnCDE. *BMC Microbiology*, 19.
- 1327 Stouthamer, A. H. (1973). A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie van Leeuwenhoek*, 39(1):545–565.
- 1329 Stouthamer, A. H. and Bettenhausen, C. W. (1977). A continuous culture study of an ATPase-negative mutant of *Escherichia coli*. *Archives of Microbiology*, 113(3):185–189.
- 1331 Szenk, M., Dill, K. A., and de Graff, A. M. R. (2017). Why Do Fast-Growing Bacteria Enter Overflow Metabolism? Testing the Membrane Real Estate Hypothesis. *Cell Systems*, 5(2):95–104.

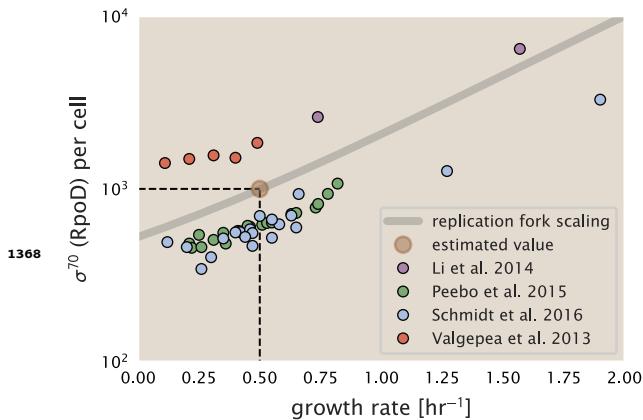
- 1333 Taheri-Araghi, S., Bradde, S., Sauls, J. T., Hill, N. S., Levin, P. A., Paulsson, J., Vergassola, M., and Jun, S. (2015). Cell-size control  
1334 and homeostasis in bacteria. - PubMed - NCBI. *Current Biology*, 25(3):385–391.
- 1335 Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X. S. (2010). Quantifying *E. coli* proteome and  
1336 transcriptome with single-molecule sensitivity in single cells. *Science (New York, N.Y.)*, 329(5991):533–538.
- 1337 Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of  
1338 protein functions and evolution. *Nucleic Acids Research*, 28(1):33–36.
- 1339 Taymaz-Nikerel, H., Borujeni, A. E., Verheijen, P. J. T., Heijnen, J. J., and van Gulik, W. M. (2010). Genome-derived minimal metabolic  
1340 models for *Escherichia coli* mg1655 with estimated in vivo respiratory ATP stoichiometry. *Biotechnology and Bioengineering*,  
1341 107(2):369–381. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/bit.22802>.
- 1342 The Gene Ontology Consortium (2018). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*,  
1343 47(D1):D330–D338.
- 1344 Valgepea, K., Adamberg, K., Seiman, A., and Vilu, R. (2013). *Escherichia coli* achieves faster growth by increasing catalytic and  
1345 translation rates of proteins. *Molecular BioSystems*, 9(9):2344.
- 1346 Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W.,  
1347 Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson,  
1348 E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero,  
1349 E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors, S. . . (2020). SciPy 1.0:  
1350 Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.
- 1351 Volkmer, B. and Heinemann, M. (2011). Condition-Dependent Cell Volume and Concentration of *Escherichia coli* to Facilitate  
1352 Data Conversion for Systems Biology Modeling. *PLOS ONE*, 6(7):e23126.
- 1353 Vollmer, W., Blanot, D., and De Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*,  
1354 32(2):149–167.
- 1355 Weber, J. and Senior, A. E. (2003). ATP synthesis driven by proton transport in F1F0-ATP synthase. *FEBS Letters*, 545(1):61–70.
- 1356 Willsky, G. R., Bennett, R. L., and Malamy, M. H. (1973). Inorganic Phosphate Transport in *Escherichia coli*: Involvement of Two  
1357 Genes Which Play a Role in Alkaline Phosphatase Regulation. *Journal of Bacteriology*, 113(2):529–539.
- 1358 You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C. W., Wang, Y.-P., Lenz, P., Yan, D., and Hwa, T. (2013). Coordination  
1359 of bacterial proteome with metabolism by cyclic AMP signalling. *Nature*, 500(7462):301–306.
- 1360 Yu, X., Liu, T., Zhu, F., and Khosla, C. (2011). In vitro reconstitution and steady-state analysis of the fatty acid synthase from  
1361 *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 108(46):18643–18648.
- 1362 Zhang, Z., Aboulwafa, M., and Saier, M. H. (2014). Regulation of crp gene expression by the catabolite repressor/activator, cra,  
1363 in *Escherichia coli*. *Journal of Molecular Microbiology and Biotechnology*, 24(3):135–141.
- 1364 Zhu, M. and Dai, X. (2019). Growth suppression by altered (p)ppGpp levels results from non-optimal resource allocation in  
1365 *Escherichia coli*. *Nucleic Acids Research*, 47(9):4684–4693.



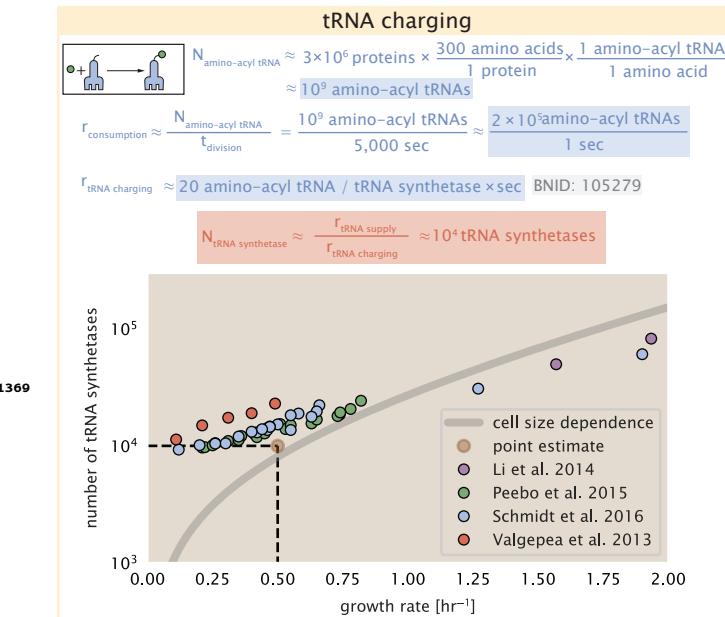
**Figure 2-Figure supplement 1.** (A) Estimate for the number of PitA phosphate transport systems needed to maintain a 3% phosphorus *E. coli* dry mass. Points in plot correspond to the total number of PitA transporters per cell. (B) Estimate of the number of CysUWA complexes necessary to maintain a 1% sulfur *E. coli* dry mass. Points in plot correspond to average number of CysUWA transporter complexes that can be formed given the transporter stoichiometry  $[CysA]_2[CysU][CysW][Sbp/CysP]$ . Grey line in (A) and (B) represents the estimated number of transporters per cell at a continuum of growth rates.



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



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



**Figure 8-Figure supplement 1.** Estimation for the number of tRNA synthetases that will supply the required amino acid demand. The sum of all tRNA synthetases copy numbers are plotted as a function of growth rate ([ArgS], [CysS], [GlnS], [GltX], [IleS], [LeuS], [ValS], [AlaS]<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>).