

# 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.**  
<sup>4</sup> **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>**

\*For correspondence:

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

<sup>5</sup> <sup>1</sup>Department of Biology, University of Washington, Seattle, WA, USA; <sup>2</sup>Department of  
<sup>6</sup> Applied Physics, California Institute of Technology, Pasadena, CA, USA; <sup>3</sup>Department of  
<sup>7</sup> Chemical Engineering, Stanford University, Stanford, CA, USA; <sup>4</sup>Department of  
<sup>8</sup> Molecular Cell Biology and Department of Physics, University of California Berkeley,  
<sup>9</sup> Berkeley, CA, USA; <sup>5</sup>Department of Physics, Brandeis University, Waltham, MA, USA;  
<sup>10</sup> <sup>6</sup>Department of Applied Physics, Stanford University, Stanford, CA, USA; <sup>7</sup>Allen Institute  
<sup>11</sup> for Cell Science, Seattle, WA, USA; <sup>8</sup>Division of Biology and Biological Engineering,  
<sup>12</sup> California Institute of Technology, Pasadena, CA, USA; <sup>9</sup>Department of Physics,  
<sup>13</sup> California Institute of Technology, Pasadena, CA, USA; \*Co-corresponding authors.  
<sup>14</sup> Address correspondence to phillips@pboc.caltech.edu and jtheriot@uw.edu

15

---

<sup>16</sup> **Abstract** Recent years have seen a deluge of experiments dissecting the relationship between  
<sup>17</sup> bacterial growth rate, cell size, and protein content, quantifying the abundance of proteins across  
<sup>18</sup> growth conditions with unprecedented resolution. However, we still lack a rigorous  
<sup>19</sup> understanding of what sets the scale of these measurements and whether protein abundances  
<sup>20</sup> should (or should not) depend on growth rate. Here, we seek to quantitatively understand this  
<sup>21</sup> relationship across a collection of *Escherichia coli* proteomic data sets covering  $\approx 4000$  proteins  
<sup>22</sup> and 31 growth conditions. We estimate the basic requirements for steady-state growth by  
<sup>23</sup> considering key processes in nutrient transport, energy generation, cell envelope biogenesis, and  
<sup>24</sup> the central dogma, from which ribosome biogenesis emerges as a primary determinant of  
<sup>25</sup> growth rate. We conclude by exploring a model of ribosomal regulation as a function of the  
<sup>26</sup> nutrient supply, revealing a mechanism tying cell size and growth rate to ribosomal content.

27

---

## 28 Introduction

29 The observed range of bacterial growth rates is enormously diverse. In natural environments,  
<sup>30</sup> some microbial organisms might double only once per year (?) while in comfortable laboratory  
<sup>31</sup> conditions, growth can be rapid with several divisions per hour (?). This six order-of-magnitude  
<sup>32</sup> difference in time scales encompasses different microbial species and lifestyles, yet even for a sin-  
<sup>33</sup> gle species such as *E. coli*, the growth rate can be modulated over a similar scale by tuning the  
<sup>34</sup> type and amount of nutrients in the growth medium. This remarkable flexibility in growth rate  
<sup>35</sup> illustrates the intimate relationship between environmental conditions and the rates at which cells  
<sup>36</sup> convert nutrients into new cellular material – a relationship that has remained a major topic of  
<sup>37</sup> inquiry in bacterial physiology for over a century (?).

38 As was noted by Jacques Monod, “the study of the growth of bacterial cultures does not con-  
<sup>39</sup> stitute a specialized subject or branch of research, it is the basic method of Microbiology.” Those  
<sup>40</sup> words ring as true today as they did when they were written 70 years ago (?). Indeed, the study  
<sup>41</sup> of bacterial growth has undergone a renaissance. Many of the key questions addressed by the

42 pioneering efforts in the middle of the last century can be revisited by examining them through  
 43 the lens of the increasingly refined molecular census that is available for bacteria such as the mi-  
 44 crobial workhorse *Escherichia coli*. In this work, we explore an amalgamation of recent proteomic  
 45 data sets to explore fundamental limits of bacterial growth.

46 Several of the evergreen questions about bacterial growth that were originally raised by micro-  
 47 biologists in the middle of the 20th century can now be reframed in light of this newly available  
 48 data. For example, what biological processes set the absolute speed limit for how fast bacterial  
 49 cells can grow and reproduce? How do cells alter the absolute numbers and relative ratios of  
 50 their molecular constituents as a function of changes in growth rate or nutrient availability? In  
 51 this paper, we address these two questions from two distinct angles. First, as a result of an array  
 52 of high-quality proteome-wide measurements of the *E. coli* proteome under myriad growth condi-  
 53 tions, we have a census that allows us to explore how the number of key molecular players change  
 54 as a function of growth rate. This census provides a window into the question of whether the rates  
 55 of central processes such as energy generation or DNA synthesis are regulated systematically as  
 56 a function of cell growth rate by altering protein copy number in individual cells. Second, by com-  
 57 piling molecular turnover rate measurements for many of the fundamental processes associated  
 58 with bacterial growth, we can make quantitative estimates to determine whether the observed pro-  
 59 tein copy numbers under varying conditions appear to be in excess of what would be minimally  
 60 required to support cell growth at the observed rates.

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

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

## 124 Uptake of Nutrients

125 We begin our series of estimates by considering the critical transport processes diagrammed in  
 126 ??(A). In order to build new cellular mass, the molecular and elemental building blocks must be scav-  
 127 enged from the environment in different forms. Carbon, for example, is acquired via the transport  
 128 of carbohydrates and sugar alcohols with some carbon sources receiving preferential treatment in  
 129 their consumption (?). Phosphorus, sulfur, and nitrogen, on the other hand, are harvested primar-  
 130 ily in the forms of inorganic salts, namely phosphate, sulfate, and ammonia (?????). All of these  
 131 compounds have different permeabilities across the cell membrane ? and most require some  
 132 energetic investment either via ATP hydrolysis or through the proton electrochemical gradient to  
 133 bring the material across the hydrophobic cell membrane. Given the diversity of biological trans-  
 134 port mechanisms and the vast number of inputs needed to build a cell, we begin by considering

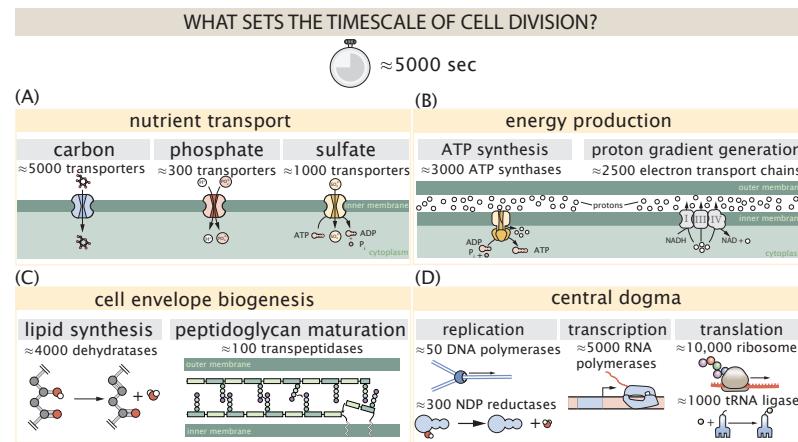
**Box 1. The Rules of Engagement for Order-Of-Magnitude Estimates**

This work relies heavily on so-called "back-of-the-envelope" estimates to understand the abundances and growth-rate dependences of a variety of molecular complexes. This moniker 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.

**The rule of "one, few, and ten".** The philosophy behind order-of-magnitude estimates is to provide a 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. 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.

**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 (?) as a repository for such information. Every value we draw from this database has an associated BioNumbers ID number, abbreviated as BNID, and we provide this reference in grey-boxes in each figure.

**Uncertainty in the data sets and the accuracy of an estimate.** The data sets presented in this work are the products of careful experimentation with the aim to report, to the best of their ability, the absolute copy numbers of proteins in the cell. These data, collected over the span of a few years, come from different labs and use different internal standards, controls, and even techniques (discussed further in Supplemental Section ??). 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  $\approx 5000 \text{ s}$ .

135 transport of some of the most important cellular ingredients: carbon, nitrogen, oxygen, hydrogen,  
136 phosphorus, and sulfur.

137 The elemental composition of *E. coli* has received much quantitative attention over the past  
138 half century (????), providing us with a starting point for estimating the copy numbers of various  
139 transporters. While there is some variability in the exact elemental percentages (with different  
140 uncertainties), we can estimate that the dry mass of a typical *E. coli* cell is  $\approx 45\%$  carbon (BioN-  
141 umber ID: 100649, see ??),  $\approx 15\%$  nitrogen (BNID: 106666),  $\approx 3\%$  phosphorus (BNID: 100653), and  
142 1% sulfur (BNID: 100655). In the coming paragraphs, we will engage in a dialogue between back-  
143 of-the-envelope estimates for the numbers of transporters needed to facilitate these chemical  
144 stoichiometries and the experimental proteomic measurements of the biological reality. Such an  
145 approach provides the opportunity to test if our biological knowledge is sufficient to understand  
146 the scale at which these complexes are produced. At the end of this section, we discuss physical  
147 limits as to the number of transporters that can be present, and comment on the plausibility of  
148 this process acting as a molecular bottleneck.

### 149 Nitrogen Transport

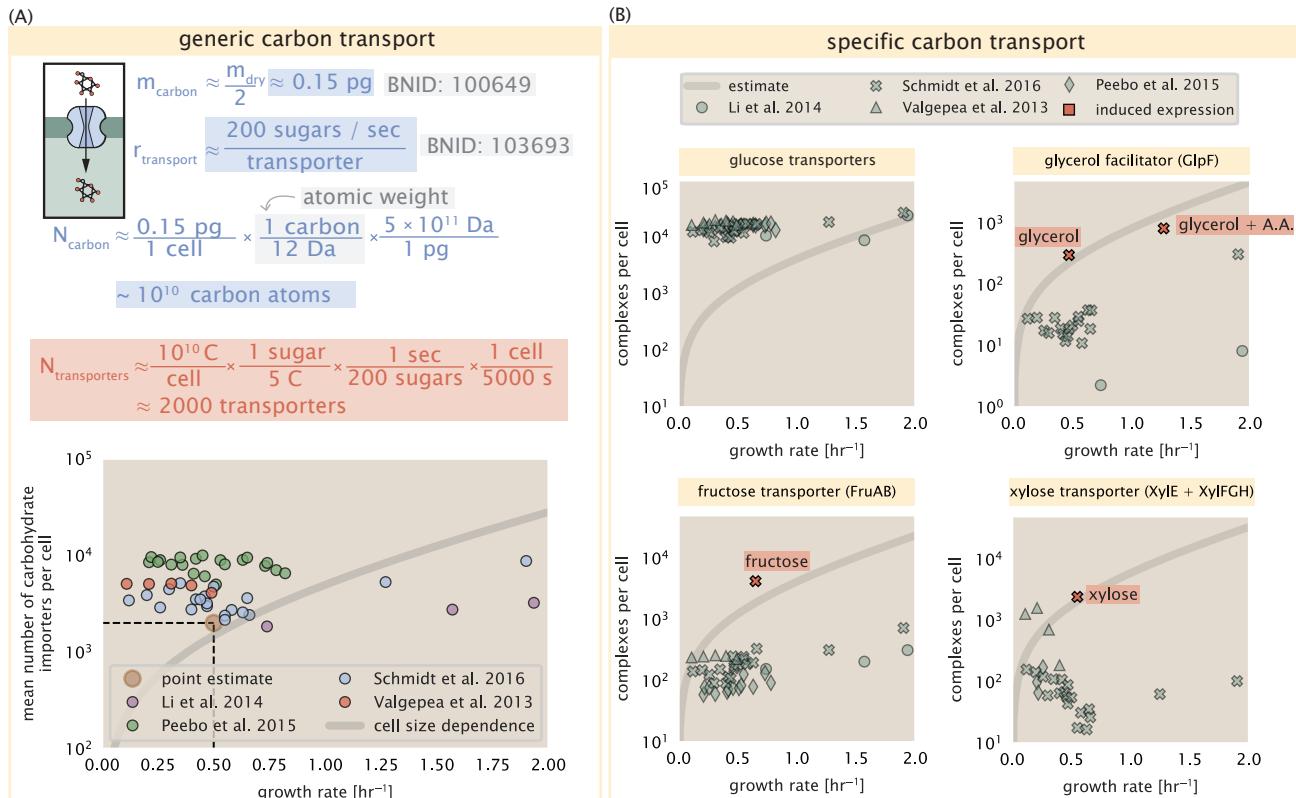
150 We must first address which elemental sources must require proteinaceous transport, meaning  
151 that the cell cannot acquire appreciable amounts simply via diffusion across the membrane. The  
152 permeability of the lipid membrane to a large number of solutes has been extensively charac-  
153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

**164 Carbon Transport**

165 We begin with the most abundant element in *E. coli* by mass, carbon. Using  $\approx 0.3$  pg as the typical  
 166 *E. coli* dry mass (BNID: 103904), we estimate that  $\sim 10^{10}$  carbon atoms must be brought into the  
 167 cell in order to double all of the carbon-containing molecules (??(A, top)). Typical laboratory growth  
 168 conditions, such as those explored in the aforementioned proteomic data sets, provide carbon as  
 169 a single class of sugar such as glucose, galactose, or xylose to name a few. *E. coli* has evolved myr-  
 170 iad mechanisms by which these sugars can be transported across the cell membrane. One such  
 171 mechanism of transport is via the PTS system which is a highly modular system capable of trans-  
 172 porting a diverse range of sugars (?). The glucose-specific component of this system transports  $\approx$   
 173 200 glucose molecules per second per transporter (BNID: 114686). Making the assumption that  
 174 this is a typical sugar transport rate, coupled with the need to transport  $\sim 10^{10}$  carbon atoms, we  
 175 arrive at the conclusion that on the order of 1,000 transporters must be expressed in order to bring  
 176 in enough carbon atoms to divide in 5000 s, diagrammed in the top panel of ??(A). This estimate,  
 177 along with the observed average number of the PTS system carbohydrate transporters present in  
 178 the proteomic data sets (????), is shown in ??(A). While we estimate 1500 transporters are needed  
 179 with a 5000 s division time, we can abstract this calculation to consider any particular growth rate  
 180 given knowledge of the cell density and volume as a function of growth rate and direct the reader  
 181 to the Supplemental Information for more information. As revealed in ??(A), experimental mea-  
 182 surements exceed the estimate by several fold, illustrating that transport of carbon into the cell is  
 183 not rate limiting for cell division. Abstracting this point estimate at 5000 s to a continuum of growth  
 184 rates (grey line in ??(A)) reveals an excess of transporters at other growth rates, though in rapid  
 185 growth regimes, the abundance is below our simple estimate.

186 The estimate presented in ??(A) neglects any specifics of the regulation of the carbon transport  
 187 system and presents a view of how many carbohydrate transporters are present on average. Using  
 188 the diverse array of growth conditions explored in the proteomic data sets, we can explore how  
 189 individual carbon transport systems depend on the population growth rate. In ??(B), we show the  
 190 total number of carbohydrate transporters specific to different carbon sources. A striking obser-  
 191 vation, shown in the top-left plot of ??(B), is the constancy in the expression of the glucose-specific  
 192 transport systems. Additionally, we note that the total number of glucose-specific transporters is  
 193 tightly distributed at  $\approx 10^4$  per cell, the approximate number of transporters needed to sustain  
 194 rapid growth of several divisions per hour. This illustrates that *E. coli* maintains a substantial num-  
 195 ber of complexes present for transporting glucose regardless of growth rate, which is known to be  
 196 the preferential carbon source (??).

197 It is now understood that a large number of metabolic operons are regulated with dual-input  
 198 logic gates that are only expressed when glucose concentrations are low (mediated by cyclic-AMP  
 199 receptor protein CRP) and the concentration of other carbon sources are elevated (??). A famed  
 200 example of such dual-input regulatory logic is in the regulation of the *lac* operon which is only  
 201 natively activated in the absence of glucose and the presence of allolactose, an intermediate in  
 202 lactose metabolism (?), though we now know of many other such examples (??). This illustrates  
 203 that once glucose is depleted from the environment, cells have a means to dramatically increase  
 204 the abundance of the specific transporter needed to digest the next sugar that is present. Several  
 205 examples of induced expression of specific carbon-source transporters are shown in ??(B). Points  
 206 colored in red (labeled by red text-boxes) correspond to growth conditions in which the specific  
 207 carbon source (glycerol, xylose, or fructose) is present. These plots show that, in the absence of the  
 208 particular carbon source, expression of the transporters is maintained on the order of  $\sim 10^2$  per cell.  
 209 However, when the transport substrate is present, expression is induced and the transporters be-  
 210 come highly-expressed. The grey lines in ??(B) show the estimated number of transporters needed  
 211 at each growth rate to satisfy the cellular carbon requirement. It is notable that in all cases, the  
 212 magnitude of induced expression (shown in red) falls close to the estimate, illustrating the ability  
 213 of the cell to tune expression in response to changing environments. Together, this generic esti-



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

214 mation and the specific examples of induced expression suggest that transport of carbon across  
 215 the cell membrane, while critical for growth, is not the rate-limiting step of cell division.

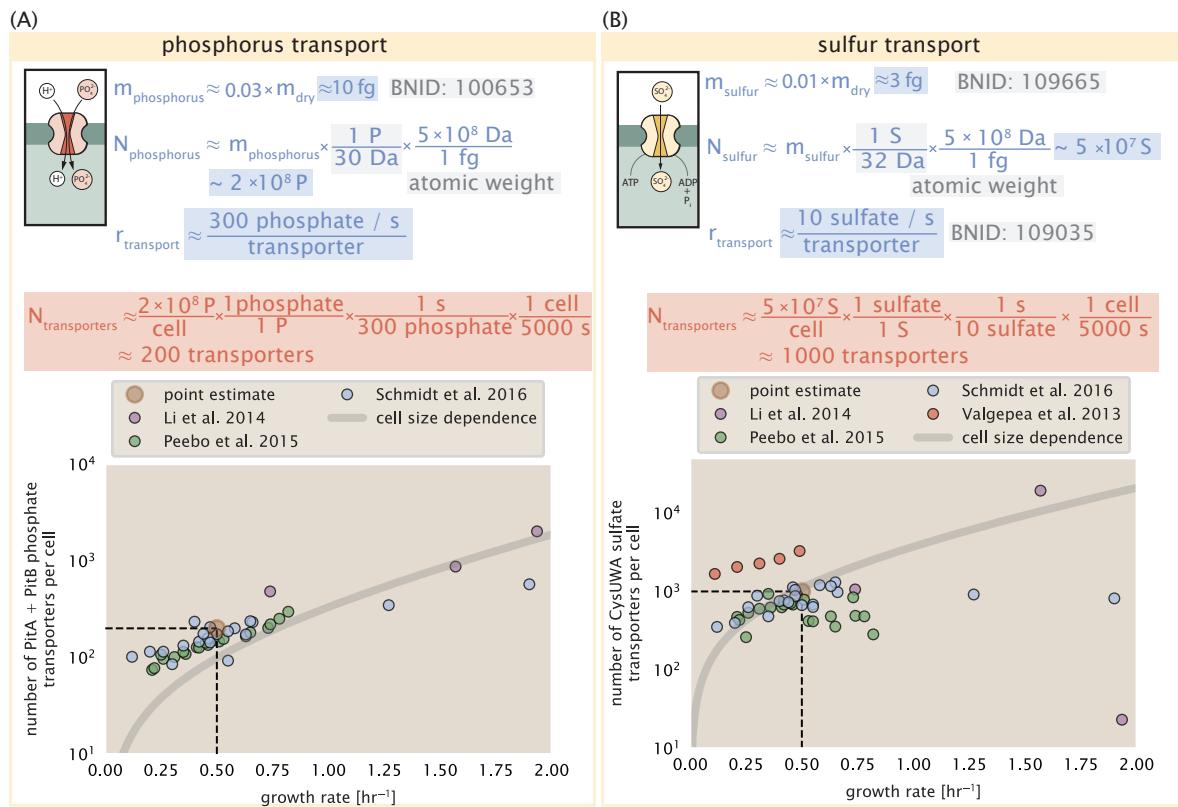
### 216 **Phosphorus and Sulfur Transport**

217 We now turn our attention towards other essential elements, namely phosphorus and sulfur. Phos-  
 218 phorus is critical to the cellular energy economy in the form of high-energy phosphodiester bonds  
 219 making up DNA, RNA, and the NTP energy pool as well as playing a critical role in the post-translational  
 220 modification of proteins and defining the polar-heads of lipids. In total, phosphorus makes up  
 221  $\approx 3\%$  of the cellular dry mass which in typical experimental conditions is in the form of inorganic  
 222 phosphate. The cell membrane has remarkably low permeability to this highly-charged and critical  
 223 molecule, therefore requiring the expression of active transport systems. In *E. coli*, the proton elec-  
 224 trochemical gradient across the inner membrane is leveraged to transport inorganic phosphate  
 225 into the cell (?). Proton-solute symporters are widespread in *E. coli* (??) and can have rapid trans-  
 226 port rates of 50 to 100 molecules per second for sugars and other solutes (BNID: 103159; 111777).  
 227 As a more extreme example, the proton transporters in the F<sub>1</sub>-F<sub>0</sub> ATP synthase, which use the pro-  
 228 ton electrochemical gradient for rotational motion, can shuttle protons across the membrane at  
 229 a rate of  $\approx 1000$  per second (BNID: 104890; 103390). In *E. coli* the PitA phosphate transport sys-  
 230 tem has been shown to be very tightly coupled with the proton electrochemical gradient with a  
 231 1:1 proton:phosphate stoichiometric ratio (??). Taking the geometric mean of the aforementioned  
 232 estimates gives a plausible rate of phosphate transport on the order of 300 per second. Illustrated  
 233 in ??(A), we can estimate that  $\approx 200$  phosphate transporters are necessary to maintain an  $\approx 3\%$  dry  
 234 mass with a 5000 s division time. This estimate is consistent with observation when we examine  
 235 the observed copy numbers of PitA in proteomic data sets (plot in ??(A)). While our estimate is very  
 236 much in line with the observed numbers, we emphasize that this is likely a slight overestimate of  
 237 the number of transporters needed as there are other phosphorous scavenging systems, such as  
 238 the ATP-dependent phosphate transporter Pst system which we have neglected.

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

### 253 **Limits on Transporter Expression**

254 So which, if any, of these processes may be rate limiting for growth? As suggested by ?? (B), induced  
 255 expression can lead to an order-of-magnitude (or more) increase in the amount of transporters  
 256 needed to facilitate transport. Thus, if acquisition of nutrients was the limiting state in cell division,  
 257 could expression simply be increased to accommodate faster growth? A way to approach this  
 258 question is to compute the amount of space in the bacterial membrane that could be occupied  
 259 by nutrient transporters. Considering a rule-of-thumb for the surface area of *E. coli* of about 5  
 260  $\mu\text{m}^2$  (BNID: 101792), we expect an areal density for 1000 transporters to be approximately 200  
 261 transporters/  $\mu\text{m}^2$ . For a typical transporter occupying about 50 nm<sup>2</sup>/dimer, this amounts to about  
 262 only 1 percent of the total inner membrane (?). In addition, bacterial cell membranes typically have



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

263 densities of  $10^5$  proteins/ $\mu\text{m}^2$  (?), implying that the cell could accommodate more transporters of a  
 264 variety of species if it were rate limiting. As we will see in the next section, however, occupancy of  
 265 the membrane can impose other limits on the rate of energy production.

### 266 Translation and Ribosomal Synthesis

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

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

### 278 tRNA Synthetases

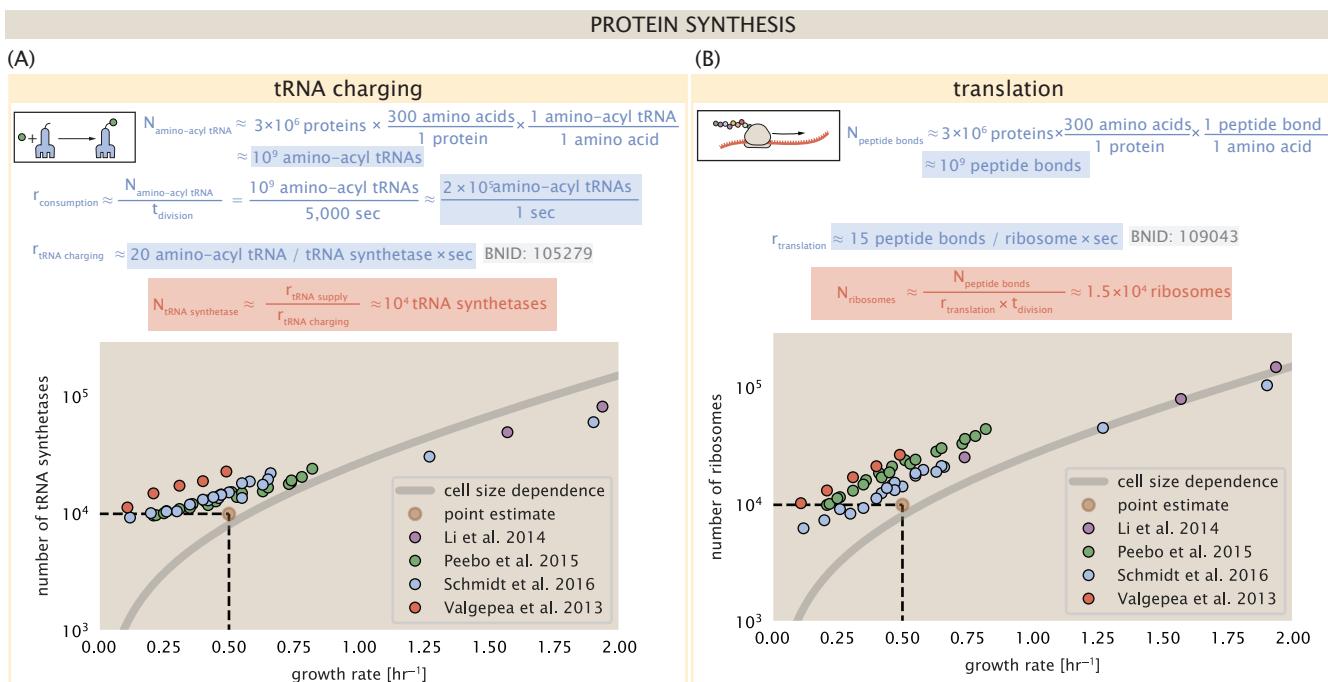
279 We begin by first estimating the number of tRNA synthetases in *E. coli* needed to convert free  
 280 amino-acids to polypeptide chains. Again using an estimate of  $\approx 3 \times 10^6$  proteins per cell at a 5000 s  
 281 division time (BNID: 115702) and a typical protein length of  $\approx 300$  amino acids (BNID: 100017), we  
 282 can estimate that a total of  $\approx 10^9$  amino acids are stitched together by peptide bonds.

283 How many tRNAs are needed to facilitate this remarkable number of amino acid delivery events  
 284 to the translating ribosomes? It is important to note that tRNAs are recycled after they've passed  
 285 through the ribosome and can be recharged with a new amino acid, ready for another round of  
 286 peptide bond formation. While some *in vitro* data exists on the turnover of tRNA in *E. coli* for  
 287 different amino acids, we can make a reasonable estimate by comparing the number of amino  
 288 acids to be polymerized to cell division time. Using our stopwatch of 5000 s and  $10^9$  amino acids,  
 289 we arrive at a requirement of  $\approx 2 \times 10^5$  tRNA molecules to be consumed by the ribosome per  
 290 second.

291 There are many processes which go into synthesizing a tRNA and ligating it with the appropriate  
 292 amino acids. As we discussed previously, there appear to be more than enough RNA polymerases  
 293 per cell to synthesize the needed pool of tRNAs. Without considering the many ways in which  
 294 amino acids can be scavenged or synthesized *de novo*, we can explore ligation the as a potential  
 295 rate limiting step. The enzymes which link the correct amino acid to the tRNA, known as tRNA  
 296 synthetases or tRNA ligases, are incredible in their proofreading of substrates with the incorrect  
 297 amino acid being ligated once out of every  $10^4$  to  $10^5$  events (BNID: 103469). This is due in part  
 298 to the consumption of energy as well as a multi-step pathway to ligation. While the rate at which  
 299 tRNA is ligated is highly dependent on the identity of the amino acid, it is reasonable to state that  
 300 the typical tRNA synthetase has charging rate of  $\approx 20$  AA per tRNA synthetase per second (BNID:  
 301 105279).

302 We can make an assumption that amino-acyl tRNAs are in steady-state where they are pro-  
 303 duced at the same rate they are consumed, meaning that  $2 \times 10^5$  tRNAs must be charged per second.  
 304 Combining these estimates together, as shown schematically in ??(A), yields an estimate of  $\sim 10^4$   
 305 tRNA synthetases per cell with a division time of 5000 s. This point estimate is in very close agree-  
 306 ment with the observed number of synthetases (the sum of all 20 tRNA synthetases in *E. coli*). This  
 307 estimation strategy seems to adequately describe the observed growth rate dependence of the  
 308 tRNA synthetase copy number (shown as the grey line in ??(B)), suggesting that the copy number  
 309 scales with the cell volume.

310 In total, the estimated and observed  $\sim 10^4$  tRNA synthetases occupy only a meager fraction of  
 311 the total cell proteome, around 0.5% by abundance. It is reasonable to assume that if tRNA charg-



**Figure 4. Estimation of the required tRNA synthetases and ribosomes.** (A) Estimation for the number of tRNA synthetases that will supply the required amino acid demand. The sum of all tRNA synthetases copy numbers are plotted as a function of growth rate ([ArgS], [CysS], [GlnS], [Glx], [IleS], [LeuS], [ValS], [AlaS]<sub>2</sub>, [AsnS]<sub>2</sub>, [AspS]<sub>2</sub>, [TyrS]<sub>2</sub>, [TrpS]<sub>2</sub>, [ThrS]<sub>2</sub>, [SerS]<sub>2</sub>, [ProS]<sub>2</sub>, [PheS]<sub>2</sub>[PheT]<sub>2</sub>, [MetG]<sub>2</sub>, [LysS]<sub>2</sub>, [GlyS]<sub>2</sub>[GlyQ]<sub>2</sub>). (B) Estimation of the number of ribosomes required to synthesize 10<sup>9</sup> peptide bonds with an elongation rate of 15 peptide bonds per second. The average abundance of ribosomes is plotted as a function of growth rate. Our estimated values are shown for a growth rate of 0.5 hr<sup>-1</sup>. Grey lines correspond to the estimated complex abundance calculated at different growth rates. See Supplemental Information XX for a more detail description of this calculation.

ing was a rate limiting process, cells would be able to increase their growth rate by devoting more cellular resources to making more tRNA synthases. As the synthesis of tRNAs and the corresponding charging can be highly parallelized, we can argue that tRNA charging is not a rate limiting step in cell division, at least for the growth conditions explored in this work.

### 316 Protein Synthesis

With the number of tRNA synthetases accounted for, we now consider the abundance of the protein synthesis machines themselves, ribosomes. Ribosomes are enormous protein/rRNA complexes that facilitate the peptide bond formation between amino acids in the correct sequence as defined by the coding mRNA. Before we examine the synthesis of the ribosome proteins and the limits that may place on the observed bacterial growth rates, let's consider replication of the cellular proteome.

While the rate at which ribosomes translates is well known to have a growth rate dependence ? and is a topic which we discuss in detail in the coming sections. However, for the purposes of our order-of-magnitude estimate, we can make the approximation that translation occurs at a rate of  $\approx 15$  amino acids per second per ribosome (BNID: 100233). Under this approximation and assuming a division time of 5000 s, we can arrive at an estimate of  $\approx 10^4$  ribosomes are needed to replicate the cellular proteome, shown in ??(B). This point estimate, while glossing over important details such as chromosome copy number and growth-rate dependent translation rates, proves to be notably accurate when compared to the experimental observations (??(B)).

331 **Discussion**

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

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

360 Of the processes illustrated in ??, we arrive at a ribosome-centric view of cellular growth rate  
 361 control. This is in some sense unsurprising given the long-held observation that *E. coli* and many  
 362 other organisms vary their ribosomal abundance as a function of growth conditions and growth  
 363 rate ?? . However, through our dialogue with the proteomic data, two additional key points emerge.  
 364 The first relates to our question of what process sets the absolute speed limit of bacterial growth.  
 365 While a cell can parallelize many of its processes simply by increasing the abundance of specific pro-  
 366 teins or firing multiple rounds of DNA replication, this is not so for synthesis of ribosomes (??(A)).  
 367 The translation time for each ribosome [ $\approx$  6 min, ?] places an inherent limit on the growth rate  
 368 that can only be surpassed if the cell were to increase their polypeptide elongation rate, or if they  
 369 could reduce the total protein and rRNA mass of the ribosome. The second point relates to the  
 370 long-observed correlations between growth rate and cell size (??), and between growth rate and  
 371 ribosomal mass fraction. While both trends have sparked tremendous curiosity and driven sub-  
 372 stantial amounts of research in their own regards, these relationships are themselves intertwined.  
 373 In particular, it is the need for cells to increase their absolute number of ribosomes under con-  
 374 ditions of rapid growth that require cells to also grow in size. Further experiments are needed to  
 375 test the validity of this hypothesis. In particular, we believe that the change in growth rate in re-  
 376 sponse to translation-inhibitory drugs (such as chloramphenicol) could be quantitatively predicted,  
 377 given one had precision measurement of the relevant parameters, including the fraction of actively  
 378 translating ribosomes  $f_a$  and changes in the metabolic capacity of the cell (i.e. the parameter  $r_{AA}$   
 379 in our minimal model) for a particular growth condition.

380 While the generation of new ribosomes plays a dominant role in growth rate control, there exist

381 other physical limits to the function of cellular processes. One of the key motivations for consid-  
382 ering energy production was the physical constraints on total volume and surface area as cells  
383 vary their size (??). While *E. coli* get larger as it expresses more ribosomes, an additional constraint  
384 begins to arise in energy production due to a relative decrease in total surface area where ATP is  
385 predominantly produced (?). Specifically, the cell interior requires an amount of energy that scales  
386 cubically with cell size, but the available surface area only grows quadratically (??(A)). While this  
387 threshold does not appear to be met for *E. coli* cells growing at 2 hr<sup>-1</sup> or less, it highlights an ad-  
388 ditional constraint on growth given the apparent need to increase in cell size to grow faster. This  
389 is also potentially relevant to eukaryotic organisms, whose mitochondria exhibit convoluted mem-  
390 brane structures that nevertheless remain bacteria-sized organelles (?). In the context of bacteria  
391 growth and energy production more generally, we have limited our analysis to the aerobic growth  
392 conditions associated with the proteomic data and further consideration will be needed for anaer-  
393 obic growth.

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

# Supplemental material 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>

\*For correspondence:

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

<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.  
<sup>424</sup> Address correspondence to phillips@pboc.caltech.edu and jtheriot@uw.edu

425 **Contents**

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

Author	Method	Reported Quantity
Taniguchi <i>et al.</i> (2010)	YFP-fusion, cell fluorescence	fg/copies per cell
Valgepea <i>et al.</i> (2012)	mass spectrometry	fg/copies per cell
Peebo <i>et al.</i> (2014)	mass spectrometry	fg/copies per fl
Li <i>et al.</i> (2014)	ribosomal profiling	fg/copies per cell <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 ?? for more details on this).

## 426 **Summary of Proteome Data: Experimental Details**

427 Here we provide a brief summary of the experiments behind each proteomic data set. The purpose  
 428 of this section is to identify how the authors arrived at absolute protein abundances. In the  
 429 following section (Section ??) we will then provide a summary of the final protein abundance mea-  
 430 surements that were used throughout the main text. Table ?? provides an overview of the publi-  
 431 cations we considered. These are predominately mass spectrometry-based, with the exception of  
 432 the work from ? which used ribosomal profiling, and the fluorescence-based counting done in ?.

### 433 **Fluorescence based measurements**

434 In the work of ?, the authors used a chromosomal YFP fusion library where individual strains have a  
 435 specific gene tagged with a YFP-coding sequence. 1018 of their 1400 attempted strains were used  
 436 in the work. A fluorescence microscope was used to collect cellular YFP intensities across all these  
 437 strains. Through automated image analysis, the authors normalized intensity measurements by  
 438 cell size to account for the change in size and expression variability across the cell cycle. Follow-  
 439 ing correction of YFP intensities for cellular autofluorescence, final absolute protein levels were  
 440 determined by a calibration curve with single-molecule fluorescence intensities. This calibration  
 441 experiment was performed separately using a purified YFP solution.

### 442 **Ribosomal profiling measurements**

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

449 In the experiments, ribosome-protected mRNA is extracted from cell lysate and selected on  
 450 a denaturing polyacrylamide gel for deep sequencing (15–45 nt long fragments collected and se-  
 451 quenced by using an Illumina HiSeq 2000 in ?). Counts of ribosome footprints from the sequencing  
 452 data were then corrected empirically for position-dependent biases in ribosomal density across  
 453 each gene, as well as dependencies on specific sequences including the Shine-Dalgarno sequence.  
 454 These data-corrected ribosome densities represent relative protein synthesis rates. Absolute pro-  
 455 tein synthesis rates are obtained by multiplying the relative rates by the total cellular protein per  
 456 cell. The total protein per unit volume was determined with the Lowry method to quantify total

<sup>457</sup> protein, calibrated against bovine serum albumin (BSA). By counting colony-forming units following  
<sup>458</sup> serial dilution of their cell cultures, they then calculated the total protein per cell.

#### <sup>459</sup> Mass spectrometry measurements

<sup>460</sup> Perhaps not surprisingly, the data is predominantly mass spectrometry based. This is largely due  
<sup>461</sup> to tremendous improvements in the sensitivity of mass spectrometers, as well as improvements in  
<sup>462</sup> sample preparation and data analysis pipelines. It is now a relatively routine task to extract protein  
<sup>463</sup> from a cell and quantify the majority of proteins present by shotgun proteomics. In general, this  
<sup>464</sup> involves lysing cells, enzymatically digesting the proteins into short peptide fragments, and then  
<sup>465</sup> introducing them into the mass spectrometer (e.g. with liquid chromatography and electrospray  
<sup>466</sup> ionization), which itself can have multiple rounds of detection and further fragmentation of the  
<sup>467</sup> peptides.

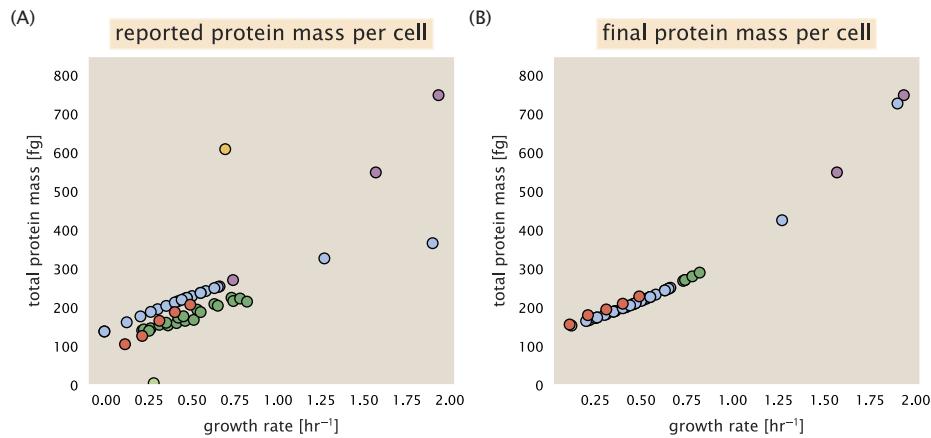
<sup>468</sup> Most quantitative experiments rely on labeling protein with stable isotopes, which allow multi-  
<sup>469</sup> ple samples to be measured together by the mass spectrometer. By measuring samples of known  
<sup>470</sup> total protein abundance simultaneously (i.e. one sample of interest, and one reference), it is pos-  
<sup>471</sup> sible to determine relative protein abundances. Absolute protein abundances can be estimated  
<sup>472</sup> following the same approach used above for ribosomal profiling, which is to multiply each relative  
<sup>473</sup> abundance measurement by the total cellular protein per cell. This is the approach taken by ? and  
<sup>474</sup> ?, with relative protein abundances determined based on the relative peptide intensities (label free  
<sup>475</sup> quantification 'LFQ' intensities). For the data of ?, total protein per cell was determined by measur-  
<sup>476</sup> ing total protein by the Lowry method, and counting colony-forming units following serial dilution.  
<sup>477</sup> For the data from ?, the authors did not determine cell quantities and instead report the cellular  
<sup>478</sup> protein abundances in protein per unit volume by assuming a mass density of 1.1 g/ml, with a 30%  
<sup>479</sup> dry mass fraction.

<sup>480</sup> An alternative way to arrive at absolute protein abundances is to dope in synthetic peptide  
<sup>481</sup> fragments of known abundance. These can serve as a direct way to calibrate mass spectrometry  
<sup>482</sup> signal intensities to absolute mass. This is the approach taken by ?. In addition to a set of shotgun  
<sup>483</sup> proteomic measurements to determine proteome-wide relative abundances, the authors also per-  
<sup>484</sup> formed absolute quantification of 41 proteins covering over four orders of magnitude in cellular  
<sup>485</sup> abundance. Here, a synthetic peptide was generated for each of the proteins, doped into each pro-  
<sup>486</sup> tein sample, and used these to determine absolute protein abundances of the 41 proteins. These  
<sup>487</sup> absolute measurements, determined for every growth condition, were then used as a calibration  
<sup>488</sup> curve to convert proteomic-wide relative abundances into absolute protein abundance per cell. A  
<sup>489</sup> more extensive discussion of the ? data set can be found in Section ??.

#### <sup>490</sup> Summary of Proteomic Data

<sup>491</sup> In the work of the main text we only used the data from ??. As shown in ??(A), the reported total  
<sup>492</sup> protein abundances in the work of ? and ? differed quite substantially from the other work. For  
<sup>493</sup> the work of ? this is in part due to a lower coverage in total proteomic mass quantified, though we  
<sup>494</sup> also noticed that most proteins appear undercounted when compared to the other data.

<sup>495</sup> ??(B) summarizes the total protein mass for each data point in our final compiled data set. We  
<sup>496</sup> note that protein abundances were all scaled so they followed a common growth rate-dependent  
<sup>497</sup> change in total protein mass. While our inclination initially was to leave reported copy numbers  
<sup>498</sup> untouched, a notable discrepancy in the scaling total protein per cell between ? and the other  
<sup>499</sup> data sets forced us to dig deeper into those measurements (compare ? and ? data in ??(A)). The  
<sup>500</sup> particular trend in ? appears to be due to assumptions of cell size and we provide a more extensive  
<sup>501</sup> discussion and analysis of that data set in section ?. As a compromise, and in an effort to treat  
<sup>502</sup> all data equally, we instead scaled all protein abundance values to a data-driven estimate of total  
<sup>503</sup> protein per cell. Here we used cell size measurements from ??, and an estimate of total protein  
<sup>504</sup> content through expected dry mass. Total protein per cell was estimated using available data on



**Figure 5. 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 ??????. Note that the data from ? only reported protein abundances per unit volume and total protein per cell was found by multiplying these by the growth-rate dependent cell size as determined by ?. (B) Adjusted total protein abundances across the proteomic data sets are summarized. Protein abundances were adjusted so that all data shared a common set of growth-rate dependent total protein per cell and cellular protein concentration following the cell size expectations of ? (see section on ?? for further details).

505 total DNA, RNA, and protein from ??, which account for the majority of dry mass in the cell. We  
506 consider these details in sections ?? and ?? that follows.

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

### 515 Estimation of Cell Size, Surface Area

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

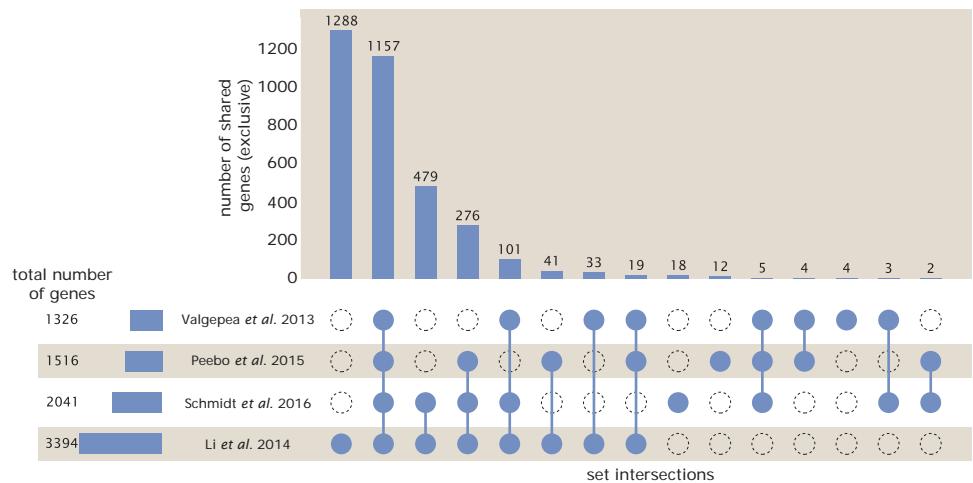
521 The average size measurements from each of their experiments are shown in Figure ??, with  
522 cell length and width shown in (A) and (B), respectively. The length data was well described by the  
523 exponential function  $0.5 e^{1.09 \cdot \lambda} + 1.76 \mu\text{m}$ , while the width data was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

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

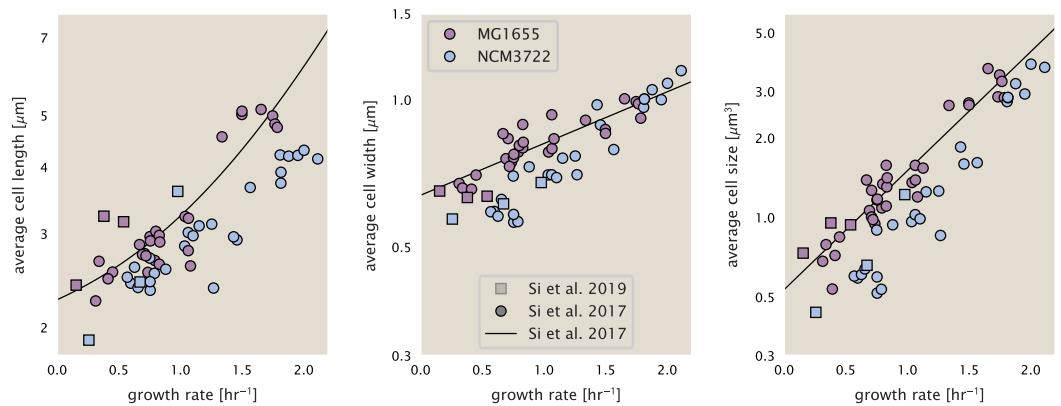
526 where  $r$  is half the cell width. A best fit to the data is described by  $0.533 e^{1.037 \cdot \lambda} \mu\text{m}^3$ . Calculation of  
527 the cell surface area is given by,

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

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



**Figure 6. Comparison of proteomic coverage across different data sets.** An UpSet diagram (?) summarizes the total number of protein coding genes whose protein abundance was reported in the data sets of ?????. The total number of genes reported in each individual data set are noted on the bottom left, while their overlap is summarized in the bar chart. Each column here refers to the intersection of a set of data sets identified by blue circles.



**Figure 7. Summary of size measurements from Si et al. 2017, 2019.** Cell lengths and widths were measured from cell contours obtained from phase contrast images, and refer to the long and short axis respectively. (A) Cell lengths and (B) cell widths show the mean measurements reported (they report 140-300 images and 5,000-30,000 for each set of samples; which likely means about 1,000-5,000 measurements per mean value reported here since they considered about 6 conditions at a time). Fits were made to the MG1655 strain data; length:  $0.5 e^{1.09 \cdot \lambda} + 1.76 \mu\text{m}$ , width:  $0.64 e^{0.24 \cdot \lambda} \mu\text{m}$ . (C) Cell size,  $V$ , was calculated as cylinders with two hemispherical ends (Equation ??). The MG1655 strain data gave a best fit of  $0.533 e^{1.037 \cdot \lambda} \mu\text{m}^3$ .

529 **Estimation of Total Protein Content per Cell**

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

538 However, even if dry mass concentration is relatively constant across growth conditions, it is  
 539 not obvious how protein concentration might vary due to the substantial increase in rRNA at faster  
 540 growth rates (?). This is a well-documented result that arises from an increase in ribosomal abun-  
 541 dance at faster growth rates (?). To proceed therefore rely on experimental measurements of total  
 542 DNA content per cell that also come from Basan *et al.*, and RNA to protein ratios that were mea-  
 543 sured in Dai *et al.* (and cover the entire range of growth conditions considered here). These are  
 544 reproduced in Figure ??(A) and (B), respectively.

545 Assuming that the protein, DNA, and RNA account for 90 % of the total dry mass, the protein  
 546 mass can then determined by first subtracting the experimentally measured DNA mass, and then  
 547 using the experimental estimate of the RNA to protein ratio. The total protein per cell is will be  
 548 related to the summed RNA and protein mass by,

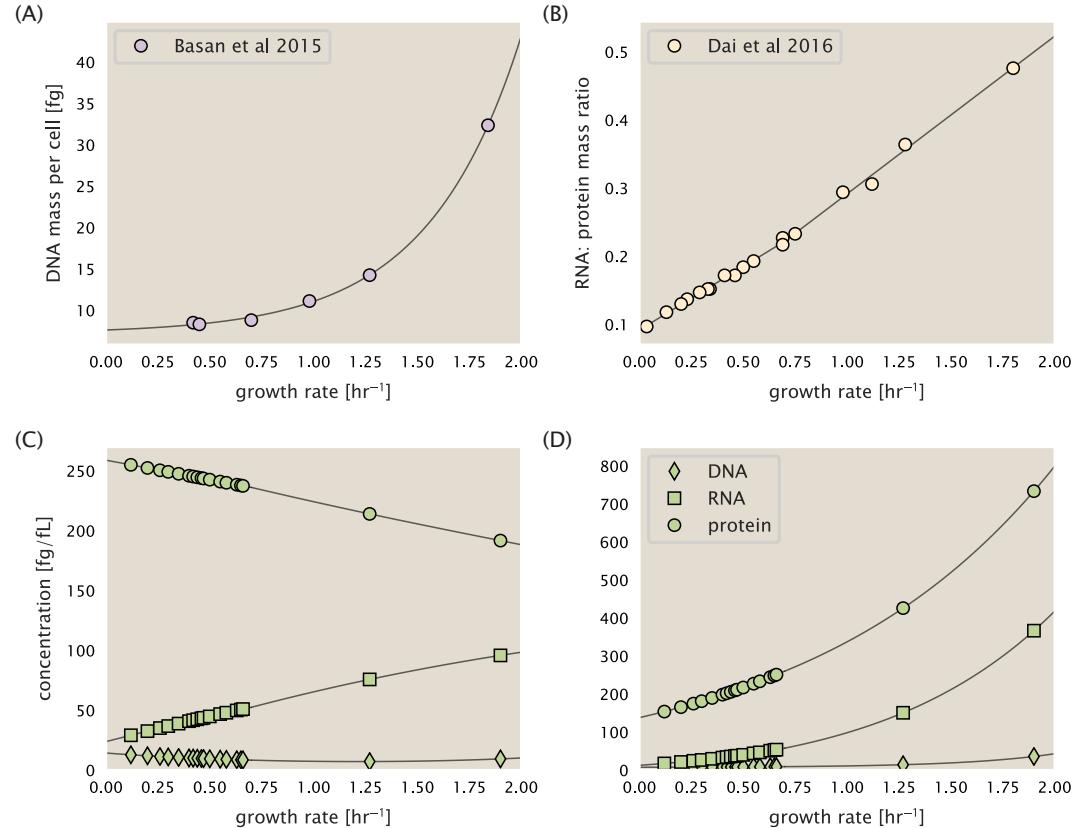
$$M_P = \frac{[M_P + M_{RNA}]}{1 + (RP_{ratio})}. \quad (3)$$

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

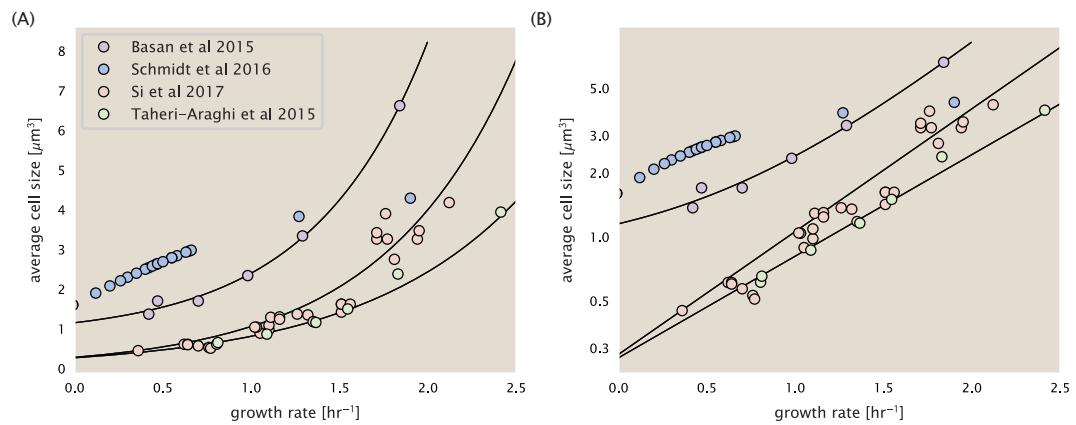
554 **Additional Considerations of Schmidt *et al.* Data Set**

555 While the data set from ? remains a heroic effort that our labs continue to return to as a resource,  
 556 there were steps taken in their calculation of protein copy number that we felt needed further con-  
 557 sideration. In particular, the authors made an assumption of constant cellular protein concentra-  
 558 tion across all growth conditions and used measurements of cell volume that appear inconsis-  
 559 tent with an expected exponential scaling of cell size with growth rate that is well-documented in *E. coli*  
 560 (??).

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



**Figure 8. 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  $0.42 e^{2.23\lambda} + 7.2$  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 ?.



**Figure 9. Measurements of cell size as a function of growth rate.** (A) Plot of the reported cell sizes from several recent papers. The data in blue come from Volkmer and Heinemann, 2011 (?) and were used in the work of Schmidt *et al.*. Data from the lab of Terence Hwa are shown in purple (?), while the two data sets shown in green and red come from the lab of Suckjoon Jun (??). (B) Same as in (A) but with the data plotted on a logarithmic y-axis to highlight the exponential scaling that is expected for *E. coli*.

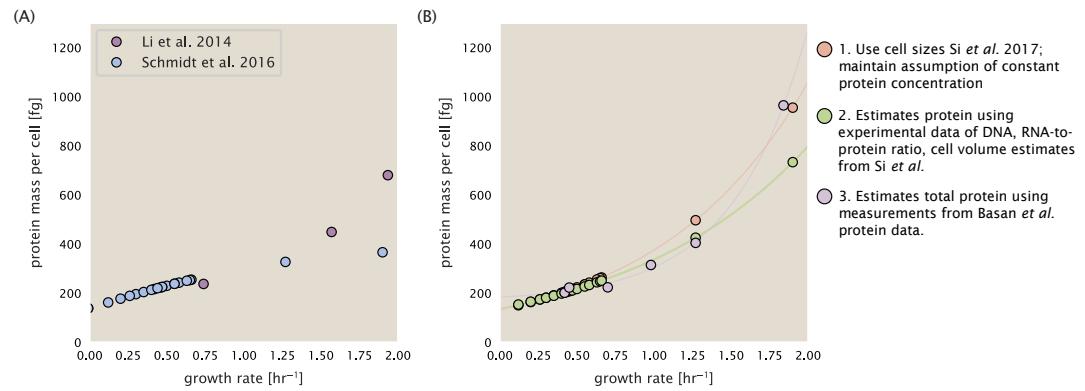
575 Since it is not obvious how measurements of cell size influenced their reported protein abundances,  
 576 in the following subsections we begin by considering this calculation. We then consider  
 577 three different approaches to estimate the growth-rate dependent total protein mass to compare  
 578 with those values reported from ?. The results of this are summarized in ??(B), with the original val-  
 579 ues from both ? and ? shown in ??(A) for reference. For most growth conditions, we find that total  
 580 protein per cell is still in reasonable agreement. However, for the fastest growth conditions, with  
 581 glycerol + supplemented amino acids, and LB media, all estimates are substantially higher than  
 582 those originally reported. This is the main reason why we chose to readjusted protein abundance  
 583 as shown in ??(B) (with the calculation described in section ??).

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

585 As noted in section ??, the authors calculated proteome-wide protein abundances by first deter-  
 586 mining absolute abundances of 41 pre-selected proteins, which relied on adding synthetic heavy  
 587 reference peptides into their protein samples at known abundance. This absolute quantitation was  
 588 performed in replicate for each growth condition. Separately, the authors also performed a more  
 589 conventional mass spectrometry measurement for samples from each growth condition, which  
 590 attempted to maximize the number of quantified proteins but only provided relative abundances  
 591 based on peptide intensities. Finally, using their 41 proteins with absolute abundances already  
 592 determined, they then created calibration curves with which to relate their relative intensity to  
 593 absolute protein abundance for each growth condition. This allowed them to estimate absolute  
 594 protein abundance for all proteins detected in their proteome-wide data set. Combined with their  
 595 flow cytometry cell counts, they were then able to determine absolute abundance of each protein  
 596 detected on a per cell basis.

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

605 For cells grown in M9 minimal media + glucose an average total mass of  $M_p = 240$  fg per cell was



**Figure 10. Alternative estimates of total cellular protein for the growth conditions considered in Schmidt et al.** (A) The original protein mass from Schmidt et al. and Li et al. are shown in purple and blue, respectively. (B) Three alternative estimates of total protein per cell. 1. *light red*: Rescaling of total protein mass assuming a growth rate independent protein concentration and cell volumes estimated from Si et al. 2017. 2. *light green*: Rescaling of total protein mass using estimates of growth rate-dependent protein concentrations and cell volumes estimated from Si et al. 2017. Total protein per cell is calculated by assuming a 1.1 g/ml cellular mass density, 30% dry mass, with 90% of the dry mass corresponding to DNA, RNA, and protein (?). See ?? for details on calculation. 3. *light purple*: Rescaling of total protein mass using the experimental measurements from Basan et al. 2015.

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,

$$M_{P,i} = [M_P]_{orig} \cdot V_i \quad (4)$$

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

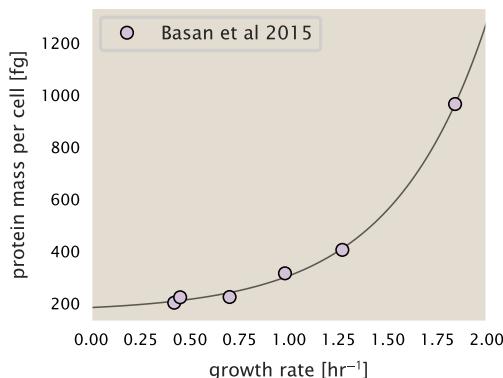
If we instead consider the cell volumes predicted in the work of Si et al., we again need to take growth in M9 minimal media + glucose as a reference with known total mass, but we can follow a similar approach to estimate total protein mass for all other growth conditions. Letting  $V_{Si,glu} = 0.6$  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,

$$M'_{P,i} = [M_P]_{Si} \cdot V_{Si,i} \quad (5)$$

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 ??(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 ?? for all data points in the ? data set. The green data points in ??(B) show this prediction, and this represents the approach used to estimate total protein per cell for all data sets.



**Figure 11. 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  $hr^{-1}$ .

### 629     **Experimental measurements of total protein from Basan *et al.* 2015.**

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

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

### 650     **Calculation of Complex Abundance**

651     All data collected quantified the abundance of individual proteins with high resolution. After cor-  
 652     recting for errors introduced from overestimated volumes and imposed boundaries on the protein  
 653     concentration, we are left with a large data set, largely comparable between one another. How-  
 654     ever, this work is focused on estimating the abundance of individual protein *complexes*, rather  
 655     than copies of individual proteins. In this section, we outline the procedure we used to annotate  
 656     proteins as being part of a macromolecular complex as well as how we computed their absolute  
 657     abundance.

658     Protein complexes, and proteins individually, often have a variety of names, both longform  
 659     and shorthand. As individual proteins can have a variety of different synonyms, we sought to en-  
 660     sure that each protein annotated in the data sets used the same synonym. To do use, we relied

661 heavily on the EcoCyc Genomic Database (?). Each protein in available data sets included an anno-  
 662 tation of one of the gene name synonyms as well as an accession ID – either a UniProt or Blattner  
 663 “b-number”. We programmatically matched up individual accession IDs between the proteins in  
 664 different data sets. In cases where accession IDs matched but the gene names were different, we  
 665 manually verified that the gene product was the same between the datasets and chose a single  
 666 synonym. All code used in the data cleaning and unification procedures can be found on the asso-  
 667 ciated [GitHub repository] (DOI:XXX) associated with this paper as well as on the associated [paper](#)  
 668 [website](#).

669 With each protein product in the data sets conforming to a single identification scheme, we  
 670 were tasked to identify the molecular complexes each protein was a member of. Additionally, we  
 671 needed to identify how many copies of each protein were present in each complex (i.e. the sub-  
 672 unit copy number) and compute the estimated abundance complex that accounted for fluctuations  
 673 in subunit stoichiometry. To map proteins to complexes, we programmatically accessed the Eco-  
 674 Cyc *E. coli* database ? using PathwayTools version 23.0 ?. With a license for PathWay Tools, we  
 675 programmatically mapped each unique protein to its annotated complexes via the BioCyc Python  
 676 package. As we mapped each protein with *all* of its complex annotations, there was redundancy  
 677 in the dataset. For example, ribosomal protein L20 (RplT) is annotated to be a component of the  
 678 50S ribosome (EcoCyc complex CPLX-03962) as well as a component of the mature 70S ribosome  
 679 (EcoCyc complex CPLX-03964).

680 In addition to the annotated complex, we collected information of how many copies of each  
 681 individual protein is present in each macromolecular complex. With this number in hand, we cal-  
 682 culated the maximum number of complexes that *could* be formed given the observed abundance  
 683 of each protein subunit as

$$N_{\text{complex}}^{(\max)}(\text{subunit}) = \frac{N_{\text{subunit}}^{(\text{observed})}}{N_{\text{subunit}}^{(\text{annotated})}}. \quad (6)$$

684 For example, the 70S mature ribosome complex has 55 protein components, all of which are  
 685 present in a single copy except L4 (RplL), which is present in 4 copies. For each ribosomal pro-  
 686 tein, we then calculate the maximum number of complexes that could be formed using ???. This  
 687 example, along with example from 5 other macromolecular complexes, can be seen in ??.

688 It is important to note that measurement noise, efficiency of protein extraction, stochastic er-  
 689 rors will mean that the precise value of each calculation will be different for each component of a  
 690 given complex. Thus, to report the total complex abundance, we computed the arithmetic mean  
 691 of  $N_{\text{complex}}^{(\max)}$  for all subunits as

$$\langle N_{\text{complex}} \rangle = \frac{1}{N_{\text{subunits}}} \sum_i^{N_{\text{subunits}}} \frac{N_i^{(\text{observed})}}{N_i^{(\text{annotated})}}. \quad (7)$$

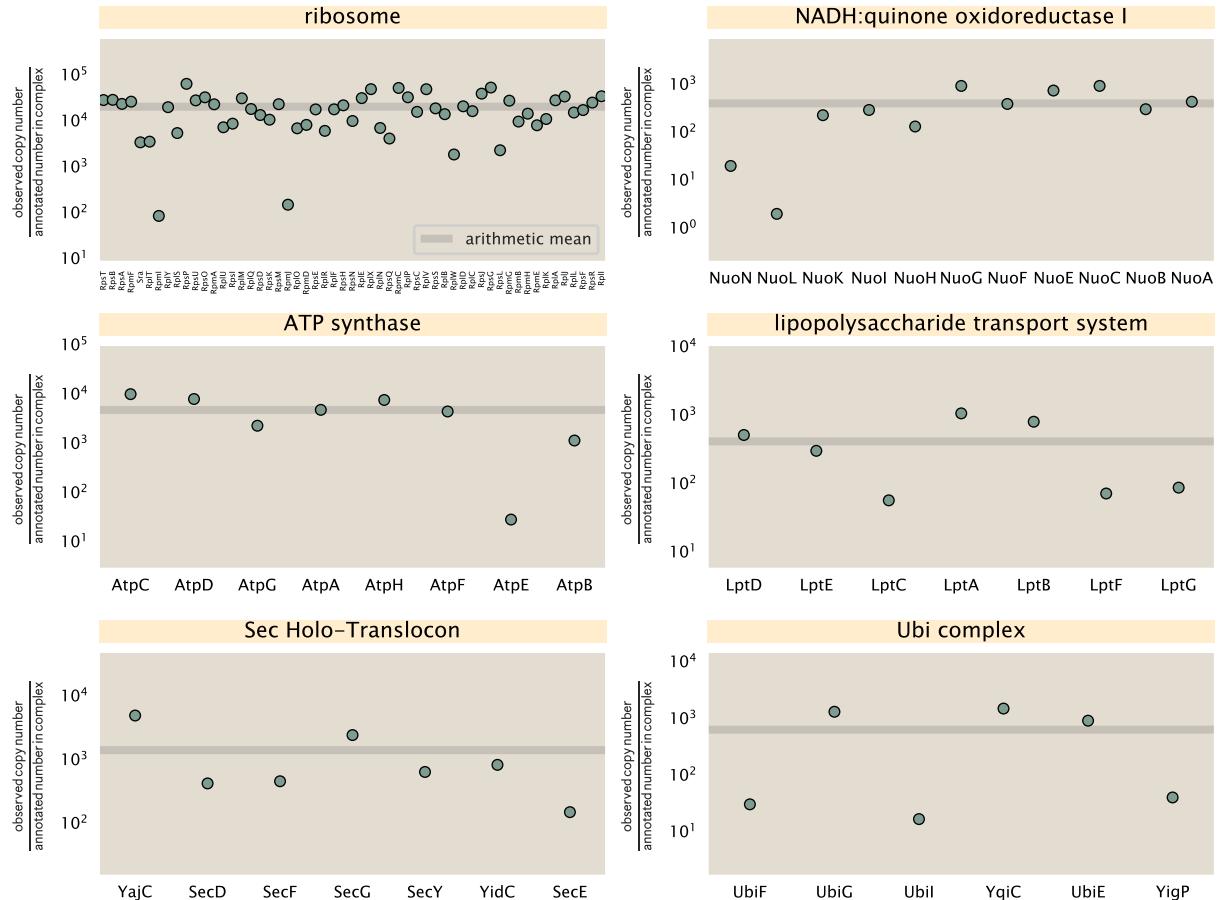
692 in ??, we show this mean value as a grey line for a variety of different complexes. Additionally, we  
 693 have built an interactive figure accessible on the [paper website](#) where the validity of this approach  
 694 can be examined for any complex with more than two subunits (thus, excluding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

### 696 Extending Estimates to a Continuum of Growth Rates

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

### 702 Estimation of the total cell mass

703 For many of the processes estimated in the main text we relied on a cellular dry mass of ≈ 300  
 704 fg from which we computed elemental and protein fractions using knowledge of fractional com-



**Figure 12. Calculation of the mean complex abundance from measurements of single subunits.** Six of the largest complexes (by number of subunits) in *E. coli*. Points correspond to the maximum number of complexes that can be formed given measurement of that individual protein. Solid grey line corresponds to the arithmetic mean across all subunits. These data correspond to measurements from ? in a glucose-supplemented minimal growth medium.

705 position of the dry mass. At modest growth rates, such as the 5000 s doubling time used in the  
 706 main text, this is a reasonable number to use as the typical cell mass is  $\approx 1$  pg and *E. coli* cells can  
 707 approximated as 70% water by volume. However, as we have shown in the preceding sections, the  
 708 cell size and therefore cell volume is highly dependent on the growth rate. This means that a dry  
 709 mass of 300 fg cannot be used reliably across all growth rates.

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

$$m_{\text{cell}} \approx \rho V(\lambda) \approx \rho a e^{A*b} \quad (8)$$

713 where  $a$  and  $b$  are constants with units of  $\mu\text{m}^3$  and hr, respectively. The value of these constants  
 714 can be estimated from the careful volume measurements performed by ??, as is described in the  
 715 previous section.

### 716 Complex Abundance Scaling With Cell Volume

717 Several of the estimates performed in the main text are implicitly dependent on the cell volume.  
 718 This includes processes such as ATP synthesis and, most prominently, the transport of nutrients.  
 719 Of the latter, we estimated the number of transporters that would be needed to shuttle enough  
 720 carbon, phosphorus, and sulfur across the membrane to build new cell mass. To do so, we used  
 721 elemental composition measurements combined with a 300 fg cell dry mass to make the point  
 722 estimate. As we now have a means to estimate the total cell mass as a function of volume, we can  
 723 generalize these estimates across growth rates.

724 Rather than discussing the particular details of each transport system, we will derive this scaling  
 725 expression in very general terms. Consider we wish to estimate the number of transporters for  
 726 some substance  $X$ , which has been measured to be make up some fraction of the dry mass  $\theta_X$ . If  
 727 we assume that, irrespective of growth rate, the cell dry mass is  $\approx 30\%$  of the total cell mass, we  
 728 can state that the total mass of substance  $X$  as a function of growth rate is

$$m_X \approx 0.3 \times \rho V(\lambda) \theta_X, \quad (9)$$

729 where we have used  $\rho V(\lambda)$  as an estimate of the total cell mass, defined in ???. To convert this to  
 730 the number of units  $N_X$  of substance  $X$  in the cell, we can use the formula weight  $w_X$  of a single  
 731 unit of  $X$  in conjunction with ??,

$$N_X \approx \frac{m_X}{w_X}. \quad (10)$$

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

$$N_{\text{transporters}}(\lambda) = \frac{0.3 \times \rho V(\lambda) \theta_X}{w_X r_X \tau}. \quad (11)$$

739 This function is used to draw the continuum estimates for the number of transporters seen in  
 740 Figures 2 and 3 as transparent grey curves. Occasionally, this continuum scaling relationship will  
 741 not precisely agree with the point estimate outlined in the main text. This is due to the fact that we  
 742 make an initial approximation made of a dry cell mass of  $\approx 300$  fg for the point estimate while we  
 743 consider more precise values in the continuum estimate. We note, however, that both this scaling  
 744 relation and the point estimates are meant to describe the order-of-magnitude observed, and not  
 745 the predict the exact values of the abundances.

746 ?? is a very general relation for processes where the cell volume is the "natural variable" of the  
 747 problem. This means that, as the cell increases in volume, the requirements for substance  $X$  also

scale with volume rather than scaling with surface area, for example. So long as the rate of the process, the fraction of the dry mass attributable to the substance, and the formula mass of the substance is known, ?? can be used to compute the number of complexes needed. For example, to compute the number of ATP synthases per cell, ?? can be slightly modified to the form

$$N_{\text{ATP synthases}}(\lambda) = \frac{0.3 \times \rho V(\lambda) \theta_{\text{protein}} N_{\text{ATP}}}{w_{AA} r_{\text{ATP}} \tau}, \quad (12)$$

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

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

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

#### 764 Number of Lipids

765 To compute the number of lipids as a function of growth rate, we make the assumption that some features, such as 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 of lipids ( $\approx 40\%$ , BNID: 100078) are independent of the growth rate. Using these approximations combined with ??, and recognizing that each membrane is composed of two leaflets, we can compute the number of lipids as a function of growth rate as

$$N_{\text{lipids}}(\lambda) \approx \frac{4 \text{ leaflets} \times 0.4 \times \eta \pi \left( \frac{\eta \pi}{4} - \frac{\pi}{12} \right)^{-2/3} V(\lambda)^{2/3}}{A_{\text{lipid}}} \quad (13)$$

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

#### 771 Number of Murein Monomers

772 In calculation of the number of transpeptidases needed for maturation of the peptidoglycan, we used an empirical measurement that  $\approx 3\%$  of the dry mass is attributable to peptidoglycan and that 774 a single murien monomer is  $m_{\text{murein}} \approx 1000 \text{ Da}$ . While the latter is independent of growth rate, the former is not. As the peptidoglycan exists as a thin shell with a width of  $w \approx 10 \text{ nm}$  encapsulating 776 the cell, one would expect the number of murein monomers scales with the surface area of this 777 shell. In a similar spirit to our calculation of the number of lipids, the total number of murein 778 monomers as a function of growth rate can be calculated as

$$N_{\text{murein monomers}}(\lambda) \approx \frac{\rho_{\text{pg}} w \eta \pi \left( \frac{\eta \pi}{4} - \frac{\pi}{12} \right)^{-2/3} V(\lambda)^{2/3}}{m_{\text{murein}}}, \quad (14)$$

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

### 780 Complex Abundance Scaling With Number of Origins

781 While the majority of our estimates hinge on the total cell volume or surface area, processes related 782 to the central dogma, namely DNA replication and synthesis of rRNA, depend on the number of 783 chromosomes present in the cell. As discussed in the main text, the ability of *E. coli* to parallelize the 784 replication of its chromosome by having multiple active origins of replication at a given is critical to

<sup>785</sup> synthesize enough rRNA, especially at fast growth rates. Derived in ? and reproduced in the main  
<sup>786</sup> text, the average number of origins of replication at a given growth rate can be calculated as

$$\langle \#ori \rangle \approx 2^{t_{cyc} \lambda / \ln 2} \quad (15)$$

<sup>787</sup> where  $t_{cyc}$  is the total time of replication and division. We can make the approximation that  $t_{cyc} \approx$   
<sup>788</sup> 70 min, which is the time it takes two replisomes to copy an entire chromosome.

<sup>789</sup> In the case of rRNA synthesis, the majority of the rRNA operons are surrounding the origin of  
<sup>790</sup> replication. Thus, at a given growth rate  $\lambda$ , the average dosage of rRNA operons per cell  $D_{rRNA}$  is

$$D_{rRNA}(\lambda) \approx N_{rRNA \text{ operons}} \times 2^{t_{cyc} \lambda / \ln 2}. \quad (16)$$

<sup>791</sup> This makes the approximation that all rRNA operons are localized around the origin. In reality,  
<sup>792</sup> the operons are some distance away from the origin, making ?? an approximation.

<sup>793</sup> In the main text, we stated that at the growth rate in question, there is  $\approx 1$  chromosome per  
<sup>794</sup> cell. While a fair approximation, ?? illustrates that is not precisely true, even at slow growth rates.  
<sup>795</sup> In estimating the number of RNA polymerases as a function of growth rate, we consider that re-  
<sup>796</sup> gardless of the number of rRNA operons, they are all sufficiently loaded with RNA polymerase such  
<sup>797</sup> that each operon produces one rRNA per second. Thus, the total number of RNA polymerase as a  
<sup>798</sup> function of the growth rate can be calculated as

$$N_{\text{RNA polymerase}}(\lambda) \approx L_{\text{operon}} D_{rRNA} \rho_{\text{RNA polymerase}}, \quad (17)$$

<sup>799</sup> where  $L_{\text{operon}}$  is the total length of an rRNA operon ( $\approx 4500$  bp) and  $\rho_{\text{RNA polymerase}}$  is packing density  
<sup>800</sup> of RNA polymerase on a given operon, taken to be 1 RNA polymerase per 80 nucleotides.