Supplemental material for: Fundamental limits on the rate of bacterial cell division

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

***For correspondence:**

[†]These authors contributed equally to this work

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

16 ## Contents

Table 1. Overview of proteomic data sets.

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

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

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

Summary of Proteome Data: Experimental Details

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

Fluorescence based measurements

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

Ribosomal profiling measurements

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

In the experiments, ribosome-protected mRNA is extracted from cell lysate and selected on a denaturing polyacrylamide gel for deep sequencing (15–45 nt long fragments collected and sequenced by using an Illumina HiSeq 2000 in ***Li et al.*** (***2014***)). Counts of ribosome footprints from the sequencing data were then corrected empirically for position-dependent biases in ribosomal density across each gene, as well as dependencies on specific sequences including the Shine-Dalgarno sequence. These data-corrected ribosome densities represent relative protein synthesis rates. Absolute protein synthesis rates are obtained by multiplying the relative rates by the total cellular

protein per cell. The total protein per unit volume was determined with the Lowry method to quantify total protein, calibrated against bovine serum albumin (BSA). By counting colony-forming units following serial dilution of their cell cultures, they then calculated the total protein per cell.

Mass spectrometry measurements

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

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

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

Summary of Proteomic Data

In the work of the main text we only used the data from ***Valgepea et al.*** (***2013***); ***Li et al.*** (***2014***); ***Peebo et al.*** (***2015***); ***Schmidt et al.*** (***2016***). As shown in ***Figure 1***(A), the reported total protein abundances in the work of ***Taniguchi et al.*** (***2010***) and ***Soufi et al.*** (***2015***) differed quite substantially from the other work. For the work of ***Taniguchi et al.*** (***2010***) this is in part due to a lower coverage in total proteomic mass quantified, though we also noticed that most proteins appear undercounted when compared to the other data.

Figure 1(B) summarizes the total protein mass for each data point in our final compiled data set. We note that protein abundances were all scaled so they followed a common growth rate-dependent change in total protein mass. While our inclination initally was to leave reported copy numbers untouched, a notable descrepency in the scaling total protein per cell between ***Schmidt et al.*** (***2016***) and the other data sets forced us to dig deeper into those measurements (compare ***Schmidt et al.*** (***2016***) and ***Li et al.*** (***2014***) data in ***Figure 1***(A)). The particular trend in ***Schmidt et al.*** (***2016***) appears to be due to assumptions of cell size and we provide a more extensive discussion

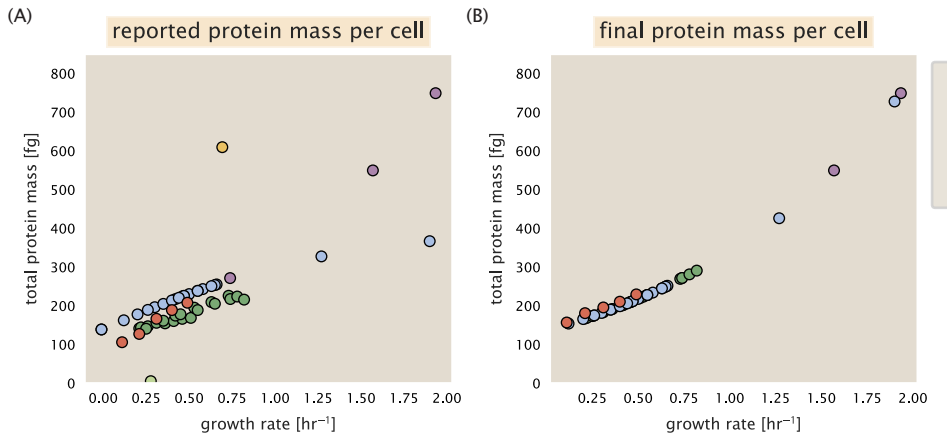

Figure 1. Summary of the growth-rate dependent total protein abundance for each data set. (A) Total protein abundance per cell as original reported in the data sets of ***Taniguchi et al.*** (***2010***); ***Valgepea et al.*** (***2013***); ***Li et al.*** (***2014***); ***Soufi et al.*** (***2015***); ***Peebo et al.*** (***2015***); ***Schmidt et al.*** (***2016***). Note that the data from ***Peebo et al.*** (***2015***) only reported protein abundances per unit volume and total protein per cell was found by multiplying these by the growth-rate dependent cell size as determined by ***Si et al.*** (***2017***). (B) Adjusted total protein abundances across the proteomic data sets are summarized. Protein abundances were adjusted so that all data shared a common set of growth-rate dependent total protein per cell and cellular protein concentration following the cell size expectations of ***Si et al.*** (***2017***) (see section on Estimation of Cell Size, Surface Area for further details).

and analysis of that data set in section Additional Considerations of Schmidt *et al.* Data Set. As a compromise, and in an effort to treat all data equally, we instead scaled all protein abundance values to a data-driven estimate of total protein per cell. Here we used cell size measurements from ***Si et al.*** (***2017***); **?**, and an estimate of total protein content through expected dry mass. Total protein per cell was estimated using available data on total DNA, RNA, and protein from ***Basan et al.*** (***2015***); ***Dai et al.*** (***2016***), which account for the majority of dry mass in the cell. We consider these details in sections Estimation of Cell Size, Surface Area and Estimation of Total Protein Content per Cell that follows.

Lastly, in ***Figure 2*** we show the total proteomic coverage and overlap of proteins quantified across each data set. Here we have used an UpSet diagram (***Lex et al., 2014***) to compare the data. Overall, the overlap in quantified proteins is quite high, with 1157 proteins quantified across all data sets. The sequencing based approach of ***Li et al.*** (***2014***) has substantially higher coverage compared to the mass spectrometry data sets (3394 genes versus the 2041 genes quantified in the work of ***Schmidt et al.*** (***2016***)). However, in terms of total protein mass, the data from ***Li et al.*** (***2014***); ***Schmidt et al.*** (***2016***); ***Peebo et al.*** (***2015***) each quantify roughly equivalent total protein mass. An exception to this is in the data from ***Valgepea et al.*** (***2013***), where we find that the total protein quantified in ***Valgepea et al.*** (***2013***) is 90-95 % of the total protein mass (when using the data from ***Schmidt et al.*** (***2016***) as a reference).

Estimation of Cell Size, Surface Area

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

The average size measurements from each of their experiments are shown in Figure ***Figure 3***, with cell length and width shown in (A) and (B), respectively. The length data was well described by

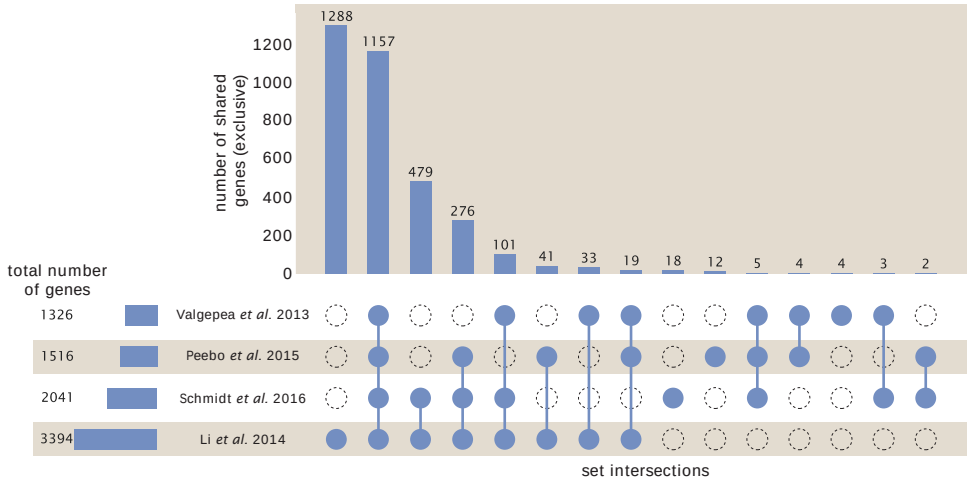

Figure 2. Comparison of proteomic coverage across different data sets. An UpSet diagram (***Lex et al., 2014***) summarizes the total number of protein coding genes whose protein abundance was reported in the data sets of ***Valgepea et al.*** (***2013***); ***Li et al.*** (***2014***); ***Schmidt et al.*** (***2016***); ***Peebo et al.*** (***2015***). The total number of genes reported in each individual data set are noted on the bottom left, while their overlap is summarized in the bar chart. Each column here refers to the intersection of a set of data sets identified by blue circles.

the exponential function 0.5 $e^{1.09 \cdot \lambda}$ + 1.76 μm, while the width data was well described by 0.64 $e^{0.24 \cdot \lambda}$ μm. In order to estimate cell size we take the cell as a cylinders with two hemispherical ends (***Si et al., 2017***; ***Basan et al., 2015***). Specifically, cell size is estimated from,

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

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

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

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

Estimation of Total Protein Content per Cell

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

However, even if dry mass concentration is relatively constant across growth conditions, it is not obvious how protein concentration might vary due to the substantial increase in rRNA at faster growth rates (***Dai et al.*** (***2016***)). This is a well-documented result that arises from an increase in ribosomal abundance at faster growth rates (***Scott et al.*** (***2010***)). To proceed therefore rely on experimental measurements of total DNA content per cell that also come from Basan *et al.*, and RNA to protein ratios that were measured in Dai *et al.* (and cover the entire range of growth conditions considered here). These are reproduced in Figure ***Figure 4***(A) and (B), respectively.

Assuming that the protein, DNA, and RNA account for 90 % of the total dry mass, the protein mass can then determined by first subtracting the experimentally measured DNA mass, and then

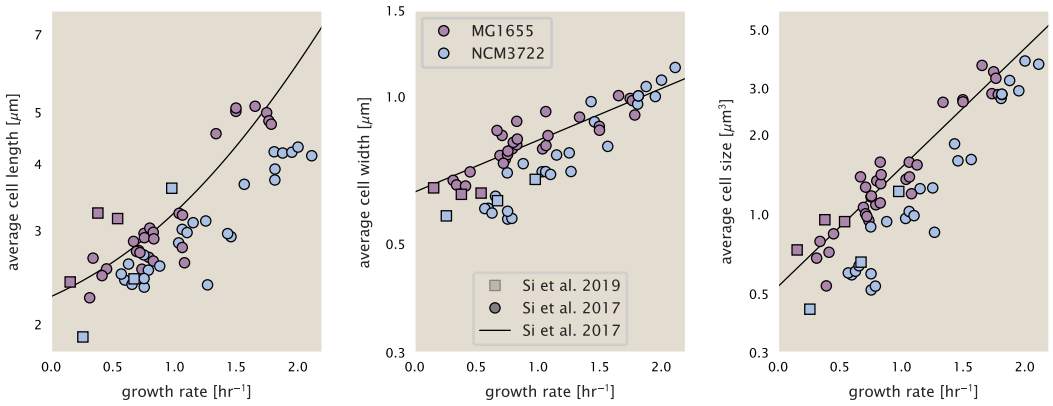

Figure 3. Summary of size measurements from Si *et al.* 2017, 2019. Cell lengths and widths were measured from cell contours obtained from phase contrast images, and refer to the long and short axis respectively. (A) Cell lengths and (B) cell widths show the mean measurements reported (they report 140-300 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 μm, width: 0.64 $e^{0.24 \cdot \lambda}$ μm. (C) Cell size, V, was calculated as cylinders with two hemispherical ends (Equation 1). The MG1655 strain data gave a best fit of 0.533 $e^{1.037 \cdot \lambda}$ μm^3.

using the experimental estimate of the RNA to protein ratio. The total protein per cell is will be related to the summed RNA and protein mass by,

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

(RP_{ratio} refers to the RNA to protein ratio as measured by Dai *et al.*. In Figure ***Figure 4***(C) we plot the estimated cellular concentrations for protein, DNA, and RNA from these calculations, and in Figure ***Figure 4***(D) we plot their total expected mass per cell. This later quantity is the growth rate-dependent total protein mass that was used to extimate total protein abundance across all data sets (and summaried in ***Figure 1***(B)).

Additional Considerations of Schmidt *et al.* Data Set

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

We begin by looking at their cell volume measurements, which are shown in blue in Figure ***Figure 5***. As a comparison, we also plot cell sizes reported in three other recent papers: measurements from Taheri-Araghi *et al.* and Si *et al.* come from the lab of Suckjoon Jun, while those from Basan *et al.* come from the lab of Terence Hwa. Each set of measurements used microscopy and cell segmentation to determine the length and width, and then calculated cell size by treating the cell is a cylinder with two hemispherical ends, as we considered in the previous section. While there is notable discrepancy between the two research groups, which are both using strain NCM3722, Basan *et al.* found that this came specifically from uncertainty in determining the cell width. This is prone to inaccuracy given the small cell size and optical resolution limits (further described in their supplemental text). Perhaps the more concerning point is that while each of these alternative measurements show an exponential increase in cell size at faster growth rates, the measurements used by Schmidt *et al.* appear to plateau. This resulted in an analogous trend in their final reported total

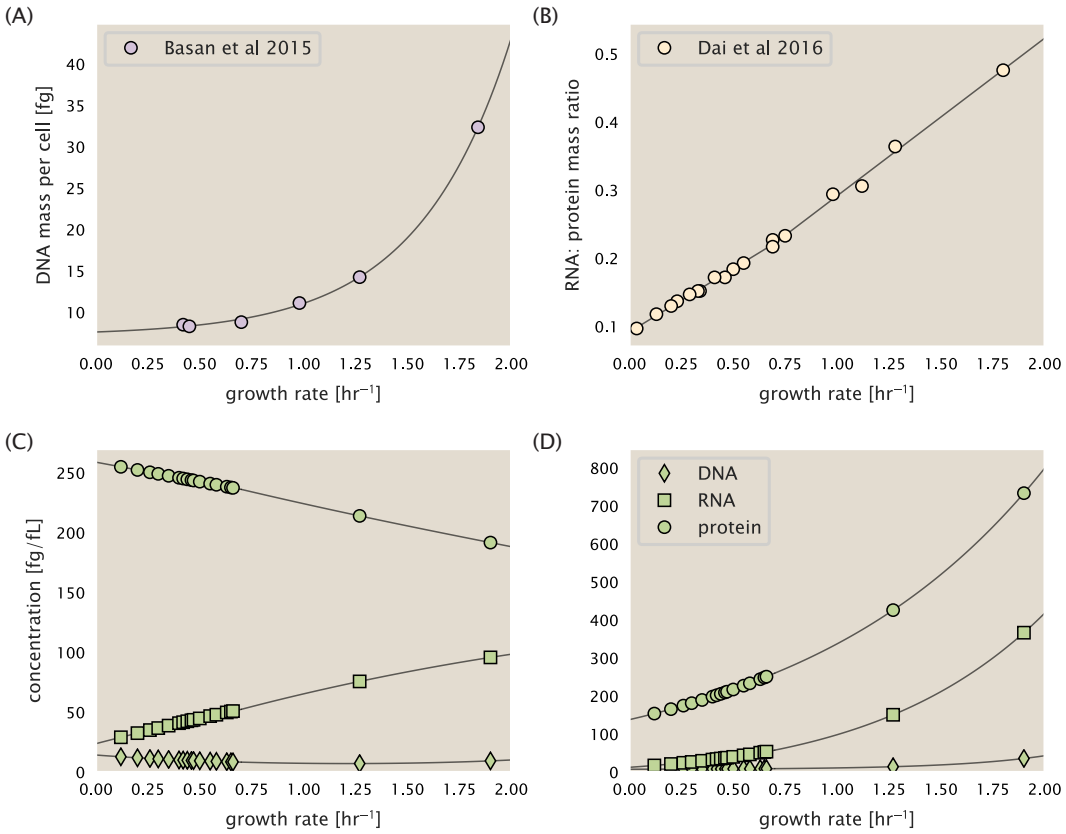

Figure 4. Empirical estimate of cellular protein, DNA, and RNA as a function of growth rate. (A) Measured DNA mass per cell as a function of growth rate, reproduced from Basan *et al.* 2015. The data was fit to an exponential curve (DNA mass in fg per cell is given by 0.42 $e^{2.23 \cdot \lambda}$ + 7.2 fg per cell, where λ is the growth rate in hr^{-1}). (B) RNA to protein measurements as a function fo growth rate. The data was for to two lines: for growth rates below 0.7 hr^{-1}, the RNA/protein ratio is $0.18 \cdot \lambda + 0.093$, while for growth rates faster than 0.7 hr^{-1} the RNA/protein ratio is given by $0.25 \cdot \lambda + 0.035$. For (A) and (B) cells are grown under varying levels of nutrient limitation, with cells grown in minimal media with different carbon sources for the slowest growth conditions, and rich-defined media for fast growth rates. (C) Predictions of cellular protein, DNA, and RNA concentration. (D) Total cellular mass predicted for protein, DNA, and RNA using the cell size predictions from Si *et al.*. Symbols (diamond: DNA, square: RNA, circle: protein) show estimated values of mass concentration and mass per cell for the specific growth rates in ***Schmidt et al.*** (***2016***).

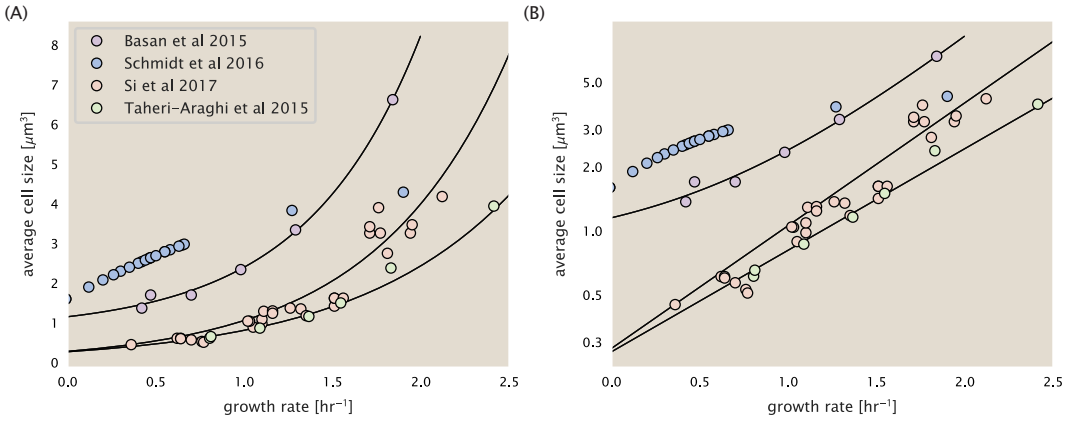

Figure 5. Measurements of cell size as a function of growth rate. (A) Plot of the reported cell sizes from several recent papers. The data in blue come from Volkmer and Heinemann, 2011 (***Volkmer and Heinemann*** (***2011***)) and were used in the work of Schmidt *et al.*. Data from the lab of Terence Hwa are shown in purple (***Basan et al.*** (***2015***)), while the two data sets shown in green and red come from the lab of Suckjoon Jun (***Taheri-Araghi et al.*** (***2015***); ***Si et al.*** (***2017***)). (B) Same as in (A) but with the data plotted on a logarithmic y-axis to highlight the exponential scaling that is expected for *E. coli*.

cellular protein per cell as shown in Figure ***Figure 6*** (purple data points), and is in disagreement with other measurements of total protein at these growth rates (***Basan et al., 2015***).

Since it is not obvious how measurements of cell size influenced their reported protein abundances, in the following subsections we begin by considering this calculation. We then consider three different approaches to estimate the growth-rate dependent total protein mass to compare with those values reported from ***Schmidt et al.*** (***2016***). The results of this are summarized in ***Figure 5***(B), with the original values from both ***Schmidt et al.*** (***2016***) and ***Li et al.*** (***2014***) shown in ***Figure 5***(A) for reference. For most growth conditions, we find that total protein per cell is still in reasonable agreement. However, for the fastest growth conditions, with glycerol + supplemented amino acids, and LB media, all estimates are substantially higher than those originally reported. This is the main reason why we chose to readjusted protein abundance as shown in ***Figure 1***(B) (with the calculation described in section Estimation of Total Protein Content per Cell).

Effect of cell volume on reported absolute protein abundances

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

While this approach provided absolute abundances, another necessary step to arrive at total cellular protein was to account for any protein loss during their various protein extraction steps. Here the authors attempted to determine total protein separately using a BCA protein assay. In personal communications, it was noted that determining reasonable total protein abundances by BCA across their array of growth conditions was particularly troublesome. Instead, they noted

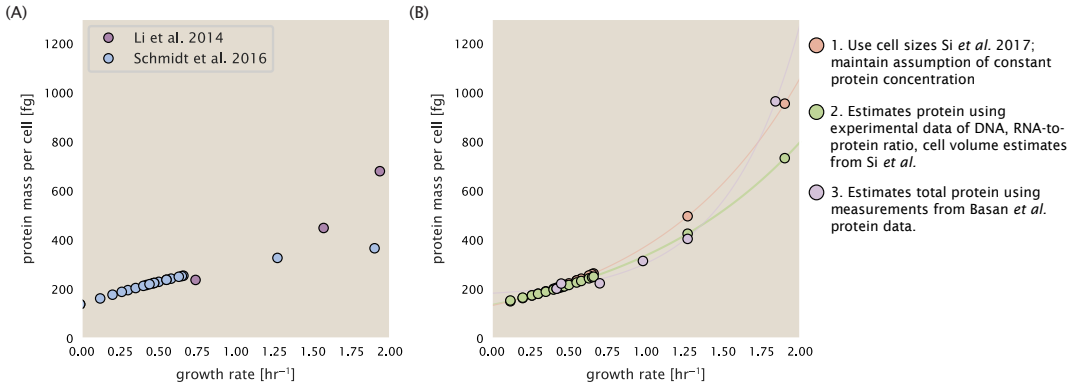

Figure 6. Alternative estimates of total cellular protein for the growth conditions considered in Schmidt *et al.* (A) The original protein mass from Schmidt *et al.* and Li *et al.* are shown in purple and blue, respectively. (B) Three alternative estimates of total protein per cell. 1. *light red*: Rescaling of total protein mass assuming a growth rate independent protein concentration and cell volumes estimated from Si *et al.* 2017. 2. *light green*: Rescaling of total protein mass using estimates of growth rate-dependent protein concentrations and cell volumes estimated from Si *et al.* 2017. Total protein per cell is calculated by assuming a 1.1 g/ml cellular mass density, 30% dry mass, with 90% of the dry mass corresponding to DNA, RNA, and protein (***Basan et al., 2015***). See Estimation of Total Protein Content per Cell for details on calculation. 3.*light purple*: Rescaling of total protein mass using the experimental measurements from Basan *et al.* 2015.

confidence in their total protein measurements for cells grown in M9 minimal media + glucose and used this as a reference point with which to estimate the total protein for all other growth conditions.

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

$$M_{P\_i} = [M_P]_{orig} \cdot V_i \tag{4}$$

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 ***Figure 6*** (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,

$$M'_{P\_i} = [M_P]_{Si} \cdot V_{Si\_i} \tag{5}$$

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 ***Figure 6***(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

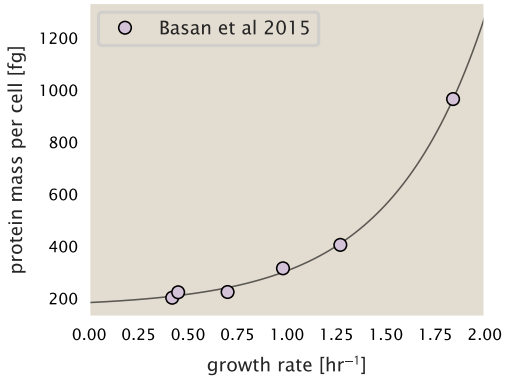

Figure 7. Total cellular protein reported in Basan *et al.* 2015. Measured protein mass as a function of growth rate as reproduced from Basan *et al.* 2015, with cells grown under different levels of nutrient limitation. The data was fit to an exponential curve where protein mass in fg per cell is given by 14.65 $e^{2.180 \cdot \lambda}$ + 172 fg per cell, where λ is the growth rate in hr^{-1}).

detailed in section Estimation of Total Protein Content per Cell for all data points in the ***Schmidt et al.*** (***2016***) data set. The green data points in ***Figure 6***(B) show this prediction, and this represents the approach used to estimate total protein per cell for all data sets.

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

One of the challenges in our estimates in the preceding sections is the need to estimate protein concentration and cell volumes. These are inherently difficult to to accurately due to the small size of *E. coli*. Indeed, for all the additional measurements of cell volume included in Figure ***Figure 5***, no measurements were performed for cells growing at rates below 0.5 hr^{-1}. It therefore remains to be determined whether our extrapolated cell volume estimates are appropriate, with the possibility that the logarithmic scaling of cell size might break down for slower growth.

In our last approach we therefore attempt to estimate total protein using experimental data that required no estimates of concentration or cell volume. Specifically, in the work of Basan *et al*, the authors measured total protein per cell for a broad range of growth rates (reproduced in Figure ***Figure 7***). These were determined by first measuring bulk protein from cell lysate, measured by the colorimetric Biuret method (***You et al.*** (***2013***)), and then abundance per cell was calculated from cell counts from either plating cells or a Coulter counter. While it is unclear why Schmidt *et al.* was unable to take a similar approach, the results from Basan *et al* appear more consistent with our expectation that cell mass will increase exponentially with faster growth rates. In addition, although they do not consider growth rates below about 0.5 hr^{-1}, it is interesting to note that the protein mass per cell appears to plateau to a minimum value at slow growth. In contrast, our estimates using cell volume so far have predicted that total protein mass should continue to decrease slightly for slower growing cells. By fitting this data to an exponential function dependent on growth rate, we could then estimate the total protein per cell for each growth condition considered by ***Schmidt et al.*** (***2016***). These are plotted as red data points in ***Figure 6***(B).

Calculation of Complex Abundance

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

abundance.

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

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

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

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

For example, the 70S mature ribosome complex has 55 protein components, all of which are present in a single copy except L4 (RplL), which is present in 4 copies. For each ribosomal protein, we then calculate the maximum number of complexes that could be formed using ***Equation 6***. This example, along with example from 5 other macromolecular complexes, can be seen in ***Figure 8***.

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

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

in ***Figure 8***, we show this mean value as a grey line for a variety of different complexes. Additionally, we have built an interactive figure accessible on the paper website where the validity of this approach can be examined for any complex with more than two subunits (thus, excluding monomers and dimers).

Extending Estimates to a Continuum of Growth Rates

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

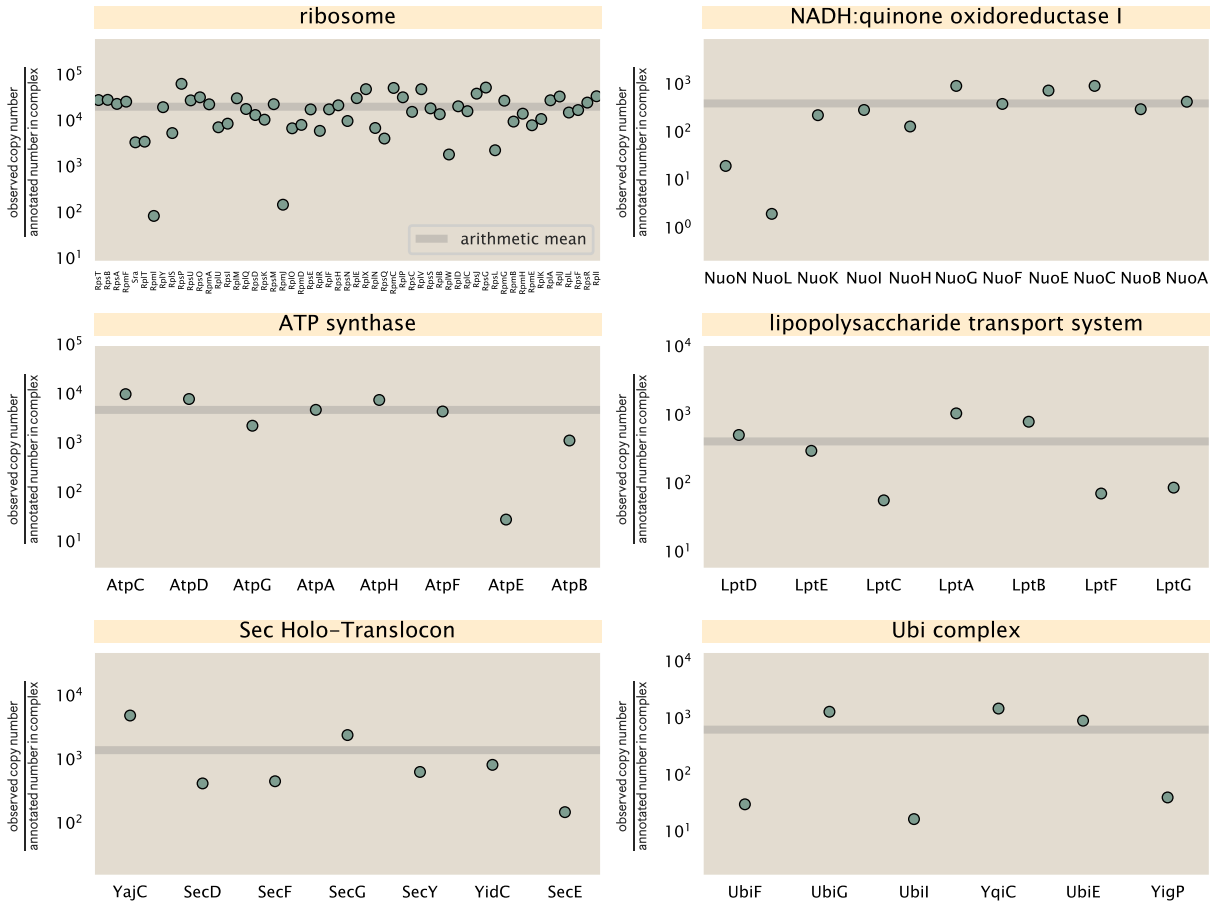

Figure 8. Calculation of the mean complex abundance from measurements of single subunits. Six of the largest complexes (by number of subunits) in *E. coli*. Points correspond to the maximum number of complexes that can be formed given measurement of that individual protein. Solid grey line corresponds to the arithmetic mean across all subunits. These data correspond to measurements from ***Schmidt et al.*** (***2016***) in a glucose-supplemented minimal growth medium.

Estimation of the total cell mass

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

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

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

where a and b are constants with units of μm^3 and hr, respectively. The value of these constants can be estimated from the careful volume measurements performed by ***Si et al.*** (***2017***), as is described in the previous section.

Complex Abundance Scaling With Cell Volume

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

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

$$m_X \approx 0.3 \times \rho V(\lambda) \theta_X, \tag{9}$$

where we have used $\rho V(\lambda)$ as an estimate of the total cell mass, defined in ***Equation 8***. To convert this to the number of units N_X of substance X in the cell, we can use the formula weight w_X of a single unit of X in conjunction with ***Equation 9***,

$$N_X \approx \frac{m_X}{w_X}. \tag{10}$$

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

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

This function is used to draw the continuum estimates for the number of transporters seen in Figures 2 and 3 as transparent grey curves. Occasionally, this continuum scaling relationship will not precisely agree with the point estimate outlined in the main text. This is due to the fact that we make an initial approximation made of a dry cell mass of ≈ 300 fg for the point estimate while we

consider more precise values in the continuum estimate. We note, however, that both this scaling relation and the point estimates are meant to describe the order-of-magnitude observed, and not the predict the exact values of the abundances.

Equation 11 is a very general relation for processes where the cell volume is the "natural variable" of the 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, ***Equation 11*** can be used to compute the number of complexes needed. For example, to compute the number of ATP synthases per cell, ***Equation 11*** can be slightly modified to the form

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

where we have included the term N_{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.

A Relation for Complex Abundance Scaling With Surface Area

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 μm, 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.

Number of Lipids

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 ($A_{\text{lipid}} \approx 0.5$ 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 ***Equation 2***, and recognizing that each membrane is composed of two leaflets, we can compute the number of lipids as a function of growth rate as

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

where η is the length-to-width aspect ratio and V is the cell volume.

Number of Murein Monomers

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 a single murien monomer is $m_{\text{murein}} \approx 1000$ 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 $w \approx 10$ nm encapsulating the cell, one would expect the number of murein monomers scales with the surface area of this shell. In a similar spirit to our calculation of the number of lipids, the total number of murein monomers as a function of growth rate can be calculated as

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

where ρ_{pg} is the density of peptidoglycan.

Complex Abundance Scaling With Number of Origins

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

$$\langle \#\text{ori} \rangle \approx 2^{t_{\text{cyc}}\lambda/\ln 2} \tag{15}$$

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

In the case of rRNA synthesis, the majority of the rRNA operons are surrounding the origin of replication. Thus, at a given growth rate λ, the average dosage of rRNA operons per cell D_{rRNA} is

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

This makes the approximation that *all* rRNA operons are localized around the origin. In reality, the operons are some distance away from the origin, making ***Equation 16*** an approximation.

In the main text, we stated that at the growth rate in question, there is ≈ 1 chromosome per cell. While a fair approximation, ***Equation 15*** illustrates that is not precisely true, even at slow growth rates. In estimating the number of RNA polymerases as a function of growth rate, we consider that regardless of the number of rRNA operons, they are all sufficiently loaded with RNA polymerase such that each operon produces one rRNA per second. Thus, the total number of RNA polymerase as a function of the growth rate can be calculated as

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

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

References

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006). Construction of Escherichia coliK-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2(1):2460.

Basan, M., Zhu, M., Dai, X., Warren, M., Sévin, D., Wang, Y.-P., and Hwa, T. (2015). Inflating bacterial cells by increased protein synthesis. *Molecular Systems Biology*, 11(10):836.

Dai, X., Zhu, M., Warren, M., Balakrishnan, R., Patsalo, V., Okano, H., Williamson, J. R., Fredrick, K., Wang, Y.-P., and Hwa, T. (2016). Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth. *Nature Microbiology*, 2(2):16231.

Datsenko, K. A. and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences*, 97(12):6640–6645.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1983–1992.

Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635.

Ojkic, N., Serbanescu, D., and Banerjee, S. (2019). Surface-to-volume scaling and aspect ratio preservation in rod-shaped bacteria. *eLife*, 8:642.

Peebo, K., Valgepea, K., Maser, A., Nahku, R., Adamberg, K., and Vilu, R. (2015). Proteome reallocation in *Escherichia coli* with increasing specific growth rate. *Molecular BioSystems*, 11(4):1184–1193.

Schaechter, M., Maaløe, O., and Kjeldgaard, N. O. (1958). Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *Microbiology*, 19(3):592–606.

446 Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold,
447 R., and Heinemann, M. (2016). The quantitative and condition-dependent *Escherichia coli* proteome. *Nature*
448 *Biotechnology*, 34(1):104–110.

449 Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and
450 gene expression: origins and consequences. *Science*, 330(6007):1099–1102.

451 Si, F., Li, D., Cox, S. E., Sauls, J. T., Azizi, O., Sou, C., Schwartz, A. B., Erickstad, M. J., Jun, Y., Li, X., and Jun, S. (2017).
452 Invariance of Initiation Mass and Predictability of Cell Size in Escherichia coli. *Current Biology*, 27(9):1278–
453 1287.

454 Soufi, B., Krug, K., Harst, A., and Macek, B. (2015). Characterization of the E. coli proteome and its modifications
455 during growth and ethanol stress. *Frontiers in Microbiology*, 6:198.

456 Taheri-Araghi, S., Bradde, S., Sauls, J. T., Hill, N. S., Levin, P. A., Paulsson, J., Vergassola, M., and Jun, S. (2015).
457 Cell-size control and homeostasis in bacteria. *Current Biology*, 25(3):385–391.

458 Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X. S. (2010). Quantifying
459 E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science (New York, N.Y.)*,
460 329(5991):533–538.

461 Valgepea, K., Adamberg, K., Seiman, A., and Vilu, R. (2013). *Escherichia coli* achieves faster growth by increasing
462 catalytic and translation rates of proteins. *Molecular BioSystems*, 9(9):2344.

463 Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P.,
464 Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J.,
465 Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold,
466 J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van
467 Mulbregt, P., and Contributors, S. . . (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in
468 Python. *Nature Methods*, 17:261–272.

469 Volkmer, B. and Heinemann, M. (2011). Condition-Dependent Cell Volume and Concentration of Escherichia
470 coli to Facilitate Data Conversion for Systems Biology Modeling. *PLOS ONE*, 6(7):e23126.

471 You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C. W., Wang, Y.-P., Lenz, P., Yan, D., and Hwa, T. (2013).
472 Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature*, 500(7462):301–306.