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

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

***For correspondence:**

[*]These authors contributed equally to this work

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

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]
Caglar *et al.* (2017)	mass spectrometry	relative abundance

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 Adjustments to Schmidt *et al.* dataset for more details on this).

Summary of Proteome Datasets.

Here we provide a brief summary of the experiments behind each proteomic data sets. The purpose of this section is to better identify the key steps taken by the authors to arrive at absolute protein abundances. In the following section (Section Summary of Final Compiled Data Set) we will then provide a summary of the final protein abundance measurements that were used for in the main text.

Table **??** provides an overview of the main data sets that 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).

The general strategy taken in these works is to quantify fractional abundance of each protein and then to convert these to absolute abundance by multiplying these fractions by the bulk measured total cellular protein abundance. Note that the work of Peebo *et al.* (2014) did not perform any measurement of cell count or volume, and thus were only able to report cellular protein concentration.

Exceptions to this are found in Schmidt *et al.* and Taniguchi *et al.*. A key distinction in the work of Schmidt *et al.* is that in addition to determining relative abundance by mass spectrometry, they also selected 41 enzyme that cover over four orders of magnitude in cellular abundance to use in absolute protein quantification. Specifically, synthetic peptides were generated for each of these 41 enzymes and used to provide a calibration between measured mass spectrometry intensities and absolute protein abundances (using stable isotope dilution (SID) and selected reaction monitoring (SRM), though the details of this are beyond the scope of this section). In the work of Taniguchi *et al.*, the authored tagged each protein with a yellow fluorescent protein (YFP) and used fluorescence as readout of cellular expression.

Figure **??** shows the distribution in reported protein abundance for a subset of the data.

An important consideration is whether the reported abundance per cell are correlated. while we expect some variability in expression of each protein due to growth rate, the reported values are nonetheless expected to be correlated. Figure **??** compares each dataset to the copy numbers from Schmidt *et al.*, grown in M9 minimal media supplemented with glucose.

Summary of Final Compiled Data Set.

[NB: in progress]

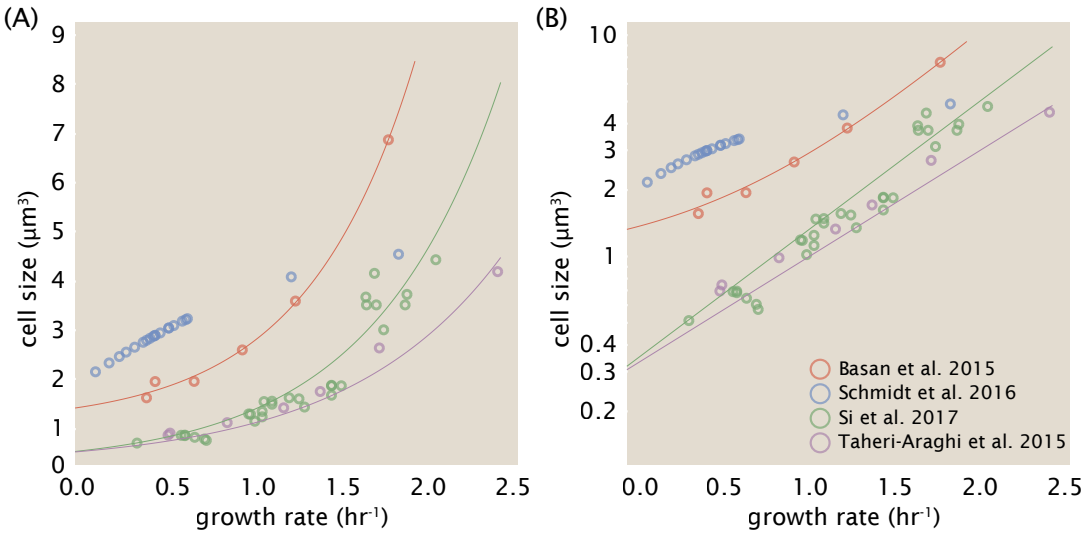

Figure 1. Measurements of cell size as a function of growth rate. (A) Plot of the reported cell sizes from several recent papers. The data in blue come from Volkmer and Heinemann, 2011 (***Volkmer and Heinemann*** (***2011***)) and were used in the work of Schmidt *et al.*. Data from the lab of Terence Hwa are shown in red (***Basan et al.*** (***2015***)), while the two data sets shown in green and purple 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*.

Adjustments to Schmidt *et al.* dataset

While the dataset from Schmidt *et al.* 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 some 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 1. As a compairon, 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. While there is a large discrepancy in cell size between the two research groups, Basan *et al.* found that this came specifically from uncertainty in determining the cell width, which is prone to inaccuracy given the small cell size and optical resolution limits (further described in their supplemental text). Perhaps the more concerning point is that while each of these alternative measurements show an exponential increase in cell size at faster growth rates, the measurements used by Schmidt *et al.* appear to plateau. This resulted in an analogous trend in their final reported total cellular protein per cell as shown in Figure 2 (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 might have influenced their reported protein abundances, we will go through this calculation in the next section. We will also show how these can adjusted to better reflect the alternative measurements of cell size shown in Figure 1. Finally, we consider several strategies to adjust the reported copy numbers, with the result summarized in Figure 2. For most growth conditions, we find that total protein expectations are not expected to change dramatically. However, for the fastest growth conditions, with glycerol + supplemented amino acids, and LB media, there is quite a bit of variability among are different estimates.

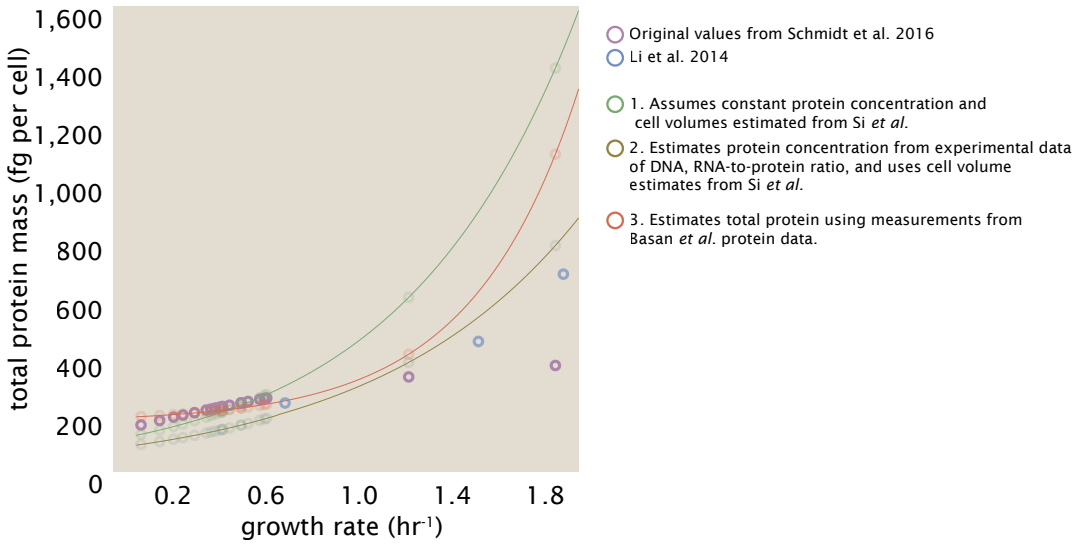

Figure 2. Alternative estimates of total cellular protein for the growth conditions considered in Schmidt *et al.* The original protein mass from Schmidt *et al.* and Li *et al.* are shown in purple and blue, respectively. *Green*: Rescaling of total protein mass assuming a growth rate independent protein concentration and cell volumes estimated from Si *et al.* 2017. *Gold*: Rescaling of total protein mass using estimates of growth rate-dependent protein concentrations and cell volumes estimated from Si *et al.* 2017. *Red*: Rescaling of total protein mass using the experimental measurements from Basan *et al.* 2015.

Effect of cell volume on reported absolute protein abundances in the work of Schmidt *et al.* .

The authors calculated proteome-wide protein abundance 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 (with proteins selected to cover the range of expected copy numbers). 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 needed to arrive at total cellular protein is to account for any protein loss during their various protein extraction steps. Here the authors attempted to determine total protein separately using a BCA protein assay. In personal communications, it was noted that determining reasonable total protein abundances by BCA across their array of growth conditions was particularly troublesome. Instead, they noted confidence in their total protein measurements for cells grown in M9 minimal media + glucose and used this as a reference point with which to estimate the total protein for all other growth conditions.

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

mass for all other growth conditions from,

$$M_{P\_i} = [M_P]_{orig} \cdot V_i \tag{1}$$

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 2 (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,

$$M'_{P\_i} = [M_P]_{Si} \cdot V_{Si\_i} \tag{2}$$

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 [] dots in Figure 2.

Reconsidering assumption that protein concentration is constant.

We next relax the assumption that cellular protein concentration is constant and instead, attempt to estimate it using experimental data. Here we first note that for across almost the entire range of growth rates considered here, protein, DNA, and RNA accounted for at least 90 % of the dry mass in measurements from the lab of Terence Hwa (***Basan et al.*** (***2015***)). They 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. Using the cell volume predictions from Si *et al.*, we can then calculate the associated mass per cell.

However, even if dry mass concentration is relatively constant across growth conditions, it is not a given that protein concentration should also be constant. In particular, we know that rRNA increases substantially at faster growth rates (***Dai et al.*** (***2016***)). This is a well-documented result that arises from an increase in the fraction of ribosomes at faster growth rates (***Scott et al.*** (***2010***)). To proceed we will use 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 3(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 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 3(C) we plot the estimated cellular concentrations for protein, DNA, and RNA from these calculations, and in Figure 3(D) we plot their total expected mass per cell.

Estimating cellular protein concentration as a function of growth rate.

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

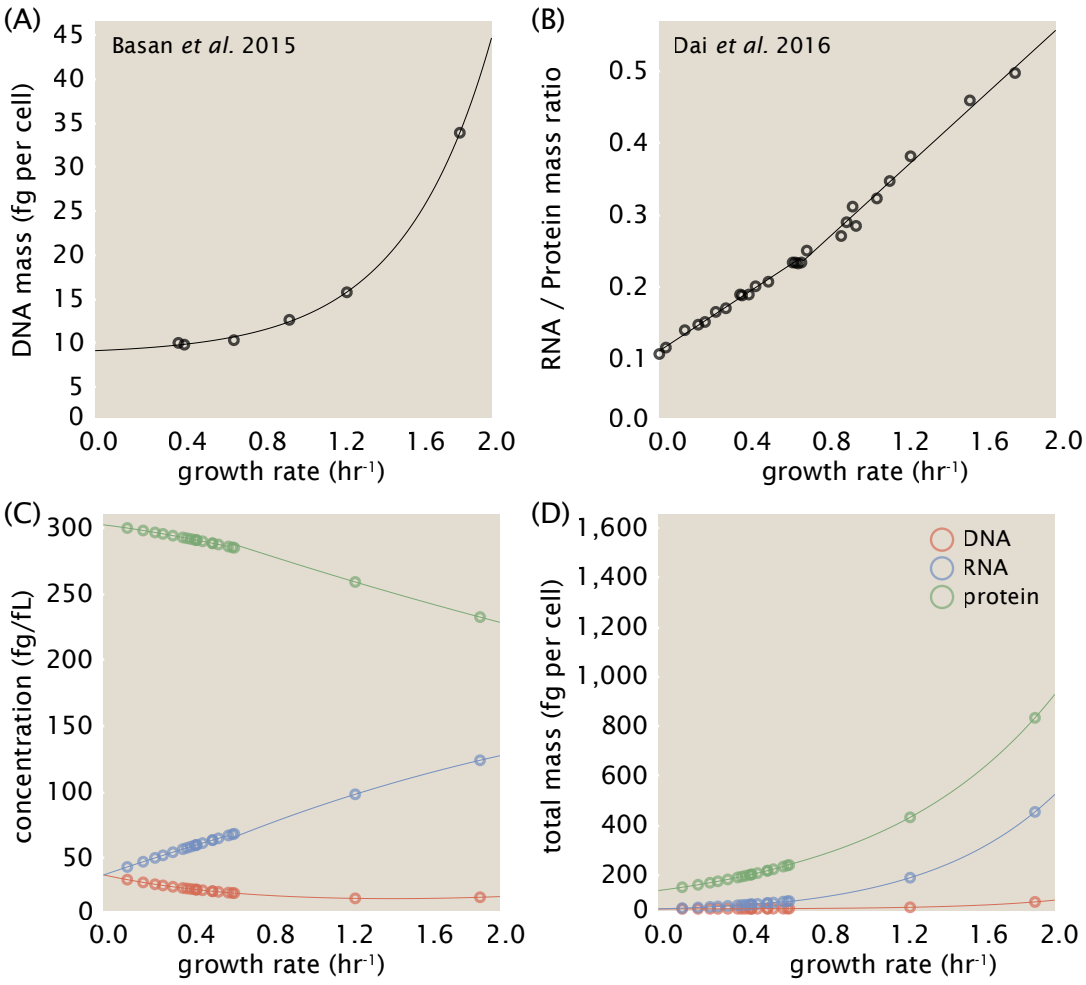

Figure 3. Empirical estimate of cellular protein, DNA, and RNA as a function of growth rate. (A) Measured DNA mass per cell as a function of growth rate, reproduced from Basan *et al.* 2015. The data was fit to an exponential curve (DNA mass in fg per cell is given by 0.42 $e^{2.23 \cdot \lambda}$ + 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.*

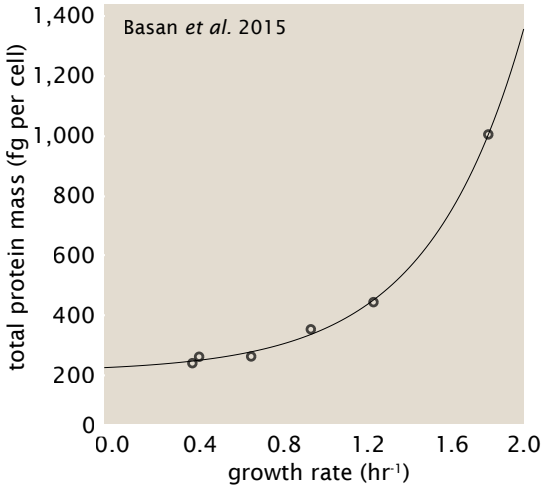

Figure 4. Total cellular protein reported in Basan *et al.* 2015. Measured protein mass as a function of growth rate as reproduced from Basan *et al.* 2015, with cells grown under different levels of nutrient limitation. The data was fit to an exponential curve where protein mass in fg per cell is given by 14.65 $e^{2.180 \cdot \lambda}$ + 172 fg per cell, where λ is the growth rate in hr^{-1}).

size of *E. coli*. Indeed, for all the additional measurements of cell volume included in Figure 1, 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 4). 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.*. These are plotted in red in Figure 2.

Estimation of cell size and surface area across all growth conditions.

In Figure 1 we looked at a number of recent cell size measurements and potential issues with the values used by Schmidt *et al.*. Since most of the proteomic data sets lack cell size measurements, we chose instead to use a common set of size measurements 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 Si *et al.* 2017, 2019 using the optimize.curve$_{f}it function from the Scipy python package (Virtanen et al., 2020).

The size data is shown in Figure 5)(A) and (B), for the cell length and width, respectively. The length data was well described by 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 (or volume) is

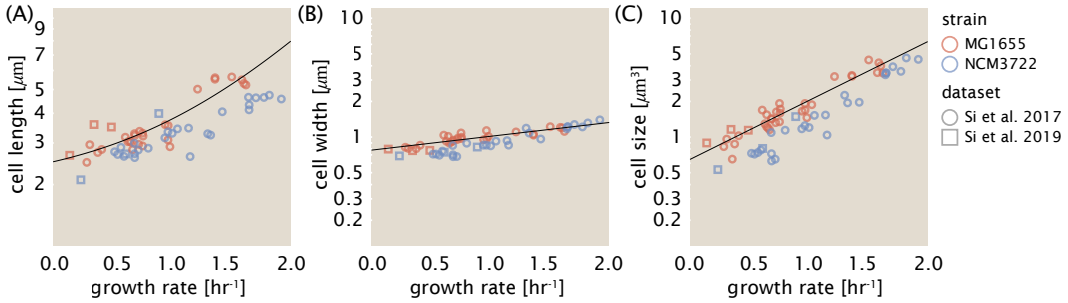

Figure 5. Summary of size measurements from Si *et al.* 2017, 2019. Cell lengths and widths were measured from cell contours obtained from phase contrast images, and refer to the long and short axis respectively. (A) Cell lengths and (B) cell widths show the mean measurements reported (they report 140-300 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 μ m, width: 0.64 $e^{0.24\cdot\lambda}$ μ m. (C) Cell size, V, was calculated as cylinders with two hemispherical ends (Equation 4). The MG1655 strain data gave a best fit of 0.533 $e^{1.037\cdot\lambda}$ μ m^3.

estimated from,

$$V = \pi \cdot r^2 \cdot (l - 2r/3), \tag{4}$$

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

$$S = \eta \cdot \pi (\frac{\eta \cdot \pi}{4} - \frac{\pi}{12})^{-2/3} V^{2/3}, \tag{5}$$

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

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 and compare and contrast the approach to the point estimate procedure.

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 $\approx$ 1 pg and *E. coli* cells can approximated as 70% water by volume. However, as we have shown in this supplemental information, 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

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.

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

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.

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

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

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

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

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

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