Supplemental Information:

N. Belliveau, G. Chure, J. Theriot, R. Phillips

April 17, 2020

1 Summary of Proteome Datasets.

Here we briefly summarize the datasets that were considered for the work of the main text. The goal of this section is to give an overview of each dataset considered, including the main experimental details, and to provide a more detailed look at how well each compares.

Table **??** provides an overview of the proteomic datasets that we found in the literatrue. 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 measusured 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 fluoresence as readout of cellular expression.

Author	Method	Strain	N datasets	Reported Quantity	f
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			protein synthesis rate	
Soufi *et al.* (2015)	Mass spectrometry			fg/copies per cell	
Schmidt *et al.* (2016)	Mass spectrometry			fg/copies per cell	
Caglar *et al.* (2017)	Mass spectrometry			relative abundance	

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 2 compares each dataset to the copy numbers from Schmidt *et al.*, grown in M9 minimal media supplemented with glucose.

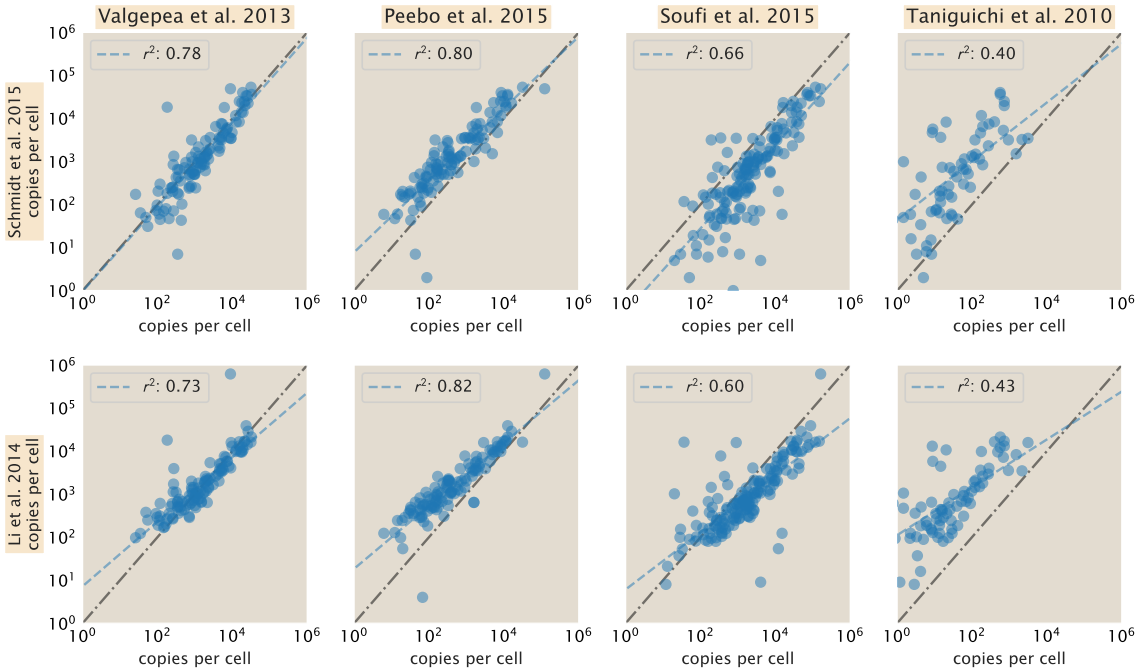

Figure 1

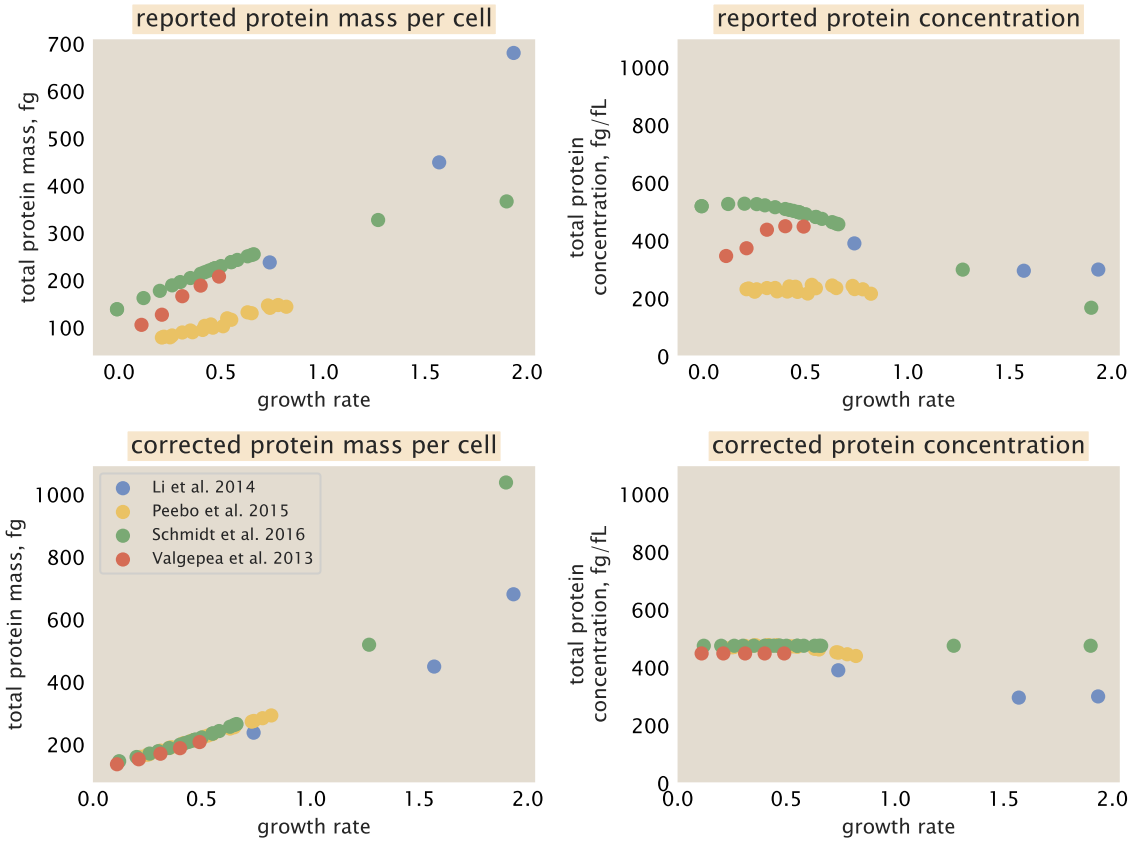

Figure 2

2 Adjustments to Copy Number Data.

The datasets encompass a range of bacterial growth conditions, different *e. coli* strains, and for those that report quantities on a cell basis, different methods to normalize by cell count and volume. It was therefore important to consider if certain discrepencies exist across the data and whether these might be reasonably dealt with to make the compilated dataset internally consistent. - give reference to what was done in example of yeast proteome data corrections. However, given the work of [cite] and others, there are well-documented expectations about how characteristics such as total protein mass per cell and cell volume should scale with growth rate. We were therefore inclined to only renormalize data in a way that took into account such expectations. Figure **??** shows the total protein mass reported as a function of growth rate for each experiment. Indeed, with the exception of the work of Peebo et al., the total mass per cell is generally consistent as a fuction of growth rate, and provide some confidence in such an approach.

In the remainder of this section we describe the rescaling that was done to each dataset, with a particular focus on correcting for discrepencies in cellular protein concentration, which may reflect differences in protein extraction efficiency. It is important to note that with the exception of the work from Peebo *et al.* (which is discussed more below), any

rescaling is only performed within the data of individual authors and not performed globally. We felt this was important in order not to bias any individuals' work since we lack any true standard of protein abundance.

2.1 Corrections to Enforce a Consistent Cellular Protein Concentration

One parameter that we do not expect to change substantially across growth conditions is cellular protein concentration. As a general rule of thumb, we expect an *e. coli* cell to have about 30% dry mass, with about 55% of this expected from protein. With a density of about 1.1 g/ml, we find that the protein concentration in a cell should be approximately 180 fg/fL. The cellular density and dry mass are essentially fixed, with the fraction of cellular protein varying from [X-Y; refs??]. Hence, this parameter provides a useful reference point that datasets should agree on. Indeed, out of concern over differences in protein extraction efficiency in growth phases like stationary phase, Schmidt *et al.* applied a correction to their measured protein abundances to ensure cellular protein concentrations were internally consistent.

From the work of Schmidt *et al.* they reported an ability to consistently get high protein yield from cells grown in M9 minimal media supplemented with glucose. In order to account any protein loss during extraction, they use their measured protein concentration from this sample as a reference for which total protein concentration in all other growth conditions should match. This is shown in Figure **??**A. One challenge in performing this calculation is that cell volume must be known; the authors use volumes that were measured by flow cytometry in previous work [cite]. These volumes are shown in Figure **??**B. While it is difficult to assess the accuracy of these numbers, we find them to be quite inconsistent with the expected scaling that is reported by Taheri-Araghi *et al.* (2015), carefully measured as a function of growth rate [and other work?].

In addition, since cell volume was not determined in all studies, and to be consistent throughout, we instead use the predicted cell volumes from Taheri-Araghi *et al.*. Dealing with each dataset seperately, we apply correction factors to correct for discrepencies in protein concentration across the different growth conditions considered [NB: I wonder if in these other datasets, the more appropriate thing to do is match to the average measured protein concentration]. Specifically, the scaling factor ϕ is given by,

$$\phi = \frac{P_i}{V_i} \cdot [P]_r \tag{1}$$

where P_i is the total protein mass in conditino i, V_i is the estimated cell volume, and $[P]_r$ is the reference protein concentration (i.e. growth in glucose for the Schmidt data).

2.2 Peebo *et al.*: Conversion from copies/ fL to copies per cell

In the work of Peebo *et al.*, the authors only report protein concentration. In order to determine protein per cell, we multiple these concentrations by expected cell volumes using the predictions from Taheri-Araghi *et al.* This is shown in Figure **??**A, where we see that reported mass is substantially lower than the other work considered here; as well as work from others [Sinauer, 1990].

Indeed, both Schmidt *et al.* and Li *et al.* reported a total protein mass of about 250 fg per cell at a growth rate of about $\lambda \approx 0.5hr^{-1}$ (M9 minimal media with glucose and MOPS minimal media, respectively). Given this descrepency, in addition to requiring that cellular protein concentration be internally consistent across the growth conditions they reported on, we also required that total cellular mass be consistent with the work Schmidt *et al.* and Li *et al.* This amounted to performing a linear regression between total protein mass and growth rate, and using this to scale the Peebo *et al.* dataset according to this trend.

3 Translation-dependent limits on the rate of cell division.

Here we consider the hypothesis that synthesis of ribosomes represents the rate-limiting process of cell division. In addition, we consider how a cell might try to achieve this rate of growth and the potential implications on the required ribosomal content of the cell.

3.1 Maximum possible growth rate is set by the time to make a ribosome.

Ribosomes take a unique position among proteins due to their role in replicating the entire pool of cellular proteins, including themselves. In order for a cell to maintain its copy number of ribosomes during division into two daughter cells, each ribosome must make all the protein subunits for a second ribosome. Since the mass of a single ribosome is about 2.5 MDa, with about 2/3 RNA and 1/3 protein, each ribosome has to make about 800 kDa of protein. In *E. coli*, this corresponds to 7,459 amino acids. At a maximal rate of 20 amino acids per second, this would take just over 6 minutes. This time for the synthesis of a ribosome sets a firm time limit on how fast a cell can double itself. Irrespective of the absolute number of ribosomes in a cell, the time to double their number is given by our calculation above. This contrasts with other proteins, where the simple solution to making more proteins is to apparently devote more ribosomes to their synthesis.

NB: I wonder if this should also include the time to make other ribosomal components in addition to core subunits.

3.2 The maximum rate of growth is set by the number of ribosomes.

While the inability to parallelize ribosome synthesis sets an inherent speed limit, this represents a somewhat unachievable rate since the ribosomes must spend some of their time also doubling all the other proteins present in the cell. A translation-limited rate of growth will then be set by the time to double the entire proteome. The only way to reach the minimum duplication time set by the ribosome would be to increase the relative fraction of ribosomes. This would effectively reduce the time needed to duplicate the non-ribosomal proteins.

In order to understand the consequence of each ribosome having to duplicate itself and devote time to double the remaining proteome, we will make use of a toy model. Specifically, lets consider a hypothetical cell that consists of two species of protein, ribosomes and non-ribosomal proteins. The cell contains R ribosomes per cell, and P non-ribosomal proteins per cell. The time τ needed to duplicate the entire proteome is simplly given by,

$$\tau = \tau_R + \tau_P, \tag{2}$$

where τ_R is the time to double a ribosome, while τ_P is the time required to double the remaining proteome. While we found that τ_R is fixed at about 6 minutes, τ_P will depend on the number of ribosomes R and can be approximated by,

$$\tau_P = \frac{N_{aa}}{r_t \cdot R}. \tag{3}$$

Here N_{aa} refers to the total number of amino acids (aa) that must be translated, while r_t refers to the rate of translation, at about 20 aa / sec. Finally, we can then calculate a translation-limited growth rate from,

$$\lambda_{\max} = \frac{ln(2)}{\tau}. \tag{4}$$

Using Equation 3 and 2, this becomes,

$$\lambda_{\max} = \frac{ln(2)}{\tau_R + \frac{N_{aa}}{r_t \cdot R}}. \tag{5}$$

We can see from this that the only way to increase the translation-limited growth rate would be to make more ribosomes, or if it were possible, to decrease the number non-ribosomal proteome.

Next, lets consider some representative values of R and N_{aa} and plot $\lambda_{\max}$. From Schmidt *et al.*, cells growth in glucose were found to have 214 fg of non-ribosomal protein mass. This corresponds to about 1.17 x 10^9 amino acids. We also estimate a ribosomal copy number, R = 20,656 per cell based on the mean copy number of individual subunit copy numbers. Using Equation 5, this corresponds to a maximum growth rate of 0.78 hr^-1, versus the measured rate of 0.58 hr^-1.

NB: I should also write this in terms of ribosome mass fraction.

As we noted earlier, the only way to divide faster than this is to make more ribosomes. One additional difficultly that arises is that in order for a cell to add proteins, it likely will need to increase in size. This may then require that other proteomic proteins also increase in proportion. However, to keep our problem simple, lets proceed with the simplifying assumption that the value of N_{aa} is sufficient to build a cell irrespective of the number of ribosomes. This in effect provides a lower bound on the total proteomic content at faster growth rates. In Figure 3 we plot the translation-limited growth rate $\lambda_{\max}$ as a function of ribosome copy number per cell. While indeed, the maximum attainable growth rate is that set by the time to make a ribosome, it could only be reached if the number of ribosomes was increased about 100 fold. In addition, since cell volume will be proportional to cell mass, the cell will need to become excessively large. We can estimate a lower bound on the required cell volume as a function of the number of ribosomes by making the following assumptions: mass density of 1.1 g/ml, a dry mass of 30% that consists of only protein and RNA. This is shown in Figure 4.

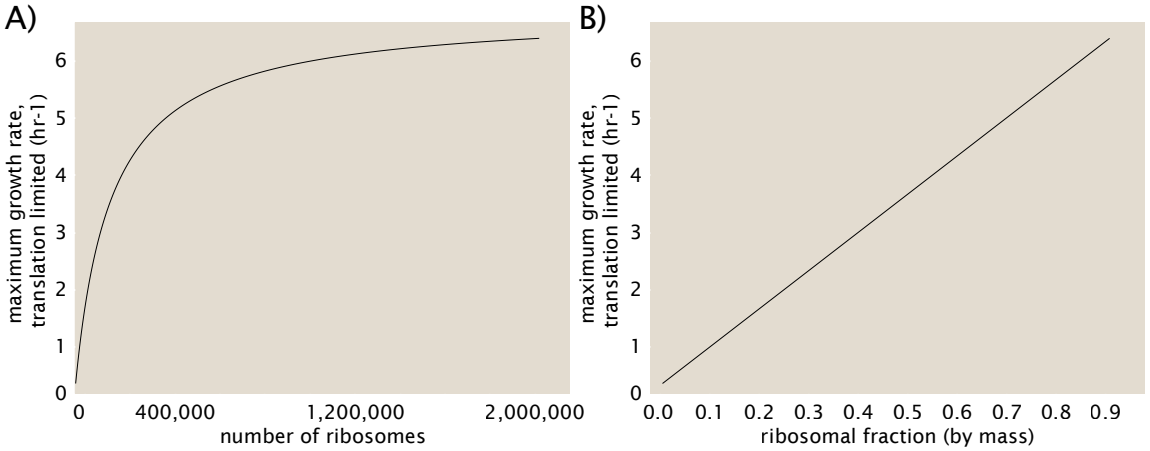

Figure 3

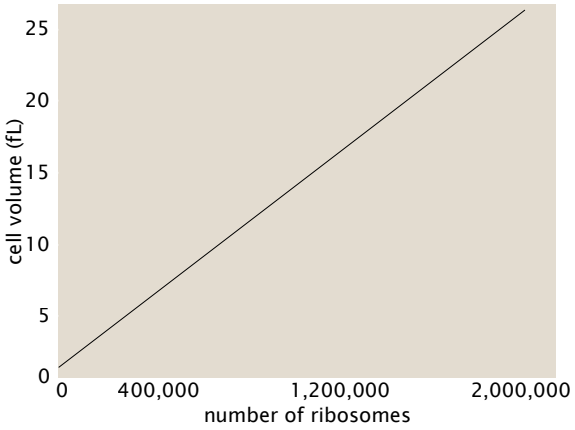

Figure 4

As a last consideration, one observation is the apparently linear dependence between the growth rate and fraction of ribosomal mass. This, along with the dramatic scaling in ribosomal copy number, are particularly relavent to the phenemonlogical growth laws reported by others on how cell size and cell mass scale with growth rate in bacteria. The linear scaling appears to be a feature irrespective of the size of the non-ribosomal mass, as shown in Figure 5. Indeed, with a bit of algerbra, we can re-write the translation-limited growth rate define by Equation 5 as a function of ribosomal mass fracted, denoted by Φ_R, as,

$$\lambda_{\max} = \frac{ln(2)}{N_A \cdot MW_R} \cdot r_t \cdot \Phi_R. \tag{6}$$

N_A is Avagadro's number, while MW_R is the molecular weight of a single ribosome.

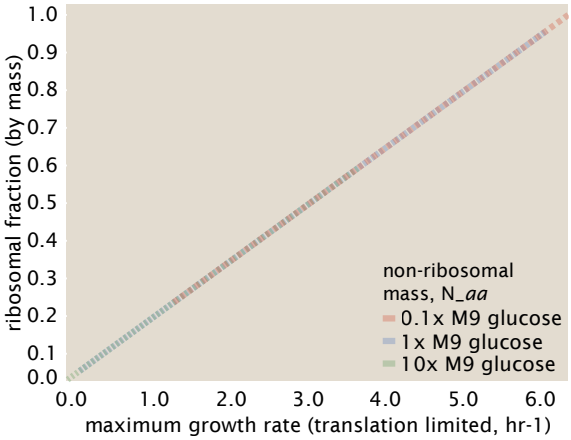

Figure 5

3.3 Growth only appears translation-limited in rich growth media.

NB: A better way would be to directly calculate the number of aa from proteomic sequences and copy number.

With some expectation on the maximum growth rate achiveable as a function of ribosomal content as discussed above, lets now take a look at our experimental data. To simplify our calculations, we approximate the number of amino acids that must be translated from the reported non-ribosomal protein mass and an average molecular weight of an amino acid of 110 Da (N_{aa} = (mass/110 Da) $\cdot$ N_A). Using Equation 5, we can then calculate the maximum rate of growth under translation limitation. In Figure 6A we plot maximum growth under translation limitation against the measured growth rates, while in Figure 6B we plot the cell cycle time that would be associated with these growth rates. The shaded regions identify the regions that should not be attainable with a translation rate of 20 aa/sec. From these two plots, it appears to we are only translation-limited in rich media (data points with growth rates great than 1 hr^{-1} in Figure 6A)). Though it should be noted that a more careful calculation of the maximum tranlation time suggested by Equation 5 needs to be undertaken.

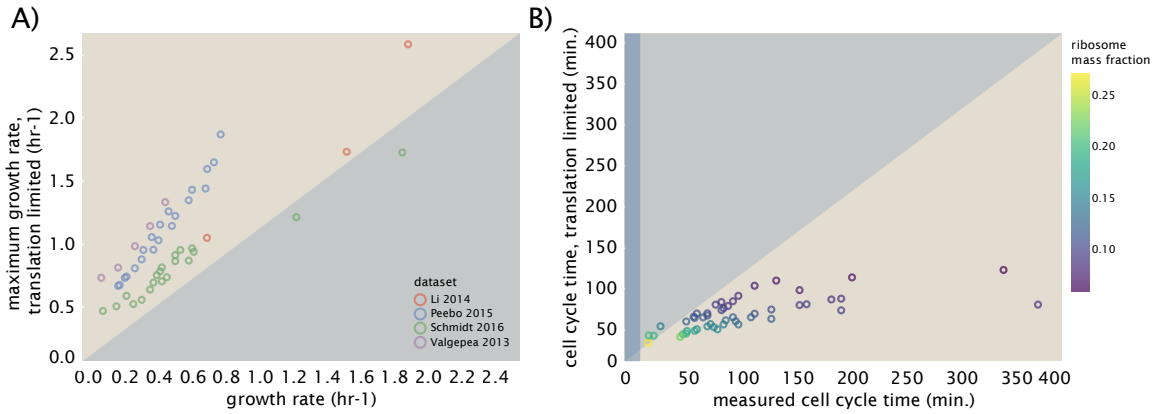

Figure 6

From Figure 6B, it is apparent that at the slower growth rates, the cell cycle time is indeed much longer than might have been expected given a translation rate of 20 aa/sec. We can actually infer what the apparent or effective translation is given the observed growth rate and show this in Figure 7. Interestingly, these translation rates are in good agreement with those measured in Dai *et al.* [1], though that work also suggets that the fraction of active ribosomes may decrease in poor nutrient conditions, further suggesting that under a certain regime cells may not be utilizaing their full ribosomal capacity (i.e. not translation-limited).

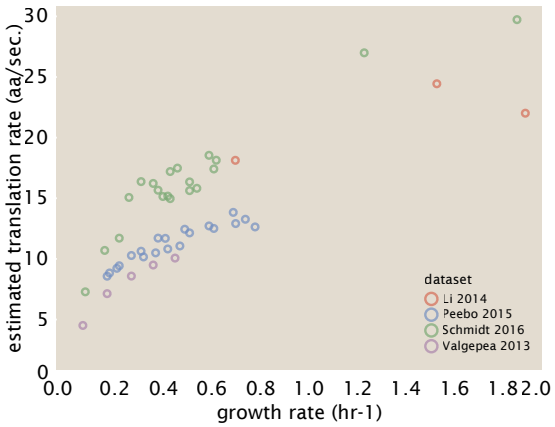

Figure 7

References

[1] Xiongfeng Dai, Manlu Zhu, Mya Warren, Rohan Balakrishnan, Vadim Patsalo, Hiroyuki Okano, James R Williamson, Kurt Fredrick, Yi-Ping Wang, and Terence Hwa. Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth. *Nature Microbiology*, 2(2):16231, December 2016.