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 literature. 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.

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.

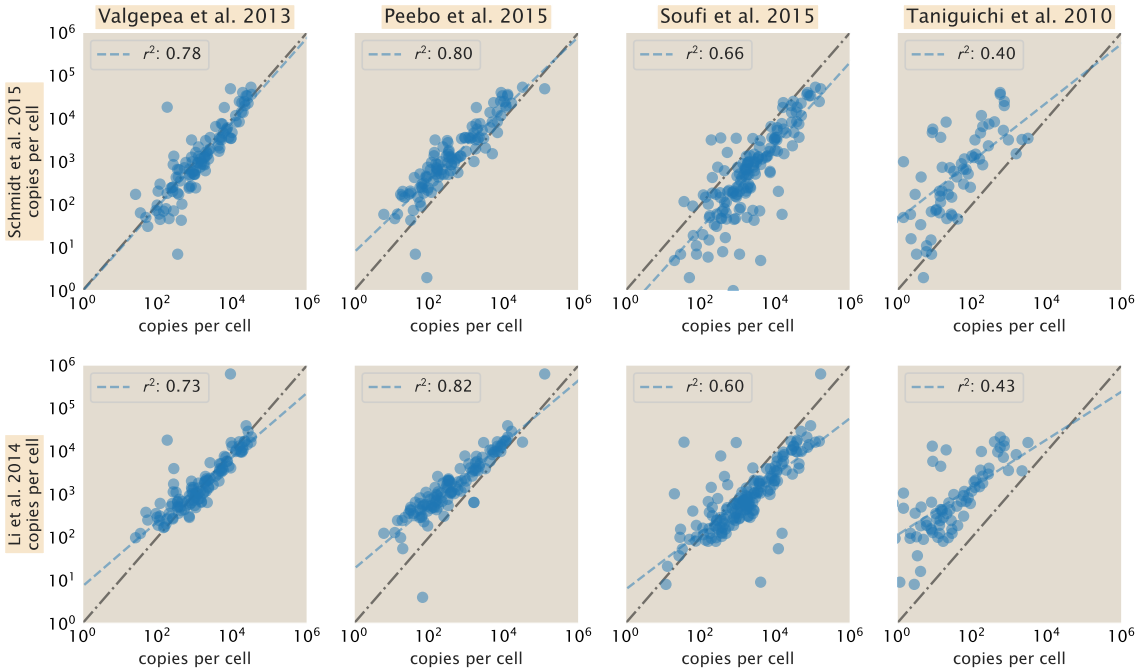


Figure 1

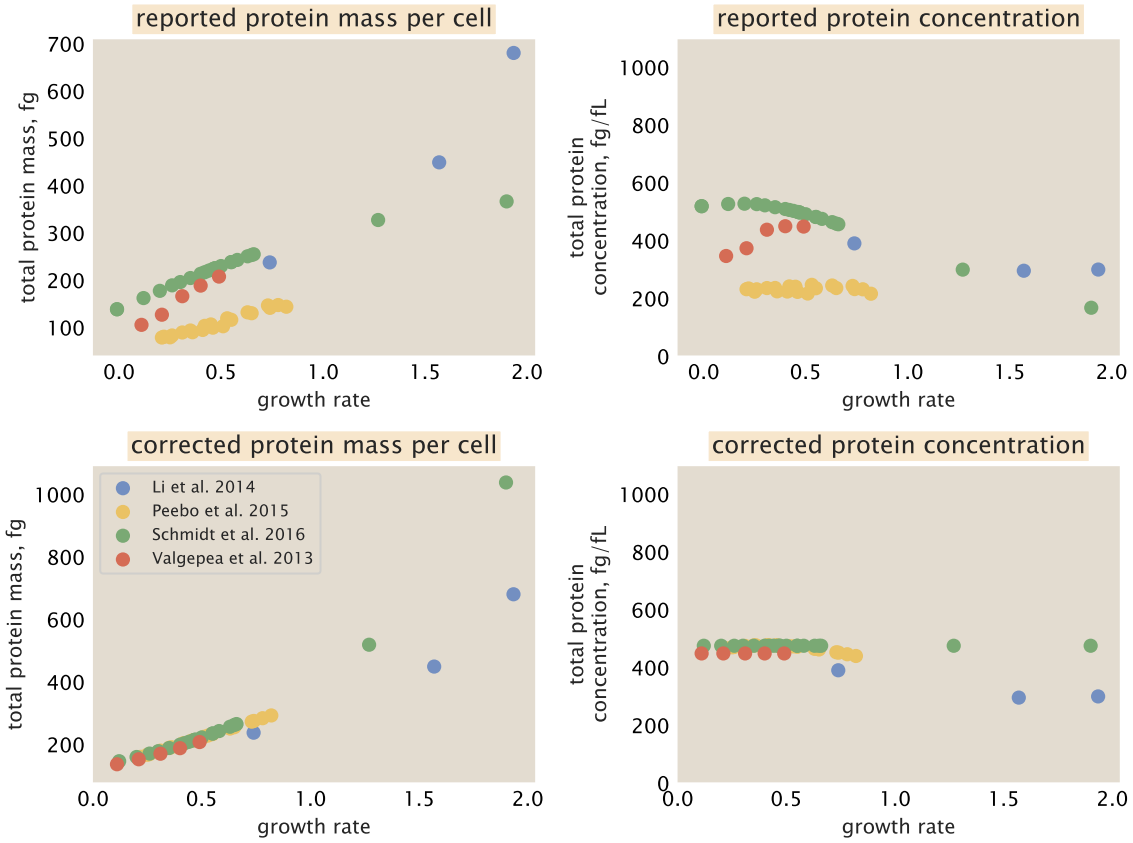


Figure 2

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 discrepancies 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 function 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 discrepancy, 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 the process of translation sets the speed limit of bacterial growth. We begin by considering the synthesis of the ribosome itself, finding that it sets a strict limit on division time, and then from there we consider how the abundance of other cellular proteins limits this achievable growth rate further.

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 cellular proteome. In order for a cell to maintain its own pool of ribosomes during division into two daughter cells, a primary requirement is that 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 translation rate of 20 amino acids per second, this would take just over 6 minutes. Growing any faster would result in a drop in the average number of ribosomes as the cell divides and highlights a strict time limit on how fast a cell can double itself. This result is irrespective of the absolute number of ribosomes, and 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 include the time to make other ribosomal components in addition to core subunits.

3.2 The translation-limited growth rate is set by the fraction of ribosomal mass.

While the inability to parallelize ribosomal synthesis sets an inherent speed limit, this represents a somewhat unachievable growth rate since ribosomes must also spend some of their time doubling the remaining proteome. A translation-limited rate of growth is therefore set by the time to double the entire proteome.

In order to understand the consequence of each ribosome having to duplicate itself and devote time to double the remaining proteome, we consider a hypothetical cell that consists of only two species of protein: ribosomes and non-ribosomal proteins. The cell is taken to contain R ribosomes per cell, and P non-ribosomal proteins per cell. The time τ needed to duplicate the entire proteome is simply 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 proteins. While we found that τ_R is fixed at about 6 minutes, τ_P will depend on the number of ribosomes R available 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 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 Equation 5 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 of non-ribosomal proteins.

Next, let's consider some representative values for R and N_{aa}, and determine $\lambda_{\max}$. From Schmidt *et al.*[1], cells grown 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 subunits that were reported. Using Equation 5, this corresponds to a maximum growth rate of 0.78 hr^{-1}, versus the measured rate of 0.58 hr^{-1}, suggesting cells are growing somewhat below their maximal rate.

As we noted above, the only way to divide faster than this limit of 0.78 hr^{-1} would be to make more ribosomes. We therefore consider how this maximal growth rate might vary as a function of ribosomal copy number. One difficultly that arises is that in order for a cell to add more ribosomes, it will need to increase in size. This might require that other proteins also increase in number to maintain a specific proportion relative to the cell size. 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 us with a lower bound on the total proteomic content at faster growth rates and a lower limit on the achievable growth rate. In Figure 3 we plot $\lambda_{\max}$ 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, we see that it can only be achieved if the number of ribosomes was increased about 100 fold.

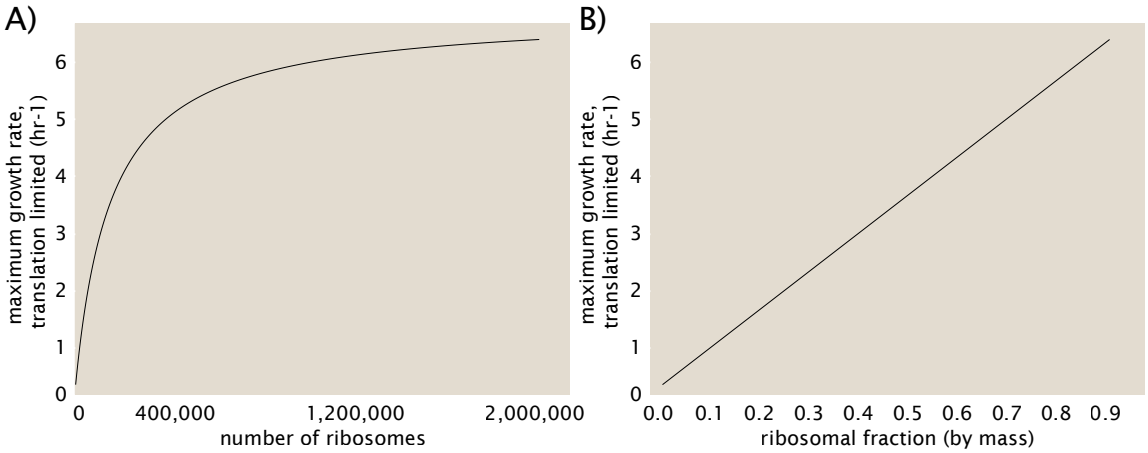


Figure 3: Expectations on the maximum growth rate as a function of ribosome abundance. A) Plot of the translation-limited growth rate in Equation 5, with the 1.17 x 10^9 amino acids, and R from about 2,000 to 2,000,000 copies per cell. B) Related to part A, but instead showing the translation-limited growth rate as a function of ribosomal mass fraction.

To consider what this might mean with respect to cell size, note that cell volume will be proportional to cell mass. We can estimate a lower bound on the required cell volume as a function of the number of ribosomes by assuming a mass density of 1.1 g/ml, and a dry mass of 30% that consists of only protein and RNA. This is shown in Figure 4, and shows that the cell volume would need to be excessively large in order to hold such a large number of ribosomes, with a volume of about 25 fL when $\lambda_{\max}$ is equal to the time to generate a ribosome.

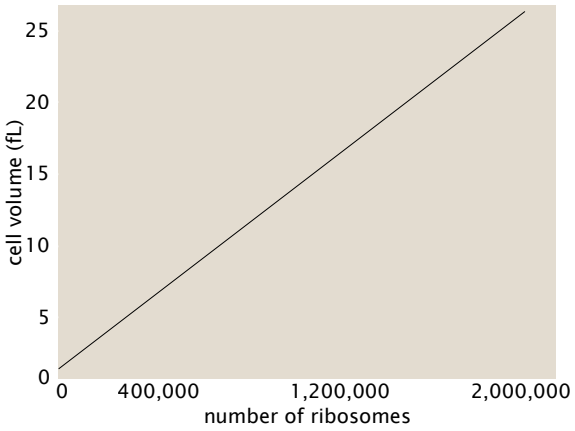


Figure 4: Estimated scaling of cell size with ribosomal copy number. As a first approximation, the cell mass it taken to consist of 214 fg non-ribosomal protein, and a ribosomal mass based on 1/3 corresponding to protein, and 2/3 corresponding to RNA. The cell volume is then calculated assuming a 30 % dry mass, and cell mass density of 1.1 g/ml.

As a last consideration, one additional observation from Figure 3B is an apparently linear dependence between $\lambda_{\max}$ and the fraction of ribosomal mass. This, along with the dramatic

scaling in ribosomal copy number, are particularly relevant to the phenomenological 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 algebra, we can re-write the translation-limited growth rate defined by Equation 5 as a function of ribosomal mass fraction, denoted by Φ_R, as,

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

L_R refers to the number of amino acids that make a single ribosome (L_R = 7,459 aa for *E. coli*). As a sanity check, we can quickly see that if $\Phi_R = 1$, we are once again limited only by the time required to double a ribosome and the rate of translation.

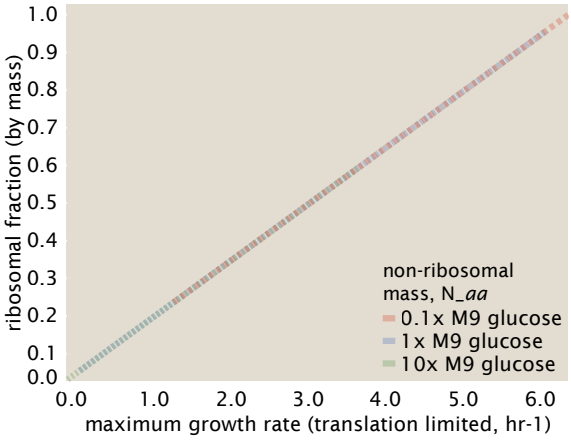


Figure 5: Effect of ribosomal mass fraction on translation-limited growth rate. Following the approach result from Figure 3B, we recalculate the maximum growth rate as the total non-ribosomal mass is either reduced or increased ten fold (i.e. $N_{aa} = [0.1\text{x}N_{aa}, N_{aa}, 10\text{x}N_{aa}]$).

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

With some expectation on the maximum growth rate achievable 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 by assuming an average molecular weight for an amino acid of 110 Da ($N_{aa} = (\text{mass}/110\text{ Da}) \cdot N_A$). Using Equation 5, we can then calculate the maximum rate of growth under translation limitation. In Figure 6A we plot this maximal growth rate, $\lambda_{\max}$, 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 t_r, 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.

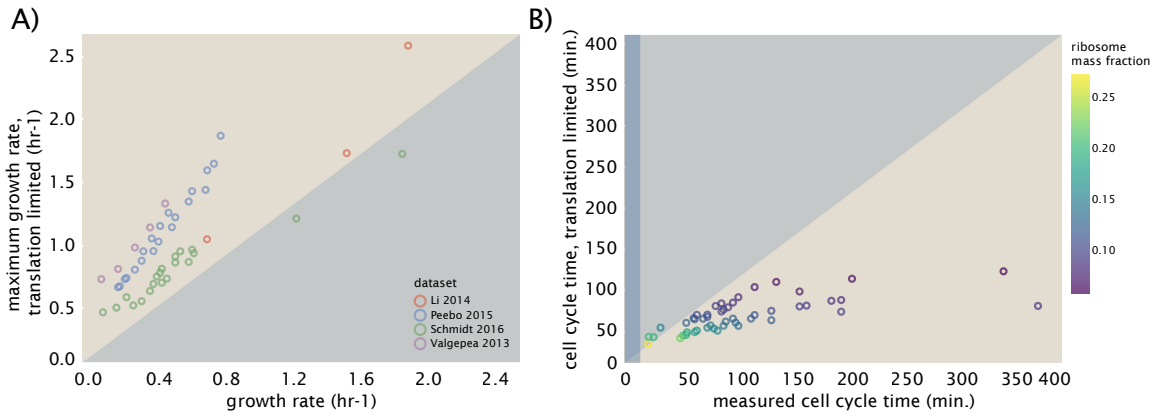
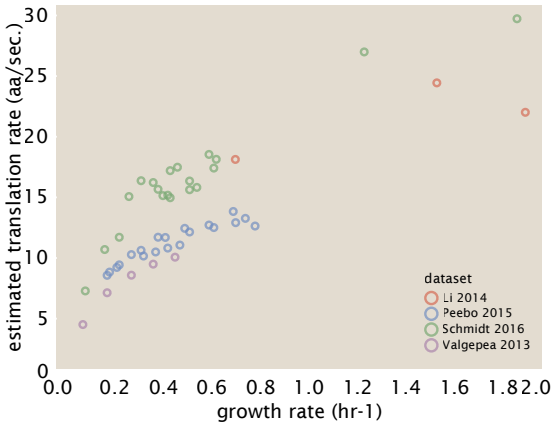


Figure 6: Comparison of translation-limited rate of growth to observed growth rates. A) Plot of maximum growth rates based on reported cell mass and calculated from Equation 5. B) Related to (A), but instead plotting the cell cycle time in minutes. The light shaded regions in (A) and (B) reflect boundaries where growth would not be possible due to a translation rate of 20 aa/sec. The dark shaded region in (B) corresponds to the maximum division rate set by doubling a ribosome. (NB: There is something weird about the fraction of ribosomal protein in Peebo, Valgepea; it is higher, and also higher than that found in Scott *et al.* - is it real??)

From Figure 6B, it is apparent that for cells with slower growth, 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 effective translation is given the observed growth rates, which we show in Figure 7. Interestingly, these translation rates are in good agreement with those measured in Dai *et al.* [2], which were also shown to decrease when grown under limited nutrient conditions.



JT suggests using measured elongation rate in Dai *et al.* to redefine translation boundaries in Figure 6.

Figure 7: Estimate of apparent translation rates based on observed growth rates and measured proteomic mass. Using the measured ribosomal and non-ribosomal mass, we use Equation 5 to estimate r_t for each of the proteomic datasets available.

References

[1] Alexander Schmidt, Karl Kochanowski, Silke Vedelaar, Erik Ahrne, Benjamin Volkmer, Luciano Callipo, Kevin Knoops, Manuel Bauer, Ruedi Aebersold, and Matthias Heinemann. The quantitative and condition-dependent *Escherichia coli* proteome. *Nature Biotechnology*, 34:104–111, 2016.

[2] 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.