Supplement for "Evolutionary tradeoffs in cellular composition across diverse bacteria"

Christopher P. Kempes[1,2,3], Lawrence Wang[1], Jan P. Amend [4,5], John Doyle [2] & Tori Hoehler[3]

[1]*Santa Fe Institute, Santa Fe, NM 87501, USA*

[2]*Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, USA*

[3]*NASA Ames Research Center, Moffett Field, CA 94035, USA*

[4]*Department of Earth Sciences, University of Southern California, Los Angeles, USA*

[5]*Department of Biological Sciences, University of Southern California, Los Angeles, USA*

Introduction

The purpose of this document is to provide extra details on our data compilation, conversion between quantities, mathematical analysis, and parameter values. We provide various plots of data in the original units, equations for conversion, and calculated or measured values of key parameters along with an analysis of statistical significance.

Data Fitting

Most of the curve fitting to compiled data in this study involves power law fits with the exception of our model for ribosomal content, although we also test the possibility of a power law here too. For these fits we performed an ordinary least squares fit to logarithmically transformed data. For this to be an appropriate model the assumption is that the error is multiplicative in the original linear space and thus additive in the logarithmic space, and it should be the case that the variance of the logarithmically transformed data is not increasing with increasing values of x. We tested this and found that for each data set there was no significant trend in variance with increasing x-value bins (e.g. the P-values were all much greater than 0.05 for tests of a positive slope with x).

For each power law fit **Table S1** gives the best fit values along with the 95% confidence intervals and the R^2 values. The P-values for the fitted exponents, and the intercepts for that matter, are all too small to report and are indistinguishable from zero.

It should be noted that the confidence intervals are the largest for the power law fits to the number of ribosomes (**Table S1**), and indeed in the main text we present a more complicated form for the number of ribosomes as a function of cell size, which is connected with a few mechanistic assumptions (please see below for the full derivation of this relationship). **Equation S41** is a function of only one free parameter under the assumption that $\eta = \phi$, and can easily be fit to the data. In order to evaluate all scales of the data equally, and for a fair comparison with the power-law fits, we also us a least squares fit of this model in logarithmic space (both the model and the measured data have been transformed logarithmically). We compare the goodness of fit from this model to that of a simple power law fit by reporting the residual sum of squares (RSS, or SSR or SSE depending on terminology), and the residual standard error, σ. We report both of these in terms of values for logarithmically transformed data.

We find that the for the power law $RSS = 90.08$ and $\sigma = 1.21$ compared with the full form for ribosomes which gives the slightly higher, but similar, values of $RSS = 95.73$ and $\sigma = 1.24$ (see **Table S1** for fitted values of $\eta = \phi$). The power law is a slightly better fit, however, it does implement two degrees of freedom (both of which are significant in terms of P-values) while our full form, which has a mechanistic underpinning, requires only fits for ϕ and compares well with measured values. Furthermore, we have not capitalized on the full fitting capabilities of **Equation S41**: our aim was to predict the ribosome requirement from other observations, namely the growth rate, μ, and total protein volume, thus highlighting the interconnection of component requirements. This implies that the values of μ_0, β_B, P_0, and β_P are being enforced from fits to two other data sources. We have already reduced the degrees of freedom for the full ribosome model from other data, and yet still find fits that are comparable to an unconstrained power law. This suggests that the component requirements are connected in the manner that we have proposed. Allowing any of these parameters (e.g. P_0 and β_P) to vary freely provides a reduction of error between the full ribosome model and data.

In addition, it should be noted that fits of **Equation S41** to data are effectively insensitive to η (given a fixed scaling in μ and N_p) and that the assumption that the ribosomes do not degrade, $\eta = 0$, allows for a fitting of ϕ alone with nearly identical RSS and σ to the $\eta = \phi$ assumption.

Data Compilation

We searched the literature for combinations of terms related to a particular component (e.g. "ribosome") along with generic terms that are related to total quantity, for example "count", "number", "volume", "abundance", "total", "amount", "concentration", etc. We also relied on previous compilations of protein abundance and genome size. Not every datum is useful because it cannot be paired with a cell size, which is a fundamental aspect of this study, and other data cannot be used because they are only given in relative units or abundances (e.g florescence measurements). We relied on various conversions from original units as detailed below.

Because the scale of change in volume can be rather different for each parameter the plots in the main text, which are all on the same y-scale, make it difficult to visually inspect the data and its variation. Thus **Figure S1** provides a plot of each data set within the natural range of y-values.

Growth Rate

Although the model presented in the main text and in Kempes et al. [1] can be derived from a few simple principles, the resulting equational form is fairly complicated and it is useful to provide the power-law approximation, $\mu = \mu_0 V_c^{\beta_{mu}}$, to this relationship for several purposes. For cell sizes sufficiently larger than the asymptotic lower bound, we numerically solved for the best-fit power-law approximation, where we find that $\mu \approx \mu_0 V_c^{\beta_B - 1}$, with $\beta_B \approx 1.64$ (the value extracted from a best fit of μ to the data from [2], which compares well with the direct OLS fit of $\beta_B \approx 1.7$ to the metabolic data from [2]) $\mu_0 = 4 \times 10^7$. The approximation fits the full form of μ with $R^2 = 0.9999$ for cell volumes between 10^{-19} m^3 and 10^4 m^3.

Genome Size

The original units of genome size are typically reported as either total number of base pairs or the total mass of DNA. In figure **S2** we report our compiled data for the genome length in base pairs as a function of cell volume. For the data that were reported in mass units we used the following conversion

$$L = M_{DNA}/m_{bp} \tag{S1}$$

where M_{DNA} is the total measured mass of DNA and $m_{bp} = 1.023 \times 10^{-21}$ g is the mass of a single basepair [3].

Given the total genome length, L, and the volume of a nucleotide, v_N, the total volume of DNA is given by

$$V_{DNA} = \bar{v}_N L, \tag{S2}$$

and thus, scales identically to genome length. For v_N we assumed that the length of a nucleotide is given by 0.33 nm [4] and that the radius is given by 1.2 nm [4], and thus, $v_N \approx 1.47 \times 10^{-27}$ m^{-3}.

Conversion between cell mass and volume.

Many previous studies report allometric relationships in terms of cell mass dependence. However, the most common direct measurement of size is cell volume (often estimated from cell dimensions) which is convenient given our interest in the overall space constraints of cells. We assume that most of the studies that report cell mass have used a simple constant density conversion to calculate this quantity from measured cell volumes (e.g. [2]), and so we convert back to volume using a standard cellular density of $\delta_c = 1.1 \times 10^6$ g m^{-3} (e.g. [5]). Future efforts to convert from cell volume to

mass may need to consider the shifts in cell density associated with cell volume as illustrated in this study.

Dry Weight Scaling

For later conversions it is useful to define the conversion between overall cell volume and the total dry weight. Past work has shown that the dry mass of a cell, M_d, scales with overall volume following

$$M_d = v_0 V_c^{\beta_c} \quad \textbf{(S3)}$$

where the constants vary between $\beta_c = 0.86$ and $v_0 = 435$ (dry fg μm$^{-\beta_c}$) in *E. coli* [6] to $\beta_c = 0.91$ and $v_0 = 162$ (dry fg μm$^{-\beta_c}$) for a broader assessment of bacteria [7]. We use an average of these values defined as $\beta_c^{ave} = (0.91 + 0.86)/2$ and $v_0^{ave} = (162 \times 435)^{1/2}$.

Protein Mass

Protein amounts are typically reported in either total mass per cell units, total number units, or percent of total cellular dry weight. Figure **S3** gives our compilation for the total protein mass of a cell against cell volume.

When protein values were reported in percent of cellular dry weight we estimated the cellular dry weight using the cell volume and the conversion found in Equation **S3**. We converted from the number of proteins to total protein mass using

$$M_p = \bar{m}_p N_p \quad \textbf{(S4)}$$

where the average mass of a protein ranges from 4.98×10^{-20} g [8] to 6.64×10^{-20} g [9] and N_p is the number of proteins. For the average mass of a protein, we used the average of the two, so $\bar{m}_p = 5.81 \times 10^{-20}$ g.

For conversion to total protein volume reported in the main text we used

$$V_p = \bar{d}_p M_p \quad \textbf{(S5)}$$

where the average density of a protein is given by $\bar{d}_p = 1.37 \times 10^6$ g m^3 [10]. For calculations in the main text (Equations 6 and 7), it is important to note that the average length of a protein is $\bar{l}_p = 975$ bp [11] which is similar to the reported average gene length of 924 bp found to be approximately invariant across bacteria [12].

Ribosome Number

The basic premise for the number of ribosomes is that the ribosomes must be able to replicate all of proteins and all of the ribosomes within the division cycle. Below we present several perspectives on this requirement, each of which adds a layer of detail to the model, ranging from a zeroth-order approximation for the entire life cycle to a model which accounts for dynamic production and degradation of both ribosomes and proteins integrated over a life-cycle. We present all of these models because each may be useful in different contexts where certain approximations or layers of detail might be most relevant and/or appropriate.

Simple life-cycle average

The total number of base pairs that the ribosome must process in a life-cycle is given by

$$N_p\bar{l}_p + N_r\bar{l}_r \tag{S6}$$

where $\bar{l}_p$ and $\bar{l}_r$ are the average length, in base pairs, of protein and ribosome transcripts, respectively, that must be processed by the ribosome, and the N terms are how many proteins and ribosomes the cell actually has. The total translational capacity of all of the ribosomes is then given by the rate at which a ribosome can process a base pair, r_r, multiplied by how much time the cell has to divide, t_d and the total number or ribosomes in the cell, or

$$t_d r_r N_r. \tag{S7}$$

Here we assume that the ribosome processing rate r_r is a constant, taken to be maximal, and independent of concentration effects. For the cell to successfully double all proteins and ribosomes it must be the case that the total translational capacity (Equation **S7**) meet or exceed the total requirements (Equation **S6**):

$$t_d r_r N_r \geq N_p\bar{l}_p + N_r\bar{l}_r. \tag{S8}$$

which becomes

$$t_d r_r N_r - N_r\bar{l}_r \geq N_p\bar{l}_p \tag{S9}$$

$$N_r \geq \frac{N_p\bar{l}_p}{t_d r_r - \bar{l}_r}. \tag{S10}$$

Equation S10 depends strongly on overall cell volume given that $t_d = \ln(2)/\mu$ (where μ depends on cell volume) given our earlier presentation of N_p. However, the connection to cell size can be difficult to observe in the above form given the complicated relationship for μ and the several layers of subsequent dependencies of features like N_p. The full form can be found by noting that, $V_p = P_0 V_c^{\beta_p}$, and $N_p = V_p/\bar{v}_p$ given the average volume of a protein, $\bar{v}_p$. All of these considerations

taken together lead to

$$N_r \geq \frac{P_0 V_c^{\beta_p} \bar{l}_p}{\bar{v}_p \left(\ln(2) \left(\frac{(B_m/E_m)(1-\beta_B)\ln[\epsilon]}{\ln\left[\frac{1-(B_m/B_0)(V_c d_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_c d_c)^{1-\beta_B}}\right]} \right)^{-1} r_r - \bar{l}_r \right)} \quad \textbf{(S11)}$$

which we can simplify using the power-law approximation for μ (see **Equation 5**) as

$$N_r \geq \frac{P_0 V_c^{\beta_p} \bar{l}_p}{\bar{v}_p \left(\ln(2) \left(\mu_0 V_c^{\beta_\mu} \right)^{-1} r_r - \bar{l}_r \right)}. \quad \textbf{(S12)}$$

Most of the ribosome data is reported in number units (Figure **S4**) which we convert to volume using

$$V_r = \bar{v}_r N_r \quad \textbf{(S13)}$$

where the average volume of a ribosome ranges from 2.68×10^{-24} m^3 [13] to 3.40×10^{-24} m^3 [14]. For the average volume of a ribosome, we used the average of the two, so $\bar{v}_r = 3.04 \times 10^{-24}$ m^{-3}. For the average length of a ribosome we used $\bar{l}_r = 4566$ bp [15], and took the maximum ribosome processing rate to be $r_r = 63$ bp s^{-1} [15].

Temporal Dynamics of Ribosome and Protein Synthesis

It should be noted that as the cell grows the number of ribosomes is changing and newly synthesized ribosomes can contribute to the biosynthesis of proteins or more ribosomes. Similarly, the number of proteins is also changing in time as is the total requirement for protein biosynthesis. For proteins there are also many degradation terms that contribute to the total amount of biosynthesis which we discuss later. The capacity for biosynthesis, and thus the rate of change in both ribosomes and proteins, will depend on the current number of ribosomes in the system and how this current capacity is partitioned between these two components. Thus, the explicit temporal dynamics of both ribosome and protein synthesis can be described by the temporal derivatives

$$\frac{dN_r}{dt} = \gamma \frac{r_r N_r}{\bar{l}_r} \quad \textbf{(S14)}$$

$$\frac{dN_p}{dt} = (1-\gamma) \frac{r_r N_r}{\bar{l}_p} \quad \textbf{(S15)}$$

where γ is the fraction of total ribosomal biosynthetic capacity dedicated to making more ribosomes. Solving these equations provides the number of ribosomes and proteins as a function of time:

$$N_r(t) = N_{r,0} e^{\gamma r_r t / \bar{l}_r} \quad \textbf{(S16)}$$

where $N_{r,0}$ is the initial number of ribosomes at the beginning of a division cycle and the relationship for the number of proteins is a bit more complicated

$$N_p(t) = \frac{(1-\gamma)\,\bar{l}_r N_{r,0}\left(e^{\gamma r_r t/\bar{l}_r}-1\right)+\gamma\bar{l}_p N_{p,0}}{\gamma\bar{l}_p} \tag{S17}$$

where $N_{p,0}$ is the initial number of proteins. It should be noted that the data collected for our scaling relationships represent average quantities, and so it is useful to define the average values for N_r and N_p over a life cycle:

$$\bar{N}_r(t) \equiv \frac{1}{t_d}\int_0^{t_d} N_r(t) \tag{S18}$$

$$= \frac{N_{r,0}\bar{l}_r\left(e^{\gamma r_r t_d/\bar{l}_r}-1\right)}{\gamma r_r t_d} \tag{S19}$$

and

$$\bar{N}_p(t) \equiv \frac{1}{t_d}\int_0^{t_d} N_p(t) \tag{S20}$$

$$= \frac{(1-\gamma)\,\bar{l}_r^2 N_{r,0}\left(e^{\gamma r_r t_d/\bar{l}_r}-1\right)+\gamma^2\bar{l}_p N_{p,0} r_r t_d - l_r N_{r,0} r_r t_d\,(\gamma-1)\,\gamma}{\gamma^2\bar{l}_p r_r t_d}. \tag{S21}$$

Given our expression for the number of ribosomes in time, Equation **S16**, we can bound the value for the partitioning of total biosynthetic resources dedicated to ribosome production. Considering are earlier requirements, at a minimum the ribosomes must be doubled in the time to divide, or

$$N_r(t_d) \geq 2N_{r,0} \tag{S22}$$

$$N_{r,0}e^{\gamma r_r t/\bar{l}_r} \geq 2N_{r,0} \tag{S23}$$

$$\gamma \geq \frac{\ln(2)\,\bar{l}_r}{r_r t_d}. \tag{S24}$$

$$\tag{S25}$$

This bound on γ can be used to find a rough estimate for the "ribosome catastrophe" by noting that $\gamma < 1$ so that there remain transcriptional resources for proteins. We might expect that γ should be as small as possible to allow for the maximum translational resources to be dedicated to protein synthesis, and substituting the lower bound into the average number of ribosomes gives the simple result that

$$\bar{N}_r = \frac{N_{r,0}}{\ln(2)} \tag{S26}$$

which could be anticipated from standard considerations of exponential growth.

Similarly, it must also be the case that all of the proteins are replicated in the devision cycle and that

$$N_p\left(t_d\right) \geq 2N_{p,0} \quad \textbf{(S27)}$$

$$\frac{\left(1-\gamma\right)\bar{l}_r N_{r,0}\left(e^{\gamma r_r t/\bar{l}_r}-1\right)+\gamma\bar{l}_p N_{p,0}}{\gamma\bar{l}_p} \geq 2N_{p,0} \quad \textbf{(S28)}$$

and, importantly, this condition can be manipulated to give us a relationship between $\bar{N}_r$ and $\bar{N}_p$ which is our ultimate goal. Equation **S28** can be rewritten as

$$N_{r,0}\left(e^{\gamma r_r t_d/\bar{l}_r}-1\right) \geq \frac{\gamma\bar{l}_p N_{p,0}}{\bar{l}_r-\gamma\bar{l}_r} \quad \textbf{(S29)}$$

and taking the minimum value for γ from Equation **S25** leads to

$$N_{r,0} \geq \frac{\gamma\bar{l}_p N_{p,0}\ln\left(2\right)}{\bar{r}_r t_d-\ln\left(2\right)\bar{l}_r}. \quad \textbf{(S30)}$$

Again taking the minimum value for the γ in Equation **S21** gives the initial number of proteins as a function of the average

$$N_{p,0}=\frac{\bar{l}_p\bar{N}_p\ln\left(2\right)^2+N_{r,0}\left(\ln\left(2\right)-1\right)\left(\bar{r}_r t_d-\bar{l}_r\ln\left(2\right)\right)}{\bar{l}_p\ln\left(2\right)^2}. \quad \textbf{(S31)}$$

and substituting for $N_{p,0}$ in Equation **S30** gives

$$\frac{N_{r,0}}{\ln\left(2\right)} \geq \frac{N_{r,0}\left(\ln\left(2\right)-1\right)}{\ln\left(2\right)^2}+\frac{\bar{l}_p\bar{N}_p}{\bar{r}_r t_d-\bar{l}_r\ln\left(2\right)} \quad \textbf{(S32)}$$

which can be simplified to give our final bound for the number of ribosomes:

$$\bar{N}_r \geq \frac{\bar{l}_p\bar{N}_p}{\frac{\bar{r}_r t_d}{\ln(2)}-\bar{l}_r}. \quad \textbf{(S33)}$$

This result is nearly identical to the average perspective present in **Equation S10** but where the $\bar{r}_r t_d$ differs by a factor of $\ln\left(2\right)$.

Biosynthetic Dynamics with Degradation

It should be noted that proteins are constantly being degraded and re-synthesized during cellular function and growth. We can add this effect to **Equations S15 and S14** by including a constant degradation rate, η for ribosomes and ϕ for proteins,

$$\frac{dN_r}{dt}=\gamma\frac{r_r N_r}{\bar{l}_r}-\eta N_r \quad \textbf{(S34)}$$

$$\frac{dN_p}{dt} = (1-\gamma)\frac{r_r N_r}{\bar{l}_p} - \phi N_p \tag{S35}$$

which can be solved to find the number of ribosomes and proteins as a function of time:

$$N_r = N_{r,0} e^{\frac{t\left(\gamma\bar{r}_r - \eta\bar{l}_r\right)}{\bar{l}_r}} \tag{S36}$$

$$N_p(t) = \frac{e^{-t\phi}\left(\bar{l}_r\left(\bar{l}_p\bar{N}_{p,0}(\phi-\eta) + (\gamma-1)\bar{N}_{r,0}\bar{r}_r\left(1 - e^{t\left(-\eta+\frac{\gamma\bar{r}_r}{\bar{l}_r}+\phi\right)}\right)\right) + \gamma\bar{l}_p\bar{N}_{p,0}\bar{r}_r\right)}{\bar{l}_p(\bar{l}_r(\phi-\eta) + \gamma\bar{r}_r)}. \tag{S37}$$

Following the definitions and procedures of **Equations S21-S25** we find that

$$\bar{N}_r = \frac{\bar{l}_r\bar{N}_{r,0}e^{\frac{\gamma\bar{r}_r t_d}{\bar{l}_r} - \eta t_d} - \bar{l}_r\bar{N}_{r,0}}{\gamma\bar{r}_r t_d - \eta\bar{l}_r t_d} \tag{S38}$$

$$\bar{N}_p = \frac{e^{-t_d\phi}\left(\bar{l}_p\bar{N}_{p,0}\left(e^{t_d\phi}-1\right)(\bar{l}_r(\phi-\eta)+\gamma\bar{r}_r) + \frac{(\gamma-1)\bar{l}_r\bar{N}_{r,0}\bar{r}_r\left(-\eta\bar{l}_r + \bar{l}_r\phi e^{t_d\left(-\eta+\frac{\gamma\bar{r}_r}{\bar{l}_r}+\phi\right)} + \bar{l}_r(\eta-\phi)e^{t_d\phi} + \gamma\bar{r}_r\left(1-e^{t_d\phi}\right)\right)}{\eta\bar{l}_r - \gamma\bar{r}_r}\right)}{\bar{l}_p t_d\phi(\bar{l}_r(\phi-\eta)+\gamma\bar{r}_r)} \tag{S39}$$

and

$$\gamma \geq \frac{\bar{l}_r(\eta t_d + \ln(2))}{\bar{r}_r t_d}. \tag{S40}$$

Which, given several steps of substituting following the same algebra outlined above, the number of ribosomes required, considering the protein degradation rate, is given by

$$\bar{N}_r \geq \frac{\bar{l}_p\bar{N}_p\left(\frac{t_d\phi}{\ln(2)}+1\right)}{\frac{\bar{r}_r t_d}{\ln(2)} - \bar{l}_r\left(\frac{\eta t_d}{\ln(2)}+1\right)} \tag{S41}$$

which reduces to **Equation S33** if ϕ and η are small. Noting that $\mu = \ln(2)/t_d$, **Equation S41** can be rewritten as

$$\bar{N}_r \geq \frac{\bar{l}_p\bar{N}_p\left(\frac{\phi}{\mu}+1\right)}{\frac{\bar{r}_r}{\mu} - \bar{l}_r\left(\frac{\eta}{\mu}+1\right)} \tag{S42}$$

and substituting either the full form (**Equation 3**) or power-law approximation (**Equation 4**) for the growth rate μ leads to relationships for $\bar{N}_r$ that are explicitly dependent on cell volume:

$$\bar{N}_r \geq \frac{\bar{l}_p\bar{N}_p\left(\frac{\phi\ln\left[\frac{1-(B_m/B_0)(V_c d_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_c d_c)^{1-\beta_B}}\right]}{(B_m/E_m)(1-\beta_B)\ln[\epsilon]}+1\right)}{\frac{\bar{r}_r\ln\left[\frac{1-(B_m/B_0)(V_c d_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_c d_c)^{1-\beta_B}}\right]}{(B_m/E_m)(1-\beta_B)\ln[\epsilon]} - \bar{l}_r\left(\frac{\eta\ln\left[\frac{1-(B_m/B_0)(V_c d_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_c d_c)^{1-\beta_B}}\right]}{(B_m/E_m)(1-\beta_B)\ln[\epsilon]}+1\right)} \tag{S43}$$

or

$$\bar{N}_r \geq \frac{\bar{l}_p \bar{N}_p \left(\frac{\phi}{\mu_0 V_c^{\beta_B - 1}} + 1 \right)}{\frac{\bar{r}_r}{\mu_0 V_c^{\beta_B - 1}} - \bar{l}_r \left(\frac{\eta}{\mu_0 V_c^{\beta_B - 1}} + 1 \right)}. \quad \textbf{(S44)}$$

The final step of substitution required for these relationships to be fully determined by cell volume is the scaling of N_p which is equal to $\frac{V_p}{\bar{v}_p}$ and so **Equations S43 and S44** become

$$\bar{N}_r \geq \frac{\bar{l}_p \frac{P_0 V_c^{\beta_p}}{\bar{v}_p} \left(\frac{\phi \ln\left[\frac{1-(B_m/B_0)(V_c d_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_c d_c)^{1-\beta_B}} \right]}{(B_m/E_m)(1-\beta_B)\ln[\epsilon]} + 1 \right)}{\frac{\bar{r}_r \ln\left[\frac{1-(B_m/B_0)(V_c d_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_c d_c)^{1-\beta_B}} \right]}{(B_m/E_m)(1-\beta_B)\ln[\epsilon]} - \bar{l}_r \left(\frac{\eta \ln\left[\frac{1-(B_m/B_0)(V_c d_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_c d_c)^{1-\beta_B}} \right]}{(B_m/E_m)(1-\beta_B)\ln[\epsilon]} + 1 \right)} \quad \textbf{(S45)}$$

or

$$\bar{N}_r \geq \frac{\bar{l}_p \frac{P_0 V_c^{\beta_p}}{\bar{v}_p} \left(\frac{\phi}{\mu_0 V_c^{\beta_B - 1}} + 1 \right)}{\frac{\bar{r}_r}{\mu_0 V_c^{\beta_B - 1}} - \bar{l}_r \left(\frac{\eta}{\mu_0 V_c^{\beta_B - 1}} + 1 \right)}. \quad \textbf{(S46)}$$

these forms could also be written in terms of **Equation 9** but, as noted in the main text, then the scaling of genome size $\frac{\bar{n}_p \bar{v}_p}{\bar{l}_p}$ with cell size must be accounted for separately.

Measured Degradation Rates

Given the half-life, $t_{1/2}$, for proteins we can find the specific degradation rates as $\phi = \ln(2)/t_{1/2}$. The reported minimum and maximum values from [16] would give ϕ between 4.58×10^{-6} (s^{-1}) and 1.60×10^{-5} (s^{-1}) with a reported median of 8.37×10^{-6} (s^{-1}). Other reported typical rates for degradation include 3.03×10^{-5} (s^{-1}) [17] and 2.53×10^{-5} (s^{-1}) [18]. These rates are comparable but slightly smaller than the best fit of $\phi = 6.20 \times 10^{-5}$ (s^{-1}) from fits of V_r to measured data. For these fits we assumed that since protein degradation rates are often measured over the entire set of proteins it is fair to take $\eta = \phi$ (note that V_r is relatively insensitive to η and the approximation of no ribosome degradation $\eta = 0$ does not alter the fitted values of ϕ by much nor the goodness of fit). In testing the number of ribosomes we considered both choices for modeling growth rate, the power-law approximation and the asymptotic form, that lead to **Equations S43 and S44**. For practically fitting the data, the power-law approximation for μ gives a better fit because it lacks the lower asymptote where growth rate goes to zero. However, we consider that **Equation S43** with the measured value for $\eta = \phi$ is the fundamental prediction for the lower bound on the number of ribosomes and the data show that this limit is rarely violated.

For predictions of the lower bound on cell size in the main text we use the ribosome form with the power-law approximation for μ, however all forms of μ are approximately the same for

large cells, and thus predictions of the "ribosome catastrophe" and upper bound on bacterial cell size are insensitive to these choices for μ.

Cellular Envelope

The cellular envelope consists of the membrane and peptidoglycan layers. Gram-negative bacteria are maximizing for membrane thickness as opposed to peptidoglycan thickness, whereas gram-positive bacteria are maximizing for peptidoglycan thickness as opposed to membrane thickness. For the cellular envelope we take the effective thickness to be: 1.) In mycoplasma, for example, there is just a single layer of membrane to minimize the total volume. The smallest measured value we have found is $r_{env} = 4.5$ nm thick [19]; 2.) For Gram-negative bacteria, we add the outer membrane and peptidoglycan layers to the inner membrane thickness, where a good estimate for the total thickness is 40.5 nm [19]. 3.) For Gram-positive bacteria, we subtract the thickness of the outer membrane from the total thickness of the Gram-negative envelope: $40.5 - 13 = 27.5$ nm [20]. It should be noted that the peptidoglycan layer is a significant portion of the effective envelope thickness for the latter two cases, and this may be variable in species, especially at the small end where it is likely greatly reduced.

To calibrate the average percentage of the envelope volume that is occupied by proteins we considered values for *E. coli*. Here we calculate the total envelope volume, including the proteins, from overall cell volume, and subtract from this the volume of membrane bound proteins. The membrane bound proteins makeup about 30% of the proteome [21] and taking cell volumes of 0.7 to 6 μm^3, along with observed scaling relationship between cell volume and protein volume, would give $\bar{p}_p = 0.147$ to 0.151.

It should be noted that the methods for measuring protein content may not completely extract all membrane bound proteins and so measured values of V_p may be underestimates which would effect the value of $\bar{p}_p$. However, the purpose of $\bar{p}_p$ is to avoid double counting the envelope volume with the volume already counted in the protein volume. Thus if V_p does not include some of the membrane-bound proteins then the volume of these proteins is not subtracted from V_{env} and $V_{env} + V_p$ should be accurately represented given the calculated value of $\bar{p}_p$.

Similarly, it should be noted that there may be overlap in some studies between the volume being measured for the total protein pool and the proteins of ribosomes. However, for most of the range of cell sizes this should only introduce small errors into our analysis because the volume of ribosomes in an order to two orders of magnitude smaller than the protein volume.

tRNA and mRNA

In the main text (**Equations 15 and 16**) we present the proportionality between both $tRNA$ and $mRNA$ and the total number of ribosomes and here it is useful to explicitly give the full form with the volume-dependent relationship for N_r substituted into these equations. Taking

$$V_{tRNA} = \bar{v}_{tRNA}\bar{n}_{tRNA}N_r \tag{S47}$$

$$V_{mRNA} = \bar{v}_{mRNA}\bar{n}_{mRNA}N_r \tag{S48}$$

and using **Equation S43** would give

$$V_{tRNA} = \frac{\bar{v}_{tRNA}\bar{n}_{tRNA}\bar{l}_p\frac{P_0V_c^{\beta_p}}{\bar{v}_p}\left(\frac{\phi\ln\left[\frac{1-(B_m/B_0)(V_cd_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_cd_c)^{1-\beta_B}}\right]}{(B_m/E_m)(1-\beta_B)\ln[\epsilon]}+1\right)}{\frac{\bar{r}_r\ln\left[\frac{1-(B_m/B_0)(V_cd_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_cd_c)^{1-\beta_B}}\right]}{(B_m/E_m)(1-\beta_B)\ln[\epsilon]}-\bar{l}_r\left(\frac{\eta\ln\left[\frac{1-(B_m/B_0)(V_cd_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_cd_c)^{1-\beta_B}}\right]}{(B_m/E_m)(1-\beta_B)\ln[\epsilon]}+1\right)} \tag{S49}$$

$$V_{mRNA} = \frac{\bar{v}_{mRNA}\bar{n}_{mRNA}\bar{l}_p\frac{P_0V_c^{\beta_p}}{\bar{v}_p}\left(\frac{\phi\ln\left[\frac{1-(B_m/B_0)(V_cd_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_cd_c)^{1-\beta_B}}\right]}{(B_m/E_m)(1-\beta_B)\ln[\epsilon]}+1\right)}{\frac{\bar{r}_r\ln\left[\frac{1-(B_m/B_0)(V_cd_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_cd_c)^{1-\beta_B}}\right]}{(B_m/E_m)(1-\beta_B)\ln[\epsilon]}-\bar{l}_r\left(\frac{\eta\ln\left[\frac{1-(B_m/B_0)(V_cd_c)^{1-\beta_B}}{1-\epsilon^{1-\beta_B}(B_m/B_0)(V_cd_c)^{1-\beta_B}}\right]}{(B_m/E_m)(1-\beta_B)\ln[\epsilon]}+1\right)} \tag{S50}$$

and the power-law from **Equation S44** (which considers the power-law approximation for μ) could also be substituted for N_r to find the alternative forms of

$$V_{tRNA} = \frac{\bar{v}_{tRNA}\bar{n}_{tRNA}\bar{l}_p\frac{P_0V_c^{\beta_p}}{\bar{v}_p}\left(\frac{\phi}{\mu_0V_c^{\beta_B-1}}+1\right)}{\frac{\bar{r}_r}{\mu_0V_c^{\beta_B-1}}-\bar{l}_r\left(\frac{\eta}{\mu_0V_c^{\beta_B-1}}+1\right)} \tag{S51}$$

$$V_{mRNA} = \frac{\bar{v}_{mRNA}\bar{n}_{mRNA}\bar{l}_p\frac{P_0V_c^{\beta_p}}{\bar{v}_p}\left(\frac{\phi}{\mu_0V_c^{\beta_B-1}}+1\right)}{\frac{\bar{r}_r}{\mu_0V_c^{\beta_B-1}}-\bar{l}_r\left(\frac{\eta}{\mu_0V_c^{\beta_B-1}}+1\right)} \tag{S52}$$

where in all cases it should be noted that $\frac{P_0V_c^{\beta_p}}{\bar{v}_p}$ has been substituted for N_p.

We use the average gene length (in nucleotides), $\bar{l}_p$, to estimate the average volume of an mRNA as $\bar{v}_{mRNA} \approx 1.43\times10^{-24}$ m^3. Similarly, taking the length of a tRNA to be ≈ 85 nucleotides gives an approximate volume of $\bar{v}_{tRNA} \approx 3.10 \times 10^{-26}$ m^3 [22].

To calculate the number of mRNA per ribosome we use the number of mRNA per protein, 540 [23], and the number of proteins per ribosome in *E. coli*, 588 , calculated from our two cross species correlations to find $\bar{n}_{mRNA} \approx 1.08$. For tRNA it has been reported that $\bar{n}_{tRNA} = 9.3$ [15].

Cellular Density

Given that our analysis describes the full cellular composition for every size of bacteria (including the inferred volume of water: $V_w = V_c - (V_{DNA} + V_p + V_r + V_{env} + V_{tRNA} + V_{mRNA})$) it is possible to calculate cellular density across the range of bacteria. Taking the densities of cellular components to be $\bar{d}_p = 1.37 \times 10^6$ g m^3 [10,24,25], $\bar{d}_{DNA} = 2 \times 10^6$ g m^3 [24,26], $\bar{d}_{RNA} = 2 \times 10^6$ g m^3 [24], $\bar{d}_{mem} = 1.05 \times 10^6$ g m^3 [27], for proteins, DNA, RNA, and membrane respectively, and noting that ribosomes are a 2:1 RNA to protein ratio [28], we have $d_r = 1.79 \times 10^6$ g m^3, so we have that the total density of the cell is

$$d_c = \frac{\bar{d}_{DNA}V_{DNA} + \bar{d}_pV_p + \bar{d}_rV_r + \bar{d}_{mem}V_{env} + \bar{d}_{rna}V_{tRNA} + \bar{d}_{rna}V_{mRNA} + d_wV_w}{V_c} \qquad \textbf{(S53)}$$

where the density of water is given by $d_w = 1 \times 10^6$ g m^3. **Figure S5** gives the calculated cellular density across the range of bacteria where we find that density is greatest for the smallest and largest bacteria with a minimum value of 1.05×10^6 g m^{-3} for an intermediate cell size of 4.86×10^{-18} m^3.

References

1. Kempes, C. P., Dutkiewicz, S. & Follows, M. J. Growth, metabolic partitioning, and the size of microorganisms. *Proc of the Natl Acad of Sci* **109**, 495–500 (2012).

2. DeLong, J., Okie, J., Moses, M., Sibly, R. & Brown, J. Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proc Natl Acad Sci USA* **107**, 12941–12945 (2010).

3. Dolezel, J., Bartos, J., Voglmayr, H. & Greilhuber, J. Nuclear dna content and genome size of trout and human. *Cytometry. Part A: the journal of the International Society for Analytical Cytology* **51**, 127–128 (2003).

4. Mandelkern, M., Elias, J. G., Eden, D. & Crothers, D. M. The dimensions of dna in solution. *Journal of molecular biology* **152**, 153–161 (1981).

5. Baldwin, W. W., Myer, R., Powell, N., Anderson, E. & Koch, A. L. Buoyant density of escherichia coli is determined solely by the osmolarity of the culture medium. *Archives of microbiology* **164**, 155–157 (1995).

6. Loferer-Krößbacher, M., Klima, J. & Psenner, R. Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis. *Applied and Environmental Microbiology* **64**, 688–694 (1998).

7. Norland, S., Heldal, M. & Tumyr, O. On the relation between dry matter and volume of bacteria. *Microbial Ecology* **13**, 95–101 (1987).

8. Phillips, R., Kondev, J., Theriot, J. & Garcia, H. *Physical biology of the cell* (Garland Science, 2012).

9. Neidhardt, F., Umbarger, H. & Eds Neidhardt, F. *Chemical composition of Escherichia coli. In Escherichia coli and Salmonella typhimurium. Cellular and molecular biology. Second Edition.* (American Society for Microbiology, 1996).

10. Erickson, H. P. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol Proced Online* **11**, 32–51 (2009).

11. Dill, K. A., Ghosh, K. & Schmit, J. D. Physical limits of cells and proteomes. *Proceedings of the National Academy of Sciences* **108**, 17876–17882 (2011).

12. Xu, L. *et al.* Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. *Molecular biology and evolution* **23**, 1107–1108 (2006).

13. Gabashvili, I. S. *et al.* Solution structure of the e. coli 70s ribosome at 11.5 å resolution. *Cell* **100**, 537–549 (2000).

14. Zhu, J., Penczek, P. A., Schröder, R. & Frank, J. Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: Procedure and application to the 70s escherichia coli ribosome. *Journal of structural biology* **118**, 197–219 (1997).

15. Bremer, H., Dennis, P. P. & Eds. Neidhardt, F. *Modulation of chemical composition and other parameters of the cell by growth rate. In Escherichia coli and Salmonella typhimurium. Cellular and molecular biology. Second Edition.* (American Society for Microbiology, 1996).

16. Maier, T. *et al.* Quantification of mrna and protein and integration with protein turnover in a bacterium. *Molecular systems biology* **7**, 511 (2011).

17. Jayapal, K. P. *et al.* Multitagging proteomic strategy to estimate protein turnover rates in dynamic systems. *Journal of proteome research* **9**, 2087–2097 (2010).

18. Trötschel, C., Albaum, S. P. & Poetsch, A. Proteome turnover in bacteria: current status for corynebacterium glutamicum and related bacteria. *Microbial biotechnology* **6**, 708–719 (2013).

19. Gan, L., Chen, S. & Jensen, G. J. Molecular organization of gram-negative peptidoglycan. *Proceedings of the National Academy of Sciences* **105**, 18953–18957 (2008).

20. Bayer, M. Zones of membrane adhesion in the cryofixed envelope of escherichia coli. *Journal of structural biology* **107**, 268–280 (1991).

21. Kalwarczyk, T., Tabaka, M. & Holyst, R. Biologistics – diffusion coefficients for complete proteome of escherichia coli. *Bioinformatics* **28**, 2971–2978 (2012).

22. Spirin, A. S. *Ribosomes* (Springer Science & Business Media, 1999).

23. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature biotechnology* **25**, 117–124 (2007).

24. Delgado, F. F. *et al.* Intracellular water exchange for measuring the dry mass, water mass and changes in chemical composition of living cells. *PloS one* **8**, e67590 (2013).

25. Görisch, S. M., Lichter, P. & Rippe, K. Mobility of multi-subunit complexes in the nucleus: accessibility and dynamics of chromatin subcompartments. *Histochemistry and cell biology* **123**, 217–228 (2005).

26. Panijpan, B. The buoyant density of dna and the g + c content. *Journal of Chemical Education* **54**, 172–173 (1977).

27. Rickwood, D., Eds Birnie, G. & Rickwood, D. *In centrifugal separations in molecular and cell biology* (Butterworths, London, 1978).

28. Melnikov, S. *et al.* One core, two shells: bacterial and eukaryotic ribosomes. *Nature structural & molecular biology* **19**, 560–567 (2012).

29. Shuter, B. J., Thomas, J., Taylor, W. D. & Zimmerman, A. M. Phenotypic correlates of genomic dna content in unicellular eukaryotes and other cells. *American Naturalist* **122**, 26–44 (1983).

30. West, G. B. & Brown, J. H. The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *Journal of Experimental Biology* **208**, 1575–1592 (2005).

31. Milo, R. What is the total number of protein molecules per cell volume? a call to rethink some published values. *Bioessays* **35**, 1050–1055 (2013).

32. Dethlefsen, L. & Schmidt, T. M. Performance of the translational apparatus varies with the ecological strategies of bacteria. *Journal of bacteriology* **189**, 3237–3245 (2007).

33. Tempest, D. & Hunter, J. The influence of temperature and ph value on the macro-molecular composition of magnesium-limited and glycerol-limited aerobacter aerogenes growing in a chemostat. *Journal of general microbiology* **41**, 267–273 (1965).

34. Tempest, D., Hunter, J. & Sykes, J. Magnesium-limited growth of aerobacter aerogenes in a chemostat. *Journal of general microbiology* **39**, 355–366 (1965).

35. Simon, M. & Azam, F. Protein content and protein synthesis rates of planktonic marine bacteria. *Marine ecology progress series. Oldendorf* **51**, 201–213 (1989).

36. Valgepea, K., Adamberg, K., Seiman, A. & Vilu, R. Escherichia coli achieves faster growth by increasing catalytic and translation rates of proteins. *Molecular BioSystems* **9**, 2344–2358 (2013).

37. Seybert, A., Herrmann, R. & Frangakis, A. S. Structural analysis of mycoplasma pneumoniae by cryo-electron tomography. *Journal of structural biology* **156**, 342–354 (2006).

38. Fegatella, F., Lim, J., Kjelleberg, S. & Cavicchioli, R. Implications of rrna operon copy number and ribosome content in the marine oligotrophic ultramicrobacterium sphingomonassp. strain rb2256. *Applied and environmental microbiology* **64**, 4433–4438 (1998).

39. Luef, B. *et al.* Diverse uncultivated ultra-small bacterial cells in groundwater. *Nature communications* **6**, 6372 (2015).

Table S1: Power law fit parameters and statistics for each data source.

Property	**Best Fit Parameters**	95% **Confidence Interval**	R^2
Genome Volume			
D_0	3.00×10^{-17} (m^3 DNA $\cdot$ $\left(\mathrm{m}^3 \text{ Cell}\right)^{-\beta_D}$)	$\{9.31 \times 10^{-18}, 9.64 \times 10^{-17}\}$	0.60
β_D	0.21	$\{0.18, 0.24\}$	
Protein Volume			
P_0	3.42×10^{-7} (m^3 Protein $\cdot$ $\left(\mathrm{m}^3 \text{ Cell}\right)^{-\beta_P}$)	$\{3.33 \times 10^{-8}, 3.51 \times 10^{-6}\}$	0.87
β_P	0.70	$\{0.64, 0.75\}$	
Ribosome Volume			
R_0	1.54×10^{-7} (m^3 Ribosomes $\cdot$ $\left(\mathrm{m}^3 \text{ Cell}\right)^{-\beta_R}$)	$\{2.57 \times 10^{-10}, 9.30 \times 10^{5}\}$	0.62
β_R	0.73	$\{0.59, 0.88\}$	
Ribosome Volume (Full Model)			
$\eta = \phi$	6.19×10^{-5} (s^{-1})	$\{1.20 \times 10^{-5}, 3.19 \times 10^{-4}\}$	NA

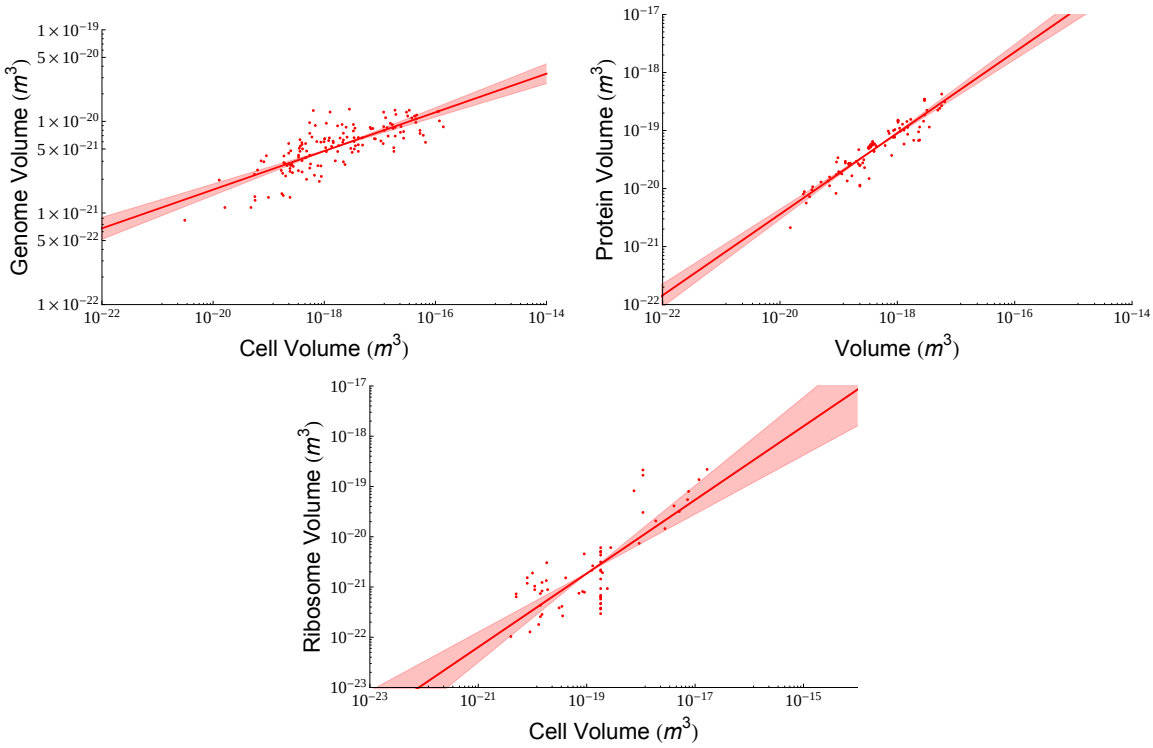

Figure S1: The scaling of component volume plotted in the natural range of y-values for total (a) genome volume, (b) protein volume, and (c) ribosomal volume.

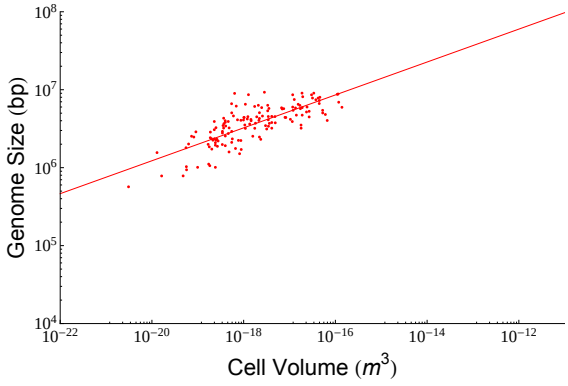

Figure S2: The volume-dependent scaling of genome length. The data are from [2,29,30].

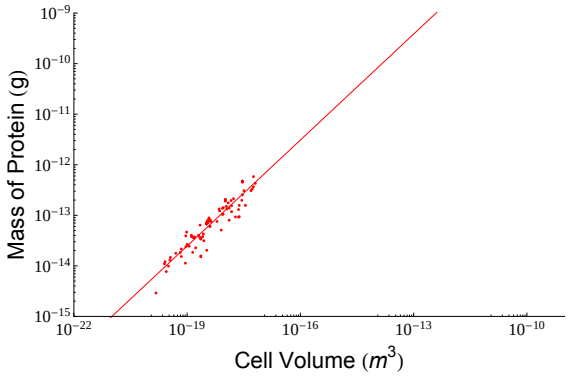

Figure S3: The volume-dependent scaling of protein content. The data compilation includes, but is not limited to, [31–36].

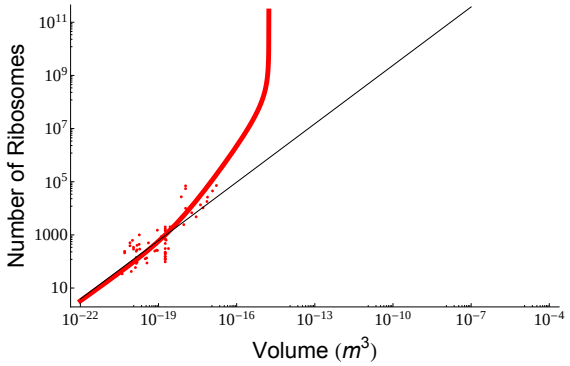

Figure S4: The volume-dependent scaling for the number of ribosomes. The black curve is a best fit power law and the red curve is the prediction from Equation **S46** of the main text.The data compilation includes, but is not limited to, [15,37–39].

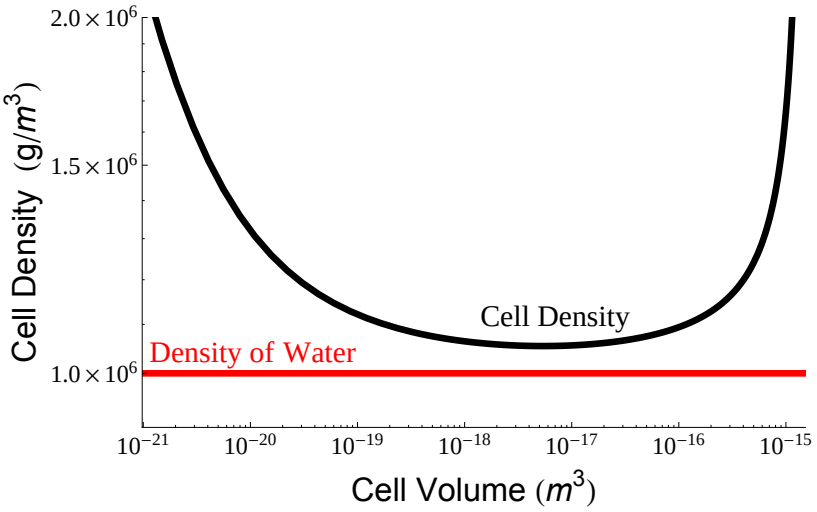

Figure S5: The volume-dependent scaling for the calculated total cellular density. The black curve is the calculated density and the red curve is the reference value for the density of water.