Fundamental limits on the rate of bacterial cell division

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

***For correspondence:**

[†]These authors contributed equally to this work

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

Abstract Recent years have seen a deluge of experiments dissecting the relationship between bacterial growth rate, cell size, and protein content, quantifying the abundance of proteins across growth conditions with unprecedented resolution. However, we still lack a rigorous understanding of what sets the scale of these quantities and when protein abundances should (or should not) depend on growth rate. Here, we seek to quantitatively understand this relationship across a collection of *Escherichia coli* proteomic data sets covering ≈ 4000 proteins and 31 growth conditions. We estimate the basic requirements for steady-state growth by considering key processes in nutrient transport, energy generation, cell envelope biogenesis, and the central dogma, from which ribosome biogenesis emerges as a primary determinant of growth rate. We conclude by exploring a model of ribosomal regulation as a function of the nutrient supply, revealing a mechanism that ties cell size and growth rate to ribosomal content.

Appendix for: Fundamental limits on the rate of bacterial cell division

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

***For correspondence:**

[†]These authors contributed equally to this work

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

42 ## Contents

Calculation of Complex Abundance

All protein data quantified the abundance of individual proteins per cell. However, this work requires estimates on the abundance of individual protein *complexes*, rather than the copy number of individual proteins. In this section, we outline the approach we used to annotate proteins as being part of a macromolecular complex and how we computed their absolute abundances per cell.

Protein complexes, and proteins individually, often have a variety of names, both longform and shorthand. As individual proteins can have a variety of different synonyms, we sought to ensure that each protein annotated in the data sets used the same synonym. To do use, we relied heavily on the EcoCyc Genomic Database (***Keseler et al., 2017***). Each protein in available data sets included an annotation of one of the gene name synonyms as well as an accession ID – either a UniProt or Blattner "b-number". We programmatically matched up individual accession IDs between the proteins in different data sets. In cases where accession IDs matched but the gene names were different, we manually verified that the gene product was the same between the datasets and chose a single synonym. All code used in the data cleaning and unification procedures can be found on the associated [GitHub repository] (DOI:XXX) associated with this paper as well as on the associated paper website.

With each protein conforming to a single identification scheme, we then needed to identify the molecular complexes each protein was a member of. Additionally, we needed to identify how many copies of each protein were present in each complex (i.e. the subunit copy number) and compute the estimated abundance complex that accounted for fluctuations in subunit stoichiometry. To map proteins to complexes, we accessed the EcoCyc *E. coli* database ***Keseler et al.*** (***2017***) using PathwayTools version 23.0 ***Karp et al.*** (***2019***). With a license for PathWay Tools, we mapped each unique protein to its annotated complexes via the BioCyc Python package. As we mapped each protein with *all* of its complex annotations, there was redundancy in the dataset. For example, ribosomal protein L20 (RplT) is annotated to be a component of the 50S ribosome (EcoCyc complex `CPLX-03962`) as well as a component of the mature 70S ribosome (EcoCyc complex `CPLX-03964`).

In addition to the annotated complex, we collected information on the stoichiometry of each macromolecular complex. For a complex with N_{subunits} protein species, for each protein subunit i we first calculate the number of complexes that *could* be formed given the measured protein copy numbers per cell,

$$N_{\text{complex}}(\text{subunit i}) = \frac{P^{(\text{measured})}_{\text{subunit i}}}{m_{\text{subunit i}}}. \tag{1}$$

Here, $P^{(\text{measured})}_{\text{subunit i}}$ refers to the measured protein copy number of species i, and m refers to the number of monomers present for that protein in the complex. For example, the 70S mature ribosome complex has 55 protein components, all of which are present in a single copy except L4 (RplL), which is present in 4 copies (m = 4). For each ribosomal protein, we then calculate the maximum number of complexes that could be formed using ***Equation 1***. This example, along with example from 5 other macromolecular complexes, can be seen in ***Figure 1***.

It is important to note that measurement noise, efficiency of protein extraction, and differences in protein stability will mean that the precise value of each calculation will be different for each component of a given complex. Thus, to report the total complex abundance, we use the arithmetic mean of across all subunits in the complex,

$$\langle N_{\text{complex}} \rangle = \frac{1}{N_{\text{subunits}}} \sum_{i}^{N_{\text{subunits}}} \frac{P^{(\text{measured})}_{i}}{m_{\text{subunit i}}}. \tag{2}$$

in ***Figure 1***, we show this mean value as a grey line for a variety of different complexes. Additionally, we have built an interactive figure accessible on the paper website where the validity of this approach can be examined for any complex with more than two subunits (thus, excluding monomers and dimers).

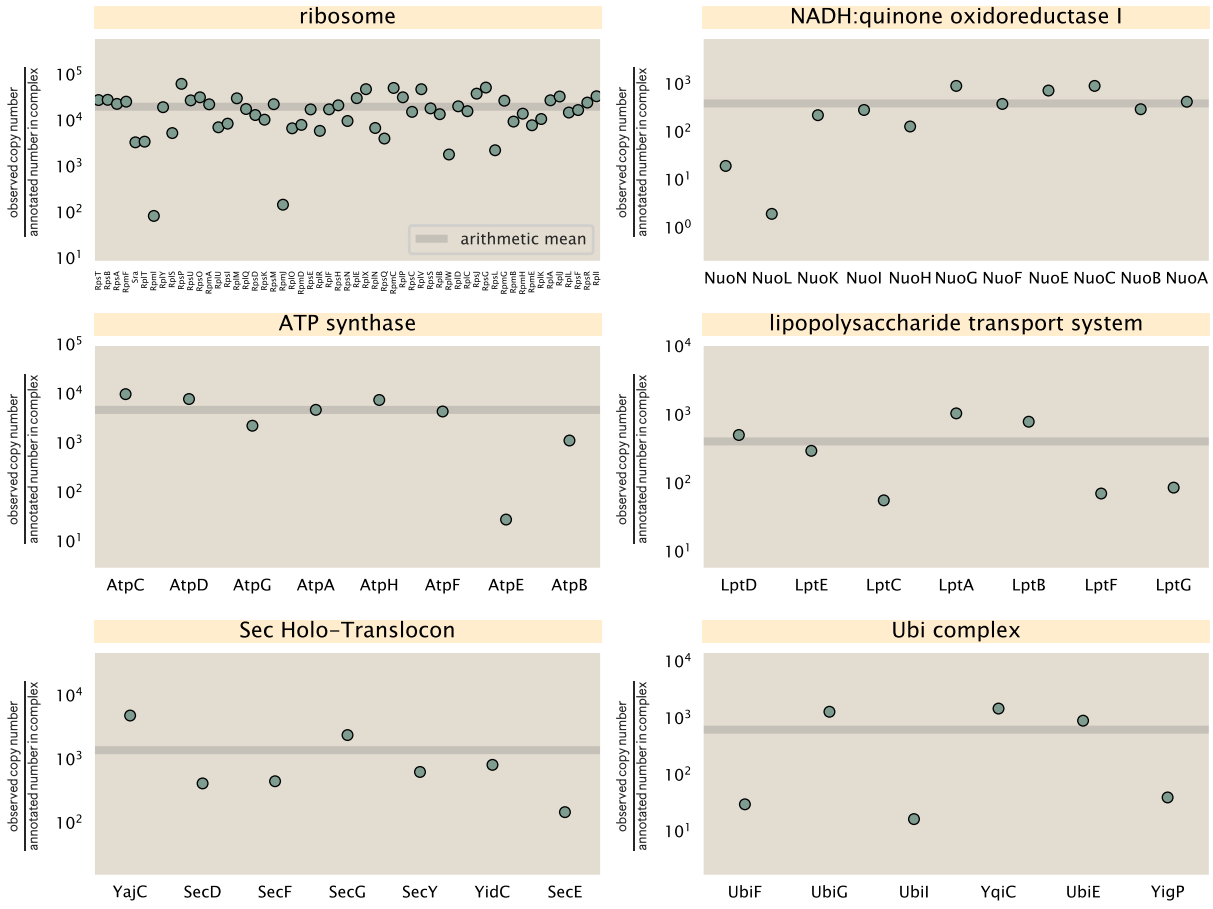

Figure 1. Calculation of the mean complex abundance from measurements of single subunits. Six of the largest complexes (by number of subunits) in *E. coli*. Points correspond to the maximum number of complexes that can be formed given measurement of that individual protein. Solid grey line corresponds to the arithmetic mean across all subunits. These data correspond to measurements from ***Schmidt et al.*** (***2016***) in a glucose-supplemented minimal growth medium.

References

Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S. M., and Subhraveti, P. (2019). The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, 20(4):1085–1093.

Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., Latendresse, M., Muñiz-Rascado, L., Ong, Q., Paley, S., Peralta-Gil, M., Subhraveti, P., Velázquez-Ramírez, D. A., Weaver, D., Collado-Vides, J., Paulsen, I., and Karp, P. D. (2017). The EcoCyc database: reflecting new knowledge about Escherichia coliK-12. *Nucleic Acids Research*, 45(D1):D543–D550.

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