doi:10.1038/nature22998

Ribosomes are optimized for autocatalytic production

Shlomi Reuveni[1], Måns Ehrenberg[2] & Johan Paulsson[1]

Many fine-scale features of ribosomes have been explained in terms of function, revealing a molecular machine that is optimized for error-correction, speed and control. Here we demonstrate mathematically that many less well understood, larger-scale features of ribosomes—such as why a few ribosomal RNA molecules dominate the mass and why the ribosomal protein content is divided into 55–80 small, similarly sized segments—speed up their autocatalytic production.

Ribosomes translate sequences of nucleic acids into sequences of amino acids[1]. Their features are therefore typically explained in terms of how they affect translation[1]. However, in recent years it has also become clear that ribosomes are exceptional as products of the ribosomal machinery[2–4]. Not only do ribosomal proteins (r-proteins) make up a large fraction of the total protein content in many cells[5], but the autocatalytic nature of ribosome production introduces additional constraints. Specifically, the ribosome doubling time places a hard bound on the cell doubling time, because for every additional ribosome to share the translation burden there is also one more to make[2,3]. Even for the smallest and fastest ribosomes, it takes at least 6 min, and typically much longer, for one ribosome to make a new set of r-proteins (Supplementary Information); and this estimate does not account for the substantial time that is invested in the synthesis of ternary complexes[4]. This bound seems to explain the observed limits on bacterial growth, because ribosomes must also spend much of their time making other proteins[6–11], and shows that ribosomes are under very strong selective pressure to minimize the time they spend reproducing. Similar principles might also apply to some eukaryotes, because the ribosomes of eukaryotes are larger and slower[1,12]. In fact, even organisms in which cell doubling times are not limited by ribosome doubling times would benefit from faster ribosome production, allowing ribosomes to spend more of their time producing the rest of the proteome. This efficiency constraint was recently shown to have broad physiological consequences for cells[3,4,11,13–15], and here we demonstrate mathematically that it might also explain many broader features of the ribosome (Fig. 1).

Why do ribosomes have so many small r-proteins?

The total time τ it takes one ribosome to elongate a set of r-proteins for a new ribosome is proportional to the total combined length of the r-proteins. However, the average time that each ribosome must dedicate to that process also depends on how the total protein mass is divided up into individual segments. Because so many ribosomes produce r-proteins in parallel and because each ribosome consists of many r-proteins, complete sets of r-proteins will form by chance more quickly than individual ribosomes could elongate that amount of protein on their own. For example, if ribosomes consisted of two similarly sized r-proteins, complete pairs would start to form $\tau/2$ time units after production was initiated, and those newly made ribosomes could then share the translation burden. Similarly, if ribosomes contained n proteins of equal genomic length, then the nascent peptides that cannot contribute to new ribosomes would be about n times shorter and mature r-proteins would be released n times faster (Fig. 2a). Increasing n thus reduces the minimum fraction of time φ that ribosomes must spend on their own production:

$$\varphi \geq n[2^{\tau/(nT_{\mathrm{gen}})} - 1] \quad (1)$$

where T_{gen} is the cell generation time (see Supplementary Information). This expression asymptotically approaches the previously determined bound on growth[4,10,11] $\tau\ln(2)/T_{\mathrm{gen}}$ in the limit of large n, and shows that dividing the ribosomal protein content into a large number of small proteins can reduce the time taken to make new ribosomes by as much as 30% (Fig. 2b), but with diminishing returns at larger n. Because increasing n also has disadvantages that are not accounted for above, there should be an optimal number n_{opt} of r-proteins.

One disadvantage of dividing the ribosomal protein complement into many small segments is that the different proteins will not be made in exactly the same numbers in individual cells. Even if production rates for all r-proteins were perfectly matched on average, the probabilistic nature of expression in individual cells would inevitably lead to more of some r-proteins and fewer of others. The assembly of complete ribosomes is thus limited by the r-protein that is present in the lowest number, creating a surplus of all other r-proteins and thereby reducing efficiency. Because each r-protein runs the risk of being under-produced, this surplus should increase with n. However, for the types of distributions often observed for gene expression, the average surplus is an exceedingly damped function of n, closely following $\sqrt{\ln(n)}$ (Extended Data Fig. 1 and Supplementary Information). Combining this diminishing disadvantage with the diminishing returns of shorter

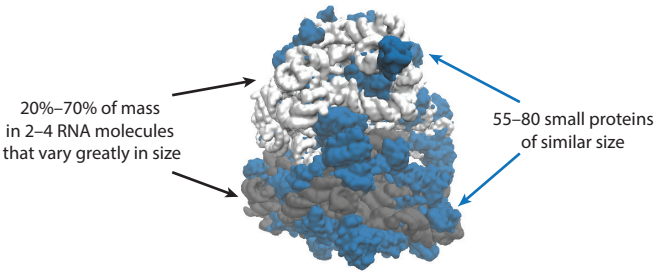


Figure 1 | Many unusual features of ribosomes are not well understood. With the exception of mitochondrial ribosomes, most of the mass in ribosomes is in a few RNA molecules that are very large on average, but vary greatly in size (grey), whereas the proteins (blue) are unusually small, numerous and similar to each other in size.

[1]Department of Systems Biology, HMS, Harvard University, 200 Longwood Avenue, Boston, Massachusetts 02115, USA. [2]Department of Cell and Molecular Biology, Uppsala University, Uppsala Biomedicinska Centrum (BMC) Husargatan 3, Uppsala, Sweden.

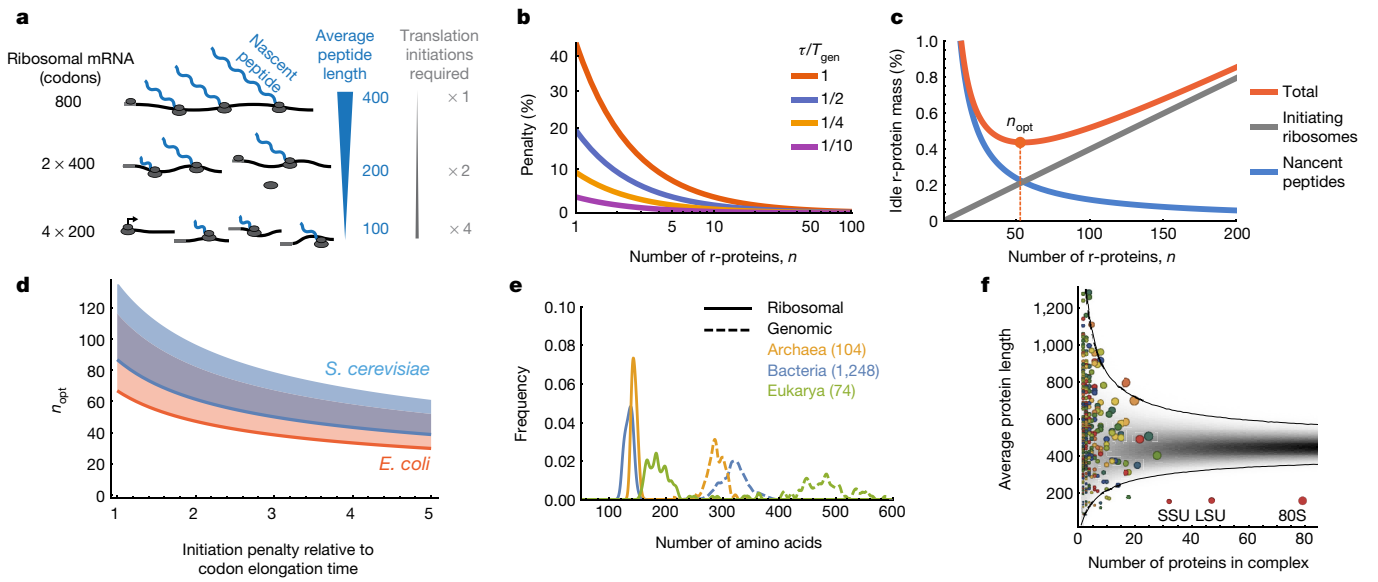


Figure 2 | Optimal number of r-proteins for ribosome biogenesis. **a**, Nascent r-protein peptides (blue) cannot participate in translation before their own translation is complete. Subdividing the protein mass into more and smaller pieces reduces the inactive fraction, but also increases the number of translation initiations that are required. Peptide length is given in terms of the number of amino acids. Grey circles represent ribosomes. **b**, The penalty, due to finite number of r-proteins, that is incurred on the minimum fraction of time (see equation (1)) that ribosomes must spend on their own production, plotted as a function of the number of r-proteins n for different values of τ/T_{gen} (1, red; 1/2, blue; 1/4, yellow; 1/10, magenta). The penalty is calculated relative to $\tau\ln(2)/T_{gen}$, the asymptotic value in the limit of large n: $n(2^{\tau/(nT_{gen})}-1)/[\tau\ln(2)/T_{gen}]-1$. The maximum penalty occurs for $\tau/T_{gen}=1$ and $n=1$ and is $1/\ln(2)\approx 1.44$. **c**, The fraction of dedicated r-protein mass that is elongationally idle (red), due to a combination of nascent r-protein peptides (blue) and ribosomes occupied in initiation (grey), is minimized for an optimal number of r-proteins n_{opt}. **d**, The optimal number of r-proteins n_{opt} (see equation (2) and **c**) for *E. coli* (orange) and *S. cerevisiae* (blue), in the range of the expected ratios between initiation and elongation times (see main text). The shaded areas above the lines correspond to values of n_{opt} that would be accessible if ribosomes were able to produce slightly fewer of the larger subunits, which bind only at the last step of initiation (Supplementary Information). **e**, Across organisms, the average lengths of proteins (in terms of the number of amino acids) in ribosomes are 2–3 times shorter than in genomes[21]. Distribution of the ribosomal (solid lines) and genomic (dashed lines) averages are computed across 104 archaea (orange), 1,248 bacteria (blue) and 74 eukarya (green). **f**, The average length of proteins in *S. cerevisiae* multiprotein complexes versus the number of proteins in the complex. Each coloured circle corresponds to a known complex[22] (the area is proportional to total number of amino acids). The different colours are to aid presentation only. Grey shading represents the predicted probability density when drawing random genes from the genome. Most of the 403 complexes shown fall within the expected 99% confidence interval (black lines). The main statistical outlier is the ribosome 80S, which includes both the small and large subunits (SSU and LSU) independently, and falls approximately 7 standard deviations below the mean.

nascent peptides in equation (1) produces no effective upper limit on n: all $n > n_{opt}$ are virtually equally optimal (Extended Data Fig. 1; Supplementary Information), even without expression control. *E. coli* cells also reduce this problem further via negative feedback loops[16,17] that prevent surplus accumulation, and by expressing the r-proteins in operons such that terminating ribosomes skip directly to the next start site[18], possibly ensuring that r-proteins in the same operon are expressed in almost identical numbers[16,17,19].

However, there is another countervailing selective force that limits n for a given total ribosome mass. For every additional r-protein, one more translation initiation is required, sequestering ribosomes from the elongation process and creating another form of overhead (Fig. 2a). With τ_{oh} as the initiation overhead time during which a ribosome is occupied making ribosomal proteins without actually elongating, and again τ as the time it takes a ribosome to elongate another set of r-proteins,

$$n_{opt} = A\sqrt{\tau/\tau_{oh}} \qquad (2)$$

where $A \approx 1$ (Supplementary Information). The square root reflects the fact that an optimal number of r-proteins n_{opt} minimizes the total idleness, which is due to the length of nascent peptides, which decreases proportionally to $1/n$, and to ribosome sequestration from initiation, which increases proportionally to $n\tau_{oh}/\tau$ (Fig. 2c).

The exact initiation penalties—the times that ribosomes are sequestered to initiate translation—are not known *in vivo*. However, we need to know only how these times increase with n, and only relative to elongation rates. An estimate can be determined in *E. coli* from an observed average spacing of about 15 codons[20] between initiating ribosomes, but this might greatly exaggerate the relevant initiation penalty for several reasons. First, the previous ribosome must move away some distance from the initiation site before the next one can bind, but during that time the latter is free to bind other transcripts. If messenger RNA (mRNA) levels are abundant, then the previous ribosome can on average move away a longer distance before another ribosome binds the same transcript, increasing the spacing without decreasing efficiency. Second, the total time for which ribosomes must search the cytoplasm for mRNAs to produce n r-proteins should be roughly independent of n: producing twice as many r-proteins of half the size requires twice as many initiations, but also doubles the total concentration of the target mRNA for a given total investment in mRNA and thus approximately halves the time it takes each diffusing ribosome to find an r-protein mRNA. Therefore, only the time that the ribosome spends bound to the mRNA without elongating would be relevant here. Last, as noted above, most r-proteins in *E. coli* do not require typical initiations because terminating ribosomes continue directly to the next start site[18] in the operon. The relevant penalty due to initiation should consequently be a small fraction of the time it takes to elongate 15 codons, possibly as low as one or two codon equivalents.

Owing to the square-root effect, initiation penalties in the broad range of 1–5 codons lead to a prediction for n_{opt} of approximately $40 \lesssim n_{opt} \lesssim 85$ (Fig. 2d), consistent with the 56 r-proteins observed for *E. coli*, even though the first-principles derivation does not take into account the typical size of proteins or any structural properties. Formulated in terms of size, this prediction means that r-proteins

should be about 2–4 times smaller than average proteins. The average r-protein length in *E. coli* is indeed only about 130 amino acids, compared to about 315 amino acids over the genome[21]. This is approximately 6.5 standard deviations below average, given that there are 56 r-proteins, and the probability of observing such low averages when drawing genes randomly from the genome is less than 10^{-17} (Extended Data Fig. 2). Similar principles are observed in ribosomes across organisms (Fig. 2e, Extended Data Fig. 2, Supplementary Information), but generally not in other multiprotein complexes (Fig. 2f). There may also be other advantages of shorter r-proteins, such as the reduced risk of premature termination due to limited ribosome processivity, but this effect should be small (Supplementary Information).

Why are r-proteins so similar in length?

The efficiency argument above also has direct consequences for how similar in length ribosomal proteins should be to maximize the efficiency of their own production. For example, if 99% of the ribosomal protein mass was in a single r-protein, then it would not help to divide the rest into small pieces even if doing so would reduce the average length. Specifically, for a given production rate, the probability of finding a ribosome in the process of translating a particular r-protein is proportional to its length. This relationship has been shown experimentally[19] and arises because longer r-proteins take proportionally longer time to translate. Therefore, on average, if one of the r-proteins is twice as long as another, it not only has nascent peptides that are twice as long, but also occupies twice as many ribosomes, contributing four times more to the average length of nascent peptides. The average amino-acid length over all r-protein nascent peptides in cells is then not simply half the average genomic codon length, $\langle L\rangle/2$, as is the case when the r-proteins have identical lengths, but approximately $\langle L^2\rangle/(2\langle L\rangle)$—a function of the average square of the genomic codon length. Because by definition of variances $\sigma_L^2 = \langle L^2\rangle - \langle L\rangle^2$, the efficiency for a given average length is reduced by variation in length (Fig. 3a). With a coefficient of variance for the length distribution of r-proteins of $\mathrm{CV}_L = \sigma_L/\langle L\rangle$, equation (1) generalizes to (Supplementary Information)

$$\varphi \geq n\left\{2^{\tau/(nT_{\mathrm{gen}})}\left[1+\frac{1}{2}\left(\frac{\tau \ln(2)}{nT_{\mathrm{gen}}}\right)^2 \mathrm{CV}_L^2\right]-1\right\} \qquad (3)$$

Inspecting this expression reveals that, to maximize the efficiency of their own production, the r-proteins should not only be short, but also of similar length (Fig. 3b). We find that CV_L is substantially lower for r-proteins than for the genome overall for a wide range of organisms (Fig. 3c), whether or not we account for the fact that the r-proteins also are smaller on average (Extended Data Fig. 3, Supplementary Information). For bacteria, similar results hold for Gram-positives, whereas for Gram-negatives the effect is substantial except for the single large r-protein S1. However, S1 is only transiently associated with the ribosome and did not appear in initial crystal structures[22]. It has also been shown to be involved in the selective initial binding to some mRNAs[23] and is not produced in an operon with other r-proteins[16,17]. We therefore propose, supported by biochemical evidence[24], that cells use S1 to create specialized ribosomes, for example, with different balances between speed and accuracy. Regardless, this single, notable exception to the length-variation rule clearly has a different role in ribosomes from that of other r-proteins.

Why are ribosomes so rich in RNA?

Another unusual feature of ribosomes is that so much of their mass is based in ribosomal RNA (rRNA)[1,12]. This feature could also be explained by selection to minimize the time that each ribosome is occupied in self-production. With τ_{pol} as the translation time of the proteins of one RNA polymerase, each ribosome could produce $T_{\mathrm{gen}}/\tau_{\mathrm{pol}}$ new polymerases during each generation period of length T_{gen}, and each of these polymerases could produce rRNA for $T_{\mathrm{gen}}/\tau_{\mathrm{rRNA}}$ new ribosomes per generation period, where τ_{rRNA} is the time required to make one set of rRNA. The ratio of the time that ribosomes must spend producing additional RNA polymerase for one set of rRNA and that spent making one set of r-proteins is therefore (Supplementary Information)

$$\frac{\text{time for rRNA}}{\text{time for r-proteins}} = \frac{(T_{\mathrm{gen}}/\tau)\ln(2)}{(T_{\mathrm{gen}}/\tau_{\mathrm{pol}})(T_{\mathrm{gen}}/\tau_{\mathrm{rRNA}})} \qquad (4)$$

This ratio is typically very small, despite rRNA dominating the ribosome mass, and even when accounting for inactive RNA polymerases (Supplementary Information). For example, numbers for fast-growing *E. coli* (Fig. 4a) suggest that the time ribosomes invest in r-protein synthesis could be two orders of magnitude higher than for an equivalent mass of rRNA (Supplementary Information). The consideration of other auxiliary costs—such as the cost of making the proteins that are required to synthesize the nucleotides needed for rRNA versus that of the proteins that are required to produce the nucleotides, mRNA, amino acids, charged transfer RNA (tRNA), and initiation and elongation factors needed for r-proteins—further supports these conclusions (Supplementary Information).

The above analysis suggests a great efficiency advantage of using rRNA over protein, whenever chemically possible, and so could explain why ribosomes defy the general rule that enzymes are made mostly of protein[25] (Fig. 1). This finding does not mean that the role of rRNA is merely to ensure appropriate overall dimensions of the ribosome;

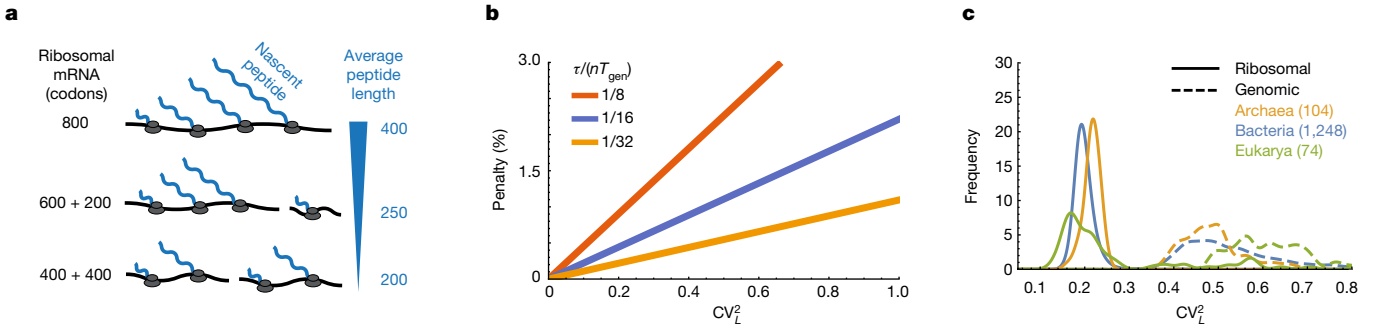


Figure 3 | Similarly sized r-proteins increase the efficiency of ribosome biogenesis. a, Dividing up a long r-protein (for example, 800 codons) into two unequal parts (600 and 200 codons) does not reduce the average length of the nascent peptides as much as does dividing it into equal parts (400 and 400 codons). **b**, The penalty that results from $\mathrm{CV}_L > 0$ that is incurred on the minimum fraction of time that ribosomes must spend on their own production (see equation (3)), normalized by $n[2^{\tau/(nT_{\mathrm{gen}})} - 1]$ (see equation (1)), is plotted as a function of the square of the coefficient of variance for the length distribution of r-proteins CV_L^2 for different values of $\tau/(nT_{\mathrm{gen}})$ (1/8, red; 1/16, blue; 1/32, yellow). **c**, Relative frequency distributions of the normalized variance (CV_L^2; see equation (3)) of protein length in ribosomes (solid lines) and genomes (dashed lines), normalized to integrate to 1, showing substantially lower CVs for ribosomes. The distributions are shown for 104 archaea (orange), 1,248 bacteria (blue) and 74 eukarya (green).

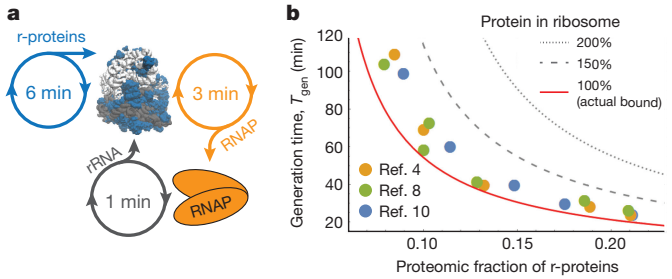


Figure 4 | Compared to r-proteins, rRNA production requires much less ribosome involvement. a, Ribosome auto-production takes up ribosome time to make r-proteins (blue) and RNA polymerases (RNAPs; orange) for rRNA production (grey). An RNA polymerase can be translated in half the time (3 min) compared to a set of r-proteins (6 min), and in fast-growing *E. coli* each polymerase can produce 20 sets of rRNA per generation. **b**, Generation time T_{gen} versus the proteomic fraction of *E. coli* r-proteins. Filled circles correspond to previously published experimental data: yellow, ref. 4; green, ref. 8; blue, ref. 10. Because most proteins in *E. coli* are stable, the proteomic fraction of r-proteins is roughly equal to the fraction of time φ that ribosomes spend translating r-proteins. The red solid line is the lower bound on the generation time set by ribosome production: $T_{gen} \geq \tau \ln(2)/\varphi$ (see equation (1) in the limit of large n). For the same ribosome speed, cells with hypothetical ribosomes in which rRNA mass is replaced by r-protein mass (150% or 200% the actual amount, shown as grey dashed and dotted lines, respectively) could not divide as quickly.

however, it does provide a fundamental reason for why proteins must be used sparingly in the ribosome, for example, to increase accuracy or speed up translation, whereas rRNA should be used wherever possible without compromising function. If even one-quarter of the rRNA mass were replaced with r-protein without increasing translation rates, many bacteria would not be able double as quickly as they do (Fig. 4b).

Why are rRNAs so few, large and varying in size?

Considering the principles for r-proteins above, it might seem that rRNA should also be produced in small and uniformly sized pieces to maximize the efficiency of ribosome production. However, in most organisms, nascent rRNA already participates in ribosome assembly by binding r-proteins during their transcription[26,27], eliminating the need to release the rRNA in smaller pieces. In fact, producing many small rRNAs could greatly reduce efficiency because any differences in relative rRNA abundances would not only waste the rRNA that are in surplus, but also the r-proteins that are bound to those rRNA. Because wasting r-protein has a greater (almost two orders of magnitude) effect on ribosome efficiency than does wasting rRNA, the associated loss in efficiency should dominate over any gains from producing rRNA in small pieces. This finding suggests that cells should produce one single, large rRNA. In *E. coli*, the entire rRNA mass is indeed produced as a single transcript, which is then cut at various stages of assembly, creating three rRNA molecules for the two ribosomal subunits. This process ensures that the rRNA molecules are made in strict stoichiometric proportions and minimizes the waste of r-proteins from binding to surplus rRNA. Similar mechanisms occur across a broad range of cell types, including bacteria and mammalian cells, creating rRNA molecules that are generally very large. Such mechanisms also completely relieve the selective pressure for r-proteins to be of similar length (see above). The rRNA varies greatly in length, much more than does genome-wide mRNA and consistent with uniformly distributed random cuts. Although the theory does not predict specifically that the size distribution of rRNA must be broad, the principle of minimizing the time for making new ribosomes is consistent with the rRNA having the opposite size characteristics to r-proteins.

Why do r-proteins differ between ribosomes?

The efficiency perspective might also explain differences between ribosomes. For example, bacterial ribosomes—which are arguably under the most severe selective pressure for fast biogenesis, possibly along with the less well understood archaea—are the smallest and fastest (in terms of translation rates) ribosomes, with the shortest r-proteins and the largest percentage of rRNA by mass (as much as 70%). Mitochondrial ribosomes—which are predicted to be at the opposite extreme and an exception to our theory because they are not present in high abundances and are made by cyto-ribosomes rather than by self-production—have the largest protein mass, the longest r-proteins and as little as 20% rRNA by mass[1,28–31] (Extended Data Fig. 4). Phylogenetic studies suggest that mitochondria originated from bacteria[32,33] and that, over evolutionary time, the rRNA fraction of the mito-ribosomes was gradually replaced by larger r-proteins. Although rRNA may have originated in a hypothetical pre-protein RNA world[25]—the most common explanation for why ribosomes contain so much RNA—it is not necessary to invoke such explanations based on 'frozen accidents' in evolution. In addition to its many specific roles[1], rRNA serves a fundamental purpose in ribosomes by providing appropriate dimensions and an assembly backbone at a ribosome-sequestration cost that is two orders of magnitude lower than would be possible using protein. When this cost is less important, as it is to some extent for higher organisms but particularly for mitochondrial ribosomes, it seems that rRNA is indeed replaced by protein.

Cytosolic ribosomes in eukaryotes are between these extremes. Although the theory cannot predict the amounts of protein and RNA that these ribosomes should contain, because these amounts depend on the pressure for efficiency versus other functionality, it does make a clear prediction about the relative changes in the numbers and sizes of r-proteins. Rather than predicting larger numbers of r-proteins for greater selective pressures, it predicts an optimal number, set by the relative initiation penalty and the total protein mass. This optimal number could potentially be reached by any organism, because translational efficiency is always important. Although initiation complexes can build up slowly for some controlled genes in eukaryotes, elongation is also slower, and observed distances between translating ribosomes can be as low as a few codons[34]. This result strongly suggests that the relative initiation penalty is similar to or lower than that in *E. coli*, and probably equivalent to one or two elongation steps. In this case, equation (2) predicts that any increase in the total protein mass within the ribosome should be achieved by similar changes in the numbers and sizes of the r-proteins, or by a slightly larger increase in the numbers, because the initiation penalty seems to be slightly lower. Eukaryotic ribosomes have a total protein mass that is 1.7–1.8 times higher than that of bacteria, owing to 1.4 times as many r-proteins with an average length of 1.2–1.3 times that of bacterial r-proteins. These similar figures could be coincidental, given the uncertainty in the estimates of initiation penalties, but the theory definitely predicts a much larger number of r-proteins in eukaryotes than in bacteria, given their larger size.

Outlook

Decades of structure–function analyses have shown how individual parts of the ribosome are optimized for translation. Here we expand that approach to consider functionality in a broader cellular context. We show that many features that previously were hard to explain—the unusually small, numerous and similarly sized r-proteins, and the high rRNA complement in a few large molecules—increase overall ribosome efficiency by reducing the time that ribosomes are sequestered for their own production. This increased efficiency could, for example, enable cells to use fewer ribosomes or ribosomes that trade speed for some other functionality. Rather than being relics of an evolutionary past, the unusual features of ribosomes may reflect an additional layer of functional optimization that acts on the collective properties of their parts.

Data Availability A list of genomes that were analysed to support the findings of this study is provided as Supplementary Information. Genomes are available via UniProt (http://www.uniprot.org/proteomes/).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 February; accepted 2 June 2017.

1. Rodnina, M. V., Wintermeyer, W. & Green, R. *Ribosomes Structure, Function, and Dynamics* (Springer, 2011).
2. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38** (Suppl.), D750–D753 (2010).
3. Dill, K. A., Ghosh, K. & Schmit, J. D. Physical limits of cells and proteomes. *Proc. Natl Acad. Sci. USA* **108,** 17876–17882 (2011).
4. Klumpp, S., Scott, M., Pedersen, S. & Hwa, T. Molecular crowding limits translation and cell growth. *Proc. Natl Acad. Sci. USA* **110,** 16754–16759 (2013).
5. Liebermeister, W. *et al.* Visual account of protein investment in cellular functions. *Proc. Natl Acad. Sci. USA* **111,** 8488–8493 (2014).
6. Schaechter, M., Maaløe, O. & Kjeldgaard, N. O. Dependency on medium and temperature of cell size and chemical composition during balanced grown of *Salmonella typhimurium*. *J. Gen. Microbiol.* **19,** 592–606 (1958).
7. Maaløe, O. An analysis of bacterial growth. *Dev. Biol. Suppl.* **3,** 33–58 (1969).
8. Forchhammer, J. & Lindahl, L. Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of *Escherichia coli* 15. *J. Mol. Biol.* **55,** 563–568 (1971).
9. Ehrenberg, M. & Kurland, C. G. Costs of accuracy determined by a maximal growth rate constraint. *Q. Rev. Biophys.* **17,** 45–82 (1984).
10. Bremer, H. & Dennis, P. P. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *Ecosal Plus* **3,** http://doi.org/10.1128/ecosal.5.2.3 (2008).
11. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science* **330,** 1099–1102 (2010).
12. Melnikov, S. *et al.* One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.* **19,** 560–567 (2012).
13. Scott, M., Klumpp, S., Mateescu, E. M. & Hwa, T. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol. Syst. Biol.* **10,** 747 (2014).
14. Maitra, A. & Dill, K. A. Bacterial growth laws reflect the evolutionary importance of energy efficiency. *Proc. Natl Acad. Sci. USA* **112,** 406–411 (2015).
15. Maitra, A. & Dill, K. A. Modeling the overproduction of ribosomes when antibacterial drugs act on cells. *Biophys. J.* **110,** 743–748 (2016).
16. Nomura, M., Gourse, R. & Baughman, G. Regulation of the synthesis of ribosomes and ribosomal components. *Annu. Rev. Biochem.* **53,** 75–117 (1984).
17. Zengel, J. M. & Lindahl, L. Diverse mechanisms for regulating ribosomal protein synthesis in *Escherichia coli*. *Prog. Nucleic Acid Res. Mol. Biol*. **47,** 331–370 (1994).
18. Yamamoto, H. *et al.* 70S-scanning initiation is a novel and frequent initiation mode of ribosomal translation in bacteria. *Proc. Natl Acad. Sci. USA* **113,** E1180–E1189 (2016).
19. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157,** 624–635 (2014).
20. Brandt, F. *et al.* The native 3D organization of bacterial polysomes. *Cell* **136,** 261–271 (2009).
21. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43,** D204–D212 (2015).
22. Sengupta, J., Agrawal, R. K. & Frank, J. Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA. *Proc. Natl Acad. Sci. USA* **98,** 11991–11996 (2001).
23. Qu, X., Lancaster, L., Noller, H. F., Bustamante, C. & Tinoco, I. Jr. Ribosomal protein S1 unwinds double-stranded RNA in multiple steps. *Proc. Natl Acad. Sci. USA* **109,** 14458–14463 (2012).
24. Sauert, M., Temmel, H. & Moll, I. Heterogeneity of the translational machinery: variations on a common theme. *Biochimie* **114,** 39–47 (2015).
25. Cech, T. R. Evolution of biological catalysis: ribozyme to RNP enzyme. *Cold Spring Harb. Symp. Quant. Biol.* **74,** 11–16 (2009).
26. Shajani, Z., Sykes, M. T. & Williamson, J. R. Assembly of bacterial ribosomes. *Annu. Rev. Biochem.* **80,** 501–526 (2011).
27. Woolford, J. L. Jr & Baserga, S. J. Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* **195,** 643–681 (2013).
28. O'Brien, T. W. Evolution of a protein-rich mitochondrial ribosome: implications for human genetic disease. *Gene* **286,** 73–79 (2002).
29. Sharma, M. R. *et al.* Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins. *Cell* **115,** 97–108 (2003).
30. Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343,** 1485–1489 (2014).
31. Greber, B. J. *et al.* Ribosome. The complete structure of the 55S mammalian mitochondrial ribosome. *Science* **348,** 303–308 (2015).
32. Sagan, L. On the origin of mitosing cells. *J. Theor. Biol.* **14,** 225–274 (1967).
33. Andersson, S. G. E. *et al.* The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396,** 133–140 (1998).
34. Myasnikov, A. G. *et al.* The molecular structure of the left-handed supra-molecular helix of eukaryotic polyribosomes. *Nat. Commun.* **5,** 5294 (2014).

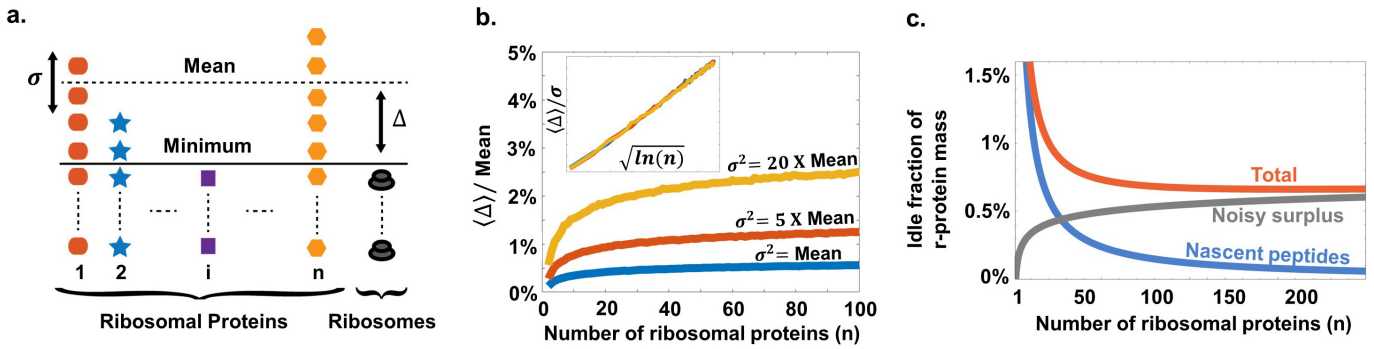
Supplementary Information is available in the online version of the paper.

Acknowledgements S.R. was supported by a James S. McDonnell Foundation fellowship. S.R. and J.P. were supported by NSF-DMS grant PD127334 and NIH grant R01GM095784. J.P. and M.E. were supported by HFSP grant RGP0042 and M.E. was further supported by the Swedish Research Council and the Wallenberg Foundation (RiboCORE). We are grateful to R. Ward, A. Hilfinger, R. Milo, R. Jajoo and M. Landon for discussions.

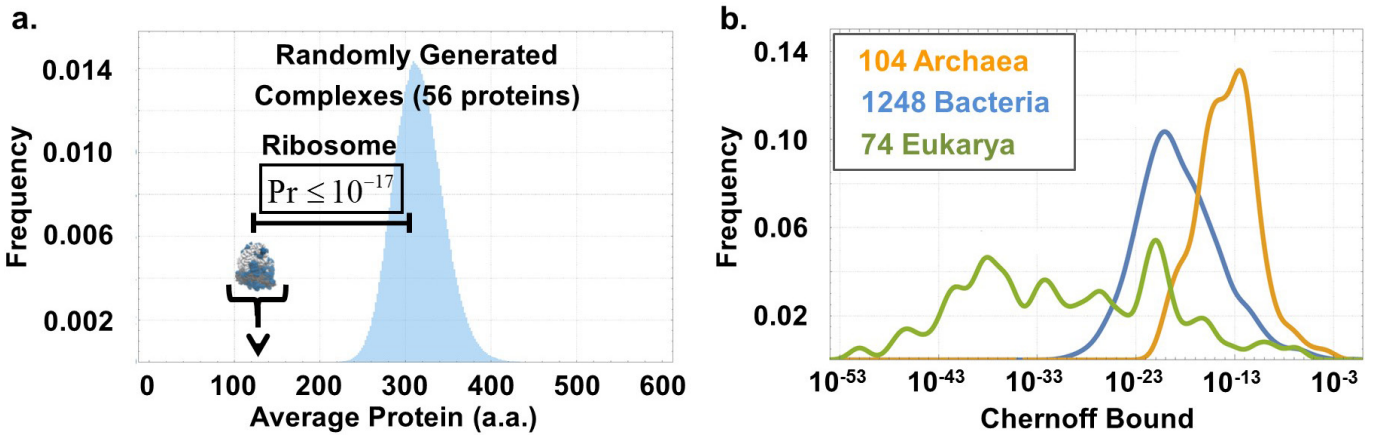
Author Contributions S.R. and J.P. conceived the work, derived results and wrote the paper. M.E. contributed extensive advice and ideas.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.P. (johan_paulsson@hms.harvard.edu).

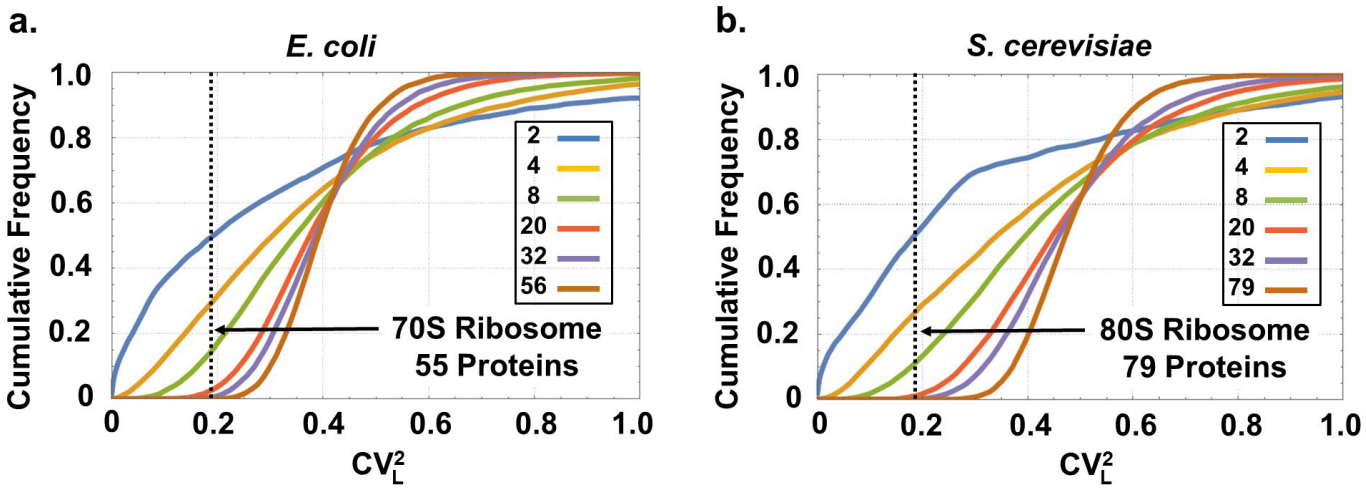
Reviewer Information *Nature* thanks I. Golding, M. Oeffinger, S. Klumpp and the other anonymous reviewer(s) for their contribution to the peer review of this work.



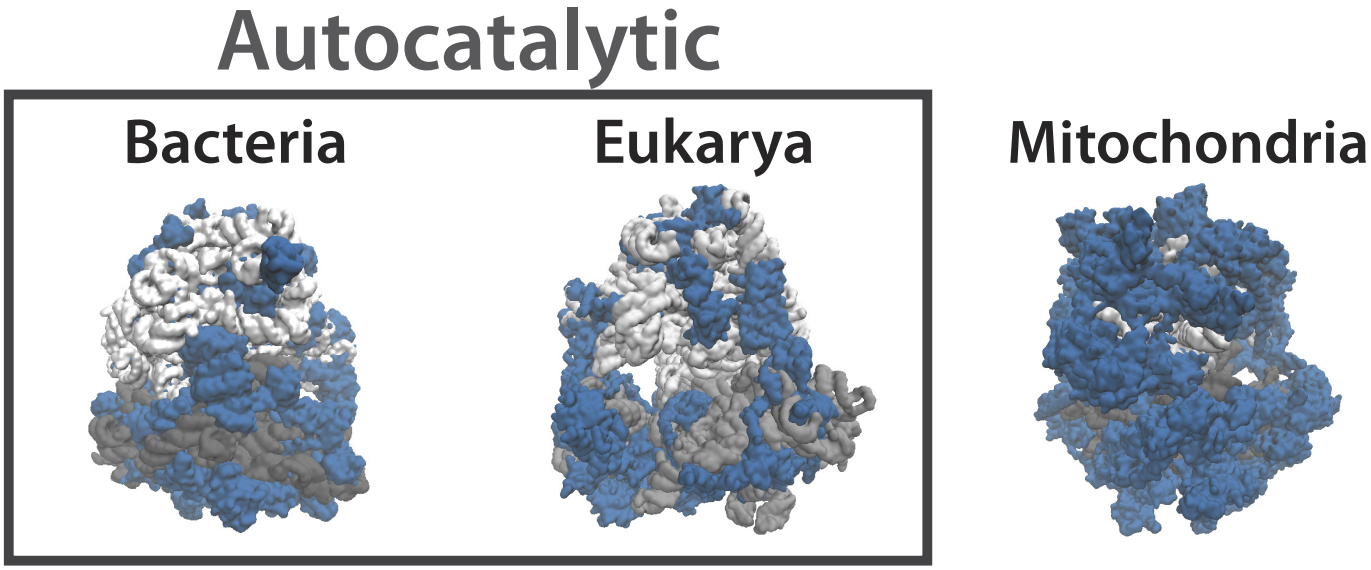
Extended Data Figure 1 | Stochastic production of individual r-proteins may not substantially limit efficiency in ribosome biogenesis. **a**, Schematic of the way in which stochastic gene expression creates temporary shortages of some r-proteins and surpluses of others. The number of complete ribosomes assembled is then limited by the r-protein that is present in lowest abundance, and the average value of the difference between the minimum and the mean $\langle \Delta \rangle$ is the mean number of unmatched r-proteins. **b**, The relative mean free r-protein pool (from **a**) increases very slowly with the number of ribosomal proteins n. For illustrative purposes we used negative binomial distributions with mean μ and different variances σ^2 (Supplementary Information), because this distribution has been observed and predicted in many studies of stochastic gene expression. The inset shows a curve collapse for $\langle \Delta \rangle / \sigma \propto \sqrt{\ln(n)}$. We obtained similar results for more complete kinetic models. **c**, The fraction of the r-protein mass in the form of nascent peptides or free pools arising from noisy expression (assuming Poisson noise for simplicity, but with similar results for many other noise models). There is technically an optimal number of r-proteins n_{opt}, but the curves around and above this value are very shallow, meaning that there is practically no upper limit.



Extended Data Figure 2 | The r-proteins are statistical outliers in terms of size compared to the rest of the genome. a, Distribution of the average protein length in 10^6 random samples of 56 proteins taken from the genome of *E. coli*; the mean length of a protein in the bacterial ribosome is also marked (inset ribosome with arrow). The probability of generating an average protein length as small as that seen in the bacterial ribosome is vanishingly small (at most 10^{-17}). **b**, The Chernoff upper bound (Supplementary Information) on the probability that the average length of r-proteins could arise at random, that is, with no size selection, is computed for: 104 archaea (orange), 1,248 bacteria (blue) and 74 eukarya (green). The value for *E. coli* is 10^{-17} (see **a**); vanishingly small probabilities are attributed to all organisms that we examined.



Extended Data Figure 3 | The r-proteins are unusually similar to each other in size even when conditioning on the average size. Cumulative frequency of CV_L^2 for randomly generated protein complexes of varying numbers of proteins is shown. See Supplementary Information for details. **a**, **b**, Only complexes in which the average protein length is identical (±5 amino acids) to that seen in the ribosomes of *E. coli* (**a**) and *S. cerevisiae* (**b**) entered the statistics. As the number of proteins in a complex gradually increases from 2 to the number of proteins in the ribosome—55 in *E. coli* (protein S1 was excluded; see main text) and 79 in *S. cerevisiae*—the occurrence of coefficients of variation (CV_L) as low as those seen for the set of r-proteins becomes extremely rare.



Extended Data Figure 4 | Differences between different types of ribosomes are qualitatively as expected from considering efficiency in their biogenesis. Bacterial ribosomes contain about 55 r-proteins (blue) that average about 130 amino acids and make up 30%–35% of the ribosome mass. Eukaryotic ribosomes contain about 80 r-proteins (blue) that average about 165 amino acids and make up about 45% of the ribosome mass. In both these cases, ribosomes catalyse their own production. By contrast, in mitochondrial ribosomes, which do not produce themselves, rRNA (grey) is more scarce and r-proteins are much larger, making up as much as 80% of the ribosome mass.