Molecular Biology and Evolution, 2026, 43, 1–13
https://doi.org/10.1093/molbev/msag062
Published: 8 March 2026
Discoveries

MBE OXFORD

Stoichiometry-induced differential selection on codon optimization among ribosomal protein genes in bacterial species

Xuhua Xia [1,2,*]

[1]Department of Biology, University of Ottawa, Marie-Curie Private, Ottawa, ON K1N 9A7, Canada
[2]Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON K1H 8M5, Canada
***Corresponding author:** E-mail: xxia@uottawa.ca.
Associate editor: Laasya Samhita

Abstract

While it is well established that highly expressed genes in bacteria exhibit stronger codon optimization than lowly expressed ones, whether codon–anticodon adaptation shows fine-scale differentiation among highly expressed genes themselves remains unexplored. Ribosomal proteins, which are expressed in stoichiometric amounts and often cotranscribed in polycistronic operons, provide an ideal system for testing such differential selection. Here, I demonstrate that in *Escherichia coli*, *Bacillus subtilis*, and *Vibrio natriegens*, codon usage is more optimized in long ribosomal protein genes compared to short ones. This pattern persists even among genes within individual operons, such as *S10*, *spc*, and *α* operons. A ribosome in *E. coli* or *B. subtilis* needs four copies of L7/L12 (encoded by *rplL*) but only one copy each of the other ribosomal proteins. This high demand for L7/L12 leads to my prediction that the *rplL* gene should be translated more actively than its operonic partner, *rplJ*, encoding L10. This prediction is also strongly supported by empirical evidence from representative bacterial species. Actively translated mRNAs are protected from endonucleolytic cleavage and degradation. If *rplL* mRNA is more actively translated than *rplJ* mRNA, then *rplL* mRNA would be degraded less and become more abundant than *rplJ* mRNA, which is true. These results demonstrate that translation optimization reflects functional stoichiometry and protein length constraints. This is the first demonstration of natural selection operating predictably and precisely among ribosomal protein genes in the same operon, fine-tuning translational output to achieve efficient ribosomal assembly.

Keywords ribosomal proteins, differential selection, codon optimization, stoichiometry and selection, translation initiation and elongation

Introduction

The study of codon–anticodon adaptation features several landmarks that have contributed significantly to translation elongation and its relationship to translation initiation. The pioneering studies by Ikemura (1981, 1992) identified differential tRNA availability as a key factor driving optimal codon usage in unicellular bacteria and yeasts. Gouy and Gautier (1982) and subsequent papers demonstrated stronger codon usage bias in highly expressed genes (HEGs) than lowly expressed genes (LEGs), presumably because HEGs have experienced stronger codon-optimizing selection than LEGs. Robinson et al. (1984), Sorensen et al. (1989), and Curran and Yarus (1989) performed early experiments that altered codon usage, measured the resulting change in translation efficiency, and concluded that optimal codon usage increases protein production in *Escherichia coli*. This was consistent with subsequent experiments on HIV-1 (Human Immunodeficiency Virus type 1) protein-coding genes (Haas et al. 1996; Ngumbela et al. 2008). Thus, the conclusion that optimizing selection drives codon adaptation, especially in HEGs, in rapidly replicating organisms is well established.

However, the conclusion above was challenged by an experiment (Kudla et al. 2009) that altered synonymous codon usage of the same *E. coli* green fluorescent protein (GFP) but revealed little association between protein production and codon adaptation. This surprising finding was subsequently found to be confounded by two factors that were not considered by Kudla et al. (2009). First, optimizing codon usage has little effect on protein production when translation initiation efficiency is poor, but a strong effect on protein production with efficient translation initiation (Tuller et al. 2010; Xia 2015). Second, Kudla et al. (2009) used codon adaptation index (CAI) (Sharp and Li 1987; Xia 2007) to quantify codon adaptation. However, CAI does not take background mutation bias into consideration and can lead to a biased measure of codon adaptation (Xia 2015). When

Received: October 16, 2025. **Revised:** February 24, 2026. **Accepted:** February 26, 2026

these two factors were incorporated into a reanalysis of the experimental data by Kudla et al. (2009) , and when an index of translation elongation (I_{TE}), which is a generalized version of CAI taking into consideration of background mutation bias and numerically illustrated in detail in Xia (2018a), was used to quantify codon adaptation, nearly 20% of the variation in protein production can be attributed to codon usage differences (Xia 2015). Such studies led not only to theoretical refinement of both deterministic (Xia 1998) and stochastic models (Bulmer 1991) but also to practical applications. For example, the COVID-19 (Coronavirus disease 2019) vaccine mRNA by Pfizer and Moderna was extensively codon optimized (Xia 2021).

The translation rate is about 15 codons per second in highly expressed *E. coli* genes during the exponential phase, which implies about 65 ms per codon (Dai et al. 2016; Irshad and Sharma 2023), although the average rate for all genes could be slower (Guet et al. 2008). Early experimental studies on *E. coli* at 37 °C showed a range from 12 amino acids per second at slow growth to 21 amino acids per second at rapid growth (Dennis and Bremer 1974; Dennis and Nomura 1974; Young and Bremer 1976). This rate is close to the transcription rate of about 85 nucleotides per second measured with the 5S rRNA in *E. coli* (Shen and Bremer 1977; Ryals et al. 1982). Because rRNA and mRNA are translated with the same RNA polymerase, one may extrapolate the transcription rate from 5S rRNA to mRNA. It seems that codon-optimized genes should also have optimized transcription to avoid the collision between ribosomes and RNA polymerase unless there is an anticollision mechanism between the two.

A minor codon in *E. coli* often takes 1.5 times as long to decode as the major codon (Gardin et al. 2014). A major codon is a translationally optimal codon, operationally characterized with two features: (i) overused by HEGs and (ii) decoded by the most abundant tRNA (Xia 2018a). For mass-produced proteins that could reach half a million copies per generation per cell, a descendant cell with many more major codons than its peers should conceivably be favored by natural selection.

How precisely can the theory of codon–anticodon adaptation predict empirical patterns of codon adaptation? It is now trivial to compare HEGs and LEGs and report differences in codon adaptation between the two groups. However, it would be phenomenal if the theory could make fine-scale predictions, such as differential selection that operates on highly expressed ribosomal protein (RP) genes and generates differential patterns of codon optimization among them.

E. coli is probably the best species for testing the fine-scale predictions of the codon–anticodon adaptation. It has 55 RPs encoded by 54 protein-coding genes (Wittmann 1982; Wittmann-Liebold 1986; Korobeinikova et al. 2012), with L7 and L12 encoded by the same gene *rplL* sharing the same primary sequence (Wittmann 1976). L7 is the acetylated form of L12. The two are often referred to as the L7/L12 protein. RP L8 is not from a single gene but is a complex of L7/L12 +L10. The proteins in the small ribosomal subunit are encoded by 21 unique genes (*rpsA* to *rpsU*, encoding S1 to S21, respectively), and those in the large ribosomal subunit are encoded by 33 unique genes (*rplA* to *rplF*, *rplI* to *rplY*, and *rpmA* to *rpmJ*, encoding L1 to L6, L9 to L25, and L27 to L36, respectively). RP L26 was found to be identical to S20 (Wittmann 1976, 1982). Subsequent structural determination revealed that S20 exists only in the small ribosomal subunit (Wimberly et al. 2000; Harms et al. 2001; Schuwirth et al. 2005).

Two features of *E. coli* RPs are directly related to differential selection. First, with only one exception (ie L7/L12, which is the only multicopy protein in bacterial ribosomes), all RPs exist in single copies in *E. coli* ribosomes (Davydov et al. 2013). It has long been known that RPs, except for L7/L12, are produced in equimolar amounts (Lindahl and Zengel 1979). Second, the L7/L12 protein exists in four copies (in two dimers) in the large ribosomal subunit in *E. coli* (Diaconu et al. 2005; Ilag et al. 2005; Gordiyenko et al. 2010; Davydov et al. 2013), and four or six copies in other bacterial species (Diaconu et al. 2005; Ilag et al. 2005; Mandava et al. 2012; Davydov et al. 2013), although there is no strong evidence that L7 and L12 exist in the exact 1:1 ratio (Brot and Weissbach 1981; Asato 2005; Vila-Sanjurjo 2008; Davydov et al. 2013). *E. coli* mutants with only one dimer of L7/L12 (instead of two dimers) in the large subunit ribosome suffer from much reduced growth and replication (Mandava et al. 2012).

The two features of *E. coli* RPs immediately lead to two inferences and associated testable predictions. The first involves the possible mechanisms by which long and short RPs are produced in equimolar amounts. The longest RP (S1) has 557 aa residues in contrast to the shortest L36 with only 38 aa residues. Both empirical data (Lu et al. 2007) and theoretical models (Valleriani et al. 2011) show that protein production decreases with sequence lengths in bacterial species, everything else being equal. This length-dependent effect is attributed to two factors (Valleriani et al. 2011). First, the RNA degradation in bacterial species is mainly carried out by endonucleases such as RNase E and RNase III (or their equivalents). Ribosomes on long mRNA have a higher chance of experiencing an endonucleolytic cleavage and failing to complete the translation. Second, ribosomes traveling along a long mRNA are more likely to experience traffic jams or collisions than short ones. This suggests that long mRNAs may have difficulties in producing the same number of proteins as short ones. There is indeed experimental evidence that S1, the longest RP, is rate limiting. Two *E. coli* mutants, MA261 and KK101, exhibit much reduced growth rates, with doubling times of 260 and 496 min, respectively. Addition of putrescine (a polyamine) at 100 μg/mL to the culture stimulated the synthesis of RP S1 markedly but had little effect on other RPs (Kashiwagi et al. 1989). The doubling time is reduced from 260 to 128 min for the MA261 strain and from 496 to 135 min for the KK101 strain. This result is consistent with the hypothesis that the concentration of RP S1 is rate limiting in *E. coli*, ie increasing S1 concentration in the slow-growing mutants increases the replication rate. However, there is no study that increases the S1 level above the wild type to see if the replication rate is increased.

In order to achieve the 1:1 stoichiometric relationship (ie a ribosome needs one copy of each RP except for L7/L12), long ribosomal mRNAs may be selected to take one or more of the following paths. First, the mRNA of long RPs, such as S1, could have more efficient translation initiation than that of short RPs, such as L36. Second, long RPs could have more abundant mRNAs (ie higher transcription and lower degradation) than short RPs. For example, RP genes of different lengths could be placed in different operons, with genes of long coding sequences (CDSs) having stronger promoters (and consequently more mRNAs) than genes of short CDSs. The longest S1 is encoded by its own operons, whereas many other RPs share the same operon; eg the S10

operon contains 11 RP genes in the order of *rpsJ*, *rplC*, *rplD*, *rplW*, *rplB*, *rpsS*, *rplV*, *rpsC*, *rplP*, *rpmC*, and *rpsQ* (Zengel et al. 1980; Lindahl et al. 1983; Zurawski and Zurawski 1985). Is the S1 mRNA more abundant than the mRNA of other RP genes? Third, the mRNA of long RPs may have more optimized codon usage and can be translated more efficiently than short ones. This could reduce the chance of an mRNA experiencing endonucleolytic cleavage before a ribosome reaches the stop codon. Also, optimal codon usage essentially allows all ribosomes to travel along the mRNA at roughly the same high speeds, but suboptimal codon usage would be equivalent to allowing ribosomes to travel at a mixture of high and low speeds. The latter is more likely to cause traffic jams than the former.

The second inference is more straightforward. L7/L12 is the only multicopy protein in the *E. coli* ribosome, so its encoding gene should be under strong selection pressure to meet the 4:1 stoichiometric requirement in ribosome assembly. One would therefore predict that the encoding gene, *rplL*, should exhibit more optimized codon usage than all other RPs. In particular, the multicopy feature of L7/L12 in bacterial ribosomes appears universal (Diaconu et al. 2005; Ilag et al. 2005; Mandava et al. 2012; Davydov et al. 2013), which implies that the selection for increased production of the L7/L12 should have operated for hundreds of millions of years. Such long-term directional selection most likely has left footprints in differential codon optimization between *rplL* (encoding L7/L12) and other RP genes. Confirming this prediction not only has scientific merit but also industrial significance, ie the L7/L12 abundance may be rate limiting, so engineering an additional *rplL* gene into *E. coli* should enhance the growth and replication rate.

Results

Genes encoding long RPs exhibit more optimized codon usage than short ones

I outlined in the introduction three paths for a long CDS to produce an equimolar amount of protein as a short CDS and argued that improved translation elongation rate through codon optimization is the most likely of the three. I used the index of translation elongation (I_{TE}) (Xia 2015, 2017b, 2018a) to measure codon optimization. I_{TE} is advantageous over CAI in that it takes background mutation into consideration, as illustrated in the Materials and methods section.

Consistent with the prediction, I_{TE} increases asymptotically with the CDS length of RPs (Fig. 1) as predicted. Gram-negative bacteria such as *E. coli* (Fig. 1a) and *Vibrio natriegens* (Fig. 1c) share a RP S1 much longer than all the rest, which might bias the model fitting. For this reason, I also repeated the analysis for the subset of genes encoding the large RPs. The results (Fig. 1b and d) are also consistent with the prediction. The model species for the Gram-positive bacteria, *Bacillus subtilis*, does not have the large S1 gene, but the analysis was also done for all RPs (Fig. 1e) and the subset of genes encoding the large subunit RPs (Fig. 1f) for the sake of completeness.

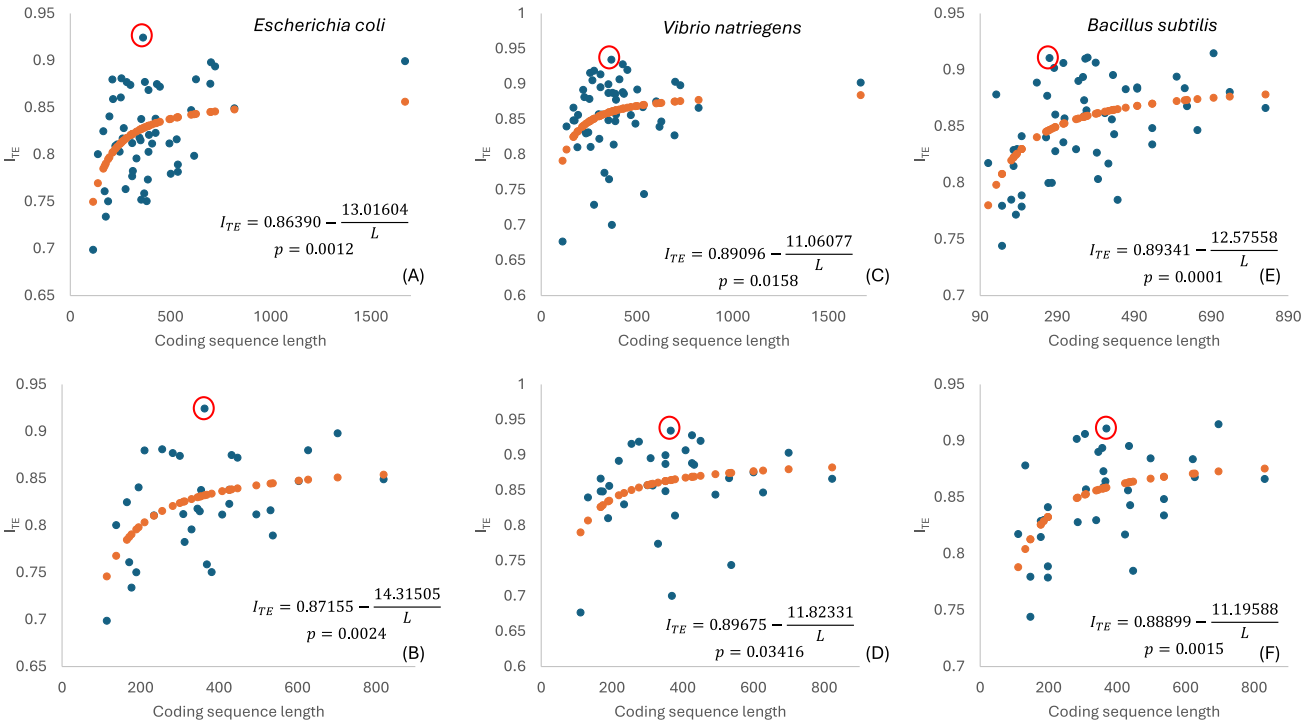

Figure 1 I_{TE} (a measure of codon optimization) increases asymptotically with RP gene length. The top three figures (a, c, and e) include all RP genes. The bottom three figures (b, d, and f) include only genes encoding the large subunit RPs. The two figures on the left (a and b), middle (c and d), and right (e and f) are for *E. coli*, *V. natriegens*, and *B. subtilis*, respectively. The *P*-values are from two-tailed tests. The circled point denotes *rplL* encoding the only multicopy L7/L12 protein in bacterial ribosomes. The equations are the fitted models between I_{TE} and L (CDS length). The *P*-values are from LRTs, with the null model being $y = \alpha$ and the general model specified in Equation (1).

Numerically, the increase of I_{TE} with the CDS length of RPs should be asymptotic instead of linear because I_{TE}, like CAI, has a maximum value of 1. One simple mathematical function for such an asymptotic relationship is

$$I_{TE} = \alpha - \frac{\beta}{L} \quad (1)$$

where α and β are parameters to be estimated from the empirical data of observed I_{TE} and sequence length (L). The fitted relationship is shown for individual species in Fig. 1.

What is remarkable is that the positive association between I_{TE} and the CDS length is maintained even for RP genes within individual operons. In the *E. coli* K12 reference genome (NC_000913), the *S10* operon contains 11 RP genes (*rpsJ*, *rplC*, *rplD*, *rplW*, *rplB*, *rpsS*, *rplV*, *rpsC*, *rplP*, *rpmC*, *rpsQ*) (Zengel et al. 1980; Lindahl et al. 1983; Zurawski and Zurawski 1985), the *spc* operon contains 10 (*rplN*, *rplX*, *rplE*, *rpsN*, *rpsH*, *rplF*, *rplR*, *rpsE*, *rpmD*, *rplO*) (Cerretti et al. 1983), and the α operon contains 5 (*rpmJ*, *rpsM*, *rpsK*, *rpsD*, *rplQ*) (Cerretti et al. 1983; Meek and Hayward 1984). All three operons are packed next to each other along the same strand in the *E. coli* reference genome (NC_000913). One may also fit the null model (M0: $y = \alpha$) and the alternative model (M1) in Equation (1) and then use the likelihood ratio test (LRT) to obtain statistical significance. However, because LRT assumes a large sample size, and because the sample size in Table 1 is apparently small, the resulting *P*-value would need to be corrected for small sample size by parametric bootstrap likelihood ratio test (pbLRT) (Efron 1979, 1982), as shown in Table 1. I illustrate the rationale of pbLRT with the S10 operon in Table 1, with sample size n = 11. We first fit models M0 ($y = \alpha$) and M1 in Equation (1) to obtain the log-likelihood (lnL) for the two models, designated $\ln L_{M0}$ and $\ln L_{M1}$, respectively, and calculate the likelihood ratio chi-square, designated as $G^2 = 2\times(\ln L_{M1} - \ln L_{M0}) = 5.8744$ (Table 1). We then generate 20,000 samples based on model M0, estimate parameters α and β in Equation (1), calculate G_i^2 for each sample *i*, and count the number of G_i^2 values greater than or equal to G^2. This number is 700 out of 20,000. The *P*-value from pbLRT for the S10 operon genes is then 700/20,000 = 0.035 (Table 1). The lack of a significant correlation for the *spc* operon genes (Table 1) could be attributed to the small variation in sequence length, with the minimum and maximum sequence length being 177 and 537, respectively. In contrast, the corresponding values are 114 and 618, respectively, for the α operon and 189 and 819, respectively, for the *S10* operon.

The point inside a red circle in Fig. 1 indicates the *rplL* gene encoding L7/L12, the only multicopy protein in bacterial ribosomes. Because each ribosome in *E. coli* needs multiple copies (four copies in *E. coli* and *B. subtilis*) of L7/L12 but only one copy of other ribosome proteins, I predicted that the *rplL* gene should be translated more efficiently. Specifically, it should have a higher I_{TE} value than other RP genes. This prediction is consistent with the results in Fig. 1, as the *rplL* gene in all three bacterial species shows the highest I_{TE} among RPs.

Table 1 pbLRTs of the relationship between CDS length and I_{TE} for three operons (S10, α, and spc).

pbLRT[a]	S10	α operon	spc
n	11	5	10
$\ln L_{M0}$	18.3193	9.9619	18.9035
$\ln L_{M1}$	21.2564	14.5091	20.5513
G^2	5.8744	9.0945	3.2955
P	0.0154	0.0026	0.0695
p′	0.0350	0.0279	0.1173

[a]Information needed for pbLRT.
n, sample size; $\ln L_{M0}$ and $\ln L_{M1}$, the log-likelihood for the null model and the alternative model; $G^2 = 2\times(\ln L_{M1} - \ln L_{M0})$, likelihood ratio chi-square; p, uncorrected significance value from a conventional LRT; p′, *P*-value from pbLRT.

Does S1 in *E. coli* exhibit increased translation initiation efficiency or have more mRNA than others?

RP S1 is the longest among RPs, with 557 residues. The next longest is L2 (encoded by *rplB*), which has only 273 residues. Given that translation efficiency can decrease with increasing sequence length, substantiated by both empirical evidence (Lu et al. 2007) and theoretical models (Valleriani et al. 2011), it is not surprising that longer mRNAs of RPs exhibit more optimized codon usage (Fig. 1). One may ask if S1 mRNA exhibits increased translation initiation efficiency, which is often characterized by (i) the secondary structure stability near the translation initiation region (TIR; flanking and including the Shine–Dalgarno (SD) sequence and the start codon) and (ii) the presence and position of SD and the pairing strength between SD and anti-SD (aSD) sequences.

Previous studies have shown a weakening of secondary structure in bacterial and phage mRNA near TIR, especially in HEGs (Prabhakaran et al. 2015; Xia 2018b, 2023). The minimum folding energy (MFE) in this region has been used as a proxy for translation initiation efficiency. However, there are differences in specifying the boundary of TIR among different studies. Kudla et al. (2009) used a window from site −4 to +37 as a proxy of translation initiation efficiency, whereas others (Prabhakaran et al. 2015; Xia 2023) used a sliding window and found that the most changes in MFE occur 20 nt upstream and downstream of the start codon. I calculated MFE for three windows, 30 + 10, 20 + 20, and 10 + 30, where the first number is the number of nucleotides upstream of the start codon, and the second number is the number of nucleotides in the CDS, including the start codon (Table 2).

The TIR region in the S1 mRNA does exhibit reduced secondary structure stability (MFE = 0 means no formation of secondary structure, and increasingly negative MFE means increasingly stronger secondary structure), suggestive of an increased translation initiation efficiency relative to an average gene. How likely do these MFE values for S1 belong to the empirical distribution of MFE values of other RP genes? To address the question, one first needs to fit a distribution to the empirical MFE values of other RP genes. One cannot use a normal distribution because the maximum MFE value is 0. One may take $x = |\text{MFE}|$ so that all *x* values are non-negative and then fit a gamma distribution to *x*. However, the probability density of a gamma distribution is zero when $x = 0$, but the MFE value of an RNA sequence can realistically be zero (ie no secondary structure formation). This implies that we need to fit a zero-inflated continuous distribution specified in two parts: (i) a probability mass π for $x = 0$ and (ii) a gamma distribution for $x > 0$. The

Table 2 Mean and standard deviation (Std) of MFE of the TIR in *E. coli* RP genes and other genes (non-RP).

TIR[a]	S1 (*rpsA*)	Mean		Std				
		RP[b]	Non-RP[c]	RP[b]	Non-RP[c]	α_{RP}	β_{RP}	π_{RP}
30 + 10	0	−4.51926	−4.58774	2.72131	2.79374	4.08103	1.19597	0.05556
20 + 20	−2.5	−4.36333	−4.17273	2.43549	2.57906	2.45565	1.82562	0.01852
10 + 30	−1.4	−4.88352	−4.52485	2.74738	2.73662	2.96331	1.70229	0.01852

A mixture distribution is fitted to $x = |\text{MFE}|$ for RP genes, with parameters α and β for the gamma distribution when $x > 0$, and π as the probability mass for $x = 0$.
[a]TIR, translation initiation region in the form of (the number of nucleotides upstream of the start codon) + (the number of nucleotides in the 5′ end of the CDS).
[b]RP, ribosomal proteins. [c]Non-RP, nonribosomal proteins.

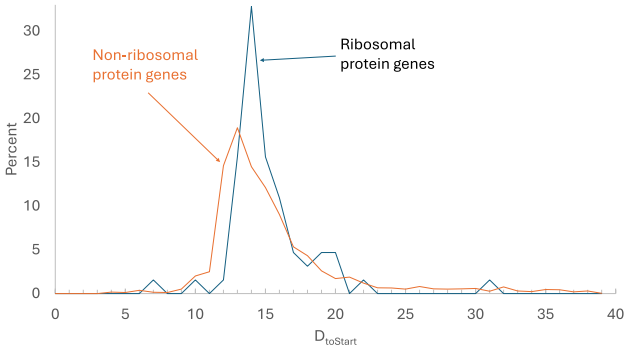

Figure 2 Frequency distribution of $D_{toStart}$ (the number of nucleotides between the 3′ end of the 16S rRNA to the start codon, given the constraint imposed by the base pairing between SD and aSD) in *E. coli* genes. The $D_{toStart}$ values for ribosomal protein genes are constrained within a narrower range than those for the rest of *E. coli* genes.

log-likelihood function (lnL) for estimating the α, β, and π parameters is

$$\boldsymbol{lnL} = \boldsymbol{k}\ln\boldsymbol{\pi} + (\boldsymbol{n}-\boldsymbol{k})\ln(1-\boldsymbol{\pi}) + \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}-\boldsymbol{k}} \ln[\boldsymbol{g}(\boldsymbol{x_i}|\boldsymbol{\alpha},\boldsymbol{\beta})] \qquad (2)$$

where k is the number of observations of $x = 0$ out of a total of n observations and $g(x_i|\alpha, \beta)$ is the probability density function of the gamma distribution. The estimated parameters for the distribution of MFE values of the 53 RP genes (excluding S1) are shown in Table 2. The probabilities that the three MFE values for the S1 mRNA, ie 0, −2.5, and −1.4 in Table 2, belong to the empirical distribution of the MFE values for the RP genes are 0.05556, 0.27054, and 0.05358, respectively, given the fitted distribution.

There is no significant difference in MFE between RP genes and non-RP genes for *E. coli* in Table 2. The three *t*-tests corresponding to the three TIR positions yield $p_{30+10} = 0.8579$, $p_{20+20} = 0.5892$, and $p_{10+30} = 0.3386$. The corresponding *P*-values for *B. subtilis* are 0.9900, 0.9766, and 0.5894, respectively.

The presence of a well-positioned SD with optimal base pairs between the SD and aSD has previously been considered as a good indicator of efficient translation initiation (Prabhakaran et al. 2015; Xia 2023). The *base pairing* between SD and aSD defines a distance from the 3′ end of 16S rRNA, known as $D_{toStart}$ (Prabhakaran et al. 2015; Xia 2023), which is strongly constrained, especially for RP genes (Fig. 2). The optimal $D_{toStart}$ is likely 14 bases, assuming that RP genes are under strong stabilizing selection to reach optimal $D_{toStart}$. Paradoxically, S1 mRNA does not have an SD, defined as ≥4 contiguous bases complementary to aSD in the 16S rRNA. In fact, S1 mRNA in *E. coli* represents a prime example of bacterial mRNA with an SD-independent TIR that enables extremely efficient translation (Skorski et al. 2006). The S1 protein enhances not only its own translation but also the SD-independent translation of many other genes (Aliprandi et al. 2008; Duval et al. 2013; Lund et al. 2020), which explains why S1 abundance is a limiting factor of replication rate and that increasing S1 production increases replication rate in *E. coli* mutants (Kashiwagi et al. 1989).

Table 3 Mean and standard deviation (Std) of gene expression (in RPKM, reads per kilobase per million mapped reads) of the TIR.

Strain[a]	S1 (*rpsA*)	Mean		Std	
		RP[b]	Non-RP[c]	RP[b]	Non-RP[c]
Wild-type	3784.43	4795.75	158.78	2595.11	1601.16
rnb mutant	3701.35	4692.70	159.78	2589.89	1631.78
pnr mutant	3059.10	3851.52	164.42	2098.88	1349.85
pnp mutant	2740.67	4290.89	171.78	2057.09	1593.97

[a]Strain, wild-type and three mutant strains defective in the ribonucleases RNase II, RNase R, and PNPase, respectively. [b]RP, ribosomal proteins.
[c]Non-RP, nonribosomal proteins.

One may also ask if the S1 protein might have a higher mRNA concentration than other RPs. However, while RPs are much more expressed than an average gene in *E. coli* (Table 3, $P < 0.0001$, two-tailed *t*-test), the S1 protein does not have higher mRNA abundance than other RPs (Table 3). Because highly expressed long mRNA might sequester too many ribosomes to impact global translation, it seems reasonable for such mRNA to have an increased translation elongation rate (Fig. 1) instead of having many mRNA molecules.

In short, S1 mRNA has long known to be efficiently translated (Skorski et al. 2006; Aliprandi et al. 2008; Duval et al. 2013; Lund et al. 2020) and has a weak secondary structure in its TIR (Table 2), consistent with the previous observation that weak secondary structure in the TIR is associated with efficient translation initiation in both bacteria and bacteriophage(Kudla et al. 2009; Prabhakaran et al. 2015; Xia 2023). This is also consistent

with the interpretation that SD and various other mechanisms that enhance translation initiation are mostly for preventing the start codon from being embedded in secondary structure (Nakamoto 2006). However, it is puzzling that RP genes do not have a mean MFE closer to zero than other protein genes.

Ribosomal gene *rplL* encoding L7/L12 exhibits more optimized codon usage than *rplJ* encoding L10 in the same operon

One might argue that the high codon optimization in *rplL*, shown by the high I_{TE} values (Fig. 1), may not be caused by its multicopy requirement in each ribosome. Instead, *rplL* might happen to reside on a genomic segment where the mutation bias and tRNA-mediated selection happen to act in the same direction, whereas other RP genes might be in genomic regions where mutation and the selection act in opposite directions. For this reason, a fair comparison should be performed between *rplL* and its neighboring genes, or ideally against a RP gene on the same operon.

Given the reasoning above, one can predict that I_{TE} should be higher for *rplL* than for *rplJ* in *E. coli*, given the 4:1 stoichiometric relationship. This prediction is consistent with empirical evidence (Table 4). The difference between the two genes in I_{TE} is highly significant based on a paired-sample *t*-test ($t = 4.4527$, $DF = 5$, $P = 0.0067$, two-tailed test). The effect size, shown in the last column in Table 4, is more biologically meaningful than the *P*-value, which could be made infinitely small by just adding more bacterial species, because the difference is consistent in more than 30 randomly chosen bacterial species I examined and is likely universal. The effect size is relatively small, especially for the three Gram-positive species (the last three in Table 4). Note that I_{TE} varies from about 0.7 to about 0.9 among different RP genes that are produced in equimolar amounts. Therefore, the small ΔI_{TE} appears insufficient to guarantee a 4:1 stoichiometric ratio between the two proteins.

The three Gram-negative species (the first three in Table 4) have greater ΔI_{TE} than the three Gram-positive species. This might suggest phylogenetic inertia, so a phylogeny-based inference (Felsenstein 1985; Harvey and Pagel 1991) would seem appropriate. However, all six species are highly diverged from each other with little homology at their synonymous sites in aligned sequences (Fig. 3). Indeed, it would be silly to argue that the similarity in ΔI_{TE} between *B. subtilis* and the two other Gram-positive species is due to phylogenetic similarity between them. However, one could deduce from Table 4 and Fig. 3 that *rplL* I_{TE} is consistently greater than *rplJ* I_{TE} in both Gram-negative and Gram-positive bacteria. Within Gram-positive bacteria, the difference is also consistent in species differing dramatically in GC content (percentage of nucleotides G and C).

Table 4 The *rplL* gene encoding L7/L12 is more codon optimized, as measured by I_{TE}, than the *rplJ* gene encoding L10.

Species	Genome	*rplJ* I_{TE}	*rplL* I_{TE}	ΔI_{TE}
E. coli	NC_000913	0.81175	0.92437	0.11262
V. natriegens	NZ_CP009977	0.84386	0.93439	0.09053
H. influenzae	NZ_CP085952	0.8412	0.93634	0.09514
B. subtilis	NC_000964	0.88445	0.91091	0.02646
M. smegmatis	NZ_CP080274	0.82087	0.86506	0.04419
M. abscessus	GCF_000069185	0.81307	0.84564	0.03257

The last column shows the difference in I_{TE}. Both genes are in the same operon.

Multiple experiments (Pettersson and Kurland 1980; Liljas and Sanyal 2018) have demonstrated a much reduced translation efficiency when L7/L12 is reduced below the normal level in *E. coli*. However, increasing the production of L7/L12 proteins above the normal level does not stimulate faster growth in *E. coli* (Hofmann et al. 2025). Thus, the level of L7/L12 appears to be at the optimal level in *E. coli*, given the normal cellular expression of other genes. Ribosomes are made of multiple ribosomal RNAs and proteins. Changing the abundance of one component may not only fail to increase growth and reproduction but also have a negative effect.

The *rplL* mRNA is more abundant than *rplJ*

The *rplJ* and *rplJ* are in the same operon, together with two sigma factors B and C, forming the *rplJL-rpoBC* operon. This L10-L7/L12 operon arrangement is widely represented in bacterial lineages, which makes sense because they are the component proteins forming the stalk of the large ribosomal subunit and should be on the same operon and coregulated in expression (Liljas and Sanyal 2018). However, cistrons in a polycistronic mRNA are typically translated independently (Huber et al. 2019). Occasionally, the same ribosome, after translating the upstream cistron, can reinitiate and translate the downstream cistron (Levin-Karp et al. 2013; Saito et al. 2020; Brown and Wade 2025). One is prone to assuming the same mRNA abundance for genes in the same operon. However, an mRNA molecule transcribed from a polycistronic operon in bacteria is often cleaved at intercistronic regions by one of two endonucleases, RNase E and RNase III, generating independently translated mRNAs (Mackie 2013; Jeon et al. 2020; Jeon et al. 2025), especially when an AU-rich element (i.e., an element rich in nucleotides A and U) exists in the intercistronic region (Clarke et al. 2014). For example, the *rplJL-rpoBC* operon is cleaved by RNase III into *rplJL* and *rpoBC* mRNAs (Barry et al. 1980), so *rplJL* might also be cleaved into *rplJ* and *rplL* mRNAs. *rplJ* and *rplL* have an intercistronic region of ~66 bases in many bacterial species. Empirical studies on *Streptomyces coelicolor* support the cleavage between *rplJ* and *rplL* (Blanco et al. 2001), although no equivalent study has been carried out in *E. coli.* Therefore, *rplL* and *rplJ* are likely translated independently and produce different amounts of proteins to approximate the 4:1 stoichiometric ratio.

Given that *rplJ* and *rplL* could be processed into independent mRNAs, their abundance can differ due to differential degradation, even if their production rate is the same (being on the same transcription unit). In *E. coli*, ribosomes protect mRNA from endonucleolytic cleavage by endonucleases, such as RNase E (Braun et al. 1998) and SmrB (Saito et al. 2022). Consistent with this mRNA-protection effect of ribosomes, factors such as puromycin that encourage ribosome release from mRNA tend to destabilize mRNA, but factors such as chloramphenicol that freeze ribosomes on the mRNA tend to stabilize mRNA (Braun et al. 1998). Thus, if *rplL* mRNA is translated more actively than *rplJ* mRNA, then *rplL* mRNA should be more

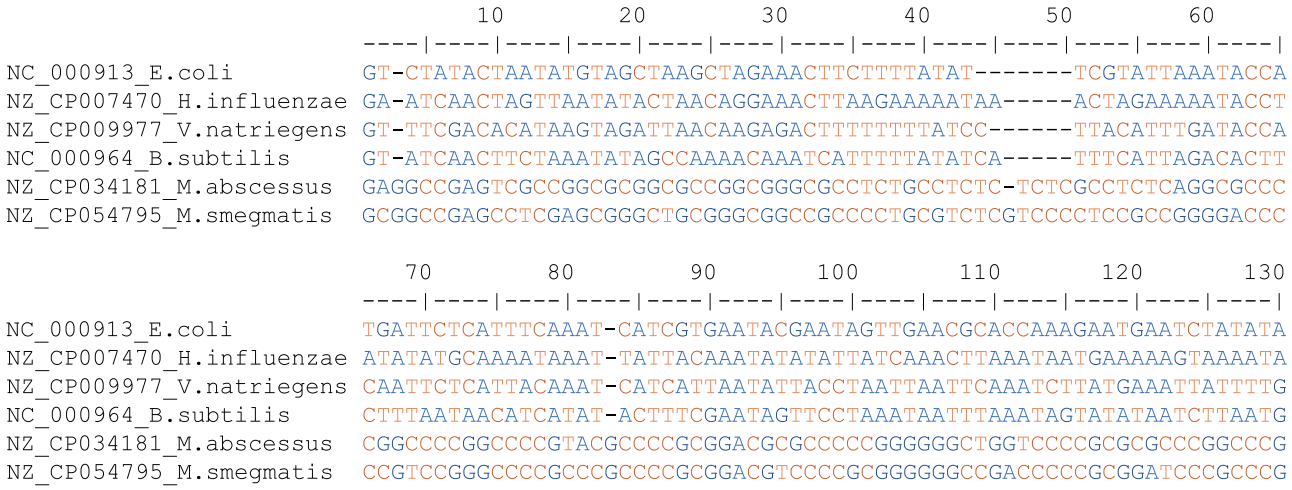

Figure 3 Third codon positions extracted from aligned *rplL* genes in the six species in Table 4. The last two species are GC rich. The sequence names are in the form of "Accession_Species name."

Table 5 mRNA abundance (in RPKM) in the four genes on the *rplJL-rpoBC* operon.

Strains	*rplJ*	*rplL*	*rpoB*	*rpoC*
Wild-type	5544.11	8061.62	1373.04	1645.25
rnb mutant	6442.26	9075.65	1224.94	1381.83
pnr mutant	5314.55	6845.82	789.20	1116.71
pnp mutant	6053.07	6887.81	1095.55	1365.67

abundant in the cell than *rplJ* mRNA, although their production rate is the same (being on the same transcription unit). This expectation is substantiated by the transcriptomic data (Table 5). The difference in mRNA abundance between *rplJ* and *rplL* (Table 5) is significant ($P = 0.01088$, one-tailed *t*-test).

I have also calculated MFE for the same three windows (30 + 10, 20 + 20, and 10 + 30) as specified in Table 2. The three window-specific MFE values are −8.5, −6.3, and −4, respectively, for the *rplJ* mRNA, and −4.1, −2.12 and −2.4, respectively, for the *rplL* mRNA. Thus, the secondary structure in TIR is weaker for *rplL* TIR than for *rplJ* TIR, consistent with the association of weaker secondary structure in TIR and translation initiation efficiency (Kudla et al. 2009; Prabhakaran et al. 2015; Xia 2023).

In short, *rplL* has more optimized codon usage and weaker secondary structure at its TIR than *rplJ.* These differences suggest that *rplL* is more efficiently translated than *rplJ.* More efficient translation initiation implies better mRNA protection against endonucleolytic cleavage. Consistent with this interpretation, *rplL* mRNA is more abundant than *rplJ*, although they share the same mRNA production, being on the same operon. All these results point to an enhanced production of L7/L12 relative to L10 as a mechanism to meet the 4:1 stoichiometric relationship.

Discussion

The results presented here are consistent with the interpretation that differential selection operates on highly expressed ribosomal genes to achieve functional stoichiometry (Li et al. 2014; Li 2015). This finding extends the classical framework of codon–anticodon adaptation (Ikemura 1981; Gouy and Gautier 1982; Sharp and Li 1987; Bulmer 1991; Ikemura 1992; Xia 1998) by showing how such differential selection can act specifically on biosynthetic processes to optimize cell growth and reproduction. Many protein complexes have their component proteins in certain stoichiometric relationships. Detecting differential selection on translation optimization among those component proteins will shed light on how biosynthetic systems can be optimized in the pharmaceutical industry.

Achieving equimolar protein production with different CDS lengths

The equimolar production of nearly all RPs, despite large variation in CDS length, poses a significant challenge to translational coordination. Long mRNAs tend to have reduced translation efficiency, which is substantiated by both experimental data (Lu et al. 2007) and theoretical modeling (Valleriani et al. 2011). This length-dependent effect is attributed to two factors (Valleriani et al. 2011). First, the RNA degradation in bacterial species is mainly carried out by endonucleases, such as RNase E and RNase III (or their equivalents). Ribosomes upstream of the cleavage site will then fail to complete the translation. Long mRNAs have a higher chance of being cleaved than short ones, leading to ribosomes failing to complete the translation. Second, ribosomes traveling along a long mRNA are more likely to experience traffic jams or ribosomal stalling than short ones (Mohammad et al. 2016; Bonnin et al. 2017). Ribosomal stalling may not only affect translation efficiency of the stalled ribosome but also cause endonucleolytic cleavage of the mRNA (Saito et al. 2022) and translation abortion of all ribosomes on the mRNA mediated by the tmRNA-SmpB complex (Keiler et al. 1996; Keiler 2015), RF3 (Zaher and Green 2011), ArfA (Chadani et al. 2010, 2012), and YaeJ/ArfB (Chadani et al. 2011) working either jointly or independently. These factors suggest that long mRNAs may have difficulties in producing the same number of proteins as short ones, everything else being equal. Thus, long

ribosome proteins are expected to be rate limiting and consequently subject to strong selection to optimize their translation initiation and codon usage.

There is indeed experimental evidence that S1, the longest RP, is rate limiting. Two *E. coli* mutants, MA261 and KK101, exhibit much reduced growth rates, with doubling times of 260 and 496 min, respectively. When putrescine is added to the culture, which stimulates the synthesis of RP S1 markedly but has little effect on other RPs (Kashiwagi et al. 1989), the doubling time is reduced from 260 to 128 min for the MA261 strain and from 496 to 135 min for the KK101 strain. This result is consistent with the hypothesis that the concentration of RP S1 is rate limiting in *E. coli*, ie increasing S1 concentration increases the replication rate. These observations suggest that longer RP mRNAs should have more efficient translation initiation, elongation, or both. The observed asymptotic increase in I_{TE} with the CDS length of RPs (Fig. 1) provides a mechanistic solution. Codon optimization accelerates translation elongation and potentially increases the number of proteins translated before mRNA cleavage or ribosome stalling.

I should mention here an alternative hypothesis for the increase of I_{TE} with mRNA length (Kudla et al. 2009). Long and highly expressed mRNAs could sequester too many ribosomes and consequently reduce the global availability of ribosomes. Such long and highly expressed mRNAs therefore need to have optimized codon usage to achieve efficient elongation, not because the codon optimization would increase protein production but because of the need to avoid excessive ribosome retention by these long and highly expressed mRNAs.

An alternative way for long RPs to have equimolar amounts as short ones is by having more active mRNA transcription, more efficient translation initiation (so that more ribosomes are translating along the mRNA than a short mRNA), or more efficient translation elongation. By studying RP genes in the same operon, one may remove the effect of differential mRNA abundance. The *S10*, *spc*, and *α* operons in *E. coli* encode 11, 10, and 5 RP genes, respectively. I_{TE} increases with CDS length within each of these operons (Table 1). It is remarkable that natural selection could operate so precisely among genes within the same operon.

The relationship between gene expression and sequence length can be affected by many factors (Coghlan and Wolfe 2000; Jansen and Gerstein 2000; Urrutia and Hurst 2003; Li et al. 2007). This is the first study that takes into consideration the stoichiometry of the component proteins in ribosomes in predicting the evolutionary trajectory of RP genes. The new conceptual framework results in more precise and testable predictions and a less ambiguous interpretation of empirical results.

The exceptional case of L7/L12

The unique multicopy requirement of the L7/L12 protein provides a clear case of directional selection on translation efficiency to achieve stoichiometric protein production. The consistently high I_{TE} of *rplL* (encoding L7/L12) relative to other RP genes (Fig. 1), including its operonic partner *rplJ* (Table 4), reflects a long-term evolutionary response to the need for multiple copies of L7/L12 per ribosome. This pattern is conserved in both Gram-negative and Gram-positive bacteria with diverse GC content (Fig. 3), indicating that the multicopy nature of L7/L12 imposes a universal selection pressure on synonymous sites in *rplL*.

The modest magnitude of ΔI_{TE} between *rplL* and *rplJ* (Table 4) may seem insufficient to generate a 4:1 stoichiometric ratio. The *rplL* mRNA may also have more efficient translation initiation. The minimum requirement for translation initiation is that the start codon is accessible (Nakamoto 2006). A start codon would not be accessible if it were embedded in a stable secondary structure. For example, the start codon AUG in the replicase gene of MS2 phage is embedded in a stem formed by part of the replicase gene and the upstream coat gene. Only when the coat gene is translated, separating the secondary structure as a consequence, can the downstream replicase gene be translated (Jou et al. 1972). Stable secondary structure in sequences flanking the start codon has been experimentally shown to inhibit translation initiation (Osterman et al. 2013), and mRNAs in bacterial species and unicellular eukaryotes tend to have much weaker secondary structure near the start codon than elsewhere, especially those from HEGs (Xia 2018b). Previous studies (Kudla et al. 2009; Tuller et al. 2010; Goodman et al. 2013; Prabhakaran et al. 2015; Xia 2015) demonstrated the relationship between translation initiation efficiency and the structural stability in sequences flanking the initiation signals (ie SD sequence and the start codon). In general, the weaker the secondary structure in sequences flanking the SD sequence and the start codon, the higher the translation initiation efficiency. Secondary structure stability is often measured by the MFE. MFE at the TIR is consistently greater (weaker secondary structure) in *rplL* than in *rplJ.* For example, MFE equals −0.6 and −6.3 for *rplL* and *rplJ*, respectively, in *E. coli*, and −0.6 and −3.8 for *rplL* and *rplJ*, respectively, in *B. subtilis*.

Actively translated mRNAs are protected by ribosomes from cleavage by endonucleases (Braun et al. 1998) (Saito et al. 2022). If *rplL* is translated more actively than *rplJ*, then *rplL* mRNA will become more abundant than *rplJ* mRNA because *rplL* mRNA would have reduced degradation relative to *rplJ* mRNA, although the two are produced at the same rate, being in the same transcription unit. This expectation is supported by empirical evidence (Table 5).

All these lines of evidence suggest that L7/L12 is more efficiently produced than L10 to approximate the 4:1 stoichiometric ratio. An alternative way of achieving the 4:1 ratio, assuming that L7/L12 cannot be produced more than L10 (Petersen 1989), is to degrade L10 (Petersen 1990). This alternative might be used under special circumstances, but it seems wasteful.

Bioengineering considerations

From an applied perspective, these findings have implications for synthetic biology and biotechnology. Translation optimization involves the optimization of the translation machinery and mRNA (Farookhi and Xia 2024; Xia 2026). The most rapidly replicating bacterial species, such as the Gram-negative *V. natriegens* or the Gram-positive *Clostridium perfringens*, have more *rrn* (ribosomal rRNA) operons and more tRNA genes in their genomes than their relatives (Xia 2026). A strong positive correlation is observed between the number of tRNA genes and *rrn* operons among diverse bacterial species with different doubling times (Xia 2026). The number of tRNA genes for amino acids is associated with the usage of the amino acids in proteins (Xia 1998, 2026).

The design of the COVID mRNA vaccine by Pfizer and Moderna incorporated optimization in translation initiation, elongation,

and termination, as well as mRNA stability (Xia 2021). Optimization of the translation machinery has lagged behind. This study shed light on the optimization of the RP genes. To achieve optimal protein production in proportion to stoichiometry, one could engineer an extra copy of the *rplL* gene in the *E. coli* genome so as to meet its multicopy requirement in each ribosome, as previously hinted (Mandava et al. 2012). Among all other RPs that should be produced in equimolar amounts, one may place long genes and short genes in separate operons and engineer stronger promoters, more efficient translation initiation sites, and more optimized codon usage for longer genes than for shorter genes.

In conclusion, this study demonstrates that synonymous codon usage is shaped not only by overall gene expression level but also by functional and structural constraints within protein complexes. Codon optimization serves as an adaptive mechanism to balance translation rates, maintain stoichiometric precision, and meet differential demand among subunits. The results refine our understanding of how natural selection acts at the most granular level of translation, where even small synonymous changes contribute to the efficiency and fidelity of the bacterial proteome.

Materials and methods

To test the first prediction, ie longer RP genes should be more codon optimized than shorter ones in *E. coli*, I downloaded from GenBank the reference genomes for *E. coli* K12 (Accession NC_000913), *V. natriegens* (NZ_CP009977, NZ_CP009978), *Haemophilus influenzae* (NZ_CP085952), *B. subtilis* (NC_000964), *Mycolicibacterium smegmatis* (NZ_CP080274), and *Mycobacteroides abscessus* (GCF_000069185) and extracted all RP genes using DAMBE (Xia 2018c) from the GenBank flat files (.gbff files). DAMBE can extract CDSs, upstream and downstream of CDSs, rRNA and tRNA sequences, sequences between CDSs, and sequences upstream and downstream of CDSs. For .gbff files for eukaryotic species, DAMBE can also extract introns, exons, and intron–exon junctions. To download a long list of GenBank files of bacterial genomes, one can enter the accessions into a text file, one accession per line, and use NCBI's Batch Entrez function to download all of them. These downloaded files are in GenBank flat file (.gbff) format, equivalent to the previous .gbk or .gb file format. They are all plain text files but contain gene annotations in the FEATURES table.

To extract all RPs in a species, one can first extract all CDSs and then save all RP genes into a separate file. To extract RP genes from a downloaded .gbff file, say E_coli.gbff, one would open E_coli.gbff into DAMBE. In the next dialog, choose "CDS" to extract all CDSs and the "gene" option to use gene names as sequence IDs. Using gene names as sequence IDs eases the extraction of RP genes because of their standard naming convention in annotated GenBank files. To save RP genes, click "Save a subset of sequences," provide a sequence file name, enter "^rps," and click "Find." This will find all genes with names starting with "rps," so genes with gene IDs "rpsA," "rpsB," …, "rpsU" (which are standard names for small ribosomal subunit protein genes in GenBank files for bacterial genomes) will all be highlighted. Click the right arrow to move them into the list of "Chosen Sequences." Do the same to add "rplA," "rplB," … "rplY" and "rpmA," "rpmB," …, "rpmJ" (both "rpl" and "rpm" indicate large ribosomal subunit protein genes) to the list of "Chosen sequences." Click "Go!" to save the RP genes.

I computed the index of translation elongation (I_{TE}) (Xia 2015) implemented in DAMBE (Xia 2017b). I_{TE} is a generalized CAI (Sharp and Li 1987; Xia 2007). Both I_{TE} and CAI have a minimum value of 0 and a maximum value of 1. I did not use CAI because it does not take background mutation into consideration. The problem of ignoring background mutation bias can be illustrated with the Ala codon subfamily GCR (where R stands for either A or G) (Xia 2015). The frequencies of GCA and GCG in *E. coli* HEGs, as compiled and distributed with EMBOSS (Rice et al. 2000), are 1973 and 2,654, respectively, with 57.36% being GCG. This may lead one to think that the *E. coli* translation machinery prefers GCG over GCA, ie GCG is a major codon. Indeed, CAI will treat GCG as a major codon. However, GCG is used even more frequently in *E. coli* non-HEGs, accounting for 62.91% among GCR codons. Thus, GCA is relatively more frequent in HEGs than in non-HEGs in *E. coli*. This suggests that mutation bias favors GCG, but the *E. coli* translation machinery prefers GCA, leading to relatively more frequent use of GCA in HEGs than in non-HEGs. This interpretation is consistent with tRNA genes in *E. coli.* The *E. coli* reference genome (NC_000913) encodes three $tRNA^{Ala}$ genes for GCR codons, all with a UGC anticodon forming a perfect Watson–Crick base pair with codon GCA. This is consistent with the interpretation that GCA is the true major codon, ie preferred by HEGs relative to non-HEG and decoded by the most abundant tRNA. I_{TE} takes the background mutation bias into account and should therefore be preferable over CAI. I_{TE} is reduced to CAI when there is no background mutation bias (Xia 2015; 2018a). One can also compute CAI with DAMBE so that comparisons between CAI and I_{TE} can be made. Both CAI and I_{TE} functions are accessed by clicking "Seq.Analysis | Codon usage" and followed by choosing either CAI or I_{TE} (index of translation elongation). For *E. coli*, one needs to choose "Escherichia coli" to use the *E. coli* reference codon usage table.

The first prediction is then reduced to a numerical relationship, ie I_{TE} should increase with increasing sequence length. I have also downloaded the reference genome of *B. subtilis* (Accession NC_000964), the model species of the Gram-positive bacteria, and carried out a similar analysis to extend the generality of the conceptual framework. Also analyzed is the genome from *V. natriegens* (NZ_CP009977), known to be the fastest replicating bacterial species (Eagon 1962; Yin et al. 2020) with a highly optimized translation machinery (Farookhi and Xia 2024; Xia 2026). I calculated I_{TE} based on R-ending and Y-ending subfamilies of codons (where R and Y are purine and pyrimidine, respectively) as was done before (Xia 2015) for *E. coli* genes.

The prediction that L7/L12, the only multicopy protein in bacterial ribosomes (Davydov et al. 2013), should have better codon adaptation than other RP genes can be tested in multiple ways, controlling for other potentially confounding variables. First, it should have a higher I_{TE} than others given the same sequence length or the same mRNA level. Second, L7/L12 and L10 are on the same operon (Barry et al. 1979; Fiil et al. 1979; Post et al. 1979; Christensen et al. 1984; Climie and Friesen 1988), so we can exclude the potential effect of differential mRNA levels by comparing I_{TE} values between L10 and L7/L12. Because the multicopy feature of the L7/L12 and its sharing the same operon with L10 appear ubiquitous in bacterial species (Oliveira et al. 1994; Yakhnin et al. 2015; Aseev et al. 2024), the comparison can be

made across multiple species. I used the nine species used in Farookhi and Xia (2024) for this comparison. These nine species include fast-replicating and slow-replicating bacterial species.

Translation initiation efficiency is often related to two features: (i) the MFE of the sequences near the start codon (Kozak 1980; Kudla et al. 2009; Prabhakaran et al. 2015; Xia 2018b) and (ii) the position and pairing strength between the SD sequences and the aSD sequence at the 3′ tail of the 16S rRNA. I used DAMBE (Xia 2017b), which includes the Vienna RNA secondary structure library (Hofacker 2003), to measure MFE and to characterize SD/aSD base pairing. Open a GenBank file in DAMBE, and extract CDSs together with 30 nucleotides upstream of the CDSs in the next dialog. Click "Sequences | Sequence manipulation | Delete specific segments," and keep 30 nucleotides at the 5′ end of the CDS together with the 30 nucleotides upstream of the CDS. Click "Structure | Minimum folding energy." In the next dialog, check the "Sliding window" checkbox and specify 40 nucleotides as the window size. Click "OK" to obtain the window-specific MFE.

For measuring mRNA expression, I used the transcriptomic data from *E. coli* K12 in a previous paper (Pobre and Arraiano 2015). The data set includes transcriptomic data from a wild-type strain and three mutants, in four files in the NCBI SRA database (SRR1536586 to SRR1536589). The read quality is high, as previously reanalyzed with the software ARSDA (Xia 2017a). The gene-specific mRNA abundance has already been available in the supplementary files of these two papers. However, the expression values (RPKM, reads per 1,000 bases per million mapped reads) in these two papers were not standardized exactly to 1,000,000 mapped reads per experiment, so my standardized RPKM values differ slightly from their values.

Acknowledgments

This research was funded by a Discovery Grant from the Natural Science and Engineering Research Council (NSERC, RGPIN-2024-05641) of Canada. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results. I thank members in XiaLab for their comments and discussion. Two anonymous reviewers provided excellent feedback that improved the paper in multiple ways.

Funding

This research was funded by a Discovery Grant from the Natural Science and Engineering Research Council (NSERC, RGPIN-2024-05641) of Canada.

Conflicts of Interest

The author declares no conflict of interests.

Data Availability

The extracted RP gene sequences from downloaded bacterial genomes were compiled in a set of FASTA files in a zip file available at https://dambe.bio.uottawa.ca/Bacteria_RP.zip. The newly compiled .ITE files used to compute I_{TE} of protein-coding genes using DAMBE are available in zipped format at http://dambe.bio.uottawa.ca/Files_for_ITE.zip. They should be unzipped and copied into the codon folder in DAMBE's installation folder.

References

Aliprandi P *et al.* S1 ribosomal protein functions in translation initiation and ribonuclease RegB activation are mediated by similar RNA-protein interactions: an NMR and SAXS analysis. *J Biol Chem.* 2008:283:13289–13301. https://doi.org/10.1074/jbc.M707111200.

Asato Y. Control of ribosome synthesis during the cell division cycles of E. coli and Synechococcus. *Curr Issues Mol Biol*. 2005:7:109–117. https://doi.org/10.21775/cimb.007.109.

Aseev LV, Koledinskaya LS, Boni IV. Extraribosomal functions of bacterial ribosomal proteins—an update, 2023. *Int J Mol Sci*. 2024:25:2957. https://doi.org/10.3390/ijms25052957.

Barry G, Squires C, Squires CL. Attenuation and processing of RNA from the rplJL–rpoBC transcription unit of Escherichia coli. *Proc Natl Acad Sci U S A.* 1980:77:3331–3335. https://doi.org/10.1073/pnas.77.6.3331.

Barry G, Squires CL, Squires C. Control features within the rplJL-rpoBC transcription unit of Escherichia coli. *Proc Natl Acad Sci U S A.* 1979:76:4922–4926. https://doi.org/10.1073/pnas.76.10.4922.

Blanco G, Sánchez C, Rodicio MR, Méndez C, Salas JA. Identification of a growth phase-dependent promoter in the rplJL operon of Streptomyces coelicolor A3(2). *Biochim Biophys Acta*. 2001:1517:243–249. https://doi.org/10.1016/S0167-4781(00)00280-3.

Bonnin P, Kern N, Young NT, Stansfield I, Romano MC. Novel mRNA-specific effects of ribosome drop-off on translation rate and polysome profile. *PLoS Comput Biol*. 2017:13:e1005555. https://doi.org/10.1371/journal.pcbi.1005555.

Braun F, Le Derout J, Régnier P. Ribosomes inhibit an RNase E cleavage which induces the decay of the rpsO mRNA of Escherichia coli. *EMBO J.* 1998:17:4790–4797. https://doi.org/10.1093/emboj/17.16.4790.

Brot N, Weissbach H. Chemistry and biology of E. coli ribosomal protein L12. *Mol Cell Biochem.* 1981:36:47–63. https://doi.org/10.1007/BF02354831.

Brown KM, Wade JT. Translational coupling of neighboring genes in prokaryotes. *J Bacteriol.* 2025:207:e00255-00225. https://doi.org/10.1128/jb.00255-25.

Bulmer M. The selection-mutation-drift theory of synonymous codon usage. *Genetics*. 1991:129:897–907. https://doi.org/10.1093/genetics/129.3.897.

Cerretti DP, Dean D, Davis GR, Bedwell DM, Nomura M. The spc ribosomal protein operon of Escherichia coli: sequence and cotranscription of the ribosomal protein genes and a protein export gene. *Nucleic Acids Res*. 1983:11:2599–2616. https://doi.org/10.1093/nar/11.9.2599.

Chadani Y *et al.* Ribosome rescue by Escherichia coli ArfA (YhdL) in the absence of trans-translation system. *Mol Microbiol.* 2010:78:796–808. https://doi.org/10.1111/j.1365-2958.2010.07375.x.

Chadani Y, Ito K, Kutsukake K, Abo T. Arfa recruits release factor 2 to rescue stalled ribosomes by peptidyl-tRNA hydrolysis in *Escherichia coli*. *Mol Microbiol.* 2012:86:37–50. https://doi.org/10.1111/j.1365-2958.2012.08190.x.

Chadani Y, Ono K, Kutsukake K, Abo T. *Escherichia coli* YaeJ protein mediates a novel ribosome-rescue pathway distinct from SsrA- and ArfA-mediated pathways. *Mol Microbiol.* 2011:80:772–785. https://doi.org/10.1111/j.1365-2958.2011.07607.x.

Christensen T, Johnsen M, Fiil NP, Friesen JD. RNA secondary structure and translation inhibition: analysis of mutants in the rplJ leader. *Embo J*. 1984:3:1609–1612. https://doi.org/10.1002/j.1460-2075.1984.tb02018.x.

Clarke JE, Kime L, Romero AD, McDowall KJ. Direct entry by RNase E is a major pathway for the degradation and processing of RNA in Escherichia coli. *Nucleic Acids Res.* 2014:42:11733–11751. https://doi.org/10.1093/nar/gku808.

Climie SC, Friesen JD. In vivo and in vitro structural analysis of the rplJ mRNA leader of Escherichia coli. Protection by bound L10-L7/L12. *J Biol Chem.* 1988:263:15166–15175. https://doi.org/10.1016/S0021-9258(18)68160-8.

Coghlan A, Wolfe KH. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. *Yeast*. 2000:16:1131–1145. https://doi.org/10.1002/1097-0061(20000915)16:12<1131::AID-YEA609>3.0.CO;2-F.

Curran JF, Yarus M. Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. *J Mol Biol*. 1989:209:65–77. https://doi.org/10.1016/0022-2836(89)90170-8.

Dai X *et al.* Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth. *Nat Microbiol*. 2016:2:16231. https://doi.org/10.1038/nmicrobiol.2016.231.

Davydov II *et al.* Evolution of the protein stoichiometry in the L12 stalk of bacterial and organellar ribosomes. *Nat Commun*. 2013:4:1387. https://doi.org/10.1038/ncomms2373.

Dennis PP, Bremer H. Differential rate of ribosomal protein synthesis in Escherichia coli B/r. *J Mol Biol*. 1974:84:407–422. https://doi.org/10.1016/0022-2836(74)90449-5.

Dennis PP, Nomura M. Stringent control of ribosomal protein gene expression in Escherichia coli. *Proc Natl Acad Sci U S A.* 1974:71:3819–3823. https://doi.org/10.1073/pnas.71.10.3819.

Diaconu M *et al.* Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell*. 2005:121:991–1004. https://doi.org/10.1016/j.cell.2005.04.015.

Duval M *et al. Escherichia coli* ribosomal protein S1 unfolds structured mRNAs onto the ribosome for active translation initiation. *PLoS Biol*. 2013:11:e1001731. https://doi.org/10.1371/journal.pbio.1001731.

Eagon RG. *Pseudomonas natriegens*, a marine bacterium with a generation time of less than 10 minutes. *J Bacteriol*. 1962:83:736–737. https://doi.org/10.1128/jb.83.4.736-737.1962.

Efron B. Bootstrap methods: another look at the jackknife. *The Annals of Statistitics*. 1979:7:1–26. https://doi.org/10.1214/aos/1176344552.

Efron B. *The jackknife, the bootstrap and other resampling plans*. Society for Industrial and Applied Mathematics; 1982.

Farookhi H, Xia X. Differential selection for translation efficiency shapes translation machineries in bacterial species. *Microorganisms*. 2024:12:768. https://doi.org/10.3390/microorganisms12040768.

Felsenstein J. Phylogenies and the comparative method. *Amer. Nat*. 1985:125:1–15. https://doi.org/10.1086/284325.

Fiil NP, Bendiak D, Collins J, Friesen JD. Expression of Escherichia coli ribosomal protein and RNA polymerase genes cloned on plasmids. *Mol Gen Genet.* 1979:173:39–50. https://doi.org/10.1007/BF00267689.

Gardin J *et al.* Measurement of average decoding rates of the 61 sense codons in vivo. *Elife*. 2014:3:e03735. https://doi.org/10.7554/eLife.03735.

Goodman DB, Church GM, Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. *Science*. 2013:342:475–479. https://doi.org/10.1126/science.1241934.

Gordiyenko Y *et al.* Mass spectrometry defines the stoichiometry of ribosomal stalk complexes across the phylogenetic tree. *Mol Cell Proteomics.* 2010:9:1774–1783. https://doi.org/10.1074/mcp.M000072-MCP201.

Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* 1982:10:7055–7074. https://doi.org/10.1093/nar/10.22.7055.

Guet CC *et al.* Minimally invasive determination of mRNA concentration in single living bacteria. *Nucleic Acids Res*. 2008:36:e73. https://doi.org/10.1093/nar/gkn329.

Haas J, Park E-C, Seed B. Codon usage limitation in the expression of HIV-1 envelope glycoprotein. *Curr Biol.* 1996:6:315–324. https://doi.org/10.1016/S0960-9822(02)00482-7.

Harms J *et al.* High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*. 2001:107:679–688. https://doi.org/10.1016/S0092-8674(01)00546-3.

Harvey PH, Pagel MD. *The comparative method in evolutionary biology*. Oxford University Press; 1991.

Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res*. 2003:31:3429–3431. https://doi.org/10.1093/nar/gkg599.

Hofmann JL, Yang TS, Sunol AM, Zia RN. Ribosomal L12 stalks recruit elongation factors to speed protein synthesis in Escherichia coli. *Commun Biol*. 2025:8:940. https://doi.org/10.1038/s42003-025-08366-4.

Huber M *et al.* Translational coupling via termination-reinitiation in archaea and bacteria. *Nat Commun*. 2019:10:4006. https://doi.org/10.1038/s41467-019-11999-9.

Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol.* 1981:146:1–21. https://doi.org/10.1016/0022-2836(81)90363-6.

Ikemura T. Correlation between codon usage and tRNA content in microorganisms. In: Hatfield DL, Lee BJ, Pirtle RM, editors. *Transfer RNA in protein synthesis*. CRC Press; 1992. p. 87–111.

Ilag LL *et al.* Heptameric $(L12)_6$/L10 rather than canonical pentameric complexes are found by tandem MS of intact ribosomes from thermophilic bacteria. *Proc Natl Acad Sci U S A.* 2005:102:8192–8197. https://doi.org/10.1073/pnas.0502193102.

Irshad IU, Sharma AK. Decoding stoichiometric protein synthesis in E. coli through translation rate parameters. *Biophys Rep (N Y)*. 2023:3:100131. https://doi.org/10.1016/j.bpr.2023.100131.

Jansen R, Gerstein M. Analysis of the yeast transcriptome with structural and functional categories: characterizing highly expressed proteins. *Nucleic Acids Res*. 2000:28:1481–1488. https://doi.org/10.1093/nar/28.6.1481.

Jeon HJ *et al.* Translation initiation control of RNase E-mediated decay of polycistronic gal mRNA. *Front Mol Biosci*. 2020:7:586413. https://doi.org/10.3389/fmolb.2020.586413.

Jeon HJ, N MPA, Wang X, Lim HM. Rho-dependent termination and RNase E-mediated cleavage: dual pathways for RNA 3′ end processing in polycistronic mRNA. *J Bacteriol.* 2025:207:e00437-00424. https://doi.org/10.1128/jb.00437-24.

Jou WM, Haegeman G, Ysebaert M, Fiers W. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature*. 1972:237:82–88. https://doi.org/10.1038/237082a0.

Kashiwagi K, Sakai Y, Igarashi K. Polyamine stimulation of ribosomal synthesis and activity in a polyamine-dependent mutant of *Escherichia coli*. *Arch Biochem Biophys.* 1989:268:379–387. https://doi.org/10.1016/0003-9861(89)90598-5.

Keiler KC. Mechanisms of ribosome rescue in bacteria. *Nat Rev Microbiol*. 2015:13:285–297. https://doi.org/10.1038/nrmicro3438.

Keiler KC, Waller PR, Sauer RT. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science*. 1996:271:990–993. https://doi.org/10.1126/science.271.5251.990.

Korobeinikova AV, Garber MB, Gongadze GM. Ribosomal proteins: structure, function, and evolution. *Biochemistry (Mosc).* 2012:77:562–574. https://doi.org/10.1134/S0006297912060028.

Kozak M. Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits. *Cell*. 1980:19:79–90. https://doi.org/10.1016/0092-8674(80)90390-6.

Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*. 2009:324:255–258. https://doi.org/10.1126/science.1170160.

Levin-Karp A *et al.* Quantifying translational coupling in E. coli synthetic operons using RBS modulation and fluorescent reporters. *ACS Synth Biol*. 2013:2:327–336. https://doi.org/10.1021/sb400002n.

Li GW. How do bacteria tune translation efficiency? *Curr Opin Microbiol*. 2015:24:66–71. https://doi.org/10.1016/j.mib.2015.01.001.

Li GW, Burkhardt D, Gross C, Weissman JS. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*. 2014:157:624–635. https://doi.org/10.1016/j.cell.2014.02.033.

Li S-W, Feng L, Niu D-K. Selection for the miniaturization of highly expressed genes. *Biochem Biophys Res Commun.* 2007:360:586–592. https://doi.org/10.1016/j.bbrc.2007.06.085.

Liljas A, Sanyal S. The enigmatic ribosomal stalk. *Q Rev Biophys.* 2018:51:e12. https://doi.org/10.1017/S0033583518000100.

Lindahl L, Archer R, Zengel JM. Transcription of the S10 ribosomal protein operon is regulated by an attenuator in the leader. *Cell*. 1983:33:241–248. https://doi.org/10.1016/0092-8674(83)90353-7.

Lindahl L, Zengel JM. Operon-specific regulation of ribosomal protein synthesis in Escherichia coli. *Proc Natl Acad Sci U S A.* 1979:76:6542–6546. https://doi.org/10.1073/pnas.76.12.6542.

Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol*. 2007:25:117–124. https://doi.org/10.1038/nbt1270.

Lund PE, Chatterjee S, Daher M, Walter NG. Protein unties the pseudoknot: S1-mediated unfolding of RNA higher order structure. *Nucleic Acids Res.* 2020:48:2107–2125. https://doi.org/10.1093/nar/gkz1166.

Mackie GA. RNase E: at the interface of bacterial RNA processing and decay. *Nat Rev Microbiol.* 2013:11:45–57. https://doi.org/10.1038/nrmicro2930.

Mandava CS *et al.* Bacterial ribosome requires multiple L12 dimers for efficient initiation and elongation of protein synthesis involving IF2 and EF-G. *Nucleic Acids Res*. 2012:40:2054–2064. https://doi.org/10.1093/nar/gkr1031.

Meek DW, Hayward RS. Nucleotide sequence of the rpoA-rplQ DNA of Escherichia coli: a second regulatory binding site for protein S4? *Nucleic Acids Res*. 1984:12:5813–5821. https://doi.org/10.1093/nar/12.14.5813.

Mohammad F, Woolstenhulme CJ, Green R, Buskirk AR. Clarifying the translational pausing landscape in bacteria by ribosome profiling. *Cell Rep*. 2016:14:686–694. https://doi.org/10.1016/j.celrep.2015.12.073.

Nakamoto T. A unified view of the initiation of protein synthesis. *Biochem Biophys Res Commun.* 2006:341:675–678. https://doi.org/10.1016/j.bbrc.2006.01.019.

Ngumbela KC *et al.* Quantitative effect of suboptimal codon usage on translational efficiency of mRNA encoding HIV-1 *gag* in intact T cells. *PLoS One*. 2008:3:e2356. https://doi.org/10.1371/journal.pone.0002356.

Oliveira SC, Zhu Y, Splitter G. Sequences of the rplJL operon containing the L10 and L7/L12 genes from Brucella abortus. *Gene*. 1994:140:137–138. https://doi.org/10.1016/0378-1119(94)90744-7.

Osterman IA, Evfratov SA, Sergiev PV, Dontsova OA. Comparison of mRNA features affecting translation initiation and reinitiation. *Nucleic Acids Res*. 2013:41:474–486. https://doi.org/10.1093/nar/gks989.

Petersen C. Long-range translational coupling in the rplJL-rpoBC operon of Escherichia coli. *J Mol Biol*. 1989:206:323–332. https://doi.org/10.1016/0022-2836(89)90482-8.

Petersen C. Escherichia coli ribosomal protein L10 is rapidly degraded when synthesized in excess of ribosomal protein L7/L12. *J Bacteriol*. 1990:172:431–436. https://doi.org/10.1128/jb.172.1.431-436.1990.

Pettersson I, Kurland CG. Ribosomal protein L7/L12 is required for optimal translation. *Proc Natl Acad Sci U S A.* 1980:77:4007–4010. https://doi.org/10.1073/pnas.77.7.4007.

Pobre V, Arraiano CM. Next generation sequencing analysis reveals that the ribonucleases RNase II, RNase R and PNPase affect bacterial motility and biofilm formation in E. coli. *BMC Genomics*. 2015:16:72. https://doi.org/10.1186/s12864-015-1237-6.

Post LE, Strycharz GD, Nomura M, Lewis H, Dennis PP. Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit beta in Escherichia coli. *Proc Natl Acad Sci U S A.* 1979:76:1697–1701. https://doi.org/10.1073/pnas.76.4.1697.

Prabhakaran R, Chithambaram S, Xia X. *Escherichia coli* and *Staphylococcus* phages: effect of translation initiation efficiency on differential codon adaptation mediated by virulent and temperate lifestyles. *J Gen Virol*. 2015:96:1169–1179. https://doi.org/10.1099/vir.0.000050.

Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet*. 2000:16:276–277. https://doi.org/10.1016/S0168-9525(00)02024-2.

Robinson M *et al.* Codon usage can affect efficiency of translation of genes in *Escherichia coli*. *Nucleic Acids Res*. 1984:12:6663–6671. https://doi.org/10.1093/nar/12.17.6663.

Ryals J, Little R, Bremer H. Temperature dependence of RNA synthesis parameters in Escherichia coli. *J Bacteriol*. 1982:151:879–887. https://doi.org/10.1128/jb.151.2.879-887.1982.

Saito K *et al.* Ribosome collisions induce mRNA cleavage and ribosome rescue in bacteria. *Nature*. 2022:603:503–508. https://doi.org/10.1038/s41586-022-04416-7.

Saito K, Green R, Buskirk AR. Ribosome recycling is not critical for translational coupling in Escherichia coli. *Elife*. 2020:9:e59974. https://doi.org/10.7554/eLife.59974.

Schuwirth BS *et al.* Structures of the bacterial ribosome at 3.5 A resolution. *Science*. 2005:310:827–834. https://doi.org/10.1126/science.1117230.

Sharp PM, Li WH. The codon adaptation index–a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*. 1987:15:1281–1295. https://doi.org/10.1093/nar/15.3.1281.

Shen V, Bremer H. Rate of ribosomal ribonucleic acid chain elongation in Escherichia coli B/r during chloramphenicol treatment. *J Bacteriol*. 1977:130:1109–1116. https://doi.org/10.1128/jb.130.3.1109-1116.1977.

Skorski P, Leroy P, Fayet O, Dreyfus M, Hermann-Le Denmat S. The highly efficient translation initiation region from the *Escherichia coli* rpsA gene lacks a Shine-Dalgarno element. *J Bacteriol*. 2006:188:6277–6285. https://doi.org/10.1128/JB.00591-06.

Sorensen MA, Kurland CG, Pedersen S. Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol*. 1989:207:365–377. https://doi.org/10.1016/0022-2836(89)90260-X.

Tuller T, Waldman YY, Kupiec M, Ruppin E. Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci U S A.* 2010:107:3645–3650. https://doi.org/10.1073/pnas.0909910107.

Urrutia AO, Hurst LD. The signature of selection mediated by expression on human genes. *Genome Res.* 2003:13:2260–2264. https://doi.org/10.1101/gr.641103.

Valleriani A, Zhang G, Nagar A, Ignatova Z, Lipowsky R. Length-dependent translation of messenger RNA by ribosomes. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2011:83:042903. https://doi.org/10.1103/PhysRevE.83.042903.

Vila-Sanjurjo A. Modification of the ribosome and the translational machinery during reduced growth due to environmental stress. *EcoSal Plus*. 2008:3. 10.1128/ecosalplus.1122.1125.1126.

Wimberly BT *et al.* Structure of the 30S ribosomal subunit. *Nature*. 2000:407:327–339. https://doi.org/10.1038/35030006.

Wittmann H-G. Structure, function and evolution of ribosomes. *Eur J Biochem.* 1976:61:1–13. https://doi.org/10.1111/j.1432-1033.1976.tb09992.x.

Wittmann HG. Components of bacterial ribosomes. *Annu Rev Biochem*. 1982:51:155–183. https://doi.org/10.1146/annurev.bi.51.070182.001103.

Wittmann-Liebold B. Ribosomal proteins: their structure and evolution. In: Hardesty B, Kramer G, editors. *Structure, function, and genetics of ribosomes*. Springer New York; 1986. p. 326–361.

Xia X. How optimized is the translational machinery in *Escherichia coli*, *Salmonella typhimurium* and *Saccharomyces cerevisiae*? *Genetics*. 1998:149:37–44. https://doi.org/10.1093/genetics/149.1.37.

Xia X. An improved implementation of codon adaptation index. *Evol Bioinform Online.* 2007:3:53–58. https://doi.org/10.1177/117693430700300028.

Xia X. A Major controversy in Codon-anticodon adaptation resolved by a new codon usage index. *Genetics*. 2015:199:573–579. https://doi.org/10.1534/genetics.114.172106.

Xia X. ARSDA: a new approach for storing, transmitting and analyzing transcriptomic data. *G3: Genes|Genomes|Genetics*. 2017a:7:3839–3848. https://doi.org/10.1534/g3.117.300271.

Xia X. DAMBE6: new tools for microbial genomics, phylogenetics, and molecular evolution. *J Hered*. 2017b:108:431–437. https://doi.org/10.1093/jhered/esx033.

Xia X. Bioinformatics and translation elongation. In: *Bioinformatics and the cell: modern computational approaches in genomics, proteomics and transcriptomics*. Springer, Cham; 2018a. p. 197–238.

Xia X. Bioinformatics and translation initiation. In: *Bioinformatics and the cell: modern computational approaches in genomics, proteomics and transcriptomics*. Springer, Cham; 2018b. p. 173–195.

Xia X. DAMBE7: new and improved tools for data analysis in molecular biology and evolution. *Mol Biol Evol.* 2018c:35:1550–1552. https://doi.org/10.1093/molbev/msy073.

Xia X. Detailed dissection and critical evaluation of the Pfizer/BioNTech and Moderna mRNA vaccines. *Vaccines (Basel)*. 2021:9:734. https://doi.org/10.3390/vaccines9070734.

Xia X. Optimizing protein production in therapeutic phages against a bacterial pathogen, *Mycobacterium abscessus*. *Drugs Drug Candidates.* 2023:2:189–209. https://doi.org/10.3390/ddc2010012.

Xia X. Evolution of translational machinery in fast- and slow-growing bacteria. *Microorganisms*. 2026:14:377. https://doi.org/10.3390/microorganisms14020377.

Yakhnin H, Yakhnin AV, Babitzke P. Ribosomal protein L10(L12)4 autoregulates expression of the Bacillus subtilis rplJL operon by a transcription attenuation mechanism. *Nucleic Acids Res*. 2015:43:7032–7043. https://doi.org/10.1093/nar/gkv628.

Yin M *et al.* Changes in *Vibrio natriegens* growth under simulated microgravity. *Front Microbiol*. 2020:11:2040. https://doi.org/10.3389/fmicb.2020.02040.

Young R, Bremer H. Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of growth rate. *Biochem J*. 1976:160:185–194. https://doi.org/10.1042/bj1600185.

Zaher HS, Green R. A primary role for release factor 3 in quality control during translation elongation in *Escherichia coli*. *Cell*. 2011:147:396–408. https://doi.org/10.1016/j.cell.2011.08.045.

Zengel JM, Mueckl D, Lindahl L. Protein L4 of the E. coli ribosome regulates an eleven gene r protein operon. *Cell*. 1980:21:523–535. https://doi.org/10.1016/0092-8674(80)90490-0.

Zurawski G, Zurawski SM. Structure of the Escherichia coli S10 ribosomal protein operon. *Nucleic Acids Res*. 1985:13:4521–4526. https://doi.org/10.1093/nar/13.12.4521.