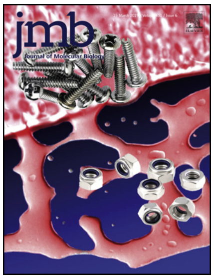

Fast Translation within the First 45 Codons Decreases mRNA Stability and Increases Premature Transcription Termination in *E. coli* ☆

Steen Pedersen [1,2], Thilde Bagger Terkelsen [1], Mette Eriksen [1], Magnus Krarup Hauge [1], Casper Carstens Lund [1], Kim Sneppen [2] and Namiko Mitarai [2]

*1 - **Department of Biology,** University of Copenhagen, Ole Maaløes vej 5, DK-2200 Copenhagen N, Denmark*
*2 - **Center for Models of Life,** Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark*

***Correspondence to Namiko Mitarai:** mitarai@nbi.dk.*
https://doi.org/10.1016/j.jmb.2019.01.026
Edited by Suckjoon Jun

Abstract

We show here that the specific use of fast or slowly translated codons in the early coding region of a gene may influence both the mRNA stability and premature transcription termination. We first inserted a pair of nearly identical 42-base-pair (bp)-long sequences into codon 3 of the *Escherichia coli lacZ* gene. The only difference between the two inserts was that the first base in one was moved to become the last base in the other, providing a difference in the reading frame, one of which had the biased codons typical for ribosomal protein genes and which previously was shown to be faster translated than average. This insert reduced the mRNA stability and increased premature transcription termination and together resulted in a hundred-fold difference in *lacZ* expression. We next generated *lacZ* variants with 7, 14 or 21 fast translated, ribosomal-type codons inserted into codon 13 of *lacZ*. This gave progressively more unstable mRNAs and also progressively increased transcription termination up to 90%. By modeling, based on estimates of the translation rate of individual codons, we can explain these observations by an increased susceptibility of the mRNA to degradation, determined by the length and degree of the early mRNA being uncovered by ribosomes. Thus, we suggest that the translation rate differences among the synonymous codons early in a gene enable a "velocity code" within the amino acid coding ability, where the translation rate differences encode the mRNA stability and the premature termination of the RNA polymerase.

☆ This article is part of the special issue on "Biovisualization" edited by Donald E Ingber, Janet H Iwasa, Jodie Jenkinson, Gaël McGill, Seán O'Donoghue, Charles Reilly, Alain Viel.

Introduction

The first gene to become sequenced was the *lacI* gene of *Escherichia coli* [1]. It showed no significant bias in the use of synonymous codons. In contrast, the four ribosomal protein genes next to be sequenced had a very strong codon bias [2]. The ribosomal proteins are expressed about hundred-fold more than the Lac repressor. It was therefore suggested that the biased codons in the ribosomal genes contributed to the high expression, possibly because they were translated fast [3]. This codon bias was also evident in the several hundred *E. coli* genes that became sequenced during the next years. The codon adaptation index (CAI) was defined that quantified this codon bias [4]. Early experiments showed a substantial translation rate difference among the individual codons [5] that in some cases were caused by translation of a rarely used codon [6].

That the ribosomal-type codons were translated fast was shown experimentally by measuring the translation chain growth rate of four highly expressed ribosomal-type mRNAs and comparing the rate to that of mRNA from two genes, *lacI* and *bla*, with a non-biased codon usage. The result was that the ribosomes were translating the non-biased codons with approximately half the speed of that of the biased codons used for encoding the highly expressed proteins [7]. By determining the translation speed of two almost identical sequences inserted in two different reading frames, with either ribosomal-type biased or non-biased codons, we suggested that the difference in translation speed was due to the codon bias and not due to hypothetical differences in mRNA structure [8]. Finally, the translation speed of four individual codons was measured. The span in the velocity was 5-fold and two codons read by the same tRNA differed 3-fold, the fastest being the one preferred in the ribosomal genes [9].

Inspecting the sequences of hundreds of genes from *E. coli*, Bulmer [10] noticed that the CAI was substantially lower in the initial part of an mRNA, even in those for the highly expressed genes. He suggested that the biased codon usage in the later part of the genes was selected for and that the early part of the genes drifted toward having non-biased codon usage because of the absence of a selective pressure. This suggestion was later corrected because the rate of synonymous codon changes was lower in the start of genes and the non-biased codons were suggested to diminish the tendency to form mRNA structures in the early part of the gene [11] or restrict the ribosome entry to the mRNA to increase the overall translation efficiency by minimizing queue formation in the distal part of the gene [12,13]. We modeled the half-lives found for mRNA with early sequence-variants of the *lacZ* gene [14] and this pointed to that local variations in the ribosome density early on the mRNA influenced the mRNA half-life [15].

To analyze the function of the early, non-biased codon usage further, we have here constructed several variants of the *lac* mRNA with inserts of biased or non-biased codons early in the *lacZ* gene. These changes resulted in an up to 10-fold variation in both the functional half-life of the *lacZ* mRNA and, more surprisingly, in the degree of premature transcription termination, measured as the failure of the RNA polymerase to reach the distal *lacA* gene and together altered the expression hundred-fold. We argue here, based on experimental results and modeling, that the mechanism for this change in expression involves a translation rate difference between the synonymous codons. In other words, the specific use of synonymous codons early in a gene generates a "velocity code," within the amino acid coding ability, where the translation rate differences give mRNA stretches of varying length that is uncovered by ribosomes, thereby exposing the mRNA to varying rates of degradation. In addition, we also observed varying premature terminations of the RNA polymerase.

Results

To find the reasons for having non-biased codons in the early mRNA we constructed several variants of the *lacZ* gene with inserts of the ribosomal-type biased codons in the early coding region. As described in Methods, we first introduced a unique restriction enzyme site, *Bst*EII in codon 3 of *lacZ* and then inserted two 14-codon-long sequences into this site. These sequences were identical, except that the first base in the insert, in pSP1779, was moved to become the last base in the other, pSP1780. The pSP1779 insert had a biased ribosomal-codon-type, GAA-rich insert, and pSP1780 had a frame-shifted insert dominated by codon AAG, not frequently found in the ribosomal-type genes. Similarly, we inserted 7, 14 or 21 codon GAA-rich sequences into codon 13 of pSN4 [15]. Finally, we supplemented the *lacZ* gene in these plasmids with the *lacY* and *lacA* genes to give pMET29, pMET30 and pMET31, respectively. For cloning reasons, a few additional biased codons were included in the inserts. The relevant sequences of these inserts are presented in Table 1.

The inserts in pMET29, pMET30, pMET31 and pSP1779 were dominated by codon GAA, which previously was measured to be translated with two to three times the average rate [9,12]. The reading frame in the pSP1780 insert was dominated by codon AAG. The CAI [4] of these inserts is above 0.74 in the fast inserts and 0.07 in the slow reading frame insert. The CAI value of 0.74 is higher than in any natural gene of *E. coli,* and the CAI of 0.07 is lower than in the least expressed genes. As the CAI is a coarse indicator of the translation rate [7,8,12], this suggests that the inserts in these variants are translated with different rates. In addition, the GA-rich sequences we inserted have the advantage of not being able to give fold back structures with the Shine–Dalgarno sequence as this will reduce the translational initiation frequency [16,17].

We first measured the expression and the functional mRNA half-life of these variants of *lacZ*. As described in the Methods section, we used three different methods, A, B and C, for determining the functional *lacZ* mRNA half-lives. The half-lives of the *lacZ* mRNA from pSP1779 and of pSP1780 were first determined by method A, removal of the inducer followed by measuring the time course of the residual capacity to synthesize β-galactosidase by pulse labeling. The half-life of the pSP1779 *lacZ* mRNA was too short to be determined by this method, whereas that of pSP1780 was approximately 35 s. To estimate the short half-life of the pSP1779 variant, we used method C (see Fig. 1 B). The results are given in Table 2. The expression from pSP1780 is seen to be approximately one fourth of that for the wild-type *lacZ*, in agreement with the values found by method A, but 40-fold higher

Table 1. Relevant sequences of the *lacZ* variants mRNA

Plasmid	*lacZ* coding sequence
pMET1 Codon 1-20	AUG ACC uag AUU* ACG GAU UCA CUG GCC GUC GUU UUA CAA CGU CGU GAC UGG GAA AAC CCU
pJEM1 Codon 1-20	AUG guc acc AUU* ACG GAU UCA CUG GCC GUC GUU UUA CAA CGU CGU GAC UGG GAA AAC CCU
pSN4 Codon 1-20	AUG ACC AUG AUU ACG GAU UCA CUG GCC GUC GUU ggu cac cgu cgu agg cgg gga AAC$^{\$}$ CCU
pCARO1 codon 1-20	AUG ACC AUG AUU ACG GAU UCA CUG GCC GUC GUU ggu cac cgu cgu aga cgg gga AAC$^{\$}$ CCU
pTT1 Codon 1-20	AUG ACC AUG AUU ACG GAU UCA CUG GCC GUC GUU ggu cac cgu cgu agg cgg cga AAC$^{\$}$ CCU
pTT2 Codon 1-20	AUG ACC AUG AUU ACG GAU UCA CUG GCC GUC GUU ggu cac cgu cgu cga ccg gga AAC$^{\$}$ CCU
pJEM2/ pSP1779 Codon 1-20	AUG guc acg gaa gaa gcu auc gaa gaa gcu auc gaa gaa gcu auc gaa gaa guc acc AUU*
pJEM3/ pSP1780 Codon 1-20	AUG guc acg aag aag cua ucg aag aag cua ucg aag aag cua ucg aag aag guc acc AUU*
pSP1781	AUG ACC AUG AUU ACG GAU UCA CUG GCC GUC GUU ggu cac (gga aga agc uau)3 cga aga agu cac cgu cgu agg cgg gga AAC$^{\$}$
pSP1782	AUG ACC AUG AUU ACG GAU UCA CUG GCC GUC GUU ggu cac (gaa gaa gcu auc)3 gaa gaa ggu cac cgu cgu agg cgg gga AAC$^{\$}$
pMAP1058/ pMET29 Codon 1-26	AUG ACC AUG AUU ACG GAU UCA CUG GCC GUC GUU ggu cac gaa gaa gcu gaa gaa ggu cac cgu cgu agg cgg gga AAC$^{\$}$
pMAP1059 / pMET30 Codon 1-33	AUG ACC AUG AUU ACG GAU UCA CUG GCC GUC GUU ggu cac (gaa)3 gcu (gaa)4 gcu (gaa)3 ggu cac cgu cgu agg cgg gga AAC$^{\$}$
pMAP1060/ pMET31 Codon 1-40	AUG ACC AUG AUU ACG GAU UCA CUG GCC GUC GUU ggu cac (gaa)3 gcu (gaa)4 gcu (gaa)4 gcu (gaa)5 ggu cac cgu cgu agg cgg gga AAC$^{\$}$

[a] Designates codon 4 in the *lacZ* wild-type sequence. The sequence distal to codon 4 in these variants is the *lacZ* wild-type sequence.
[b] Designates codon 19 in the *lacZ* wild-type sequence. The sequence distal to codon 19 in these variants is the *lacZ* wild-type sequence.
Lower case font: sequences different from the *lacZ* wild type. The codons of special interest are marked with the following: red, expected slowly translated codons; orange, expected translation with average rate; and green, expected fast translated codons. The base generating the different reading frame in pJEM2/pSP1779 and pJEM3/pSP1780 is underlined.

than that of pSP1779. This indicated that the apparent half-life of the *lacZ* mRNA encoded by pSP1779 was close to 1 s (but see later).

Our attempts to determine the mRNA stability in the three variants with 7, 14 or 21 codons inserted in codon 13 by either method A or B were unsuccessful, indicating that the half-lives of these variant mRNAs were much shorter than that of the parental pSN4 *lacZ* variant, 35 s. To estimate these, we used method C and the results are shown in Fig. 1 C and Table 2. As seen, there is a progressive decrease in the β-galactosidase expression when inserting longer and longer stretches of the ribosomal-gene-type, fast-translated codons.

Our early failed efforts to model these results made it clear that an additional mechanism than the effect on the half-lives was affecting the expression levels. The *lac* mRNA is cleaved between *lacZ* and

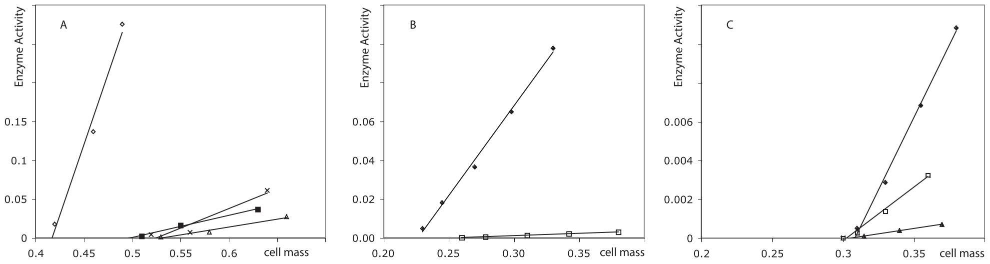

Fig. 1. Differential expression of LacZ enzymatic activity after induction. (A) Wild type (diamonds), pSN4 (squares), pTT1 (triangles) and pTT2 (crosses). (B) pSP1779 (squares) and pSP1780 (diamonds). (C) pMET19 (diamonds), pMET30 (squares) and pMET31 (triangles).

Table 2. The expression levels of β-galactosidase and the Lac-transacetylase

Strain (lacZ variant)	Relative LacZ activity (SEM)	Relative LacA activity (SEM)	LacZ mRNA half-life (s) 116 s × relative lacZ activity/ relative lacA activity (SEM)
MET20 (WT)	1 (0.10)	1 (0.10)	116 (17.0)
SP1779	0.0087 (0.0016)	0.123 (0.019)	8.2 (2.0)
SP1780	0.33 (0.07)	0.915 (0.158)	41.4 (11.8)
SP1781	0.018 (0.0019)	0.37 (0.005)	5.6 (1.0)
SP1782	0.074 (0.0011)	0.56 (0.07)	15.3 (3.0)
MET29	0.016 (0.0033)	0.179 (0.030)	10.2 (2.7)
MET30	0.0066 (0.0018)	0.110 (0.017)	7.0(2.2)
MET31	0.0029 (0.0008)	0.084 (0.014)	4.0 (1.2)

The fully induced β-galactosidase level for the pMET20 strain is approximately 19,000 Miller units, and that for the LacA enzyme is 440 units.

lacY–lacA [18]. The *lacA* mRNA part thus has an independent half-life from the *lacZ* mRNA part of the mRNA [14,19]. Consequently, *lacA* expression should be unaffected by changes in *lacZ*, unless an event in *lacZ* terminated the RNA polymerase. We determined such premature transcription termination by measuring the expression of the distal *lacA* gene and found it to vary up to 10-fold among the different *lacZ* variants. The expression of *lacZ*, the half-lives and the *lacA* expression are shown in Table 2. By an accidental error, one inserted sequence generated a frameshift, which gave rise to a UGA stop codon in codon 23. To compare our results with termination in a polar *lacZ* mutant, we also measured the LacA activity in this strain and found that 97% of all transcripts were terminated. This is very close to the value expected from previous studies [20].

Finally, adjusting for the premature transcription termination, we calculate the mRNA functional half-lives in the last column of Table 2. We analyze the results with pMET29, 30 and 31 in Table 3. As seen from this table, we are able to calculate the mRNA half-lives for the 14- and 21-bp insert variants mRNA from the half-lives of the parental variant in pSN4 and the 7-bp insert in pMET29 that are within the bounds set by the SEM of the determinations.

The parameter affecting the mRNA half-life: translation rate or mRNA structure

We previously demonstrated that alteration of three codons at position 16–18 to AGG, CGG and GGA (expected to be slowly translated) in the *lacZ* variant pSN4 destabilized the mRNA 4-fold [15]. If this was a result of an increased ribosome spacing downstream after this bottleneck, as we suggested, altering these three codons to three different but still slowly translated codons should not change the mRNA stability substantially. Alternatively, if we unknowingly had created a sequence in pSN4 that destabilized the mRNA by forming a target for degradation, for instance, a particular RNA structure, the new mRNA with these perturbed sequences should each have different stability and probably closer to that of the wild type.

To test whether the mRNA stabilities were correlated to these three slowly translated codons or to hypothetical mRNA structures, we made three additional alterations of codons 16–18: pTT1 (AGG, CGG, CGA), pTT2 (CGA, CCG, GGA) and pCARO1 (AGA, CGG, GGA). Analysis of, for example, the mRNA codon 1 to codon 33 by the MFOLD algorithm http://unafold.rna.albany.edu/?q=mfold [21] indicated that the region in pSN4 and in pCARO1 can form the same potential structure, whereas these in pTT1

Table 3. Analysis of half-life data from Table 2

	Insert	0 bp	7 bp	14 bp	21 bp
1	Half-life	35 s	10.2 s	7.0 s	4.0 s
2	$A + \varepsilon$	1/35 = 0.029	1/10.2 = 0.098	1/7.0 = 0.143	1/3.9 = 0.23
3	A	0	0.069	0.114	0.201
4	Calc α's	–	$\alpha_7 = 0.069$	$\alpha_{14} = 2\alpha_7 = 0.138$	$\alpha_{21} = 3\alpha_7 = 0.207$
5	Calc. half-life: $1/(\alpha_n + \varepsilon)$	= 1/0.029 = 35 s	= 1/(0.069 + 0.029) = 10.2 s	= 1/(0.138 + 0.029) = 6.0 s	= 1/(0.207 + 0.029) = 4.2 s

Row1. The experimentally determined half-lives from either pSN4 [15] or pMET29, MET30 and MET31 from Table 2. These *lacZ* variants have 0, 7, 14 or 21 bp inserted into codon 13 of pSN4.

Row 2. We separate the degradation constant into one, ε, from the parental construct pSN4 without inserts and three, A_7, A_{14} and A_{21}, resulting from the insert of 7, 14 and 21 bp, respectively. The degradation constants, in s^{-1}, are proportional to the inverse of the half-life.

Row 3. The degradation constants A_n from experiments.

Row 4. We now try to see if the rate of degrading the mRNA due to the inserted sequence of 2 or 3 units of 7 fast codons can be estimated from the rate of degradation, α_7 for one 7-bp unit. If so, the four degradation rates, α_0, α_7, α_{14} and α_{21} should be ε, $\varepsilon + \alpha_7$, $\varepsilon + 2\alpha_7$ and $\varepsilon + 3\alpha_7$, respectively.

Row 5. The calculated half-lives for the 2- or 3-unit inserts, including the experimental degradation rates for the parental *lacZ* and the 7-bp insert variant.

Table 4. Half-life of lacZ mRNAs with perturbations of slowly translated codons at codons 16, 17 and 18

lacZ variant	Half-life (s)		
	Method A	Method B	Method C
pMAP217, wild-type half-life	116	116	116
pSN4 altered codons 16–18	28	54	8
pCARO1 altered codons 16–18	28	26	17
pTT1 altered codons 16–18	25	44	8
pTT2 altered codons 16–18	ND	51	16

The mRNA half-lives of the different mRNA are given normalized to the value of pMAP217 measured in the same experiment. In pMAP217, 36 fast-translated GAA codons were inserted at codon 927 in *lacZ*. The half-lives of the *lacZ* mRNA in pMAS2 and in pMAP217 are identical [15]. As described in the text, we used three methods to measure the functional half-lives. ND, not determined.

and TT2 are different from each other and different from those in pCARO1 and pSN4 (not shown). The stability of these new *lacZ* variants was measured by one or more of the methods A, B, C as shown in Table 4.

All four codon 16–18 sequences consist of codons expected to be translated slowly [12] and are seen to give mRNAs with similar half-lives. This supports the conclusion that the most important parameter for this mRNA destabilization is the slow translation rate of these three codons, giving an increased distance between ribosomes distal to codon 18 and not by other features of the mRNA sequence.

We have previously developed a mathematical model for the translation process that allows taking variations in the translation speed of the individual codons into account. For instance, the model can calculate the fraction of time; a moving window of, for instance, five codons will not be covered by a ribosome [12]. The results for the wild-type and the pSN4 variant are shown in Fig. 2 a and d, demonstrating that we expect the ribosome occupancy to vary between 40% (wild type) and 25% (pSN4) in the region distal to the slow codons at position 16–18.

The new variants constructed here, pCARO1, pTT1 and pTT2, will have the same modeled ribosome occupancy as pSN4 because we assign identical rates to their codon 16–18 sequences. Table 4 shows that they also have similar half-lives. Figure 2b and c shows that the fast-translated insert in the pSP1779 mRNA is approximately 2-fold less covered with ribosomes compared to the more slowly translated insert in pSP1780. This agrees with our model that mRNA uncovered by ribosomes is more subject to degradation. The region immediately distal from the insert in pSP1780 is seen to be less occupied by ribosomes compared to pSP1779 and to the wild type. The short half-life of the pSP1779 variant indicates that the unoccupied coding region in 1–20 is more important for the stability than later codons on the message.

Finally, Fig. 2e and f shows that the early mRNA region is progressively less covered by ribosomes when we insert 7 and 21 fast-translated codons into codon 13 in pMET29 and pMET31. We therefore conclude that the ribosome occupancy not only in the region codons 20–45 [15] but also and especially in the region 1–20 influences the mRNA stability.

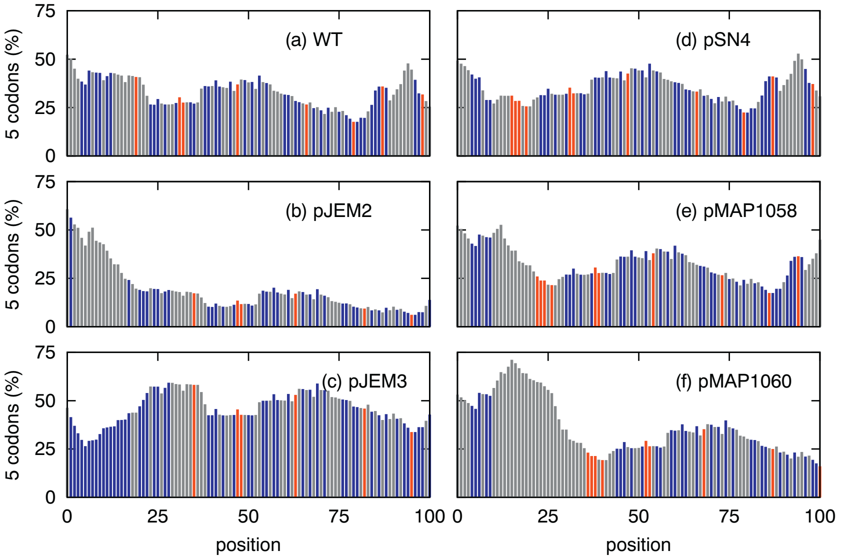

Fig. 2. Ribosome non-occupancy of the first 100 codons for the indicated *lacZ* mRNA variants. The ordinate that gives the fraction of the time a window of 5 codons will be uncovered by a ribosome in the steady state. The approximate translation rates for the codons are indicated: gray, high rate; blue, average rate; and red, low rate [12].

Discussion

This work indicates that the translation rate of individual codons in the early coding region is a major parameter that affects the mRNA half-life and premature transcription termination. Previously, we presented evidence that the translation rate is codon specific. First, it was shown that the *lacI* mRNA and the *bla* mRNA with non-biased codons were translated with approximately half the rate of four ribosomal-type genes, *rpsA, fus, tuf* and *tsf*, in the same experiments [7]. To try to exclude that the rate difference was caused by different mRNA structures, we measured the translation rate on two inserted sequences that were almost identical except for the reading frame that had ribosomal-type biased codons in one and non-biased codons in the other. Again, the biased ribosomal gene reading frame was translated about twice as fast as the non-biased sequence [8], and similarly, a specific codon GAA was found to be translated with three times of the rate of the other codon read by the same tRNA, GAG (22 codons/s *versus* 6.5 codons/s) [9]. In general, the codon preferred in the ribosomal-type genes pairs best with the anti-codon according to Watson–Crick rules, whereas the other codons read by the same tRNA have wobble pairing at the third codon position, indicating that the K_m for tRNA binding to the ribosomal A-site is determining the translation rate. The translation of the CGA arginine codon involves an inosine–adenine pairing at the wobble position and was determined to be one third of the average translation rate [9]. Finally, changing the modifications of the uridine base in the wobble base in the tRNA that reads GAA and GAG had a substantial effect on the translation rate of these codons [22]. The concentration of a tRNA has also been shown to affect the translation rate [6]. These observations make it unlikely that mRNA structures are very important for the determination of the translation velocity.

This does not exclude that other mechanism also might have an impact on the translation rate. Such include proline-rich sequences [23], positive amino acid side-chain residues [24], specific amino acid sequences [25] and mRNA structures (e.g., Refs. [26,27]). In addition, the codon-specific rate of translation has been shown to be involved in protein folding [28]. The influence of mRNA structures on the expression level has been shown experimentally several times (e.g., Refs. [16,17,29]). In these cases, the ribosome accessibility to the mRNA binding site was lowered. We think that the first three, proline-rich sequences, positive side chains and specific amino acid sequences, represent special cases and that eventual mRNA structures are formed mainly before binding of the first ribosome. The codon-dependent translation rate may therefore affect the expression in the time the mRNA is devoid of structures and open for translation.

Also, we developed a mathematical model for the translation process, and this indicated for the first time an interdependence of the translation rate and the mRNA half-life [15]. The mechanism behind this, we suggested, was that different codon-specific translation rates caused local mRNA segments to be covered by ribosomes to a varying degrees, thus exposing the uncovered mRNA segments to degradation. We suggested that the mRNA region from codons 20 to 45 was particularly sensitive to such degradation. After this study, we realized that several of the inserts from the study by Petersen [14] had rare and slowly translated codons late in the insert. These would act as a bottleneck as in our pSN4 variant and make the ribosomes queue before these slow codons. This queue would nullify any difference in codon-specific translation rates before the bottleneck and be the reason why we failed to detect any influence of early codons on the mRNA half-life. We show here, by inserting stretches of rapidly translated codons into codon 3 and into codon 13, that also the early part of the mRNA, if unprotected by ribosomes, is a target for degradation. This is visualized by our mathematical model (Fig. 2).

We unexpectedly found premature transcription termination when we inserted stretches of fast-translated codons into either codon 3 or codon 13 (Table 2). One might have expected that slowly translated codons might make the distance between the RNA polymerase and the first ribosome large enough to activate Rho-dependent termination. Instead, we observed that it was stretches of fast-translated codons that led to transcription termination. The degree of premature transcription termination clearly is increasing with the length of the fast-translated inserts. As mentioned above, long codon GAA sequences late in the gene did not affect mRNA half-life or the premature transcription termination even if it should generate a long stretch of ribosome-free mRNA. We envisage that two mechanisms are at play that affect the expression yield: premature transcription termination and degradation of the mRNA.

We note that following:

1) The CAI increases gradually and reaches its final value for a particular gene around codon 50 [10].
2) At approximately this position, we cease to observe an effect on the mRNA half-life by codon choice, that is, mRNA stretches not covered by ribosomes [15].
3) Stretches of ribosomal-gene-type, fast-translated codons in approximately the first 50 codons will lead to premature transcription termination (data presented here).

4) The NusA and the NusG proteins join the transcribing RNA polymerase gradually after its initiation [30].
5) The parameters that affect the premature transcription termination cease to act after binding of the first ribosome. Fast translating leading ribosomes might come too close to the RNA polymerase and prevent binding of the Nus proteins to the RNA polymerase. This may lead to transcription termination.
6) The parameter that determines the mRNA half-life involves the length of the mRNA between each initiating ribosome (not only the leading ribosome) and the mRNA's 5′ terminal tri-phosphate group because the RppH enzyme has been shown to initiate the mRNA degradation [31]. They also observed an influence of the initial codons on the mRNA half-life, which was also observed in our previous study [32].

Our current summary of the parameters that influence the expression yield is depicted in Fig. 3.

The argumentation that the *lacA* and the *lacZ* mRNA are degraded independently of each other is based on experiments with the wild type. The constructs that show a decreased expression of LacA have short mRNA half-lives, meaning that they are translated by few ribosomes. As an alternative to the model presented in Fig. 3, one might suggest that the mRNA structure that is split before *lacA* is not cleaved in these constructs, possibly because a degradosome that had been tailing the leading ribosome after this terminated at the *lacZ* stop codon might degrade the first part of the structure before the last part was synthesized by the RNA polymerase. If so, the degradosome would follow immediately after the RNA polymerase and give a reduced LacA expression. In this scenario, the reduction of the LacA expression is proportional to the probability to have only one ribosomes translating lacZ mRNA. This probability should vary as 1/(N +1) with the average number, *N*, of translating ribosomes per lacZ mRNA, when normalized to the cases where a mRNA translated once or more. This number *N* can be estimated to be between 30 and 75 in the wild type. Modeling this, based on the values in Table 2, we show that the reduction in the LacA expression already is significant at mRNA half-life at 4 to 10 translations, making this model less likely.

Finally, our ability to calculate the half-lives of the 14- and 21-bp inserts from those of the parental and 7-bp insert is in agreement with and support our suggestion that the specific use of fast or slowly translated codons early in a gene is a “velocity code” within the amino acid coding ability, where the translation rate differences give mRNA stretches of varying lengths that is uncovered by ribosomes, thereby exposing the mRNA to varying rates of degradation.

Methods

Strains and plasmids

The strain used in all experiments were MC1000 *recA1*/F′ *lacI*^{q1} *lacZ*:: Tn*5* constructed by N. Fiil.

All plasmids were derivatives of pBR322, either pMAS2 carrying *lacZ* only [8] or pMET20, which is pMAS2 supplemented with the wild-type *lacY* and *lacA* genes, thus carrying the *lac* wild-type operon [32]. We find that the mRNA half-lives of our *lacZ* variants are the same whether the plasmid coded for the complete *lac* operon or for the *lacZ* part only.

The *lacZ* variants were constructed by recombineering [33] screening for increased *lacZ* expression from a pMAS20 derivative, pMET1, with a stop codon in codon 3 as described [32]. pJEM1 has a *Bst*EII recombinered into pMET1. Our *lac* variants come in two versions, one derived from pMAS2 without the *lacY* and *lacA* and one where the *lacZ* variant was part of the complete *lacZYA* operon. These pairs are as follows: pMAS2, pMET20 (wild type); pJEM2, pSP1779 (14 biased, fast codons inserted into codon 3); pJEM3, pSP1780 (14 non-biased codons inserted into codon 3); pMAP1058,

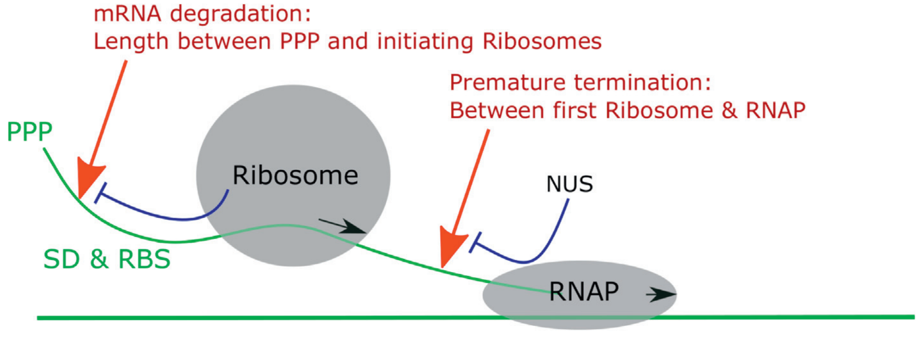

Fig. 3. Schematic overview of the mechanisms affecting the mRNA half-life and the premature transcription termination.

pMET29 (7 fast codons inserted into codon 13 of pSN4, [15]); pMAP1059, pMET30 (14 fast codons inserted into codon 13 of pSN4); and pMAP1060, pMET31 (21 fast codons inserted into codon 13 of pSN4). Table 1 gives the sequences of the relevant early part of these *lacZ* mRNAs.

Enzymatic assays

The LacZ and LacA enzyme activities were measured on essentially as described [34], except that cells were opened by sonic treatment. Columns 2 and 3 in Table 2 give the expression levels as the enzyme activity per cell mass for a fully induced culture for the LacZ and LacA enzymes, respectively.

Measurement of the functional *lacZ* mRNA stability

To measure the functional *lacZ* mRNA stability *in vivo*, we have used three methods.

In the most direct and the only to give the time for the half-life, method A, the *lac* operon was induced for 5 min, whereupon the bacteria were collected on a filter. After two washes and re-suspension in inducer-free medium, the capacity to synthesize β-galactosidase at different time points was followed by pulse labeling and a chase to complete the labeled nascent peptides. We performed such measurements on mixtures of two cultures that each encoded a *lacZ* gene, one serving as an internal reference and with an increased molecular weight that allowed separation by gel electrophoresis from the *lacZ* variant to be measured. Examples of such measurements are given in Fig. 4 and summarized in Table 4.

For mRNA half-lives that are shorter than about 30 s, method A could not be used because of the lengthy filtering, washing and resuspension procedure. To estimate the stability of such unstable mRNA, we used a second method, B. In this, we determined the amount of [^{35}S] methionine-labeled β-galactosidase protein in a pulse-chase experiment during steady-state induction relative to the radioactivity in the RNA polymerase subunits β and β′. These proteins were separated on 7.5% acrylamide one-dimensional PAGE gels. To ensure a reproducible steady state, the cells had been growing exponentially for more than 10 doubling times before induction and labeling. Because all other parameters —cellular growth rate, plasmid copy number, promoter and ribosome binding site sequences—were identical for all the *lacZ* variants we use, the β-galactosidase expression should only depend on differences in half-life or of an eventual premature transcription termination of the mRNA. All our *lacZ* variants have 23 methionine residues and their synthesis rate could be directly estimated from their β-galactosidase expression, relative to that of the two RNA polymerase subunits.

In cases where the expression was further reduced to indicate a half-life of below approximately 10 s, we were even unable to use method B because of an insufficient labeling of the β-galactosidase protein compared to the background labeling. In such cases, we used a third method, C, where one measures the expression of enzymatic activity in “Miller units” [34]. When estimating the mRNA half-lives by method C, one should have a constant or known specific activity of the encoded enzymes. Low expression is difficult to measure. However, we were able to determine the effective half-life with methods B and C for the *lacZ* variants with a medium or wild-type expression level (Table 4). The half-lives found by method C were only slightly lower than those found by method B,

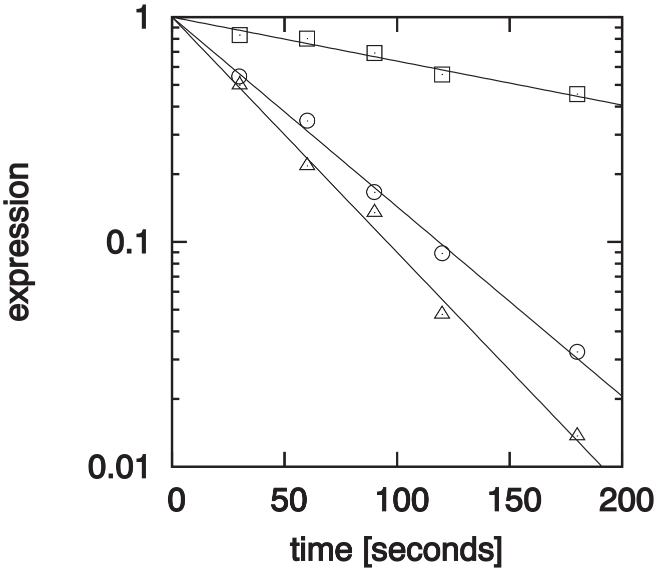

Fig. 4. An example of the determination the functional mRNA half-life by method A for the wild type (squares), pSN4 (circles) and pTT1 (triangles). The figure shows the time course for the expression of the β-galactosidase protein after removal of the inducer.

indicating that there was at most a slight reduction in enzyme-specific activity compared to the wild-type enzyme.

CRediT authorship contribution statement

Steen Pedersen: Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Writing - original draft, Writing - review & editing. **Thilde Bagger Terkelsen:** Data curation, Formal analysis, Investigation. **Mette Eriksen:** Data curation, Formal analysis, Investigation. **Magnus Krarup Hauge:** Data curation, Formal analysis, Investigation. **Casper Carstens Lund:** Data curation, Formal analysis, Investigation.
Kim Sneppen: Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Writing - review & editing. **Namiko Mitarai:** Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Writing - review & editing.

Acknowledgments

We thank June Eyde Madsen, Carolin Runge and Dr. Margit Pedersen for constructing the lacZ variants pJEM1,2,3; and pCARO1; and pMAP 1058, 1059 and 1060. We thank Dr. Stanley Brown and Dr. Michael A. Sørensen of the Department of Biology, University of Copenhagen, for materials and comments, and Tonny D. Hansen for excellent technical assistance. The work was supported by the Lundbeck foundation, the University of Copenhagen MOLPHYSX program, by CMOL and BASP funded by the Danish Research Council.

Received 13 November 2018;
Received in revised form 11 January 2019;
Accepted 16 January 2019
Available online 25 January 2019

Keywords:
translation velocity;
synonymous codons;
mRNA half-life;
premature transcription termination

Abbreviation used:
CAI, codon adaptation index.

References

[1] P.J. Farabaugh, Sequence of the *lacI* gene, Nature 274 (1978) 765–769.
[2] L.E. Post, G.D. Strycharz, M. Nomura, H. Lewis, P.P. Dennis, 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. 76 (1979) 1697–1701.
[3] M. Gouy, C. Gautier, Codon usage in bacteria: correlation with gene expressivity, Nucleic Acids Res. 10 (1982) 7055–7074.
[4] P.M. Sharp, W.H. Li, The codon adaptation index; a measure of directional synonymous codon usage bias, and its potential applications, Nucleic Acids Res. 15 (1987) 1281–1295.
[5] S. Varenne, J. Buc, R. Lloubes, C. Lazdunski, Translation is a non-uniform process: effect of tRNA availability on the rate of elongation of nascent polypeptide chains, J. Mol. Biol. 180 (1984) 549–576.
[6] R. Misra, P. Reeves, Intermediates in the synthesis of TolC protein include an incomplete peptide stalled at a rare Arg codon, FEBS 152 (1985) 151–155.
[7] S. Pedersen, *Escherichia coli* ribosomes translate *in vivo* with variable rate, EMBO J. 3 (1984) 2895–2898.
[8] M.A. Sørensen, C.G. Kurland, S. Pedersen, Codon usage determines translation rate in *Escherichia coli*, J. Mol. Biol. 207 (1989) 365–377.
[9] M.A. Sørensen, S. Pedersen, The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate, J. Mol. Biol. 222 (1991) 265–280.
[10] M. Bulmer, Codon usage and intragenic position, J. Theor. Biol. 133 (1988) 67–71.
[11] A. Eyre-Walker, M. Bulmer, Reduced synonymous substitution rate at the start of enterobacterial genes, Nucleic Acids Res. 21 (1993) 4599–4603.
[12] N. Mitarai, K. Sneppen, S. Pedersen, Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization, J. Mol. Biol. 382 (2008) 236–245.
[13] T. Tuller, A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman, Y. Pilpel, An evolutionarily conserved mechanism for controlling the efficiency of protein translation, Cell 141 (2010) 344–354.
[14] C. Petersen, The functional stability of the *lacZ* transcript is sensitive towards sequence alterations immediately downstream of the ribosome binding site, Mol. Gen. Genet. 209 (1987) 179–187.
[15] M. Pedersen, S. Nissen, N. Mitarai, S.L. Svenningsen, K. Sneppen, S. Pedersen, The functional half-life of an mRNA depends on the ribosome spacing in an early coding region, J. Mol. Biol. 407 (2011) 35–44.
[16] M.H. de Smit, J. van Duin, Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis, Proc. Natl. Acad. Sci. U. S. A. 87 (1990) 7668–7672.
[17] M.N. Hall, J. Gabay, M. Débarbouillé, M. Schwartz, A role for mRNA secondary structure in the control of translation initiation, Nature 295 (1982) 616–618.
[18] V.J. Cannistraro, M.N. Subbarao, D. Kennell, Specific endonucleolytic cleavage sites for decay of *Escherichia coli* mRNA, J. Mol. Biol. 192 (1986) 257–274.
[19] G.J. Murakawa, C. Kwan, J. Yamashita, D.P. Nierlich, Transcription and decay of the *lac* messenger: role of an intergenic terminator, J. Bacteriol. 173 (1991) 28–36.
[20] W.A. Newton, J.R. Beckwith, D. Zipser, S. Brenner, Nonsense mutants and polarity in the Lac operon of *Escherichia coli*, J. Mol. Biol. 14 (1965) 290–296.
[21] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, Nucleic Acids Res. 31 (2003) 3406–3415.

[22] M.K. Krüger, S. Pedersen, T.G. Hagervall, M.A. Sørensen, The modification of the Wobble Base of $tRNA_{Glu}$ modulates the translation rate of glutamic acid codons in vivo, J. Mol. Biol. 284 (1998) 621–631.

[23] M.Y. Pavlov, R.E. Watts, Z. Tan, V.W. Cornish, M. Ehrenberg, A.C. Forster, Slow peptide bond formation by proline and other *N*-alkylamino acids in translation, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 50–54.

[24] J. Lu, C. Deutsch, Electrostatics in the ribosomal tunnel modulate chain elongation rates, J. Mol. Biol. 384 (2008) 73–86.

[25] K. Ito, S. Chiba, K. Poglianoc, Divergent stalling sequences sense and control cellular physiology, Biochem. Biophys. Res. Commun. 393 (2010) 1–5.

[26] S. Bhattacharyya, W.M. Jacobs, B.V. Adkar, J. Yan, W. Zhang, E.I. Shakhnovich, Accessibility of the Shine–Dalgarno sequence dictates N-terminal codon bias in *E. coli*, Mol. Cell 70 (2018) 894–905.

[27] C. Chen, H. Zhang, S.L. Broitman, M. Reiche, I. Farrell, B.S. Cooperman, Y.E. Goldman, Dynamics of translation by single ribosomes through mRNA secondary structures, Nat. Struct. Mol. Biol. 20 (2013) 582–588.

[28] C.-H. Yu, Y. Dang, Z. Zhou, C. Wu, F. Zhao, M.S. Sachs, Y. Liu, Codon usage influences the local rate of translation elongation to regulate co-translational protein folding, Mol. Cell 59 (2015) 744–754.

[29] K.I. Sørensen, K.E. Baker, R.A. Kelln, J. Neuhard, Nucleotide pool-sensitive selection of the transcriptional start site *in vivo* at the *Salmonella typhimurium pyrC* and *pyrD* promoters, J. Bacteriol. 175 (1993) 4137–4144.

[30] R.A. Mooney, S.E. Davis, J.M. Peters, J.L. Rowland, A.Z. Ansari, R. Landick, Regulator trafficking on bacterial transcription units *in vivo*, Mol. Cell 33 (2009) 97–108.

[31] J. Richards, D.J. Luciano, J.G. Belasco, Influence of translation on RppH-dependent mRNA degradation in *Escherichia coli*, Mol. Microbiol. 86 (2012) 1063–1072.

[32] M. Eriksen, K. Sneppen, S. Pedersen, N. Mitarai, Occlusion of the ribosome binding site connects the translational initiation frequency, mRNA stability and premature transcription termination, Front. Microbiol. 8 (2017) 362–368.

[33] J.A. Sawitzke, L.C. Thomason, N. Costantino, M. Bubunenko, S. Datta, D.L. Court, Recombineering: *in vivo* genetic engineering in *E. coli*, *S. enterica*, and beyond, Methods Enzymol. 421 (2007) 171–199.

[34] J.H. Miller, Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, New York, NY, 1972.