DNA Research, 2018, 25(2), 195–205
doi: 10.1093/dnares/dsx049
Advance Access Publication Date: 17 November 2017
Full Paper

OXFORD

Full Paper

Genome scale analysis of *Escherichia coli* with a comprehensive prokaryotic sequence-based biophysical model of translation initiation and elongation

Gilad Shaham[1] and Tamir Tuller[1,2,*]

[1]Department of Biomedical Engineering, The Engineering Faculty, Tel Aviv University, Israel and [2]The Sagol School of Neuroscience, Tel-Aviv University, Tel-Aviv, Israel

*To whom correspondence should be addressed. Tel. +972 3 6405839. Fax. +972 3 6407308. Email: tamirtul@post.tau.ac.il

Edited by Dr. Mikita Suyama

Received 15 March 2017; Editorial decision 2 November 2017; Accepted 4 November 2017

Abstract

Translation initiation in prokaryotes is affected by the mRNA folding and interaction of the ribosome binding site with the ribosomal RNA. The elongation rate is affected, among other factors, by the local biophysical properties of the coding regions, the decoding rates of different codons, and the interactions among ribosomes. Currently, there is no comprehensive biophysical model of translation that enables the prediction of mRNA translation dynamics based only on the transcript sequence and while considering all of these fundamental aspects of translation. In this study, we provide, for the first time, a computational simulative biophysical model of both translation initiation and elongation with all aspects mentioned above. We demonstrate our model performance and advantages focusing on *Escherichia coli* genes. We further show that the model enables prediction of translation rate, protein levels, and ribosome densities. In addition, our model enables quantifying the effect of silent mutations on translation rate in different parts of the transcript, the relative effect of mutations on translation initiation and elongation, and the effect of mutations on ribosome traffic jams. Thus, unlike previous models, the proposed one provides comprehensive information, facilitating future research in disciplines such as molecular evolution, synthetic biology, and functional genomics. A toolkit to estimate translation dynamics of transcripts is available at: https://www.cs.tau.ac.il/~tamirtul/transim

Key words: translation, biophysical model, transcript evolution, UTR, mRNA folding

1. Introduction

All living cells share the ability to translate mRNA transcripts into proteins. The codes related to this process partially appear in different regions of the transcript including the UTRs and the coding region. For example, it was shown that the mRNA folding at the 5′UTR and the 5′end of the coding region regulates translation initiation efficiency.[1,2] In prokaryotes, translation efficiency is related to the hybridization of the ribosomal RNA and the ribosome biding sites upstream of the start codon.[3] It was also shown that different codons have different decoding rate; thus, codon usage biase regulate translation elongation.[4–6] In addition, it was shown that there is high concentration of functional silent codes related to various gene expression steps at the 5′end of the coding regions.[7]

© The Author 2017. Published by Oxford University Press on behalf of Kazusa DNA Research Institute.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Thus, having the ability to harness the nucleotide composition of the transcript to the biophysical dynamics of mRNA translation should have fundamental contribution to disciplines such as molecular evolution, functional genomics, synthetic biology, biotechnology, and human health.[8–13]

Yet, similar to other cellular processes, despite decades of study and various technological advancements, it is still very challenging to predict the dynamics of mRNA translation based only on the transcript nucleotide composition. Specifically, today there are biophysical predictive models that consider mostly translation elongation[14–18] or translation initiation.[19–22] In addition, there are machine learning based models that statistically integrate various features of the 5′UTR and the coding region to predict gene expression measurements such as protein levels.[23–25] However, there is no complete user-friendly model/toolkit, which incorporates both translation initiation and elongation dynamics, based only on the transcript nucleotide composition.

In this study, we make an additional step towards developing a computational biophysical model that predicts the dynamics of translation based solely on the transcript nucleotide composition. As we show, the model is specifically important as it enables, for the first time, to study the relation between transcript mutations and the biophysics/dynamics of translation. Translation is a fundamental intracellular process directly related to organismal fitness; thus, we provide an approach that should promote improved and novel understanding of transcript evolution at a nucleotide resolution. We specifically demonstrate how our new model can provide important, and previously unpredicted, information.

2. Materials and methods

The research scheme is shown in Fig. 1: Using the mRNA sequence, the mRNA folding and mRNA-rRNA interactions are calculated, resulting in the relevant Gibbs free energy predicted values (A). These free energy terms are described in the translation initiation efficiency calculator subsection and are then used to predict the initiation rates (B); which is the rate by which ribosomes approach the start codon. The codon translation elongation rates (C) are calculated using ribosome profiling data; each codon has its own elongation rate and when we later refer to the elongation rate of a gene, we consider the elongation rates of all its codons and the dynamic of the translation process (see Section 5). Rather than using a mean field approximation of the model totally asymmetric exclusion process (TASEP)[16] such as the ribosome flow model,[14] we decided to directly use the stochastic 'high resolution' TASEP model to simulate translation using the estimated initiation rates and elongation rates (D). TASEP provides the predictions of translation rate (which is the overall translation rate when considering both the initiation and the elongation steps), ribosomal density, termination counts and occurrences of ribosome jamming (E). Finally, we study various novel questions related to the relations between nucleotide composition of the transcript and translation aspects (F).

3. Translation initiation efficiency calculator

The translation initiation calculator is based on the ribosome binding site (RBS) mRNA folding calculator.[19,21] In general, to compute the nominal initiation rate, we implemented a thermodynamic model of bacterial initiation, which predicts the Gibbs free energy of ribosome binding. The input to the initiation rate calculator is the mRNA sequence and its output is the initiation rate, predicted on a proportional scale span over 8 orders of magnitude (0.001 to 100, 000+ au). The thermodynamic model calculates the difference in Gibbs free energy before and after the 30 S complex assembles onto an mRNA transcript. Five free energy terms are calculated and summed together:

$$\Delta G_{tot} = \Delta G_{\text{mRNA:rRNA}} + \Delta G_{\text{start}} + \Delta G_{\text{spacing}} - \Delta G_{\text{standby}} - \Delta G_{\text{mRNA,}}$$

where

- ΔG_{mRNA}—The folding energy of the mRNA subsequence prior to binding with the 30S complex.
- $\Delta G_{\text{mRNA:rRNA}}$— The energy released when the last nine nucleotides of the 16S rRNA cofolds and hybridizes with the mRNA sub-sequence at the 16S rRNA-binding site.
- ΔG_{start}—The energy released when the tRNAfMet's anticodon hybridizes to the start codon.
- $\Delta G_{\text{standby}}$— The energy released when the standby site is folded.
- $\Delta G_{\text{spacing}}$—An energetic penalty for a non-optimal distance between the 16S rRNA-binding site and the start codon.

These free energy terms are illustrated in Fig. 1. See the Supplementary Text for more details regarding these energy terms. We used the Vienna RNA package[26] version 2.2.7 to perform the necessary folding and resultant free energy calculations.

4. Normalizing initiation rate

The initiation rate of each transcript is calculated using the following exponential formula:

$$K\exp(-\beta\Delta G_{\text{tot}}),$$

where $\beta = 0.45$ mol/kcal and $K = 2500$. This initiation rate is predicted on a proportional scale from 0.001 to 100, 000+ au; we then normalize it to initiation time measured in seconds. We took the list of 302 essential genes from the *Escherichia coli* chromosome (PEC) database[27,28] and we calculated the average initiation rate of 709.773 ([au] units which are proportional to rate) for 138 genes in our list that are classified as essential genes. We then took the mean initiation time of 178 out of the 302 essential genes available at the Transimulation Protein Biosynthesis Server,[29,30] with an average 12.179 s. Hence, we multiply the two values to obtain a factor of approximately $8,\ 644\ [\text{au}\cdot\text{s}]$ and divide each individual transcript initiation rate by this factor in order to obtain the initiation rate in $[\text{s}^{-1}]$ units.

5. Modelling translation elongation

The elongation modeling was based on a TASEP.[14,16] The TASEP is a stochastic flow model of translation elongation, whose output is the predicted translation rate, the ribosomal density and the number of termination events. An mRNA transcript with N codons is modeled as a chain of sites, each of which is labeled by the index i, where $i = 1 \ldots N$. The first and last codons $i = 1$ and $i = N$, are associated with the start and stop codons, respectively. At any time t, $M(t)$ ribosomes are bound to the mRNA.

The initiation time as well as the time a ribosome spends translating each codon are exponentially distributed with a codon dependent rate. In addition, ribosomes span over l codons and if two ribosomes are adjacent, the trailing one is delayed until the ribosome in front of it has proceeded onwards. Unless otherwise specified, all the reported results in this paper use $l = 11$. A free ribosome will attach to codon $i = 1$ with mean rate λ, if the first $(l+1)/2$ codons on the mRNA are empty. An attached ribosome located at codon i will move to the

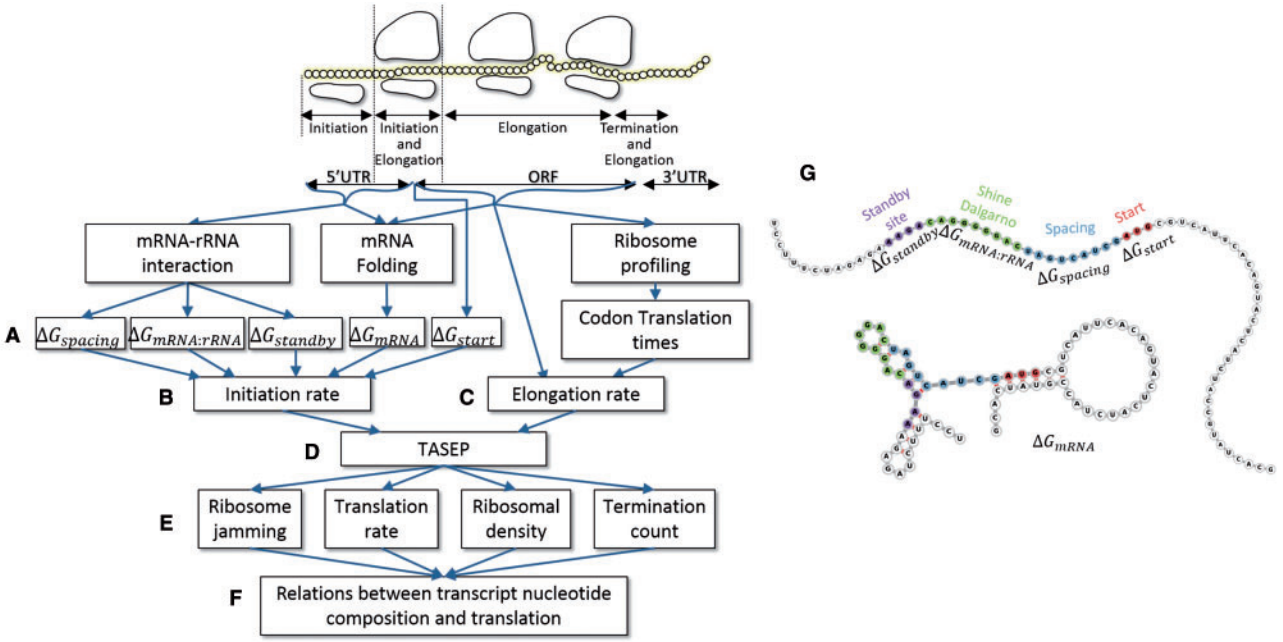

Figure 1. (A) Based on the 5′UTR and the beginning of the ORF, the Gibbs free energy is calculated. (B) Summing together the free energies results with the expected initiation rate. (C) Using codon translation times and the mRNA ORF, the per-codon elongation rate is calculated. (D) TASEP calculation is performed using the initiation rates and elongation rates from the previous steps. (E) TASEP results with translation rate, ribosomal density, termination count and ribosome jamming. (F) We study the relation between transcript nucleotide composition and translation based on the simulation results. (G) Illustration, as described by the RBS calculator where ΔG_{mRNA} is the folding energy. The location of $\Delta G_{standby}$ is depicted in purple, $\Delta G_{mRNA:rRNA}$ is in green, $\Delta G_{spacing}$ is in blue and ΔG_{start} is in red.

next codon $i+1$ with mean rate λ_i, provided codon $i+(l+1)/2$ is not covered by another ribosome. In case $i+(l+1)/2 > N$ (ribosome is bulging out of the mRNA strand) an attached ribosome will move to the next codon with mean rate λ_i. In the case of $i = N$ the jump attempt is in fact a termination step.

6. TASEP elongation rate

We obtained the per-codon *E. coli* typical decoding time (TDR).[31] We calculated the mean TDR of the codons by multiplying each codon TDR by the empirical probability of the codon to appear in the entire sequence data (sequence data source is described in *E. coli* Sequence Data subsection below). We then normalized the mean TDR to be 8 codons per second.[32]

We estimate the elongation rate for each transcript by calculating TASEP with a constant initiation rate. We chose the mean normalized initiation rate (0.2884) to be this constant.

7. TASEP traffic jams

A ribosome can only move if the ribosome ahead of it does not cover the next codon. During the simulation, we choose to either attempt initiation or, based on the existing ribosomes, perform translation and move a ribosome to the next position. In cases where such movement cannot occur, we consider this event as a traffic jam. If the simulation attempts initiation, but cannot, since the first $(l+1)/2$ codons are not empty, we count this as 'initiation jam occurrence' and we measure the total time difference between the failed initiation attempt and the next time an initiation is successfully performed as 'initiation jam time'. If the simulation attempts to move a ribosome to site i, but cannot—since the next $(l+1)/2$ codons are not empty, we count this as 'jam occurrence'. We then ensure that a ribosome would move at the same time the subsequent ribosome clears the next $(l+1)/2$ codon sites, i.e. when it is possible for ribosome to proceed. We also measure the mean time the delayed ribosome had to wait as 'mean jam time'.

8. Data

8.1. *Escherichia coli* sequence data

We took *E. coli* str. K-12 substr. MG1655 mRNA sequences, including 5′UTR and ORF annotations from RegulonDB version 9.1 (http://regulondb.ccg.unam.mx).[33] The two files, UTR_5_3_sequence.txt and Gene_sequence.txt, were downloaded directly from the Downloadable Experimental Datasets web page on 31 July 2016.

For our analysis of the initiation and translation rates, we chose either operons with a single gene or the first gene in operons that contain multiple genes. Overall, we found 1,500 genes with 2,035 unique 5′UTR and ORF sequences. For some of these transcripts, we could not calculate initiation rate (either the start codon is not ATG, CTG, GTG, or TTG or we did not find a suitable ribosome-binding site); after eliminating those, we were left with 1,904 transcripts. Finally, the initiation calculation sometimes reaches non-feasible initiation time (e.g. millions seconds); thus, to remove extreme outliers we discarded the transcripts with top 50 initiation times from our analysis. Out of the transcripts left, we took only those whose initiation time was lower or equal to 5 times standard deviation away from the initiation time median. Thus, we were left with 1,835 transcripts. Additional details regarding the transcripts discarded from the analysis are available in the supplementary material and Supplementary Table S2.

Unless otherwise specified, the results in this paper are calculated per transcript. When comparing the model results with per-gene expression levels, we used the mean results across the different transcripts for the same gene, prior to calculating the correlation.

8.2. *Escherichia coli* protein abundance

The *E. coli* str. K-12 substr. MG1655 protein abundance (PA) integrated dataset, a weighted average of all *E. coli* datasets, published in 2015, was downloaded from PaxDB,[34] accessed on 7 May 2016. We excluded any gene with zero reads. Note that similar conclusions were obtained when we analyzed different databases separately (refer to supplementary material for more details).

8.3. *Escherichia coli* mean mRNA levels

The mean mRNA levels were obtained from Taniguchi *et al.*[35] Data include mean mRNA levels and mRNA levels per cell cycle.

8.4. *Escherichia coli* protein per mRNA levels

We used the protein abundance, described above and divided it by the corresponding mean mRNA levels, described in the previous subsection to calculate the protein abundance per mean mRNA.

8.5. *Escherichia coli* ribo-seq

Two ribosomal profiling datasets were obtained from the Gene Expression Omnibus: GSE35641, *E. coli* MG1655 rep1-2.[36] Transcript sequences were obtained from EnsEMBL for *E. coli* (K-12 MG1655 release 121, accessed 28 July 2015). We trimmed 3′ adaptors from the reads using Cutadapt[37] version 1.8.3, and utilized Bowtie[38] version 1.1.1 to map them to the *E. coli* transcriptome. In the first phase, we discarded reads that mapped to rRNA and tRNA sequences with Bowtie parameters '-n 2-seedlen 21-k 1–norc'. In the second phase, we mapped the remaining reads to the transcriptome with Bowtie parameters '-v 2 -a –strata –best –norc -m 200'. We filtered out reads longer than 34 nt and shorter than 26 nt. Unique alignments were first assigned to the ribosome occupancy profiles. For reads that are mapped to more than one position, the best alignments in terms of number of mismatches were kept. Then, multiple aligned reads were distributed between locations according to the distribution of unique ribosomal reads in the respective surrounding regions. To this end, for each position i, a 100 nt window was used to compute the read count density RCD_i (total read counts in the window divided by length, based on unique reads) in vicinity of the M multiple aligned positions in the transcriptome, and the fraction of a read assigned to each position was $\mathrm{RCD}_i / \sum_{j=1}^{M} \mathrm{RCD}_j$.

8.6. Functional annotation

We performed functional annotation of specific groups of genes using DAVID bioinformatics resources version 6.7 (https://david.ncifcrf.gov/).[39,40] We set the background list to be the genes in the selected dataset, as described in the *E. coli* Sequence Data subsection.

8.7. Simulating single-nucleotide mutations

For each transcript in our selected dataset, we simulate a single nucleotide mutation in the range of -35 to -1 and $+3$ to $+34$ (where 0 is the first nucleotide in the ORF). We excluded the start codon change, as most mutations would result with an invalid start codon. Furthermore, mutations that introduced premature STOP codons were not allowed.

8.8. Correction for multiple comparisons

The enrichment analysis P-values were for multiple comparisons using Bonferroni correction. All the correlation P-values pass Benjamini and Hochberg correction for false discovery rate.

9. Results

9.1. A toolkit to estimate translation dynamics of transcripts

We created a software tool that enables the estimation of the initiation rate and translation rate, as well as the sequence features affecting those rates, such as folding energies, ribosome binding efficiencies, and elongation rate. The tool is executed via command line interface, where one should provide either a single mRNA sequence or a list of different mRNA sequences and execution parameters via an input file.

Special attention was given to optimize the code, thus allowing almost immediate results for a single mRNA calculation and performing large-scale calculations within reasonable computational time.

We named the toolkit Translation Simulator, or Transim, and it is available to download for academic use at the authors' website—http://www.cs.tau.ac.il/~tamirtul/transim.

9.2. Higher correlation of predicted translation rate with experimental measurements than the correlation of predicted initiation rate with experimental measurements in endogenous genes

As a first step, we aimed at checking how well our biophysical model, and different components of the model, can predict measured (i.e. 'true') gene expression levels. Today, there are no *direct* large-scale measurements of translation (measurements of protein levels or mRNA levels, e.g. are related to the transcription step, and the mRNA degradation step; protein levels are related also to the protein degradation step). Thus, we used the measure most correlated/related to translation rate that is available today, protein abundance (PA) divided by mRNA levels which is the estimation of the number of proteins generated per mRNA molecule.

We calculated Spearman correlation of protein abundance (PA) divided by mRNA levels (protein per mRNA) with the calculated initiation rates, elongation rates and translation rates. As can be seen in Fig. 2, the correlation with the translation rate was higher and more significant than that with the initiation rate and elongation rate alone. In addition, the partial correlation of translation with PA divided by mRNA given the initiation or the elongation is significant: $R(\text{translation, PA/mRNA} \mid \text{initiation}) = 0.265$, $(P = 4.23 \times 10^{-7})$, $R(\text{translation, PA/mRNA} \mid \text{elongation}) = 0.1932$ $(P = 2.55 \times 10^{-4})$. These results support the usage of the complete model (which includes both the initiation and the elongation steps).

In the future, it will be interesting to further validate our model with direct experimental measurements of translation.[41]

9.3. About the distributions of initiation rates, elongation rates, and translation rates

It is worth noting that the distribution of the translation rate is different from the initiation rate and elongation rate, as seen in Fig. 3 and Supplementary Fig. S11. Specifically, all distributions are not normal (KStest P-value $< 10^{-10}$ for initiation and translation and P-value is $5.86 \times \cdot 10^{-6}$ for elongation) with positive skewness of 6 and 6.6 for initiation and translation, respectively due to right tails, and small negative skewness (-1.28) for elongation. The positive skewness of the initiation and translation may be related to traffic jams and/or extremely low initiations rate in some genes at the initiation step (extreme and unexpected delays in the initiation times contribute to very high and non-typical initiation times, resulting in the right tail).[18]

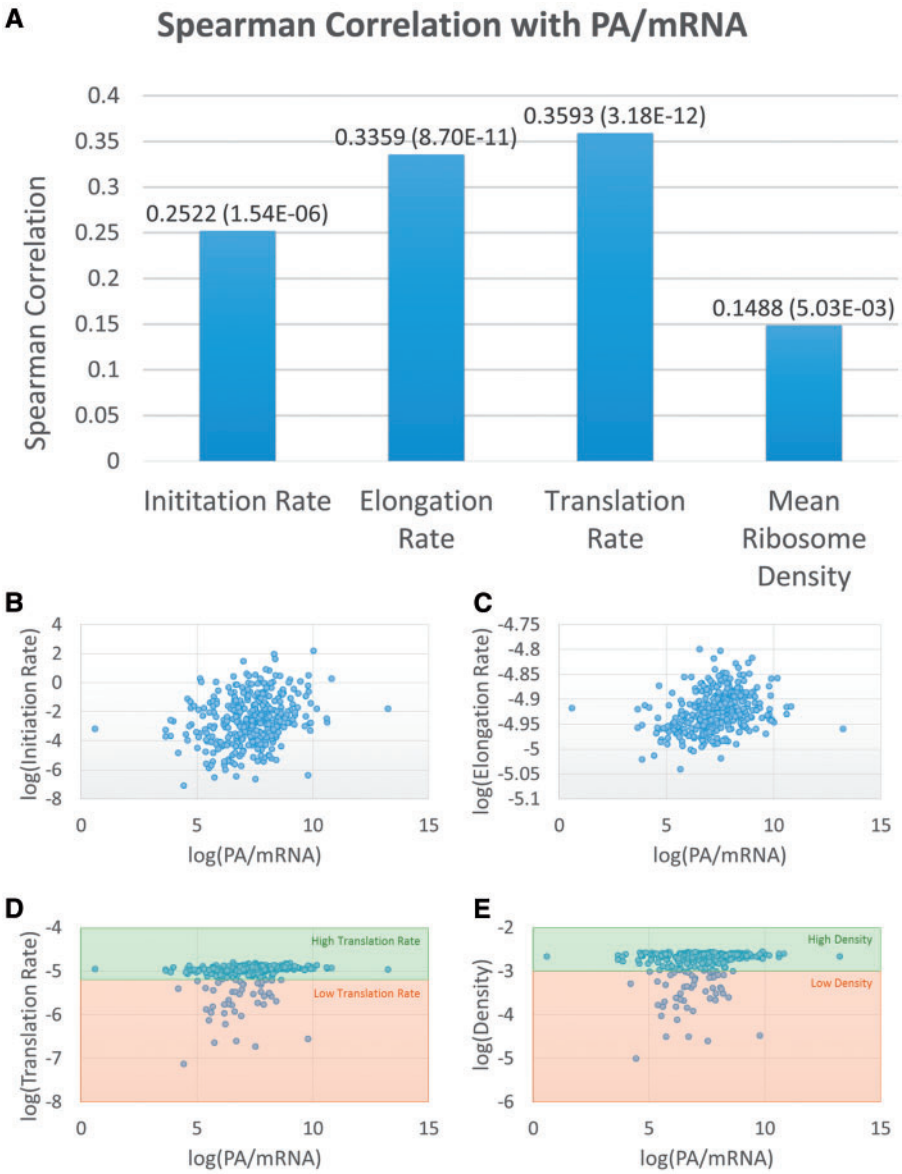

Figure 2. (A) Spearman correlation of initiation rate, elongation and translation rate with protein abundance (PA) per mRNA levels. The values in parenthesis indicate the *P*-value. (B)–(E) Scatter plots for each of the cases. In the case of the complete translation model there are two 'regions' in the figure: one related to low initiation rate where the initiation rate is low and rate limiting and the density is low; the second region related to higher initiation rate where the elongation rate becomes more rate limiting and the density is higher.

As can be seen (Fig. 3), the mean initiation time per transcript is 54.4 s while the mean elongation time is 138.1 s and the mean translation time is 174.2 s. This is expected, as translation includes both initiation and elongation. While there are cases with very low initiation time (the minimum is 0.04 s), the minimum elongation and translation times are 96.3 and 98.6 s, respectively. In the cases where initiation rates are very fast, we would expect elongation to be the rate-limiting factor. We also see the maximum initiation time found is 1,629 s, while the maximum translation time is 1,739.7 s. The maximum elongation time is 155.3 s, but 80% of the initiation rates are below 52.2 s and 80% of the translation rates are below 155.7 s and we expect the initiation to be rate-limiting factor when it is slow.

The fact that the typical initiation rate is significantly lower than the typical translation rate demonstrates that, in practice, not only the initiation step affects translation rate but that also the elongation rate tcontributes to the final translation rate;[7,42] this insight provides a better understanding to previous models in the field.[43]

While translation rate can be significantly affected by the initiation rate, fast initiation rates (i.e. low initiation time), can increase the translation rate, but up to a point. In these cases, the elongation rate becomes a limiting factor, effectively creating a maximum threshold for the translation rate. This limit in translation rate may be important for biological processes, such as prevention of depletion of the ribosome pool. Hence, genes that have high and efficient initiation rate may have codons that partially limit their overall translation rate and ribosome consumption.[7] Limiting the translation elongation rate may also assist in allowing the resultant protein to cotranslationally fold correctly.[44–48]

The results reported here suggest that changes in the codon composition of a gene should have significant effect on translation, even if the

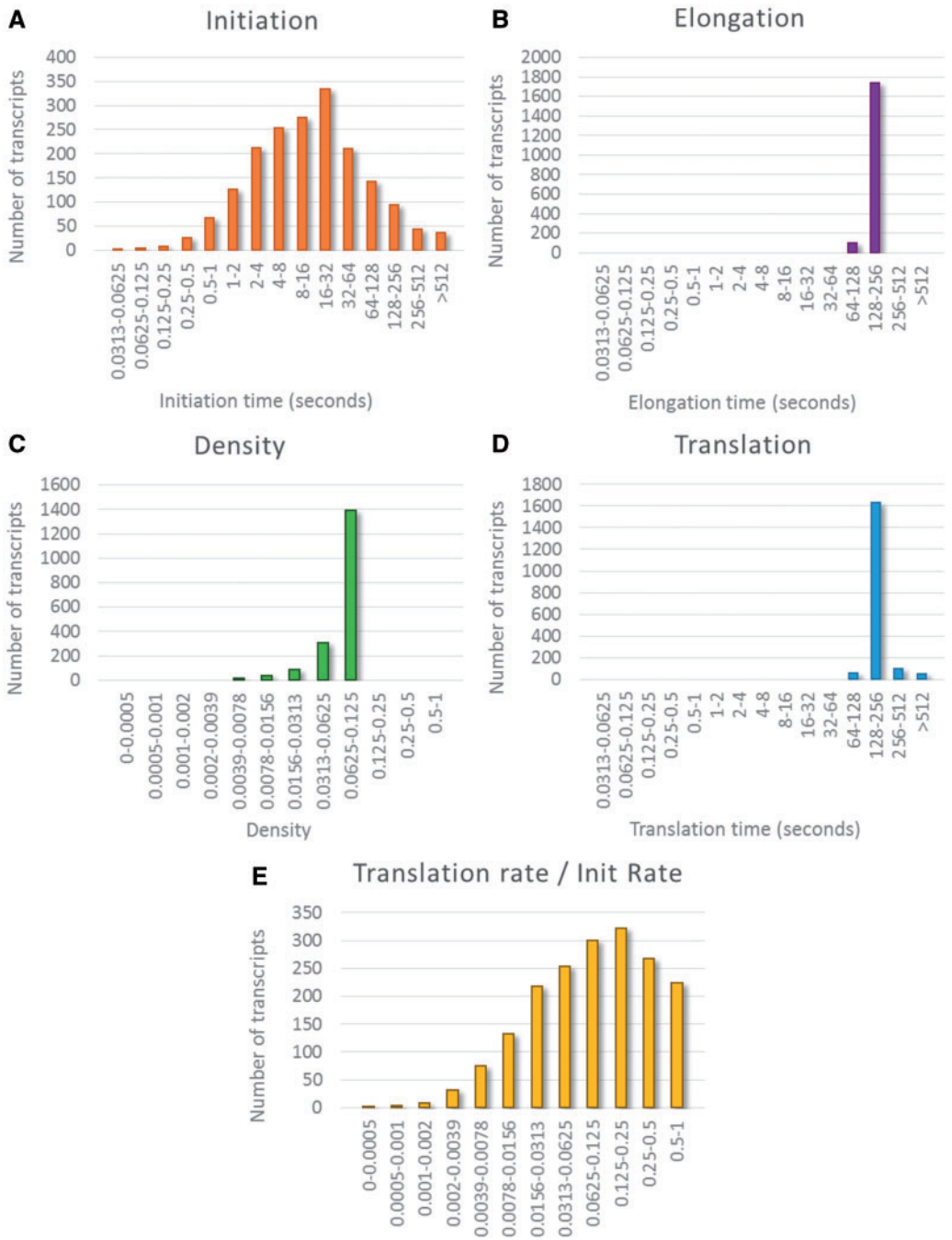

Figure 3. The distribution of various translation variables: initiation rates (A), elongation rates (B), ribosome densities (C), translation rates (D), translation rate to initiation rate ratio (E). Note that the bin sizes are not uniform, but grow exponentially.

initiation rate is optimal/maximal and this can be further validated experimentally via the design of relevant libraries of gene variants and the measurements of translation of the different variants.

9.4. Only specific locations of single nucleotide mutation affect initiation and translation rates

At the next step, we aimed at understanding the effect of mutations in different regions surrounding the start codon on translation rate in endogenous transcripts. This was done under the assumption that the expression levels of each transcript is not extremely high such that the mutations should not affect the translation rate of other transcripts.[49] Thus, for each transcript in our dataset and each nucleotide in the range of −35 to −1 and +3 to +34 (where 0 is the first nucleotide of the ORF), we changed the actual nucleotide to all three other possible nucleotides. We then calculated the new initiation rate and translation rate per each change. For each position, we calculated the mean result and compared it to the mean of the unmodified transcript.

Results are available in Fig. 4. On average, there is a clear reduction in initiation rate when mutations occur at the RBS site (which is generally 5–13 nucleotides upstream of the start codon, with a typical distance of about 8–10 bases relative to the start codon),[50–52]

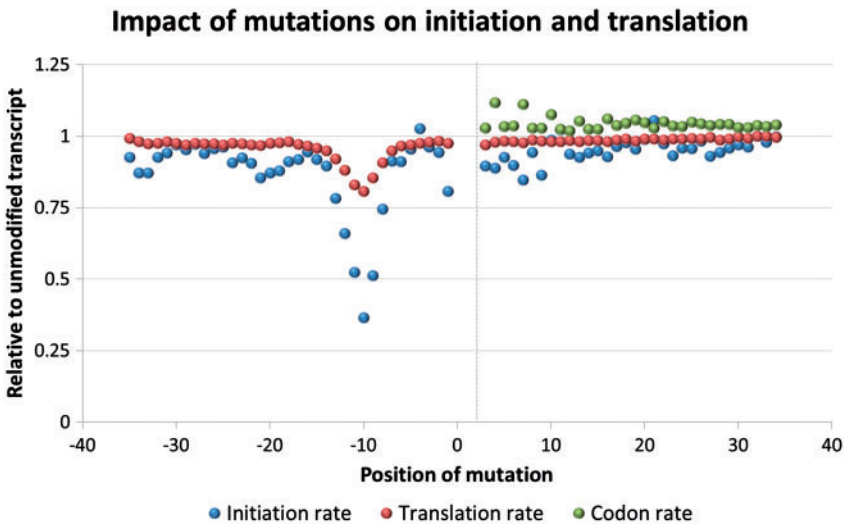

Figure 4. Impact of single single nucleotide mutation on the mean initiation rate, the mean change in codon decoding rate for the position where the mutation occurs, and the mean translation rate.

although the translation rate is less affected. Apart from the RBS site, changes in the first nucleotides of the ORF also tend to slightly decrease mean initiation rate. When we examine this in more detail (Supplementary Figs S1–S4), while we still see the same effect in the RBS site, we also see that for translation, only a small group of transcripts had an increase in translation rate, and some had a decrease, but for most transcripts, no change occurred in translation rate. For initiation, we also see several groups, and a large portion remains unchanged. These results demonstrate that translation rate tend to be immune to single point mutations in the transcript. We also see slight increase in codon rate in that region due to the mutation, this occurs since codons are usually 'less optimal' (in terms of their translation rate) in the beginning of the ORF.[7]

The fact that the most prominent effect is seen in the RBS site is in agreement with previous studies, which have demonstrated that mutations in the Shine-Delgarno regions have the strongest effect on translation rate.[53,54] This could be further validated in large-scale experiments in the future.

We further examined the impact of mutation type on the initiation rate (Supplementary Fig. S10). The type of mutation tends to have significant position-based effect most prominent in the RBS: G to C and G to T seem to decrease the initiation rate while the inverse mutation C to G and T to G have the opposite effect and tend to increase the initiation rate in that region. This inverse effect does not always exist, for example, G to A tends to reduce the initiation rate, but A to G tends to have an almost neutral effect. It is likely these changes in the initiation rate are due to changes in the ribosome binding to the mRNA, since the mutation can change the binding site to be closer or farther from the Shine-Delgarno sequence.

The result reported here provides an estimation of different type of mutations in different regions surrounding the start codon on translation. This estimation can be further validated experimentally via the design of relevant libraries of gene variants and the measurements of translation of the different variants.

9.5. Correlation between initiation rate, elongation rate, ribosome density, and translation rate

Next, we aimed at understanding the relations between initiation rate, elongation rate, ribosome density and translation rate in endogenous transcripts. Thus, we calculated the Spearman correlation between these values. The results are shown in Table 1.

We can see weak correlation between elongation rate and initiation rate and strong relationship between initiation rate and ribosome density. We also find that translation rate is highly correlated with all three other measurements. Please note that mathematically there can be cases where some of these correlations do not exist (refer to the supplementary material for more details). In endogenous/real sequences, a correlation can be explained by the fact that the translation process depends on ribosome initiation and elongation; in addition, sequences with high translation rates will require more ribosomes in average (due to higher initiation rate). The magnitude of the correlation, provided here for the first time, allows estimation of the relative 'contribution' of initiation and elongation to translation rate variability in endogenous/real transcripts. In general, when increasing the initiation rate while keeping the elongation rate constant, we expect to see an increase in ribosome density (as more ribosomes are 'pumped' into the coding region). Higher elongation rate, for a constant initiation rate, is expected to decrease the ribosome density (as it corresponds to faster removal of ribosomes from the coding region); indeed the partial Spearman correlation between translation elongation and ribosome density when controlling for the initiation rate $r(\text{elongation},\ \text{density}|\text{initiation})$ is negative ($r = -0.044$; P-value is borderline significant 0.0597). Increase in initiation and elongation rates (when they are rate limiting) is expected to correspond to an increase in the translation rate. Our analysis demonstrates that genes that tend to have increased/higher initiation rate tend to also have higher elongation rate; the increase in initiation rate in highly expressed genes is more 'dominant' in terms of the effect on ribosome density as the correlation between translation rate/protein levels and ribosome density is positive. This supports the approach of various studies that measure translation efficiency via measurements that are related to ribosome density.[55]

It should be interesting to study in the future other organisms to see if the positive relation between initiation and elongation rate vs. ribosome density is universal.

9.6. Traffic jams study in different gene groups and the way they are effected by mutations

One advantage of our model is the fact that it provides estimations related to variables such as ribosome collisions/traffic jams that

Table 1. Spearman correlation between initiation rate, elongation rate, ribosome density, and translation rate

	Elongation rate	Ribosome density	Translation rate
Initiation rate	0.1137 (1.034×10^{-6})	0.7822 ($<2.2 \times 10^{-308}$)	0.6487 (1.125×10^{-219})
Elongation rate	—	0.1155 (6.984×10^{-7})	0.6454 (1.046×10^{-216})
Ribosome density	—	—	0.6514 ($<2.2 \times 10^{-308}$)

The values in parenthesis indicate the respective *P*-value (all *P*-values pass correction for false discovery rate).

cannot be accurately predicted based on models that estimate only the initiation or only the elongation. For native/endogenous transcripts, we see that most of the jam time can be attributed to the initiation step: across all transcripts, the mean time for initiation jamming is 8.7% of the total initiation time, while for the elongation jam time, the mean is up to 2.4% of the total time (Fig. 5). Furthermore, the maximum initiation jam time over all gene is 96.9% of the total initiation time, whereas the maximum elongation jam time at a codon is at most 8%. When we measure the mean elongation jam time across all codons, we see the mean is almost 20% out of the elongation time and for 25% of the genes this value is above 27%.

In addition, the Spearman correlation between initiation rate and initiation jam time is 0.72 (P-value $= 3.61 \times 10^{-293}$) and between initiation rate and mean elongation jam time it is 0.39 (P-value $= 1.7 \times 10^{-69}$). The correlation between translation rate and initiation jam time is 0.56 and with mean elongation jam time it is 0.54 (P-value $< 2.2 \times 10^{-308}$).

At the next step, we analyzed the effect of mutations on jamming (Supplementary Figs S5–S8). We found that the initiation jam time relative change has significantly high correlation with the initiation rate, where for the same position a change occurs—a similar number of transcripts decrease or increase in the same magnitude. We determined that the correlation between the relative change of initiation rate and initiation jam time is 0.93 (P-value $< 2.2 \times 10^{-308}$). If we check the same correlation per each mutation position, the minimum correlation is 0.738, and if we exclude the two extreme mutation positions, we see the minimum correlation is 0.856 and a total of 61 out of 67 position have a correlation of above 0.9 (all *P*-values are lower than 2.2×10^{-308}). While we have seen in previous sections that translation rate was mostly unaffected by single nucleotide mutations, there is greater impact on the jam time during elongation. This change may indeed further propagate towards changes in initiation rate and initiation jam time, as a change in jam time during elongation may free or occupy codons at the beginning of the ORF, depending on whether the jam time decreased or increased.

Previous studies have shown that a conserved region (codons 18–50) exists where the codons tend to be slower.[7,56] It was previously suggested that both slower initiation and slower elongation rates at the beginning of the coding region should contribute (independently) to decreased jamming.[7,56] To examine this, we performed the following simulation: first, we changed the codons at positions 18–50 to be the fastest synonymous codons; as a result, the mean traffic jam increased by 6.2%. This value is very significant since increase in traffic jams/ribosome-density should be proportional to the decrease in the number of free ribosomes, which consequently affects the organism growth rate.[49] The effect was

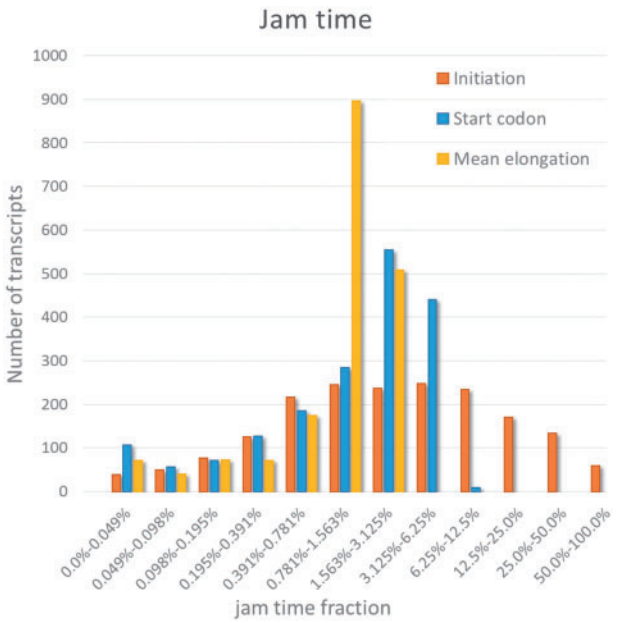

Figure 5. The fraction of time ribosome jam occurs for initiation, on start codon, and mean jam time across the entire ORF. Note that the bin sizes are not uniform but grow exponentially.

even higher when we changed the codons in positions 51–100 to be the slowest (Traffic jam increased by 25.2%) or when we changed all codons downstream the first 50 codons to be slowest (Traffic jam increased by 68.2%).

In this sub-section, we provide some estimation of traffic jams and the effect of codon distribution/bias on traffic jams. These estimation and relations can further validated experimentally via the design of relevant libraries of gene variants and the measurements of the profile or ribosome densities (e.g. using ribo-seq) of the different variants.

9.7. Genes with relatively extreme translation initiation and translation elongation rate

Our model enables recognizing and studying groups of genes with unique translation features. In this subsection, we demonstrate this via the analyses of various group of genes with extreme translation elongation and initiation levels.

We first considered the group of genes with high initiation rate but with relatively low elongation rate. These genes are interesting because, in general, there is a *positive* correlation between translation initiation and elongation (see previous sections). Relative slower elongation for these genes may be related to additional constraints such as co-translational folding.[57] We found 16 genes to be in the top 10% of the initiation rate and the bottom 10% of the elongation rate. When we raised the threshold to 20%, we found 63 genes (see Supplementary Table S1) and 267 genes were found with a threshold of 40%. Using functional annotation (Gene Ontology: GO; http://www.geneontology.org) we found that, among others, the latest group of genes includes 37 genes that are related to organelle inner membrane (GO: 0019866; P-value $= 6.84 \times 10^{-19}$; all the *P*-values reported in this section were adjusted for multiple comparison—see the Methods section) and 57 genes that are related to 'plasma membrane' (GO: 0005886; P-value $= 6.84 \times 10^{-9}$).

One explanation for these results may be related to the suggestion that membrane proteins undergo various maturation aspects during the elongation step which requires slower elongation (while the initiation may remain efficient).[7,58–60]

The second gene group we study includes genes with relatively low initiation rate (bottom 40%) and high elongation rate (40%). We found 170 such genes (see Supplementary Table S1); the group was enriched with genes relate to purine nucleotide biosynthetic process (GO: 0006164/ecx00230; $P = 2.52 \times 10^{-3}$) and genes related to translation/ribosome (GO: 0006412; $P = 0.03$). It is possible that these are highly expressed genes with constraints on the nucleotide composition near the start codon that effect the evolution of and optimality of their initiation (e.g. due to specific amino acids at the 5′end or signals related to transcription); it is also possible that some of them have non-canonical initiation regulation.

The third gene group we study includes genes with relatively high initiation rate (bottom 40%) and high elongation rate (40%). We found 228 such genes (see Supplementary Table S1); this group, naturally, includes highly expressed genes and enriched with housekeeping gene groups/functions such as amino acid metabolism (e.g. GO: 0008652 with $P = 1.44 \times 10^{-5}$) translation/ribosome (e.g. GO: 0003735 with $P = 7.25 \times 10^{-6}$), protein folding/chaperone (GO: 0006457 with $P = 0.0067$).

The groups of genes described here can be further experimentally studied in the future, by introducing silent mutations into the coding sequences and the UTRs of these genes and examining their effect on the improvement of translation rate, the functionality of the related proteins, and the organism fitness.

9.8. Mutations with reciprocal effect on initiation and elongation

In this section we aimed at further demonstrating how the new model can answer new important questions related to the effect on transcript evolution and mRNA translation regulation and biophysics. The (non intuitive) reciprocal effect of mutations on elongation and initiation in different regions is a simple example of such a question which was already discussed in the past[7] but with no clear quantitative analyses.

Thus, mutations causing reciprocal effect in initiation and elongation are of interest, given that we naturally see/expect significant correlation between elongation and initiation rate. To study these types of mutations, for each possible mutation occurring in the beginning of the ORF (positions +3 until +34), we looked for cases where the initiation rate was increased by at least 25%, while the codon translation rate was reduced by at least 25% (which we name 'type 1 mutations'). We also looked for cases where initiation rate decreased by at least 25% and codon translation rate was increased by 25% (which we name 'type 2 mutations'). For each of the above cases such occurrence happens in 4.6% of the ORF mutations. More details of per-position changes are available in Supplementary Fig. S9. Interestingly, we found that, on average, in positions 3–17 of the coding region there are higher levels of mutations of type 2. Conversely, there are more occurrences of type 1 mutations in positions 18–34. Using Wilcoxon signed-rank test[61] we compared, per each mutation position, the relative initiation rate and the relative codon decoding rate. In about half of the positions within the range of 3 and 17, and in all of the positions in the range of 18–34, we found significant differences. These results demonstrate that many point mutations can have reciprocal effect on initiation and elongation and that the first nucleotides of the coding region tend to be relatively more 'optimized' to the translation initiation stage than the elongation stage. Previous studies have suggested that the beginning of the coding sequence includes 'late initiation' signals, which are part of the initiation step but encoded in the coding region.[7] Our model enables the prediction of the exact boundaries of this region.

In this sub-section, we provide estimations of the effect of each nucleotide/region on initiation and elongation. These estimations and relations can be further validated experimentally via the design of relevant libraries of gene variants and the measurements of the profile or ribosome densities (e.g. using ribo-seq) or better performing high resolution tracking of ribosome movements[62] over the different variants.

10. Discussion

The model described in this study allows predicting, for the first time, translation dynamics based on the transcript features alone. As we demonstrate here, the model enables performing analyses that were not possible within the framework of previous models (e.g. statistical/machine learning models, or models that consider only the initiation or elongation step). Specifically, we demonstrate how the model can be used for predicting expression levels of endogenous and heterologous genes and predicting various aspects related to the translation dynamics. In addition, the model is very useful for studying the molecular evolution of translation aspects via studying the effect of mutations on translation dynamics.

Among others, we demonstrate a close association between changes in initiation rate and jam time during initiation. Within the framework of our model, it is easy to explain such relations, as increase or decrease in initiation rate can directly alter jam time. Within the cell, it is also possible that changes affecting the initiation jam time (such as those occurring in the ORF) can influence the initiation rate. We also demonstrate that mutations at the RBS region have the strongest effect on translation and we show that membrane genes tend to have relatively slower elongation and high initiation. Furthermore, we show that around 4–5% of the mutations near the beginning of the ORF *increase* initiation rate by at least 25% while *decreasing* by at least 25% the elongation efficiency. We also find 4–5% of these mutations *increase* elongation while *reducing* initiation rate by the same factor. Our analysis provides an estimation of the ratio between translation rate and initiation rate in the *E. coli* endogenous genes and found that its median is 10.13%. This result further implies that elongation (and the coding region itself) and not only initiation has central effect on the translation rate.

It is important to emphasize that this model connects, for the first time, the impact of *mutations* (possibly silent) in the transcripts with *fundamental biophysical phenomena* (e.g., ribosome allocation, traffic jams, translation rate, etc.), which are directly related to the organism fitness. Thus, with this model (and additional relevant adjustments), we are able to understand the effect of silent mutations on fitness and thus estimate their corresponding selection pressure; this is an important aspect that should be integrated into any realistic molecular evolution models. While existing models (including non-physical ones such as regressors) may similarly predict the expression levels of endogenous genes,[23,25,63] it is important to emphasize that our model's novel potential to predict the impact of single mutations on initiation and elongation biophysics, as well as other translation aspects, based on the mRNA sequence alone.

These results can only be derived from a combined model, such as the one presented here. The model's application can be downloaded and used in future quantitative studies related to translation in the field.

Future work can extend the current model to support wider and more complex aspects of translation. For instance, whole cell simulations, taking into account additional constraints, such as the tRNA pool size and ribosome pool size may provide additional insights. mRNA and protein degradation rates can be incorporated to better correlate mRNA levels with protein abundance, etc.

We chose to calculate the initiation rates only for the first gene in each operon. Extending the calculations to genes in other positions of the operon may yield additional interesting results but this requires modifications in the model due to additional possible interactions of the mRNA molecules with itself and with the ribosome.[64] Furthermore, although we provide some initial results for *B. subtilis* (see in the Supplementary material), our analysis focused on *E. coli* simply since it is the prokaryote with the most abundance relevant large-scale measurements. It is important to emphasize that the model (with some small adjustment) can be used on other prokaryotes and in the future, it will be interesting to comprehensively study other prokaryotes with similar models. In addition, it will be interesting to develop similar model for Eukaryotes and in different tissues; specifically, for this domain currently there is no initiation model while elongation model exists and is relatively similar between the two domains.[65]

Finally, in this work, we assume steady state of the translation process. A more dynamic model, taking into account mRNA folding during ribosomes detachment and mRNA decay rate may result in simulations of higher accuracy and a better understanding of the biophysics of translation.

Funding

This research was partially supported by a research grant from the Israeli Ministry of Science, Technology and Space and by a grant from the Ela Kodesz Institute for Medical Physics and Engineering.

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at *DNARES* online.

References

1. Hannig, G. and Makrides, S.C. 1998, Strategies for optimizing heterologous protein expression in Escherichia coli, *Trends Biotechnol.*, **16**, 54–60.
2. Kozak, M. 1999, Initiation of translation in prokaryotes and eukaryotes, *Gene*, **234**, 187–208.
3. Shine, J. and Dalgarno, L. 1975, Determinant of cistron specificity in bacterial ribosomes, *Nature*, **254**, 34–8.
4. Ikemura, T. 1981, Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system, *J. Mol. Biol.*, **151**, 389–409.
5. Gouy, M. and Gautier, C. 1982, Codon usage in bacteria: correlation with gene expressivity, *Nucleic Acids Res.*, **10**, 7055–74.
6. Dana, A. and Tuller, T. 2014, The effect of tRNA levels on decoding times of mRNA codons, *Nucleic Acids Res.*, **42**, 9171–81.
7. Tuller, T. and Zur, H. 2015, Multiple roles of the coding sequence 5' end in gene expression regulation, *Nucleic Acids Res.*, **43**, 13–28.
8. Gustafsson, C., Govindarajan, S. and Minshull, J. 2004, Codon bias and heterologous protein expression, *Trends Biotechnol.*, **22**, 346–53.
9. Kimchi-Sarfaty, C., Oh, J. M., Kim, I.-W., et al. 2007, A "Silent" Polymorphism in the MDR1 gene changes substrate specificity, *Science*, **315**, 525–8.
10. Bahir, I., Fromer, M., Prat, Y. and Linial, M. 2009, Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences, *Mol. Syst. Biol.*, **5**, 311.
11. Zhang, F., Saha, S., Shabalina, S.A. and Kashina, A. 2010, Differential arginylation of actin isoforms is regulated by coding sequence-dependent degradation, *Science*, **329**, 1534–7.
12. Fredrick, K. and Ibba, M. 2010, How the sequence of a gene can tune its translation, *Cell*, **141**, 227–9.
13. Zur, H. and Tuller, T. 2016, Predictive biophysical modeling and understanding of the dynamics of mRNA translation and its evolution, *Nucleic Acids Res*, **44**, 9031–49.
14. Reuveni, S., Meilijson, I., Kupiec, M., Ruppin, E. and Tuller, T. 2011, Genome-scale analysis of translation elongation with a ribosome flow model wasserman, *PLoS Comput. Biol.*, **7**, e1002127.
15. Sharp, P.M. and Li, W.H. 1987, The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications, *Nucl. Acids Res.*, **15**, 1281–95.
16. Shaw, L.B., Zia, R.K.P. and Lee, K.H. 2003, Totally asymmetric exclusion process with extended objects: a model for protein synthesis, *Phys. Rev. E*, **68**, 021910.
17. dos Reis, M., Savva, R. and Wernisch, L. 2004, Solving the riddle of codon usage preferences: a test for translational selection, *Nucleic Acids Res.*, **32**, 5036–44.
18. Dana, A. and Tuller, T. 2014, Mean of the typical decoding rates: a new translation efficiency index based on the analysis of ribosome profiling data, *G3 Bethesda Md*, **5**, 73–80.
19. Salis, H.M., Mirsky, E.A. and Voigt, C.A. 2009, Automated design of synthetic ribosome binding sites to control protein expression, *Nat. Biotechnol.*, **27**, 946–50.
20. Gu, W., Zhou, T. and Wilke, C.O. 2010, A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes , *PLoS Comput. Biol.*, **6**, e1000664.
21. Salis, H.M. 2011, *The Ribosome Binding Site CalculatorMethods in Enzymology*. Elsevier, pp. 19–42.
22. Zur, H. and Tuller, T. 2013, New universal rules of eukaryotic translation initiation fidelity *PLoS Comput. Biol.*, **9**, e1003136.
23. Vogel, C., Abreu Rde, S., Ko, D., et al. 2010, Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line, *Mol. Syst. Biol.*, **6**, 400.
24. Huang, T., Wan, S., Xu, Z., et al. 2011, Analysis and prediction of translation rate based on sequence and functional features of the mRNA Kudla, *PLoS One*, **6**, e16036.
25. Zur, H. and Tuller, T. 2013, Transcript features alone enable accurate prediction and understanding of gene expression in *S. cerevisiae*, *BMC Bioinformatics*, **14**, S1.
26. Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., et al. 2011, ViennaRNA Package 2.0, *Algorithms Mol. Biol.*, **6**, 26.
27. Hashimoto, M., Ichimura, T., Mizoguchi, H., et al. 2004, Cell size and nucleoid organization of engineered Escherichia coli cells with a reduced genome: engineered *E. coli* with a reduced genome, *Mol. Microbiol.*, **55**, 137–49.
28. Kato, J. and Hashimoto, M. 2007, Construction of consecutive deletions of the *Escherichia coli* chromosome, *Mol. Syst. Biol.*, **3**, 132.
29. Siwiak, M. and Zielenkiewicz, P. 2010, A Comprehensive, quantitative, and genome-wide model of translation, *PLoS Comput. Biol.*, **6**, e1000865.
30. Siwiak, M. and Zielenkiewicz, P. 2013, Transimulation -protein biosynthesis web service, *PLoS One*, **8**, e73943.
31. Dana, A. and Tuller, T. 2014, Properties and determinants of codon decoding time distributions, *BMC Genomics*, **15**, S13.
32. Guet, C.C., Bruneaux, L., Min, T.L., et al. 2008, Minimally invasive determination of mRNA concentration in single living bacteria, *Nucleic Acids Res.*, **36**, e73.

33. Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., et al. 2016, RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond, *Nucleic Acids Res*., **44**, D133–43.
34. Wang, M., Weiss, M., Simonovic, M., et al. 2012, PaxDb, a database of protein abundance averages across all three domains of life, *Mol. Cell. Proteomics*, **11**, 492–500.
35. Taniguchi, Y., Choi, P.J., Li, G.-W., et al. 2010, Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells, *Science*, **329**, 533–8.
36. Li, G.-W., Oh, E. and Weissman, J. S. 2012, The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria, *Nature*, **484**, 538–41.
37. Martin, M. 2011, Cutadapt removes adapter sequences from high-throughput sequencing reads, *Embnet. J*., **17**, 10.
38. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. 2009, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol*., **10**, R25.
39. Huang, D.W., Sherman, B.T. and Lempicki, R.A. 2008, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat. Protoc*., **4**, 44–57.
40. Huang, D.W., Sherman, B.T. and Lempicki, R.A. 2009, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, *Nucleic Acids Res*., **37**, 1–13.
41. Schwanhäusser, B., Busse, D., Li, N., et al. 2011, Global quantification of mammalian gene expression control, *Nature*, **473**, 337–42.
42. Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M. F. and von der Haar, T. 2014, Translation elongation can control translation initiation on eukaryotic mRNAs, *Embo J*., **33**, 21–34.
43. Jacques, N. and Dreyfus, M. 1990, Translation initiation in *Escherichia coli*: old and new questions, *Mol. Microbiol*., **4**, 1063–7.
44. Purvis, I.J., Bettany, A.J., Santiago, T.C., et al. 1987, The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis, *J. Mol. Biol*., **193**, 413–7.
45. Brunak, S. and Engelbrecht, J. 1996, Protein structure and the sequential structure of mRNA: alpha-helix and beta-sheet signals at the nucleotide level, *Proteins*, **25**, 237–52.
46. Thanaraj, T.A. and Argos, P. 1996, Ribosome-mediated translational pause and protein domain organization, *Protein Sci.* , **5**, 1594–612.
47. Thanaraj, T.A. and Argos, P. 1996, Protein secondary structural types are differentially coded on messenger RNA, *Protein Sci.* ., **5**, 1973–83.
48. Adzhubei, A.A., Adzhubei, I.A., Krasheninnikov, I.A. and Neidle, S. 1996, Non-random usage of "degenerate" codons is related to protein three-dimensional structure, *FEBS Lett*., **399**, 78–82.
49. Raveh, A., Margaliot, M., Sontag, E.D. and Tuller, T. 2016, A model for competition for ribosomes in the cell, *J. R Soc. Interface*, **13**, 20151062.
50. Ringquist, S., Shinedling, S., Barrick, D., et al. 1992, Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site, *Mol. Microbiol*., **6**, 1219–29.
51. Chen, H., Bjerknes, M., Kumar, R. and Jay, E. 1994, Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs, *Nucl. Acids Res*., **22**, 4953–7.
52. Ma, J., Campbell, A. and Karlin, S. 2002, Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures, *J. Bacteriol*., **184**, 5733–45.
53. Hui, A. and de Boer, H.A. 1987, Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A*, **84**, 4762–6.
54. de Smit, M.H. and van Duin, J. 1994, Translational initiation on structured messengers: another role for the shine-dalgarno interaction, *J. Mol. Biol*., **235**, 173–84.
55. Ingolia, N. ., Ghaemmaghami, S., Newman, J.R.S. and Weissman, J.S. 2009, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling, *Science*, **324**, 218–23.
56. Tuller, T., Carmi, A., Vestsigian, K., et al. 2010, An evolutionarily conserved mechanism for controlling the efficiency of protein translation, *Cell*, **141**, 344–54.
57. Hardesty, B., Tsalkova, T. and Kramer, G. 1999, Co-translational folding, *Curr. Opin. Struct. Biol*., **9**, 111–4.
58. Wolin, S.L. and Walter, P. 1993, Discrete nascent chain lengths are required for the insertion of presecretory proteins into microsomal membranes, *J. Cell Biol*., **121**, 1211–9.
59. Alberts, B. 2002, *Molecular Biology of the Cell*, 4th ed. Garland Science, New York, NY.
60. Murakami, A., Nakatogawa, H. and Ito, K. 2004, Translation arrest of SecM is essential for the basal and regulated expression of SecA, *Proc. Natl. Acad. Sci. USA*, **101**, 12330–5.
61. Wilcoxon, F. 1945, Individual comparisons by ranking methods, *Biom. Bull*., **1**, 80.
62. Uemura, S., Aitken, C.E., Korlach, J., Flusberg, B.A., Turner, S.W. and Puglisi, J.D. 2010, Real-time tRNA transit on single translating ribosomes at codon resolution, *Nature*, **464**, 1012–7.
63. Shaham, G. and Tuller, T. 2014, Most associations between transcript features and gene expression are monotonic, *Mol. Biosyst*, **10**, 1426–40.
64. Tian, T. and Salis, H.M. 2015, A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons, *Nucleic Acids Res*., **43**, 7137–51.
65. Jackson, R.J., Hellen, C.U.T. and Pestova, T.V. 2010, The mechanism of eukaryotic translation initiation and principles of its regulation, *Nat. Rev. Mol. Cell Biol*., **11**, 113–27.