

SCIENTIFIC REPORTS



OPEN

Sensitivity of mRNA Translation

Gilad Poker¹, Michael Margaliot¹ & Tamir Tuller²

Received: 24 November 2014

Accepted: 09 July 2015

Published: 04 August 2015

Using the dynamic mean-field approximation of the totally asymmetric simple exclusion process (TASEP), we investigate the effect of small changes in the initiation, elongation, and termination rates along the mRNA strand on the steady-state protein translation rate. We show that the sensitivity of mRNA translation is equal to the sensitivity of the maximal eigenvalue of a symmetric, nonnegative, tridiagonal, and irreducible matrix. This leads to new analytical results as well as efficient numerical schemes that are applicable for large-scale models. Our results show that in the usual endogenous case, when initiation is more rate-limiting than elongation, the sensitivity of the translation rate to small mutations rapidly increases towards the 5' end of the ORF. When the initiation rate is high, as may be the case for highly expressed and/or heterologous optimized genes, the maximal sensitivity is with respect to the elongation rates at the middle of the mRNA strand. We also show that the maximal possible effect of a small increase/decrease in any of the rates along the mRNA is an increase/decrease of the same magnitude in the translation rate. These results are in agreement with previous molecular evolutionary and synthetic biology experimental studies.

During mRNA translation, complex molecular machines called ribosomes attach to the 5' end of the messenger RNA (mRNA) and then scan it in a sequential manner. At each elongation step, a nucleotide triplet (codon) is read and the ribosome waits until a freely diffusing transfer RNA (tRNA), carrying the corresponding amino-acid, binds to the ribosome. The process ends when the ribosome reaches the 3' end of the mRNA, detaches, and releases the chain of amino-acids that folds into a functioning protein¹.

Translation is a crucial step in gene expression, and it is becoming increasingly clear that understanding this process is vital in order to reveal how biological systems develop, evolve, and function. Indeed, mRNA translation is the most extensively regulated step in mammals², and a 100-fold range of translational efficiency was detected between different genes^{3,4}. This clearly has a strong effect on the protein abundance that cannot be predicted by measuring mRNA abundances alone. In particular, two important dynamical aspects of the translation process are: (1) certain codons are “slower” than others due to factors such as low abundance of tRNA molecules with the corresponding anti-codon, folding of the mRNA, and interactions of the translated protein and the ribosome⁵; and (2) many ribosomes scan along the same mRNA chain in parallel and “traffic jams” can form behind a slowly moving ribosome.

TASEP is the standard mathematical model for translation^{6–8}. In this model, particles move along a chain of n consecutive sites. Each site can be either occupied by a particle or free. A particle attaches to the first site with probability α (but only if this site is free), hops from site i to site $i + 1$ with probability γ_i (but only if site $i + 1$ is free), and hops from the last site of the chain with probability β . In the homogeneous TASEP, all the transitions rates γ_i are assumed to be equal and, without loss of generality, scaled to one. In the context of translation, the particles [chain] model the ribosomes [mRNA molecule].

The dynamic mean-field approximation of TASEP⁹, sometimes called the *ribosome flow model* (RFM)¹⁰, is a set of n ordinary differential equations:

¹School of Elec. Eng.-Systems, Tel Aviv University, Israel. ²Dept. of Biomedical Eng., Tel Aviv University, Israel. Correspondence and requests for materials should be addressed to T.T. (email: tamirtul@post.tau.ac.il)

$$\begin{aligned}
\dot{x}_1 &= \lambda_0(1 - x_1) - \lambda_1 x_1(1 - x_2), \\
\dot{x}_2 &= \lambda_1 x_1(1 - x_2) - \lambda_2 x_2(1 - x_3), \\
\dot{x}_3 &= \lambda_2 x_2(1 - x_3) - \lambda_3 x_3(1 - x_4), \\
&\vdots \\
\dot{x}_{n-1} &= \lambda_{n-2} x_{n-2}(1 - x_{n-1}) - \lambda_{n-1} x_{n-1}(1 - x_n), \\
\dot{x}_n &= \lambda_{n-1} x_{n-1}(1 - x_n) - \lambda_n x_n.
\end{aligned} \tag{1}$$

Here $x_i(t) \in [0, 1]$ is the normalized occupancy level at site i at time t , and $\lambda_i > 0$ is a parameter that controls the transition rate from site i to the consecutive site $i+1$. To explain this model, consider for example the equation

$$\dot{x}_1 = \lambda_0(1 - x_1) - \lambda_1 x_1(1 - x_2).$$

The term $\lambda_0(1 - x_1)$ is the rate at which ribosomes attach to the beginning of the chain. This is given by the product of the initiation rate λ_0 and the term $(1 - x_1)$. This means that as x_1 increases, *i.e.*, as site 1 becomes fuller, the effective binding rate decreases. In particular, when $x_1(t) = 1$ the site is completely full and the effective binding rate is zero. The term $(1 - x_1)$ thus reflects a “soft version” of the simple exclusion principle of TASEP. The term $\lambda_1 x_1(1 - x_2)$ is the rate in which ribosomes move from site 1 to site 2. This is proportional to the occupancy level at site 1, and to $(1 - x_2)$ representing again the simple exclusion principle. The symmetry between the x_i and $(1 - x_i)$ terms also preserves the so called particle-hole symmetry of TASEP¹¹. The term $R(t) = \lambda_n x_n(t)$ is the rate of ribosomes exiting from the last site, that is, the protein translation rate at time t .

Unlike TASEP, the RFM is a deterministic, synchronous, and continuous-time model. Nevertheless, it has been shown that for the range of parameters that are relevant for translation RFM and TASEP provide highly correlated predictions¹⁰.

The state-space of the RFM is the unit cube $C^n = \{x \in \mathbb{R}^n: x_i \in [0, 1]: i = 1, \dots, n\}$. For $a \in C^n$, let $x(t, a)$ denote the solution at time t of the RFM emanating from $x(0) = a$. For a set S , let $\text{int}(S)$ denote the interior of S . It is known^{12,13} that the RFM admits a unique equilibrium $e = e(\lambda_0, \dots, \lambda_n) \in \text{int}(C^n)$, and that every trajectory of the RFM converges to e , that is, $\lim_{t \rightarrow \infty} x(t, a) = e$ for all $a \in C^n$. In particular, $R(t)$ converges to the *steady-state protein translation rate* $R := \lambda_n e_n$. In other words, the ribosomal densities and thus the translation rate always converge to steady-state values that depend on the rates, but not on the initial conditions.

Substituting e for x in (1) yields

$$\begin{aligned}
\lambda_0(1 - e_1) &= \lambda_1 e_1(1 - e_2) \\
&= \lambda_2 e_2(1 - e_3) \\
&\vdots \\
&= \lambda_{n-1} e_{n-1}(1 - e_n) \\
&= \lambda_n e_n,
\end{aligned} \tag{2}$$

and this gives

$$R = \lambda_i e_i(1 - e_{i+1}), \quad i = 0, 1, \dots, n, \tag{3}$$

where we define $e_0 := 1$ and $e_{n+1} := 0$. From this it is possible to obtain an elegant continued fraction (CF) equation for R . For example, for $n = 2$, (3) yields

$$1 - \frac{R/\lambda_0}{1 - \frac{R/\lambda_1}{1 - R/\lambda_2}} = 0.$$

Solving the CF equation for R provides a mapping $(\lambda_0, \dots, \lambda_n) \rightarrow R$ (note that since $e \in C^n$ is unique, there is a single feasible solution R).

It has been recently shown¹⁴ that the $(n+2) \times (n+2)$ matrix:

$$A := \begin{bmatrix} 0 & \lambda_0^{-1/2} & 0 & 0 & \dots & 0 & 0 \\ \lambda_0^{-1/2} & 0 & \lambda_1^{-1/2} & 0 & \dots & 0 & 0 \\ 0 & \lambda_1^{-1/2} & 0 & \lambda_2^{-1/2} & \dots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & \dots & \lambda_{n-1}^{-1/2} & 0 & \lambda_n^{-1/2} \\ 0 & 0 & 0 & \dots & 0 & \lambda_n^{-1/2} & 0 \end{bmatrix} \tag{4}$$

has real and distinct eigenvalues: $\zeta_1 < \zeta_2 < \dots < \zeta_{n+2}$, with

$$\zeta_{n+2} = R^{-1/2}. \tag{5}$$

Furthermore, if we let $q_i, i \in \{0, \dots, n\}$, denote the $(i+1) \times (i+1)$ principal minor of $(\zeta_{n+2}I - A)$ then

$$q_i = R^{-(i+1)/2} e_i e_{i-1} \dots e_1. \tag{6}$$

Equation (5) provides a *spectral representation* of the mapping from the λ_i 's to R . As we will show below, this provides a useful framework for studying the dependence of the steady-state translation rate R in the RFM on the (generally inhomogeneous) initiation, elongation, and termination rates. Note that since A is a (componentwise) nonnegative matrix, ζ_{n+2} is also the Perron root¹⁵ of A , denoted $\rho(A)$.

Note that the matrix A should not be confused with the transition matrix used in the stochastic analysis of TASEP, as that matrix decodes the transitions between all possible particle configurations and is thus of dimensions $2^n \times 2^n$. This limits its use to very short TASEPs only.

Let $\mathbb{R}_{++}^k := \{x \in \mathbb{R}^k: x_i > 0, i = 1, \dots, k\}$. Recently, Eq. (5) has been applied¹⁴ to prove that the mapping $(\lambda_0, \dots, \lambda_n) \rightarrow R$ is strictly concave on \mathbb{R}_{++}^{n+1} . This means that the problem of maximizing R , subject to an affine constraint on the rates, is a *convex optimization problem*. Maximizing R , given the limited bimolecular budget, is important because it is known that translation is one of the most energy consuming processes in the cell¹. Thus, it is natural to assume that evolution optimized this process. Also, maximizing the translation rate is a major challenge in synthetic biology and specifically in heterologous gene expression¹⁶. Several other tools from the field of systems and control theory have been used to analyze the RFM^{13,17,18}.

Results

A number of interesting papers studied the effect of slow codon configurations on the steady-state translation rate using simulations and various approximations of TASEP^{19–22}. The spectral representation of R provides a new, exact, and computationally efficient approach for studying this issue in the RFM using algorithms that compute the largest eigenvalue of a symmetric and tridiagonal matrix. For example, Fig. 1 shows R , computed via (5), for various slow rate configurations in an RFM with $n = 1000$.

Here, however, we use (5) to analyze a different (albeit related) notion, namely, the *sensitivities*:

$$s_i(\lambda_0, \dots, \lambda_n) := \frac{\partial}{\partial \lambda_i} R(\lambda_0, \dots, \lambda_n), i = 0, 1, \dots, n.$$

A relatively large value of s_k indicates that a small change in the rate λ_k will have a strong effect on the translation rate R . In other words, the sensitivities can be used to determine which rates are the most important in terms of their effect on the translation rate. The advantage of the spectral representation of R is that determining s_i becomes the classical eigenvalue sensitivity problem. Indeed, since A is (componentwise) nonnegative and irreducible¹⁵, there exists an eigenvector $v \in \mathbb{R}_{++}^{n+2}$ (that is unique up to scaling) such that $Av = \zeta_{n+2}v$. By known results from linear algebra²³,

$$\frac{\partial}{\partial \lambda_i} \zeta_{n+2} = \frac{v' \left(\frac{d}{d\lambda_i} A \right) v}{v'v}, \tag{7}$$

so

$$s_i = \frac{2R^{3/2} v_{i+1} v_{i+2}}{\lambda_i^{3/2} v'v}. \tag{8}$$

This provides a way to compute, in an efficient and numerically stable way, the sensitivities for large-scale inhomogeneous RFMs using standard algorithms for computing the eigenvalues and eigenvectors of symmetric tridiagonal matrices.

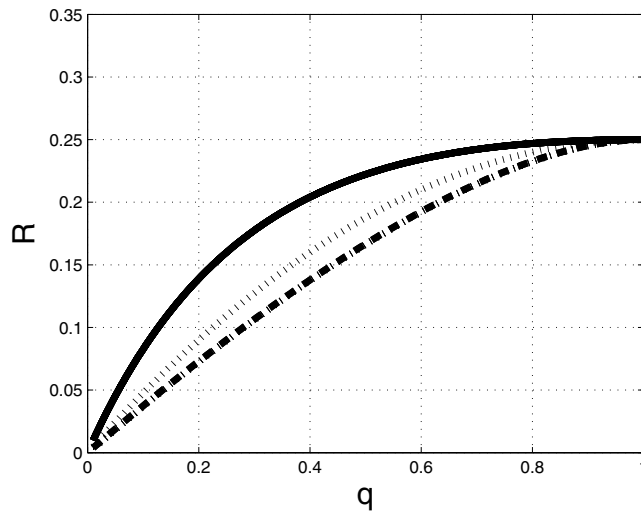


Figure 1. Steady state translation rate R in a RFM with $n = 1000$ and rates $\lambda_i = 1$, except for a configuration of slow rates with rate q . Solid line: $\lambda_{500} = q$; Dotted line: $\lambda_{500} = \lambda_{501} = q$; Dashed line: $\lambda_{499} = \lambda_{500} = \lambda_{501} = q$. As expected, R is an increasing function of q . Note that a cluster of consecutive slow sites considerably reduces R .

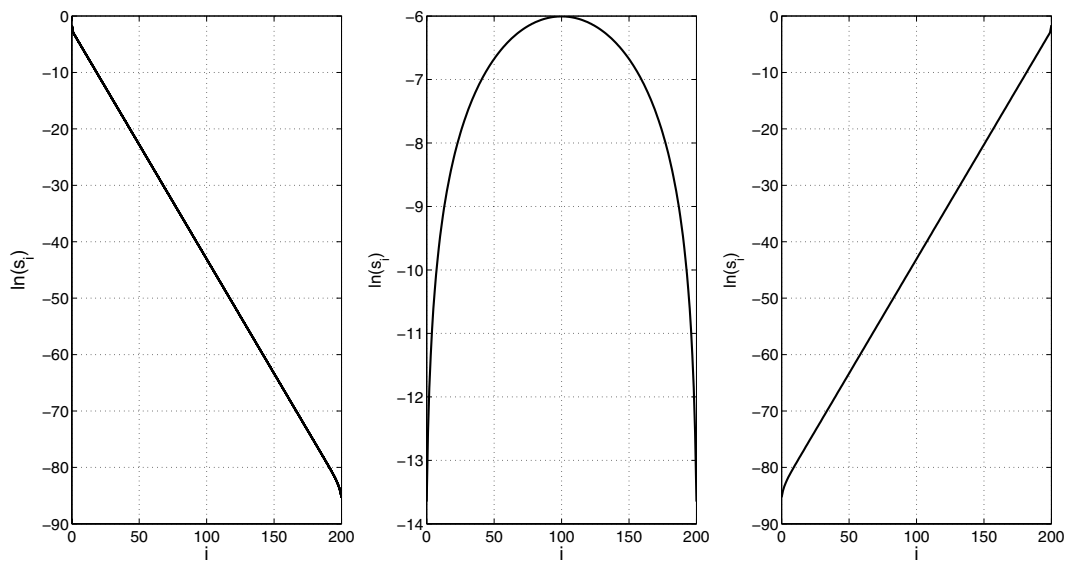


Figure 2. $\ln(s_i)$ as a function of i in three HRFMs with length $n = 200$. Left: $\lambda_0 = 0.4$ and $\lambda_{200} = 1$. Here λ_0 is rate limiting and thus the sensitivities of sites close to site 1 are larger. Note that $\ln(s_i)$ decays linearly with i . Middle: $\lambda_0 = \lambda_{200} = 1$. Here the maximal sensitivity is with respect to $\lambda_{n/2}$ and it decreases as we move towards the edges of the chain. Right: $\lambda_0 = 1$ and $\lambda_{200} = 0.4$.

Figure 2 depicts $\ln(s_i)$ as a function of i for three homogeneous RFMs (HRFMs) (*i.e.*, $\lambda_i = 1$ for all $i \in \{1, \dots, n - 1\}$), with size $n = 200$. The left sub-figure shows the case where $\lambda_0 = 0.4$ and $\lambda_{200} = 1$, so λ_0 is the rate limiting factor. The sensitivity s_0 is maximal and the sensitivities decrease as i increases. This regime describes the typical case in endogenous genes where initiation is the rate limiting factor²⁴. Such genes indeed demonstrate selection for increased robustness to transcription errors in ORF features that affect the translation rate (*e.g.* mRNA folding and adaptation to the tRNA pool)⁵. Similarly, our results may also explain the evolutionary selection for unusual codon usage bias at the ORF 5' end²⁵.

The right sub-figure shows the symmetric case where $\lambda_0 = 1$ and $\lambda_{200} = 0.4$. The middle sub-figure depicts the case where *all* the λ_i s are one. The plot demonstrates what is known in the TASEP literature as the *edge effect*²⁶: the maximal sensitivity is in the center of the chain, and it decreases as we move toward

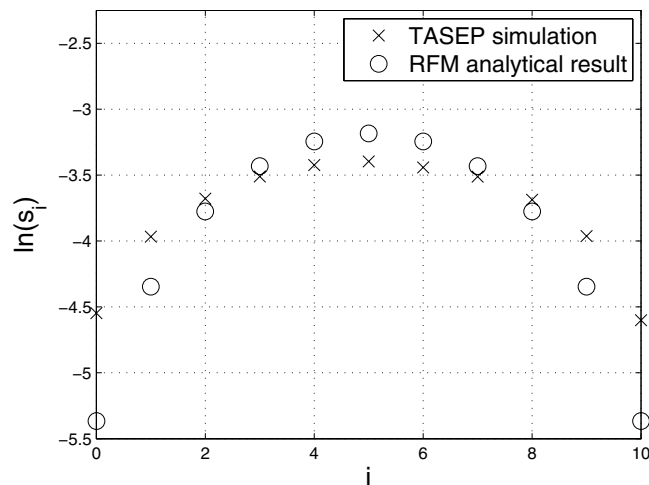


Figure 3. $\ln(s_i)$ as a function of i for a THRFM with $n = 10$. The analytical results for the RFM (o) are compared to TASEP sensitivity (x) calculated using Monte Carlo simulations.

the edges. This suggests that in order to maximize the translation rate in heterologous gene expression¹⁶, more attention should be devoted to tuning the codons in the middle of the coding sequence.

The spectral approach provides an upper bound on the sensitivities, namely,

$$s_i \leq 1, \quad \text{for all } i \in \{0, 1, \dots, n\} \quad (9)$$

(see the Methods section for details). This means that the maximal possible effect of a small increase/decrease in any of the rates is an increase/decrease of the same magnitude in the translation rate. This agrees with a recent experimental study on the change in protein abundance resulting from perturbing the codons of heterologous genes²⁷. In this study, 25 variants of the viral gene *HRSVgp04* were generated and the corresponding protein levels were measured in *S. cerevisiae*. In each variant *only* codons 41–80 of the ORF were perturbed, without changing the encoded protein; thus, mRNA levels and translation initiation were expected to be identical in all variants. For each variant, the predicted change in the corresponding $\lambda_i s$ (*i.e.* transition rates) was computed²⁸. An average change of 33.3% in the transition rate led to a 27.5% change in the protein levels.

We now focus on two special cases where it is possible to obtain *exact closed-form* expressions for the sensitivities.

Totally homogeneous ribosome flow model (THRFM). Suppose that $\lambda_i = \lambda_c$ for all $i \in \{0, \dots, n\}$. In other words, the initiation, termination, and all transition rates are equal, with λ_c denoting their common value. We refer to this case as the THRFM. In this case (see the Methods section)

$$R = \frac{\lambda_c}{4} \cos^{-2}\left(\frac{\pi}{n+3}\right), \quad (10)$$

and the sensitivities are

$$s_i = \frac{\sin\left(\frac{i+1}{n+3}\pi\right)\sin\left(\frac{i+2}{n+3}\pi\right)}{2(n+3)\cos^3\left(\frac{\pi}{n+3}\right)}, \quad i = 0, \dots, n. \quad (11)$$

This provides a closed-form expression for the graph shown in the middle plot of Fig. 2.

Equation (5) implies that for any $c > 0$, $R(c\lambda_0, \dots, c\lambda_n) = cR(0_0, \dots, \lambda_n)$. By Euler's homogeneous function theorem, $R = \sum_{i=0}^n \lambda_i s_i$, and for the THRFM this gives $\sum_{i=0}^n s_i = \frac{1}{4} \cos^{-2}\left(\frac{\pi}{n+3}\right)$. Thus, $\lim_{n \rightarrow \infty} \sum_{i=0}^n s_i = 1/4$.

Figure 3 compares the analytical formula (11) to simulations of the corresponding TASEP (see the methods section for details). It may be seen that the results agree reasonably well.

An exact measure of the edge effect in the THRFM is given by the ratio $\Delta(n) := \frac{s_{n/2}}{s_0}$. Substituting (11) and simplifying yields

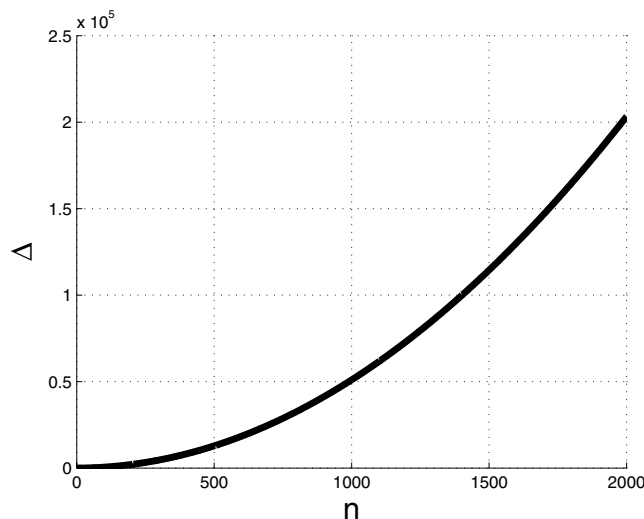


Figure 4. The function $\Delta(n) := \frac{s_{n/2}}{s_0}$ that measures the edge effect as a function of n .

$$\Delta(n) = \left(8 \cos\left(\frac{\pi}{n+3}\right) \sin^2\left(\frac{\pi/2}{n+3}\right) \right)^{-1}$$

(see Fig. 4). Note that $\lim_{n \rightarrow \infty} \Delta(n) = \infty$. In other words, although all the sensitivities decrease with n (see (11)) the edge effect actually becomes more prominent. One explanation for this edge effect is that a site at the center of the chain has more “close” neighbors than a site located towards one of the edges of the chain. Thus, when all the rates are equal the protein translation rate is more sensitive to the rates at the center of the chain.

Homogeneous and symmetric ribosome flow model (HSRFM). Another case where closed-form expressions for the sensitivities can be derived is when all the elongation rates are equal: $\lambda_1 = \lambda_2 = \dots = \lambda_{n-1} = \lambda_c$, and also

$$\lambda_n = \lambda_c - \lambda_0. \tag{12}$$

We refer to this case as the HSRFM. Here (see the Methods section for the details)

$$R = \lambda_c \alpha (1 - \alpha), \tag{13}$$

where $\alpha := \lambda_0/\lambda_c$. If $\alpha = 1/2$ then

$$s_i = \begin{cases} 0, & i = 0 \text{ or } i = n, \\ \frac{1}{4(n+1)}, & 1 \leq i \leq n-1. \end{cases} \tag{14}$$

If $\alpha \neq 1/2$,

$$s_i = \begin{cases} 1 - 2\alpha, & i = 0, \\ \frac{\alpha(1 - 2\alpha)}{1 - \mu^{n+1}} \mu^i, & 1 \leq i \leq n-1, \\ 2\alpha - 1, & i = n, \end{cases} \tag{15}$$

where $\mu := \lambda_0/\lambda_n$. This can be explained as follows. If $\lambda_0 < \lambda_n$ (so $\alpha < 1/2$ and $\mu < 1$) then $\frac{s_{i+1}}{s_i} = \mu$, so $\ln(s_{i+1}) - \ln(s_i) = \ln(\mu)$, for all $i \in \{1, \dots, n-2\}$. In other words, inside the chain the sensitivity strictly decreases with i , with $\ln(s_i)$ decaying linearly with i . This is reasonable, as in this case the initiation rate is the limiting factor. Note that $s_n < 0$. This is due to (12), as increasing λ_n means decreasing λ_0 , and since this is the rate limiting factor, this decreases R . The case $\lambda_0 > \lambda_n$ is symmetric.

Figure 5 depicts $\ln(s_i)$ in (15) for an HSRFM with $n = 10$, $\lambda_c = 1$, and $\lambda_0 = 0.3$. The results are also compared to TASEP simulations (see the Methods section for details). It may be seen that the results agree with these simulations, except for large values of i , that probably require a larger number of simulations

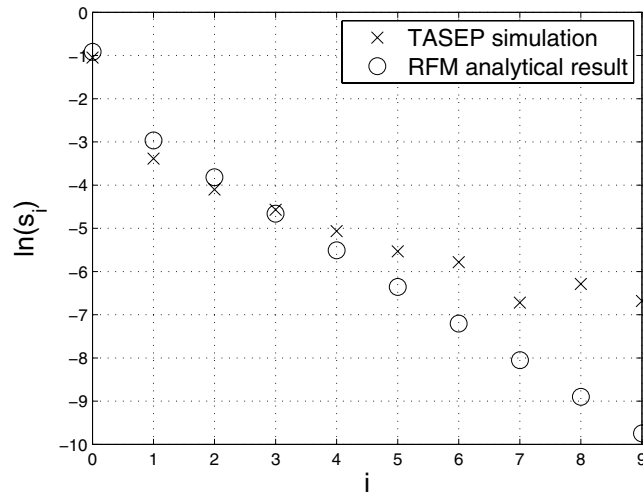


Figure 5. $\ln(s_i)$ as a function of i for an HSRFM with $n = 10$, $\lambda_c = 1$ and $\lambda_0 = 0.3$. The last value is not shown, as $s_{10} = -s_0 < 0$. The analytical results for the RFM (o) are compared to TASEP sensitivity (x) calculated using Monte Carlo simulations.

to obtain reliable averages. However, the sensitivity results for TASEP, obtained via Monte Carlo simulations, are problematic because the small change in the steady-state current, resulting from a small change in one of the transition rates, is masked by the inherent stochasticity of the process. Comparing Fig. 5 to Fig. 2 shows that the explicit equations for the HSRFM actually provide reasonable approximations to the general behavior of the HRFM in the case where the initiation rate is the limiting factor (and thus also in the symmetric case where termination rate is the limiting factor).

For a more reliable comparison between the sensitivity of RFM and that of TASEP, we used the closed-form formula for the steady-state current in the *homogenous* TASEP with N sites given by $J = Z(N-1)/Z(N)$, where

$$Z(k) = \sum_{j=1}^k \frac{j(2k-j-1)!}{k!(k-j)!} \frac{\alpha^{-j-1} - \beta^{-j-1}}{\alpha^{-1} - \beta^{-1}}$$

(see e.g. the survey paper⁹). Here α [β] is the entry [exit] probability and all the internal hopping rates γ_i are equal and normalized to one. We numerically calculated the sensitivity of J with respect to small changes in α , and compared this to the sensitivity of the corresponding RFM, i.e. the RFM with $n = N$, $\lambda_0 = \alpha$, $\lambda_i = 1$, $i = 1, \dots, n$, and $\lambda_n = \beta$. Figure 6 depicts $\frac{\Delta J}{\Delta \alpha}$ and $\frac{dR}{d\lambda_0}$ for the case $N = n = 30$, $\beta = 0.3$, and various values of α . It may be seen that the sensitivities agree well when $\alpha \ll 0.3$ and when $\alpha \gg 0.3$. The value $\alpha = 0.3$ corresponds to a phase transition in TASEP and then the sensitivity of the RFM, that is based on a mean-field approximation, differs from that of TASEP. In particular, it seems that the sensitivity near $\alpha = 0.3$ in TASEP changes more smoothly than the sensitivity in the corresponding RFM.

Discussion

Steady-state properties of the dynamic mean-field approximation of TASEP can be represented in a spectral form. Using this, we studied the sensitivity of the steady-state translation rate to perturbations in the initiation, transition, and termination rates. In this context, the problem reduces to the sensitivity analysis of the Perron root of a symmetric, nonnegative, tridiagonal, and irreducible matrix. This leads to: (1) highly efficient numerical computation of the sensitivities that is thus applicable for large-scale inhomogeneous RFMs; and (2) exact, closed-form expressions for the sensitivities in some special, yet important, cases.

Our results are based on analyzing the RFM which is the dynamic mean-field approximation of TASEP. However, simulations suggest that these results agree with the sensitivities in TASEP, which is the standard model for mRNA translation. The correspondence between the two models breaks down at parameter values that correspond to a phase transition in TASEP. In the remainder of this section, we discuss the relationship between the analytical results for the RFM sensitivity and known experimental findings, and suggest some future research directions.

First, a typical regime in the case of endogenous genes is when the initiation rate is the bottleneck rate¹. The analytical results imply that in this case mutations at the 5' end of the ORF will have a higher effect on the translation rate than mutations downstream of this region. The error rate in the process of gene transcription is estimated to be 1 in every 10^4 nucleotides¹. Thus, on average, one in every 67 windows with a length of 50 codons includes an error. Considering the fact that there are thousands of

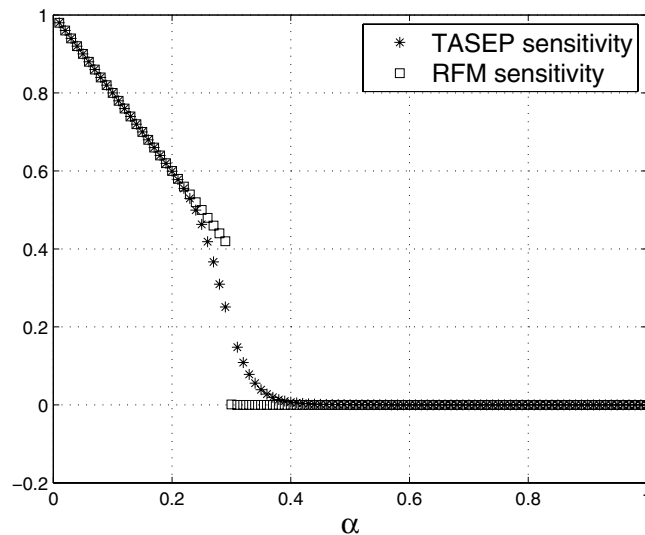


Figure 6. Sensitivity of the steady-state current J in the homogeneous TASEP (*) and of the corresponding homogenous RFM (\square) as a function of α .

copies of mRNA molecules in the cell (for example, the number of mRNA molecules in *S. cerevisiae* [*E. coli*] is around 60000²⁹ [1380]³⁰), and that genes are transcribed and translated continuously suggests that the probability of an error that reduces the organism's fitness is non-negligible. Thus, evolution should shape this region such that transcriptional mutations will have a lower effect on translation rates than in the rest of the ORF. Indeed, a recent study has shown that this may be the case⁵. This study included a computational analysis of all 40 nucleotide windows in endogenous ORFs in two model organisms: the prokaryote *E. coli* and the eukaryote *S. cerevisiae*. For each window, and for each coding sequence the effect of all possible point transcriptional mutations on features known to affect translation elongation rate (e.g. adaptation of codons to the tRNA pool²⁸ and local strength of mRNA folding^{5,25}) was computed; such a profile of each gene is actually a prediction of the mean effect of mutations on the local nominal translation rate (i.e. the different λ_i for all the sites i in our model) for that gene. At the next step, the mean of this profile over all the coding sequences of the organisms was computed to obtain the genomic profile of robustness of nominal translation rate to mutations (transcription errors). It was found that the robustness to mutations is higher at the beginning of the coding sequences (i.e. mutations at the beginning of the ORF tend to have a lower effect on the nominal translation rate). This suggests that there is an increased selection for robustness to the effect of mutations on translation rate at the beginning of genes. This result agrees with the current study: if sensitivity of mRNA translation to small changes in λ_i is higher at the beginning of the coding sequence it makes sense that evolution shaped these regions such that the effect of mutations on λ_i will be lower.

Second, another study supports our analytical results which imply that mutations (small changes in λ_i) at the 5' end of the ORF will have a higher effect on the translation rate than mutations downstream of this region. A recent synthetic biology experiment (under review) estimated the distinct and causal effects of different parts of the transcript in the eukaryote *S. cerevisiae*. Reporter libraries of the viral *HRSVgp04* gene have been generated for studying the effect of mutations in codons 2–41 and codons 42–81 of the ORF on translation rate and thus protein levels. The first library, named *L2–41*, measured the protein levels of various variants of different codons in positions 2–41 (i.e. different λ_i) but with the same codons elsewhere. The second library, *L42–81*, measured the protein levels of various variants of different codons in positions 42–81 but with the same codons elsewhere. It is important to mention that the size of these regions is close to the typical chunk corresponding to a single state-variable in the RFM¹⁰ (around 25 codons). It was found that the variability in protein levels is higher in the first library than in the second one; this suggests that the λ_i s in the first codons have a higher effect on translation rate, supporting the analysis reported here.

Third, we show that in the RFM the maximal possible effect of a small increase/decrease in any of the elongation rates along the mRNA is an increase/decrease of the same magnitude in the translation rate. Using library *L42–81* described above, it was possible to estimate the expected effect on translation elongation rate via applying a novel measure of per codon elongation rate²⁸ (corresponding to the λ_i s in the RFM) which is based on local ribosome density measurements using ribo-seq experiments³¹. Average changes of 33% in the rate led to an actual 27% average effect on the measured protein levels. This agrees well with our bound $s_i \leq 1$, suggesting that s_i can be helpful for designing synthetic genes, as it provides an estimate of the effect of synonymous mutations on the resulting protein levels.

Finally, we show that when all the rates are more or less equal, as may be the case for highly expressed and/or heterologous optimized genes, the maximal sensitivity is at the middle of the mRNA sequence. A common biotechnological approach for protein production is heterologous gene expression^{16,32} in which a target protein is expressed in a host cell. Often, the corresponding gene and host cell are optimized in order to maximize the protein levels of the target protein. In such cases, it makes sense to maximize all gene expression steps, and to make all of them uniformly efficient such that there are no bottlenecks. Furthermore, it also makes sense to design the heterologous gene such that these features will be maintained under different perturbations of the intracellular gene expression machinery. Our results suggest that perturbations near the middle of the ORF are expected to have a more prominent effect on the translation rate; this fact should be taken into consideration when designing heterologous genes.

It is important to mention that the conclusions reported here may also be relevant for other intracellular processes such as transcription of RNA, DNA replication, or intracellular trafficking over microtubule¹ that involve macro-molecules movement over a polymer. For example, transcription has been modeled using a mathematical model that is similar to the RFM³³ (in this case, the model describes the movement of RNA polymerase instead of ribosomes), and TASEP has posed the grounds for the development of theoretical tools describing molecular motors, with a good qualitative agreement when compared to state of the art experiments³⁴. We believe that the approach reported here should have practical contributions to the study of gene expression in various biomedical disciplines including evolutionary biology, functional genomics, and synthetic biology.

An interesting research topic is based on further validating and studying the results reported here both computationally and experimentally. Experiments similar to the ones described above where genes with different initiation rate and (constant) elongation rate should be engineered and the effect on translation rate (or protein levels) of small changes in the decoding rates in different parts of the coding sequence should be measured. Comprehensive computational models of gene expression that may include additional aspects that are not included in the RFM (e.g. mRNA degradation, the stochastic nature of the process, the fact that it may not converge to a steady state, the size of ribosomes, frame shifts) can be employed to understand the effect of small changes in the elongation rates on the translation rate and protein levels *in vivo*.

Methods

Derivation of Eq. (9). Pick $i \in \{0, 1, \dots, n\}$. Since $v_k > 0$ for all $k \in \{1, \dots, n + 2\}$, Eq. (8) implies that $s_i > 0$, *i.e.* an increase in any of the rates increases the steady-state translation rate. To determine an upper bound on s_i , perturb λ_i to $\tilde{\lambda}_i = \lambda_i + \epsilon$, with $\epsilon > 0$. This yields a perturbed matrix \tilde{A} that is identical to A except for entries $(i + 1, i + 2)$ and $(i + 2, i + 1)$ that are:

$$\tilde{\lambda}_i^{-1/2} = (\lambda_i + \epsilon)^{-1/2} = \lambda_i^{-1/2} - \epsilon \lambda_i^{-3/2} / 2 + o(\epsilon^2).$$

Thus, $\tilde{A} = A + E$, where E is a matrix with zero entries except for entries $(i + 1, i + 2)$ and $(i + 2, i + 1)$ that are $-\epsilon \lambda_i^{-3/2} / 2 + o(\epsilon^2)$. By Weyl’s inequality¹⁵, $\rho(\tilde{A}) \geq \rho(A) - \epsilon \lambda_i^{-3/2} / 2 + o(\epsilon^2)$. This yields $\frac{d\rho(A)}{d\lambda_i} \geq -\lambda_i^{-3/2} / 2$, so $s_i \leq \left(\frac{R}{\lambda_i}\right)^{3/2}$. Since $R \leq \lambda_i$ for all i^{10} , this yields (9).

Derivation of the sensitivities in the THRFM. For the THRFM, the matrix A in (4) becomes $A = \lambda_c^{-1/2} B$, where $B \in \mathbb{R}^{(n+2) \times (n+2)}$ is a tridiagonal Toeplitz matrix with zeros on the main diagonal, and ones on the super- and sub-diagonal. It is well-known³⁵ that the Perron root and Perron eigenvector of B are $\rho(B) = 2 \cos\left(\frac{\pi}{n+3}\right)$, and $v(B) = \left[\sin\left(\frac{\pi}{n+3}\right) \sin\left(\frac{2\pi}{n+3}\right) \dots \sin\left(\frac{(n+2)\pi}{n+3}\right)\right]$. Therefore, $\rho(A) = 2\lambda_c^{-1/2} \cos\left(\frac{\pi}{n+3}\right)$, and using $R = \rho^{-2}(A)$ yields (10). Substituting these values in (8) and using the fact that $\sum_{i=1}^{n+2} \sin^2\left(\frac{i\pi}{n+3}\right) = (n + 3) / 2$ yields (11).

Derivation of the sensitivities in the HSRFM. In the HSRFM, the matrix A in (4) becomes $A = \lambda_c^{-1/2} C$, where

$$C := \begin{pmatrix} 0 & \alpha^{-1/2} & 0 & \dots & 0 & 0 \\ \alpha^{-1/2} & 0 & 1 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ & & \vdots & & & \\ 0 & 0 & \dots & 1 & 0 & (1 - \alpha)^{-1/2} \\ 0 & 0 & \dots & 0 & (1 - \alpha)^{-1/2} & 0 \end{pmatrix},$$

with $\alpha = \lambda_0 / \lambda_c$. For $\alpha < 1$, $\rho(C) = (\alpha(1 - \alpha))^{-1/2}$, and $v(C)$ is given by

$$v_i = \begin{cases} 1, & i = 1, \\ \mu^{(i-1)/2} \alpha^{-1/2}, & 2 \leq i \leq n+1, \\ \mu^{n/2}, & i = n+2, \end{cases} \quad (16)$$

where $\mu := \alpha/(1 - \alpha) = \lambda_0/\lambda_n$. Indeed, a calculation shows that $Cv = \rho v$. Thus, v is an eigenvector corresponding to the eigenvalue ρ , and since $v \in \mathbb{R}_{++}^{n+2}$, this eigenvalue is the spectral radius. Note that if $\alpha = 1/2$ then $v'v = 2(1 + n)$, and otherwise $v'v = \frac{2(1 - \mu^{n+1})}{1 - \mu}$. Thus, $R = \rho^{-2}(A) = \lambda_c \alpha(1 - \alpha)$. The sensitivities in the HSRFM can now be determined. Here s_i , $1 \leq i \leq n - 1$, were calculated using (8). Due to the additional coupling in (12), s_0 and s_n cannot be computed using (8), so we used (7). This yields (14) and (15).

TASEP Simulation. Every site i , $i = 1, \dots, n$, in the TASEP chain is associated with “hopping-times”, where the time between two consecutive hopping times is exponentially distributed with rate λ_i . Hopping times into the first site are distributed with rate λ_0 . A simulation begins with an empty chain, and continues for 10^8 time steps. At each step all sites are scanned. If site i 's hopping time is equal to the simulation time, then the next hopping time is calculated. Also, if site i is occupied and site $i + 1$ is free, the particle hops from site i to site $i + 1$. The steady-state average occupancy of each site is the number of time steps it was occupied divided by the total simulation time, with the first 10^6 steps excluded from the calculation. For sensitivity simulations, TASEP with $n = 10$ sites is simulated 11 times. At each simulation, a different rate λ_i is perturbed by $\epsilon = 0.01$, while the rest of the rates are left unchanged. Sensitivity for this iteration is then calculated as $\Delta J/\Delta \lambda_i$ where $\Delta J = J(\lambda_0, \dots, \lambda_i + \epsilon, \dots, \lambda_n) - J(\lambda_0, \dots, \lambda_i, \dots, \lambda_n)$ is the change in the average steady-state current, and $\Delta \lambda_i = \epsilon$. Results are averaged over 1000 iterations. As noted above, these results must be taken with care, as the small change in J corresponding to the ϵ perturbation may be masked by the inherent stochasticity of the process. This is why we also added a comparison to the analytic closed-form for J available for the *homogeneous* TASEP. The MATLAB code for the simulations is available in the supplementary material accompanying this paper (and can be downloaded from: http://www.cs.tau.ac.il/~tamir-tul/supp_material_code.zip).

References

- Alberts, B. *et al. Molecular Biology of the Cell*. 4th edn (Garland Science, 2002).
- Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–42 (2011).
- Vasquez, J.-J., Hon, C.-C., Vanselow, J. T., Schlosser, A. & Siegel, T. N. Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. *Nucleic Acids Res.* **42**, 3623–3637 (2014).
- Rojas-Duran, M. F. & Gilbert, W. V. Alternative transcription start site selection leads to large differences in translation activity in yeast. *RNA* **18**, 2299–2305 (2012).
- Tuller, T. *et al.* Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biol.* **12**, R110 (2011).
- Schadschneider, A., Chowdhury, D. & Nishinari, K. *Stochastic Transport in Complex Systems: From Molecules to Vehicles* 1st edn (Elsevier, 2011).
- Zia, R. K. P., Dong, J. & Schmittmann, B. Modeling translation in protein synthesis with TASEP: A tutorial and recent developments. *J. Statistical Physics* **144**, 405–428 (2011).
- MacDonald, C. T., Gibbs, J. H. & Pipkin, A. C. Kinetics of biopolymerization on nucleic acid templates. *Biopolymers* **6**, 1–25 (1968).
- Blythe, R. A. & Evans, M. R. Nonequilibrium steady states of matrix-product form: a solver's guide. *J. Phys. A: Math. Gen.* **40**, R333–R441 (2007).
- Reuveni, S., Meilijson, I., Kupiec, M., Ruppim, E. & Tuller, T. Genome-scale analysis of translation elongation with a ribosome flow model. *PLoS Comput Biol.* **7**, e1002127; doi: 10.1371/journal.pcbi.1002127 (2011).
- Lakatos, G., Chou, T. & Kolomeisky, A. Steady-state properties of a totally asymmetric exclusion process with periodic structure. *Phys. Rev. E* **71**, 011103 (2005).
- Margaliot, M. & Tuller, T. Stability analysis of the ribosome flow model. *IEEE/ACM Trans Comput Biol Bioinform.* **9**, 1545–1552 (2012).
- Margaliot, M., Sontag, E. D. & Tuller, T. Entrainment to periodic initiation and transition rates in a computational model for gene translation. *PLoS One* **9**, e96039; doi: 10.1371/journal.pone.0096039 (2014).
- Poker, G., Zarai, Y., Margaliot, M. & Tuller, T. Maximizing protein translation rate in the nonhomogeneous ribosome flow model: A convex optimization approach. *J. R. Soc. Interface* **11**, 20140713 (2014).
- Horn, R. A. & Johnson, C. R. *Matrix Analysis* 2 edn. (Cambridge University Press, 2013).
- Gustafsson, C., Govindarajan, S. & Minshull, J. Codon bias and heterologous protein expression. *Trends Biotechnol.* **22**, 346–353 (2004).
- Margaliot, M. & Tuller, T. Ribosome flow model with positive feedback. *J. R. Soc. Interface* **10**, 20130267 (2013).
- Zarai, Y., Margaliot, M. & Tuller, T. Explicit expression for the steady-state translation rate in the infinite-dimensional homogeneous ribosome flow model. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **10**, 1322–1328 (2013).
- Chou, T. & Lakatos, G. Clustered bottlenecks in mRNA translation and protein synthesis. *Phys. Rev. Lett.* **93**, 198101 (2004).
- Pierobon, P., Mobilia, M., Kouyos, R. & Frey, E. Bottleneck-induced transitions in a minimal model for intracellular transport. *Phys. Rev. E* **74**, 031906 (2006).
- Dong, J. J., Schmittmann, B. & Zia, R. K. P. Inhomogeneous exclusion processes with extended objects: The effect of defect locations. *Phys. Rev. E* **76**, 051113 (2007).
- Kolomeisky, A. B. Asymmetric simple exclusion model with local inhomogeneity. *J. Phys. A: Math. Gen.* **31**, 1153–1164 (1998).
- Magnus, J. R. On differentiating eigenvalues and eigenvectors. *Econometr. Theor.* **1**, 179–191 (1985).

24. Malys, N. & McCarthy, J. Translation initiation: variations in the mechanism can be anticipated. *Cell Mol. Life Sci.* **68**, 991–1003 (2011).
25. Tuller, T. & Zur, H. Multiple roles of the coding sequence 5' end in gene expression regulation *Nucleic Acids Res.* **43**, 13–28 (2015).
26. Dong, J. J., Zia, R. K. P. & Schmittmann, B. Understanding the edge effect in TASEP with mean-field theoretic approaches. *J. Phys. A: Math. Gen.* **42**, 015002 (2009).
27. Ben-Yehzekel, T., Atar, S., Zur, H., Diamant, A., Goz, E., Marx, T., Cohen, R., Dana, A., Feldman, A., Shapiro, E., & Tuller, T. Rationally designed, heterologous *S. cerevisiae* transcripts expose novel expression determinants. *RNA Biol.* (2015) Jul 15:0. (PMID: 26176266) [Epub ahead of print].
28. Dana, A. & Tuller, T. The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.* **42**, 9171–81 (2014).
29. Zenklusen, D., Larson, D. R. & Singer, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* **15**, 1263–1271 (2008).
30. Neidhardt, F. C. *et al. Escherichia coli and Salmonella typhimurium - Cellular and Molecular Biology* 1st edn, vol. 1 (ASM Press, 1996).
31. Ingolia, N. T. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.* **15**, 205–213 (2014).
32. Binnie, C., Cossar, J. D. & Stewart, D. I. Heterologous biopharmaceutical protein expression in *Streptomyces*. *Trends Biotechnol.* **15**, 315–20 (1997).
33. Edri, S., Gazit, E., Cohen, E. & Tuller, T. The RNA polymerase flow model of gene transcription. *IEEE Trans Biomed Circuits Syst.* **8**, 54–64 (2014).
34. Ciandrini, L., Romano, M. C. & Parmeggiani, A. Stepping and crowding of molecular motors: Statistical kinetics from an exclusion process perspective. *Biophysical J.* **107**, 1176–1184 (2014).
35. Noschese, S., Pasquini, L. & Reichel, L. Tridiagonal Toeplitz matrices: properties and novel applications. *Numer. Linear Algebra Appl.* **20**, 302–326 (2013).

Acknowledgements

This work was supported by a research grant from the Israeli Ministry of Science, Technology, and Space. We also thank Yoram Zarai for helpful comments.

Author Contributions

G.P., M.M. and T.T. performed the research and wrote the paper.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Poker, G. *et al.* Sensitivity of mRNA Translation. *Sci. Rep.* **5**, 12795; doi: 10.1038/srep12795 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>