

Deciphering mRNA Sequence Determinants of Protein Production Rate


Juraj Szavits-Nossan*

*SUPA, School of Physics and Astronomy, University of Edinburgh,
Peter Guthrie Tait Road, Edinburgh EH9 3FD, United Kingdom*

Luca Ciandrini

*L2C, Université de Montpellier, CNRS, Montpellier, France
and DIMNP, Université de Montpellier, CNRS, Montpellier, France*

M. Carmen Romano

*SUPA, Institute for Complex Systems and Mathematical Biology, Department of Physics, Aberdeen AB24 3UE, United Kingdom
and Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen AB24 3FX, United Kingdom* (Received 2 August 2017; revised manuscript received 18 January 2018; published 19 March 2018)

One of the greatest challenges in biophysical models of translation is to identify coding sequence features that affect the rate of translation and therefore the overall protein production in the cell. We propose an analytic method to solve a translation model based on the inhomogeneous totally asymmetric simple exclusion process, which allows us to unveil simple design principles of nucleotide sequences determining protein production rates. Our solution shows an excellent agreement when compared to numerical genome-wide simulations of *S. cerevisiae* transcript sequences and predicts that the first 10 codons, which is the ribosome footprint length on the mRNA, together with the value of the initiation rate, are the main determinants of protein production rate under physiological conditions. Finally, we interpret the obtained analytic results based on the evolutionary role of the codons' choice for regulating translation rates and ribosome densities.

DOI: [10.1103/PhysRevLett.120.128101](https://doi.org/10.1103/PhysRevLett.120.128101)

Translation is one of the major steps in protein biosynthesis. During this process, the nucleotide sequence of a messenger RNA (mRNA) is translated into a functional protein. Each nucleotide triplet, called a codon, codes for a specific amino acid, the protein's building block. There is experimental evidence that the rate at which a certain mRNA is translated depends on its specific codon sequence [1–4], especially in the case of eukaryotes. Identifying sequence features that determine the protein production rate, also commonly referred to as the translation rate or efficiency, is a fundamental open question in molecular biology [1,3].

Translation is performed by molecular motors called ribosomes, which move unidirectionally along the mRNA. The amino acids are delivered to the ribosome by molecules called transfer RNAs (tRNAs), which are specific to the codon and the amino acid they deliver (Fig. 1). The dwelling time of a ribosome on a specific codon depends primarily on the abundance of the corresponding tRNA [5,6]. Understanding how codon sequences determine protein production rates can potentially unlock many synthetic applications [7,8].

The standard biophysical model of translation is known as the totally asymmetric simple exclusion process (TASEP), which captures the concurrent motion of ribosomes on the mRNA [9,10]. In this model, ribosomes progress along the mRNA codon by codon, provided that

the codon that a ribosome moves onto is not occupied by another ribosome. The protein production rate can then be identified as the ribosomal current of this driven lattice gas. Because of the net current, TASEP is not in equilibrium, and its steady state is only known for a few special cases [11,12]. Unfortunately, most biologically relevant variants of TASEP can only be studied numerically [13–15], and efficient methods of exploring a large number of parameters are lacking. The steady state of the TASEP with nonuniform hopping rates is a long outstanding problem in nonequilibrium statistical physics [15,16]. An analytic prediction of protein production rates is also needed in order to interpret recently developed ribosome profiling experiments that are capable of monitoring ribosome positions along the mRNA [17]. Such analytical prediction is missing in previous models of ribosome dynamics, implicitly or explicitly based on the TASEP [18–22], and it is key to decipher sequence determinants of protein production rates.

In this Letter, we develop a versatile analytic method to solve TASEP-based models of translation. Our analytic approach, integrated with simulations and experimental data, allows us to efficiently identify the main features of the mRNA codon sequence that determine the rate of protein production. The genome-wide comparison of our analytic predictions with numerical simulations shows an

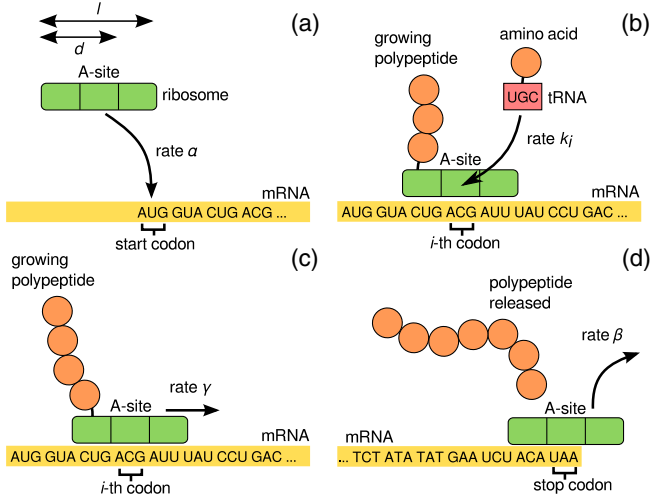


FIG. 1. A sketch of the mRNA translation process involving initiation (a), elongation consisting of tRNA delivery (b), and translocation (c), followed by termination (d). We emphasize that this is an oversimplified scheme of the process, and that actual ribosomes cover $\ell = 10$ codons. At each elongation step, the ribosome receives an amino acid from a tRNA that matches the codon occupied by the A site of the ribosome (the “reading” site). After the amino acid is added to the growing polypeptide, the ribosome translocates one codon forward, and the process is repeated.

excellent agreement for the model organism *S. cerevisiae* (baker’s yeast).

Stochastic model of mRNA translation.—We focus on the model for translation introduced in [21,23,24]. The mRNA is represented by a one-dimensional lattice consisting of L discrete sites (codons), where site 1 designates the start codon. Ribosomes are represented by particles that occupy $\ell = 10$ lattice sites, which is the ribosome footprint length measured in ribosome profiling experiments [17]. We identify the position $1 \leq i \leq L$ of a ribosome with the position of its A site, which is located $d = 5$ lattice sites from the trailing end of the ribosome [Fig. 1(a)]. A ribosome that waits for a tRNA at position i is labeled by 1_i , and a ribosome that has already received the correct tRNA and is ready to move is labeled by 2_i . The set of labels of all translating ribosomes on the lattice is called a configuration C of the system. For example, $C = 1_1 2_{13}$ denotes a lattice with 2 ribosomes, one at site 1 waiting for a tRNA and another one at site 13 that has already received the correct tRNA.

Ribosomes initiate translation at rate α by binding to the mRNA so that their A site is at the start codon, provided the sites $1, \dots, \ell - d + 1$ are empty [Fig. 1(a)]. A ribosome at site i makes the transition $1_i \rightarrow 2_i$ at rate k_i , dependent on tRNA abundances [Fig. 1(b)], and moves one site forward at rate γ [Fig. 1(c)]. Because of steric interactions between particles, the particle at site i can move only if there is no ribosome at site $i + \ell$. Termination occurs when the particle at site L receives the last amino acid, releases

the final protein, and detaches from the mRNA, which we integrate into a single step occurring at rate β [Fig. 1(d)].

Our main goal is to compute the rate of protein production as a function of the parameters of the model, which are α , β , γ , and k_i for each of the L codons used. We assume that translation takes place under steady-state conditions, so that the ribosomal current is constant along the mRNA and is equal to the rate of protein production, which in turn is equal to rate at which ribosomes load onto the mRNA and initiate translation. To this end, we define the codon occupation number τ_i to be equal to 1 if the i th codon is occupied by an A site, and 0 if otherwise. By definition, the exact steady-state ribosomal current J then reads

$$J = \alpha \sum_C \left(\prod_{i=1}^{\ell} (1 - \tau_i(C)) \right) P(C), \quad (1)$$

where $\tau_i(C)$ denotes the i th codon occupation number for configuration C , $P(C)$ denotes the steady-state probability that the lattice is in configuration C , and the summation goes over all configurations C . Other quantities of interest that we compute are the local and total particle densities $\rho_i = \langle \tau_i \rangle$ and $\rho = (1/L) \sum_{i=1}^L \rho_i$, respectively.

Series expansion method for computing $P(C)$.—In order to find $P(C)$ one has to solve the steady-state master equation $M\mathbf{P} = 0$, where \mathbf{P} is a column vector whose \mathcal{N} elements are the steady-state probabilities $P(C)$ of being in configuration C , and \mathcal{N} denotes the total number of configurations. The transition rate matrix $M_{C,C'}$ is given by $W_{C' \rightarrow C}$ for $C \neq C'$ and $-e(C)$ for $C = C'$, where $W_{C' \rightarrow C}$ is the transition rate from C' to C and $e(C) = \sum_{C' \neq C} W_{C \rightarrow C'}$ is the total exit rate from C . The exact solution of the master equation can be formally written as

$$P(C) = \frac{\det M^{(p,p)}}{\sum_{q=1}^{\mathcal{N}} \det M^{(q,q)}}, \quad (2)$$

where p is the position of configuration C in the column vector \mathbf{P} and $\det M^{(p,p)}$ is a determinant of the matrix obtained by removing p th row and p th column from M (see Supplemental Material [25] for details). Unfortunately, calculating this determinant is feasible only for unrealistically small system sizes.

To circumvent this problem, we exploit the fact that $\det M^{(p,p)}$ is a multivariate polynomial in the variables $\alpha, k_1, \dots, k_L, \beta$, and γ [26]. In the biological literature it is often assumed that the initiation rate α is a major limiting step of the translation process, mainly determined by the presence of secondary structures [27,28]. For this reason, we use α as an expansion parameter, and we assume that $\alpha \ll \gamma, \beta, k_1, \dots, k_L$. By collecting terms with the same power of α , we can rewrite $P(C)$ as an univariate polynomial $f(C)$ in the variable α with unknown coefficients $f_n(C)$

$$P(C) = \frac{f(C)}{\sum_C f(C)}, \quad f(C) = \sum_{n=0}^{K(C)} f_n(C) \alpha^n, \quad (3)$$

where the coefficients $f_n(C)$ depend on the transition rates k_1, \dots, k_L, γ , and β (in order to ease the notation, we leave out the explicit dependence on those parameters). Since we expect that $f(C)$ can be well approximated by the first few terms, the value of $K(C)$ in Eq. (3) is irrelevant in our study.

In order to find the unknown coefficients $f_n(C)$, we insert Eq. (3) into the master equation $M\mathbf{P} = 0$, collect all the terms with the same n th power of α , and equate their sum to zero. For a given power n , the resulting equation is similar to the original master equation in which $P(C)$ is replaced by $f_n(C)$ unless the coefficient multiplying $P(C)$ is α , in which case $P(C)$ is replaced by $f_{n-1}(C)$. Starting with $n = 0$, we note the equations for the coefficients $f_0(C)$ have the same form as the original master equation, but with $\alpha = 0$, leading to the trivial solution $f_0(C) = 0$ if $C \neq \emptyset$. Since the terms $f(C)$ are not normalized, we have the freedom to choose any value for $f_0(\emptyset)$, which we set to 1.

For $n \geq 1$, the equation for $f_n(C)$ reads

$$\begin{aligned} e_0(C)f_n(C) + f_{n-1}(C) \sum_{C'} I_{C,C'} \\ = \sum_{C'} I_{C',C} f_{n-1}(C') + \sum_{C'} (1 - I_{C',C}) W_{C' \rightarrow C} f_n(C'), \end{aligned} \quad (4)$$

where $I_{C,C'} = 1$ if $W_{C \rightarrow C'} = \alpha$, and is 0 otherwise, and $e_0(C) = \sum_{C'} (1 - I_{C,C'}) W_{C \rightarrow C'}$ is the total exit rate from C excluding the rate α .

A key observation in our analysis is that $f_n(C) = 0$ whenever the number of particles in C is larger than n . This follows from the result $f_0(C) = 1(0)$ if $C = \emptyset$ ($C \neq \emptyset$) in conjunction with the hierarchical structure of Eq. (4), which connects configurations differing in the number of particles by no more than one (a proof for $n = 1$ is presented in the Supplemental Material [25]). This allows us to write Eq. (4) taking into account configurations with only one particle and discarding all the others, which yields

$$f_1(1_1) = \frac{1}{k_1} f_0(\emptyset), \quad f_1(1_L) = \frac{\beta}{k_L} f_1(2_L), \quad (5a)$$

$$f_1(1_i) = \frac{\gamma}{k_i} f_1(2_{i-1}), \quad 2 \leq i \leq L, \quad (5b)$$

$$f_1(2_i) = \frac{k_i}{\gamma} f_1(1_i), \quad 1 \leq i \leq L-1. \quad (5c)$$

The solution to Eqs. (5) is given by

$$f_1(1_i) = \frac{1}{k_i}, \quad f_1(2_i) = \begin{cases} \frac{1}{\gamma} & i = 1, \dots, L-1 \\ \frac{1}{\beta} & i = L \end{cases}. \quad (6)$$

The equations for $f_2(C)$ involving configurations with one and two particles are presented in the Supplemental Material [25].

Once we determine the coefficients $f_n(C)$ up to a desired order n , we can compute the steady-state average of any observable $\mathcal{O}(C)$ by inserting $P(C)$ from Eq. (3), and expanding $\langle \mathcal{O} \rangle$ around $\alpha = 0$,

$$\langle \mathcal{O}(C) \rangle = \frac{\sum_{n=0}^K \sum_C \mathcal{O}(C) f_n(C) \alpha^n}{\sum_{n=0}^K \sum_C f_n(C) \alpha^n} = \sum_{n=0}^{\infty} c_n \alpha^n. \quad (7)$$

For example, the first three coefficients c_0 , c_1 and c_2 are given by

$$c_0 = \frac{a_0}{b_0}, \quad c_1 = \frac{a_1 - c_0 b_1}{b_0}, \quad c_2 = \frac{a_2 - c_1 b_1 - c_0 b_2}{b_0}, \quad (8)$$

where a_n and b_n are defined as

$$a_n = \sum_C \mathcal{O}(C) f_n(C), \quad b_n = \sum_C f_n(C). \quad (9)$$

Notice that the expansion in Eq. (7) is slightly different for the current J due to an extra α in Eq. (1) and is given by $J = \sum_{n=0}^{\infty} c_n \alpha^{n+1}$. Using the expressions for $f_0(C)$ and $f_1(C)$ computed earlier yields

$$J = \alpha \left[1 - \sum_{i=1}^{\ell} \left(\frac{1}{k_i} + \frac{1}{\gamma_i} \right) \alpha + O(\alpha^2) \right], \quad (10)$$

$$\rho = \frac{1}{L} \sum_{i=1}^L \left(\frac{1}{k_i} + \frac{1}{\gamma_i} \right) \alpha + O(\alpha^2), \quad (11)$$

$$\rho_i = \left(\frac{1}{k_i} + \frac{1}{\gamma_i} \right) \alpha + O(\alpha^2), \quad (12)$$

where $\gamma_i = \gamma + (\beta - \gamma) \delta_{i,L}$. These equations constitute our main result. Equation (10) shows that if the initiation rate α is small compared to k_1, \dots, k_L , and γ , then the protein production rate J depends predominately on the initiation rate, along with the translocation and elongation rates of the first 10 codons, corresponding to the ribosome footprint ℓ . The importance of the first 10 codons is a direct consequence of the excluded volume interactions: any ribosome already present in that region will prevent a new ribosome from binding the mRNA. It is also important to emphasize that (10)–(12) are exact series expansions around $\alpha = 0$; the approximation is made only when the series is truncated.

Independent Particle Approximation (IPA).— Interestingly, the excluded volume interactions between particles have no effect on the first two terms in the series expansion of J . This motivates us to ask how the expansion in Eq. (10) would look like if we assumed that all particles

are independent, i.e., not experiencing any exclusion interaction. In our model, the IPA amounts to replacing $P(C)$ with

$$P^{\text{IPA}}(C) = \frac{1}{Z_L} \prod_{j=1}^{N(C)} w_1(\theta_{X(j)}), \quad (13)$$

where $N(C)$ is the number of particles in a configuration C , θ is one of the two particle states 1 and 2, $X(i)$ is the position of the i th particle on the lattice, and Z_L is the normalization constant. The weights $w_1(1_i)$ and $w_1(2_i)$ for $i = 1, \dots, L$ are obtained by solving the master equation for a single particle, and they are given by α/k_i and α/γ_i , respectively. The corresponding expressions for the current J and local density ρ_i read

$$J^{\text{IPA}} = \frac{\alpha}{\prod_{i=1}^{\ell} (1 + p_i)}, \quad \rho_i^{\text{IPA}} = \frac{p_i}{1 + p_i}, \quad (14)$$

where $p_i = \alpha(1/k_i + 1/\gamma_i)$. We will use these results later in order to determine the importance of ribosome collisions in real genetic sequences. We also note that the IPA in Eq. (13) provides a good approximation to $f_n(C)$ for $n = N(C)$ when the particles in C are far apart from each other.

Application to mRNA translation in yeast.—We now apply our results to the transcriptome of *S. cerevisiae* using realistic model parameters. The values of α in the range $0.005\text{--}4.2 \text{ s}^{-1}$, with the median value of 0.09 s^{-1} , have been previously estimated in Ref. [21] using genome-wide experimental values of the ribosomal density. We assume that the rates k_i are mainly proportional to the gene copy number of tRNAs delivering the corresponding amino acid [29], the rates are normalized so that the average codon translation rate is equal to the experimental value of 10 codons/s [30], and the estimates of all elongation rates, along with the distribution of α , are presented in the Supplemental Material [25]. The translocation rate γ is fixed to $\gamma = 35 \text{ codons/s}$ [31], and termination is assumed to be fast and comparable to translocation [30], $\beta \approx \gamma$, so that $\gamma_i = \gamma \forall i$ in Eqs. (10)–(12).

In total, we analyzed 5836 gene sequences; for each gene we calculated J , ρ , and ρ_i for $1 \leq i \leq L$ up to and including the second order of the perturbative expansion at the corresponding physiological value of α . The results were then compared to the exact values obtained numerically with stochastic simulations using the Gillespie algorithm [32] by calculating the percent error ϵ .

For the protein production rate J , with the zeroth order of the perturbative expansion (predicting that $J = \alpha$), we obtain an error of $\epsilon < 5\%$ for only 11% of the genes. Remarkably, that percentage jumps to 80.7% when the first-order coefficients in Eq. (6) are taken into account (Fig. 2). Including the second-order coefficients [computed numerically from Eq. (4)] does not significantly improve results, due to a large value of $\alpha \gtrsim 0.15 \text{ s}^{-1}$ in about 20% of

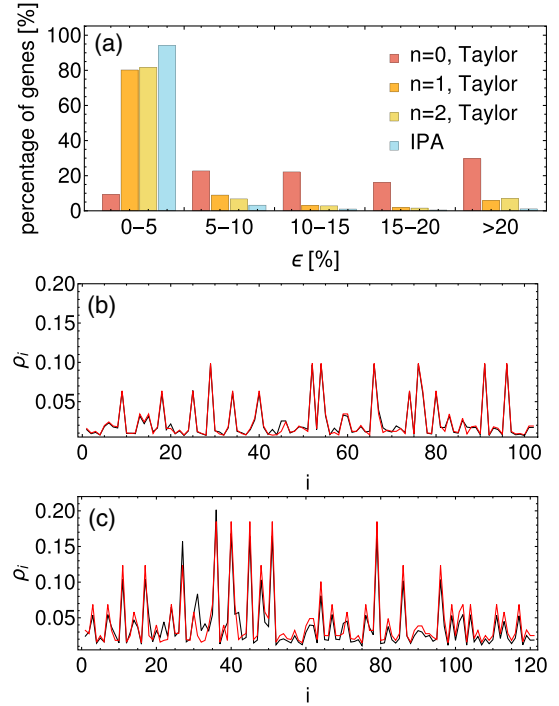


FIG. 2. (a): A histogram of the percent error ϵ measuring the discrepancy between the protein production rate obtained by stochastic simulations and the perturbative expansion including the independent-particle approximation (IPA). (b) and (c): Ribosomal density profile obtained by stochastic simulations (black) compared to Eq. (12) (red) for two values of the initiation rate α , one close to the median value $\approx 0.09 \text{ s}^{-1}$ (gene YAL045C, b) and the other that is close to the 90th percentile value $\approx 0.23 \text{ s}^{-1}$ (gene YGL034C, c).

the genes. Since the coefficients in Eq. (7) typically alternate in sign, truncating the series will ultimately lead to a wrong result when the value of α is large enough. On the other hand, the IPA does not suffer from this problem and leads to an error $\epsilon < 5\%$ in 94% of genes, whereby only 1% of genes have $\epsilon > 20\%$. The success of the IPA also suggests that ribosome collisions and traffic jams have a minor effect on the rate of translation, which is in accordance with recent experimental evidence [27,33]. This is also apparent from the density profile ρ_i , which is very well approximated by the linear approximation in Eq. (12), even for larger values of α [Figs. 2(b) and 2(c)].

Identifying determinants of mRNA translation.—Our analytic prediction allows us to decompose the contributions from initiation and elongation to the rate of translation J , thereby addressing a long-standing question about the main determinants of protein production rate. Remarkably, the expressions for the current J , obtained with both the first-order and the independent particle approximation, involve only the first $\ell = 10$ codons. This result therefore strongly indicates that, together with the initiation rate, the first 10 codons of the mRNA are the key determinants of the protein production rate by preventing a new ribosome

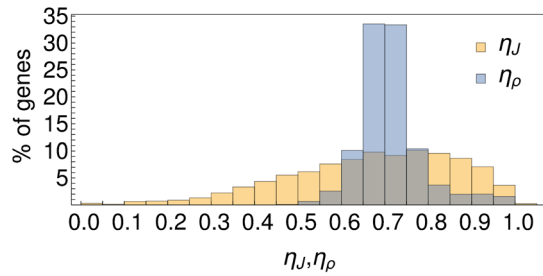


FIG. 3. Histogram of η_J (yellow) and η_ρ (blue) for the *S. cerevisiae* genome.

from binding, which effectively decreases the initiation rate.

If we assume that the cell maximizes the rate of protein production, given the above result, we would expect to find a signature in the genome for selecting efficient *fast* codons at the beginning of each gene. We test this hypothesis by computing the ribosomal current for the fastest (J^F) and the slowest (J^S) set of first ℓ synonymous codons for each gene, using the IPA. A score $\eta_J = (J - J^S)/(J^F - J^S)$ is then assigned to each gene, which is 1 (0) when the sequence corresponds to the fastest (slowest) codon sequence. On the other hand, one might assume that the cell not only tries to maximize protein production rates, but at the same time, it tries to minimize the ribosome density ρ on mRNAs. This assumption is motivated by the fact that ribosomes are limiting [22,34] and highly costly in terms of cellular energy resources [35], and therefore, ribosome queues are to be avoided. Hence, we also compute $\eta_\rho = (\rho - \rho^S)/(\rho^F - \rho^S)$ for each gene, where ρ^F and ρ^S denote the ribosome density for the fastest and slowest set of synonymous codons, respectively.

Figure 3 shows the histogram of η_J (yellow) and η_ρ (blue) computed for 5836 genes of *S. cerevisiae*. Both histograms show an average of 0.7, suggesting the selection of fast codons near the start codon to maximize J , as well as an overall selection of fast codons along the mRNA to minimize ρ . However, the width of the distribution of η_ρ is substantially smaller than the one of η_J . This might indicate that the optimization of the protein production rate is strongly dependent on the particular gene since different proteins are needed at different concentrations. In contrast, the minimization of the number of ribosomes on mRNAs could be a more general constraint.

Conclusions.—We have presented an analytic method that allows us to quantify sequence determinants of protein production rates, using a model for translation that is based on an inhomogeneous exclusion process. Our results demonstrate that the rate of protein production is largely determined by the initiation rate and the elongation rates of the first 10 codons (the ribosome footprint length on the mRNA), which control how fast ribosomes load onto the mRNA, whereby ribosome collisions and queues have a minor effect under physiological conditions.

M. C. R. and L. C. contributed equally to this work. J. S. N. was supported by the Leverhulme Trust Early Career Fellowship. M. C. R. was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) BB/N017161/1 and the Scottish Universities Life Sciences Alliance. L. C. thanks the CNRS for having granted him a demi-délégation.

*jszavits@staffmail.ed.ac.uk

- [1] H. Gingold and Y. Pilpel, *Mol. Syst. Biol.* **7**, 481 (2011).
- [2] A. J. Kemp, R. Betney, L. Ciandrini, A. C. Schwenger, M. C. Romano, and I. Stansfield, *Mol. Microbiol.* **87**, 284 (2013).
- [3] D. Chu, E. Kazana, N. Bellanger, T. Singh, M. F. Tuite, and T. von der Haar, *EMBO J.* **33**, 21 (2014).
- [4] B. Gorgoni, L. Ciandrini, M. R. McFarland, M. C. Romano, and I. Stansfield, *Nucleic Acids Res.* **44**, 9231 (2016).
- [5] S. Varenne, J. Buc, R. Llobes, and C. Lazdunski, *J. Mol. Biol.* **180**, 549 (1984).
- [6] M. A. Sorensen, C. G. Kurland, and S. Pedersen, *J. Mol. Biol.* **207**, 365 (1989).
- [7] T. E. Goroehowski, I. Avciilar-Kucukgoze, R. A. Bovenberg, J. A. Roubos, and Z. Ignatova, *ACS Synth. Biol.* **5**, 710 (2016).
- [8] G. Boël *et al.*, *Nature (London)* **529**, 358 (2016).
- [9] C. T. MacDonald, J. H. Gibbs, and A. C. Pipkin, *Biopolymers* **6**, 1 (1968); C. T. MacDonald and J. H. Gibbs, *Biopolymers* **7**, 707 (1969).
- [10] H. Zur and T. Tuller, *Nucleic Acids Res.* **44**, 9031 (2016).
- [11] B. Derrida, M. R. Evans, V. Hakim, and V. Pasquier, *J. Phys. A* **26**, 1493 (1993); G. M. Schütz and E. Domany, *J. Stat. Phys.* **72**, 277 (1993).
- [12] R. A. Blythe and M. R. Evans, *J. Phys. A* **40**, R333 (2007).
- [13] T. Chou and G. Lakatos, *Phys. Rev. Lett.* **93**, 198101 (2004).
- [14] L. B. Shaw, J. P. Sethna, and K. H. Lee, *Phys. Rev. E* **70**, 021901 (2004).
- [15] R. K. P. Zia, J. J. Dong, and B. Schmittmann, *J. Stat. Phys.* **144**, 405 (2011).
- [16] J. Schmidt, V. Popkov, and A. Schadschneider, *Europhys. Lett.* **110**, 20008 (2015).
- [17] N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, and J. S. Weissman, *Science* **324**, 218 (2009).
- [18] M. A. Gilchrist and A. Wagner, *J. Theor. Biol.* **239**, 417 (2006).
- [19] N. Mitarai, K. Sneppen, and S. Pedersen, *J. Mol. Biol.* **382**, 236 (2008).
- [20] S. Reuveni, I. Meilijson, M. Kupiec, E. Ruppim, and T. Tuller, *PLoS Comput. Biol.* **7**, e1002127 (2011).
- [21] L. Ciandrini, I. Stansfield, and M. C. Romano, *PLoS Comput. Biol.* **9**, e1002866 (2013).
- [22] P. Shah, Y. Ding, M. Niemczyk, G. Kudla, and J. B. Plotkin, *Cell* **153**, 1589 (2013).
- [23] L. Ciandrini, I. Stansfield, and M. C. Romano, *Phys. Rev. E* **81**, 051904 (2010).
- [24] S. Klumpp, Y. Chai, and R. Lipowsky, *Phys. Rev. E* **78**, 041909 (2008).

- [25] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.120.128101> for derivation of the exact solution of a stationary master equation, proof that many $f^{(n)}(C) = 0$, second-order equations, estimates of the codon elongation rates and distribution of translation initiation rates for the yeast genome.
- [26] J. Szavits-Nossan, *J. Phys. A* **46**, 315001 (2013).
- [27] G. Kudla, A. W. Murray, D. Tollervey, and J. B. Plotkin, *Science* **324**, 255 (2009).
- [28] H. M. Salis, E. A. Mirsky, and C. A. Voigt, *Nat. Biotechnol.* **27**, 946 (2009).
- [29] R. Percudani, A. Pavesi, and S. Ottonello, *J. Mol. Biol.* **268**, 322 (1997).
- [30] Y. Arava, Y. Wang, J. D. Storey, C. L. Liu, P. O. Brown, and D. Herschlag, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3889 (2003).
- [31] A. Savelsbergh, V. I. Katunin, D. Mohr, F. Peske, M. V. Rodnina, and W. Wintermeyer, *Mol. Cell* **11**, 1517 (2003).
- [32] D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
- [33] B. S. Laursen, H. P. Sørensen, K. K. Mortensen, and H. U. Sperling-Petersen, *Microbiol. Mol. Biol. Rev.* **69**, 101 (2005).
- [34] P. Greulich, L. Ciandrini, R. J. Allen, and M. C. Romano, *Phys. Rev. E* **85**, 011142 (2012).
- [35] K. C. Keiler, *Nat. Rev. Microbiol.* **13**, 285 (2015).