Queueing Phase Transition: Theory of Translation

M. Carmen Romano,[1,2] Marco Thiel,[1] Ian Stansfield,[2] and Celso Grebogi[1]

[1]*Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen, AB24 3UE, United Kingdom*
[2]*Institute of Medical Sciences, Foresterhill, University of Aberdeen, Aberdeen, AB25 2ZD, United Kingdom*

(Received 21 January 2009; published 15 May 2009)

We study the current of particles on a lattice, where to each site a different hopping probability has been associated and the particles can move only in one direction. We show that the queueing of the particles behind a slow site can lead to a first-order phase transition, and derive analytical expressions for the configuration of slow sites for this to happen. We apply this stochastic model to describe the translation of mRNAs. We show that the first-order phase transition, uncovered in this work, is the process responsible for the classification of the proteins having different biological functions.

DOI: 10.1103/PhysRevLett.102.198104 PACS numbers: 87.10.−e, 05.45.−a

Nonequilibrium statistical physics is a main subject of much current research. Within nonequilibrium statistical physics, the totally asymmetric exclusion process plays a paradigmatic role [1]. It describes a driven lattice gas in one dimension, in which particles are injected with a probability α, then hop from one lattice site to the next with certain probability p, until they reach the end of the lattice, where they are then released at a rate β. The totally asymmetric exclusion process has been intensively studied not only to describe physical driven diffusive processes, but also biological processes such as inhomogeneous growth [2] and translation [3–5].

In this Letter, we generalize the totally asymmetric exclusion process, associating to every lattice site i a different hopping probability p_i. Additionally, we consider a sequential rather than a random update rule for the reasons stated later. We show that, depending on the configuration of the hopping probabilities in the lattice, the current of particles can undergo a phase transition of first order. This phase transition, as it will be shown later, has direct experimental consequences. We derive analytical expressions that yield the necessary conditions for the configuration to have a first-order phase transition. In the absence of clusters of slow sites (sites with a low hopping probability), the position of the slowest site in the lattice determines whether the current will be subject to a phase transition. In the presence of clusters of slow sites, the situation is more complex: even though the sites constituting one cluster are faster than the slowest site of the lattice, the cluster can become rate limiting for the current of particles, depending on the size of the cluster and the ratio between $p_{\min}$ and q. Here $p_{\min}$ denotes the smallest hopping probability and q the hopping probability of each of the sites forming the cluster. Crucially, the position in the lattice of the site or sites constituting the rate limiting step determines the existence of a first-order phase transition.

The numerical implementation of the generalized totally asymmetric exclusion process is given by the following two rules: (i) advance with probability p_i if site $i+1$ is free; (ii) wait if site $i+1$ is occupied. The positions of the particles are updated at every Monte Carlo time step in sequential order, starting from the rightmost site. This update rule is dictated by the experimental process that led to this work, as argued below. By iterating these two rules we compute the current J of particles, i.e., how many particles per unit time hop from one site to the next, for lattices of arbitrary length and involving slow sites configurations. Figure 1(a) shows J versus the initiation rate α for four different configurations of slow sites: equally spaced, randomly distributed, four clusters, and one-cluster. The value of J for a certain value of α depends strongly on the configuration of slow sites; sequences of the same length and the same number of fast and slow sites, reach a different value of J depending on the position of the slow sites. The difference in J for large values of α between the configuration in which the slow sites are equally spaced and the one in which all slow sites are forming one single cluster is about threefold. Hence, the optimal configuration [4] for the current is achieved when the distance between slow sites is maximized.

Importantly, also the rate of change of J with respect to α depends on the configuration of slow sites. We obtain two qualitatively different types of behavior: for the randomly distributed slow sites and for the one-cluster configuration, there is a discontinuity in the derivative of J with respect to α (type-I), whereas the other two configurations (equally spaced slow sites and four-cluster configuration) show an overall smooth dependence of J versus α (type-II) [Fig. 1(a)]. The cause for the different types of transitions lies in the position of the slow sites with respect to the beginning of the lattice. In the sequence with slow sites positioned at regular distances, and in the sequence with four clusters of slow sites, the first slow site is at position $i = 1$. By contrast, for both the random and the one-cluster configuration, the first slow site is at position $i > 1$. If the slow sites that are rate limiting are at the beginning of the lattice, it is not possible to form a queue of particles. Conversely, if the rate limiting slow sites are at position $i > 1$, as soon as the initiation rate increases beyond a critical value, a queue can form. This leads to

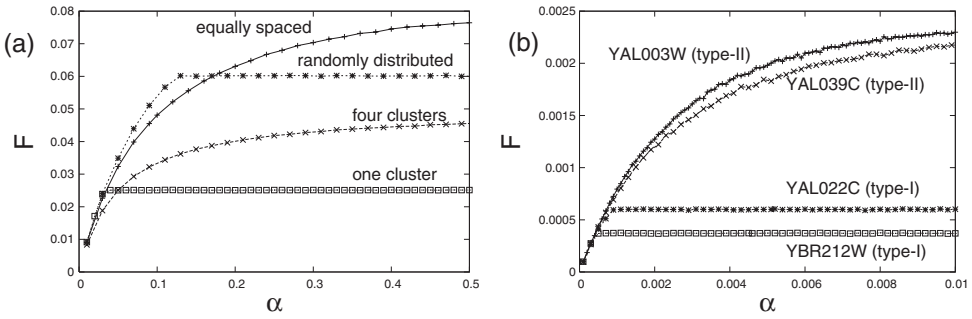

FIG. 1. Current of particles relative to the initiation rate α: (a) four different configurations of a lattice containing 1% of rare codons, and (b) four real mRNAs of baker's yeast.

an abrupt increase of the mean density, which is reflected by a discontinuity of the rate of change of the current with respect to α.

The queueing theory developed in this work is motivated by the biological process of translation, in which a protein is synthesized using a messenger RNA (mRNA) molecule as a template. Our theory allows us to classify the natural process of protein production into two large and fundamental classes depending on whether the translation undergoes the phase transition. In the queueing theory, the lattice corresponds to the mRNA, and each site i represents one codon specifying one amino acid. The particles that hop from one site to the next are ribosomes, molecular machines that carry out the translation. The initiation rate α gives the probability with which new ribosomes start the translation. Every site i has a different probability value p_i associated with it, determined by the concentration of the corresponding transfer RNA (tRNA) molecule in the cytoplasm; these are the molecules that bring the amino acids to the ribosomes, so that the sequence of codons—the mRNA—can be translated into a sequence of amino acids—the protein. A codon is commonly referred to as ''rare'' if it is translated by a low abundance tRNA. Hence, a rare codon represents a slow site on the lattice. The current of ribosomes on one specific mRNA yields the translation rate or the number of proteins per unit time being produced. Each elongation step consists of two main events: (i) finding the correct tRNA molecule, and (ii) formation of the peptide bond (transpeptidase reaction) and downstream movement of the ribosome by one codon (translocation reaction). Once the ribosome has recruited the correct tRNA, it remains charged. Crucially, the time needed for event (i) is much larger than the time needed for event (ii) [6]. Both transpeptidase and translocation reactions are almost instantaneous, in fact too fast to be measured [6]. Hence, once the ribosome finds the correct tRNA, the time that the ribosome needs to move to the next codon is negligible. Therefore, if two ribosomes occupying two adjacent codons find their respective tRNAs within the same Monte Carlo time step, both of them will move forward. This is precisely implemented by the ordered sequential update rule used in this work [7,8].

Our theory predicts two main types of mRNAs purely based on the dynamics of the ribosomes: type-I mRNAs, that can undergo a first-order phase transition, and type-II mRNAs, in which the phase transition does not occur. To test the prediction of this theory, we use experimental data for the distribution and configuration of the ''hopping'' probabilities p_i of ribosomes from 500 yeast mRNA sequences [9]. In Fig. 1(b) we show J depending on α for four real mRNAs [10]. We can clearly classify them into type-I and type-II sequences. Figures 2(a) and 2(b) show the density profiles (mean occupancy time at each codon) for two representative yeast mRNAs of type-I and type-II, respectively, for a value of α close to one, so that both mRNAs carry the maximal number of ribosomes that they can hold. The ribosomal ''traffic jam'' in the type-I sequence is very pronounced and it is caused by the cluster of

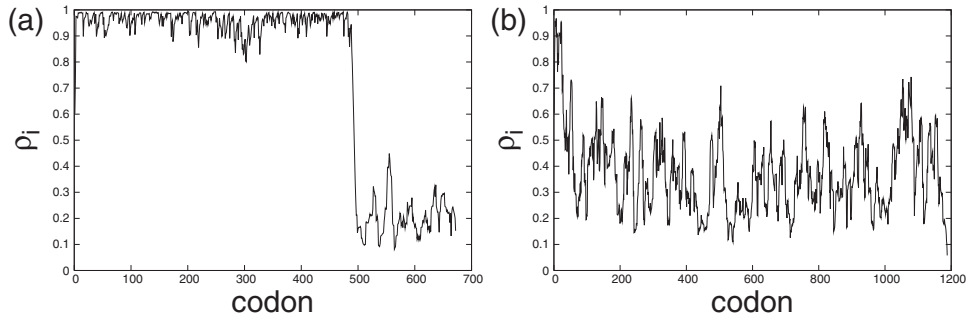

FIG. 2. Density profiles ρ_i or mean occupancy times at each codon for two real yeast proteins: (a) type-I (YAR042W), and (b) type-II (YBR210W).

rare codons (with the lowest possible value of p_i) situated at $i = 485, \ldots, 493$. This queue in front of the cluster of rare codons is a consequence of the first-order phase transition. Conversely, the density profile for the type-II sequence is more homogeneous in the sense that it does not present any appreciable traffic jam, indicating that J depends smoothly on α.

The key result regarding the translation process is that our classification purely according to the phase transition matches the *biological function* of the proteins encoded by type-I and type-II mRNAs. All ribosomal proteins (proteins that form part of ribosomes) are classified exclusively as type-II mRNAs (no phase transition), whereas all type-I mRNAs (first-order phase transition) translate into nonribosomal ones. A rapidly dividing cell contains proportionately more ribosomes than a slow-growing cell. Hence, during periods of rapid growth, the rate of ribosome synthesis is high. Therefore, it will be advantageous for ribosomal protein synthesis to be able to respond to increased numbers of free ribosomes (representing higher values of α). This implies a smooth dependence of the translation rate with α, i.e., a type-II response. By contrast, nonribosomal proteins whose mRNAs exhibit a type-I response will be translated at a constant rate once α increases past a critical value. This will lead to a noncontinuous change of the current of ribosomes with respect to α, i.e., a first-order phase transition. The positioning of rare codons is thus able to insulate type-I sequences from variations in the value of α, possibly a cellular mechanism to reduce noise in gene expression. This constitutes a strong validation of our theoretical results, showing that the specific configuration of slow codons in mRNAs constitutes a regulatory mechanism by which the translation machinery responds differently to a high value of α in different types of proteins.

To show analytically the origin of the first-order phase transition of the current with respect to the initiation rate, we introduce a generalized totally asymmetric exclusion process with a sequential update rule defined by the following exact equations

$$u_i = p_i P_{i,i+1}(\underline{\bullet}, \circ) + p_i P_{i,i+1}(\underline{\bullet}, \bullet) u_{i+1|i}, \qquad (1)$$

where u_i denotes the effective probability that the particle at site i moves to site $i+1$, $P_{i,i+1}(\underline{\bullet}, \circ)$ is the conditional probability that site $i+1$ is free given that site i is occupied, $P_{i,i+1}(\underline{\bullet}, \bullet)$ is the conditional probability that site $i+1$ is occupied given that site i is also occupied, and $u_{i+1|i}$ is the effective conditional probability that the particle at site $i+1$ advances one step if site i is occupied. The first term accounts for the situation in which the next site is free and the particle hops; the second term accounts for the situation in which the next particle is occupied, but the particle occupying it also moves, so that the particle on site i can also advance. A comparison of the solution obtained from Eq. (1) with the one obtained with the numerical implementation of rules (i) and (ii) shows that they agree.

By applying the mean field approximation, i.e., neglecting correlations between neighboring sites, the analysis of Eq. (1) becomes much more tractable. Equation (1) is then approximated as follows

$$u_i = p_i(1 - \rho_{i+1}) + p_i \rho_{i+1} u_{i+1}, \qquad i = 1, \ldots, N, \quad (2)$$

where ρ_i is the probability that the site i is occupied. Even though there are some deviations with respect to the numerical implementation of the rules (i) and (ii), the qualitative behavior is reproduced correctly. Once the steady state has been reached, the current is conserved along the sequence, i.e., $u_i \rho_i = u_{i+1} \rho_{i+1} = J, \forall\ i$, since premature termination events are not included in the process [11]. A particle on the last site is not hindered by any other particle and hence, $u_N = p_N$. Also note that the initiation rate α is equivalent to having a reservoir of particles ready to hop, i.e., $p_0 = \alpha$ and $\rho_0 = 1$, so that $u_0 = J$. Fixing a value for the current J and iterating Eq. (2) backwards by starting from site N, we calculate at each step the values of u_i and ρ_i until we reach the first site. We then calculate the initiation rate α which corresponds to the fixed value of J by using $J = u_0 = \alpha(1 - \rho_1) + \alpha \rho_1 u_1$. This leads to a polynomial of Nth order in J with coefficients that depend on $p_1, \ldots, p_N$ and α. We consider three main cases for the configuration of slow sites on the lattice, in which we obtain the polynomial in J by finding closed forms for the recurrence relations [Eq. (2)] using homographic functions.

Case A: We assume that there are M slow sites with hopping probabilities $\{q_1, \ldots, q_M\}$, where q_1 is at position $i > 1$. We further assume that the sites of the lattice between the slow ones are fast; i.e., the first site is also fast. The distances N_i between two slow sites q_{i-1} and q_i are assumed to be large enough, so that we can neglect terms of the order of N_i or higher in the polynomial of J. For the fast sites, we choose for simplicity hopping probability $p = 1$. In this case, we obtain a polynomial in J with the following roots: $\alpha, q_1, q_2, \ldots, q_M$. Combining these solutions with the fact that the current J is limited by the slowest site of the lattice [12], we obtain the following solution for the current: $J = \alpha$, if $\alpha < \alpha_c$, and $J = q_{\min}$, otherwise, where $q_{\min} = \min\{q_1, \ldots, q_M\}$. Hence, the derivative of J with respect to α has a discontinuity at the critical value $\alpha_c = q_{\min}$, indicating a first-order phase transition [Fig. 3(a)]. This transition is accompanied by a sudden increase of the mean density of particles on the lattice, and a sudden decrease in the mean speed of the particles. If $i_{\min}$ indicates the position of the slowest site of the lattice, then for $\alpha > \alpha_c$ the region $i = 1, \ldots, i_{\min}$ is completely occupied,

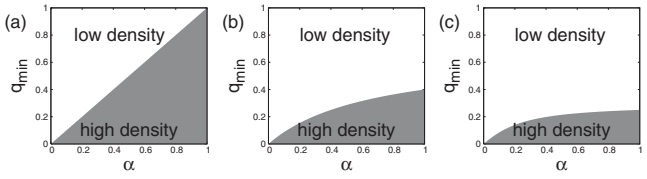

FIG. 3. Phase transition diagrams: (a) case A, (b) case B, (c) case C with $q_{\min} = q_2$. In figures (b) and (c), $q_1 = 0.4$.

whereas the region $i = i_{\min} + 1, \ldots, N$ has a low density [Fig. 2(a)].

Case B: We consider the same configuration as before, but now the first site of the lattice is slow; i.e., q_1 is at position $i = 1$. In this situation we obtain two different subcases. (i) If $q_1 = q_{\min}$, then $J = \frac{\alpha q_1}{\alpha + q_1 - \alpha q_1}$, for $0 \leq \alpha \leq 1$. In this case the dependence of J on α is smooth and there is no phase transition in the current of particles. (ii) If by contrast, $q_1 > q_{\min}$, then $J = \frac{\alpha q_1}{\alpha + q_1 - \alpha q_1}$, if $\alpha < \alpha_c$, and $J = q_{\min}$, otherwise, where $\alpha_c = (q_1 q_{\min})/(q_1 - q_{\min} + q_1 q_{\min})$. In this case, we have a first-order phase transition, because the rate limiting step is not the first site of the lattice [Fig. 3(b)].

Case C: A more complicated situation arises when the slow sites are not isolated but appear in clusters. Assume that we have a cluster of size S at the beginning of the sequence, i.e., $i = 1, \ldots, S$, where the sites have hopping probability q_1. At a large distance from that cluster, we have an isolated slow site with hopping probability q_2. Even in the case where $q_2 < q_1$, the slow cluster at the beginning might constitute the rate limiting step for the particles. Applying the procedure outlined above, we obtain the following polynomial:

$$\begin{aligned}
&\gamma^S(u_+ u_- \alpha q_2 - u_- \alpha q_2) + u_+(1 - u_-)\alpha q_2 \\
&+ [\gamma^S(\alpha q_2 + u_+ u_- \alpha(q_2 - 1) + u_- q_2 - \alpha u_+ q_2 + u_- \alpha \\
&- u_+ u_- q_2 - u_- \alpha q_2) + u_+ u_-(q_2 + \alpha - \alpha q_2) \\
&+ u_+(\alpha q_2 - q_2 - \alpha) + \alpha q_2(u_- - 1)]J \\
&+ [\gamma^S(u_+ u_-(1 - \alpha) + \alpha u_+ + u_-(\alpha - 1) - \alpha) \\
&\quad + u_+ u_-(\alpha - 1) + u_+(1 - \alpha) - u_- \alpha + \alpha]J^2 = 0,
\end{aligned}$$

where $u_\pm = \frac{1}{2}[q_1 + q_1 J \pm \sqrt{q_1(q_1 + 2q_1 J + q_1 J^2 - 4J)}]$ are the fixed points of the map $u_i = q_1(1 - J/u_{i+1} + J)$ and $\gamma = \frac{-2u_- - q_1 J + J(u_+ - u_-) + J(4 - q_1)}{2u_+ + q_1 J + J(u_+ - u_-) - J(4 - q_1)}$. To exemplify the effect of the cluster of slow sites at the beginning of the lattice, we discuss the solutions for $S = 2$: (i) if $q_2 < q_1/(2 - q_1)$, then

$$J = \begin{cases} \frac{1}{2}\frac{-2\alpha q_1 + 2\alpha + q_1 - \sqrt{-4\alpha^2 q_1 + 4\alpha^2 + q_1^2}}{\alpha q_1 - \alpha - q_1 + 1} & \text{if } \alpha < \alpha_c, \\ q_2, & \text{otherwise,} \end{cases}$$

where $\alpha_c = [q_2(q_1 + q_2 q_1 - q_2)]/(q_2^2 q_1 + 2q_2 q_1 + q_1 - q_2^2 - 2q_2)$. This means that for $\alpha > \alpha_c$, the single slow site determines the current of the particles in spite of the presence of the cluster. In this case, we have again a discontinuity in $\partial J/\partial \alpha$, indicating a first-order phase transition [Fig. 3(c)]. (ii) By contrast, if $q_2 > q_1/(2 - q_1)$, then we obtain $J = \frac{1}{2}\frac{-2\alpha q_1 + 2\alpha + q_1 - \sqrt{-4\alpha^2 q_1 + 4\alpha^2 + q_1^2}}{\alpha q_1 - \alpha - q_1 + 1}$ for $0 < \alpha < 1$. Therefore, in the latter case, even though $q_2 < q_1$, the cluster determines the current of the particles and J depends smoothly on α.

In conclusion, we have shown that, depending on the specific configuration of slow sites in the queueing process, the current of particles can undergo a first-order phase transition. We have derived analytical expressions for the conditions that the configuration must obey in order to do so, and have validated our theoretical results with experimental data of the concentration of tRNA molecules in *S. cerevisiae*. Our theory predicts the classification of mRNA sequences into two fundamental types purely based on the dynamics of the ribosome traffic, and this classification matches perfectly the biological function of each type, providing thus the link between the first-order phase transition and the biology. This first-order phase transition is manifestly the key factor distinguishing the way in which the translation of the different types of mRNA responds to ribosome availability. Our results are not only relevant for nonequilibrium statistical physics, but also find applications in synthetic biology since, by rearranging synonymous codons, the translatability of a given mRNA under different conditions of free ribosome availability can be controlled.

The authors thank E. Wilson and D. Smith for helpful discussions and SULSA; BBSRC [BB/F00513X/1 and BB/G010722]; NRP; and RCUK for financial support.

[1] B. Derrida, E. Domany, and D. Mukamel, J. Stat. Phys. **69**, 667 (1992); M. R. Evans, J. Phys. A **30**, 5669 (1997); J. Krug, Phys. Rev. Lett. **67**, 1882 (1991); R. J. Harris and R. B. Stinchcombe, Phys. Rev. E **70**, 016108 (2004); G. Tripathy and M. Barma, Phys. Rev. E **58**, 1911 (1998).

[2] M. Kardar, G. Parisi, and Y.-C. Zhang, Phys. Rev. Lett. **56**, 889 (1986).

[3] L. B. Shaw, R. K. P. Zia, and K. H. Lee, Phys. Rev. E **68**, 021910 (2003).

[4] T. Chou and G. Lakatos, Phys. Rev. Lett. **93**, 198101 (2004).

[5] J. J. Dong, B. Schmittmann, and R. K. P. Zia, Phys. Rev. E **76**, 051113 (2007).

[6] S. Varenne *et al.*, J. Mol. Biol. **180**, 549 (1984); S. Schilling-Bartetzko, A. Bartetzko, and K. H. Nierhaus, J. Biol. Chem. **267**, 4703 (1992); N. Bilgin *et al.*, Biochimie **70**, 611 (1988); T. Pape, W. Wintermeyer, and M. V. Rodnina, EMBO J. **17**, 7490 (1998).

[7] D. A. Smith, and R. E. Wilson, J. Phys. A **40**, 2651 (2007).

[8] Because of the different update rule, we have a transcritical bifurcation, instead of a saddle-node bifurcation for Eq. (2). This leads only to the existence of a first-order phase transition, and not of a second-order one for $p_i = 1$.

[9] These probabilities were estimated as $p_i = (\mathrm{GCN})_i / \sum_{i=1}^{T} (\mathrm{GCN})_i$, where $(\mathrm{GCN})_i$ is the gene copy number of the ith tRNA and T is the total number of different tRNAs. The abundance of tRNAs is known to be correlated with tRNA gene copy numbers in yeast [R. Percudani, A. Pavesi, and S. Ottonello, J. Mol. Biol. **268**, 322 (1997)].

[10] In these computations we have taken into consideration that a ribosome occupies 9 codons [S. Wolin and P. Walter, EMBO J. **7**, 3559 (1988)].

[11] L. Rajkowitsch *et al.*, J. Mol. Biol. **335**, 71 (2004).

[12] Note that $J = u_i \rho_i \leq p_i \rho_i \leq p_i$.