

Clustered Bottlenecks in mRNA Translation and Protein Synthesis

Tom Chou¹ and Greg Lakatos²

¹Department of Biomathematics, UCLA, Los Angeles, California 90095, USA

²Department of Physics, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

(Received 23 October 2003; published 1 November 2004)

Using a model based on the totally asymmetric exclusion process, we investigate the effects of slow codons along messenger RNA. Ribosome density profiles near neighboring clusters of slow codons interact, enhancing suppression of ribosome throughput when such bottlenecks are closely spaced. Increasing the slow codon cluster size beyond ~ 3 –4 codons does not significantly reduce the ribosome current. Our results are verified by both extensive Monte Carlo simulations and numerical calculation, and provide a biologically motivated explanation for the experimentally observed clustering of low-usage codons.

DOI: 10.1103/PhysRevLett.93.198101

PACS numbers: 87.16.Ac, 05.10.-a, 87.10.+e

During protein synthesis, ribosome molecules initiate at the 5' end of messenger RNA (mRNA), scan (“elongate”) along the mRNA sequence, and terminate with the completed protein product at the 3' termination end [Fig. 1(a)]. Each elongation step requires reading (translating) a nucleotide triplet (codon) and the binding of a freely diffusing transfer RNA (tRNA) molecule carrying the amino acid specific to that codon [1]. Besides being a critical final stage of gene expression *in vivo*, control of protein synthesis is vital for protein adaptation and evolution [2–4], for the control of viral parasitism [5], and for high yield, cell-free, synthetic *in vitro* protein production [6].

Protein expression can be regulated by exploiting relative concentrations of tRNA in cytoplasm that determine local ribosome translation rates. For example, viruses have no transcriptional machinery and can only regulate protein production by utilizing existing host tRNA abundances [7] or by insertion or deletion of specific codons. Technologically, tRNA abundances have also been exploited by modifying mRNA codons to those preferred in mammalian systems, thereby optimizing expression levels of green fluorescent protein [8].

“Slow” codons (those with rare corresponding tRNA and/or amino acids) along mRNA are known to inhibit protein production [9,10]. Such “bottleneck” or “defect” codons typically include CTA (Leu), ATA (Ile), ACA (Thr), CCT and CCC (Pro), CGG, AGA, and AGG (Arg), but arise infrequently (about 4% in *E. coli*) [11,12]. Slow codons can appear throughout the mRNA, at or near the initiation site, termination site, and/or in the interior “elongation” region [13]. Statistics indicate a higher occurrence of rare codons near the 5' initiation site of *E. coli* genes [14]. Even more striking is the proclivity for rare codons to cluster [11]. Rare codon clusters (~ 2 –5) occur frequently in *E. coli*, *Drosophila*, yeast, and primates [11].

Although the strength, number, and positioning of bottlenecks can affect local ribosome densities and overall translation rates, there has been no quantitative model

describing how various bottleneck motifs control ribosome throughput and protein synthesis. In this Letter, we consider a simple physical model of how specific codon usages that give rise to local delays in elongation can be used to suppress protein synthesis. We develop a new algorithm that allows accurate and fast numerical calculation of the steady-state ribosome throughput. Within a nonequilibrium stochastic model, our results quantify the biological effects of slow codon clustering.

We model mRNA translation by ribosome particles using a nonequilibrium totally asymmetric exclusion process (TASEP) [15,16] with a few carefully distributed slow sites [cf. Fig. 1(b)]. In the TASEP, ribosome particles attach (initiate) at the first lattice site with rate α , only if the first site is empty. Interior ribosome particles can move forward with rate p_i from site i to site $i + 1$ only if site $i + 1$ is empty. For each step a ribosome moves forward, a codon is read, and an amino-acid-carrying

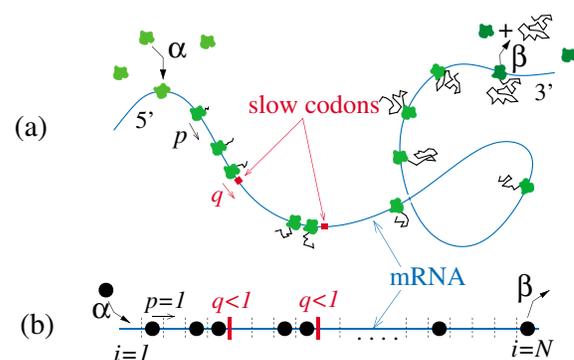


FIG. 1 (color). (a) The mRNA translation/protein synthesis process. Ribosomes move unidirectionally along mRNA as tRNA (not shown) deliver the appropriate amino acid to the growing chain. Codons with low concentrations of corresponding tRNA result in bottlenecks that locally suppress ribosome motion across it. (b) A simple totally asymmetric exclusion process on N lattice sites used to model mRNA translation in the presence of slow codons bottlenecks or defects.

tRNA delivers its amino acid to the growing polypeptide chain. No motion is allowed if the site in front of a particular ribosome is occupied. Each ribosome that reaches the last site $i = N$ (the 3' termination site) has polymerized a complete protein and detaches with rate β .

In the case of *uniform* $p_i = 1$, the protein production rate (e.g., the steady-state ribosome current) and ribosome density along the mRNA are known exactly in terms of α and β [15,16]. In the long chain ($N \rightarrow \infty$) limit, the steady-state results reduce to simple forms illustrating the fundamental physical regimes. The current may be entry-rate limited ($\alpha < 1/2$, $\alpha < \beta$), where the ribosome density is low and the steady-state current $J(\alpha) = \alpha(1 - \alpha)$ depends only on α . If β is sufficiently small ($\beta < 1/2$, $\beta < \alpha$), the density is high and the current $J(\beta) = \beta(1 - \beta)$ is a function of only the rate-limiting exit step. When both $\alpha, \beta > 1/2$, the rate-limiting processes are the uniform internal hopping rates, and $J = 1/4$. For typical mRNA ($N \sim 100$ – 1000 codons), these simple analytic forms for J are extremely accurate. We shall restrict our subsequent analyses to the $N \rightarrow \infty$ limit.

There is no general theory for computing steady-state particle (ribosome) currents when the internal hopping rates p_i vary with lattice position i . Generally, one solves 2^N system of equations for all the configurational probabilities. However, determining the sparse $2^N \times 2^N$ transition matrix is time-consuming as it requires testing all configurations for connectivity. However, specific motifs $\{p_i\}$ (such as isolated defects [17] and periodic variations) can be treated with approximations and simulation. Since slow codons are also rare, with typical probabilities of 0.03, we will consider simple configurations of a few, identical bottlenecks (with hopping rate $q < 1$) distributed within an otherwise uniform (with rate 1) lattice. Finite numbers of “fast” defects ($q > 1$) do not affect steady-state currents since the rate-limiting hops $p_i = 1$ dominate the lattice. Figures 2(a)–2(c) show hypothetical placements of defects near the 5' initiation end, the mRNA interior, and the 3' termination end. Within a finite-segment (of length $n = 5, 6, 4$ sites in Fig. 2, respectively) straddling the bottlenecks, we explicitly enumerate all 2^n distinct states according to the algorithm indicated in Fig. 3. This generates a *banded* (of width 2^{n-1}) $2^n \times 2^n$ matrix coupling the probabilities P_j ($1 \leq j \leq 2^n$) that the segment is in state j .

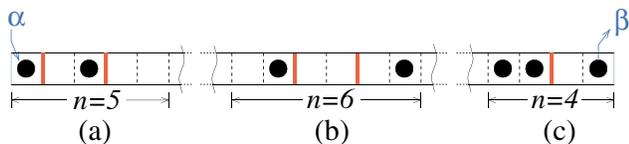


FIG. 2 (color). Placements of slow defects, or rare-usage codons (thick red segments). (a) Two defects near the initiation site straddled by an $n = 5$ lattice segment. (b) Two defects in the chain interior, away from the boundaries, $n = 6$. (c) A single slow defect near the termination end of the chain, $n = 4$.

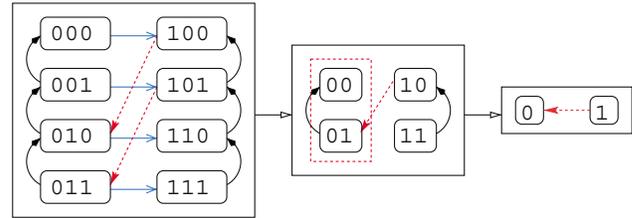


FIG. 3 (color). Matrix-generating algorithm for a three site model. Each possible occupancy of the lattice is associated with a bit pattern, and the state is enumerated with the corresponding decimal value; i.e., since 011 is the binary representation of 3, we label this state 3. Next, divide the states into groups where the first lattice site is occupied (1-states), and where the first lattice site is empty (0-states). Regardless of the number of lattice sites in the TASEP, the transitions between the two classes of states *always* occur between the first half of the 1-states and the second half of the 0-states (dashed red arrows). To determine the remaining transitions we call the algorithm recursively on both the 1-states and the 0-states, making sure to ignore the highest order bit (i.e., the leftmost lattice site). Finally, we add the transitions between each 0-state and each 1-state resulting from injection at the left edge of the lattice (blue arrows). With the connectivity of the states fully enumerated, one readily assigns the appropriate rate p_i to each of the transitions.

Now consider an interior segment [Fig. 2(b)]. The mean density in the site immediately to the left (right) of the segment is denoted σ_- (σ_+). The transition matrix contains the parameters σ_{\pm} since entry and exit into the enumerated segment is proportional to σ_- and $1 - \sigma_+$, respectively. The steady-state current can be calculated from the particle flux out of the rightmost site of the segment: $J(\sigma_-, \sigma_+, \{p_i\}) = (1 - \sigma_+) \sum_{j=\text{odd}} P_j(\sigma_-, \sigma_+, \{p_i\})$, since the odd states correspond to those with a particle at the last site in the segment. The singular eigenvector of the $2^n \times 2^n$, sparse, banded transition matrix was computed (up to $n \approx 18$) using an implicit restarted Arnoldi iteration method [18] via MATLAB. The effects of particle correlations surrounding a bottleneck are accounted for provided n is larger than the density boundary layer thickness. Since the densities are uniform far from the defects, we assume that they are also σ_- (σ_+) far to the left (right) of the segment. Thus, the mean-field currents well to the left and right of the segment are $J_- = \sigma_-(1 - \sigma_-) = J_+ = \sigma_+(1 - \sigma_+)$. Since the only physical solution is $\sigma_- = 1 - \sigma_+$, we equate $J(\sigma_- = 1 - \sigma_+, \sigma_+, \{p_i\}) = \sigma_+(1 - \sigma_+)$ and solve for σ_+ numerically, determining J . For $\alpha < 1/2$ and defects near the initiation end, we simply equate $J(\alpha, \sigma_+, \{p_i\}) = \sigma_+(1 - \sigma_+)$ and solve for σ_+ . For $\beta < 1/2$ and defects near the termination end, σ_- is determined from $\sigma_-(1 - \sigma_-) = J(\sigma_-, \beta, \{p_i\})$. We used this improved, systematic finite-segment mean-field theory (FSMFT) to compute currents of TASEPs with various placements of internal defects. In practice, segment lengths that include only 2 to 3 sites on each side of a defect were sufficient for obtaining extremely accurate

results. Efficient continuous-time Monte Carlo (MC) simulations using the Bortz-Kalos-Lebowitz algorithm [19,20] were performed on lattices of size $N \geq 1000$ to verify all numerical results.

MC simulations show that for $\alpha, \beta > 1/2$, the currents J are insensitive to the position of defects. The behavior for $\alpha, \beta < 1/2$ resembles an interior defect near the initiation or the termination end. Slow initiation and/or termination rates can be effectively described by defects ($q < 1$) near the ends of the lattice. Therefore, we restrict our analysis to large $\alpha, \beta \gg 1/2$. Hopping across a single interior defect (with rate $q < 1$) is the overall rate-limiting step. The single-defect-reduced steady-state current $J_1(q)$ found from both FSMFT and MC simulations are shown in Fig. 4(a). The $n = 4$ FSMFT yields currents within 2% of those computed from MC simulations. The least-accurate $n = 0$ FSMFT gives $J = q/(q + 1)^2$ and is equivalent to previous treatments of a single defect [17]. Larger segments n yield increasingly unwieldy algebraic expression for $J_1(q)$. The FSMFT (which is exact as if $n = N$) is asymptotically correct for $q \rightarrow (0, 1)$ and is a systematic expansion in $J = \sum_{i=1}^{\infty} a_i q^i$ about $q = 0$. Coefficients up to a_{n+1} can be shown to be given correctly by a $2n$ -segment MFT; i.e., for $n \geq 1$, $J \sim q - 3q^2/2 + \mathcal{O}(q^3)$ rather than the $J \sim q - 2q^2 + \mathcal{O}(q^3)$ predicted by the 0-segment MFT [17].

The current J will diminish upon addition of successive, contiguous defects. The steady-state current across m equivalent, contiguous interior bottlenecks can be expressed as the power series $J_m(q \rightarrow 0) \sim \frac{(m+1)}{(4m-2)}q + \mathcal{O}(q^2)$. As m increases, the current is determined by the rate-limiting m segment which resembles a uniform chain of hopping rates q with relatively fast injection from, and extraction into, the remainder of the uniform lattice. The current approaches $J = q/4$, that of a long, uniform chain with hopping rates q in the maximal current regime. Figure 4(b) plots $J_m(q)/q$ for various q as a function of m defects using a highly accurate 14-segment MFT. For

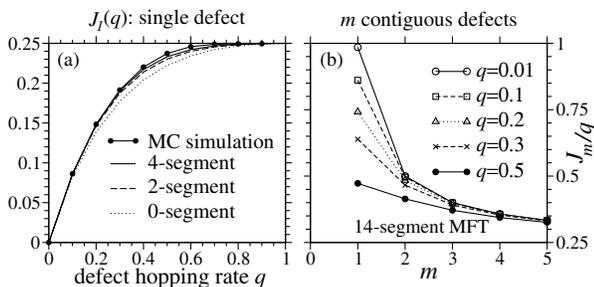


FIG. 4. (a) Comparison of steady-state currents $J_1(q)$ for a single defect derived from MC simulations with those obtained from n -segment MFT. For $n \geq 4$, the FSMFT results are within 2% of those from MC simulations. The boundary injection or extraction rates were not rate limiting ($\alpha = \beta = 10$). (b) Further reduction of steady-state current as successive, identical defects is added. The first few defects cause most of the reduction in current.

all $q < 1$, $J_m(q)/q$ approaches $1/4$ as $m \rightarrow N$. For strong bottlenecks ($q \leq 0.3$), the largest decrement in $J_m(q)$ occurs as $m = 1 \rightarrow 2$. Therefore, one may consider the effects of placing only two bottlenecks in the mRNA interior. Figure 5(a) shows the expected currents $J_2(q; k)$ across a chain containing two defects spaced k sites apart. Results for $k \leq 10$ were computed using an $n = 14$ FSMFT, while those for $k > 10$ were obtained from MC simulations. The largest reduction in the current occurs when two defects are spaced as closely as possible. The current $J_2(q; k \rightarrow \infty) \rightarrow J_1(q)$ eventually approaches that for a single defect. A finite number of multiple defects, if spaced far apart, will not significantly decrease J relative to the case of a single defect. As k increases, the density downstream of the first defect recovers to the bulk value σ_- before encountering the second defect. This behavior is clearly shown (using both FSMFT and MC simulations) in Fig. 5(b) for a pair of defects ($q = 0.15$). For two identical defects with small q , we find $J_2(q; k) \sim kq/(k + 1) + \mathcal{O}(q^2)$. Therefore, the current for two identical bottlenecks is at most a factor of 2 smaller than that for a single defect. This maximal contrast occurs when $q \rightarrow 0$ and when the two bottlenecks are adjacent to each other. Figure 5(c) plots the variation in $J_2(q; k)/J_1(q)$ as a

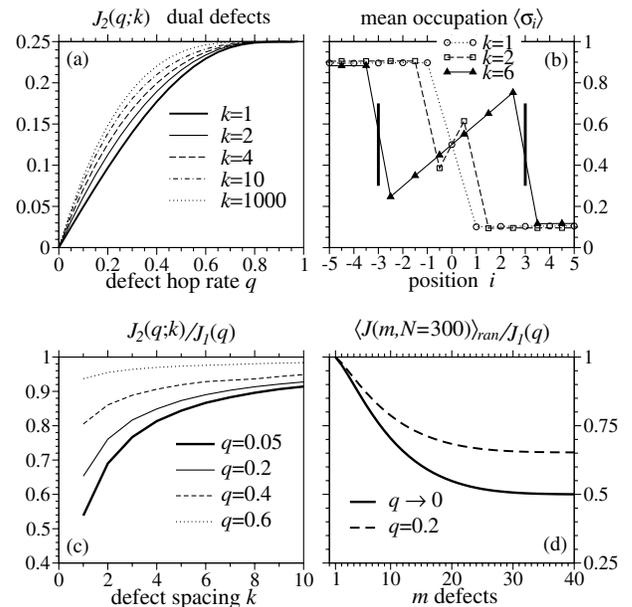


FIG. 5. (a) Steady-state currents across a chain with two identical interior defects spaced k sites apart. The current is suppressed most when the defects are closely spaced. (b) The density profiles near a pair of defects ($q = 0.15$) of various spacings k . The thick vertical bars denote the defect positions for $k = 6$. For larger k , density boundary layers heal, allowing particle buildup behind the second barrier, enhancing the current. (c) The dependence of the normalized steady-state current $J_2(q; k)/J_1(q)$ as a function of the defect separation. (d) The upper bound for the mean current of a lattice with m defects randomly distributed within the central $N = 300$ sites. A lower bound as $m \rightarrow N - 1$ is $\langle J(m \rightarrow N - 1) \rangle_{\text{ran}} \rightarrow q/4$.

function of k for various q . Note that the defect spacing dependence of J is a consequence of the nonlinear exclusion dynamics. If the lattice particles were independent (no exclusion), their complete traversal time is simply the sum of the individual waiting times and would not depend on the defect arrangement.

Typically, a cell has a fixed number of rare codons with which to regulate translation. From the two-defect current J_2 , we estimate the expected current for m randomly distributed bottlenecks. The total number of ways m defects can be placed on $N - 1$ interior sites, such that the minimum pair spacing is $\geq k$ is

$$Z_k(m, N) = \binom{N - 1 - (m - 1)(k - 1)}{m}.$$

The probability that the minimum interdefect spacing equals k is thus $Q_k(m, N) = Z_1^{-1}(Z_k - Z_{k+1})$. Since the current produced by a defect configuration with a minimum defect spacing k is $\leq J_2(q; k)$, we find the upper bound:

$$\langle J \rangle_{\text{ran}} \leq \sum_{k=1}^{\text{int}\{(N-1)/(m-1)\}} Q_k(m, N) J_2(q; k). \quad (1)$$

This upper bound will be very accurate if the defect density is low enough that one can neglect the probability that more than two defects each separated by k sites form a single cluster, particularly for small k . Although the most likely minimum defect spacing is $k = 1$, at low defect densities, the total probability of closely spaced defects remains small and the weight of $Q_k(m, N)$ at larger k dominates the statistics of $\langle J \rangle_{\text{ran}}$. The disorder-averaged current $\langle J(m, N = 300) \rangle_{\text{ran}}$ [normalized by $J_1(q)$] is shown in Fig. 5(d). For very small m , the current is approximately that of a single defect. The current is most sensitive to the number of random defects at $m \approx 10$, corresponding to $m/N \approx 0.03$, approximately the fraction of slow codons observed *in vivo*. Nonetheless, the observation of enhanced, nonrandom clustering suggests that other biological regulation pathways exist and would yield currents measurably below the upper bound (1).

We have found that not only can a single defect directly inhibit elongation across it, but also that a few bottlenecks, properly distributed, can further slow protein production by a factor of $\sim 2-4$ [21]. Although maximal current reduction is achieved by clustering defects as tightly as possible, successive addition beyond a handful of defects does little to reduce the current. Defects which are all spaced more than a handful of sites apart will not reduce the throughput more than a single defect. Our mathematical results are qualitatively consistent with the idea that a single, localized region provides the rate-limiting step for translation. Since initiation, which requires assembly of numerous ribosome parts, is typically

rate limiting, the existence of slow codons near the start codon [14] (forming the equivalent of two closely spaced defects near the start) suggests that their role is to suppress protein synthesis. Conversely, a finite number of well-separated defects at appropriate junctures provides pause points for, say, local, successive protein folding with *minimal* reduction in current.

The authors thank L. Shaw for valuable discussions, and NSERC Canada (G. L.), NSF DMS-0206733 (T. C.), and NIH K25 AI058672 (T. C.) for support.

-
- [1] B. Alberts *et al.*, *Molecular Biology of the Cell* (Garland Publishing, New York, 1994).
 - [2] M. Bulmer, *Genetics* **129**, 897 (1991).
 - [3] G. Marais and L. Duret, *J. Mol. Evol.* **52**, 275 (2001).
 - [4] J. R. Powell and E. N. Moriyama, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7784 (1997).
 - [5] C. G. Kurland, *FEBS Lett.* **285**, 165 (1991).
 - [6] *Cell-Free Translation Systems*, edited by A. S. Spirin (Springer-Verlag, New York, 2003).
 - [7] E. G. Strauss, J. H. Strauss, and A. J. Levine, in *Fundamental Virology*, edited by B. N. Fields, D. M. Knipe, and P. M. Howley (Lippincott-Raven, Philadelphia, 1996), pp. 141–159.
 - [8] R. Y. Tsien, *Annu. Rev. Biochem.* **67**, 509 (1998).
 - [9] M. Robinson *et al.*, *Nucleic Acids Res.* **12**, 6663 (1984).
 - [10] M. A. Sorensen, C. G. Kurland, and S. Pedersen, *J. Mol. Biol.* **207**, 365 (1989).
 - [11] D. A. Phoenix and E. Korotkov, *FEMS Microbiol. Lett.* **155**, 63 (1997).
 - [12] W. Konigsberg and G. N. Godson, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 687 (1983).
 - [13] L. B. Shaw, J. P. Sethna, and K. H. Lee, *Phys. Rev. E* **70**, 021901 (2004).
 - [14] G. F. T. Chen and M. Inouye, *Nucleic Acids Res.* **18**, 1465 (1990).
 - [15] G. M. Schütz and E. Domany, *J. Stat. Phys.* **72**, 277 (1993); B. Derrida, *Phys. Rep.* **301**, 65 (1998).
 - [16] C. T. MacDonald and J. H. Gibbs, *Biopolymers* **7**, 707 (1969); T. Chou, *Biophys. J.* **85**, 755 (2003); L. B. Shaw, R. K. P. Zia, and K. H. Lee, *Phys. Rev. E* **68**, 021910 (2003).
 - [17] A. B. Kolomeisky, *J. Phys. A* **31**, 1153 (1998).
 - [18] W. E. Arnoldi, *Q. Appl. Math.* **9**, 17 (1951); Y. Saad, *Numerical Solution of Large Eigenvalue Problems* (Halsted Press, New York, 1992).
 - [19] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, *J. Comput. Phys.* **17**, 10 (1975).
 - [20] G. Lakatos and T. Chou, *J. Phys. A* **36**, 2027 (2003).
 - [21] Ribosomes are structurally larger than nucleotide triplets, occluding $w \sim 10$ codons. Although our FSMFT algorithm cannot be easily adapted to larger particles, the steady-state properties of a $w > 1$ TASEP [16,20] are qualitatively the same as the standard TASEP where each particle occupies exactly one lattice site, as verified by extensive MC simulations.