Slip of grip of a molecular motor on a crowded track: Modeling shift of reading frame of ribosome on RNA template

Bhavya Mishra[1], Gunter M. Schütz[2] and Debashish Chowdhury[1]

[1] *Department of Physics, Indian Institute of Technology, Kanpur 208016, India.*
[2] *Institute of Complex Systems II, Forschungszentrum Jülich, 52425 Jülich, Germany.*

arXiv:1605.03434v2 [physics.bio-ph] 30 May 2017

PACS 87.16.-b – Subcellular structure and processes
PACS 05.40.-a – Fluctuation phenomena, random processes, noise, and Brownian motion
PACS 05.70.Ln – Nonequilibrium and irreversible thermodynamics

Abstract – We develop a stochastic model for the programmed frameshift of ribosomes synthesizing a protein while moving along a mRNA template. Normally the reading frame of a ribosome decodes successive triplets of nucleotides on the mRNA in a step-by-step manner. We focus on the programmed shift of the ribosomal reading frame, forward or backward, by only one nucleotide which results in a fusion protein; it occurs when a ribosome temporarily loses its grip to its mRNA track. Special "slippery" sequences of nucleotides and also downstream secondary structures of the mRNA strand are believed to play key roles in programmed frameshift. Here we explore the role of an hitherto neglected parameter in regulating -1 programmed frameshift. Specifically, we demonstrate that the frameshift frequency can be strongly regulated also by the density of the ribosomes, all of which are engaged in simultaneous translation of the same mRNA, at and around the slippery sequence. Monte Carlo simulations support the analytical predictions obtained from a mean-field analysis of the stochastic dynamics.
Key words: ribosome traffic, master equation, programmed frame shift.

Introduction. – A protein is a linear hetero-polymer made of a sequence of monomeric subunits, called amino acids, each of which is linked to its immediate neighbor by a peptide bond. Nature normally uses 20 different types of amino acids to make proteins in living cells. The particular sequence of the types of amino acids in a protein is determined by the sequence of nucleotides, the monomeric subunits, of the corresponding template mRNA molecule. The actual synthesis of the protein, as directed by the mRNA template, is carried out by a molecular machine [1], called ribosome [2–4] and the process is referred to as translation (of genetic message). Translation is broadly divided into three stages: Initiation, elongation and termination. Elongation of the growing protein by the ribosome takes place in a step-by-step manner, the addition of each amino acid monomer to it is accompanied by a forward stepping of the ribosome on its mRNA template by one codon, each codon being a triplet of nucleotides. Thus, a ribosome is also a molecular motor [2, 4] that exploits the mRNA template as its track and moves forward along it, by three nucleotides in each step, converting chemical energy into mechanical work.

At each position of the ribosome its "reading frame" decodes a triplet of nucleotides on the mRNA template and then slides to the next triplet as the ribosome steps forward by a codon. This reading frame is established in the initiation stage of translation and must be maintained faithfully during the course of normal elongation of the protein. However, in all kingdoms of life, on many template mRNA strands there are some special "slippery" sequences of nucleotides where a ribosome can lose its grip on its track, resulting in a shift of its reading frame either backward or forward by one or more nucleotides. These processes are referred to as ribosomal frameshift [5, 6]. The most commonly occurring, and extensively studied, cases correspond to a shift of the reading frame backward or forward by a single nucleotide on the mRNA track. These are referred to as -1 frameshift and +1 frameshift, respectively, and will be the main focus of our study in this letter.

The rate of frameshift at any arbitrary position on the mRNA track has been found to be negligibly small. In contrast to such a random frameshift, a "programmed" frameshift at a specific location on the mRNA track is known to occur with much higher *rate* and have important

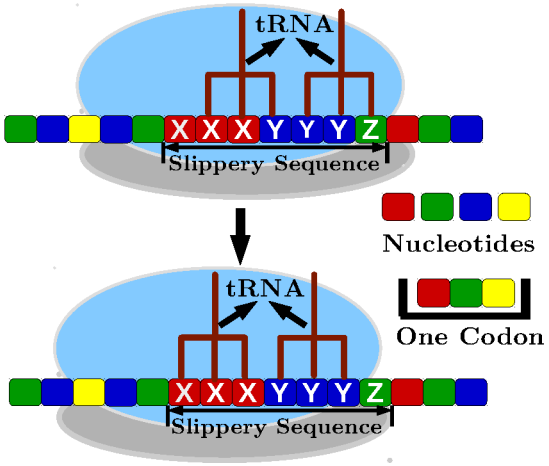

Fig. 1: (Color online) A schematic representation of -1 programmed ribosomal frameshift (PRF) from initial reading frame to -1 reading frame.

biological functions. In this letter we consider, exclusively, programmed frameshift and ignore random frameshift altogether. After suffering a programmed frameshift, the ribosome resumes its operation but decodes the template using the shifted reading frame (see Fig. (1)) thereby producing a "fusion" protein. The classic example of such a fusion product of -1 frame shift is the gag-pol fusion protein of the human immunodeficiency virus (HIV) [5, 6]. Programmed frameshift, a mode of non-conventional translation [7] happens to be one of the several modes of genetic recoding [6]. A common feature of these recoding phenomena is the context-dependent temporary alteration of the readout of mRNA. Understanding the principles followed by nature for encoding and decoding genetic message would be incomplete without understanding the causes and consequences of all such dynamic recoding.

Programmed -1 frameshift requires two key ingredients: (a) A slippery sequence (usually about seven nucleotide long) on the mRNA, and (b) a downstream secondary structure (usually a pseudoknot [8, 9]) of the mRNA [10]. Normally the pseudoknot on the mRNA template is located about 6 nucleotides downstream from the slippery sequence. In order to enter the segment of mRNA template that forms the pseudoknot, a ribosome has to unwind the secondary structure. Thus, undoubtedly, the pseudoknot acts as a roadblock against the forward movement of the ribosome that suffers a long pause on the slippery site. However, such long pauses are necessary, but may not be sufficient, for inducing -1 frameshift of the ribosomes because not all types of road blocks on mRNA can induce frameshift [11]. Moreover, not every ribosome suffers frameshift at a given slippery sequence. Even more intriguing is the fact that the same pseudoknot, that puts such an insurmountable hurdle on the path of a ribosome in the 0 reading frame, allows it to pass through more easily after the shift to the -1 reading frame.

Thus it is evident that the composition and length of the slippery sequence, its distance from the downstream pseudoknot, the energetic stability of the pseuodoknot and the kinetics of its unfolding and refolding etc. collectively encode the program that determines not only the spatial location and timing of the frame shift but also its frequency or efficiency. A combination of all these stimulators and signals are believed to alter the normal free energy landscape of the ribosome, thereby affecting the stability of various intermediate states as well as the kinetics of transitions among them that favor frameshift over the other alternative pathways [5, 6]. Several competing models have been developed to account for the mechanisms of stimulation, regulation and control of frameshift; the models differ in (a) their hypothesis as to the sub-step of the mechano-chemical cycle in which the frameshift is assumed to occur, and (b) the assumed structural, energetic and kinetic cause of the slippage [12]. Instead of committing at present to any specific structural model for frameshift, we capture the effects of the slippery sequence and the downstream pseudoknot by physically motivated generic alterations of the kinetic rates in a reduced minimal model of the elongation kinetics.

The main aim of this letter is to demonstrate how an hitherto neglected control parameter, namely the average inter-ribosome separation, or equivalently the mean number density of the ribosomes, on the mRNA track can up- or down-regulate the efficiency of ribosomal frameshift *in vivo*. In the past indications for the interplay of the inter-ribosome separation and the kinetics of unwinding of the pseudoknot have emerged from experimental studies of frameshift *in vivo* [13] as well as in experiments with synthetic mRNA secondary structures *in vitro* [14].

Because of the superficial similarities between the simultaneous movement of multiple ribosomes along a single mRNA track and vehicular traffic along a single-lane highway, the former is often referred to as ribosome traffic. Various aspects of this traffic-like phenomenon have been modelled by totally asymmetric simple exclusion process (TASEP) [15] and its biologically motivated extensions. The TASEP, an abstract model of self-driven interacting particles, is one of the simplest models of driven non-equilibrium systems in statistical physics. Since the pseudoknot acts as a bottleneck against the flow of ribosome traffic, the spatio-temporal organization of the ribosomes exhibit some of the key characteristics of TASEP with quenched defects; these include, as we show here, phase segregation of the ribosomes.

In contrast to the earlier TASEP-based models of translation [16–28] the model introduced here treats individual nucleotides, rather than triplets of nucleotides (codons), as the basic unit of the mRNA track. Moreover, unlike the earlier TASEP-based models of translation, we do not focus on the phase diagram of the model. Instead,the most important result of our investigation is the up- and down-regulation of ribosomal frameshift caused by the long

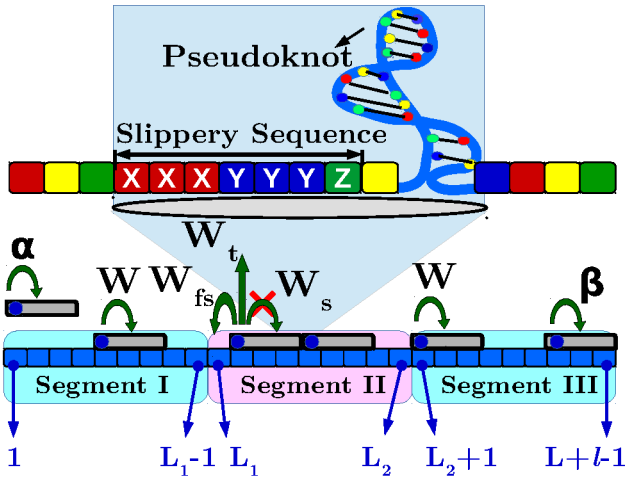

Fig. 2: (Color online) A schematic depiction of the model where a secondary structure of the mRNA (a cartoon of a pseudoknot) is shown. The mapping of the tortuous contour of the mRNA template onto a linear (one-dimensionl) chain is shown explicitly. Each site of the chain represents a single nucleotide. The first seven sites of the segment II form a "slippery" sequence while the next 2-3 nucleotides would correspond to the spacer region whereas the remaining part of the segment II remains folded in the pseudoknot. From the specific site $L_1 + 1$ programmed -1 frameshift can take place. The segments I and III are the segments of the mRNA that are located before and after, respectively, of the pseudoknot.

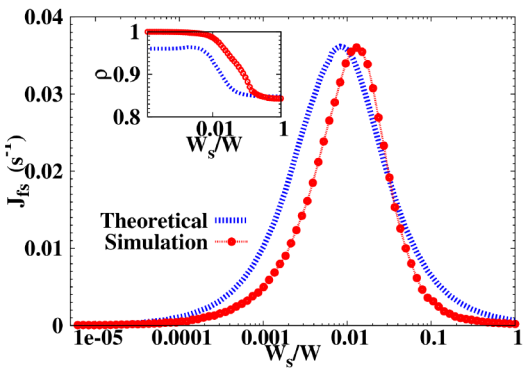

Fig. 3: (Color online) Frameshift flux J_{fs} is plotted against the parameter W_s/W. In the inset, the probability of coverage of the site $i = L_1$, which is targetted for successful -1 frameshift, is plotted also against W_s/W. The theoretical predictions derived under MFA are drawn by the dashed curve (blue). The numerical data obtained from MC simulations have been plotted with dots; the curve connecting these dots (red) serves merely as a guide to the eye. The initiation rate $\alpha = 100s^{-1}$, termination rate $\beta = 10s^{-1}$ and $W = 83$ s^{-1}. The system size is $N = L + \ell - 1 = 1000 + \ell - 1$, with the rod size $\ell = 18$, $L_1 = 399, L_2 = 450$.

queue that stretches upstream from the pseudoknot to the slippery sequence and beyond. We also suggest new applications of existing experimental techniques to test some of the new predictions made in this paper on the basis of our theoretical calculations.

Model. – In our model mRNA is treated as a linear chain (also called a lattice) of sites, each of which represents a single nucleotide. Each nucleotide is identified with an integer index j. The total length of the chain is $L + \ell - 1$, in the units of nucleotide length, although only the segment starting from $j = 1$ to $j = L$ gets translated by the ribosomes. A specific site S on this chain denotes the *second* nucleotide of the slippery sequence [6]. A ribosome is modelled as a rigid rod of length ℓ also in the units of nucleotide length, i.e., each ribosome can cover simultaneously ℓ successive nucleotides on its mRNA template. According to the convention followed consistently throughout this paper, the instantaneous position of a ribosome on the mRNA template is denoted by the integer index that labels the leftmost site *covered* by the ribosome at that instant of time. We make a clear distinction between the terms *occupied* and *covered*: A site j is *occupied* by a ribosome, i.e., its instantaneous position is j, if the ribosome is decoding the triplet of nucleotides $j, j+1, j+2$, while all the sites $j, j+1, ...j+\ell-1$ remain *covered* by it simultaneously at that instant.

In this generic model developed here for ribosomal frameshift the lattice is divided into three segments. The segment I is ranging from site 1 to site $L_1 - 1$ ($1 \leq j \leq L_1 - 1$), segment II from site L_1 to site L_2 ($L_1 \leq j \leq L_2$) and segment III from site $L_2 + 1$ to site $L + \ell - 1$ ($L_2 + 1 \leq j \leq L + \ell - 1$) (see Fig. 2). The segment II represents the stretch of the mRNA template that is folded in the form of the pseudoknot. All the numerical data presented here have been obtained with the choice $L_1 = 399$ and $L_2 = 450$. Inside segment II seven nucleotides, from site $j = L_1$ to $j = L_1 + 6$ represent a slippery sequence whose second site (i.e., $j = L_1 + 1$) is the special site from where the ribosomal frameshift is assumed to take place. The next 2-3 nucleotides would correspond to the spacer region between the slippery sequence and the pseudoknot while the remaining nucleotides of segment II form the pseudoknot. The choice of the numerical value 41 nucleotide for the length of the mRNA in the pseudoknot is only a typical one that lies between estimated lengths of the smallest and largest pseudoknots [9].

Translation initiation is captured in this model as follows: If the first ℓ sites of the lattice are empty, a ribosome can occupy the position $j = 1$ (and, thus, cover the sites $1, 2, ..., \ell$). This event occurs with a rate (i.e., probability per time unit) α. Similarly, termination of translation is described as the detachment of the ribosome from the lattice when its position is L, i.e., it *covers* the last ℓ sites of the lattice that are labelled by $L, L+1, ..., L+\ell-1$; the rate of this event is β.

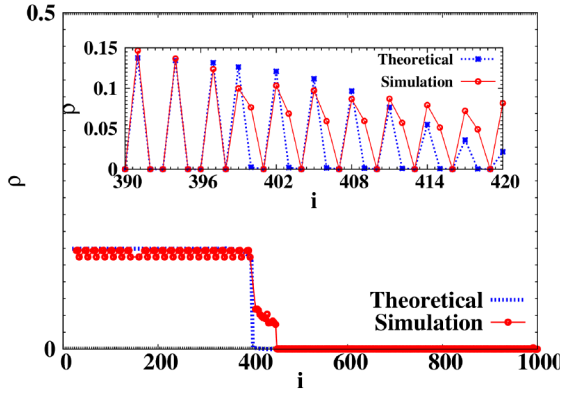

Fig. 4: (Color online) Occupational density profile in steps of 3 nucleotides. The dashed line (blue) is the theoretical prediction under MFA and the solid line (red) connecting the discrete symbols ⊙ is obtained from MC simulation with initiation rate $\alpha = 100s^{-1}$, termination rate $\beta = 10s^{-1}$, $W = 83s^{-1}$, $W_{fs} = 13.3s^{-1}$ and $W_s = 0.00074s^{-1}$. System size $N = L + \ell - 1 = 1000 + \ell - 1$ with the rod size $\ell = 18$; the special site is located at $i = 400$. The inset shows the occupational density profile of a small region from $i = 390$ to $i = 420$ in steps of 1 nucleotide; the dashed line (blue) with $*$ and the solid line (red) connecting the discrete symbols ⊙ have been obtained from MF theory and MC simulation, respectively.

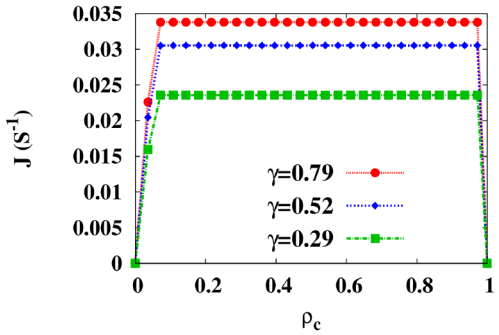

Fig. 5: (Color online) Net forward flux (J) plotted against the coverage density ρ_c for three different values of the jump rate ratio $\gamma = W_s/W$. System size is $N = 1000 + \ell - 1$, with the rod size $\ell = 18$, and $W = 83$ s^{-1}.

During the elongation stage a ribosome moves forward by three nucleotides upon successful completion of each elongation cycle. However, at the site S, which represents the *second* nucleotide of the slippery sequence, a ribosome can slip backward on its track by one single nucleotide; the rate of this -1 frameshift event, that we define below, is normally much less than unity. So far as the termination of translation by a frameshifted ribosome is concerned, we assume, for the sake of simplicity, that the ribosome detaches from the lattice when its position is $L - 1$, i.e., it covers the sites $L - 1, L, ..., L + \ell - 2$. Thus, in this model the *full length* of a protein synthesized using the non-shifted frame and that of a fusion protein would consist of the same number of amino acids.

In this scenario the overall rate of translation of each codon in the original (unshifted) frame is captured by an "effective rate" W of "hopping" of a ribosome by three steps in the forward direction, i.e., from position j to $j + 3$. In principle, the effective rate W can be expressed in terms of the actual rates of the individual transitions among the five distinct mechano-chemical states in the elongation stage (for details please see the supplementary information given in [29]).

Next we assume that the effective hopping rate W in the pseudoknot segment II gets reduced exponentially to the value W_s, i.e.,

$$W_s/W = \gamma = \exp(-b\ \Delta\tilde{G}) \tag{1}$$

where $-\Delta\tilde{G}$ is a dimensionless parameter. This choice is motivated by the plausible identification $\Delta\tilde{G} = (\Delta G)/(k_B T)$ where ΔG is the free energy barrier against forward movement of the ribosome by a single nucleotide within the pseudoknot region. Since a typical pseudoknot is not a mere hairpin, the effects of its structural complexity on the effective barrier is captured by the parameter b in the exponential. For the numerical data plotted graphically in this paper $b = 3$ was chosen. As $\Delta\tilde{G} \to 0$ the choice of the form of W_s in (1) implies $W_s \to W$; this is consistent with the fact that in this limit the difference between the segment II and the other two segments disappears. In the opposite limit $\Delta\tilde{G} \to \infty$, $W_s \to 0$ as would be expected on physical grounds that it is practically impossible to unzip an extremely stiff psudoknot. The rate of frameshift W_{fs} depends upon two factors: (a) Strength of the pseudoknot, expressed by $\Delta\tilde{G}$, and (b) the frequency W_{fs0} which is related to the breaking the bonds between tRNA and codons.

Motivated by the above physical considerations we make our next postulate: We assume that the rate of -1 frameshift in the slippery region is given by

$$W_{fs} = W_{fs0} \exp(a\ \Delta\tilde{G}) \tag{2}$$

where a is the parameter that indirectly captures the complexity of pseudoknot structure. Note that in the limit $\Delta\tilde{G} \to 0$, $W_{fs} \to W_{fs0}$ which is a non-vanishing (but, presumably small) rate of frameshift caused by the slippery sequence in the absence of a downstream psudoknot. Moreover, the form (2) is based on the assumption of a sharp increase of W_{fs} with increasing $\Delta\tilde{G}$. For the numerical data plotted graphically in this paper $a = 3$ was chosen. [8].

Similarly, the detachment (premature termination) rate

W_t of a rod from the special site is assumed to be

$$W_t = W \exp(-c\ \Delta\tilde{G}) \tag{3}$$

Rods can detach prematurely from the special site only if the nearest neighbour site from the rightmost part of the rod in the forward direction is already covered by another rod. In an alternative version of this model, defined in the supplementary material, the rods can detach prematurely if the nearest neighbor of the left most part of the rod in backward direction, i.e., the site targetted for -1 frameshift, is already covered by another rod. The choice of the form (3) is motivated by our postulate that (a) in the absence of the pseudoknot (i.e., in the limit $\Delta\tilde{G} \to 0$) both the forward hopping and premature detachment at the slippery site are equally probable, and that (b) for very stiff pseudoknots (i.e., $\Delta\tilde{G} \to \infty$) practically a ribosome stalls (no forward movement because $W_s \to 0$) and, therefore, no possibility of premature detachment [30]. Although, in principle, the two parameters a and c in (1) and (3) are not necessarily equal, we use $a = c$ just for the sake of simplicity.

The kinetics is implemented by the following rules:
(a) A new rod can attach at site 1, if and only if all initial ℓ sites are empty. The rate of attachment at site 1 is α.
(b) If there is a rod at site $i = L$ or $i = L - 1$, then it can detach from the lattice, and the rate of detachment is β.
(c) Inside segment I and III rods can jump in forward direction by +3 nucleotides only if the target site is empty and the rate of forward jump is W.
(d) Inside segment II there is one special site $i = L_1 + 1$. Except at this special site a rod can jump forward, by step size +3, if the target site is empty and the rate of forward jump in segment II is W_s.
(e) From the special site $i = L_1 + 1$ the following movements are possible:
(e1) A Rod can jump forward with rate W_s, with step size +3, if the target site is empty.
(e2) A Rod can slip back with step size -1 with rate W_{fs}, if the target site is empty.
(e3) A Rod can detach, with rate W_t, from the lattice if the site $L_1 + 1 + \ell$ (i.e., the site immediately in front of its forward edge) is occupied by another rod.

In the analytical treatment of this model $p(j|\underline{i})$ is the conditional probability of finding the site j empty, given that the site i is already occupied, where j is the target site when a rod tends to move by ± 1 nucleotides. Similarly, $q(k|\underline{i})$ is the conditional probability of finding the site k empty, given that the site i is already occupied, where k denotes the target site when the rod tends to move by 3 nucleotides. $P(i,t)$ is the occupational probability and it is defined as the probability of finding the left edge of the rod at site i at time t.

In this work we are interested in the frameshift flux (J_{fs}) from the special point which is defined as the total number of ribosomes that undergo frameshift per unit time from this site. Thus, J_{fs} is given by

$$J_{fs} = W_{fs} p(L_1|\underline{L_1 + 1}) P(L_1 + 1). \tag{4}$$

We can get the value of $P(i,t)$ at each individual site i and time t by solving master equations under mean field approximation (MFA). At site $i = 1$ one has

$$\begin{aligned} \frac{dP(i,t)}{dt} &= \alpha[1 - \sum_{a=1}^{\ell} P(a,t)] \\ &\quad - Wq(i + \ell + 2|\underline{i})P(i,t), \end{aligned} \tag{5}$$

since a new rod can attach only when all initial ℓ sites are empty. Therefore a summation up to ℓ is taken in the gain part of Eq. 5.

At the special site $i = L_1 + 1$ one gets

$$\begin{aligned} \frac{dP(i,t)}{dt} &= W_s q(i + \ell - 1|\underline{i - 3})P(i - 3,t) \\ &\quad - [W_s q(i + \ell + 2|\underline{i}) + W_{fs} p(i - 1|\underline{i})]P(i,t) \\ &\quad - W_t P(i + \ell, t)P(i,t), \end{aligned} \tag{6}$$

and at site $i = L_1$

$$\begin{aligned} \frac{dP(i,t)}{dt} &= W_s q(i + \ell - 1|\underline{i - 3})P(i - 3,t) \\ &\quad + W_{fs} p(i|\underline{i + 1})P(i + 1,t) \\ &\quad - W_s q(i + \ell + 2|\underline{i})P(i,t). \end{aligned} \tag{7}$$

At the site of termination of translation

$$\frac{dP(i,t)}{dt} = WP(i - 3,t) - P(i,t)\beta, \tag{8}$$

where $i = L$ for non-frameshifted ribosomes where $i = L-1$ in case of frameshifted ribosome. Note that when the rod is located at $i = L - 4$ or at $i = L - 3$ it does not face any exclusion at its target site $L - 1$ or L, respectively; consequently, there is no factor of q in the gain term on the right hand side of (8).

For the same reason, for all the sites $i \geq L - \ell + 1$, exclusion effect appears neither in the gain terms nor in the loss terms in the master equations

$$\frac{dP(i,t)}{dt} = WP(i - 3,t) - WP(i,t). \tag{9}$$

At all other sites in segment I and III

$$\begin{aligned} \frac{dP(i,t)}{dt} &= Wq(i + \ell - 1|\underline{i - 3})P(i - 3,t) \\ &\quad - Wq(i + \ell + 2|\underline{i})P(i,t). \end{aligned} \tag{10}$$

In segment II

$$\begin{aligned} \frac{dP(i,t)}{dt} &= W_s q(i + \ell - 1|\underline{i - 3})P(i - 3,t) \\ &\quad - W_s q(i + \ell + 2|\underline{i})P(i,t). \end{aligned} \tag{11}$$

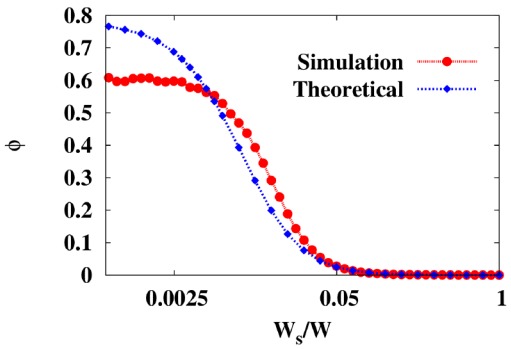

Fig. 6: (Color online) The fraction of proteins that are 'fusion' of two proteins arising from PRF are plotted as a function of the jump rate ratio $\gamma = W_s/W$. System size is $N = 1000+\ell-1$, with the rod size $\ell = 18$, and $W = 83$ s^{-1}

.

Results. – In this work we have followed two different approaches for analyzing the model to obtain quantitative results. All the effective rate constants W, W_s and W_{fs} chosen for calculation were obtained using the relation between the effective rate constant and basic rate constants of the 5-state original model whose numerical values are given in Table. 1 of the supplementary information given in [29].

In our first approach we have obtained the steady state occupational density profile by solving the Eqs. (5) - (11) iteratively, under mean field approximation (MFA), by the standard Runge-Kutta integration scheme until the steady state was attained. In our second approach we carried out Monte-Carlo simulations (MCS) of the same mechanisms as in our theoretical model. In both the approaches a very high initiation rate $\alpha \approx 100s^{-1}$ and very low termination rate $\beta \leq 10s^{-1}$ have been chosen to ensure a high density of the rods on the track.

Realistically, typical values of ℓ would be about 30 nucleotides. In order to study our model with $\ell = 30$, we would need a proportionately large values of L_1, L_2 and L. A comparison of our preliminary test results obtained for those realistic sizes with those for the shorter values reported in this manuscript showed no qualitative difference.Therefore, with the shorter values of the sizes, namely, $L_1 = 399, L_2 = 450, L = 1000$, we got lot more data for averaging which are reported in this paper.

In Fig. (3), we show the variation in frameshift flux J_{fs} with jump rate ratio $\gamma = W_s/W$. The theoretical results predicted under MFA agree well with the corresponding data from MC simulation in the regime where the coverage density of the ribosomes is sufficiently low. However, with increasing coverage density increasing deviation of the MFA from the MC data is observed. This is not unexpected because the correlations that are significant at high densities are neglected under MFA.

From Eq. (4), J_{fs} depends on two factors, (a) the parameter W_{fs} which, in turn, reflects the complexity of the pseudoknot and (b) the variable $p(L_1|\underline{L_1+1})$ which indicates the availability of an empty target site for frameshift. As the psudoknot strength $\Delta\tilde{G}$ increases the parameter W_{fs} also increases (equivalently, γ decreases), thereby causing an increase of J_{fs}. However, the increase of $\Delta\tilde{G}$ also suppresses W_s, leading to the increasingly dense queue of ribosomes just behind the special site, i.e., a concomitant decrease of $p(L_1|\underline{L_1+1})$. Therefore, with further increase of $\Delta\tilde{G}$ (i.e., decrease of γ) the flux J_{fs} attains its maximum beyond which it decreases sharply because the target site becomes inaccessible at such high density of ribosomes.

In Fig. (4), we show the occupational density profile (in steps of 3 nucleotides i.e. for $j = 1, 4, 7, 10, \ldots$) for $W_{fs} = 13.3s^{-1}$ and $W_s = 0.00074s^{-1}$. For this value of $W_s \ll W$ the entire segment II behaves as a static bottleneck which creates a high density region in segment I and a low density region in segment III. In the inset we show the occupational density profile (in steps of 1 nucleotide, i.e., for $j = 1, 2, 3, \ldots$) for a small region from $i = 390$ to $i = 420$. Since a rod can jump only in steps of 3 nucleotides, it can occupy only those sites which are of the form $3k + 1$, where $k = 1, 2, 3, \ldots$. Everywhere else occupational density will always remain zero behind the special point. After passing through the special site, a rod can occupy site $3k + 1$ (non-frameshifted rod) as well as site $3k$ (frameshifted rod) Thus, the structure observed in the occupational density profile is merely an artefact of the hopping of each rod, normally, by three sites at a time.

In Fig. (5), we show the variation in flux J with coverage density ρ_c for three different values of γ. The flat top of the flux is a well-known feature of the TASEP with a blockage [31, 32].

Another interesting quantity that characterizes PRF is the fraction ϕ of the proteins synthesized that are actually *fusion* of two proteins "conjoined at birth" because of PRF. Let n_{fs} be the total number of ribosomes that undergo frameshift from the special point and detach from the site $L - 1$ (i.e. the designated site of termination of translation by frameshifted ribosomes) after completing synthesis of a fusion protein without suffering premature detachment. Similarly, n_{nfs} is the total number of ribosomes that complete synthesis of a full length single protein without undergoing frameshift and detach from the designated termination site L. In terms of n_{fs} and n_{nfs}, the fraction ϕ is defined as

$$\phi = \frac{n_{fs}}{n_{fs} + n_{nfs}} \tag{12}$$

Thus, ϕ is also a measure of the efficiency of programmed -1 frameshift. The fraction ϕ is plotted against the ratio W_s/W in Fig.6. The deviation of the mean-field prediction from the corresponding computer simulation data increases with decreasing W_s/W. This deviation is caused by the increasing crowding of ribosomes where correlations, which

are neglected in MFA, are quite significant.

The fraction ϕ expectedly vanishes in the limit $W_s = W$. As W_s/W decreases frameshits become more likely which is reflected in the increase of ϕ. However, as the ratio W_s/W decreases further the number of frameshift events n_{fs} begins to decrease because of the increasing unavilability of an empty target site that is needed for -1 frameshift. But, concomitantly n_{nfs} also decreases because of the inability to unwind a stiffer pseudoknot. Consequently, ϕ saturates as $\gamma = W_s/W \to 0$.

Conclusion. – Programmed ribosomal frameshift is one of the most prominent modes of recoding of genetic information. In this paper we have demonstrated that the density of ribosomes at and around the slippery sequence is an important parameter that determines the frequency of programmed ribosomal -1 frameshift. In Figs. (3) and (6) we have demonstrated the effects by varying the stiffness of the pseudoknot that, in turn, controls the ribosome density.

The suppression of -1 programmed frameshift by a trailing ribosome in a dense ribosome traffic on a mRNA track is similar to the suppression of diffusive backtracking of a RNA polymerase (RNAP) motor by another trailing very closely on a DNA track [33]. Unlike the slippery sequence and the pseudoknot on the mRNA template, this hitherto neglected parameter may control the frequency of -1 frameshift dynamically because the density of the ribosomes on the mRNA can be up- or down-regulated by several different signals and pathways.

For laboratory experiments, the stiffness of a pseudoknot can be varied artificially [14] . Using such synthetic mRNA strands our theoretical prediction can be tested experimentally by a combination of ribosome profiling technique [34, 35] (for measuring the ribosome density) and FRET (for the frequency of frameshift) [36].

From the perspective of TASEP-based modeling of ribosome traffic on mRNA template, our work is a significant progress. In all the earlier models of this type each site on the lattice (chain) corresponds to a single codon, i.e., a triplet of nucleotides. In contrast, in the models developed here each site represents a single nucleotide; this modification was necessary to capture -1 frameshift where the ribosome steps backward on its mRNA track by a single nucleotide. We intend to use this prescription for modeling the mRNA track in our future model of programmed +1 frameshift. The pseudoknot segment of the mRNA track is, effectively, an extended "blockage" against the forward movement of the ribosomes. Not surprisingly, the density profile (see Fig.4) and the net flux (see Fig.5) of the ribosomes display the well known characteristics of TASEP with static blockage [31, 32].

Acknowledgments. – This work has been supported by J.C. Bose National Fellowship (DC), "Prof. S. Sampath Chair" Professorship (DC) and by UGC Junior Research Fellowship (BM).

Supplementary Information. –

5-state model of elongation cycle. A ribosome consists of two interconnected subunits called large subunit (LSU) and small subunit (SSU). A tRNA molecule can transit through the intersubunit space. The three binding sites for the tRNA, arranged sequentially from the entry to the exit along its path, are denoted by A, P and E, respectively. One end of the tRNA molecule that interacts with the SSU constitutes the anticodon that binds with the codon on the mRNA track by complementary base-pairing. The other end of the tRNA, that interacts with the LSU, is aminoacylated, i.e., charged with the corresponding amino acid monomer.

The elongation cycle can be broadly divided into a sequence of three major processes: aatRNA selection, peptide bond formation, and translocation. However, aatRNA selection itself proceeds in two steps: (i) Preliminary selection based on the differences in the codon-anticodon binding free energies for cognate and non-cognate aatRNA, (ii) kinetic discrimination between cognate and near-cognate aatRNA that is usually referred to as kinetic proofreading. Once an aatRNA passes through these two-step "molecular identification" screening, it is accommodated at the A-site and the system awaits the onset of the next major process, namely, peptide bond formation. The third major process, namely translocation, also consists of multiple sub-steps during which the two ribosomal subunits execute well orchestrated movements on the template mRNA while concomitant movement of the tRNAs take place on the triplet binding sites E,P,A on the ribosome.

An elongation cycle is captured in our model by a multi-step kinetic process involving N distinct mechano-chemical states. The number $N = 5$ and the allowed transitions among them, as shown schematically in Fig. (7), capture all the major steps in the elongation cycle that have been established so far by structural and kinetic measurements [1]. The rates of the allowed transitions indicated by the arrows on this diagram are also shown symbolically next to the corresponding arrows. In principle, frameshift can

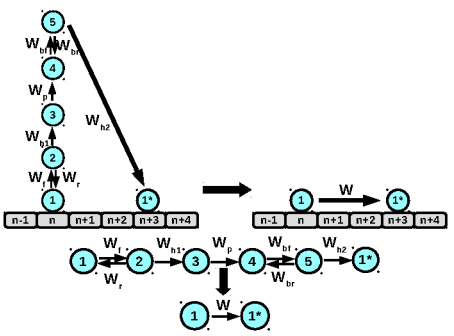

Fig. 7: Schematic diagram for conversion of a 5-state model into a single state model

take place from any of the five states of this kinetic model. Therefore, corresponding to the 5-state model of elongation cycle, one can envisage at least five different kinetic models of -1 frameshift; the only difference between these different models is the step in which frameshift is assumed to take place. The details of all these five different models are given elsewhere [37].

Although the slippery sequence and the downstream pseudoknot have been implicated in the -1 frameshift the exact step of the longation cycle where it occurs and the structural dynamics causing this frameshift remain controversial. Therefore, in this reduced minimal model the five distinct mechano-chemical states are collapsed onto a single effective state labelled by the instantaneous position of the ribosome.

Rate Exp.	Num.Value	Source
W_f	$100s^{-1}$	Pape T. et al. [39]
W_r	$25s^{-1}$	Pape T. et al. [39]
W_{h1}	$80s^{-1}$	Rodnina M.V. et al. [40]
W_p	$80s^{-1}$	Pape T. et al. [39]
W_{bf}	$25s^{-1}$	Sharma A.K. et al. [20]
W_{br}	$25s^{-1}$	Sharma A.K. et al. [20]
W_{h2}	$60s^{-1}$	Wen J.D. et al. [41]

Table 1: Parameters in the 5-state model of elongation cycle and their numerical values

Reducing the 5-state model to a 1-state model. By using effective rate constant method we have converted the five state model into an equivalent one state generic model [38] Fig. (7). The effective rate constants are,

$$W_4 = W_{bf}\frac{W_{h2}}{W_{h2}+W_{br}} \tag{13}$$

$$W_2 = W_f\frac{W_{h1}}{W_{h1}+W_r} \tag{14}$$

and

$$[W_{1\to 1^*}]^{-1} = [W]^{-1} = \frac{1}{W_2}+\frac{1}{W_{h1}}+\frac{1}{W_p} + \frac{1}{W_4}+\frac{1}{W_{h2}}. \tag{15}$$

Calculation of the conditional probabilities p and q. Consider N rods, each of size ℓ, distributed over a periodic lattice of L sites.

We first calculate the conditional probability $p(i-1 \mid \underline{i})$ that the site $i-1$ is not covered, given that a ribosome is located at the site i ((i.e., covers the sites $i, i+1, ..., i+\ell-1$). In order to do so we follow [42, 43] and map the process to the so-called zero-range process (ZRP) [44]. In this mapping one labels the N rods consecutively by integers $\alpha = 1, 2, \ldots, N$ and the size of the gap between two consecutive rods, i.e. the number of empty sites between them, becomes a particle occupation number n_α on site α of the ZRP. The total number of particles in the ZRP

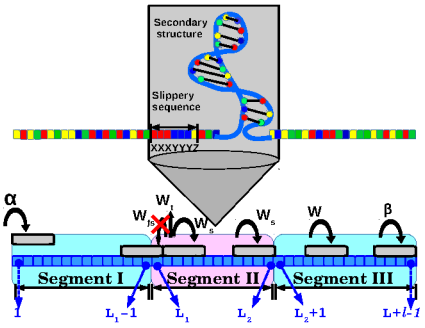

Fig. 8: Schematic diagram for Model B.

is thus $N_{ZRP} = L - N\ell$, corresponding to the number of vacant sites in the TASEP, and the number of sites in the ZRP is given by $L_{ZRP} = N$. This means that the particle number density $\rho = N/L$ of the TASEP is related to the particle number density $c = N_{ZRP}/L_{ZRP}$ of the ZRP by $c = 1/\rho - \ell$. It is convenient to work in the grand canonical ensemble where the probability of finding n ZRP-particles on (any) given site is $P_n(z) = z^n(1-z)$ and $z = c/(1+c)$ [43].

Under this mapping the conditional probability $p(i-1 \mid \underline{i})$ becomes the probability of a finding at least on particle (gap size > 0) on a site in the zero-range process, i.e., $p(i-1 \mid \underline{i}) = \sum_{n=1}^{\infty} P_n(z)$. Thus

$$p(i-1 \mid \underline{i}) = z = \frac{1-\ell\rho}{1-(\ell-1)\rho}. \tag{16}$$

Next we calculate the conditional probability $q(i+\ell+2 \mid \underline{i})$ which is the conditional probability that the site $i+\ell+2$ is not covered, given that the site i is already occupied (i.e., the sites $i, i+1, i+2, ..., i+\ell-1$ are covered). In the language of the ZRP this is the probability of finding any occupation

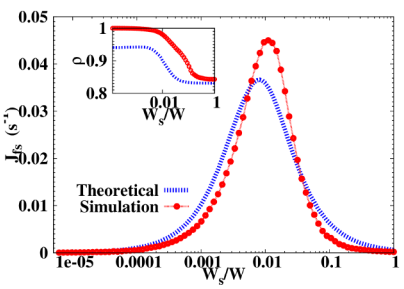

Fig. 9: Frameshift flux J_{fs} for model B. In inset, coverage probability at target site $i = L_1$ for frameshift. Dashed line is for theoretical values under MFA and dashed line with ⊙ is for MC simulation with initiation rate $\alpha = 100s^{-1}$, and termination rate $\beta = 10s^{-1}$. System size $N = L + \ell - 1 = 1000 + \ell - 1$ and rod size $\ell = 18$.

number larger than 2, i.e., $q(i+\ell+2 \mid \underline{i}) = \sum_{n=3}^{\infty} P_n(z)$. Thus

$$q(i+\ell+2 \mid \underline{i}) = z^3 = \left(\frac{1-\ell\rho}{1-(\ell-1)\rho}\right)^3. \qquad (17)$$

Models with interactions between rods that are more detailed than a hard-core repulsion can be treated with the methods of [45].

Model B. The model B differs from model A only in the rule for detachment of a ribosome. More specifically, a ribosome can detach from the special site, with rate W_t, if and only if the target site under possible -1 frameshift is already covered by another ribosome (see Fig. (8)).

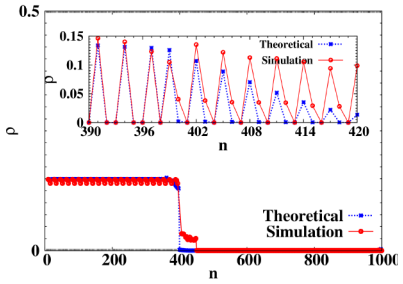

Fig. 10: Occupational density profile in steps of 3nts for model B. The dashed line with $*$ is the theoretical values under MFA and the solid line with $\odot$ is for MC simulation with initiation rate $\alpha = 100s^{-1}$, termination rate $\beta = 10s^{-1}$, $W_{fs} = 13.3s^{-1}$ and $W_s = 0.00074s^{-1}$. System size $L+\ell-1 = 1000+\ell-1$ and rod size $\ell = 18$. The inset shows the occupational density profile of a small region from $i = 390$ to $i = 420$ in steps of 1 nt. The dashed line with $*$ is for theoretical MF calculation and the solid line with $\odot$ is for MC simulation.

The master equations for model B are similar to those for Model A, except at the special site ($i = L_1 + 1$); at this special site the master equation is

$$\begin{aligned}\frac{dP(i,t)}{dt} &= W_s q(i+\ell-1|\underline{i-3})P(i-3,t) \\ &- [W_s q(i+\ell+2|\underline{i}) + W_{fs}p(i-1|\underline{i})]P(i,t) \\ &- W_t P(i-\ell,t)P(i,t).\end{aligned} \qquad (18)$$

All the conditional probabilities for model B are identical to those for Model A. For this model the frameshift flux J_{fs} is given by

$$J_{fs} = W_{fs}p(L_1|\underline{L_1+1})P(L_1+1,t). \qquad (19)$$

Results for model B. The results obtained for Model B are also similar to the corresponding results for Model A. In Fig. (9) we show the variation of frameshift flux J_{fs} with the jump rate ratio γ. In the inset we show the variation of the coverage density at the target site $i = L_1$ with the variation of the ratio γ.

In Fig. (10) we show the coverage density profile. We observe a discontinuous jump between the high- and low-density phases at the bottleneck. Both the results are very similar to the corresponding results reported for model A in the main text of this letter.

REFERENCES

[1] Frank J. (Editor), *Molecular machines in Biology: Workshop of the Cell* (Cambridge University Press, New York) 2010.
[2] Chowdhury D., *Phys. Rep.*, **529** (2013) 1.
[3] Chowdhury D., *Biophys. J.*, **104** (2013) 2331.
[4] Kolomeisky A. B., *J. Phys. Condens. Matter*, **25** (2013) 463101.
[5] Farabaugh P. J., *Programmed Alternative Reading of the Genetic Code.* (Springer) 1997.
[6] Atkins J. F. and Gesteland R. F. (Editors), *Recoding: Expansion of Decoding Rules Enriches Gene Expression.* (Springer) 2010.
[7] Firth A. E. and Brierley I., *J. Gen. Virol.*, **93** (2012) 1385.
[8] Brierley I. *et al.*, *Nat. Rev. Microbiol.*, **5** (2007) 598.
[9] Giedroc D. P. and Cornish P. V., *Virus Res*, **139** (2009) 193.
[10] Kim H. K. *et al.*, *Proc. Natl. Acad. Sci. USA*, **111** (2014) 5538.
[11] Brierley I. *et al.*, *Biochem. Soc. Trans.*, **36** (2008) 684.
[12] Tinoco I. jr *et al.*, *Biopolymers*, **99** (2013) 1147.
[13] Lopinski J. D.*et al.*, *Mol. Cell Biol.*, **20** (2000) 1095.
[14] Tholstrup J. *et al.*, *Nucleic Acids Res.*, **40** (2012) 303.
[15] Schütz G. M., *Phase Transitions and Critical Phenomena*, edited by Domb C. and Lebowitz J. L., Vol. **19** (Academic Press) 2001.
[16] Chowdhury D. *et al.*, *Phys. Rep.*, **329** (2000) 199.
[17] Schadschneider A. *et al.*, *Stochastic Transport in Complex Systems: From molecules to vehicles* (Elsevier) 2011.
[18] Chowdhury D. *et al.*, *Phys. of Life Rev.*, **2** (2005) 318.
[19] Sharma A. K. and Chowdhury D., *Phys. Biol.*, **8** (2011) 026005.
[20] Sharma A. K. and Chowdhury D., *J. Theor. Biol.*, **289** (2011) 36.
[21] Chou T. *et al.*, *Rep. Prog. Phys.*, **74** (2011) 116601.
[22] Zia R.K.P. *et al.*, *J. Stat. Phys.*, **144** (2011) 405.
[23] von der Haar T., *Comput. Struct. Biotechnol. J.*, **1** (2012) e201204002.
[24] Appert-Rolland C. *et al.*, *Phys. Rep*, **593** (2015) 1.
[25] Chou T. and Lakatos G., *Phys. Rev. Lett.*, **93** (2004) 198101.
[26] Lakatos G. and Chou T., *J. Phys. A*, **36** (2003) 2027.
[27] Mitarai N. and Pedersen S., *Phys. Biol.*, **10** (2013) 056011.
[28] Turci F. *et al.*, *Phys. Rev. E*, **87** (2013) 012705.
[29] Mishra B., Schütz, G.M. and Chowdhury D., arXiv:**1605.03434** (2016).
[30] Chen J. *et al.*, *Nature*, **512** (2014) 328.
[31] Janowsky S.A. and Lebowitz J.L., *Phys. Rev. A*, **45** (1992) 618.
[32] Schütz G. M., *J. Stat. Phys.*, **71** (1993) 471.
[33] Sahoo M. and Klumpp S., *EPL*, **96** (2011) 60004.
[34] Ingolia N. T., *Nat. Rev. Genet.*, **15** (2014) 205.
[35] Ingolia N. T., *Cell*, **165** (2016) 22.

[36] Tinoco I. jr and Gonzalez R. L. jr, *Genes Dev.*, **25** (2011) 1205.
[37] Mishra B., *Ph.D. Thesis, IIT, Kanpur*, **2016** (in preparation) .
[38] Cleland W. W., *Biochemistry*, **14** (1975) 3220.
[39] Pape T. *et al.*, *EMBO J*, **17** (1998) 7490.
[40] Rodnina M. V. *et al.*, *EMBO J*, **14** (1995) 2613.
[41] Wen J. D. *et al.*, *Nature*, **452** (2008) 598.
[42] Shaw L.B. *et al.*, *Phys. Rev. E*, **68** (2003) 021910.
[43] Schönherr G. and Schütz, G.M., *J. Phys. A: Math. Gen.*, **37** (2004) 8215.
[44] Evans M.R. and Hanney T., *J. Phys. A:Math. Gen.*, **38** (2005) R195.
[45] Buschle J., Maass P. and Dieterich W., *J. Stat. Phys.*, **99** (2000) 273.