

# The effect of mRNA diffusion on stochastic bursts in gene transcription

V. P. Zhdanov<sup>1)</sup>

*Department of Applied Physics, Chalmers University of Technology, S-41296 Göteborg, Sweden*

*Borshkov Institute of Catalysis RAS, 630090 Novosibirsk, Russia*

Submitted 13 February 2007

The positive feedback between messenger ribonucleic acid (mRNA) and regulatory-protein production may result in bistability and stochastic “bursts” in gene transcription. The temporal kinetics of such bursts have already been simulated in detail. We (i) show that the applicability of the temporal models may be limited due to relatively slow mRNA and/or protein diffusion and (ii) present the first 3D Monte-Carlo simulations indicating that with explicit mRNA diffusion the bursts may be much more irregular, the periods between bursts may be much shorter, and, depending on the circumstances, the burst window may be reduced or extended.

PACS: 05.40.–a, 87.16.–b

Life on our planet is intimately connected with genes (segments of the DNA sequence) determining the identity of species. For this reason, the understanding of the performance of genes is of high interest from the points of view of various branches of natural sciences and also from the viewpoint of philosophy. In particular, the answers to such basic questions as “how did we get here?” and “where do we go from here?” are inherently related to our knowledge of the gene evolution and functions.

At present, the key principles of the gene performance are well established [1]. The expression of the information encoded in genes is known to occur via a templated polymerization called transcription, in which the genes are used as templates to guide the synthesis of shorter molecules of ribonucleic acid (RNA). Later on, many of these molecules (or, more specifically, mRNA) serve to direct the synthesis of proteins on ribosomes. The whole process of gene expression can be regulated at all the steps. Specifically, the gene transcription, performed by RNA polymerase (RNAP) during its association with DNA, is often controlled by master regulatory proteins. Such proteins associate with DNA and either facilitate or suppress the RNA synthesis.

The positive and negative feedbacks between mRNA and protein production related to the same gene or different genes may result in complex kinetic behaviour including bistability and oscillations [2, 3]. Mathematically, these phenomena can be described by using conventional mean-field (MF) kinetic equations. In cells, most genes exist at single or low copy numbers, the number of mRNA and regulatory proteins is often low, and

accordingly the gene-transcription kinetics may exhibit stochastic features, e.g., transcriptional bursts related to bistability. In experiments, the mRNA and protein expression has traditionally been characterized by using average data for cell populations and accordingly the stochastic features have often been smeared. At present, the stochastic bursts can however be directly observed (see e.g. experiments with mRNA [4] and proteins [5]). The corresponding simulations are focused on *temporal* kinetics (see reviews [2, 3], recent articles [6–8], and references therein). In this Letter, we present the first 3D *spatio-temporal* Monte-Carlo (MC) simulations of transcriptional bursts.

To motivate our study in more detail, it is appropriate to recall that according to hydrodynamics the coefficient of diffusion of large biological molecules is given by  $D = k_B T / 6\pi\eta r$ , where  $r$  is the molecular radius, and  $\eta$  is the viscosity. For mRNA or protein diffusion in water, this equation typically yields  $D \simeq 2 \cdot 10^{-7} \text{ cm}^2/\text{s}$ . Inside cells, the diffusion coefficient is however usually lower by about one order of magnitude due to macromolecular crowding [9], i.e.,  $D \simeq 2 \cdot 10^{-8} \text{ cm}^2/\text{s}$  or even lower. The conventional expression for the time scale, characterizing diffusion, is  $\tau = \mathcal{L}^2/D$ , where  $\mathcal{L}$  is the appropriate distance. For the typical cell size,  $\mathcal{L} \simeq 5 \cdot 10^{-3} \text{ cm}$ , we obtain  $\tau \simeq 10^3 \text{ s}$  or  $\simeq 20 \text{ min}$ . This time is in the range of the time scales of elementary reaction steps related to the transcription. Thus, the diffusion of regulatory proteins or especially of mRNA may really influence the transcription kinetics.

In our analysis, we assume that the gene exists in a single copy, the association of RNAP with DNA does not limit gene transcription, and the mRNA ( $R$ ) pro-

<sup>1)</sup>e-mail: zhdanov@fy.chalmers.se

duction is activated if  $n$  regulatory sites are occupied by regulatory proteins ( $P$ ). In addition, we assume that the growth of the cell is negligibly slow and the gene expression occurs under steady-state conditions (the rapid growth may significantly influence the kinetics of gene expression [7]). In this case, the conventional MF kinetic equations for the  $R$  and  $P$  numbers are as follows [2, 7]

$$\frac{dN_R}{dt} = k_0 + k_1 \left( \frac{N_P}{K_P + N_P} \right)^n - k_R N_R, \quad (1)$$

$$\frac{dN_P}{dt} = k_s N_R - k_P N_P. \quad (2)$$

The first two terms in the right-hand part of Eq. (1) describe the  $R$ -production rate [ $k_0$  and  $k_1$  are the rate constants of the basal and  $P$ -regulated gene transcription,  $(N_P/(K_P + N_P))^n$  is the probability that all the regulatory sites are occupied by  $P$  (due to this factor the feedback between the  $R$  and  $P$  production is positive), and  $K_P$  is the  $P$  association-dissociation constant]. The third term in the right-hand part of Eq. (1) takes the  $R$  degradation into account ( $k_R$  is the degradation rate constant). The two terms in the right-hand part of Eq. (2) describe the  $P$  synthesis and degradation, respectively ( $k_s$  and  $k_R$  are the corresponding rate constants).

Under steady-state conditions, the analysis of Eqs. (1) and (2) is trivial. In particular, Eq. (2) yields  $N_P = (k_s/k_P)N_R$ . Substituting this expression into Eq. (1), one obtains a non-linear equation. With appropriate kinetic parameters, it predicts bistability provided that  $n \geq 2$ , and accordingly, with increasing and subsequent decreasing a governing parameter, one can observe a hysteresis loop with stepwise jumps from one stable solution to another. Mathematically, these jumps represent a saddle-node bifurcation.

The fluctuations of  $N_R$  and  $N_P$  are, to some extent, equivalent to temporal variation of governing parameters. If  $N_R$  and/or  $N_P$  are relatively small, the fluctuations may result in irregular switches between the low- and high-reactive states or, in other words, in stochastic transcriptional bursts. The corresponding temporal kinetics can be obtained by MC simulations of the events described above (see e.g. Refs. [7, 8]).

To study stochastic spatio-temporal kinetics, we should introduce spatial details into the model. In general, such details are numerous and accordingly, with increasing scrutiny, the model may rapidly become far from generic. In our simulations, we ignore many inferior details in order to keep the model relatively simple and to not obscure the main message. Our main assumptions are as follows. (i) We focus on  $R$  (mRNA)

diffusion. (ii) The  $P$  diffusion is assumed to be fast and accordingly  $P$  are considered to be distributed in the cell at random. Thus, the  $P$  diffusion is not treated explicitly. (iii) The cell is considered to consist of a nucleus, ribosomes, surrounding a nucleus, and the other cytoplasm (see Fig.2 below). (iv) The fine structure of ribosomes is neglected. This means that  $R$  may react (to produce  $P$ ) at any point of the region representing ribosomes. (v) In our simulations below,  $N_P$  is typically not too low, so that the contribution of  $P$  attachment and detachment to the  $P$  balance is minor, and accordingly we neglect this contribution. In addition, the  $P$  attachment and detachment are assumed to be fast and accordingly to be close to equilibrium, so that the effect of  $P$  on the  $R$  production rate can be described by using the conventional steady-state approximation, i.e., by employing the MF approximation like in Eq. (1) (this reasonable approximation used earlier to describe temporal stochastic transcriptional kinetics [2, 7] makes simulations fairly transparent).

To describe  $R$  diffusion, we use the lattice approximation. Specifically, the cell is represented by a sphere described into a  $L \times L \times L$  cubic lattice with  $L = 100$ . The sites with  $\mathcal{R} < 20$ ,  $20 \leq \mathcal{R} < 30$ , and  $30 \leq \mathcal{R} < 50$  ( $\mathcal{R}$  is the dimensionless radial coordinate) represent, respectively, the nucleus, ribosomes, and other cytoplasm (the gene is considered to be located in the center of the sphere). The real size,  $a$ , of regions corresponding to such sites is much larger than the  $R$  or  $P$  size. For this reason, we do not impose any restrictions on the occupation of sites, i.e., each site may contain an arbitrary number of  $R$ . The  $R$  diffusion is performed via jumps to nearest-neighbour sites.

With the specification above, we have five parallel processes, including  $R$  and  $P$  production and degradation, and  $R$  diffusion, running with the rates (cf. Eqs. (1) and (2))

$$W_1 = k_0 + k_1 [N_P/(K_P + N_P)]^n, \quad W_2 = k_R N_R, \\ W_3 = k_s N_R^{(\text{rib})}, \quad W_4 = k_P N_P, \quad W_5 = k_{\text{dif}} N_R.$$

In these expressions,  $N_R^{(\text{rib})}$  is the number of  $R$  located on the sites representing ribosomes.  $k_s$  is the  $P$  synthesis rate constant (note that this constant is related to  $N_R^{(\text{rib})}$  and accordingly not fully identical to that in Eq. (2)).  $k_{\text{dif}}$  is the  $R$  diffusion rate constant (the  $R$  diffusion coefficient is expressed via this rate constant as  $D_R = k_{\text{dif}} a^2 / 6$ ). The other parameters are the same as in Eqs. (1) and (2). The total rate of the processes is  $W_t = \sum_i W_i$ , where  $1 \leq i \leq 5$ .

For given numbers  $N_P$  and  $N_R$ , following the standard prescriptions for MC simulations [10], we choose

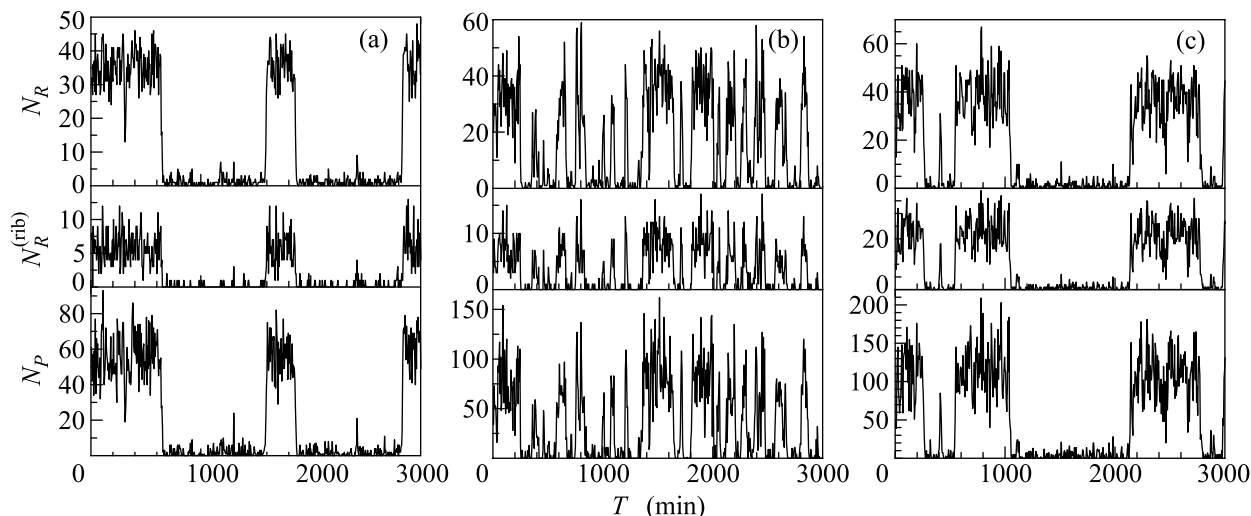


Fig.1. Numbers of mRNA in the cell and in the region representing ribosomes,  $N_R$  and  $N_R^{(\text{rib})}$ , and protein number,  $N_P$ , as a function of time: (a) for rapid  $R$  diffusion without preference of location in any region (the  $R$ -diffusion and degradation rate constants are  $k_{\text{dif}} = 10^5 \text{ min}^{-1}$  and  $k_R = 0.52 \text{ min}^{-1}$ ); (b) for relatively slow  $R$  diffusion without preference of location in any region ( $k_{\text{dif}} = 10^3 \text{ min}^{-1}$  and  $k_R = 0.72 \text{ min}^{-1}$ ); (c) for relatively slow  $R$  diffusion with preference of location in the region representing ribosomes ( $k_{\text{dif}} = 10^3 \text{ min}^{-1}$  and  $k_R = 0.65 \text{ min}^{-1}$ ). The interval between the data points is 5 min

and execute one of the possible processes [i.e., increase or decrease  $N_P$  by one, add  $R$  on the central site (this step mimics  $R$  production on the gene), remove one of  $R$ , or execute a diffusion jump of one of  $R$  to a randomly chosen nearest-neighbour sites] with probabilities  $W_i/W_t$ . If the jump should take place between the nucleus and ribosome, the trial is executed with the probability  $p_m$  ( $p_m \leq 1$ ) in order to take into account that the  $R$  diffusion between these regions is partly suppressed by the membrane separating the regions. After each MC trial, time is incremented by  $|\ln(\rho)|/W_t$ , where  $\rho$  ( $0 < \rho \leq 1$ ) is a random number.

To show the model predictions, we use the following biologically reasonable set of parameters:  $n = 3$ ,  $K_P = 20$ ,  $k_1 = 50 \text{ min}^{-1}$ ,  $k_0 = 0.5 \text{ min}^{-1}$ ,  $k_s = 100 \text{ min}^{-1}$ , and  $k_P = 10 \text{ min}^{-1}$  (in reality, the values of these rate constants may vary in a wide range). The duration of the MC runs is chosen to be 3000 min (this value is comparable e.g. with the duration of the cycle of stem cells). To illustrate general trends, the  $R$ -diffusion and degradation rate constants,  $k_{\text{dif}}$  and  $k_R$ , are employed as governing parameters.

First, we have executed simulations in the case when the  $R$  penetration via the membrane separating the nucleus and ribosome is relatively fast so that one can set  $p_m = 1$ .  $k_{\text{dif}}$  was varied in a wide range from  $10^5 \text{ min}^{-1}$  down to  $10^3 \text{ min}^{-1}$ . For  $k_{\text{dif}} = 10^5 \text{ min}^{-1}$ , the  $R$  diffusion is much faster than the reaction steps,  $R$  are distributed at random in the cell, and the well-

developed transcriptional bursts are observed at  $k_R \simeq 0.52 \text{ min}^{-1}$  (Fig.1a). As expected, these bursts are similar to those obtained earlier in the temporal MC simulations with nearly the same kinetic parameters (cf. Fig.1a and Fig.2 in Ref. [7](a)). Despite this similarity, the spatio-temporal model is found to predict a qualitatively new feature compared to the temporal simulations. Specifically, the spatio-temporal model indicates that the stochastic-burst window in the space of kinetic parameters is appreciably narrower compared to that observed in the temporal simulations. In the former case, for example, the  $R$  degradation rate constant may be changed in the range  $\pm 4\%$ . In the latter case, the corresponding range is  $\pm 10\%$ . We believe that this difference is related to the details of  $P$  production. In the temporal simulations, each  $P$  may react to produce  $P$ . In the spatio-temporal model, in contrast,  $P$  is produced only provided that  $R$  is located in the region representing ribosomes.

The typical transcriptional bursts obtained for a more realistic value of the rate constant of  $R$  diffusion,  $k_{\text{dif}} = 10^3 \text{ min}^{-1}$ , are exhibited in Fig.1b. In this case, the bursts are much more irregular and the periods between bursts are much shorter compared to the predictions of the temporal simulations or of the spatio-temporal model with rapid  $R$  diffusion [cf. panels (a) and (b) in Fig.1]. Another new feature compared to the situation with rapid  $R$  diffusion is that the burst window is now somewhat wider than that obtained in the

temporal simulations. In particular, the  $R$  degradation rate constant may now be changed in the range  $\pm 15\%$ . Both these features, can be explained taking into account that, due to  $R$ -diffusion limitations resulting in the formation of the  $R$ -concentration gradients, the  $R$  concentration is higher in the center of the cell (Fig.2a), i.e., the situation in different regions is different, and ac-

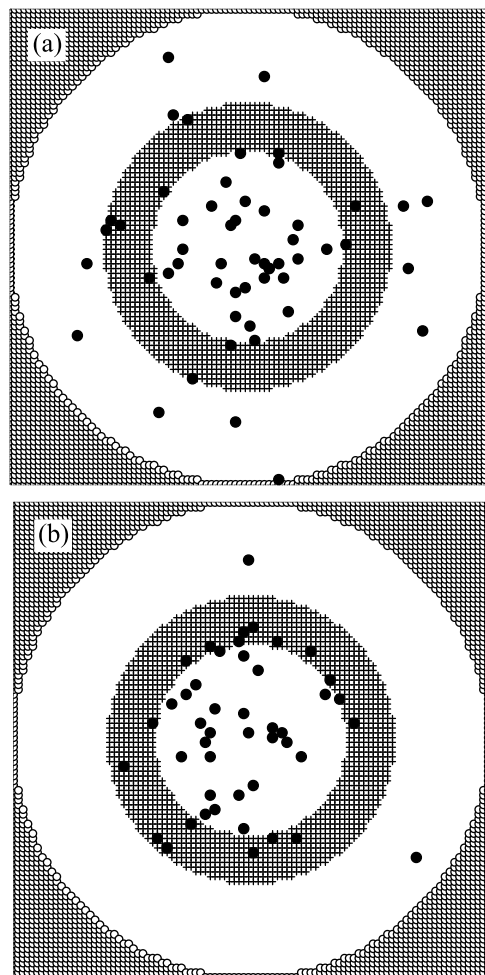


Fig.2. Cross-sections of the spherical cell. The central white area represents the nucleus. The plus signs show the region representing ribosomes. The peripheral white area represent the other cytoplasm. Filled circles indicate projections of mRNA on the cross-section plane. Panel (a) exhibits the snapshot at  $t = 2830$  min for the MC run shown in Fig.1b. Panel (b) corresponds to  $t = 2400$  min for the MC run shown in Fig.1c

ordingly the conditions for the formation of bursts can be met more easily. Thus, the burst window becomes wider. On the other hand, the conditions are met locally and accordingly the bursts become more irregular.

For  $p_m < 1$  (e.g., for  $p_m = 0.1$ ), the kinetics were found (not shown) to be qualitatively similar to those described above).

The simulations presented above were performed assuming that there is no preference of location of  $R$  and  $P$  in any region and that the rate of  $R$  diffusion is the same in all the regions. In reality, the physical properties of different regions are different and it may influence the  $R$  and/or  $P$  distribution and diffusion rates. In MC simulations, both these factors can easily be taken into account. For example, let us consider that  $R$  prefers to stay in the region representing ribosomes. This feature can be described by introducing a reducing factor,  $p_{red} < 1$ , for the  $R$  jumps from the site representing ribosomes to the other sites. Following this line, we have performed simulations with  $p_{red} = 0.1$ . For rapid  $R$  diffusion with  $k_{dif} = 10^5 \text{ min}^{-1}$ , the results (not shown) were found to be close to those predicted by the temporal model. In particular, inside the burst window, the  $R$  degradation rate constant is changed in the range  $\pm 10\%$ . For relatively slow  $R$  diffusion (with  $k_{dif} = 10^3 \text{ min}^{-1}$ ), the bursts remain well developed (see e.g. Figs.1c and 2b showing the kinetics and typical lattice pattern for  $K_P = 30$ ,  $k_s = 50 \text{ min}^{-1}$ , and  $k_R \simeq 0.65 \text{ min}^{-1}$ , the other parameters are as those in the examples above) and the burst window is about the same ( $\pm 10\%$ ). These features can be explained taking into account that now an applicable part of  $P$  is located in the region, representing ribosomes, and accordingly may produce  $P$ . Thus, the situation is close to that described by the temporal model.

The model modification introduced in the paragraph above is just an example of a few steps towards increasing scrutiny of the simulations. Many other modifications can be introduced into the model as well. For example, protein diffusion can be described explicitly. The mRNA and protein degradation can be simulated in more detail. The cell shape can be changed, etc. In our present simulations, as already noted in the introduction, we have ignored many such details in order to keep the model generic.

In summary, our Monte-Carlo simulations indicate that the mRNA diffusion may appreciably influence stochastic transcriptional bursts. This effect is found to depend on the details of the model. Compared to the predictions of the temporal model, the bursts may be much more irregular and the periods between bursts may be much shorter. Under appropriate circumstances, the burst window may be appreciably narrower. Under other circumstances, despite the gradients in the mRNA distribution, the burst window may be the same or even wider. All these findings extend the conceptual basis for

the understanding of complex kinetics of gene expression.

This work is a part of the FP6-project STREP NANOCUES funded by the European Commission.

1. B. Alberts, A. Johnson, J. Lewis et al., *Molecular Biology of the Cell*, Garland, New York, 2002.
2. L. Mariani, M. Löhning, A. Radbruch, and T. Höfer, *Progr. Biophys. Mol. Biol.* **86**, 45 (2004).
3. M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, *Nature Rev. Genet.* **6**, 451 (2005).
4. I. Golding and E. C. Cox, *Proc. Nat. Acad. Sci. USA* **101**, 11310 (2004).
5. L. Cai, N. Friedman, and X. S. Xie, *Nature* **440**, 358 (2006).
6. T. S. Zhou, L. N. Chen, and K. Aihara, *Phys. Rev. Lett.* **95**, 178103 (2005); A. Lipshtat, A. Loinger, N. Q. Balaban, and O. Biham, *Phys. Rev. Lett.* **96**, 188101 (2006); V. Lemaire, C. F. Lee, J. Lei et al., *Phys. Rev. Lett.* **96**, 198102 (2006).
7. V. P. Zhdanov, *Chem. Phys. Lett.* **424**, 394 (2006); *JETP Lett.* **84**, 632 (2006).
8. V. P. Zhdanov, *JETP* **104**, 162 (2007).
9. R. J. Ellis, *Trends Biochem. Sci.* **26**, 597 (2001).
10. D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press, Cambridge, 2000.