Distance matters: the impact of gene proximity in bacterial gene regulation

Otto Pulkkinen and Ralf Metzler

1Department of Physics, Tampere University of Technology, FI-33101 Tampere, Finland
2Institute for Physics & Astronomy, University of Potsdam, D-14476 Potsdam-Golm, Germany

Following recent discoveries of colocalization of downstream-regulating genes in living cells, the impact of the spatial distance between such genes on the kinetics of gene product formation is increasingly recognized. We here show from analytical and numerical analysis that the distance between a transcription factor (TF) gene and its target gene drastically affects the speed and reliability of transcriptional regulation in bacterial cells. For an explicit model system we develop a general theory for the interactions between a TF and a transcription unit. The observed variations in regulation efficiency are linked to the magnitude of the variation of the TF concentration peaks as a function of the binding site distance from the signal source. Our results support the role of rapid binding site search for gene colocalization and emphasize the role of local concentration differences.

PACS numbers: 87.16.-b, 87.10.-e, 05.40.-a

Suppose you live in a small town and start spreading a rumor. The time after which the rumor reaches a specific person depends on your mutual distance, either the physical distance due to word-of-mouth in the pre-telecommunications era or the topological distance in modern social networks [1]. This distance dependence is immediately intuitive for random propagation in large systems. Conversely, diffusion of signaling molecules on the microscopic scales of biological cells was observed to be fast [2], so one might assume that spatial aspects can be neglected. Yet recent studies strongly suggest that even in relatively small bacterial cells distances matter with respect to both speed and reliability of genetic regulation by DNA-binding proteins, so-called transcription factors (TFs) [3, 4]. Thus, the distance-dependence of the search time of a given TF for its target binding site on a downstream gene was proposed to affect the ordering of genes on the DNA, in particular, promote gene colocalization, i.e., the tendency of genes interacting via TFs to be close together along the chromosome [5].

Transcriptional regulation, the change in gene transcription rate caused by binding of regulatory proteins such as TFs, is the most prominent form of gene regulation in bacteria [6]. Since TFs are proteins themselves, their production consists of the inherently stochastic processes of transcription (conversion of the TF gene's code to RNA) and translation (conversion of the RNA code to proteins). Although a certain averaging of noise occurs due to long protein lifetimes, the noise in the TF production propagates to downstream genes regulated by this TF [7]. The contributions of individual stochastic steps to the total noise in protein production (magnifying glass 1 in Fig. 1) were characterized [8], and accurate theoretical models for TF-regulated expression exist in the case of known TF density at the regulatory site [9, 10].

Recently, considerably effort has been invested on explaining the efficiency of transcriptional regulation, especially the remarkable measured speed at which TFs find their binding sites [11, 12, 13]. This speed is due to facilitated diffusion [14, 15], in which free TF diffusion in three dimensions is interspersed by periods of one-dimensional sliding along the DNA (Fig. 1, magnifying glass 2). Facilitated diffusion of lac repressor molecules has indeed been observed in living E. coli cells [13]. In this context, the colocalization hypothesis certainly makes sense: a shorter search time effects more efficient regulation [4].

Concurrently, the importance of increased local protein concentration due to binding to DNA, occurrence of multiple binding sites, formation of protein complexes, and cellular compartmentalization for prokaryotic and eukaryotic gene regulation has been emphasized [20].

Here we show that high local TF concentrations due to

Figure 1: Three stochastic phases in transcriptional regulation: 1. Transcription factor (TF) production. 2. TFs perform facilitated diffusion in the nucleoid (inside the dashed line) containing the DNA. Diffusion is purely 3D in the cytosol outside the nucleoid. 3. TFs find the operator of the transcription unit (TU) gene by sliding along the DNA. The irregularly shaped blue objects depict other molecules which affect the facilitated diffusion and binding affinity of the TF.
gene proximity alone is sufficient for efficient gene regulation. Specifically, we extend the viewpoint of TF search time optimization due to colocalization to effects on the entire cascade of TF and TU gene expression, including the noise in TF production, facilitated diffusion of TF, and TF binding at TU by first binding non-specifically to the DNA and then sliding to its specific binding site (magnifying glass 3 in Fig. 1). To our knowledge, this is the first complete, quantitative approach including all relevant subprocesses in TF-mediated gene interaction.

The time-dependent intracellular concentration of a protein may be modeled by stationary shot-noise \[ \rho(t) = V_C^{-1} \int_{-\infty}^{t} e^{-\gamma(t-s)} dN_B(s) \], where \( N_B \) is a compound Poisson process of protein production with combined transcription and translation rate \( \alpha \), exponentially distributed translational burst sizes \( B_i \), \( i = 0, \pm 1, \pm 2, \ldots \), of mean \( b \) \[ \[12\] \], and a combined degradation and dilution rate \( \gamma \). \( V_C \) is the (average) cell volume. The intermediate translation step is excluded because of short mRNA lifetimes. Under the typical fast mixing assumption of molecules in the cell, the number of proteins \( M(\Omega, t) \) in a subdomain \( \Omega \) of relative volume \( v_\Omega = V_\Omega/V_C \), at given time \( t \), is therefore a Poisson random variable of intensity \( \int_\Omega \rho(t)d^3r \), with Laplace transform

\[
\langle e^{-\lambda M} \rangle = \exp \left[ -a \int_{-\infty}^{t} \frac{b(1-e^{-\lambda})v_\Omega e^{-\gamma(t-s)}ds}{1+b(1-e^{-\lambda})v_\Omega e^{-\gamma(t-s)}} \right] = [1+bv_\Omega(1-e^{-\lambda})]^{-a/\gamma}.
\]

This is but the negative binomial distribution with parameters \( a/\gamma \) and \( bv_\Omega/(1+bv_\Omega) \). In particular, the mean and the variance of the number \( M(\Omega, t) \) of proteins are

\[
\langle M \rangle = abv_\Omega/\gamma, \quad \langle M^2 \rangle - \langle M \rangle^2 = abv_\Omega(1+bv_\Omega)/\gamma.
\]

Bursty protein production (large \( b \)) clearly effects a greater variance than a simple Poissonian production of individual molecules. We note that the negative binomial distribution \[ \[11\] \] has been previously found for the number of proteins in a two-stage model of stationary expression in the fast translation limit \[ \[24\] \].

To study the expression of a gene controlled by a constitutive TF, we expand the mathematical model in two ways: (i) we introduce a position dependent kernel \( \phi(r, t) \) in the shot noise \( \rho(t) \), to include time delays in transcription, translation, protein folding, and, notably, facilitated diffusion of TFs to their target site. The coordinate \( r \) is the point of observation, namely, a point in the neighborhood of the target site (blue operator near gene \( b \) in Fig. 1). (ii) We allow a time-dependent transcription rate \( \alpha(t) \), such that the mean number of protein production events in a time interval \([t_0, t_1]\) equals \( \int_{t_0}^{t_1} \alpha(s)ds \). The corresponding, time-inhomogeneous compound Poisson process will be denoted by \( N_{\alpha, B} \). In particular, the rate \( \alpha(t) \) may be chosen to be a random process, to model fluctuations of the promoter \[ \[8\] \[25\] \] or operator state \[ \[26\] \], leading to transcriptional bursts \[ \[27\] \], see below. The resulting process reads \( \rho(r, t) = \int_{-\infty}^{t} \phi(r, t-s)dN_{\alpha, B}(s) \). Moreover, following Berg \[ \[11\] \] instead of a continuous exponential distribution, we will also include a discrete, geometric distribution for the burst sizes \( B \).

Even if the time evolution of the protein density \( \rho(r, t) \) is no longer Markovian, we can write down its Laplace transform because, for a given protocol \( \alpha \), protein production is still a time-inhomogeneous Poisson process:

\[
\langle e^{-\lambda \phi} \rangle = \exp \left[ -\int_{-\infty}^{t} \alpha(s) \left( \frac{b(e^{\lambda \phi(r,t-s)} - 1)}{b(e^{\lambda \phi(r,t-s)} + 1) + 1} - ds \right) \right].
\]

The corresponding formula for the protein number is obtained by substituting \( \lambda \rightarrow 1 - e^{-\lambda} \) and \( \phi(r, t) \rightarrow \phi(\Omega, t) = \int_{\Omega} \phi(r, t)d^3r \). In particular, the average of the protein number \( M(\Omega, t) \) and its variance can be immediately calculated from the Laplace transform, yielding

\[
\langle M(\alpha) \rangle = b \int_{-\infty}^{t} \alpha(s)\phi ds, \tag{4a}
\]

\[
\langle M^2(\alpha) \rangle - \langle M(\alpha)^2 \rangle = b \int_{-\infty}^{t} \alpha(s)[1 + (2b - 1)\phi]ds. \tag{4b}
\]

with \( \phi = \phi(\Omega, t-s) \). Eqs. \[ \[24\] \] follow as a special case of \[ \[4a\] \] and \[ \[4b\] \] for a constant transcription rate, large burst size, and infintely fast mixing of molecules in a homogeneous cell volume, i.e., \( \phi(\Omega, t) = V_\Omega e^{-\gamma t} \).

Let us now consider the effect of a TF (here, a repressor) to the transcription rate \( \alpha_{\text{TF}} \) of a transcription unit (TU) gene under control of the TF. We first assume a given density of unbound TF within the sliding distance along the DNA from the operator site, and study the local kinetics of the TF. We explicitly describe the local kinetics of the repressor molecules through facilitated diffusion \[ \[10\] \[26\] \] near the binding site by considering three states of the operator (magnifying glass 3 in Fig. 1): transcription occurs at a constant rate \( \alpha \) whenever there is no repressor bound to the DNA at the target. Then, the repressor is either performing a local search by sliding in the vicinity of the target without binding to it specifically, or TF molecules, the mean number of which is determined by the given density, are just hovering in the surrounding space. The gene is considered silent when a repressor is bound at the operator. The linear Markov dynamics of TF binding can be explicitly solved by standard methods (see Supplementary Material (SM) \[ \[28\] \]). For example, the stationary protein level is obtained by averaging over \( \alpha \) in Eq. \[ \[4a\] \], but its variance will consist of three terms instead of the two in Eq. \[ \[4b\] \] because of time correlations in the transcription rate \( \alpha \).

Introducing the equilibrium constant \( K_{\text{BP}} \) for specific TF binding to the operator and assuming fast binding and unbinding, we integrate out the fast local search state in the three-state model. This leads to a simpler
model with telegraph noise at the operator, i.e., the gene is either silent or being transcribed at some effective rate \(a_{\text{eff}}\). The transitions between these two states occur without intermediates at rate \(r_{\text{on}}\) from silent to active and \textit{vice versa} with rate \(r_{\text{off}}\). Matching the stationary mean and the variance of the protein numbers in both processes, we relate the parameters of the telegraph model to the ones depicted in the magnifying glass 3 of Fig. 1 [28]. This is the description of a mesoscopic region around the target. Its length is 34 nm (100 base pairs), its diameter is that of DNA (2.4 nm) plus 30 nm (e.g., the length of lac repressor is 14 nm). With Eq. (3),

\[
p_{\text{on}}(r, t) = \frac{1}{1 + K_{\text{SP}}} \left( 1 + \int_{0}^{\infty} e^{-\lambda - t_{\alpha}^{\text{TF}}(s)} \int_{0}^{s} d\lambda \right),
\]

where \(\alpha = b_{\text{TF}}(\exp\{\lambda K\phi(\Omega, t - s)\} - 1)\) and \(K = (1 + K_{\text{SP}})K_{\text{NS}}\). The lower limit of the inner integral can be set to zero in our scenario (\(\alpha_{\text{TF}}(t < 0) = 0\)).

Eq. (6) is a central result of this study. It is general and allows quantitative analyses of various transcriptional and translational repression scenarios in any cellular structure and geometry. In particular, it takes into account the transcriptional pulsing [21] of the TU gene induced by the binding of the repressor. Eq. (6) even allows us to model RNAP binding and mRNA degradation by setting \(b_{\text{TF}} = 1\) and introducing, as the TF production rate, a new stochastic process \(\alpha_{\text{TF}}(t) = \alpha_{\text{TF}}N_{\text{mRNA}}(t)\) with a constant translation rate \(\gamma_{\text{TF}}\), and the number of transcripts \(N_{\text{mRNA}}\) given by an immigration-death process (equivalently an M/M/\infty queue) with mRNA production rate \(\alpha_{\text{TF}}\) and mRNA degradation rate \(\gamma_{\text{mRNA}}\). Since \(\gamma_{\text{mRNA}}\) is of the same order as typical TF search times in \textit{E. coli} [21, 29], inclusion of TF mRNA dynamics may be necessary in some cases. The scenario can be even further extended to include translational pulsing by modulating the immigration-death process \(N_{\text{mRNA}}\) with telegraph noise [25]. However, the expectation of Eq. (6) is yet to be solved for these \(\alpha_{\text{TF}}\) [25]. In the examples below, we use an approximation with a constant transcription and translation rate yielding on average 500 TF molecules per cell under stationary conditions. This number is in the ballpark of TF abundances for various levels of \textit{E. coli} regulation networks [32]. Special cases with low and high TF abundances will be studied separately.

We assume the TF gene to be in the center of a spherical nucleoid and the TU gene at a radial distance \(r\) from it. There is recent evidence [3] that the spatial distribution of TFs is highly inhomogenous. TFs bind to the DNA nonspecifically, hence under many growth conditions the TF concentration is higher in the nucleoid than in the surrounding volume. Inhomogeneities were also observed to affect fold repression. We thus assume that the diffusion constant \(D_{\text{N}}\) within the nucleoid is much smaller than in the surrounding cytosol due to crowding and nonspecific binding to the DNA (see SM [28] for comparison with the model in Ref. [3]). The nucleoid is surrounded by the volume \(V_{\text{C}} - V_{\text{N}}\), where \(V_{\text{C}} = 4\pi R_{\text{C}}^{3} / 3 = 1\) \(\mu\text{m}^{3}\) and \(V_{\text{N}} = 4\pi R_{\text{N}}^{3} / 3 = 0.2\) \(\mu\text{m}^{3}\) are the cell and nucleoid volumes. The density \(\rho_{\text{TF}}(r, t)\) is subject to the radial diffusion equation. In Eq. (6), \(\phi(\Omega, t) \approx V_{\text{TF}}\phi(r, t)\) obeys

\[
\frac{\partial \phi}{\partial t} = D_{\text{N}} \left( \frac{\partial^{2} \phi}{\partial r^{2}} + \frac{2 \partial \phi}{r \partial r} \right) - \gamma \phi, \quad \text{for } 0 \leq r \leq R_{\text{N}}
\]

\[
\frac{\partial \phi}{\partial t} = -\frac{4\pi R_{\text{N}}^{2} D_{\text{N}} \partial \phi}{V_{\text{C}} - V_{\text{N}} \partial r} - \gamma \phi, \quad \text{for } r = R_{\text{N}},
\]
with a dilution rate $\gamma = 1/20 \text{ min}^{-1}$ due to cell growth and with the condition that the TFs are initially uniformly distributed in the close vicinity (say, within a radius $R_1 = 20 \text{ nm}$) of the TF gene. This is justified from the observed localization of transcripts near their transcription site in bacteria \[33\]. The explicit solution of Eqs. \[7\] for our spherical geometry is \[32\]

$$
\phi(r, t) = \frac{e^{-\gamma t}}{V_C} + \frac{3}{2\pi} \sum_{n=1}^{\infty} e^{-(2n^2 + 1)\gamma t} \frac{\sin(q_n r)}{R_N r} \times
\frac{k^2 \psi_n^4 + 3(k + 3) \psi_n^2 + 9}{k^2 \psi_n^4 + 9(k + 1) \psi_n^2} \cdot \frac{\sin(\theta_n) - \theta_n \cos(\theta_n)}{R_1 \theta_n^2},
$$

where $\psi_n = q_n R_N$, $\theta_n = q_n R_1$ and $k = (V_C - V_N)/V_N$, and the $q_n$ are the positive solutions of $(3 + k R_1 q^2) \tan(q R_1) = 3 q R_1$. Eq. \[5\] is our other central result.

Fig. 2 shows the probabilities \[6\] as function of time for short and long distances between the TF and TU genes. Accordingly, the distance impacts vastly the regulation efficiency: the response is significantly stronger for short and long distances between the TF and TU genes. The distance dependence of gene regulation efficiency and stochasticity in bacteria. Intracellular structure and non-specific binding to the DNA are taken into account in Eqs. \[5\] for the TF-TU system, in particular, the variance of the expression levels in the long time limit can be small, even if the transient response shows considerable variation. The same applies to expression fluctuations. Experimental observations \[36\] show that the protein level fluctuations are, in general, determined by the mean expression level, and are independent of system details. The dependence of protein number fluctuations on the TF-TU distance under stationary conditions needs to be explored further.

Concluding, we established a quantitative model for the distance dependence of gene regulation efficiency and stochasticity in bacteria. Intracellular structure and non-specific binding to the DNA are taken into account in terms of an inhomogeneous diffusion rate. The binding at the target is facilitated by a local search process, which was modeled by an intermediate fast degree of freedom. Significant spatial effects in the regulation efficiency was demonstrated, strongly supporting the regulation hypothesis for gene colocalization. We note that more precise models, for instance, with multiple TFs sliding simultaneously near the target can be solved, as well. The expressions are more elaborate (except for infinite numbers) but the binding probabilities show roughly the same behavior as above. It will be of interest to compare transient response to internal and external signals, as the gene location is known to depend on the type of signal \[37\].

We acknowledge funding from the Academy of Finland.

Figure 2: Transient response to a change in TF transcription rate. The circles and squares are the probabilities \[6\] for TF-TU gene distances $r = 0.05$ and $0.33 \text{ \mu m}$, and the solid and dashed lines show the corresponding mean field approximations (see main text). The inset shows the variation of TF concentration around the target site at various TF-TU distances. The equilibrium constants are $K_{NS} = 10$, $K_{SP} = 1000$, and the rest of the parameters as described in the main text.
(FiDiPro scheme).

[1] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, Nature Physics 6, 888 (2010); C. Castellano, S. Fortunato, and V. Loretto, Rev. Mod. Phys. 81, 591 (2009).
[2] J. Elf, G.-W. Li, and X. S. Xie, Science 316, 1191 (2007).
[3] T. E. Kuhlman and E. C. Cox, Mol. Syst. Biol. 8, 610 (2012).
[4] Z. Wunderlich and L. A. Mirny, Nucleic Acids Res. 36, 3570 (2008).
[5] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, Molecular Biology of the Cell, 5th ed., (Garland Science, New York, 2008).
[6] G. Kolesov, Z. Wunderlich, O. N. Laikova, M. S. Gefand, and L. A. Mirny, Proc. Natl. Acad. Sci. USA 104, 13948 (2007); M. di Stefano, A. Rosa, V. Belcastro, D. di Bernardo, and C. Micheletti, PLoS Comput. Biol. (at press).
[7] A. Eldar and M. B. Elowitz, Nature 467, 167 (2010).
[8] J. Paulsson, Phys. Life Rev. 2, 157 (2005). See also References therein.
[9] T. B. Kepler and T. C. Elston, Biophys. J. 81, 3116 (2001).
[10] N. Friedman, L. Cai, and X. S. Xie, Phys. Rev. Lett. 97, 168302 (2006).
[11] O. G. Berg, J. Theor. Biol. 71, 587 (1978).
[12] A translational burst refers to the number of times the transcripts (mRNA) are translated into proteins.
[13] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf, Science 336, 1595 (2012); X. S. Xie, P. J. Choi, G. Li, N. K. Lie, and G. Lia, Ann. Rev. Biophys. 37, 417 (2008).
[14] B. van den Broek, M. A. Lomholt, S.-M. J. Kalisch, R. Metzler, and G. J. L. Wuite, Proc. Natl. Acad. Sci. USA 105, 15738 (2008).
[15] A. D. Riggs, S. Bourgeois, and M. Cohn, J. Mol. Biol. 53, 401 (1970).
[16] P. H. von Hippel and O. G. Berg, J. Biol. Chem. 264, 675 (1989).
[17] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj, J. Phys. A 42, 434013 (2009); M. Sheinman, O. Bénichou, Y. Kafri, and R. Voituriez, Rep. Prog. Phys. 75, 026601 (2012).
[18] M. Coppey, O. Bénichou, R. Voituriez, and M. Moreau, Biophys. J. 87, 1640 (2004); M. Slutsky and L. A. Mirny, Biophys. J. 87, 4021 (2004); T. Hu, A. Y. Grosberg, and B. I. Shklovskii, Biophys. J. 90, 2731 (2006); S. E. Halford and J. F. Marko, Nucleic Acids Res. 32, 3040 (2004); A. B. Kolomeisky, Phys. Chem. Chem. Phys. 13, 2088 (2011).
[19] M. Bauer and R. Metzler, Biophys. J. 102, 2321 (2012); J. Reingruber and D. Holcman, Phys. Rev. E 84, 020901 (2011).
[20] K. V. Klenin, H. Merlitz, J. Langowski, and C. X. Wu, Phys. Rev. Lett. 96, 018104 (2006); M. Sheinman and Y. Kafri, Phys. Biol. 6, 0160033 (2009).
[21] G. Goffano, D. Marenduzzo, and E. Orlandini, Phys. Rev. E 85, 021919 (2012); M. A. Lomholt, B. v. d. Broek, S.-M. J. Kalisch, G. J. L. Wuite, and R. Metzler, Proc. Natl. Acad. Sci. USA 106, 8204 (2009).
[22] C. Loverdo, O. Bénichou, R. Voituriez, A. Biebricher, I. Bonnet, and P. Desiolles, Phys. Rev. Lett. 102, 188101 (2009); Y. Meroz, I. Eliazar, and J. Klafter, J. Phys. A 42, 434012 (2009). E. F. Koslover, M. A. Díaz de la Rosa, and A. J. Spakowitz, Biophys J. 101, 856 (2011); C. A. Brackley, M. E. Cates, and D. Marenduzzo, Phys. Rev. Lett. 109, 168103 (2012); M. Bauer and R. Metzler, PLoS ONE 8, e53956 (2013).
[23] P. Dröge and B. Müller-Hill, Bioessays 23, 179 (2001).
[24] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. USA 105, 17256 (2008).
[25] B. Kaufmann and A. van Oudenaarden, Curr. Opin. Genetics Dev. 17, 107 (2007).
[26] J. S. van Zon, M. J. Morelli, and P. R. ten Wolde, Biophys. J. 91, 4350 (2006).
[27] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, Cell 123, 1025 (2005); L.-h. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, Nat. Genet. 43, 554 (2011).
[28] Supplementary Material.
[29] F. W. Crawford and M. A. Suchard, J. Math. Biol. 65, 553 (2012); F. W. Crawford, E-print arXiv:1301.1305. See also Refs. therein.
[30] J. Yu, J. Xiao, X. Ren, K. Lao, and X. S. Xie, Science 311, 1600 (2006).
[31] P. J. Choi, L. Cai, K. Frieda, and X. S. Xie, Science 322, 442 (2008).
[32] H. S. Carslaw and J. C. Jaeger, Conduction of Heat in Solids, Oxford University Press, Oxford (1959).
[33] P. M. Llopis, A. F. Jackson, O. Slisarenko, I. Surovtsev, J. Heinritz, T. Emonet, and C. Jacobs-Wagner, Nature 466, 77 (2010).
[34] I. Lozada-Chávez, V. E. Angarica, J. Collado-Vides, and B. Contreras-Moreira, J. Mol Biol. 4, 627 (2008).
[35] S. C. Janga, H. Salgado, and A. Martinez-Antonio, Nucleic Acids Res. 37, 3680 (2009).
[36] H. Salmon, N. Brenner, C.-k. Tung, N. Elyahu, E. Stolovicki, L. Moore, A. Libchaber, and E. Braun, PRL 108, 238105 (2012); Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, X. S. Xie, Science 329, 533 (2010).
[37] S. C. Janga, H. Salgado, J. Collado-Vides, and A. Martinez-Antonio, J. Mol. Biol. 368, 263 (2007).