Suppression and enhancement of transcriptional noise by DNA looping

Jose M. G. Vilar[1,2,*] and Leonor Saiz[3,†]

[1]*Biophysics Unit (CSIC-UPV/EHU) and Department of Biochemistry and Molecular Biology,*
University of the Basque Country UPV/EHU, P.O. Box 644, 48080 Bilbao, Spain
[2]*IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain*
[3]*Department of Biomedical Engineering, University of California, 451 East Health Sciences Drive, Davis, California 95616, USA*

DNA looping has been observed to enhance and suppress transcriptional noise but it is uncertain which of these two opposite effects is to be expected for given conditions. Here, we derive analytical expressions for the main quantifiers of transcriptional noise in terms of the molecular parameters and elucidate the role of DNA looping. Our results rationalize paradoxical experimental observations and provide the first quantitative explanation of landmark individual-cell measurements at the single molecule level on the classical *lac* operon genetic system [Choi, L. Cai, K. Frieda, and X. S. Xie, Science **322**, 442 (2008)].

I. INTRODUCTION

Gene expression, the process that leads to functional biomolecules from the information encoded in genes, is carried out by inherently stochastic events [1]. Very often, the underlying stochasticity is not effectively averaged out and noise, in the form of random fluctuations, propagates throughout the system [2–8]. A key regulatory step is the binding of transcription factors to DNA, which control how effectively the RNA polymerase transcribes the genes [9]. DNA looping has a widespread prominent role in this type of situations because it allows transcription factors to bind simultaneously single and multiple DNA sites and to contact the RNA polymerase from distal sites [9–11].

Typical approaches to studying transcriptional noise have been based on, among other methodologies, stochastic simulations [12], linear noise approximations [13], Langevin dynamics [14], and analytical solutions of the master equation [15]. The Fano factor, $F = \sigma^2/\mu$, and coefficient of variation, $c_v = \sigma/\mu$, are the quantities most commonly used to characterize transcriptional noise in terms of the mean, μ, and standard deviation, σ, of the number of mRNA transcripts. Despite intensive research, explicit analytical expressions for these quantities, whether exact or approximate, remain scarce except for a few notable exceptions [15–17]. The presence of DNA looping is particularly challenging to deal with because it introduces additional complexity that, so far, has been possible to study only through stochastic simulations [18–20] or numerical calculations [21]. Without explicit analytical expressions, it is difficult to comprehend and fully understand how the different molecular components impact transcriptional noise and how to manipulate them to affect the system behavior. Experiments show that DNA looping can both enhance [22] and suppress [23] transcriptional noise but it is uncertain which of these two opposite effects is to be expected for given conditions.

Here, we obtain explicit analytical expressions for the Fano factor and coefficient of variation in terms of the molecular

parameters and find conditions that determine whether DNA looping enhances or suppresses noise. We focus explicitly on the mode of regulation of the *lac* operon, the proverbial *E. coli* genetic system that regulates and produces the enzymes needed to metabolize lactose [24,25]. In this system, the main regulator is the *lac* repressor [26], which upon binding to the main operator prevents the RNA polymerase from transcribing the three genes used in lactose metabolism. We also consider the presence of an auxiliary operator where the repressor can bind specifically without preventing transcription (Fig. 1). Because the repressor has two DNA binding domains, it can bind two operators simultaneously by looping the intervening DNA.

II. MATHEMATICAL DESCRIPTION

The canonical description of this type of multistate transcription systems considers that there is a set of transcriptional states s and that mRNA, m, is produced at a rate g_s for each transcription state [10]. Typically, the mRNA degradation rate λ is independent of s. We consider five transcriptional states, which are labeled as illustrated in Fig. 1. The transcription rates g_s can be expressed as the components of the vector

$$\mathbf{g} \equiv (g_{\text{on}} \quad g_{\text{on}} \quad 0 \quad 0 \quad 0)^{\text{T}}, \tag{1}$$

which specifies that transcription takes place at a rate g_{on} only when the main operator is free.

Transitions between transcriptional states result from the binding and unbinding of the repressor. We specify the transition rates $k_{s,s'}$ from the state s to the state s' through the elements of the matrix

$$\mathbf{k} \equiv \begin{pmatrix} 0 & n_X k_{\text{on}} & n_X k_{\text{on}} & 0 & 0 \\ k_{\text{off}-\text{A}} & 0 & 0 & k_{\text{loop}} & n_X k_{\text{on}} \\ k_{\text{off}-\text{M}} & 0 & 0 & k_{\text{loop}} & n_X k_{\text{on}} \\ 0 & k_{\text{off}-\text{M}} & k_{\text{off}-\text{A}} & 0 & 0 \\ 0 & k_{\text{off}-\text{M}} & k_{\text{off}-\text{A}} & 0 & 0 \end{pmatrix}, \tag{2}$$

where k_{on} is the association rate of the repressor for an operator; $k_{\text{off}-\text{M}}$ and $k_{\text{off}-\text{A}}$ are the dissociation rates of the repressor from the main and auxiliary operators, respectively; k_{loop} is the rate of loop formation when the repressor is bound to one operator; and n_X is the average number of active repressors

*j.vilar@ikerbasque.org
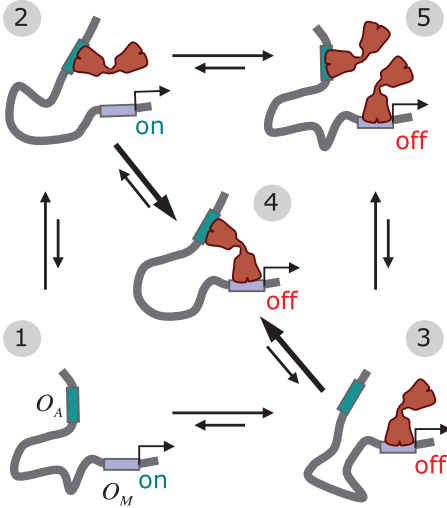†lsaiz@ucdavis.edu

FIG. 1. (Color online) Transcriptional states of the *lac* operon (labeled with encircled numbers as used in the main text) and the possible transitions among them (arrows). DNA is represented by a wavy thick line with boxes for the main (O_M) and auxiliary (O_A) operators, and the repressor is cartooned bound to the operators.

(see Appendix). This description was originally developed in Ref. [18] and has been shown to accurately describe the *lac* operon under an exhaustive range of experimental conditions [27,28].

To obtain more compact expressions, we express the dissociation and the looping rates in terms of the repressor-operator association constants, K_M and K_A, and looping local concentration, n_L, as $k_{\text{off-M}} = k_{\text{on}}/K_M$, $k_{\text{off-A}} = k_{\text{on}}/K_A$, and $k_{\text{loop}} = n_L k_{\text{on}}$.

The time evolution of the joint probability $P(m,s)$ of the number m of mRNA molecules and the system state s is governed by the master equation,

$$\frac{dP(m,s)}{dt} = \sum_{s'} [k_{s',s} P(m,s') - k_{s,s'} P(m,s)]$$
$$+ g_s [P(m-1,s) - P(m,s)]$$
$$+ \lambda [(m+1)P(m+1,s) - m P(m,s)], \quad (3)$$

which takes into account the transitions between transcriptional states, mRNA production, and mRNA degradation.

III. RESULTS

To compute the Fano factor and the coefficient of variation, we need to consider only the steady state. We proceed by rewriting the joint probability in terms of the marginal probability $P_s = \sum_m P(m,s)$ and conditional probability $P(m|s) = P(m,s)/P_s$, which we use to compute the conditional averages $\langle m \rangle_s = \sum_m m P(m|s)$ and $\langle m^2 \rangle_s = \sum_m m^2 P(m|s)$.

The steady-state expression of the marginal probability P_s is obtained by solving $0 = \sum_{s'} [k_{s',s} P_{s'} - k_{s,s'} P_s]$, which follows straightforwardly from Eq. (3). We write the solution using the statistical weights Z_s. In vector form, the marginal probabilities are given by

$$\mathbf{P} = \mathbf{Z}/\|\mathbf{Z}\|_1, \quad (4)$$

where $\|\mathbf{Z}\|_1$ is the expression of the partition function using the one norm. In the case of the *lac* operon, the statistical weights are

$$\mathbf{Z} \equiv \begin{pmatrix} 1 & n_X K_A & n_X K_M & n_X n_L K_A K_M & n_X^2 K_A K_M \end{pmatrix}^{\mathrm{T}}. \quad (5)$$

The steady-state equation for the conditional probability follows from substituting $P(m,s) = P(m|s)P_s$ in Eq. (3) with $dP(m,s)/dt = 0$. After using the detailed balance principle, $k_{s',s} P(s')/P(s) = k_{s,s'}$, and algebraic manipulations, we obtain

$$0 = \sum_{s'} \left(k_{s,s'} - \delta_{s',s} \sum_{s''} k_{s,s''} \right) P(m|s')$$
$$+ g_s [P(m-1|s) - P(m|s)]$$
$$+ \lambda [(m+1)P(m+1|s) - m P(m|s)], \quad (6)$$

where $\delta_{s',s}$ represents the Kronecker delta.

The conditional averages $\langle m \rangle_s$ and $\langle m^2 \rangle_s$ are obtained implicitly by multiplying Eq. (6) by m and m^2, respectively, and summing over all the values of m, resulting in $0 = \sum_{s'} W_{s,s'} \langle m \rangle_{s'} + g_s - \lambda \langle m \rangle_s$ and $0 = \sum_{s'} W_{s,s'} \langle m(m-1) \rangle_{s'} + 2g_s \langle m \rangle_s - 2\lambda \langle m(m-1) \rangle_s$, with $W_{s,s'} = k_{s,s'} - \delta_{s',s} \sum_{s''} k_{s,s''}$. The explicit solutions are expressed in matrix form as

$$\langle \mathbf{m} \rangle = (\lambda \mathbf{I} - \mathbf{W})^{-1} \mathbf{g},$$
$$\langle \mathbf{m}^2 \rangle = \langle \mathbf{m} \rangle + (2\lambda \mathbf{I} - \mathbf{W})^{-1} (2\mathbf{g} \circ \langle \mathbf{m} \rangle), \quad (7)$$

where $\mathbf{I}$ is the identity matrix and $\langle \mathbf{m} \rangle$, $\langle \mathbf{m}^2 \rangle$, and $\mathbf{W}$ are the vectors and matrix with elements $\langle m \rangle_s$, $\langle m^2 \rangle_s$, and $W_{s,s'}$, respectively. The symbol $\circ$ represents the Hadamard product.

The mean μ and variance σ^2 are computed from the conditional averages as

$$\mu = \langle \mathbf{m} \rangle \cdot \mathbf{P}, \quad \sigma^2 = \langle \mathbf{m}^2 \rangle \cdot \mathbf{P} - \mu^2. \quad (8)$$

The analytic expressions of μ and σ^2 in terms the molecular parameters follow straightforwardly from Eq. (8) using Eqs. (4) and (7) with Eqs. (1), (2), and (5). The mean mRNA content is given by

$$\mu = \frac{g_{\text{on}} K_A n_X + g_{\text{on}}}{\lambda \{K_A n_X [K_M (n_L + n_X) + 1] + K_M n_X + 1\}}, \quad (9)$$

which has a relatively simple expression. The general expression of the variance, in contrast, is much more intricate and is given explicitly in the Supplemental Material [29].

To simplify the general expressions, we considered first how adding a small looping contribution to the system changes the noise properties. Explicitly, we calculated the derivative of F with respect to n_L when n_L is small:

$$\left. \frac{dF}{dn_L} \right|_{n_L \to 0}$$

$$= \frac{g_{\text{on}} K_A K_M^2 n_X (\lambda K_M + k_{\text{on}} - K_M^2 k_{\text{on}} n_X^2)}{(K_A n_X + 1)(K_M n_X + 1)^2 (\lambda K_M + K_M k_{\text{on}} n_X + k_{\text{on}})^2}. \quad (10)$$

This expression shows that looping has a dual role. It decreases the Fano factor for

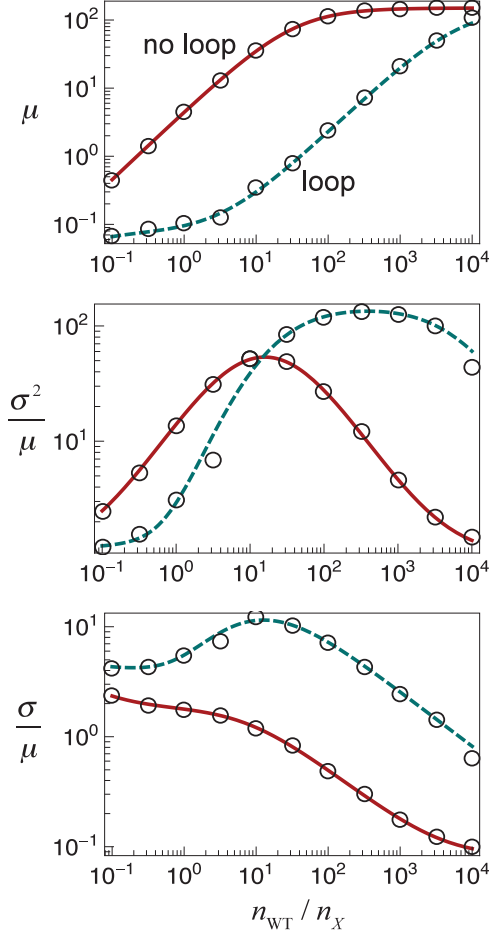$$n_X^2 K_M^2 > 1 + \lambda K_M/k_{\text{on}}, \quad (11)$$

FIG. 2. (Color online) Mean (μ), Fano factor (σ^2/μ), and co-efficient of variation (σ/μ) as functions of the inverse of the normalized average number of active repressors (n_{WT}/n_X) for a system with (dashed line) and without (continuous line) looping. Here, $n_{\text{WT}} = 10$ molec is the average number of wild-type repressors. The values of the parameters are $k_{\text{on}} = 0.0033$ molec^{-1} s^{-1}, $g_{\text{on}} = 0.5$ molec s^{-1}, $\lambda = 0.0033$ s^{-1}, $n_L = 606$ molec (looping) or $n_L = 0$ molec (no looping), $K_M = 3.3$ molec^{-1}, and $K_A = 0.33$ molec^{-1}, which correspond to the estimated experimental values. The mean and variance were computed using the expressions in the Supplemental Material [29] resulting from Eq. (8). The circles represent the results obtained from stochastic simulations [Eq. (13)].

when the occupancy of the main site is sufficiently high, and increases it otherwise. The intuitive explanation is that the Fano factor has a maximum for intermediate occupancy of the main site and looping increases the occupancy towards this maximum for low occupancies or away from it in the opposite regime. In the case of c_v, this approach leads to an expression for $(dc_v/dn_L)_{n_L \to 0}$ that is positive for all parameter values. Explicitly, the derivative of c_v with respect to n_L for small n_L is given by

$$\left. \frac{dc_v}{dn_L} \right|_{n_L \to 0}$$

$$= \frac{1}{2c_v} \left. \frac{dc_v^2}{dn_L} \right|_{n_L \to 0}$$
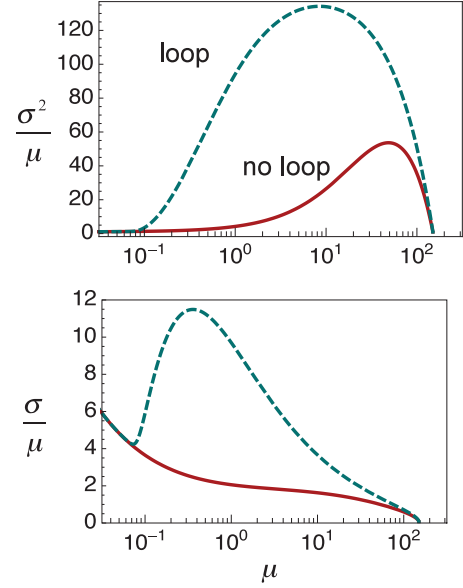
FIG. 3. (Color online) Fano factor (σ^2/μ) and coefficient of variation (σ/μ) as functions of the mean (μ) when the average number of active repressors (n_X) is changed for a system with (dashed line) and without (continuous line) looping. The values of the other parameters are the same as in Fig. 2.

$$= \frac{1}{2c_v} \frac{\lambda K_A K_M n_X}{(g_{\text{on}} K_A n_X + g_{\text{on}})(\lambda K_M + K_M k_{\text{on}} n_X + k_{\text{on}})^2}$$

$$\times \left\{ \lambda K_M^2 (\lambda + g_{\text{on}}) + K_M k_{\text{on}} [g_{\text{on}} + 2(\lambda + \lambda K_M n_X)] \right.$$

$$\left. + (K_M k_{\text{on}} n_X + k_{\text{on}})^2 \right\}, \tag{12}$$
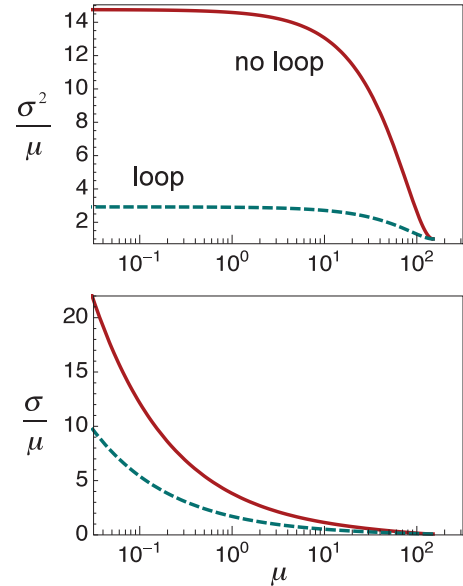
FIG. 4. (Color online) Fano factor (σ^2/μ) and coefficient of variation (σ/μ) as functions of the mean (μ) when the affinity of the main site (K_M) is changed for a system with (dashed line) and without (continuous line) looping. The values of the other parameters are the same as in Fig. 2 with $n_X = 10$ molec.
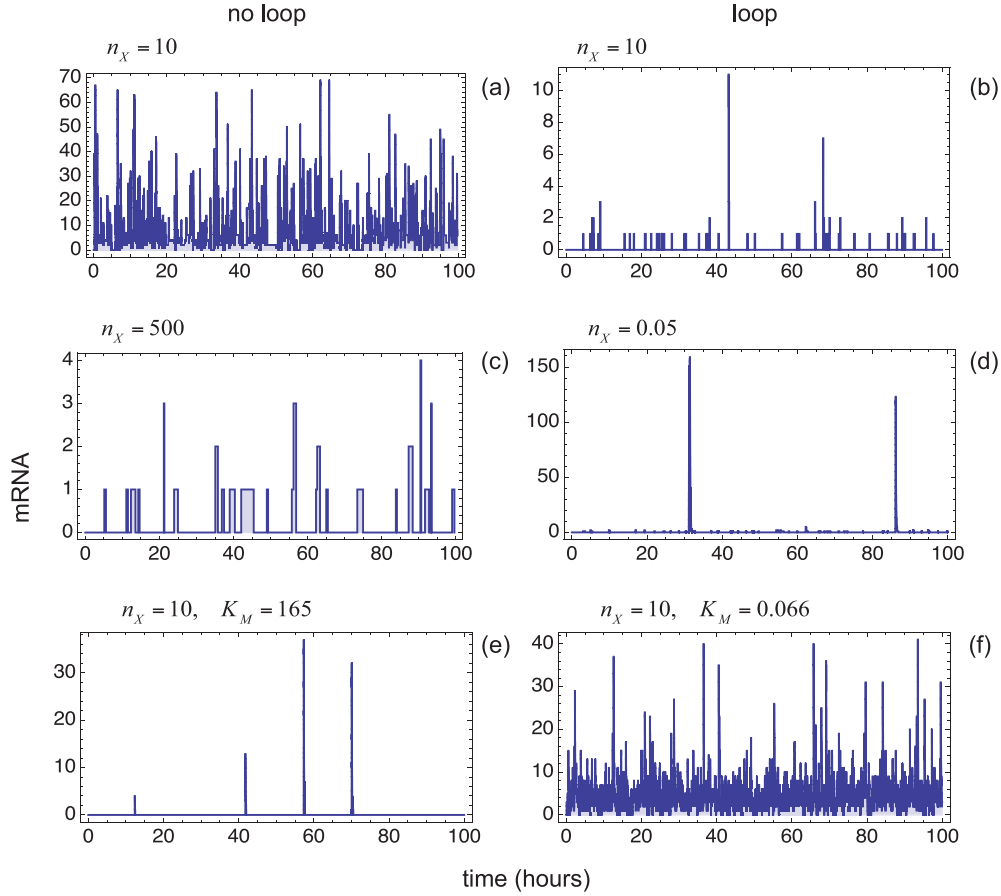
FIG. 5. (Color online) Representative time courses of the number of mRNA transcripts from stochastic simulations of the system [Eq. (13)] without (left panels) and with (right panels) looping. The values of the parameters are the same as in Fig. 2 except for those indicated in each panel. Panels (a) and (b) correspond to the wild-type case in the absence of inducer, panels (c) and (d) correspond to adjusting the mean by changing n_X, and panels (e) and (f) correspond to adjusting the mean by changing K_M.

which is always positive. Therefore, the presence of looping always increases the coefficient of variation for small values of n_L.

Experiments in the *lac* operon have reported that looping decreases F and increases c_v [23]. In agreement with the experimental data, our results show that looping increases c_v and, for the values of the parameters corresponding to the experimental conditions, Eq. (11) indicates that indeed the system is in a regime in which looping decreases F (Fig. 2). We considered the exact analytic expressions [29] for the experimental parameters of the *lac* operon, and found that these results, obtained for small values of n_L, also hold true for the actual value of the looping strength (Fig. 2).

We further confirmed our results with stochastic simulations (circles in Fig. 2) performed using the Doob-Gillespie algorithm [30,31] with transitions

$$s \xrightarrow{k_{s,s'}} s', \quad m \xrightarrow{g_s} m+1, \quad m \xrightarrow{\lambda m} m-1. \quad (13)$$
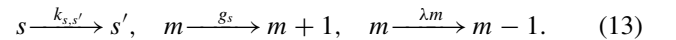
Earlier numerical calculations showed that looping increases F when it is plotted as a function of μ by changing the value of n_X [21]. Indeed, we recapitulate this result from Eq. (8) (Fig. 3). To provide analytical evidence for this effect, we considered a low affinity auxiliary operator and expanded in powers of K_A the result of substituting n_X as a function of μ. Explicitly, the expression of n_X in terms of the mean and the other parameters obtained from Eq. (9) is given by

$$n_X(\mu) = \frac{g_{on}K_A - \lambda\mu K_A - \lambda\mu K_A K_M n_L - \lambda\mu K_M}{2\lambda\mu K_A K_M} + \frac{\sqrt{4\lambda\mu K_A K_M(g_{on} - \lambda\mu) + \{K_A[\lambda\mu(K_M n_L + 1) - g_{on}] + \lambda\mu K_M\}^2}}{2\lambda\mu K_A K_M},$$

(14)

which after substitution in the expression of the Fano factor and expanding in powers of K_A leads to

$$F[n_X(\mu)] = 1 + \frac{\lambda\mu K_M(g_{on} - \lambda\mu)}{g_{on}k_{on} + \lambda^2\mu K_M}\left(1 + \frac{g_{on}K_A^2 k_{on}n_L^2}{g_{on}k_{on} + \lambda^2\mu K_M}\right) + O(K_A^3). \quad (15)$$

Therefore, our results show that looping always increases $F[n_X(\mu)]$ when the affinity of the auxiliary site is sufficiently small. [Note that Eq. (9) implies $g_{on} - \lambda\mu > 0$.] However, this increase, in contrast to what was assumed recently [21], is not a universal
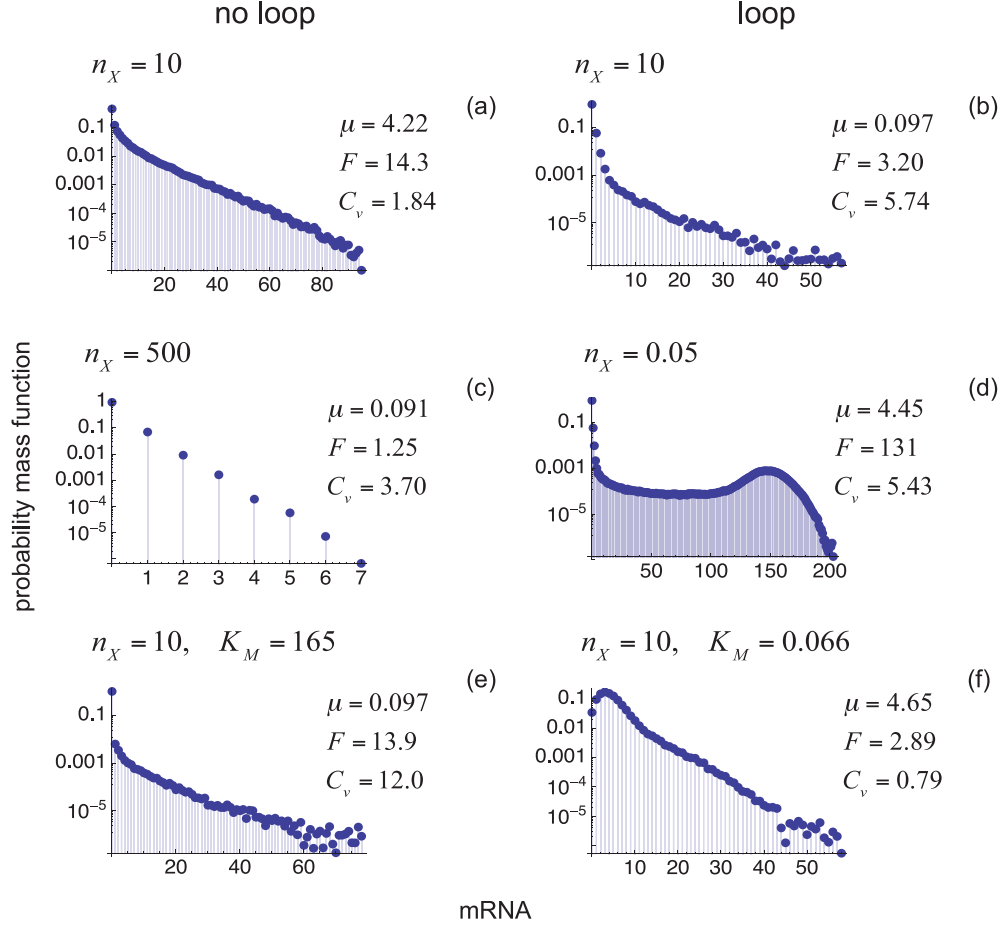
FIG. 6. (Color online) Probability mass functions for the number of mRNA transcripts from stochastic simulations for the same cases as in Fig. 5.

property of F as a function of μ. Consider, for instance, that μ is adjusted by changing K_M rather than n_X. From Eq. (9), we have

$$K_M(\mu) = \frac{(K_A n_X + 1)(g_{\text{on}} - \lambda\mu)}{\lambda\mu n_X[K_A(n_L + n_X) + 1]}. \quad (16)$$

Expanding in powers of K_A the result of substituting K_M as a function of μ, we obtain

$$F[K_M(\mu)] = 1 + \frac{(g_{\text{on}} - \lambda\mu)^2}{\lambda(g_{\text{on}} - \lambda\mu) + g_{\text{on}}k_{\text{on}}n_X}$$
$$\times \left[1 - \frac{g_{\text{on}}K_A k_{\text{on}}n_L n_X}{\lambda(g_{\text{on}} - \lambda\mu) + g_{\text{on}}k_{\text{on}}n_X}\right] + \text{O}(K_A^2), \quad (17)$$

which shows that looping always decreases $F[K_M(\mu)]$ when the affinity of the auxiliary site is sufficiently small. The full results from Eq. (8) show that looping, indeed, decreases $F[K_M(\mu)]$ for the experimentally observed values of the parameters (Fig. 4). Note that, as functions of the mean, the coefficient of variation and Fano factor are both affected by looping essentially in the same way because they are related to each other through $c_v^2(\mu) = F(\mu)/\mu$ (Figs. 3 and 4).

Simulation results confirm that the way of adjusting the mean, either through n_X or K_M, has profound differences in the system properties, including time courses (Fig. 5) and probability mass functions (Fig. 6).

IV. DISCUSSION

In summary, we have provided the first analytical expressions that quantify transcriptional noise in terms of the molecular parameters in a system with DNA looping. Our results show that looping can enhance and suppress transcriptional noise in a way that matches the experimental observations. For fixed parameter values, looping decreases the Fano factor when the occupancy of the main site is sufficiently high and increases it in the opposite limit. In this case, the coefficient of variation always increases. When the mean is fixed, both the Fano factor and coefficient of variation behave in the same way, with conditions leading to an increase of noise, as when the average number of active repressors is used to adjust the mean, and to a decrease of noise, as when the main site association constant is used to adjust the mean. Overall, our results show that DNA looping provides a highly versatile mechanism to regulate both average and noisy properties of genetic systems.

APPENDIX: ASSOCIATION RATES IN TERMS OF THE AVERAGE NUMBER OF ACTIVE REPRESSORS

Consider a cell with N_X active repressors, N_I repressors inactivated by an inducer I, and N_O repressors bound to operator DNA. Since the repressor's binding to the inducer is much faster than its binding to the operator, the average number of active n_X and inactive n_I repressors are related to each other through $n_I = n_X f_I([I])$, where $f_I([I])$ is an increasing function of the inducer concentration $[I]$ with $f_I(0) = 0$. Conservation of the number of repressors leads to $N_T - N_O = (1 + f_I([I]))n_X$, where N_T is the total number of repressors. The association rate of the repressors to the operator is given by $n_X^{(0)} k_{\text{on}}$, with $n_X^{(0)} = N_T/(1 + f_I([I]))$, for $N_O = 0$ and by $n_X^{(1)} k_{\text{on}} = n_X^{(0)} k_{\text{on}}(N_T - 1)/N_T$ for $N_O = 1$. Typical experimental values of N_T range from 10 to 900 and considering $(N_T - 1)/N_T \approx 1$ is an excellent approximation for all the experimental setups. Therefore, we use $n_X^{(1)} = n_X^{(0)} = n_X$.

[1] 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).

[2] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, Nat. Rev. Genet. **6**, 451 (2005).

[3] D. A. Charlebois, N. Abdennur, and M. Kaern, Phys. Rev. Lett. **107**, 218101 (2011).

[4] B. Munsky, G. Neuert, and A. van Oudenaarden, Science **336**, 183 (2012).

[5] H. Salman, N. Brenner, C. K. Tung, N. Elyahu, E. Stolovicki, L. Moore, A. Libchaber, and E. Braun, Phys. Rev. Lett. **108**, 238105 (2012).

[6] J. M. G. Vilar, BMC Syst. Biol. **4**, 152 (2010).

[7] G. M. Suel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz, Science **315**, 1716 (2007).

[8] L. Saiz, J. Phys.: Condens. Matter **24**, 193102 (2012).

[9] M. Ptashne and A. Gann, *Genes and Signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2002).

[10] J. M. G. Vilar and L. Saiz, Biophys. J. **104**, 2574 (2013).

[11] J. M. G. Vilar and L. Saiz, Phys. Rev. Lett. **96**, 238103 (2006).

[12] C. V. Rao, D. M. Wolf, and A. P. Arkin, Nature **420**, 231 (2002).

[13] J. Elf and M. Ehrenberg, Genome Res. **13**, 2475 (2003).

[14] P. Rue and J. Garcia-Ojalvo, Annu. Rev. Biophys. **42**, 605 (2013).

[15] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. USA **105**, 17256 (2008).

[16] J. Paulsson, Nature **427**, 415 (2004).

[17] A. Coulon, O. Gandrillon, and G. Beslon, BMC Syst. Biol. **4**, 2 (2010).

[18] J. M. G. Vilar and S. Leibler, J. Mol. Biol. **331**, 981 (2003).

[19] M. J. Morelli, P. R. Ten Wolde, and R. J. Allen, Proc. Natl. Acad. Sci. USA **106**, 8101 (2009).

[20] L. Saiz and J. M. G. Vilar, Mol. Syst. Biol. **2**, 0024 (2006).

[21] A. Sanchez, H. G. Garcia, D. Jones, R. Phillips, and J. Kondev, PLoS Comput. Biol. **7**, e1001100 (2011).

[22] Z. Hensel, X. Weng, A. C. Lagda, and J. Xiao, PLoS Biol. **11**, e1001591 (2013).

[23] P. J. Choi, L. Cai, K. Frieda, and X. S. Xie, Science **322**, 442 (2008).

[24] B. Müller-Hill, *The lac Operon: A Short History of a Genetic Paradigm* (Walter de Gruyter, Berlin, New York, 1996).

[25] F. Jacob and J. Monod, J. Mol. Biol. **3**, 318 (1961).

[26] M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu, Science **271**, 1247 (1996).

[27] J. M. G. Vilar and L. Saiz, ACS Synth. Biol. **2**, 576 (2013).

[28] L. Saiz and J. M. G. Vilar, Nucleic Acids Res. **36**, 726 (2008).

[29] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.89.062703 for supplementary text with general analytic expressions for the mean, variance, Fano factor, and coefficient of variation and for a Mathematica(R) notebook with the source code to obtain these expressions.

[30] J. L. Doob, Trans. Am. Math. Soc. **58**, 455 (1945).

[31] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).