Stochastic Model of Supercoiling-Dependent Transcription

C. A. Brackley,[1] J. Johnson,[1] A. Bentivoglio,[3] S. Corless,[2] N. Gilbert,[2] G. Gonnella,[3] and D. Marenduzzo[1]
[1]SUPA, School of Physics and Astronomy, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh EH9 3FD, United Kingdom
[2]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, The University of Edinburgh, Edinburgh EH4 2XU, United Kingdom
[3]Dipartimento di Fisica, Università di Bari and INFN, Sezione di Bari, 70126 Bari, Italy
(Received 21 October 2014; published 27 June 2016)

We propose a stochastic model for gene transcription coupled to DNA supercoiling, where we incorporate the experimental observation that polymerases create supercoiling as they unwind the DNA helix and that these enzymes bind more favorably to regions where the genome is unwound. Within this model, we show that when the transcriptionally induced flux of supercoiling increases, there is a sharp crossover from a regime where torsional stresses relax quickly and gene transcription is random, to one where gene expression is highly correlated and tightly regulated by supercoiling. In the latter regime, the model displays transcriptional bursts, waves of supercoiling, and up regulation of divergent or bidirectional genes. It also predicts that topological enzymes which relax twist and writhe should provide a pathway to down regulate transcription.

DOI: 10.1103/PhysRevLett.117.018101

The dynamics of transcription is a topic of paramount importance in cell biology and biophysics. It underpins the expression and regulation of genes, which is crucial to the development and function of all living organisms [1]. To initiate transcription of a gene, cells rely on the binding of proteins, such as polymerases and transcription factors, to the promoter—a DNA region shortly upstream of the gene [1]. As there are a finite number of copies of such proteins present within a cell, this process is inherently stochastic [2–5].

In this work, we introduce a stochastic model of gene expression, which is fundamentally different from previous studies as it couples transcription to the dynamics of DNA twist and supercoiling. Supercoiling is a topological property of DNA, arising from its chiral nature [6–8]. For B-DNA in its relaxed state, the two strands of the molecule wind around each other once approximately every 10 base pairs (bp), forming a right-handed double helix [6]. Twisting DNA away from this relaxed state, so as to over- or underwind the double helix, introduces positive or negative supercoiling respectively; if large enough, this torsional strain can lead to writhing or to DNA melting. Supercoiling thus refers to the difference in the linking number of the two DNA strands Lk, with respect to that in the relaxed state Lk_0; the global Lk is a topological invariant if the DNA is a loop or its ends are constrained [6], whereas it can vary for an open polymer whose ends can rotate.

There are several observations which strongly suggest that DNA supercoiling is intimately related to transcription and that it can regulate gene expression. First, the "twin supercoiled domain" model [9–13] is based on the long-standing theoretical observation that if rotation of the RNA polymerase and its associated transcription machinery is hindered, as is likely in the crowded intracellular environment, then gene transcription leads to the creation of positive supercoiling ahead of the tracking polymerase and negative supercoiling in its wake. For every 10 bp or so which are transcribed, the linking number changes by $\Delta\mathrm{Lk} \approx +1$ ahead of the polymerase and by $\Delta\mathrm{Lk} \approx -1$ behind it. Recent experiments have quantified supercoiling by measuring the DNA binding affinity of psoralen, a chemical which intercalates preferentially where the double helix is underwound [14,15]. These studies have shown that human chromosomes are organized into a set of supercoiling domains, whose structure is dramatically altered by inhibiting transcription.

Our model is based on these observations and incorporates the dynamics of supercoiling into a stochastic description of gene regulation. It exhibits a switch between two regimes: one where gene expression is random, and one where it is tightly regulated by supercoiling. Within our framework, this switch is triggered, e.g., by increasing the amount of supercoiling injected during each transcription event. The dynamics in the supercoiling-regulated regime help explain a number of experimental observations, such as the existence of transcriptional bursts and the abundance of bidirectional genes in the genomes of many organisms.

We model the DNA as a 1D lattice with spacing $\Delta x \equiv l \sim 15$ bp, the size of an RNA polymerase [1,13]. The DNA contains n genes, each of size λ, whose promoters are located at positions y_j $(j = 1, \ldots, n)$ on the DNA. Gene transcription is modeled as a stochastic process [16]: at each time step, for each of N polymerases, a gene is selected at random and is activated by the polymerase binding at the promoter with rate k_{on}. Once a gene is

activated, the polymerase travels along the gene body at a velocity v, so the position along the DNA of the ith polymerase which is transcribing, say, the jth gene is $x_i = y_j + vt_i$, where t_i is the time since the polymerase was activated. The total time to transcribe any gene is then $\tau = \lambda/v$, after which the polymerase unbinds from the DNA and is free to transcribe another gene. (A simpler model where a static polymerase generates supercoiling without traveling is discussed in Ref. [16]).

We couple transcription to the local supercoiling density $\sigma(x,t) = (\mathrm{Lk} - \mathrm{Lk}_0)/\mathrm{Lk}_0$, where Lk is the local linking number at position x. We propose the following diffusive dynamics for $\sigma(x,t)$:

$$\frac{\partial \sigma(x,t)}{\partial t} = \frac{\partial}{\partial x}\left[D\frac{\partial \sigma(x,t)}{\partial x} - J_{\mathrm{tr}}(x,t)\right],$$
$$J_{\mathrm{tr}}(x,t) = \sum_{i=1}^{N} J_i(t_i)\delta(x - x_i(t_i))\xi_i(t), \qquad (1)$$

where D is the effective diffusivity of supercoiling along DNA and $J_{\mathrm{tr}}(x,t)$ is the local flux of supercoiling (Fig. 1) arising due to the transcription of any of the genes [16]. We use periodic boundary conditions so that the overall level of supercoiling is conserved. (This corresponds to modeling a DNA loop.) In Eq. (1), $\xi_i(t)$ is set equal to 0 when the ith polymerase is inactive and to 1 when it is transcribing any of the n genes. The modulus of the flux is $J_i = J_0(1 + vt_i/l)$; it increases during transcription to model the fact that the positive supercoiling is racked up in front of the traveling polymerase. The sign of J_i depends on the direction of gene transcription. Because of the observation that negative supercoiling can facilitate binding of RNA polymerases and transcription factors [18,19], we further assume that k_{on} depends on the local value of σ at the promoter σ_p. For simplicity, we choose a linear coupling $k_{\mathrm{on}} = k_0 \max\{1 - \alpha\sigma_p, 0\}$, where k_0 is the polymerase binding rate for $J_0 = 0$ and α quantifies the sensitivity to σ_p. The linear dependence of k_{on} on σ_p is enough to give rise to highly nonlinear dynamics. This is because the supercoiling created when a gene is switched on favors its own transcription, as well as that of upstream genes, whereas it hinders expression of the genes downstream. These chains of positive and negative feedbacks are at the basis of the nonlinear transcription dynamics described below.

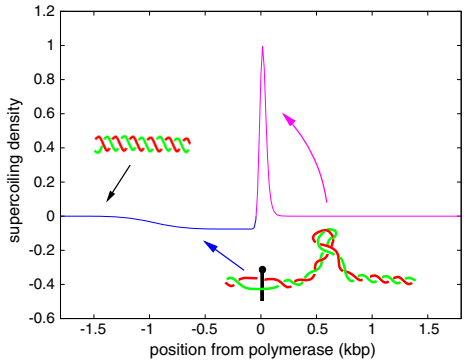

FIG. 1. Schematic of the supercoiling density close to a transcribed gene, in the frame of reference of the traveling polymerase (see also Supplemental Movie 1 [16]). The RNA polymerase creates positive supercoiling (here speculatively depicted as right-handed writhe) ahead of the gene, while it generates negative supercoiling (here speculatively depicted as DNA unwinding) behind. The supercoiling profile is obtained by solving Eq. (1) with $\bar{J}/D = 1.7$, and other parameters as in Ref. [20], except for $n = N = 1$. The gene is switched on at time $t = 0$, and the plot is for $t = \tau$; the transcription flux is here a regularized delta function [16].

There are three main dimensionless parameters in the model. The first is the product of the transcription rate and the transcription time $\Phi = (k_{\mathrm{on}}N/n)\tau$, which measures how often the gene is on. The second measures how fast supercoiling diffuses away between transcription events $\Theta = (k_{\mathrm{on}}N/n)\lambda^2/D$. The third one is $\bar{J}/D$ and identifies the supercoiling generated near the promoter while the gene is active [$\bar{J} = J_0[1 + \lambda/(2l)]$ is the average supercoiling flux during transcription]. In Ref. [16], we show that the average supercoiling at the promoter can be estimated in terms of these parameters as $\sigma_p \sim -[(\Phi/(\Phi+1)]\bar{J}/(2D)$. (This estimate should be seen as a change from the baseline value of supercoiling ~ -0.05 in bacteria.) Dimensional analysis further suggests that $\bar{J} \sim v\lambda$. The main question is then whether the average level of supercoiling generated triggers the positive feedbacks highlighted above; experiments suggest $\sigma_p \sim -0.01$ is enough to affect polymerase binding [18,21]. What is the situation inside cells? The diffusion constant of supercoils within naked DNA is $D \sim 0.1$ kbp^2/s or less [22]. Within bacteria, transcription rates are $\sim$10 RNA molecules per minute or above [23]; considering a typical gene size of 1 kbp and an elongation rate of 100 bp/s, we get $\sigma_p \sim -0.3$. This suggests that supercoiling can be relevant for transcription in prokaryotes. In eukaryotes, transcription initiation is slower due to the need for several transcription factors to colocalize at a promoter; for example, rates in yeast and humans are about 10 and 1 transcripts per hour, respectively [24,25]. Given that for eukaryotes $v \sim 25$ bp/s, while λ lies between 1.6 kbp (yeast) and 10 kbp (humans), we obtain $\sigma_p \sim -0.03$ (yeast) and $\sigma_p \sim -0.13$ (humans). Because D has not been measured for chromatin, these order-of-magnitude estimates should be viewed with caution, yet they suggest supercoiling may affect polymerase initiation in eukaryotes as well [26].

Here and in what follows, we will choose parameters which are relevant to bacterial DNA [20] and study how the system behaves upon varying $\bar{J}$. As discussed in Ref. [16], the results we report here are representative of the system's behavior in general. We start by considering a case in which all genes are read in the forward direction, from left to right.

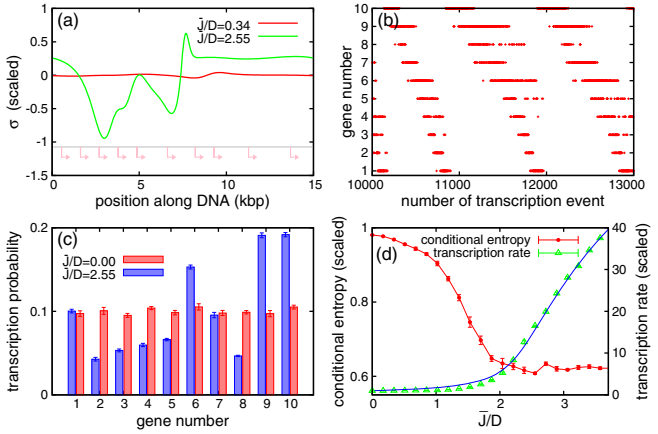

FIG. 2. (a) Snapshots of $\sigma(x,t)$ in the relaxed regime (red, $\bar{J}/D = 0.34$) and in the supercoiling-regulated regime (green, $\bar{J}/D = 2.55$) for a 15 kbp DNA. (b) Portion of the time series of the sequence of transcribed genes for $\bar{J}/D = 2.55$. A transcription wave can be seen as genes are transcribed preferentially in the order 10, 9,…, 1, 10,… (see Supplemental Movie 4 [16], genes numbered from left to right). (c) Histograms showing gene transcription probabilities for $\bar{J}/D = 0$ and $\bar{J}/D = 2.55$ (average over seven runs). The most transcribed genes for $\bar{J}/D = 2.55$ are (in order) 10, 9, and 6. (d) Plot of the conditional entropy and the overall transcription rate (scaled by $k_0 N$; the blue line is the transcription rate predicted by the semi-analytic theory in Ref. [16]). Gene positions for (a)–(d) are indicated in (a). Results for (c) and (d) were averaged over seven runs.

The genes are positioned randomly, but with the constraint that the distance between neighboring genes is > 1 kbp. For small $\bar{J}/D$, the typical values of σ generated by transcription are modest [Fig. 2(a), red curve, and Supplemental Movie 2 [16]]; we call this the *relaxed* regime. The sequence of transcription events in this regime is well described by a Poisson process: any gene is read on average the same number of times, and the total number of transcription events is $\sim k_0 NT$ where T is the total simulation time. As $\bar{J}/D$ increases, the flux of supercoiling injected by a polymerase becomes large enough to change the transcriptional dynamics significantly. Now, the scale of variation of σ is much larger [Fig. 2(a), green curve, and Supplemental Movie 3 [16]]; we call this the *supercoiling-regulated* regime. The value of σ_p is now large enough to affect k_{on} significantly, and this triggers bursts in transcription of the same gene, and waves of transcription [Fig. 2(b); see also Supplemental Movies 3 and 4 [16]]. Genes are also no longer equally expressed: those with a large gap between them and the nearest upstream neighbor are up regulated because they are less affected by the buildup of positive supercoiling during transcription [Fig. 2(c)].

As expected from the discussion above, the switch between the relaxed and supercoiling-regulated regimes is associated with a rise in overall transcription rate [Fig. 2(d)]. It is also linked to a change in the nature of the time series describing the sequence of transcribed genes which becomes non-Poissonian and displays temporal correlations (due to bursting and waves of transcription). A useful way to quantify such a change is via the "conditional entropy" and "mutual information" [27–29] (definitions are given in Ref. [16]). The conditional entropy is maximal and equal to $\log(n)$, if the transcription dynamics is a Poisson process (as is the case for $\bar{J} \to 0$), whereas it equals 0 in the limit of a maximally correlated process (e.g., when a single gene is repeatedly transcribed). Figure 2(d) shows that the conditional entropy decreases with $\bar{J}/D$ in a sigmoidal way. The mutual information is a measure of the deviation of the observed joint probability distribution for successive transcription events, from that of a random process: for the case of Fig. 2, it is close to 0 for $\bar{J} = 0$, and is higher in the supercoiling-regulated regime (Fig. S1 [16]). A semianalytic theory of transcription bursts in a single gene model reproduces well the overall transcription rate of Fig. 2. A simplified mean field theory also shows that the switch is a crossover rather than a nonequilibrium phase transition, leads to the estimate for σ_p discussed above, and further suggests that supercoiling can affect transcription if $\bar{J} k_0 \tau \alpha/(2D) \sim 1$ or larger [16].

In reality, genes can be encoded either in the forward or reverse strand of the DNA double helix [30–32]; hence, the supercoiling flux can be directed either way along the genome. To see how this affects our model, we study the case in which some of the genes are transcribed left to right, and others right to left (see Fig. 3). Figures 3(a) and 3(b) show that in the supercoiling-regulated regime (large $\bar{J}/D$),

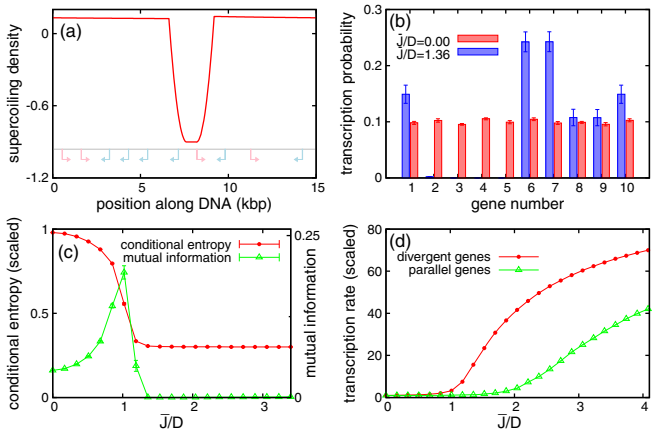

FIG. 3. (a) Plot of the average value of σ in the supercoiling-regulated phase for a 15 kbp DNA with forward and backward genes. (b) Histograms of transcription probabilities for the same system with $\bar{J}/D = 0$ (red bars) and $\bar{J}/D = 1.36$ (blue bars). The divergent pair 6, 7 is up regulated because of the trail of parallel genes in front of 6. (c) Conditional entropy [scaled by $\log(n)$] and mutual information [averaged over 200 runs for the same gene positions as in (a) and (b)]. (d) Overall transcription rate from all genes (scaled by $k_0 N$, averaged over seven runs), for the single orientation arrangement of genes in Fig. 2, and for a divergent arrangement where the genes occupy the same region of DNA, but the first five are transcribed right to left.

some gene pairs are up regulated together [see Fig. 3(b)]. These are the divergent pairs (adjacent genes which point away from each other); when either is switched on, negative supercoiling is generated between the genes, which triggers further transcription in both. Within a given run, we normally observe transcription of a single divergent gene pair, where the selection mechanism is fluctuation dependent (Fig. S4 [16]); within several runs, there is a ranking list of divergent pairs which depends quite subtly on gene position [Fig. 3(b)]. Transcription of convergent genes instead leads to a buildup of positive supercoiling, so it is always strongly down regulated.

In comparison to the case of genes which are all in the same direction, random orientations lead to a more marked peak in the mutual information and to a sharper drop in the conditional entropy [Fig. 3(c)]. Divergent transcription also yields a larger overall transcription rate [again with respect to the case of parallel genes; see Fig. 3(d)]. It is tempting to propose that this mechanism that markedly favors the transcription and coexpression of divergent pairs is among the reasons for the high abundance of such promoter pairs in the genomes of several organisms, including humans [30,31]. Furthermore, consistent with our model, divergent gene neighbors in yeast are often coexpressed, have low transcriptional noise, and, importantly, are often associated with essential genes which tend to be highly expressed [33,34].

Within a cell, the level of supercoiling is not conserved globally due to the presence of topological enzymes such as type I and type II topoisomerase, which can relax local supercoiling at a rate of $\sim 0.1-1$ supercoil/s [35]. It is therefore of interest to include these enzymes in our model; the simplest way is through a nonconserved reaction term in Eq. (1), as follows:

$$\frac{\partial \sigma}{\partial t} = \frac{\partial}{\partial x}\left[D\frac{\partial \sigma}{\partial x} - J_{\mathrm{tr}}(x,t)\right] - k_{\mathrm{topo}}\sigma, \qquad (2)$$

where k_{topo} is a relaxation rate; this is associated with a length scale $l_{\mathrm{topo}} \sim \sqrt{D/k_{\mathrm{topo}}}$, over which supercoiling-mediated regulatory interactions are screened. Figure 4 shows the effect of such enzymes in the setup corresponding to Fig. 3. Divergent gene pairs are strongly up regulated if $k_{\mathrm{topo}} = 0$, but for $k_{\mathrm{topo}} > 0$, there is a dramatic down regulation of transcription [Figs. 4(a) and 4(b)]. This is accompanied by a rise in the conditional entropy [Fig. 4(b)]; topoisomerases therefore rapidly lead to a loss of correlations in the transcription process.

In conclusion, we presented a dynamical model for supercoiling-dependent transcription, where a continuum description for the evolution of supercoiling is coupled to a stochastic transcriptional dynamics. Our model shows a crossover between two distinct regimes. When the supercoiling flux created as a polymerase transcribes a gene is small, transcription is a random process. When this flux is large, the dynamics become highly correlated. These correlations can be measured using the conditional entropy and mutual information of the transcriptional time series. For parallel genes, supercoiling drives transcriptional waves and bursts reminiscent of those observed in high-resolution dynamical experiments in both pro- and eukaryotes [36–38]. It also regulates gene expression, promoting the transcription of genes which have a larger gap separating them from their upstream neighbors. When considering genes with random orientations, transcription localizes at divergent gene pairs, which are highly up regulated. This is consistent with the observation that in yeast divergent gene pairs are often highly expressed essential genes [33] and may explain the statistically surprising abundance of bidirectional promoters within mammals [30,31]. Finally, our theory predicts that including the action of topoisomerases, which locally relax supercoiling, down regulates transcription: this agrees with the observation that inhibiting topo I can boost eukaryotic transcription rates *in vivo* [11,30]. Note that we disregard other important topological enzymes, such as the bacterial gyrase, whose role is to introduce, rather than to relax, negative supercoiling: such enzymes are known to promote transcriptional bursting [38].

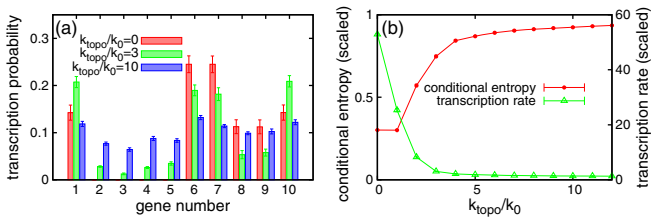

FIG. 4. (a) Histograms of transcription probabilities of the 10 bidirectional genes in Fig. 3 (with $\bar{J}/D = 2.55$), with different values of k_{topo}/k_0. (b) Conditional entropy and transcription rate for the same system as (a), as a function of k_{topo}/k_0. Results were averaged over six runs.

We foresee at least three major ways in which this work can be further pursued. First, we hope that our study will stimulate quantitative experiments measuring gene expression *in vitro*, where gene positions and directions can be controlled, e.g., via DNA editing. Second, the model could be refined by comparison with high-resolution psoralen data on supercoiling domains in both pro- and eukaryotes. Finally, it would be of interest to couple the dynamics of supercoiling to that of nucleosomes, which can at the same time create a barrier for supercoil diffusion and localize twist and writhe.

C. A. B. and D. M. acknowledge ERC for funding (ERC Consolidator Grant No. 648050, THREEDCELLPHYSICS).

[1] B. Alberts *et al.*, *Molecular Biology of the Cell* (Garland Science, New York, 2002).

[2] J. M. Raser and E. O'Shea, Science **309**, 2010 (2005).
[3] J. Paulsson, Phys. Life Rev. **2**, 157 (2005).
[4] A. Raj and A. Van Oudenaarden, Cell **135**, 216 (2008).
[5] P. Visco, R. J. Allen, and M. R. Evans, Phys. Rev. Lett. **101**, 118104 (2008).
[6] A. D. Bates and A. Maxwell, *DNA Topology* (Oxford University Press, New York, 2005).
[7] N. Gilbert and J. Allan, Curr. Opin. Genet. Dev. **25**, 15 (2014).
[8] J. F. Marko and E. D. Siggia, Phys. Rev. E **52**, 2912 (1995).
[9] L. F. Liu and J. C. Wang, Proc. Natl. Acad. Sci. U.S.A. **84**, 7024 (1987).
[10] S. P. Mielke, W. H. Fink, V. V. Krishnan, N. Groenbech-Jensen, and C. J. Benham, J. Chem. Phys. **121**, 8104 (2004).
[11] M. Dunaway and E. A. Ostrander, Nature (London) **361**, 746 (1993).
[12] F. Benedetti, J. Dorier, Y. Burnier, and A. Stasiak, Nucleic Acids Res. **42**, 2848 (2014).
[13] P. R. Cook, *Principles of Nuclear Structure and Function* (Wiley-Liss, New York, 2001).
[14] C. Naughton, N. Avlonitis, S. Corless, J. G. Prendergast, I. K. Mati, P. P. Eijk, S. L. Cockroft, M. Bradley, B. Ylstra, and N. Gilbert, Nat. Struct. Mol. Biol. **20**, 387 (2013).
[15] F. Kouzine, A. Gupta, L. Baranello, D. Wojtowicz, K. Ben-Aissa, J. Liu, T. M. Przytycka, and D. Levens, Nat. Struct. Mol. Biol. **20**, 396 (2013).
[16] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.117.018101 for additional details on the model, its derivation, and its limitations along with further numerical and analytical results for both the traveling polymerase model and other variants. Reference [17] is included in the Supplemental Material.
[17] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, England, 1995).
[18] H. Weintraub, P. F. Cheng, and K. Conrad, Cell **46**, 115 (1986).
[19] G. W. Hatfield and C. J. Benham, Annu. Rev. Genet. **36**, 175 (2002).
[20] Unless otherwise stated, parameters used are $N = n = 10$, $\alpha = 100$, $\tau = 10$ s, $v = 100$ bp/s, $k_0 = 0.001$ s^{-1}, and $D = 2.25 \times 10^3$ bp^2/s, while $\bar{J}$ is varied. Equation (1) was solved by finite difference on a lattice of $L = 1000$ sites, with $\Delta x = 15$ bp and $\Delta t = 0.01$ s. (This mapping is relevant to the bacterial case.) Technical details regarding the solution of Eq. (1) are given in Ref. [16].
[21] K. Y. Rhee, M. Opel, E. Ito, S. Hung, and G. W. Hatfield, Proc. Natl. Acad. Sci. U.S.A. **96**, 14294 (1999).
[22] M. T. van Loenhout, M. V. de Grunt, and C. Dekker, Science **338**, 94 (2012).
[23] S. T. Liang, M. Bipatnath, Y. C. Xu, S. L. Chen, P. Dennis, M. Ehrenberg, and H. Bremer, J. Mol. Biol. **292**, 19 (1999).
[24] V. Pelechano, S. Chavez, and J. E. Perez-Ortin, PLoS One **5**, e15442 (2010).
[25] D. A. Jackson, A. Pombo, and F. Iborra, FASEB J. **14**, 242 (2000).
[26] These are average values, but typically σ remains close to this provided that Θ is not too small [16].
[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[28] G. B. Brandani, M. Schor, C. E. MacPhee, H. Grubmuller, U. Zachariae, and D. Marenduzzo, PLoS One **8**, e65617 (2013).
[29] L. Barnett, J. T. Lizier, M. Harre, A. K. Seth, and T. Bossomaier, Phys. Rev. Lett. **111**, 177203 (2013).
[30] C. Naughton, S. Corless, and N. Gilbert, Transcription **4**, 162 (2013).
[31] J. M. Lin, P. J. Collins, N. D. Trinklein, Y. Fu, H. Xi, R. M. Myers, and Z. Weng, Genome Res. **17**, 818 (2007).
[32] B. S. Scruggs, D. A. Gilchrist, S. Nechaev, G. W. Muse, A. Burkholder, D. C. Fargo, and K. Adelman, Mol. Cell **58**, 1101 (2015).
[33] G. Z. Wang, M. J. Lercher, and L. D. Hurst, Genome Biol. Evol. **3**, 320 (2011).
[34] S. Meyer and G. Beslon, PLoS Comput. Biol. **10**, e1003785 (2014).
[35] K. Terekhova, K. H. Gunn, J. F. Marko, and A. Mondrago, Nucleic Acids Res. **40**, 10432 (2012).
[36] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, Cell **123**, 1025 (2005).
[37] J. R. Chubb, T. Trcek, S. M. Shenoy, and R. H. Singer, Curr. Biol. **16**, 1018 (2006).
[38] S. Chong, C. Chen, H. Ge, and X. S. Xie, Cell **158**, 314 (2014).