Theory on the dynamic memory in the transcription-factor-mediated transcription activation

R. Murugan*

Department of Biotechnology, Indian Institute of Technology, Madras, Chennai 600036, India

(Received 14 December 2010; published 29 April 2011)

We develop a theory to explain the origin of the static and dynamical memory effects in transcription-factor-mediated transcription activation. Our results suggest that the following inequality conditions should be satisfied to observe such memory effects: (a) $\tau_L \gg \max(\tau_R, \tau_E)$, (b) $\tau_{LT} \gg \tau_T$, and (c) $\tau_I \geqslant (\tau_{EL} + \tau_{TR})$ where τ_L is the average time required for the looping-mediated spatial interactions of enhancer—transcription-factor complex with the corresponding promoter—RNA-polymerase or eukaryotic RNA polymerase type II (PolII in eukaryotes) complex that is located L base pairs away from the *cis*-acting element, (τ_R, τ_E) are respectively the search times required for the site-specific binding of the RNA polymerase and the transcription factor with the respective promoter and the *cis*-regulatory module, τ_{LT} is the time associated with the relaxation of the looped-out segment of DNA that connects the *cis*-acting site and promoter, τ_T is the time required to generate a complete transcript, τ_I is the transcription initiation time, τ_{EL} is the elongation time, and τ_{TR} is the termination time. We have theoretically derived the expressions for the various searching, looping, and loop-relaxation time components. Using the experimentally determined values of various time components we further show that the dynamical memory effects cannot be experimentally observed whenever the segment of DNA that connects the *cis*-regulatory element with the promoter is not loaded with bulky histone bodies. Our analysis suggests that the presence of histone-mediated compaction of the connecting segment of DNA can result in higher values of looping and loop-relaxation times, which is the origin of the static memory in the transcription activation that is mediated by the memory gene loops in eukaryotes.

DOI: 10.1103/PhysRevE.83.041926 PACS number(s): 87.10.−e, 87.15.kj

I. INTRODUCTION

Transcription factors (TFs) play a critical role in the initiation of transcription in eukaryotes [1–6]. The TF molecule associated with a gene of interest first binds with the corresponding *cis*-acting regulatory elements (CRMs) of that gene which are located either upstream or downstream of the respective promoter sequences to form an enhancer—transcription-factor complex (ETF). This ETF then distally acts on and stabilizes the already formed promoter-RNAP complex in prokaryotes, promoter-PolII complex in eukaryotes (PR). This in turn enhances the initiation of the transcription event [4–6]. The mode of this distal action is not understood clearly. It could be either via looping out [5] of the intervening DNA segment [IDS; see Fig. 1(a) that is present between the promoter and the CRM binding site or via tracking of the ETF complex [6,7] toward the PR along the connecting DNA segment. This distal communication can also be mediated via the slithering dynamics induced by the supercoiling of the template DNA [7]. Detailed experimental studies revealed the following properties of TF-enhancer-mediated transcription initiation. (a) The transcription-factor-mediated enhancer action generally increases the probability of transcription of the associated genes rather than the rate of transcription [8]. (b) Two distally located promoters are activated by the associated enhancer with equal probabilities [9]. (c) Increase in the degree of supercoiling of the template DNA can increase the efficiency of TF-mediated enhancer action [10] whenever the distance between the *cis*-regulatory binding site and the promoter sequences is more than ∼110 base pairs (bps). The compaction of DNA in terms of higher-order chromatin structures can give rise to similar effects of DNA supercoiling [11] on the transcription-factor-mediated transcription initiation in eukaryotes. (d) The transcription-factor-mediated transcription initiation carries [6] the memory of the first-time ETF formation and ETF-PR spatial contact in the subsequent rounds of transcription initiation though *in vitro* transcription experiments suggested [12] the absence of such a functional memory. Here the term "memory" indicates the faster activation of transcription in the second and subsequent rounds than in the first round of activation. This means that whenever there are memory effects then the transcription initiation time in the second and subsequent rounds will be much lower than that of the first round of transcription.

At least two different models have been proposed to explain the mechanism of TF-mediated distal action of enhancers over the associated promoter sequences of the genes of interest, viz. tracking and looping models. According to the tracking model [6,7,13], the ETF complex performs a one-dimensional (1D) diffusion-mediated tracking or sliding along the IDS toward the PR complex and stabilizes it. Here the directionality of the dynamics of ETF toward PR might be originating from the asymmetry in the binding of TFs at the respective CRMs [13]. Whereas in the looping model [13] the IDS present between the ETF and PR is looped out and then the ETF acts distally on the PR complex through a three-dimensional (3D) mediated searching. It seems that both these modes of searching simultaneously operate in the transcription activation of the globin genes within the human β-globin gene cluster, indicating that tracking and looping are not mutually exclusive mechanisms of distal action [14,15]. Recently a random jump model for enhancer action was proposed [16], which suggested that the ETF searches for PR through 1D random jump dynamics. All the observations

*rmurugan@gmail.com

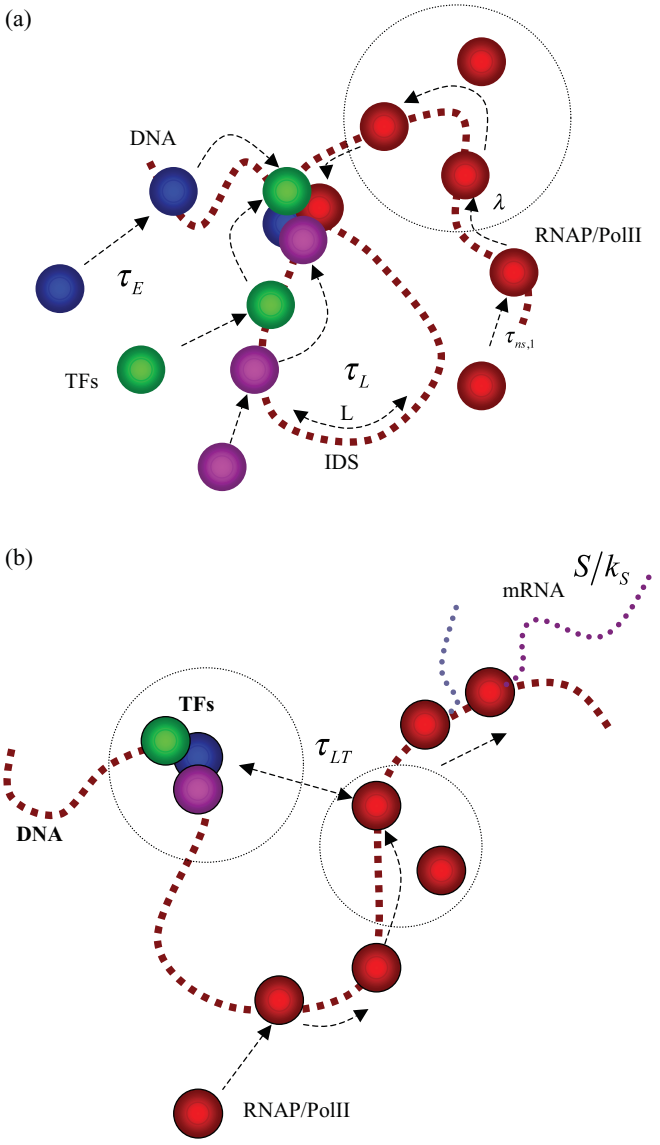

FIG. 1. (Color online) (a) Various time components in the transcription activation. The total time that is required for transcription activation can be denoted by the sum $\tau_I = \max(\tau_E, \tau_R) + \tau_L$ where (τ_E, τ_R) are respectively the time required for the binding of transcription factors with the *cis*-acting modules and RNAP/PolII with the promoter and τ_L is the time required for the spatial contact between the enhancer—transcription-factor (ETF) complex with promoter-RNAP/PolII (PR) complex via formation of the looped-out structure of the intervening DNA segment (IDS) whose length is L bps. Memory effects on the transcription can be observed only when the looping dynamics is the rate-limiting one. (b) Loop-relaxation time τ_{LT} is the time that is required to bring the IDS along with the bulky ETF and PR complexes to the original state upon escape of the RNAP/PolII into the elongation step. Memory effects on transcription can be observed only when the loop-relaxation time is higher than or comparable with the total transcription time, which is the sum of initiation, elongation ($\tau_{EL} = S/k_S$), and termination times, where S is the length of the complete transcript and k_S is the elongation speed of RNAP/PolII.

(a)—(d) pertaining to the properties of TF-mediated enhancer action could be explained successfully using the random jump model with certain critical sizes of jumps and the memory of first-time ETF formation.

A slithering mechanism was proposed [10–12,17] to explain the positive effects of supercoiling of the template DNA on the distal action of enhancers, according to which the sliding of intertwined, dynamic DNA helices within "branches" formed on the supercoiled DNA ("slithering") greatly increases the probability of ETF-PR contact. This slithering model further suggests that there is no need for a functional memory in the enhancer-mediated transcription activation. Rather, the higher-order supercoiled conformation or histone-mediated higher-order chromatin structure of the DNA template is sufficient to bring the ETF and PR closer together in 3D space, and the transcription activation level in the first round as well as in the subsequent rounds should be identical. This proposition was evidenced in the transcription activation of *glnAP2* promoter by the NtrC-dependent enhancer in the *Escherichia coli* system. The results of this *in vitro* transcription study suggested [12] that the ETF-PR interactome must be formed *de novo* during every round of transcription and no protein remained bound with the promoter sequence after the prokaryotic RNA polymerase (RNAP) escaped into the elongation step. However, one should note that the memory effects could dominate only when the ETF-PR interaction step is a rate-limiting one, which in turn is strongly dependent on the contour length of IDS present between the ETF and PR complexes. The contour length of IDS used in this *in vitro* study [12] was ~110 bps. This means that the ETF-PR interaction step with these current experimental settings might not be a rate-limiting one apart from the higher free energies associated with the DNA loop formed with this contour length of IDS, which in turn can rapidly destabilize the IDS loop upon escape of RNAP into the elongation step; this could be the reason for not observing the memory effects in this *in vitro* transcription study. The type of transcriptional memory that we are addressing here so far is a transient and dynamical one which is different from the "static" transcriptional memory [18] that can persist for many generations of the cell and plays critical roles in tissue development, differentiation, and maintenance of cell types. This means that the cells are retaining the information about the set of activated genes across many mitotic cell divisions. It seems that such persistent memory mechanisms are associated with the chromosomal "memory gene loops" formed [19] between the promoter and the 3′ ends of the responsive genes. These memory gene loops play a critical role in faster recruitment of PolII during reinduction of a gene after a transcription repression.

All these experimental observations suggested that the memory effects in the TF-enhancer-mediated transcription activation can be either static or dynamical in nature and depends on the contour length of the IDS present between the ETF and PR complexes, the relative time scale that is required for the recruitment of RNAP/PolII at the respective promoter in the second and subsequent rounds of transcription, and the mean lifetimes associated with the formation and relaxation of the IDS loop that leads to the spatial interaction between ETF and PR. In this paper we investigate the origin as well as

the conditions which are required to experimentally observe such a static or dynamical memory effects in detail.

II. THEORY

We closely follow the ideas of the recruitment model of Ptashne and Gann [20] by assuming that ETF recruits the RNAP/PolII at the promoter of the gene of interest upon getting closer in space to the promoter by forming a looped-out structure of the connecting segment of DNA. With this setting the total time that is required for the activation of transcription (τ_I) can be given by the sum of the time components as $\tau_I = \tau_E + \tau_L + \tau_R$ where τ_E is the time required for the formation of the ETF complex, τ_L is the time required for the formation of the looped structure of the connecting DNA segment and bring the ETF closer to the promoter, which depends on the length of the connecting segment, and τ_R is the time required to recruit the RNAP/PolII at the promoter sequence by the closely located ETF. The total time required to generate a complete transcript will be $\tau_T = \tau_I + \tau_{EL} + \tau_{TR}$ where τ_{EL} is the time required for elongation, which depends on the speed of RNAP/PolII, and τ_{TE} is the time required for the termination step. Whenever there is a static or dynamical memory, it is expected [17] that the time required to generate a complete transcript in the second and subsequent rounds of transcription will be much less than the time required for the first round of transcription. When the bindings of TF and RNAP/PolII are independent and parallel, then we find that $\tau_I = \max(\tau_E, \tau_R) + \tau_L$. When the mode of the distal interaction is through the slithering effect that is induced by the supercoiling of the template DNA, then τ_L in the expression of τ_I will be the time required for such a slithering-mediated distal communication of ETF with the respective promoter. When the distal communication is through the tracking mode then τ_L represents the time that is required by the ETF to reach the promoter via a one-dimensional diffusion dynamics along the contour of the IDS. When the mean lifetime of the IDS loop after the escape of the RNAP/PolII into the transcription elongation step is τ_{LT} [Fig. 1(b), then irrespective of the looping/tracking/slithering modes of searching of ETF for the promoter sequence, the dynamical memory effects will be prominently observed only when the following set of inequalities are true:

$$I_q = \{\tau_L \gg \max(\tau_R, \tau_E); \tau_{LT} \gg \tau_T; \tau_I \geqslant (\tau_{EL} + \tau_{TR})\}. \tag{1}$$

When $(\tau_E + \tau_L) < \tau_R$ then the recruitment of RNAP/PolII by ETF at the promoter will be the rate-limiting step and the level of activation in the first and subsequent rounds of transcription will be almost independent of the memory of the first-time spatial contact between the ETF and PR complexes. For shorter contour lengths of IDS we have $\tau_E \gg \tau_L$, $\tau_{LT} \ll \tau_R$, and $\tau_E \sim \tau_R$ when the transcription activation is mediated by a single transcription factor as in the case of most of the prokaryotic systems such as *E. coli*. As a result the memory effects are not prominently observed and are not required for shorter lengths of IDS in prokaryotes. Whenever the length of the transcript is very high, then $\tau_I \ll \tau_{EL}$ and $\tau_{LT} \ll \tau_T$ and the second and third inequalities in Eq. (1) will fail. This means that the transcription elongation step will be the rate-limiting one and the memory effects in transcription activation will not be prominently observed under such conditions. Here one should also note that $\tau_E \gg \tau_R$ when many combinatorial TFs are involved [21] in the transcription activation as in the case of most eukaryotic systems, and the time τ_{CR} that is required to remodel [22,23] the histone-mediated compact chromosomal structure of the template DNA also needs to be added to the sum τ_I. This means that the memory effects can be observed prominently and also are strictly required even for shorter contour lengths of IDS in the case of eukaryotes for efficient activation of transcription.

A. Search times for site-specific binding

We first compute the search times taken by the n combinatorial transcription factors to find their sequentially located *cis*-regulatory modules on the genomic DNA. Site-specific binding of protein molecules with DNA has been extensively studied both theoretically and experimentally. Berg *et al.* [24,25] had suggested that various facilitating processes such as sliding, hopping, and intersegmental transfers can enhance the rate of site-specific interactions of the protein molecule with the DNA chain over the 3D diffusion-controlled rate limit. The protein molecule that is diffusing along the DNA polymer can randomly switch between different modes of these facilitating dynamics depending on the prevailing local environment. Here the sliding mode of dynamics indicates the diffusion of the protein molecule along the DNA chain with unit base pair step size whereas the protein molecule can leap over a few bps at a time in the hopping mode. These sliding and hopping modes dominate whenever the DNA molecule is somewhat stretched and loosely packaged. On a highly condensed or supercoiled DNA chain, the diffusing protein molecule can undergo intersegmental transfers via ring closure events, which can occur whenever two distal segments of the same DNA chain come closer upon condensation. The protein molecule can leap over a few hundreds to thousands of bps during these intersegmental transfer events. All these facilitating modes reduce the overall search time that is taken by the protein molecule to locate its specific site on DNA mainly by fine-tuning the ratio of the search times spent on 1D and 3D routes. Slutsky and Mirny [26] and Murugan [27] have shown that the minimum of this overall search time can be achieved when the protein molecule spends equal amounts of time in both the 1D and 3D routes. Detailed theoretical studies of Coppey *et al.* [28] and Lomholt *et al.* [29] as well as the single molecule experimental studies of Sokolov *et al.* [30], Broek *et al.* [31], Bonnet *et al.* [32], and Wang *et al.* [33] substantiated the ideas of Berg *et al.* [24,25] and further suggested that the spatial organization and packaging [27] of the DNA molecule can significantly enhance the rate of site-specific interactions of the protein molecule with DNA. It seems that the thermally driven conformational fluctuations in the DNA binding domains of DNA binding proteins can significantly enhance the search efficiency [34–36].

With this background, consider a genomic DNA of length N bps that is embedded in a cellular volume of V_C m^3 and there is a single transcription factor ($n = 1$) which is in the process of searching for its *cis*-regulatory site on the genomic DNA via a combination of one- and three-dimensional routes.

We assume that the transcription factor nonspecifically binds with the genomic DNA and scans an average length of λ bps and then dissociates, to reassociate back at the same or different locations of the genomic DNA. The transcription factors perform on an average $N\lambda^{-1}$ numbers of association-scan-dissociation cycles before locating the *cis*-acting binding sites. With these settings from the theory of site-specific interactions of combinatorial transcription factors, we can write the expression for the total search time (τ_s) that is taken by the n such TFs to reach their sequentially located *cis*-acting sites on DNA as follows [21]:

$$\begin{aligned} &\tau_s = N\lambda^{-1}(\tau_{\lambda,n} + \tau_{ns,n}); \quad \tau_{\lambda,n} \sim n^{\alpha}\lambda^2/6D_o; \\ &\alpha \sim 2/5; \quad \tau_{ns,1} \leqslant \tau_{ns,n} \leqslant n\tau_{ns,1}. \end{aligned} \tag{2}$$

Here $\tau_{ns,n}$ is the average time taken by n TFs to nonspecifically associate with DNA through 3D diffusion-controlled collisions. When all these n TFs independently as well as simultaneously bind with the nonspecific sites of the template DNA then we find $\tau_{ns,n} \sim \tau_{ns,1}$. When there is a sequential binding in a temporally nonoverlapping manner, then $\tau_{ns,n} \sim n\tau_{ns,1}$. One should note that D_o is the diffusion coefficient associated with the 1D scanning dynamics of the TF molecule on the genomic DNA. When a single TF is involved in the activation of transcription ($n = 1$) then we find that $\tau_E \approx \tau_R \approx \tau_s$. When there is a combinatorial binding of many TFs, then we find $\tau_E \gg \tau_R$. The optimum length (λ_o) of one-dimensional scanning [21,27] that is required to achieve the minimum of the search time $\tau_{s,o}$ in Eq. (2) can be calculated by solving $\partial_\lambda \tau_s = 0$ for λ as $\lambda_o = \sqrt{6D_o\tau_{ns,n}n^{-\alpha}}$ and $\tau_{s,o} = N\sqrt{2n^{\alpha}\tau_{ns,n}/3D_o}$.

B. Looping time and loop-relaxation time

Assume that the ETF and PR complexes are already formed. The quantity that we want to calculate here is the average time τ_L required by these two complexes, which are connected by the IDS, to collide with each other via the spatial diffusion mechanism. The looping dynamics of DNA and its effects on the site-specific binding of various protein molecules have been extensively investigated earlier. Saiz *et al*. [37] have shown that in addition to the intrinsic periodicity of the DNA double helix the free energy of looping DNA has an oscillatory component of about half the helical period. Moreover, the oscillations have such an amplitude that the effects of regulatory molecules become strongly dependent on their precise DNA positioning and yet easily tunable by their cooperative interactions. Saiz and Vilar [38] have developed and experimentally tested an *ab initio* thermodynamics model on the looping-mediated distal multisite transcription regulation system which is based on the well-known *E. coli* lac Operon. Recent single-molecule experiments by van den Broek *et al*. [31] and subsequent theoretical investigations by Lomholt *et al*. [29] revealed that the site-specific association of EcoRV restriction enzyme with DNA can be enhanced significantly when the template DNA is allowed to coil and loop freely. Here one should note that under *in vivo* conditions the random loops generated from the chromosomal DNA upon remodeling could in turn enhance the collision dynamics of ETF and PR like that of the intersegmental transfers in the site-specific DNA protein interactions [29]. In such cases, those random loop regions of the genomic DNA, which are flanking the connecting segment of DNA, can enhance the rate of three-dimensional diffusion-limited collision between ETF and PR complexes. The expression for τ_L will be dependent on the type of distal communication between the ETF and PR complexes. When the mode of this distal action is via looping out of the IDS, then the free-energy potential [39–42] associated with such a thermally driven looping dynamics will be $f(r) = k_BT(aL/2r^2)$ where a is the persistent length of the DNA chain, L is the contour length of the IDS whose current radius of curvature is r, k_B is the Boltzmann constant, and T is the absolute temperature. The force generated by this potential is $F(r) = -d_r f = \omega/r^3$, where we have defined the term $\omega = k_BTa\ L$. When the TF molecule undergoes thermally driven diffusion dynamics on three-dimensional space with an average 3D diffusion coefficient of $D_r = k_BT/3\pi\eta R_P$, then the time-dependent change in r of the IDS can be described by the following Langevin-type equation:

$$d_t r = \alpha r^{-3} + \sqrt{D_r}\xi_{r,t}. \tag{3}$$

Here $\alpha = \omega/6\pi\eta R_P = D_r aL/2$, where the average diameter of the TF molecule is R_P and η is the viscosity of the medium. The random term $\xi_{r,t}$ is the Gaussian-distributed white noise with mean $\langle\xi_{r,t}\rangle = 0$ and variance $\langle\xi_{r,t}\xi_{r,t'}\rangle = \delta(t - t')$. The Fokker-Planck equation (FPE) associated with the evolution of the probability of finding a given radius of curvature r at time t can be given in the spherical coordinate system as follows:

$$\begin{aligned} \partial_t P(r,t|R,0) = &-\partial_r[\alpha r^{-3}P(r,t|R,0)] \\ &+ (D_r/2r^2)\partial_r[r^2\partial_r P(r,t|R,0)]. \end{aligned} \tag{4}$$

Here the initial condition is $P(r,0|r_0,0) = \delta(r - r_0)$ and the absorbing and reflecting boundary conditions are given as follows:

$$P(L/2\pi,t|R,0) = 0; \quad [(D_r/2)\partial_r P(r,t|R,0)]_{r=R} = 0. \tag{5}$$

The radius of curvature of the IDS, r, cannot increase beyond r=R which is induced upon binding of the transcription factors at *cis*-acting sites and acts as a reflecting boundary for the looping dynamics of the IDS in the space of radius of curvatures. The mean first-passage time (MFPT) $T(r)$ associated with the transition of r from r=R to $r = L/2\pi$ obeys the following backward-type Fokker-Planck equation [43–46]:

$$\alpha r^{-3}d_r T(r) + (D_r/2r^2)d_r[r^2 d_r T(r)] = -1. \tag{6}$$

The boundary conditions for Eq. (6) directly follow from the boundary conditions given in Eq. (5) as follows:

$$T(L/2\pi) = 0; \quad [(D_r/2)d_r T(r)]_{r=R} = 0. \tag{7}$$

The integral solution of the differential equation (6) for the absorbing and reflecting boundary conditions as given in Eqs. (7) can be written as

$$T_l(r) = \frac{1}{3D_r}\int_{L/2\pi}^{r}\{-2s + 2aL/s - e^{aL/2s^2 - aL/2R^2}(-2R^3 + 2aLR)/s^2 + (aL)^{3/2}(2\pi)^{1/2}e^{aL/2s^2} \times [\mathrm{erf}(\sqrt{2aL}/2s) - \mathrm{erf}(\sqrt{2aL}/2R)]/s^2\}ds. \quad (8)$$

Here the expression $\mathrm{erf}(x) = \sqrt{4/\pi}\int_0^x e^{-t^2}dt$ is the error function integral. When the persistence length a approaches 0, then the MFPT expression that is given in Eq. (8) reduces to the following form:

$$\lim_{a\to 0} T_l(r|L, D_r, a, R) = (1/13D_r)[(L/2\pi)^2 + 4\pi R^3/4\pi R^3/L - r^2 - 2R^3/r]. \quad (9)$$

Here one should note that $\tau_L = T_l(R)/\kappa$ in Eqs. (6)—(9), where $0 < \kappa \leqslant 1$ is the steric factor to take care of the effects of the twisting dynamics of the IDS loop that can potentially lead to the misalignment of ETF and PR complexes [5,6,13,25,26,47–51]. When there is an insulator element present in between the promoter and the *cis*-regulatory module, then we find that $\kappa \to 0$. This means that upon binding of the repressing protein molecules at the insulator elements $\tau_L \to \infty$. For a typical system a, R, and D_r will be fixed quantities and $r = R$ when the process starts with an initial curvature of the connecting segment of DNA that is caused by an already bound transcription factor at the respective *cis*-acting binding site. Similarly one can derive the expression for τ_{LT} from Eq. (6) by setting the initial value of r as $r_0 = L/2\pi$ and with the following boundary conditions:

$$T(R) = 0;\ [(D_r/2)d_r T(r)]_{r=L/2\pi} = 0. \quad (10)$$

The integral solution to the boundary conditions given by Eq. (10) can be given as

$$T_{lt}(r) = \frac{1}{3D_r}\int_{R}^{r}\{-2s + 2aL/s - e^{aL/2s^2 - 2a\pi^2/L}(-L^3/4\pi^3 + aL^2/\pi)/s^2 + (aL)^{3/2}(2\pi)^{1/2}e^{aL/2s^2} \times [\mathrm{erf}(\sqrt{2aL}/2s) - \mathrm{erf}(\pi\sqrt{2aL}/L)]/s^2\}ds. \quad (11)$$

When the persistence length a approaches 0 then the MFPT expression given in Eq. (11) reduces to the form

$$\lim_{a\to 0} T_{lt}(r|L, D_r, a, R) = (1/3D_r)(L^3/4\pi^3 R + R^2 - r^2 - L^3/4\pi^3 r). \quad (12)$$

Here one should note that $\tau_{LT} = T_{lt}(L/2\pi)$ in Eqs. (11) and (12), which follows from the fact that at the time of escape of RNAP/PolII into the elongation step of transcription, the radius of curvature of the IDS will be such that $r = L/2\pi$.

III. RESULTS AND DISCUSSION

From literature [48,49] we find the typical persistence length of DNA as $a \approx 500\ \approx 150$ bps. Here we use the transformation rule 1 bps ≈ 3.4 . The three-dimensional diffusion coefficient associated with the dynamics of the transcription factor is $D_r \approx 0.8\ \mathrm{m^2\ s^{-1}} \approx 7 \times 10^6\ \mathrm{bps^2\ s^{-1}}$ and the one-dimensional diffusion coefficient associated with the sliding of the RNAP/TF molecules along DNA is $D_o \approx 0.092\ \mathrm{m^2\ s^{-1}} \approx 8 \times 10^5\ \mathrm{bps^2\ s^{-1}}$ [52]. The typical radius of curvature due to bending of the connecting segment of DNA upon binding of protein molecules such as histone particles will be typically $R \approx 55$Å ≈ 16 bps. Using these experimentally determined parameters one can calculate the site-specific search times under various conditions as follows. Various numerical quantities used in our calculations are summarized in Table I.

A. Case I: Prokaryotes

The cellular volume [53–55] of a typical *E. coli* cell is $V_C \sim 10^{-18}\ \mathrm{m^3}$ and the concentration of a single TF molecule and a single *cis*-acting site inside this cellular volume will be ~2 nM. The maximum achievable diffusion-limited collision rate under such *in vitro* aqueous conditions will be in the order of $\sim 10^8\ \mathrm{M^{-1}\ s^{-1}}$. Due to slower diffusion dynamics of the transcription factor inside the cell, this three-dimensional collision rate limit will be ~10 times smaller at $\sim 10^7\ \mathrm{M^{-1}\ s^{-1}}$ which is equal to $\sim 10^{-2}\ \mathrm{nM^{-1}\ s^{-1}}$. Since there are $N \sim 4.6 \times 10^6$ numbers of nonspecific binding sites available for a given transcription factor inside an *E. coli* cell, the minimum time that is required for the nonspecific binding of transcription factors with the genomic DNA will be $\tau_{ns,1} \sim 5 \times 10^{-6}$ s. Upon substituting these values in Eq. (2) we find that $\lambda_o = \sqrt{6D_o\tau_{ns,1}} \sim 5$ bps and $\tau_{s,o} = \tau_E = \tau_R \sim 10$ s. In these calculations, we have assumed that the genomic DNA is homogeneously distributed all over the cellular volume. Inside an *E. coli* cell the genomic DNA is compressed and confined [53–55] inside a volume of $V_D \sim 2 \times 10^{-19}\ \mathrm{m^3}$. The probability $p_c = V_D/V_C$ that the transcription factor nonspecifically binds with DNA while it is randomly diffusing inside the entire volume of the cytoplasm will be $p_c \sim 5 \times 10^{-2}$. Upon considering these factors, we find $\tau_{ns,1} \sim 10^{-4}$ s, $\lambda_o \sim 22$bps, and $\tau_{s,o} = \tau_E = \tau_R \sim 40$ s.

TABLE I. Summary of various parameters used in the calculations.

Quantity	Prokaryote (*E. coli*) I	Eukaryote (human) II	*In vitro* (NtrC) III
$\{\tau_E,\tau_R\}$	$\sim$40 s	$\sim$10 s	$\sim$ 0.01 s
τ_L	Equation (8), depends on L		
τ_I	$\sim 10^2$ s	$\sim 10^2 n^{\alpha/2}$s	$\sim$1 s
k_S	$\sim$40 bps/s	$\sim$70 bps/s	$\sim$40 bps/s
τ_{EL}	$\sim S/k_S$s		
τ_{TR}	$\sim$1 s		
τ_{LT}	Equation (11), depends on L		
τ_T	$\sim(10^2 + S/k_S)$	$\sim(10^2 n^{\alpha/2} + S/k_S)$	$\sim(1 + S/k_S)$
S_c	$S_c \sim 4 \times 10^3$ bps	$\sim 7 \times 10^3 n^{\alpha/2}$ bps	$\sim$40bps
$\tau_{ns,1}$	$\sim 10^{-6}$ s	$\sim 10^{-7}$ s	$\sim 10^{-6}$ s

B. Case II: Eukaryotes

We consider the human cell, which contains $N \sim 3 \times 10^9$ bps of genomic DNA that is confined inside the cellular nucleus whose diameter is $d_n \sim 6$ μm. The corresponding nuclear volume will be $V_N \sim 10^{-16}$ m^3. The concentration of a single transcription factor or a single binding site on the genomic DNA in this nuclear volume will be $\sim$20 pM. We assume a maximum achievable 3D diffusion-controlled bimolecular collision rate $\sim 10^7$ M^{-1} s^{-1} $= 10^{-5}$ pM^{-1} s^{-1}. Under these conditions when $D_o \approx 8 \times 10^5$ bps^2 s^{-1} then the minimum search time required for nonspecific binding of the transcription factor with the genomic DNA will be $\tau_{ns,1} \sim 10^{-7}$ s and the optimum search time for site-specific binding of TF at its *cis*-regulatory site will be in the order of $\tau_{s,o} \sim 10^3$ s. When there are n numbers of combinatorial transcription factors involved in the activation of transcription then we find that $\tau_{s,o} \sim 10^3 n^{\alpha/2}$ s. This is the maximum possible value of $\tau_{s,o}$. When the *in vivo* three-dimensional diffusion-controlled collision rate is similar to the *in vitro* value then we find that $\tau_{s,o} \sim 10^2 n^{\alpha/2}$ s. All these calculations assume that the segment of DNA containing the CRMs is already released from the histone core particles and it is freely accessible for the one-dimensional searching of transcription factors. When the mean time that is required to remodel the chromosomal DNA upon arrival of all the n TFs closer to their respective *cis*-regulatory modules is τ_{CR} then we find that $\tau_{s,o} = \tau_E = \tau_R \sim (10^3 n^{\alpha/2} + \tau_{CR})$.

C. Case III: *In vitro* experiments

For the *in vitro* experimental setup with the NtrC- (TF-) dependent enhancer system that is used in Ref. [12] one can compute the search times associated with the site-specific binding of this TF molecule as

$$N \sim 4 \times 10^3 \text{ bps; [DNA]} \sim 2.8 \text{ nM; [NtrC]} \sim 120 \text{ nM;}$$
$$\text{[RNAP]} \sim 500 \text{ nM; } L \sim 110 \text{ bps.} \quad (13)$$

Under these *in vitro* conditions, the maximum possible three-dimensional diffusion-controlled collision rate limit will be $\sim 10^8$ M^{-1} s^{-1} and the nonspecific interaction time with respect to the settings given by Eq. (13) for both RNAP and the NtrC (transcription factor) with their respective promoter and *cis*-regulatory sites on the template DNA is $\tau_{ns,1} \sim 10^{-6}$ s. When the optimum length of one-dimensional scanning is $\lambda_o \sim 2$ bps, then the search times taken by RNAP and TF to locate their respective binding sites on the template DNA of size $\sim$4 kbps under *in vitro* conditions will be $\tau_s = \tau_E = \tau_R \sim 10$ ms.

D. Looping time and loop-relaxation time

To compute the values $\{\tau_L,\tau_{LT}\}$ we can use the following parameter settings in Eqs. (8) and (11):

$$r = L/2\pi; \quad D_r = 7 \times 10^6 \text{ bps}^2 \text{ s}^{-1}; \quad a \sim 150 \text{ bps;}$$
$$R = 16 \text{ bps;} \quad L \in (15,250) \text{ bps.} \quad (14)$$

The resulting plots of $\{\tau_L,\tau_{LT}\}$ obtained upon evaluating the integrals in Eqs. (8) and (11) at various values of L are

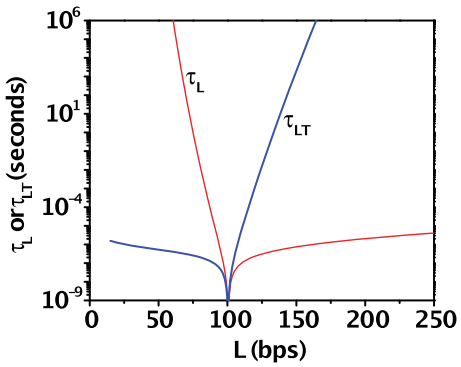

FIG. 2. (Color online) Looping and loop-relaxation times as functions of the length (L) of the DNA segment that connects the promoter and the *cis*-acting elements of the transcription factor. Equations (8) and (11) were used for computation purposes with the parameters given in Eq. (14). This result clearly suggests that when $L < 100$ then $\tau_L > \tau_{LT}$ and when $L > 100$ then $\tau_L < \tau_{LT}$ under identical conditions. This means that all the inequalities given in Eq. (1), which are required to observe the dynamical memory in the transcription activation, cannot be satisfied, and as a result the memory effects cannot be observed. In the presence of bulky histone bodies in the connecting segment of DNA, as in case of memory gene loops, both of $\{\tau_L,\tau_{LT}\}$ will be higher due to the slower dynamics of the histone-loaded chromosomal DNA segment. Under such conditions all the inequalities given in Eq. (1) will be satisfied and a static memory can be observed in the eukaryotic transcription activation mediated by the memory gene loops.

shown in Fig. 2. These results suggest that the inequalities $\tau_L < \max(\tau_R,\tau_E)$ and $\tau_{LT} \gg \tau_I$ will be true in all the cases I–III whenever the contour length of the connecting DNA segment is such that $L > 150$ bps.

E. Conditions to observe transcriptional memory

The inequality condition $\tau_I > (\tau_{EL} + \tau_{TR})$ will be true and in turn it depends on the length (S) of the transcript. When the average transcription elongation speed of the RNAP/PolII is k_S bps s^{-1}, then we find that $\tau_{EL} \sim S/k_S$. This means that there exists a critical length of the complete transcript, S_c, beyond which the condition $\tau_I < (S_c/k_S + \tau_{TR})$ is true. The experimentally observed value of k_S for prokaryotic [56] RNAP is $k_S \sim 40$bps s^{-1} and the transcription elongation rate [57] of eukaryotic PolII is $k_S \sim 70$bps s^{-1}. The experimentally observed value of the transcription termination time [58] is $\tau_{TR} \sim 1$ s. For all these experimentally observed values and $L > 100$bps, the total transcription time for various situations can be given as follows:

$$\begin{aligned}
&\text{CaseI}: \quad \tau_I \sim 10^2 \text{ s}; \ \tau_T \sim (10^2 + S/k_S)\text{s}; \quad S_c \sim 4 \times 10^3 \text{ bps},\\
&\text{CaseII}: \quad \tau_I \sim 10^2 n^{\alpha/2}\text{s}; \quad \tau_T \sim \left(10^2 n^{\alpha/2} + S/k_S\right)\text{s}; \quad S_c \sim 7 \times 10^3 n^{\alpha/2} \text{ bps},\\
&\text{CaseIII}: \quad \tau_I \sim 1\text{s}; \quad \tau_T \sim (1 + S/k_S)\,\text{s}; \quad S_c \sim 40\text{bps}.
\end{aligned} \tag{15}$$

From Eq. (15) we find that all the inequalities in the set I_q given in Eq. (1) are not true for case III and this could be the reason why the memory effects were not observed in that *in vitro* transcription experiment [12]. Our analysis also suggests that the second and third inequalities in the set I_q will be true whenever $L > 150$ and the following conditions are true:

$$\begin{aligned}
&\text{Case I}: S \leqslant (4 \times 10^3) \quad ; \text{CaseII}: S \leqslant (7 \times 10^3 n^{\alpha/\alpha 2});\\
&\text{Case III}: S \leqslant 40.
\end{aligned} \tag{16}$$

From the results presented in Fig. 2, we find that the first inequality $\tau_L \gg \max(\tau_R,\tau_E)$ will be true whenever $L < 100$ and the inequality $\tau_{LT} \gg \tau_I$ will be true whenever $L > 100$ in cases I and II. This means that the dynamical memory cannot be observed in all the cases I–III. It is apparent that all the inequalities in the set I_q given by Eq. (1) should be true in the case of experimentally observable static memory that is induced by the "memory gene loops" in eukaryotes [18,19]. Unlike looping of the connecting segment of DNA that is the origin of the dynamical memory, in the case of static memory gene loops, the connecting segment of DNA is compact and loaded with histone-mediated chromosomal structures. In such cases the looping dynamics as well as loop-relaxation dynamics of the connecting segment of DNA are much slower due to the presence of intact bulky histone bodies, which results in the higher values of $\{\tau_L,\tau_{LT}\}$. Since such higher-order chromosomal structures in the connecting segment of DNA are absent in prokaryotes, static memory is not experimentally observed or not necessary for the transcription activation in prokaryotes [12]. These results suggest that the physics of the chromosomal compaction of the genomic DNA also plays a critical role in carrying the memory of the set of expressing genes and their expression levels across many mitotic cell division cycles. In this connection one should also note that the values of the time components $\{\tau_L,\tau_{LT}\}$ could also be tuned under *in vitro* or *in vivo* conditions by introducing cognate sites for other DNA binding proteins, which are different from the TF system under consideration, in the IDS region. This is similar to the method of introducing silencing elements (Fig. 3) with the condition that there are no protein-protein interactions among those protein moleculesthat are binding at these regions.

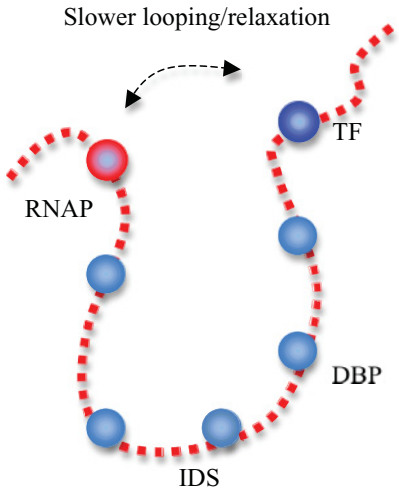

FIG. 3. (Color online) Possible methodology to increase the looping and loop-relaxation times. Here the binding sites for other DNA-binding proteins, which are different from the TF system under consideration, are introduced in the IDS region. This is similar to the method of silencing elements with the conditions that there are no protein-protein interactions among these protein molecules. Such loading of bulky protein molecules in the IDS region can slow down the looping as well as the loop-relaxation dynamics which in turn can validate the inequality conditions given in Eq. (1), which are required to observe the memory effects in transcription activation.

Molecular crowding in the IDS can also retard the looping dynamics of the IDS, which can in turn result in lower values of these time components, which is the prerequisite for observing the memory effects in transcription activation. This means that memory effects in the transcription activation play critical roles in the crowded situations of the cell nucleus. Throughout the paper we have assumed that the free-energy potential associated with the looping dynamics of intervening DNA segments scales with the radius of curvature of the loop as $f \propto 1/r^2$. The experimentally observed [59] overall stability

of DNA loops varies with the loop length L (particularly for L 100 bps) as $\Delta G_L \approx \Delta G_0 + 1.24RT \ln\left(L/L_0\right)$ where L_0 is the reference loop length and ΔG_0 is the stability of the loop when $L = L_0$. Here ΔG_L is the overall stability of the DNA loop which is mainly contributed by the stabilizing free energy that is released upon the binding of repressor or transcription factors at their cognate sites and the subsequent protein-protein interactions and the destabilizing free energy associated with the loop-relaxation dynamics. These observations [59] suggest a scaling of the form $f \propto 1/r^{\varepsilon}$, where $\varepsilon \sim 1.24$. Here one should note that the value of the scaling exponent $\varepsilon = 2$ corresponds to the looping dynamics of a flexible polymer, and the stabilizing free-energy components associated with the binding of the repressor proteins or transcription factors are ignored. Decrease of the exponent ε suggests that the stabilizing components help to overcome the entropic barriers associated with the loop-relaxation dynamics, which in turn decreases the dependency of the overall loop stability on the radius of curvature r of the loop. However, these stabilizing components will come into the picture only when r is such that $r = L/2\pi$, which is a prerequisite for the protein-protein interactions. As a result, the looping time associated with transition from $r = R$ to $r = L/2\pi$ will not be affected much. On the other hand, the loop-relaxation time will be strongly dependent on the free-energy barrier introduced upon the binding of transcription factors. This means that the first and second inequalities in Eq. (1) will be true when the number of such protein-protein interactions and their binding energies are high enough, and the memory effects will be prominently observed when the length of the corresponding mRNA transcript is such that the third inequality in Eq. (1) is true. The probability associated with the formation of DNA loops seems to be strongly dependent on the number of cognate binding sites as well as the *in vivo* concentrations of the corresponding proteins [60]. At some intermediate concentrations of the transcription factors the occupancy of the binding sites showed a discontinuous phase-transition-like behavior between the looped and open states of DNA. Further analysis showed [60] that the looping of DNA will be observed only when the interoperator protein interactions exceed a certain binding energy threshold, which approaches zero as the number of cognate sites involved in the looping increases. This also means that the dynamical memory effects in the looping-mediated transcription activation can be observed only beyond this binding energy threshold. The memory effects in looping-mediated transcription activation can lead to fluctuations in the protein synthesis from higher to lower values, as pointed out by Saiz and Vilar in Ref. [59], since the gene of interest is turned on and off for longer times. When the transcription switches slowly between active and inactive states, then there are long periods of time in which proteins are produced constantly and long periods of time without any production [59]. Together with the requirement of a binding energy threshold to observe the looping, the tradeoff between the requirement of transcriptional memory to reduce the response time of a repressed gene to an external stimulus and the requirement to control the fluctuations in the synthesized protein numbers seems to limit the number of transcription factor binding sites which can be involved in the looping-mediated transcription activation.

IV. CONCLUSIONS

We have developed a theory to explain the origin of the static and dynamical memory effects in transcription activation mediated by transcription factors. Our results suggest that the following conditions should be satisfied to observe such memory effects: (a) $\tau_L \gg \max(\tau_R, \tau_E)$, (b) $\tau_{LT} \gg \tau_T$, and (c) $\tau_I \geqslant (\tau_{EL} + \tau_{TR})$, where τ_L is the average time required for the looping-mediated spatial interactions of the enhancer—transcription-factor complex with the promoter—RNA-polymerase (PolII in eukaryotes) complex that is located L base pairs away from the *cis*-regulatory enhancer site, (τ_R, τ_E) are respectively the search times required for the site-specific binding of the RNA polymerase with the promoter and the transcription factor with the *cis*-regulatory site, τ_{LT} is the time associated with the relaxation of the looped-out segment of DNA that connects promoter and *cis*-acting site, τ_T is the time required to generate a complete transcript, τ_I is the transcription initiation time, τ_{EL} is the transcription elongation time, and τ_{TR} is the termination time. We have theoretically derived expressions for the various searching, looping, and loop-relaxation time components. Using the experimentally determined values of these time components, we further show that the dynamical memory effects cannot be experimentally observed whenever the segment of DNA that connects the *cis*-regulatory elements with the promoter is not loaded with bulky histone bodies. Our analysis suggested that the presence of histone-mediated compaction of the connecting segment of DNA can result in higher values of looping and loop-relaxation times, which is the origin of the static memory in the transcription activation that is mediated by the memory gene loops in eukaryotes.

[1] B. Lewin, *Genes VIII* (Oxford University Press, London, 2004).
[2] B. Alberts, D. Bray, J. Lewis, K. Roberts, and J. D. Watson, *Molecular Biology of the Cell* (Garland, New York, 2004).
[3] M. Ptashne, *A Genetic Switch* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004).
[4] M. Ptashne and A. Gann, *Genes and Signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
[5] J. C. Wang and G. N. Giaever, Science **240**, 300 (1988).
[6] E. M. Blackwood and J. T. Kadonaga, Science **281**, 61 (1998).
[7] D. R. Herendeen, G. A. Kassavetis, and E. P. Geiduschek, Science **256**, 1298 (1992).
[8] M. C. Walters *et al.*, Proc. Natl. Acad. Sci. USA **92**, 7125 (1995).
[9] R. Heuchel, P. Matthias, and W. Schaffner, Nucleic Acids Res. **17**, 8931 (1989).

[10] Y. Liu *et al*., Proc. Natl. Acad. Sci. USA **98**, 14883 (2001).
[11] M. A. Rubtso *et al*., Proc. Natl. Acad. Sci. USA **103**, 17690 (2006).
[12] V. Bondarenko, Y. Liu, A. Ninfa, and V. M. Studitsky, Nucleic Acids Res. **30**, 636 (2002).
[13] K. Rippe, P. H. von Hippel, and J. Langowski, Trends Biochem. Sci. **20**, 500 (1995).
[14] B. Tolhuis *et al*., Mol. Cell. **10**, 1453 (2002).
[15] J. Ling *et al*., J. Biol. Chem. **279**, 51704 (2004).
[16] R. Murugan, J. Theor. Biol. **248**, 696 (2007).
[17] V. A. Bondarenko *et al*., Biochem. Cell Biol. **81**, 241 (2003).
[18] N. J. Francis and R. E. Kingston, Nat. Rev. Mol. Cell. Biol. **2**, 409 (2001).
[19] J. H. Brickner, Curr. Biol. **20**, R20 (2010).
[20] M. Ptashne and A. Gann, Nature (London) **386**, 569 (1997).
[21] R. Murugan, J. Phys. A **43**, 195003 (2010).
[22] R. D. Kornberg and Y. Lorch, Cell **98**, 285 (1999).
[23] O. I. Kulaeva, D. A. Gaykalova, and V. M. Studitsky, Mutation Res. **618**, 116 (2007).
[24] O. G. Berg, R. B. Winter, and P. H. von Hippel, Biochemistry **20**, 6929 (1981).
[25] O. G. Berg, R. B. Winter, and P. H. von Hippel, Biochemistry **20**, 6961 (1981).
[26] M. Slutsky and L. A. Mirny, Biophys. J. **87**, 4021 (2004).
[27] R. Murugan, Phys. Rev. E **79**, 061920 (2009).
[28] M. Coppey, O. Bénichou, R. Voituriez, and M. Moreau, Biophys. J. **87**, 1640 (2004).
[29] M. A. Lomholt *et al*., Proc. Natl. Acad. Sci. USA **106**, 8204 (2009).
[30] I. M. Sokolov *et al*., Biophys. J. **89**, 895 (2005).
[31] B. van den Broek *et al*., Proc. Natl. Acad. Sci. USA **105**, 15738 (2008).
[32] I. Bonnet *et al*., Nucleic Acids Res. **36**, 4118 (2008).
[33] Y. M. Wang, R. H. Austin, and E. C. Cox, Phys. Rev. Lett. **97**, 048302 (2006).
[34] C. G. Kalodimos *et al*., Science **305**, 386 (2004).
[35] L. Hu, A. Y. Grosberg, and R. Bruinsma, Biophys. J. **95**, 1151 (2008).
[36] R. Murugan, Biophys. J. **99**, 353 (2010).
[37] L. Saiz, M. Rubi, and J. M. G. Vilar, Proc. Natl. Acad. Sci. USA **102**, 17642 (2005).
[38] L. Saiz and J. M. G. Vilar, Nucleic Acids Res. **36**, 726 (2008).
[39] R. Schleif, Annu. Rev. Biochem. **61**, 199 (1992).
[40] Y. Zhang, A. E. McEwen, D. M. Crothers, and S. D. Levene, Biophys J. **90**, 1903 (2006).
[41] P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, NY, 1979).
[42] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon, London, 1958).
[43] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 2002).
[44] H. Risken, *Fokker Plank Equations* (Springer, Berlin, 1996).
[45] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 2004).
[46] S. Redner, *A Guide to First Passage Processes* (Cambridge University Press, London, 2001).
[47] D. Shore and R. L. Baldwin, J. Mol. Biol. **170**, 957 (1983).
[48] Y. L. Zhang and D. M. Crothers, Biophys. J. **84**, 136 (2003).
[49] V. A. Bloomfield, D. M. Crothers, and I. Tinoco, *Nucleic Acids: Structures, Properties, and Functions* (University Science Books, Sausalito, CA, 1999).
[50] J. A. Schelman, Biophys. Chem. **11**, 321 (1980).
[51] J. A. Schelman, Biopolymers **13**, 217 (1980).
[52] J. Elf, G. W. Li, and X. S. Xie, Science **316**, 1191 (2007).
[53] D. S. Goodsell, Trends. Biochem. Sci. **16**, 203, (1991).
[54] T. A. Azam *et al*., J. Bacteriol. **181**, 6361 (1999).
[55] R. C. Johnson *et al*., in *The Bacterial Chromosome, edited by* N. P. Higgins (ASM Press, Washington DC, 2005).
[56] J. W. Shaevitz, E. A. Abbondanzieri, R. Landick, and S. M. Block, Nature (London) **426**, 684 (2003).
[57] X. Darzacq *et al*., Nature Struct. Mol. Biol. **14,** 796 (2007).
[58] S. J. Greive *et al*., Proc. Natl. Acad. Sci. USA **105**, 3315 (2007).
[59] L. Saiz and J. M. G. Vilar, Curr. Opin. Struct. Biol. **16**, 344 (2006).
[60] J. M. G. Vilar and L. Saiz, Phys. Rev. Lett. **96**, 238103 (2006).