Theory on the looping mediated stochastic propulsion of transcription factors along DNA

R. Murugan

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

Phone: +91 44 2257 4116. E-mail: rmurugan@gmail.com

ABSTRACT

We demonstrate that DNA-loops can stochastically propel the site-specifically bound transcription factors towards the promoters. The gradual release of elastic energy stored on the DNA-loops is the source of propulsion. The speed of looping mediated interaction of transcription factors with promoters is several times faster than the sliding mode. Elastic and entropic energy barriers associated with the looping of DNA actually shape up the distribution of distances between transcription factor binding sites and promoters. The commonly observed multiprotein binding in gene regulation is acquired through evolution to overcome the looping energy barrier. Presence of nucleosomes on the genomic DNA of eukaryotes is required to reduce the entropy barriers associated with the looping.

KEY WORDS

DNA-protein interactions, transcription factor binding, DNA-looping, *cis*-regulatory modules

INTRODUCTION

Site-specific binding of transcription factors (**TFs**) at their *cis*-regulatory motifs (**CRMs**) on the genomic DNA in the presence enormous amount of nonspecific binding sites is essential for the activation and regulation of several genes across prokaryotes to eukaryotes (1-3). Binding of TFs with their CRMs was initially thought as a single-step three-dimensional (**3D**) diffusion-controlled collision process. Kinetic experiments on *lac*-repressor-Operator system revealed a bimolecular rate in the order of $\sim 10^9$-10^{10} $M^{-1}s^{-1}$ that is ~ 10-10^2 times faster than the Smolochowski type 3D diffusion-controlled rate limit. Berg et.al. (4, 5) successfully explained this inconsistency using a two-step mechanism by establishing the key concept that TFs first bind with DNA in a nonspecific manner via 3D diffusion and then search for their cognate sites via various one-dimensional (**1D**) facilitating processes such as sliding, hopping and intersegmental transfers. Here 1D diffusion with unit base-pair step-size of TFs is the sliding, few base-pairs (bp, 1 bp = $l_d \sim 3.4$ x 10^{-10} m) step-size is called hopping and few hundred to thousand bps step-size is called intersegmental-transfer. Intersegmental transfers occur whenever two distal segments of the same DNA polymer come in nearby over 3D space via ring closure events (6-8).

Specific binding of TFs with DNA is affected by several factors (8) viz. **a**) conformational state of DNA (8, 9) **b**) spatial organization of various functionally related combinatorial CRMs along the genomic DNA (10, 11), **c**) presence of similar or other dynamic roadblock proteins (12) and semi-stationary roadblocks such as nucleosomes in eukaryotes (13-17), **d**) naturally occurring sequence mediated kinetic traps on DNA (18, 19), **e**) conformational fluctuations in the DNA binding domains of TFs (20-22) and **f**) the nonspecific electrostatic attractive forces and the counteracting shielding effects of other solvent ions and water molecules acting at the DNA-protein interface (23). Several theoretical models (7, 8, 18, 21, 24), computational (25-28) and experimental studies have been carried out to understand the effects of factors **a-f** on the kinetics of site-specific DNA-protein interactions.

In general, the searching efficiency of TFs depends on the relative amount of times spent by them on the 3D and 1D diffusions (10, 21). Clearly, neither pure 1D nor 3D diffusion is an efficient mode of searching (8, 10). Under ideal situation, maximum searching efficiency can be achieved only when TFs spend equal amount of times in both 1D and 3D diffusions (7, 10). This trade off balance between the times spent on different modes of diffusions will be modulated by the factors **a-f**. For example, presence of nucleosome roadblocks warrants more dissociations and 3D excursions of TFs rather than 1D sliding (29). Sequence specific fast conformational switching of DNA binding domains between stationary and mobile states helps TFs to overcome the sequence traps (29). Relaxed conformational state of DNA enhances more sliding rather than hopping and intersegmental transfers and so on (8). Conformational dynamics of DNA also modulates the speed of gene activation and regulation. In this context, looping of DNA is critical for the activation and expression of various genes across prokaryotes to eukaryotes (3, 30-34). Combinatorial binding of TFs with their specific CRMs on the genomic DNA activates the downstream promoters of genes via looping of the intervening DNA segment to form a synaptosome type complex (1, 35). In most of the molecular biological processes, DNA-loops are warranted for the precise protein-protein interactions which are the prerequisites for transcription and recombination (36).

The statistical mechanics of looping and cyclization of linear DNA has been studied extensively in the literature (33, 37, 38). However, it is still not clear why DNA-loops have evolved as an

3

integral part of the activation and repression of transcription and recombination although such underlying site-specific protein-protein and protein-DNA interactions can also be catered straightforwardly via a combination of 1D and 3D diffusions of TFs (4, 5, 21, 39). That is to say, upon arrival at the CRMs, TFs can directly slide or hop along the DNA polymer to reach the promoters. Schleif (31) had argued that the looping of DNA can simplify the evolution of the genomic architecture of eukaryotes by not imposing strict conditions on the spacing between the TF binding sites and the promoters. This is logical since a given set of TFs need to regulate several different genes across the genome. Therefore, placement of TF binding sites near a specific gene can be a disadvantage for other genes along the genomic evolution. Similarly, placement of TF binding sites near every gene is not an efficient genome design. The DNA loops also play critical roles in the transcription bursting (40) and memory (41). It is not clear how exactly the DNA-loop is formed between the CRMs and promoters via TFs though Rippe et.al., (32) had already taken several snapshots of the looping intermediates. In this paper, we will show that the DNA-looping combined with an asymmetric binding energy profile can stochastically propel TFs towards the promoters along DNA. Using computational tools, we further demonstrate that the looping mediated propulsion or tethered sliding of TFs along DNA can actually help in finding the direction of the promoter region and also shape up the genomic architecture.

THEORETICAL FORMULATION
Let us first list out the basic facts observed on the mechanism of distal action of CRMs-TFs system on the downstream promoters in the process of transcription activation. Firstly, both theoretical investigations (4, 5, 7, 8, 21) and experimental observations (42, 43) suggest that TFs recognize their CRMs via a combination of 1D and 3D diffusions. The key idea here is that TFs scan a random piece of DNA via 1D diffusion after each of the 3D diffusion mediated nonspecific collisions (4, 8, 21). On the contrary, the reacting molecules dissociate immediately upon each of their unfruitful collisions in the standard Smolochowski model. When the dynamics of TFs is confined within the *Onsager radius* of the DNA-protein interface, then it is categorized as the 1D diffusion. When TFs escape out of the Onsager radius and perform free 3D excursions, then we classify it as the 3D diffusion (8). The *Onsager radius* connected with the DNA-protein interface is defined (8) as the distance between the positively charged DNA binding domains of TFs and the negatively charged phosphate backbone of DNA at which the overall electrostatic energy is same as that of the background thermal energy (equals to ~1 k_BT) (**Section 1**, Supporting Materials). Secondly, transcription activation is achieved upon the distal communication between the CRMs-TFs complex with the RNAP-promoter complex (1-3). Thirdly, binding of TFs at CRMs locally bends the DNA and the DNA-loops connecting CRMs-TFs with the promoters are observed in most of the transcriptionally active genes of eukaryotes (3, 44).

Clearly, TFs activate transcription via two sequential steps viz. they bind their CRMs in the first step and then distally communicate with the promoter-RNAP complex in the second step to initiate the transcription event. To understand the role of DNA-loops in the transcription activation, we consider two possible scenarios viz. looping mediated versus a hypothetical pure 3D1D diffusion mediated distal communication between the CRMs-TFs and the promoters. In both these scenarios, TFs locate their respective CRMs via a combination of 1D and 3D diffusion in the first step. They differ only in the second step where TFs dissociate from their CRMs and communicate with the promoters via a combination of 1D and 3D diffusions in the second case whereas the distal communication will be through the DNA-loops in the first case. We denote the search time

required by TFs to locate their CRMs in the first step of transcription activation as τ_S. Clearly, those factors **a-f** listed out in the introduction section significantly modulate this quantity. We will not recalculate this here since enormous amount of literature already exists (see **Section 1** of the Supporting Materials) on the derivation of this quantity under various conditions (8, 18, 21, 45). In the following sections, we will compute the mean time required by CRMs-TFs complex to communicate with the promoter via DNA-loops in the second step of the transcription activation.

Preliminary assumptions

Upon observing the open synaptic complexes of transcriptionally active genes of eukaryotes with DNA loops, one can conclude that TFs which activate transcription via DNA-loops have at least two different DNA binding domains (DBDs) viz. one corresponds to the CRM (DBD1) (**Fig. 1**) and the another corresponds to the binding site that is located proximal to the promoter (DBD2) region. For example, the tetrameric Lac I complex binds two different Operator regions that induces looping of DNA (2, 3). However, in this case the tetramers of repressor molecules bound at these two different binding sites communicate via protein-protein interactions among them. The DNA-loop is stabilized by an octamer form of the Lac I repressor protein. Such mechanisms are common in case of multiprotein mediated DNA-looping and transcription activation. We further assume that TF reaches its specific binding site in the first step via a combination of 3D and 1D diffusions (4, 5, 7, 21, 46) in line with two-step DNA-protein interaction model and subsequently bends the DNA upon site-specifically binding their CRMs (32, 33).

Energetics of the site-specific binding of TFs and bending of DNA

Let us assume that the radius of gyration of the TF of interest is r_P. Upon binding its cognate stretch of DNA with size of X_0 bp located in between S1 to S2, the TF bends the DNA segment into a circle around its spherical solvent shell surface such that $X_0 = 2\pi r_P$ as shown in **Fig. 1A**. We set $X = 0$ at S1 and $X = X_0$ at S2 where X is the current location of the DBD2 of TF on DNA. S1 is the specific site for DBD1 and P is the specific site for DBD2 by definition. Here DNA under consideration spans over the range $(0, L)$ as in **Fig. 1B** and X is the current loop-length. The total energy required to bend a linear DNA will be the sum $E_{\text{bend}} = E_{\text{elastic}} + E_{\text{entropy}}$. For the radius of curvature r_P, one finds that $E_{\text{elastic}} \simeq aX/2r_P^2$ (measured in k_BT units) where a is the persistence length of DNA (37, 47). Clearly, E_{elastic} required to bend the DNA segment of length X into a circle will be $E_{\text{elastic}} \simeq 2\pi^2 a/X$. This energy has to be derived either solely from the site-specific binding energy of TFs or via an external energy input in the form of ATP hydrolysis (48). Noting that $E_{\text{entropy}} \simeq (3/2)\ln(\pi X/6)$ (**Eq. A1** of **Appendix A**) one finally arrives at the following expression for the overall bending energy.

$$E_{\text{bend}} \simeq 2\pi^2 a/X + (3/2)\ln(\pi X/6) \tag{1}$$

Clearly, E_{bend} attains a minimum value as $E_{\text{bend}}^{\min} = (3/2)\left[1 + \ln\left(2\pi^3 a/9\right)\right]$ at $X_C = 4\pi^2 a/3$. In the later sections, we will show that this non-monotonic behavior of the bending energy profile will restrict the possible distances between the CRMs and their corresponding promoters.

Looping mediated communication between CRMs-TFs and promoters

When TFs bind their CRMs in the first step of transcription activation, then the *site-specific binding energy* (E_{bind}) released at the DNA-TF interface dissipates partially as the elastic energy required to bend the DNA chain (E_{elastic}), partially to form specific non-covalent bonds (E_{bond}, the enthalpic component) and partially as the energy required to compensate the chain entropy loss (E_{entropy}) at the specific binding site. Clearly, $E_{\text{bind}} = E_{\text{bond}} + E_{\text{bend}}$ where $E_{\text{bend}} = E_{\text{elastic}} + E_{\text{entropy}}$.

Therefore, the overall free energy stored by the site-specific CRM-TF complex is given by $E \simeq E_{\text{bond}} + E_{\text{elastic}}$. This is the overall potential energy barrier which acts on any kind of distortion or dissociation of the site-specific CRMs-TFs complex. Conversely, E_{bend} is the potential energy barrier that resists the formation of loops out of linear DNA.

The free-energy stored in the site-specific DNA-TF complex (E) can undergo three different modes of dissipation viz. **1**) thermal induced physical dissociation of TF from DNA in which both bonding and elastic energies dissipate into the heat bath along with increase in the chain entropy, **2**) physical dissociation of only DBD2 from S2 and its re-association somewhere via looping over 3D space (which is resisted by the loop-length dependent potential energy barrier E_{bend}) while S1-DBD1 is still intact as modelled by Shvets and Kolomeisky (49), **3**) stochastic propulsion of TF on DNA via sliding of DBD2 towards the promoter which can be achieved by gradual increase in the value of X from X_0 towards L and **4**) tethered sliding of DBD2 with intact DNA-loop and DBD1-S1 interactions. In the propulsion mechanism, mainly the elastic energy dissipates that causes bulging of the DNA-loop around TF. The chain entropy does not increase much here since the intervening DNA is still under loop conformation. This is similar to the sliding of nucleosomes via bulge induced reptation dynamics of DNA (29, 50, 51). The probability associated with the spontaneous dissociation will be inversely correlated with E_{bond} and positively correlated with E_{elastic}. Generally, dissociation will be an endothermic process since $E_{\text{bond}} > E_{\text{elastic}}$. Clearly, physical dissociation will not be the most probable route of dissipation of the energy stored in the site-specific DNA-TF complex.

With this background, the DBD2 of TF needs to distally interact with the promoter in the second step and activate the transcription via looping of the intervening DNA segment that connects the CRMs and the promoter. There are two different possibilities viz. tethered sliding of DBD2 of TF with intact DBD1-S1 and a stochastic propulsion of TFs with intact DBD1-S1. Shvets and Kolomeisky (49) have recently studied another interesting model with repeated binding-unbinding of DBD2 with intact DBD1-S1. However, in their model sliding of DBD2 of TF was not allowed. All the symbols used in this paper are listed in **Table S1** of the Supporting Material. In the following sections, we will develop our stochastic propulsion and tethered sliding models in detail.

Stochastic propulsion model

When the binding energy profile of TF is such that the bonding energy near S1 is much higher than S2, then the bending energy stored in the site-specific TF-DNA complex can be gradually released via bulging of the DNA-loop around TF which in turn stochastically propels the sliding DBD2 of TF towards the promoter located at L as shown in **Fig. 1B**. There is no straightforward experimental evidence for this model. However, one can construe this idea indirectly from various other experimental studies. Particularly, Rippe et.al (32) have studied NtrC (Nitrogen regulatory protein C) system using the scanning force microscopy. In this study, they had taken snapshots of various intermediary states along the process of transcription activation from the closed to the open

promoter complex. In their model system, binding of NtrC at its specific site (CRM) activates the downstream closed complex of *glnA* promoter-RNAP-σ^{54} via looping out of the intervening DNA segment. They have shown that the transition from the inactive-closed form to an active-open promoter complex involved a gradual increase in the bending angle of the intervening DNA. This in turn is positively correlated with an increase in the radius of curvature of the intervening DNA segment which is represented as bulging of the DNA-loop in our propulsion model. Therefore, our assumption that the propulsion of TFs via increase in the radius of curvature of the bent DNA is a logical one. Here the asymmetric binding energy profile is essential to break the symmetry of the stochastic force acting on the sliding TFs (52). This is also a logical assumption since S1-DBD1 is a strong site-specific interaction and S2-DBD2 is approximately nonspecific interaction by definition. **Fig. 1C** shows another possibility in the formation of DNA-loop which is common in case of silencing mode of TFs. Based on these, the dynamical position X of TF on DNA obeys the following Langevin type stochastic differential equation (53-55).

$$dX/dt = D_C F(X) + \sqrt{2D_C}\,\Gamma_t; \quad \langle\Gamma_t\rangle = 0; \quad \langle\Gamma_t\Gamma_{t'}\rangle = \delta(t-t') \,. \tag{2}$$

In **Eq. 2**, $F(X) = -dE/dX = 2\pi^2 a/X^2$ (bp^{-1}) is the force acting on TF that is generated by the bending potential $E \sim E_{\text{elastic}} + E_{\text{bond}}$ upon bulging of the DNA-loop, Γ_t is the Δ-correlated Gaussian white noise and D_C (bp^2/s) is the 1D diffusion coefficient of the sliding of TF. The energy involved in the bonding interactions will be a constant one so that it will not contribute to the force term. Here we ignore the energy dissipation via chain entropy of bulging DNA-loop mainly because binding of TFs at their specific sites attenuates the conformational fluctuations at the DNA-TF interface (7, 20, 21). The Fokker-Planck equation describing the probability of observing a given X at time t with the condition that $X = X_0$ at $t = t_0$ can be written as follows (53, 54).

$$\partial_t P(X,t\,|\,X_0,t_0) = -D_C\left(\partial_X\left[F(X)P(X,t\,|\,X_0,t_0)\right] - \partial_X^2 P(X,t\,|\,X_0,t_0)\right). \tag{3}$$

The form of $F(X)$ suggests that it can propel the DBD2 of TF only for short distances since $\lim_{X\to\infty} F(X) = 0$ although such limit will be meaningless for $X > 2\pi^2 a$ where E_{elastic} will be close to the background thermal energy. Initial condition for **Eq. 3** will be $P(X,t_0\,|\,X_0,t_0) = \delta(X - X_0)$ where $X_0 = 2\pi r_P$ and the boundary conditions are given as follows.

$$P(L,t\,|\,X_0,t_0) = 0; \quad \left[F(X)P(X,t\,|\,X_0,t_0) - \partial_X P(X,t\,|\,X_0,t_0)\right]_{X=X_0} = 0\,. \tag{4}$$

Here X_0 acts as a reflecting boundary for a given size of TF and L is the absorbing boundary where the promoter is located. The asymmetric energy profile with respect to S1 and S2 is required for the validity of the reflecting boundary condition at X_0. Upon reaching the promoter via loop-expansion of the intervening DNA segment, TFs subsequently activate the transcription. The mean first passage time $T_B(X)$ associated with the DBD2 of TF to reach the promoter location L starting from arbitrary $X \in (X_0, L)$ obeys the following backward type Fokker-Planck equation along with the appropriate boundary conditions (6, 7).

$$D_C\left[F(X)d_X T_B(X)+d_X^2 T_B(X)\right]=-1;\ T_B(L)=0;\ d_X T_B(X_0)=0. \tag{5}$$

The integral solution of **Eqs. 5** can be expressed as follows.

$$T_B(X)=\left(2\pi^2 a/D_C\right)\int_X^L\left\{\begin{array}{l}Z/2\pi^2 a+\exp\left(2\pi^2 a/Z\right)\left[\mathrm{Ei}_1\left(2\pi^2 a/X_0\right)-\mathrm{Ei}_1\left(2\pi^2 a/Z\right)\right]\\-\left(X_0/2\pi^2 a\right)\exp\left(2\pi^2 a(X_0-Z)/X_0 Z\right)\end{array}\right\}dZ. \tag{6}$$

Here $\mathrm{Ei}_1(Y)=\int_1^\infty\left[\exp(-uY)/u\right]du$ (56) and interestingly $\lim_{L\to\infty}T_B(X)=T_N(X)$. Here $T_N(X)$ is the mean first passage time required by the DBD2 of TF to reach L via pure 1D sliding in the absence of DBD1 which is a solution of the following differential equation (6, 7, 21).

$$D_C d_X^2 T_N(X)=-1;\ T_N(L)=0;\ d_X T_N(X_0)=0;\ T_N(X)=\left(L^2-X^2\right)/2D_C - X_0(L-X)/D_C. \tag{7}$$

To obtain the target finding time, one needs to set $X = X_0$ in **Eqs. 6** and **7**. One can define the number of times the target finding rate of TF can be accelerated by the looping mediated propulsion of TF over 1D sliding as $\eta_P=\left[T_N(X_0)/T_B(X_0)\right]$ (here the subscript 'P' denotes the propulsion model) which is clearly independent of D_C of TF and solely depends on (L, a, and X_0). Explicitly one can write it as,

$$\eta_P=(L-X_0)^2\left((L^2-X_0^2)+4\pi^2 a\int_{X_0}^L\left\{\begin{array}{l}\exp\left(2\pi^2 a/Z\right)\left[\mathrm{Ei}_1\left(2\pi^2 a/X_0\right)-\mathrm{Ei}_1\left(2\pi^2 a/Z\right)\right]\\-\left(X_0/2\pi^2 a\right)\exp\left(2\pi^2 a(X_0-Z)/X_0 Z\right)\end{array}\right\}dZ\right)^{-1}. \tag{8}$$

Detailed numerical analysis (see **Section 2** of the **Supporting Material**) suggests that there exists a maximum of η_P at which $\partial\eta_P/\partial L=0$ with $L = L_{\text{opt}}$ and clearly, we have $\lim_{L\to\infty}\eta_P=1$ (**Figs. 2A** and **B**). This is logical since when $L > L_{\text{opt}}$ then $\eta_P\to 1$ and when $L < L_{\text{opt}}$ then the stored energy is not completely utilized to propel the DBD2 of TF. Further, $\lim_{L\to X_0}\eta_P=0$ since its numerator part goes to zero much faster than the denominator (**Fig. S1**). The total time required by the TFs to form a synaptosome complex via propulsion mechanism will be $\tau_P=\tau_S+T_B(X)$.

Predictions of the propulsion model

The persistence length of DNA under *in vitro* conditions is $a\sim 150$ bp and the radius of gyration for most of the eukaryotic TFs will be in the range $r_P\sim 10\text{-}15$ bp. Therefore, one can set the initial $X=2\pi r_p\sim 50\text{-}100$ bp (57, 58). Simulations (**Fig. 2A**) of the expression for η_P (**Eq. 7**) at different values of X_0 and, L from X_0 to 10^5 suggested that $L_{\text{opt}}\sim 3X_0$ (see **Figs. 2C** and **2D**). When $a\sim 150$ bp and $X_0\sim 50\text{-}100$ bp, then $L_{\text{opt}}\sim 150\text{-}300$ bp. Remarkably, this is the most probable range of the distances between the CRMs and promoters of various genes observed across several genomes (59). The efficiency of the stochastic propulsion will be maximum at L_{opt}. Although L_{opt} is not much affected by a, the maximum of η_P is positively correlated with a. This is logical since the stored elastic energy is directly proportional to the persistence length of the polymer. Remarkably,

at the optimum L_{opt} the speed of interactions between CRM-TFs complex with the promoters will be ~10-25 times faster than the normal 1D sliding.

Tethered sliding model

In this model, the tethered DBD2 of TF searches for the promoter region with intact site-specific bonding interactions at DBD1-S1. Actually, DBD2-S2 is a nonspecific type binding interaction by definition and the corresponding specific interactions occur whenever DBD2 finds the promoter region (P) and forms the site-specific DBD2-P complex. Here the tethered random walker (DBD2, which is actually tied with the DNA thread at DBD1-S1) wanders over 3D space and randomly forms nonspecific contacts with other segments of same DNA polymer analogous to the ring-closure events of intersegmental transfers. Before dissociation, there is always a possibility for the DBD2 to scan the DNA of random length for the presence of its specific site P. When the length of DNA connecting DBD1 and DBD2 is X for an arbitrary nonspecific contact of DBD2, then the potential energy barrier acting on such random scanning will be $E \simeq \left(2\pi^2 a/X\right) + \left(3/2\right)\ln\left(\pi X/6\right)$.

Interestingly, this potential energy barrier attains a minimum as $E_{\min} = \left[3/2\right]\left(1 + \ln\left(2\pi^3 a/9\right)\right)$ at $X_C = 4\pi^2 a/3$. Forward and reverse movement of such *tethered random walker* drives X to $X + 1$ or $X - 1$. Contrasting from the propulsion model, here we have not ignored the entropy component of the potential E since the interconnecting DNA segment is in free loop form. The force generated by such potential will be $F\left(X\right) = 2\pi^2 a/X^2 - 3/2X$. Upon inserting this force term in to **Eq. 5** one finally obtains the following result.

$$T_U\left(X\right) = \left(1/D_C\sqrt{2\pi a}\right)\int_X^L Z^{3/2}\exp\left(2\pi^2 a/Z\right)\left(\mathrm{Erf}\left(\sqrt{2\pi^2 a/Z}\right) - \mathrm{Erf}\left(\sqrt{2\pi^2 a/X_0}\right)\right)dZ \qquad (9)$$

Here $\mathrm{Erf}\left(v\right) = \left(2/\sqrt{\pi}\right)\int_0^v \exp\left(-z^2\right)dz$ is the error function integral (56), and $T_U\left(X\right)$ is the MFPT required by a tethered random walker to find its specific site located at L starting from X (this is the initial loop length) anywhere within (X_0, L) where X_0 is a reflecting boundary and L is an absorbing boundary. Since the potential function has a minimum at X_C, one can consider the following two different limiting regimes.

$$\lim_{X \gg X_C} T_U\left(X\right) \simeq 4\left(L^{5/2} - X^{5/2}\right)\big/5D_C\sqrt{X_0} - \left(L^2 - X^2\right)\big/D_C; \quad \lim_{X \ll X_C} T_U\left(X\right) \simeq T_B\left(X\right) \qquad (10)$$

One can define the number of times the target finding rate of TF can be accelerated by the tethered sliding of TF as $\eta_S = T_N\left(X\right)\big/T_U\left(X\right)$ (here the subscript 'S' denotes the tethered sliding model) which is clearly independent of D_C of TF and solely depends on (L, a, and X_0). Contrasting from the propulsion model, one finds that $\lim_{L\to\infty}\eta_S = 0$. In these calculations we have not included the looping mediated nonspecific association time required by the DBD2 of TF. This in fact further increases the overall MFPT of the tethered sliding model. The rate associated with the formation of the initial (nonspecific contact) loop with length X can be written as $k_{NL} \simeq k_t\exp\left(-E\right)$ where k_t (s^{-1}) is the maximum achievable rate under zero potential. Clearly, k_{NL} will be a maximum at X_C which is the most probable initial landing position of the tethered DBD2 via DNA-looping. The total time required by the CRMs-TFs system to form the synaptosome complex in this model will

9

be $\tau_{TS} = \tau_S + 1/k_{NL} + T_U(X)$ which will attain a local minimum approximately at $X = X_C$. One can also define $\eta_{NL} = k_{NL}/k_t$ which will attain the maximum value $\eta_{NL} \sim 6.7$ at X_C.

Predictions of the tethered sliding model

Tethered sliding model predicts the most probable distance of the CRMs of TFs i.e. S1 from the transcription start sites as X_C. At this distance, the rate of looping mediated synaptosome complex formation of TFs will be at maximum. Upon setting $X = X_C$ in η_S and numerically iterating L from 3000 to 10000 bp with $a \sim 150$ bp. When the left reflecting boundary was at X_0, then one finds the critical distance L_C such that $\eta_S > 1$ when $L < L_C$ and approximately $\eta_S < 1$ when $L > L_C$. Particularly when $X_0 < 100$ bp, one can define the critical distance of TSS from CRM in the tethered sliding model as $L_C \sim 3X_C$. This critical distance decreases with increase in X_0. These numerical results are demonstrated in **Fig. S3** of the Supporting Materials.

COMPUTATIONAL ANALYSIS

The core assumptions of the propulsion model are **1**) TFs have two different DNA binding domains (DBD1 and DBD2), **2**) correspondingly there should be two different binding sites (S1 and S2) in the upstream region (CRMs) of the transcription start site (TSS), **3**) out of which one that is closer to TSS should be weaker in binding strength than the one that is far away from TSS. This in turn creates the required asymmetry in the binding energy profile of TFs with the CRMs. The main prediction of the propulsion model is **4**) that $L_{opt} \sim 3X_0$ where L_{opt} is the optimum distance between the CRMs and the promoters and X_0 is the distance between the two different binding-sites of TFs (S1 and S2) within the *cis*-regulatory module. **5**) Tethered sliding model predicted the most probable distance of the CRMs (S1 corresponding to DBD1) of TFs from the transcription start site as $X_C = 4\pi^2 a/3 \sim 2000$ bp for $a \sim 150$ bp.

Datasets and analysis

To check whether such TFs-CRMs systems with properties **1**) to **5**) exist, we analyzed the upstream 5000 bp sequences of various genes of human and mouse genome. We used the position weight matrices of various transcription factors of human and mouse available with the JASPAR database and scanned upstream sequences of all the genes in the respective genome. The upstream 5000 bps sequences of various genes of human and mouse genomes were obtained from UCSC genome database (February 2009 assembly, hg19 version for human genome and December 2011 assembly, mm10 version of mouse genome) and position weight matrices (PWMs) (60, 61) of various TFs of mouse and human were obtained from the publicly available JASPAR database (62, 63). There were 21929 sequences from mouse genome and 28824 sequences from the human genome. Using the PWMs of various available TFs we generated the score table for various upstream sequences based on the following equation (60).

$$S_{v,i} = -\sum_{u=i+1}^{q} \left[\sum_{b=A}^{T} f_{b,u} \log_e \left(f_{b,u}/f_b \right) \right]; \; b = \{A, C, G, T\} \tag{11}$$

In this equation $S_{v,i}$ is the score value of PWM at i^{th} position upstream of the transcription start site on v^{th} sequence, q is the length of binding stretch of the corresponding TF, f_b is the background probability of observing base b in the corresponding genome, and $f_{b,w}$ is the probability of observing base b at position w of the specific binding sites of TFs. Here f_b was calculated from the

random sequences of the given genome available with the UCSC database. We considered only those TFs showing two different putative binding sites upstream of the promoters of various genes. This will prove our second assumption. The binding site close to TSS is S2 and the one away from TSS is S1 by definition. The distance between these sites is X_0. We also constructed the distribution of the distances of S1 and S2 from the transcription start site. There is a strong positive correlation between the score value and the binding energy of TFs (60). Therefore, the sign of the differences in the score values of these two putative binding sites of a given TF will give the information regarding the direction of the asymmetry of the binding energy profile that is required to prove our third assumption. Here the absolute distance between these binding sites will be X_0 and the distance between them from the transcription start site will be the L of our model. Checking for the relationship $L_{opt} \sim 3X_0$ will prove the forth proposition of our model. Computing the distribution of the distances of S1 from the TSS will confirm the validity of the fifth proposition predicted by the tethered sliding model.

In parallel, we also generated score table for random sequences using the same PWM from which we obtained the score distribution and the cutoff score value for the given weight matrix corresponding to a given p-value. In our calculations, we have set the p-value $< 10^{-6}$ for defining the putative specific binding sites of TFs. We used the random sequences associated with each genome that is available at UCSC database to compute the probability of occurrence of putative binding sites by chance. We considered random sequences of size 5×10^6 bps and fragmented it into 10^3 number of sequences with length of 5000 bps. Then we scanned each random sequence with the same PWM and obtained the number of putative CRMs (false positives). The probability of observing a CRM site by chance will be calculated as p_{NF} = number of false positives / 1000.

RESULTS AND DISCUSSION

The main limitation of the propulsion model is the requirement of huge energy input involved in the initial bending of DNA around the TF of interest. This needs to be derived either in the form of ATP hydrolysis or in the form of binding energy derived from the combinatorial multiprotein TFs. For example, bending of a linear DNA with size of 50-100 bp into loop requires the hydrolysis of at least 3-5 ATPs (using $E_{bend} = E_{elastic} + E_{entropy}$, 1 ATP $\sim 12\ k_BT$). Investment of such energy input is required by CRM-TF system to actively slide in a directional dependent manner towards the promoter. On the other hand, tethered sliding of TFs does not require such huge energy input since there is no restriction on the initial loop length. As a result, directional dependent movement of TFs is not possible in the tethered sliding model. However, the probability density function associated with the initial loop length will be dictated by the bending energy profile. Actually, E_{bend} will be a minimum at $X_C \simeq 4\pi^2 a/3$ where the average search time required to form the synaptosome complex will be at minimum (49). When $X < X_C$ then $E_{bend} \propto X^{-1}$. When $X > X_C$ then $E_{bend} \propto \ln(X)$. When $a \sim 150$ bp and $X_C \sim 2000$ bp then the minimum of $E_{bend} \sim 13\ k_BT$ which requires the hydrolysis of at least 1 ATP. These results are demonstrated in **Fig. 3**. Including the models presented in this paper, one can consider the following four possible modes.

1. *Propulsion mechanism*. This requires huge free energy input in the initial loop formation with a possibility of directional dependent movement of TFs towards the promoter.

2. *Tethered sliding mechanism*. This required minimal free energy input in the formation of initial loop. Although the directional dependent movement of TFs is not possible here, the free energy barrier involved in the initial loop formation stage restricts the initial landing position of DBD2 of TFs close to the promoters.
3. *Repeated binding-unbinding mode*. This mechanism is similar to tethered sliding mode with restrictions on the sliding dynamics. Here the searching for the promoters is achieved via repeated binding-unbinding of the tethered TFs. Directional dependent movement of TFs along DNA is not possible in this mode.
4. *Parallel searching of two DBDs of TFs*. Here two different DBDs of TFs (DBD1 and DBD2) search for their cognate sites on DNA (S1 and P respectively) independently through a combination of 1D and 3D diffusions. When these DBDs binds their cognate sites simultaneously, then the looping of the intervening DNA segment occurs as a result. However, this mechanism works well only for the single TF based transcription activation such as Lac I system and it is almost improbable for the combinatorial binding of TFs in the gene regulation of eukaryotes. However, this mode can be a parallel (but slow) pathway of loop formation for the above said mechanisms.

The analysis results on the upstream sequences of various genes of human and mouse are shown in **Figs. 4A** and **B**. Clearly, there are several TFs with two different putative binding sites (S1, S2) upstream of the transcription start sites. Out of these, S1 is away and S2 is close to the TSS. The distributions of the distances of S1 and S2 from the respective TSS are shown in **Fig. 5A1-2**, **B1-2**. The distributions of the distances between S1 and S2 are shown in **Figs. A3** and **B3**. The distributions of the asymmetry in the binding energy profiles of S1 and S2 are shown in **Figs. A4** and **B4**. Although our computational analysis suggested that $L \sim 3X_0$ is not a strict rule applicable to all the genes, several such CRMs-TFs systems follow the prediction of the propulsion model i.e. $L_{opt} \sim 3X_0$ where X_0 is the distance between S1 and S2, and L_{opt} is the optimum distance of S2 from the transcription start site. Although the most probable location of S2 is close to the promoter region, the most probable location of S1 seems to be around ~2500 bp away from the promoter in both mouse and human genome. This is in line with the tethered sliding model which predicted the critical distance of CRMs from the promoter to be around $X_C \sim 2000$ bp. The asymmetry in the relative binding strengths of these sites seems to be equally probable both towards as well as away from the transcription start site.

Limitations of the models
In multiprotein mediated DNA looping, there is always a possibility for two different TFs interact with S1 and P respectively and the looping is mediated via protein-protein interactions among these TFs. In both propulsion and tethered sliding models, we have assumed that the nonspecifically bound DBD2 of TF does not dissociate until reaching the promoter. Nevertheless, earlier studies suggested that this assumption is valid only for the average sliding length of TF $L_S \simeq \sqrt{2D_C/k_r}$ where k_r is the dissociation rate constant (7) that is defined as $k_r \simeq k_r^0 \exp(-\mu_{NS})$ where $k_r^0 \sim 10^6$ s^{-1} and μ_{NS} is the average nonspecific binding energy associated with DBD2 of the TF of interest. Clearly $\mu_{NS} \geq 12 \ k_BT$ is required to attain $L_S \sim 300$ bp which can be achieved via multiprotein binding.

In the absence of energy input, biological systems can overcome the looping energy barrier via three possible ways viz. **1**) multiprotein binding (38) which could be the origin of the combinatorial

TFs in the process of evolution, **2**) placing sequence mediated kinetic traps corresponding to DBD2 in between CRMs and promoters (18) and, **3**) placing nucleosomes all over the genomic DNA to decrease the $E_{entropy}$ component. All these aspects are observed in the natural systems. In multiprotein binding, the free energies associated with the DNA-protein and protein-protein interactions among TFs will be utilized in a cooperative manner for the looping of DNA. Here DBD1 and DBD2 may come from different proteins. Vilar and Saiz (38) had shown that the looping of DNA would be possible even with small concentrations of TFs when the number TFs in a combination is sufficiently large. Multiprotein binding eventually increases X_0 values. However, increasing X_0 will eventually decreases both the maximum possible acceleration of TF search dynamics and the energy barrier associated with the DNA-looping. As a result, natural systems have optimized X_0 between these two-opposing factors for maximum efficiency via manipulating the number of TFs in the combinatorial binding.

CONCLUSIONS

In summary, for the first time we have shown that DNA-loops can stochastically propel the transcription factors along DNA from their specific binding sites towards the promoters. We have shown that the source of propulsion is the elastic energy stored on the specific looped DNA-protein complex. Actually, elastic and entropic energy barriers associated with the looping of DNA shape up the distribution of distances between TF binding sites and promoters in the process of evolution. We argued that the commonly observed multiprotein binding in gene regulation might have been acquired over evolution to overcome the looping energy barrier. Presence of nucleosomes on the genomic DNA of eukaryotes is required to reduce the entropy barrier associated with the looping.

APPENDIX A

The energy component E_{entropy} that is required to compensate the chain entropy loss for a Gaussian chain can be computed as follows. Let us assume that the looping of DNA occurs when $\left|\overline{R}\right| \leq \xi$ where $\left|\overline{R}\right| \in \left(\xi, Xl_d\right)$ is the end-to-end distance vector, ξ is the minimum looping-distance (in m) and Xl_d is the maximum length of the DNA polymer. The probability density function of the vector $\overline{R}$ is $p\left(\overline{R}\right) \simeq \left(3/2\pi Xb^2\right)^{3/2} \exp\left(-3\overline{R}^2/2Xb^2\right) d\overline{R}$ (64, 65) where X is the number of monomers in the polymer and b is the average distance between the monomers. The entropy loss upon looping of DNA is $\Delta S_{\text{loop}} \simeq \ln\left(P_l/P_\Omega\right)$ (measured in k_B units) where $P_l \simeq \int_0^\xi p\left(\overline{R}\right) d\overline{R}$ is the probability of finding the loops and $P_\Omega = \int_0^{Xl_d} p\left(\overline{R}\right) d\overline{R}$ is the probability of finding all the configurations including loops. Explicitly one can write down as,

$$\Delta S_{\text{loop}} \simeq 3\ln\left(\text{Erf}\left(\sqrt{3\xi^2/2Xb^2}\right)\Big/\text{Erf}\left(\sqrt{3Xl_d^2/2b^2}\right)\right) . \tag{A1}$$

Here $\text{Erf}\left(Z\right) = \sqrt{4/\pi} \int_0^Z \exp\left(-Y^2\right) dY$ is the error function (56). When $\xi \simeq b \simeq l_d$ is very small then $\Delta S_{\text{loop}} \simeq 3\ln\left(\text{Erf}\left(\sqrt{3X/2}\right)\sqrt{6/\pi X}\right) \simeq -(3/2)\ln\left(\pi X/6\right)$ for large values of X (49). This expression for the entropy is closely linked with the Jacobson-Stockmayer factor, or J-factor associated with polymer looping (37). One finally obtains $E_{\text{bend}} \simeq 2\pi^2 a/X + (3/2)\ln\left(\pi X/6\right)$.

REFERENCES

1.   Alberts, B. 2002. Molecular biology of the cell. Garland Science, New York.
2.   Ptashne, M. 1986. A genetic switch : gene control and phage [lambda]. Cell Press ; Blackwell Scientific Publications, Cambridge, Mass.; Palo Alto, Calif.
3.   Ptashne, M., and A. Gann. 2002. Genes & signals. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
4.   Berg, O. G., R. B. Winter, and P. H. von Hippel. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry 20:6929-6948.
5.   Winter, R. B., O. G. Berg, and P. H. von Hippel. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions. Biochemistry 20:6961-6977.
6.   Murugan, R. 2004. DNA-protein interactions under random jump conditions. Phys Rev E Stat Nonlin Soft Matter Phys 69:011911.
7.   Murugan, R. 2007. Generalized theory of site-specific DNA-protein interactions. Phys Rev E Stat Nonlin Soft Matter Phys 76:011901.
8.   Niranjani, G., and R. Murugan. 2016. Generalized theory on the mechanism of site-specific DNA–protein interactions. Journal of Statistical Mechanics: Theory and Experiment 2016:053501.
9.   Koslover, E. F., M. A. Diaz de la Rosa, and A. J. Spakowitz. 2011. Theoretical and computational modeling of target-site search kinetics in vitro and in vivo. Biophys J 101:856-865.
10.   Slutsky, M., and L. A. Mirny. 2004. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. Biophys J 87:4021-4035.
11.   Wunderlich, Z., and L. A. Mirny. 2008. Spatial effects on the speed and reliability of protein-DNA search. Nucleic Acids Res 36:3570-3578.
12.   Shvets, A. A., and A. B. Kolomeisky. 2016. Crowding on DNA in Protein Search for Targets. J Phys Chem Lett 7:2502-2506.
13.   Beshnova, D. A., A. G. Cherstvy, Y. Vainshtein, and V. B. Teif. 2014. Regulation of the nucleosome repeat length in vivo by the DNA sequence, protein concentrations and long-range interactions. PLoS Comput Biol 10:e1003698.
14.   Parmar, J. J., D. Das, and R. Padinhateeri. 2016. Theoretical estimates of exposure timescales of protein binding sites on DNA regulated by nucleosome kinetics. Nucleic Acids Res 44:1630-1641.
15.   Parmar, J. J., J. F. Marko, and R. Padinhateeri. 2014. Nucleosome positioning and kinetics near transcription-start-site barriers are controlled by interplay between active remodeling and DNA sequence. Nucleic Acids Res 42:128-136.
16.   Teif, V. B., and K. Rippe. 2011. Nucleosome mediated crosstalk between transcription factors at eukaryotic enhancers. Phys Biol 8:044001.
17.   Shvets, A., M. Kochugaeva, and A. B. Kolomeisky. 2016. Role of Static and Dynamic Obstacles in the Protein Search for Targets on DNA. J Phys Chem B 120:5802-5809.
18.   Niranjani, G., and R. Murugan. 2016. Theory on the mechanism of site-specific DNA–protein interactions in the presence of traps. Physical Biology 13:046003.
19.   Lange, M., M. Kochugaeva, and A. B. Kolomeisky. 2015. Dynamics of the Protein Search for Targets on DNA in the Presence of Traps. J Phys Chem B 119:12410-12416.
20.   Kalodimos, C. G., N. Biris, A. M. Bonvin, M. M. Levandoski, M. Guennuegues, R. Boelens, and R. Kaptein. 2004. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 305:386-389.
21.   Murugan, R. 2010. Theory of site-specific DNA-protein interactions in the presence of conformational fluctuations of DNA binding domains. Biophys J 99:353-359.

22.   Zhou, H. X. 2011. Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. Proc Natl Acad Sci U S A 108:8651-8656.

23.   Ando, T., and J. Skolnick. 2014. Sliding of proteins non-specifically bound to DNA: Brownian dynamics studies with coarse-grained protein and DNA models. PLoS Comput Biol 10:e1003990.

24.   Murugan, R. 2009. Packaging effects on site-specific DNA-protein interactions. Phys Rev E Stat Nonlin Soft Matter Phys 79:061920.

25.   Khazanov, N., A. Marcovitz, and Y. Levy. 2013. Asymmetric DNA-search dynamics by symmetric dimeric proteins. Biochemistry 52:5335-5344.

26.   Marcovitz, A., and Y. Levy. 2013. Obstacles may facilitate and direct DNA search by proteins. Biophys J 104:2042-2050.

27.   Marcovitz, A., and Y. Levy. 2013. Weak frustration regulates sliding and binding kinetics on rugged protein-DNA landscapes. J Phys Chem B 117:13005-13014.

28.   Vuzman, D., Y. Hoffman, and Y. Levy. 2012. Modulating protein-DNA interactions by post-translational modifications at disordered regions. Pac Symp Biocomput:188-199.

29.   Murugan, R. 2018. Theory of Site-Specific DNA-Protein Interactions in the Presence of Nucleosome Roadblocks. Biophysical Journal 114:2516-2529.

30.   Schleif, R. 1988. DNA looping. Science 240:127-128.

31.   Schleif, R. 1992. DNA looping. Annu Rev Biochem 61:199-223.

32.   Rippe, K., M. Guthold, P. H. von Hippel, and C. Bustamante. 1997. Transcriptional activation via DNA-looping: visualization of intermediates in the activation pathway of E. coli RNA polymerase x sigma 54 holoenzyme by scanning force microscopy. J Mol Biol 270:125-138.

33.   Mulligan, P. J., Y. J. Chen, R. Phillips, and A. J. Spakowitz. 2015. Interplay of Protein Binding Interactions, DNA Mechanics, and Entropy in DNA Looping Kinetics. Biophys J 109:618-629.

34.   Lewin, R. A., D. M. Crothers, D. L. Correll, and B. E. Reimann. 1964. A Phage Infecting Saprospira Grandis. Can J Microbiol 10:75-85.

35.   Murugan, R. 2010. Theory on the mechanism of distal action of transcription factors: looping of DNA versus tracking along DNA. Journal of Physics A: Mathematical and Theoretical 43:415002.

36.   Grindley, N. D., K. L. Whiteson, and P. A. Rice. 2006. Mechanisms of site-specific recombination. Annu Rev Biochem 75:567-605.

37.   Zhang, Y., A. E. McEwen, D. M. Crothers, and S. D. Levene. 2006. Statistical-mechanical theory of DNA looping. Biophys J 90:1903-1912.

38.   Vilar, J. M., and L. Saiz. 2006. Multiprotein DNA looping. Phys Rev Lett 96:238103.

39.   Murugan, R. 2011. Theory on thermodynamic coupling of site-specific DNA–protein interactions with fluctuations in DNA-binding domains. Journal of Physics A: Mathematical and Theoretical 44:505002.

40.   Mitarai, N., I. B. Dodd, M. T. Crooks, and K. Sneppen. 2008. The generation of promoter-mediated transcriptional noise in bacteria. PLoS Comput Biol 4:e1000109.

41.   Murugan, R. 2011. Theory on the dynamic memory in the transcription-factor-mediated transcription activation. Phys Rev E Stat Nonlin Soft Matter Phys 83:041926.

42.   Elf, J., G. W. Li, and X. S. Xie. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. Science 316:1191-1194.

43.   Hammar, P., M. Wallden, D. Fange, F. Persson, O. Baltekin, G. Ullman, P. Leroy, and J. Elf. 2014. Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. Nat Genet 46:405-408.

44.   Lewin, B., J. E. Krebs, S. T. Kilpatrick, E. S. Goldstein, and B. Lewin. 2011. Lewin's genes X. Jones and Bartlett, Sudbury, Mass.

45.   Murugan, R. 2010. Theory of site-specific interactions of the combinatorial transcription factors with DNA. Journal of Physics A: Mathematical and Theoretical 43:195003.

46.   Murugan, R. 2009. Directional dependent dynamics of protein molecules on DNA. Phys Rev E Stat Nonlin Soft Matter Phys 79:041913.

47.   Zhang, Y., and D. M. Crothers. 2003. Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization. Biophys J 84:136-153.

48.   Spirin, A. S. 2009. How does a scanning ribosomal particle move along the 5'-untranslated region of eukaryotic mRNA? Brownian Ratchet model. Biochemistry 48:10688-10692.

49.   Shvets, A. A., and A. B. Kolomeisky. 2016. The Role of DNA Looping in the Search for Specific Targets on DNA by Multisite Proteins. J Phys Chem Lett 7:5022-5027.

50.   Schiessel, H., J. Widom, R. F. Bruinsma, and W. M. Gelbart. 2001. Polymer reptation and nucleosome repositioning. Phys Rev Lett 86:4414-4417.

51.   Kulic, I. M., and H. Schiessel. 2003. Nucleosome repositioning via loop formation. Biophys J 84:3197-3211.

52.   Lee, Y., A. Allison, D. Abbott, and H. E. Stanley. 2003. Minimal Brownian ratchet: an exactly solvable model. Phys Rev Lett 91:220601.

53.   Gardiner, C. W. 1985. Handbook of stochastic methods for physics, chemistry, and the natural sciences. Springer-Verlag, Berlin; New York.

54.   Risken, H. 1989. The Fokker-Planck equation : methods of solution and applications. Springer-Verlag, Berlin; New York.

55.   Kampen, N. G. v. 1981. Stochastic processes in physics and chemistry. North-Holland ; Sole distributors for the USA and Canada, Elsevier North-Holland, Amsterdam; New York; New York.

56.   Abramowitz, M., and I. A. Stegun. 1965. Handbook of mathematical functions, with formulas, graphs, and mathematical tables. Dover Publications, New York.

57.   Wingender, E., P. Dietze, H. Karas, and R. Knuppel. 1996. TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res 24:238-241.

58.   Khan, A., O. Fornes, A. Stigliani, M. Gheorghe, J. A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Cheneby, S. R. Kulkarni, G. Tan, D. Baranasic, D. J. Arenillas, A. Sandelin, K. Vandepoele, B. Lenhard, B. Ballester, W. W. Wasserman, F. Parcy, and A. Mathelier. 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res 46:D1284.

59.   Koudritsky, M., and E. Domany. 2008. Positional distribution of human transcription factor binding sites. Nucleic Acids Res 36:6795-6805.

60.   Stormo, G. D. 2000. DNA binding sites: representation and discovery. Bioinformatics 16:16-23.

61.   Kreiman, G. 2004. Identification of sparsely distributed clusters of cis-regulatory elements in sets of co-expressed genes. Nucleic Acids Research 32:2889-2900.

62.   Mathelier, A., O. Fornes, D. J. Arenillas, C. Y. Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-Hunt, A. W. Zhang, F. Parcy, B. Lenhard, A. Sandelin, and W. W. Wasserman. 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 44:D110-115.

63.   Sandelin, A., W. Alkema, P. Engstrom, W. W. Wasserman, and B. Lenhard. 2004. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32:D91-94.

64.   Doi, M., and S. F. Edwards. 1988. The theory of polymer dynamics. Clarendon Press, Oxford [Oxfordshire].

65.   Gennes, P.-G. d. 1979. Scaling concepts in polymer physics. Cornell University Press, Ithaca, N.Y.
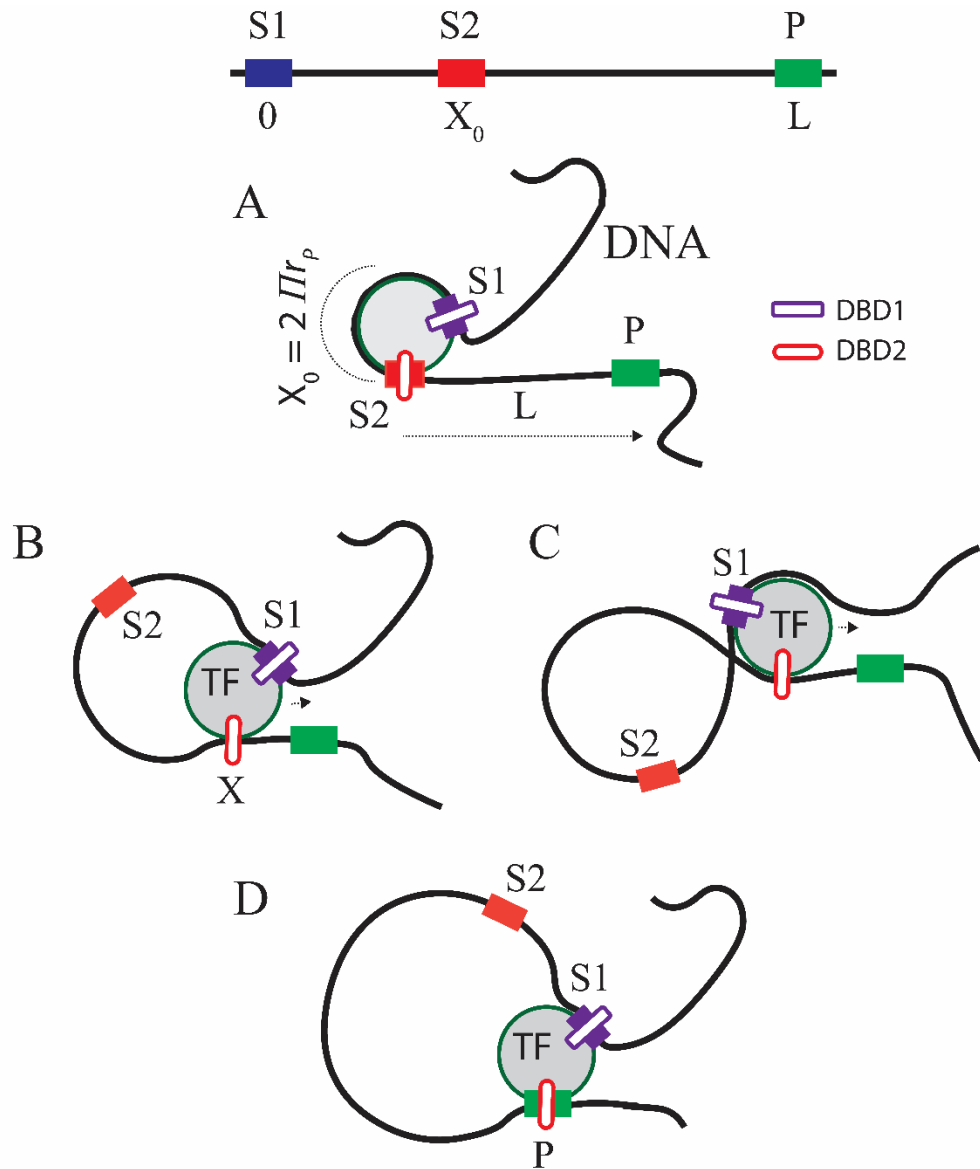
FIGURE 1. **A**. Looping mediated stochastic propulsion of TF with radius of gyration of r_P along DNA. Here TF has two binding sites corresponding to viz. its *cis*-regulatory module (DBD1) located between S1 and S2 and the promoter (DBD2). Binding of TF with its specific site (that spans for a length of X_0 from S1 ($X = 0$) to S_2 ($X = X_0$)) bends the DNA segment into a loop around it such that $X_0 = 2\pi r_P$. The bending energy stored in the site-specific complex will be incrementally released via bulging of DNA around the TF. **B**. When the binding energy near S1 is stronger than S2, then the TF can be stochastically propelled towards the promoter (P) that is located at L. Upon reaching there, DBD2 of TF interacts with the promoter to form a specific synaptic complex. **C**. DNA-loop configuration utilized for gene silencing. **D**. Synaptosome where TF is bound with both its specific binding site and the promoter via DNA-loop.
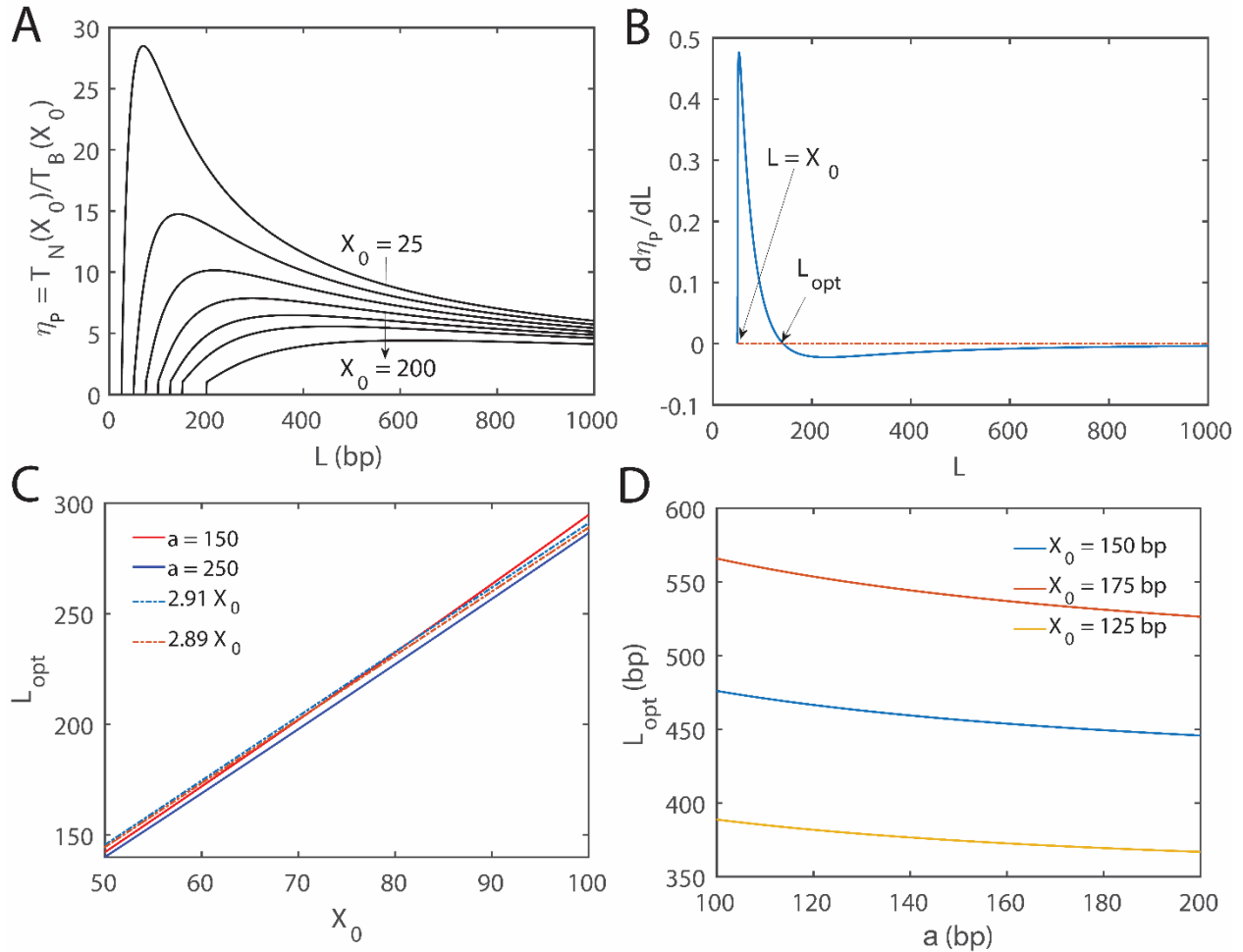
FIGURE 2. **A.** Relative efficiency of looping mediated stochastic propulsion of TFs versus normal 1D sliding along DNA. $T_N(X_0)$ is the mean first passage time that is required by TFs to reach the promoter that is located at L, starting from X_0 via 1D sliding. $T_B(X_0)$ is the mean first passage time required by TFs to reach L starting from X_0 via looping mediated stochastic propulsion mechanism. X_0 was iterated as (25, 50, 75, 100, 125, 150, 200) along the arrow while iterating L from X_0 to 1000. The efficiency of looping mediated sliding is strongly dependent on the persistence length of DNA (a), L and X_0 and it is a maximum at $L_{opt} \sim 3X_0$. **B.** Plot of $d\eta_P/dL$ with respect to L. Here the settings are $a \sim 150$ bp and $X_0 \sim 50$ bp and L was iterated from 50 to 1000 bp. Upon solving $d\eta_P/dL = 0$ for L numerically one finds that $L_{opt} \sim 142.2$ bp. **C.** Variation of L_{opt} with respect to X_0. Clearly $L_{opt} \sim 3X_0$, is slightly dependent on the persistent length a. Here we have iterated X_0 from 50 to 100 bp and $a = (150, 250)$ bp. The solution for L was searched within the interval (50, 1000) bp. **D.** Variation of L_{opt} with respect to changes in a. Here we have iterated a from 100 to 200 bp and $X_0 = (125, 150, 200)$ bp. The solution for L was searched within the interval (50, 1000) bp. The error in the approximation $L_{opt} \sim 3X_0$ seems to be < 10% over wide range of a values.
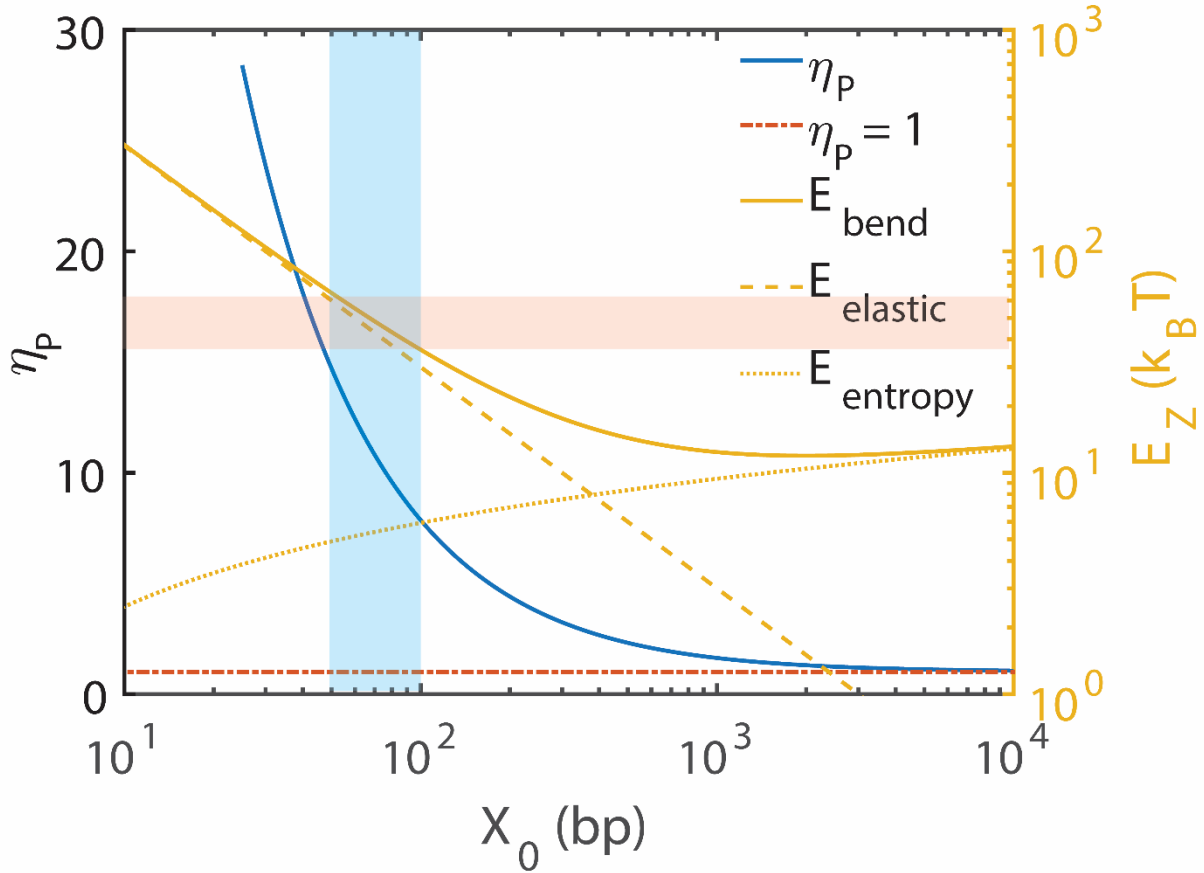
FIGURE 3. Variation of the propulsion efficiency η_P and bending energy with respect to changes in the initial loop length X_0. Here the settings are $a \sim 150$ bp, $L = 3X_0 \sim L_{opt}$. We computed $\eta_P = T_N(X_0)/T_B(X_0)$ with $L = 3X_0$ so that η_P will be close to its maximum. Here the subscript Z = (entropy, bend, elastic, enthalpy). $E_{bend} = E_{elastic} + E_{entropy}$ where $E_{elastic} \simeq 3000/X_0$ and $E_{entropy} \simeq 3\ln(\pi X_0/6)/2$ which is ~12 k_BT at $X_0 \sim 2000$ bp. $E_{elastic} \leq 1$ k_BT when $X_0 \geq 3000$ bp. Clearly, the bending energy of linear DNA is always ≥ 12 k_BT irrespective of the length. Shaded regions are the most probable X_0 values observed in the natural systems where the optimum distance between the transcription factor binding sites and promoters $L_{opt} \sim 3X_0 \sim 150\text{-}300$ bp.
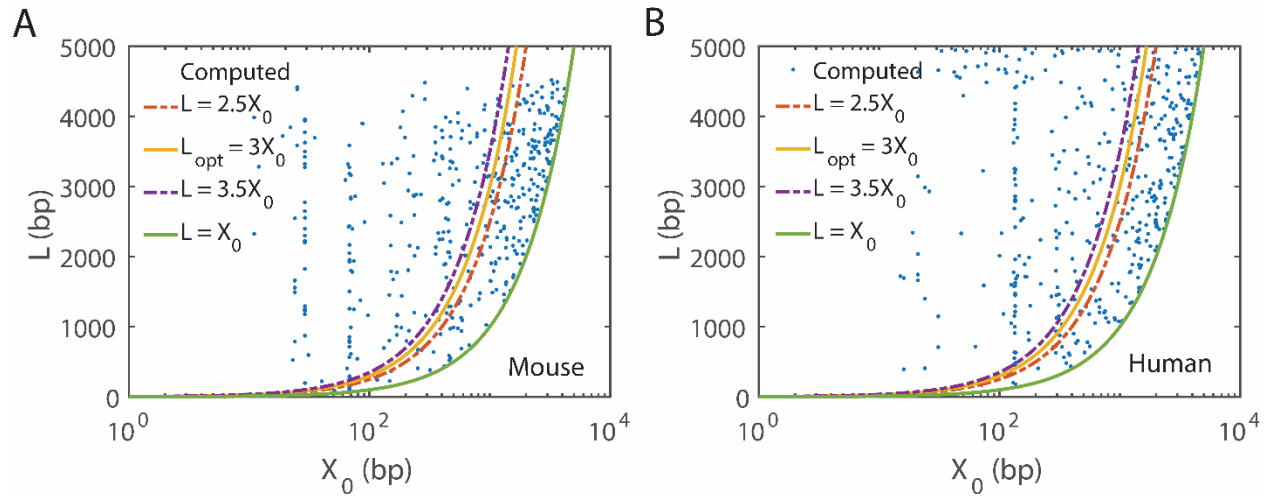
FIGURE 4. Blue dots are the computed distances between the putative *cis*-regulatory modules and L is the distance of them from the transcription start site. We considered the position weight matrices of human and mouse available with the JASPAR database and scanned the upstream 5000 bp sequences of various human and mouse genes. When there are two such putative CRMs, then the distances between them were computed along with their distances from the transcription start sites. Propulsion model suggested that $L_{opt} \sim 3X_0$. Although our computational analysis suggested that this is not a strict rule applicable to all the genes, there exist several such CRMs-TFs systems which follow the prediction of the propulsion model. **A**. Mouse. **B**. Human.
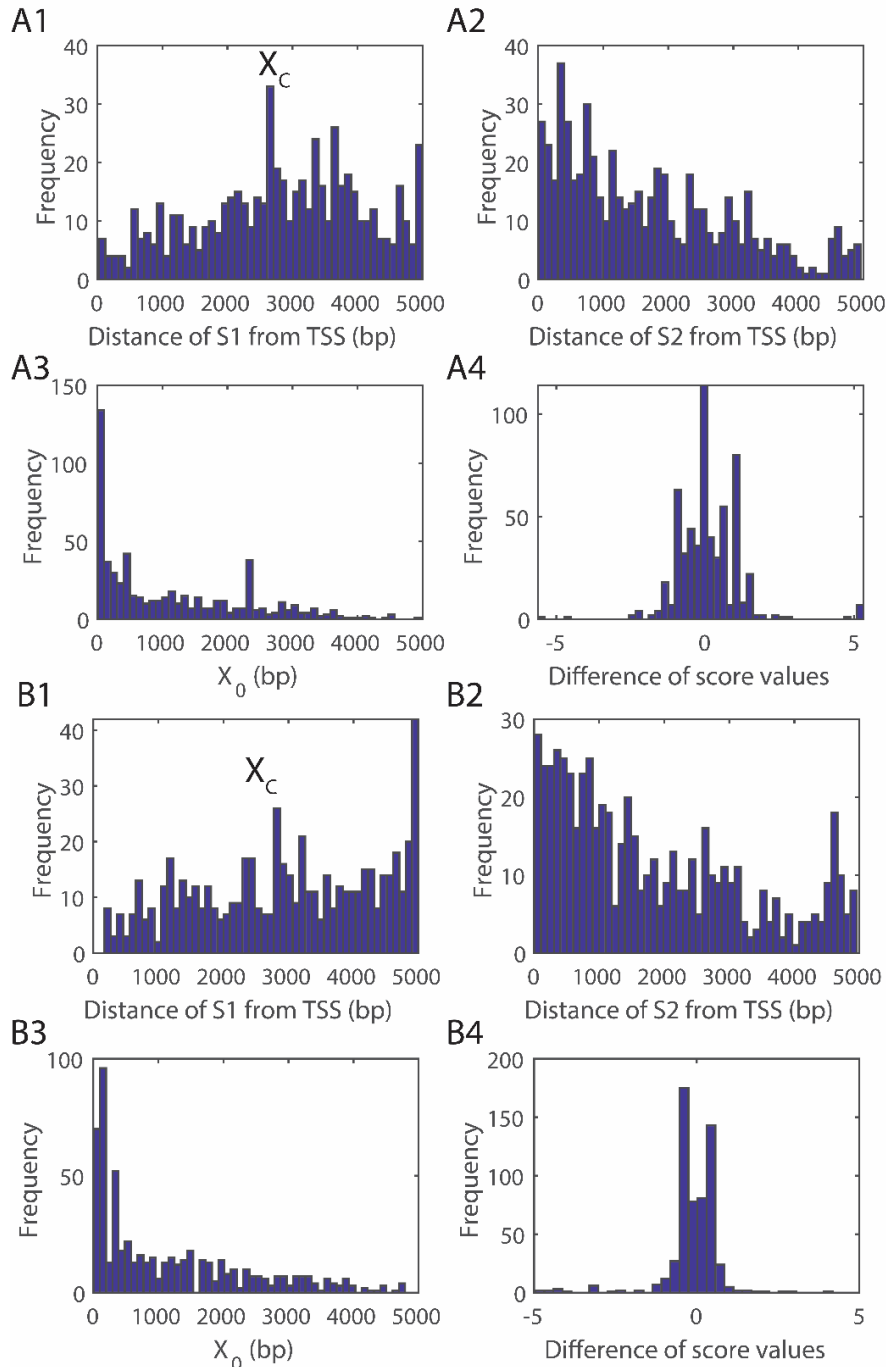
FIGURE 5. Computational data analysis results. The tethered sliding model predicted the most probable distance of the CRMs of TFs from the transcription start sites as $X_C \sim 2000$ bp for a persistence length of DNA as $a \sim 150$ bp. The distribution of the upstream location of CRMs of various TFs shows a maximum around ~2500 bp (**A1**, **B1**). Putative binding sites were defined with a p-value $< 10^{-6}$. This is in line with the tethered sliding model. Though the asymmetry in binding energy profiles of S1 and S2 is observed, there is no preferential directionality which is evident from the symmetry in the sign of the differences of the PWM score values of S1 and S2. The probability of observing a CRM site by chance will be calculated as $p_{NF} < 10^{-3}$. **A1-4**. Mouse. **B1-4**. Human.

Theory on the looping mediated stochastic propulsion of transcription factors along DNA

R. Murugan

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

Phone: +91 44 2257 4116. E-mail: rmurugan@gmail.com

Supporting Material

1. Theory of site-specific binding of TFs with CRMs

Let us consider a linear DNA of size N bps containing a *cis*-regulatory module (**CRM**) of a transcription factor (**TF**) of interest at an arbitrary location. The overall random search time or mean first passage time (**MFPT**) τ_U required by this TF to find its CRM via a combination of 3D and 1D diffusions can be written as follows (1, 2).

$$\tau_U \simeq \left[\left[P_{BTF} \right] \left(k_{fa} + k_{fn} / \left(1 + k_r \lambda \eta_U \right) \right) \right]^{-1}; \ \lambda = N/U; \ \eta_U = U^2 / 6 \left(\chi_{TF} / l_d^2 \right). \tag{S1}$$

Here, $[P_{BTF}]$ (M, mols/lit) is the concentration of TF in cytoplasm, χ_{TF} (m^2 s^{-1}) is the 1D diffusion coefficient of TFs on DNA, k_{fa} (M^{-1} s^{-1}) is the Smolochowski type bimolecular rate constant associated with the direct site-specific binding of TF with CRM via 3D diffusion, $k_{fn} \simeq k_{fa} \left(N/R_P \right)$ is the overall non-specific binding rate via 3D diffusion. In this, R_P is the radius of gyration of TF and k_r (s^{-1}) is the rate of dissociation of nonspecifically bound TF from DNA. Detailed theoretical studies suggested (1) an expression for k_{fa} as $k_{fa} \simeq k_t \left(p_R \delta / 8 \right)$. In this expression, $k_t \simeq 8 k_B T / 3\varphi$ (M^{-1} s^{-1}) is the maximum possible 3D diffusion-limited bimolecular collision rate where k_B is the Boltzmann constant, T is the absolute temperature and φ is the viscosity of the aqueous medium surrounding TFs. The numerical factor 1/8 accounts for the geometry of the random coiled and relaxed conformational state of DNA. Here (1 bps = l_d = 3.4 x 10^{-10} m). Under *in vitro* laboratory conditions (T = 298K and $\varphi \sim$ 10^{-1} kg m^{-1} s^{-1} for aqueous buffer solution), one finds that $k_t \sim$ 10^8 M^{-1}s^{-1} (3). Further, p_R is the equilibrium probability of observing a specific or nonspecific binding site to be free from other dynamic roadblocks (1, 4) which are all present on the same DNA and $\delta \simeq |\kappa| \exp \left(1.18 \omega_D \omega_P \sqrt{\Xi} \right) / R_S$ is the factor which accounts for the overall electrostatic attractive forces and the counteracting shielding effects of the solvent and other ions operating at the DNA-protein interface (1). Here $R_S \approx R_P + R_D$ is the corresponding reaction radius, R_D is the radius of the DNA cylinder and $|\kappa|$ is the **Onsager radius** *which is defined as the distance between charged reactant molecules at which the overall electrostatic energy will be the same as that of the background thermal energy.* Further ω_D and ω_P are the overall charges on the DNA backbone and

the DNA binding domains (**DBDs**) of TFs respectively and Ξ is the ionic strength of the surrounding reaction medium.

The term λ in **Eq. S1** is the minimum number of association-scan-dissociation cycles required by TFs to scan the entire DNA sequence of size N (bps) and η_U is the (averaged over initial landing positions) overall MFPT that is required to scan U bps of DNA via pure 1D diffusion. Here U (bps) is a random variable which takes different values in each of the association-scan-dissociation cycles. The probability density function associated with the 1D diffusion lengths U of TFs can be written as follows (1).

$$p(U) \simeq \left[2U \exp\left(-(U/U_A)^2\right) \middle/ U_A^2 \right]; \ U_A = \sqrt{6\chi_{TF}/l_d^2 k_r} \ . \tag{S2}$$

Here U_A is the maximum achievable 1D diffusion length of the nonspecifically bound TFs on DNA that is measured in bps. When TF moves with a jump size of k (bps) then we find that $\chi_{TF} = l_d^2 \sum_{i=-k}^{k} \left(p_{TF,i} w_{TF,i} i^2 \right) (\mathrm{m}^2 \, \mathrm{s}^{-1})$ where $i \in Z$, $w_{TF, \pm i}$ are the microscopic transition rates ($1/s$) associated with the forward and reverse movements of TFs with jump size of i, and $p_{TF, \pm i}$ are the corresponding microscopic transition probabilities (5, 6). Here the dynamics of TF with jump size k means that from the current location x (measured in m) TF can hop anywhere inside ($x - k \, l_d$, $x + k \, l_d$) with equal probabilities i.e. $1/2k$. The step length k is measured in bps. Since the dynamics at the DNA-protein interface involves segmental motion of DBDs of TFs, one can assume the protein folding rate limit (7) for $w_{TF, \pm 1} \sim 10^6 \, \mathrm{s}^{-1}$. Noting that $p_{TF, \pm 1} \sim \frac{1}{2}$ for an unbiased 1D random walk, one finds that $\chi_{TF}/l_d^2 \sim 10^6 \, \mathrm{bps}^2\mathrm{s}^{-1}$ corresponding to the sliding of TFs on DNA for which $k = 1$ bps. Approximately this is the experimental value of the 1D diffusion coefficient associated with the sliding of TFs on DNA (1, 8, 9). Using the probability density function corresponding to the sliding length, one can define the overall average search time as $\tau_S \simeq \int_0^L \tau_U p(U) dU$.

For an arbitrary jump size k with the average microscopic transition probabilities as $p_{TF, \pm i} = 1/2k$ and average transition rates $\langle w_{TF,i} \rangle = \phi_{TF}$, one finds that $\chi_{TF} = l_d^2 \phi_{TF} \left[(k+1)(2k+1)/6 \right]$. Recently, Amitai has shown (10) that $p_{TF, \pm i} \propto i^{-1/2}$ and $w_{TF, \pm i} \propto p_{TF, \pm i}$ (Eqs. 2, 4-5 of Ref. (10) where we have $|n-l| = i$ in the present context) so that $\left(p_{TF, \pm i} w_{TF, \pm i} \right) \propto i^{-1}$ for large size chromatin DNA. Upon implementing this scaling, one obtains that $\chi_{TF} \propto \left[k(k+1)/2 \right]$. Therefore, our formula for χ_{TF} with respect to k is asymptotically consistent with Ref. (10). Here the jump size k is positively correlated with the degree of condensation of DNA. Densely packed DNA allows larger jump sizes. The 3D diffusion coefficient of TFs, i.e. χ_{3D} will be always $\sim 10-10^2$ times higher than χ_{TF} and in general $\chi_{TF} \leq \chi_{3D}$ irrespective of k and presence of various 1D facilitating processes such as hopping and intersegmental transfers. Various symbols and parameters used throughout this supporting material as well as in the main text are listed in **Table S1**.

Upon mixing DNA with TF in buffer, the site-specific complex can be formed via two different pathways viz. (**a**) direct site-specific binding through 3D diffusion and (**b**) through a combination

of 1D and 3D diffusions. One can describe this by a pseudo first order scheme $DNA \xrightarrow{1/\tau_U} DNAP$. Here the concentration of specific site will be equal to the concentration of DNA, and DNAP is the site-specific complex. In τ_U of **Eq. S1**, $\left[P_{BTF}\right]k_{fa}$ is the pseudo first order rate of direct site-specific binding of TF with DNA via pure 3D diffusion and $k_{fn}\left[P_{BTF}\right]/\left(1+k_r\lambda\eta_U\right)$ is the pseudo first order rate of site-specific binding of TF via a combination of 1D and 3D diffusion.

2. Looping mediated communication between CRMs-TFs with the promoter
The mean first passage time (MFPT) associated with the DNA binding domain 2 (DBD2) of TF to reach the promoter of a gene via *DNA-loop mediated propulsion mechanism* while the DBD1 of TF complex is still tightly bound with S1 of DNA (see **Fig.1** of the main text for details) can be given as follows.

$$T_B\left(X_0\right) = \frac{L^2 - X_0^2}{2D_C} + \frac{G}{2D_C} \tag{S3}$$

Here X_0 is the initial position of DBD2 of TF on DNA or the initial loop length, L is the location of the promoter, a is the persistence length of DNA and D_C is the one-dimensional diffusion coefficient associated with the sliding of DBD2 of TF along DNA. We have assumed here that the site-specific DBD1-S1 is strong and intact (**Fig. 1A** of the main text). The function G can be defined as follows.

$$G = 4\pi^2 a \int_{X_0}^{L} \left(\exp\left(\frac{2\pi^2 a}{Z}\right) \left[\text{Ei}_1\left(\frac{2\pi^2 a}{X_0}\right) - \text{Ei}_1\left(\frac{2\pi^2 a}{Z}\right) \right] - \left(\frac{X_0}{2\pi^2 a}\right) \exp\left(\frac{2\pi^2 a\left(X_0 - Z\right)}{X_0 Z}\right) \right) dZ \tag{S4}$$

Here $\text{Ei}_1\left(Y\right) = \int_1^\infty \left[\exp\left(-sY\right)/s\right] ds$ is the E_1 exponential integral (11).

Noting that $T_N\left(X_0\right) = \left(L - X_0\right)^2/2D_C$ (from **Eq. 7** of the main text) which is the MFPT associated with the finding of the promoter by TF starting from X_0 via pure sliding dynamics, one can define η which is the number of times the DNA-loop driven searching of TF for the promoter is faster than the normal 1D sliding dynamics as follows.

$$\eta_P = \frac{T_N\left(X_0\right)}{T_B\left(X_0\right)} = \frac{\left(L - X_0\right)^2}{L^2 - X_0^2 + G} = \frac{1}{\left(L + X_0\right)/\left(L - X_0\right) + G/\left(L - X_0\right)^2} \ . \tag{S5}$$

Clearly η_P is not dependent on D_C and it depends only on the parameters (L, X_0 and a). Further, $\lim_{L \to X_0} \eta_P = 0$ since $T_N\left(X_0\right)$ approaches zero much faster than $T_B\left(X_0\right)$ as L tends towards infinity (see **Fig. S1** for details). There also exists an asymptotic limit as $\lim_{L \to \infty} \eta_P = 1$. This means that $\lim_{L \to \infty} \left[G/\left(L - X_0\right)^2\right] = 0$ (see **Fig. S2** for details). The optimum distance between CRMs and promoter i.e. L_{opt} at which η is a maximum can be obtained by solving $d\eta/dL = 0$ for L for given a and X_0. Explicitly one can write down this as follows.

3

$$\frac{d\eta_P}{dL} = \frac{2(L-X_0)\left[-\frac{(L-X_0)}{2}\frac{\partial G}{\partial L} + G + X_0(L-X_0)\right]}{\left(G + L^2 - X_0^2\right)^2} = 0.$$ (S6)

This has a trivial solution $L = X_0$. Upon ignoring this one, L_{opt} can be obtained by numerically solving the following equation for L at given a and X_0.

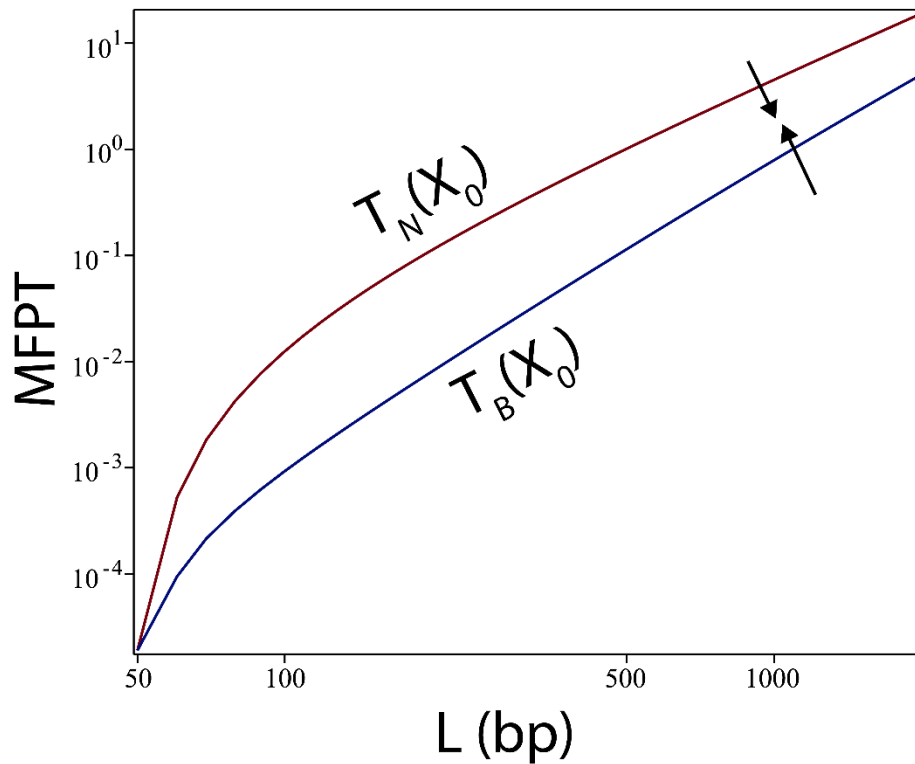$$L > X_0; \quad \therefore (L-X_0)\frac{\partial G}{\partial L} - 2G - 2X_0(L-X_0) = 0.$$ (S7)

FIG S1. Showing that $T_N(X_0)$ approaches zero much faster than $T_B(X_0)$ so that $\lim_{L \to X_0} \eta_P = 0$. Here the settings are $X_0 \sim 50$ bp and $a \sim 150$ bp and L was iterated from 50 to 2000 bp. Further we also find the limit $\lim_{L \to \infty} T_B(X_0) = T_N(X_0)$ or $\lim_{L \to \infty}\left[T_N(X_0)/T_B(X_0)\right] = 1$.
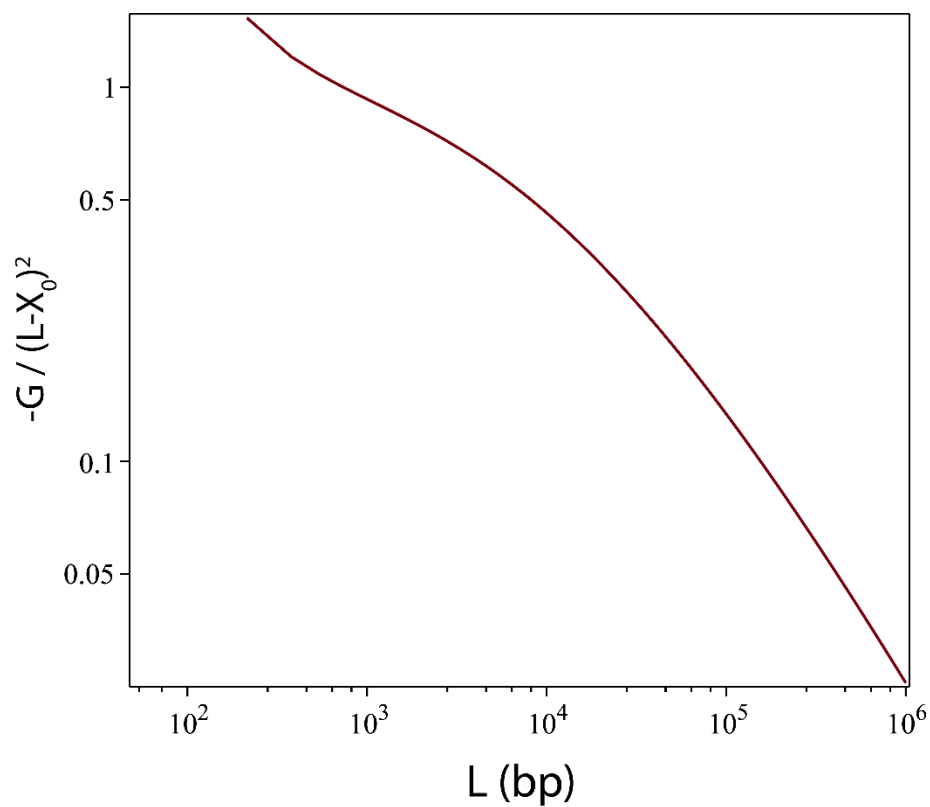
FIG S2. Showing $\lim_{L\to\infty}\left[G/\left(L-X_0\right)^2\right]\to 0$. Here the settings are $X_0 \sim 50$ bp and $a \sim 150$ bp and L was iterated from 50 to 10^6. The function G is defined as in **Eq. S4**.
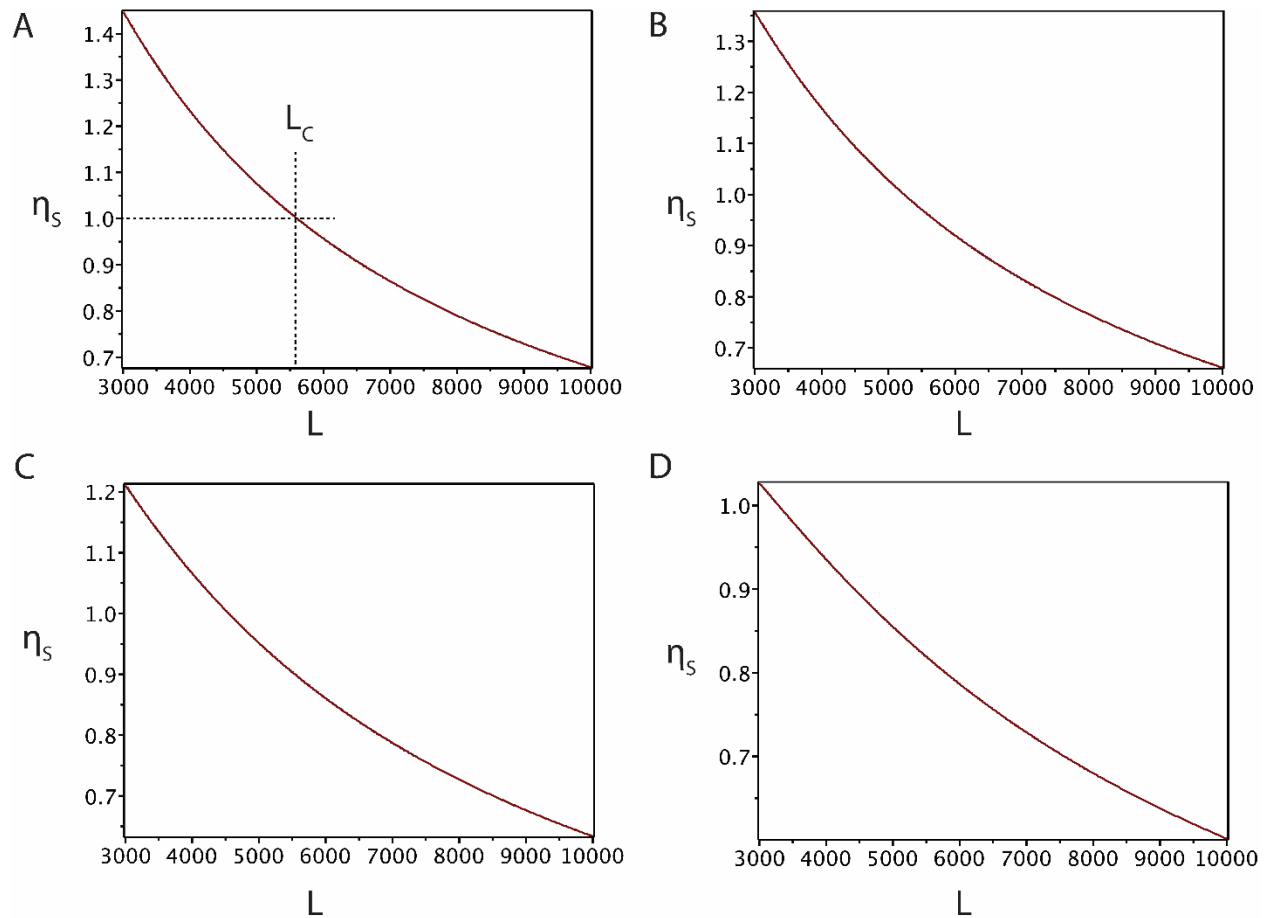
FIG S3. Variation of the critical distance (L_C) between CRMs and promoter in the tethered sliding model. Here $\eta_S = T_N(X)/T_U(X)$ as defined in **Eqs. 7** and **9**. In these MFPTs, X is the initial loop length, X_0 is the left side reflecting boundary and L is the right-side absorbing boundary. L_C is defined such that when $L < L_C$ then $\eta_S > 1$ and when $L > L_C$ then $\eta_S < 1$. Settings are (see **Eq. 9** of the main text), $X = X_C$ where $X_C = 4\pi^2 a/3$, persistence length of DNA $a \sim 150$ bp and L was iterated from 3000 to 10000 bp. **A**. $X_0 = 100$ bp. **B**. $X_0 = 250$ bp. **C**. $X_0 = 500$ bp. **D**. $X_0 = 1000$ bp. These results suggest that L_C decreases as X_0 increases. For the most probable value of $X_0 \sim 100$ bp one finds that $L_C \sim 3X_C$.

TABLE S1. List of symbols used in the main text

Symbol	Definition	Remarks
TF	Transcription factor	
CRM	*Cis*-regulatory module, specific sequence of DNA where TFs bind and then distally act on the promoters to initiate transcription.	
DBD DBD1 DBD2	DNA Binding Domain of TF proteins; DBD1 of the TF complex specifically binds with CRM; DBD2 of the TF complex specifically interacts with the promoter of the gene that needs to be activated.	
TSS	Transcription Start Site.	
MFPT	Mean First Passage Time.	s
1D, 3D	One-dimensional, three-dimensional.	
l_d	~3.4 x 10^{-10} m, equals to 1 base-pair (bp). We measure all the length parameters in bp.	bp
r_P	Radius of gyration of the TF complex.	bp
X	Length of the DNA-loop at a given time point.	bp
L	Position of the promoter.	bp
X_0	Initial DNA-loop length. It is assumed such that $X_0 \sim 2\pi r_P$ where r_P is the radius of gyration of the TF complex.	bp
S1	DNA binding sequence (CRM) corresponding to DBD1, which is a specific binding site.	
S2	DNA binding sequence present at the location of DBD2 of TFs complex, which is not a specific binding site for DBD2.	
P	Promoter sequence. This is the specific site corresponding to DBD2 by definition.	bp
a	Persistence length DNA. Typically, it is ~150 bp.	bp
E	Energy stored on the site-specific DNA-TF complex $E \sim E_{\text{bond}} + E_{\text{elastic}}$.	$k_B T$
$F(X)$	It is force generated by the potential E in the propulsion model. $F(X) = -dE/dX = 2\pi^2 a / X^2$. This in turn can propel TF towards the promoter. In case of tethered sliding model, $F(X) = 2\pi^2 a / X^2 - 3/2X$.	$k_B T$ / bp
E_{bind}	$\sim E_{\text{bond}} + E_{\text{elastic}} + E_{\text{entropy}}$, the total site-specific binding energy associated with the DNA-TF complex dissipates into these three components.	$k_B T$
E_{entropy}	Energy required to compensate the chain entropy loss of linear DNA that is converted into a loop upon binding of TF, $E_{\text{entropy}} \simeq (3/2)\ln(\pi X/6)$ where X is the size of DNA that is converted into a loop upon binding of TF.	$k_B T$
E_{elastic}	Elastic energy involved in the bending of a linear segment of DNA, $E_{\text{elastic}} \simeq aX / 2r_P^2$, where a is the persistence length of DNA, X is the length of DNA segment and r_P is the radius of curvature after bending (this is equal to the radius of gyration of the TF	$k_B T$

	complex). Elastic energy required to convert a DNA segment of size X into a loop such that $X = 2\pi r_P$ will be $E_{\text{elastic}} \simeq 2\pi^2 a/X$.	
E_{bend}	$\sim E_{\text{elastic}} + E_{\text{entropy}}$, energy required to bend a linear DNA segment into a loop. $E_{\text{bend}} \simeq 2\pi^2 a/X + (3/2)\ln(\pi X/6)$. This has a minimum with respect to X as $E_{\text{bend}}^{\min} = [3/2](1 + \ln(2\pi^3 a/9))$ at $X_C \simeq 4\pi^2 a/3$.	$k_B T$
X_C	Critical length of DNA at which the bending energy will be at minimum. Particularly, $X_C \simeq 4\pi^2 a/3$ which can be obtained by solving $dE_{\text{bend}}/dX = 0$ for X. When $X < X_C$ then $E_{\text{bend}} \propto X^{-1}$. When $X > X_C$ then $E_{\text{bend}} \propto \ln(X)$.	bp
E_{bond}	Resultant energy with respect to all types of bonding interactions present at the DNA-TF interface.	$k_B T$
D_C	One-dimensional diffusion coefficient associated with the sliding of TFs on DNA.	bp^2 s^{-1}
$T_B(X_0)$	MFPT required by TF to reach the promoter starting from its CRM (X_0) via DNA-loop mediated propulsion mechanism. $T_B(X_0) = (L^2 - X_0^2)/2D_C + G/2D_C$, where the function G is defined as in **Eq. S3**. Here X_0 is the initial loop length as well as the left side reflecting boundary and L is the right-side absorbing boundary.	s
$T_N(X_0)$	MFPT required by TF to reach the promoter starting from its CRM (X_0) via pure 1D sliding mechanism. For an arbitrary starting position X, we have $T_N(X) = (L^2 - X^2)/2D_C - X_0(L - X)/D_C$. Here $X = X_0$ is the reflecting boundary and $X = L$ is the absorbing boundary.	s
η_P	$\eta_P = [T_N(X_0)/T_B(X_0)]$. Here the subscript 'N' denotes normal sliding and 'B' denotes the propulsion model.	dimensionless
$T_U(X)$	MFPT required by TF to reach the promoter starting via tethered sliding mechanism. Here initial loop length is X, with left reflecting barrier at X_0 and right absorbing barrier at L.	s
η_S	$\eta_S = [T_N(X)/T_U(X)]$. Here the subscript 'N' denotes normal sliding and 'U' denoted the tethered sliding model.	dimensionless
L_C	In the tethered sliding model, when the left reflecting boundary was at X_0, then one finds the critical distance between CRMs and promoter L_C such that $\eta_S > 1$ when $L < L_C$ then $\eta_S < 1$ when $L > L_C$. When $X_0 < 100$ bp, then one finds that $L_C \sim 3X_C$.	bp
L_{opt}	Distance between CRM and promoter at which the ratio η_P is at a maximum. Approximately, $L_{\text{opt}} \sim 3X_0$	bp
ΔS_{loop}	Entropy change upon looping of DNA, $\Delta S_{\text{loop}} \simeq \ln(P_l/P_{all})$ where P_l is the probability of finding looped conformations of DNA and	k_B

	P_{all} is the probability of finding all the possible conformations (equal to one).	
k_r	$k_r \simeq k_r^0 \exp\left(-\mu_{NS}\right)$, dissociation rate constant related to the nonspecific DNA-TF complex where μ_{NS} is the nonspecific binding energy barrier measured in $k_B T$ and k_r^0 is the dissociation rate at zero energy barrier.	s^{-1}
k_{NL}	Rate of loop formation with nonspecific contact $k_{NL} \simeq k_t \exp\left(-E\right)$ in tethered sliding model. Here $k_t \simeq k_{on}\xi$ where k_{on} (s^{-1} M^{-1}) is the Smolochowski type 3D diffusion-controlled bimolecular collision rate limit and ξ is the concentration (mol/lit, M) of DBD2 under *in vivo* conditions. and E is the potential associated with the loop formation.	s^{-1}
L_S	$L_S \simeq \sqrt{2D_C/k_r}$, Average sliding length of nonspecifically bound TF on DNA before it dissociates where k_r is the dissociation rate constant.	bp

REFERENCES

1.	Niranjani, G., and R. Murugan. 2016. Generalized theory on the mechanism of site-specific DNA–protein interactions. Journal of Statistical Mechanics: Theory and Experiment 2016:053501.
2.	Murugan, R. 2018. Theory of Site-Specific DNA-Protein Interactions in the Presence of Nucleosome Roadblocks. Biophysical Journal 114:2516-2529.
3.	Atkins, P. W. 1978. Physical chemistry. W.H. Freeman, San Francisco.
4.	Niranjani, G., and R. Murugan. 2016. Theory on the mechanism of site-specific DNA–protein interactions in the presence of traps. Physical Biology 13:046003.
5.	Gardiner, C. W. 1985. Handbook of stochastic methods for physics, chemistry, and the natural sciences. Springer-Verlag, Berlin; New York.
6.	Murugan, R. 2007. Generalized theory of site-specific DNA-protein interactions. Phys Rev E Stat Nonlin Soft Matter Phys 76:011901.
7.	Yang, W. Y., and M. Gruebele. 2003. Folding at the speed limit. Nature 423:193-197.
8.	Hammar, P., M. Wallden, D. Fange, F. Persson, O. Baltekin, G. Ullman, P. Leroy, and J. Elf. 2014. Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. Nat Genet 46:405-408.
9.	Li, G.-W., O. G. Berg, and J. Elf. 2009. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. Nat Phys 5:294-297.
10.	Amitai, A. 2018. Chromatin Configuration Affects the Dynamics and Distribution of a Transiently Interacting Protein. Biophys J 114:766-771.
11.	Abramowitz, M., and I. A. Stegun. 1965. Handbook of mathematical functions, with formulas, graphs, and mathematical tables. Dover Publications, New York.