Directional dependent dynamics of protein molecules on DNA

Rajamanickam Murugan*
Department of Biotechnology, Indian Institute of Technology Madras (IITM), Chennai, Tamil Nadu 600 036, India
(Received 5 September 2008; revised manuscript received 11 February 2009; published 14 April 2009)

We demonstrate that a protein molecule of interest undergoing the one-dimensional Brownian dynamics along DNA can exhibit a directional dependent net transport either toward or away from its target site depending on the distribution of the initial positions of the other classes of protein molecules present on the same DNA. Directionality arises as a consequence of the confinement of the search space and dynamic reflections by other protein molecules present on the same DNA chain. Energy cost for such directionality comes from the free energy spent on setting the initial positions of the other protein molecules. In the mechanism of action of cis-acting elements on the initiation of transcription, such free-energy inputs are derived from the site-specific binding affinities of the inflowing transcriptional factors toward their cis-acting elements. If the initial distribution of other protein molecules is a random one, then the protein molecule of interest exhibits a net transport away from its target site. This directionality originates from unequal natures of enhancing and retarding effects of the randomly distributed other classes of protein molecules. The protein molecule of interest overcomes the retarding effects of the other classes of protein molecules in a dynamical manner by increasing the number of dissociation-association events when it is far away from its target site and then by switching back to the sliding dynamics due to increase in the enhancing effects as it moves closer to its target site.

DOI: 10.1103/PhysRevE.79.041913 PACS number(s): 87.10.−e, 87.15.kj

I. INTRODUCTION

Vital processes of molecular biology such as transcription of a gene and replication of the genome generally start with the interactions of a protein molecule with its specific site on the genomic DNA [1]. Transcription starts with the recognition of the promoter sequences by RNA polymerase (RNAP in prokaryotes and RNAPII in eukaryotes). DNA replication starts with the recognition of the origin of replication by DNA polymerase. It was believed that the site-specific interactions of a protein molecule with DNA are single-step three-dimensional (3D) diffusion-controlled bimolecular rate processes. Detailed experimental studies [2] on the interaction of the *lac* repressor with its operator sequence showed a bimolecular association rate of $\sim 10^{10}$ $\mathrm{mol}^{-1}\ \mathrm{s}^{-1}$. This rate is $\sim 10^2$ times faster than that of the three-dimensional diffusion-controlled collision rate of $\sim 10^8$ $\mathrm{mol}^{-1}\ \mathrm{s}^{-1}$ in aqueous medium. Such higher association rates could originate from the searching of protein molecules for their specific sites on DNA in a reduced dimensional space [3]. Effective mechanism by which a protein molecule can locate its specific site on DNA is mainly via the combination of 3D and one-dimensional (1D) routes [4–6]. Optimum combination of the searching times spent by the protein molecule of interest on both these routes to get an overall minimum target-finding time has been shown to be 50:50 [7] in case of prokaryotes. Prokaryotic genomes are also shown to be designed [8] in such as way that the genes corresponding to the DNA-binding proteins (DBPs) and their respective specific sites on the genomic DNA are generally colocalized. In such colocalized conditions, no additional driving factors are required to ensure faster searching within the physiological time scales.

If the length of the DNA is N base pairs (bps), then the maximum achievable bimolecular association rate for the nonspecific DNA-protein interactions via three-dimensional routes is $\sim N \times 10^8$ $\mathrm{mol}^{-1}\ \mathrm{s}^{-1}$. Subsequent one-dimensional searching by the protein molecule of interest for its specific site on the DNA chain is the rate-limiting one that decides how many times by which the overall site-specific association rate can be higher than that of the three-dimensional diffusion-controlled rate. If the protein molecule of interest does one-dimensional scan with unit base-pair step size (unbiased sliding walk), then the search time τ_o that is required to locate its specific site on DNA scales with the size of DNA as $\tau_o \propto N^2$. This type of searching is not an efficient one. Efficiency of such one-dimensional scanning could be enhanced [4–6] by various facilitating processes such as sliding or correlated walk, hopping, and intersegmental transfers via ring-closure events occurring between distal segments of the same DNA chain [4–6]. These facilitating processes eventually increase the one-dimensional diffusion coefficient, which in turn decreases the time that is required by the protein molecule to locate its specific site on DNA. When a sequence dependent free-energy correlation exists [4,5] along DNA that drives the protein molecule of interest toward its specific site or the protein molecule searches for its specific site on DNA via random jumps with critical jump sizes k_c that scale with N as $k_c \propto N^{2/3}$ [9], then the inefficient scaling law $\tau_o \propto N^2$ transforms to an efficient one as $\tau_o \propto N$. No such sequence dependent free-energy correlation along DNA that drives protein molecule toward its specific site or protein dynamics with critical jump sizes on DNA chain has been observed so far.

Eukaryotic systems are more complicated since their genome is comparatively more complex than prokaryotes and also larger in size. Cis-acting regulatory modules or elements (CRMs) are required for stabilizing RNAPII-promoter complex in eukaryotes. Mechanism of action of these enhancers is not yet understood clearly. According to the currently ac-

*rmurugan@gmail.com;
FAX: ++91–2257–4102.

cepted picture [1,10,11], transcription factors (TFs) first bind with these cis-acting elements to form a complex [enhancer–transcription factor (ETF) complex], which then stabilizes or interacts with the preinitiation complex (PIC) (a complex that is formed when RNAPII interacts with the promoter sequence) via distal action that subsequently results in the initiation of transcription. Mode of this distal action of ETF on PIC could be either by the one-dimensional tracking or jumping [11] of ETF along DNA toward PIC or by the three-dimensional looping out of intervening DNA segment present between ETF and PIC [10]. In both the tracking or jumping and looping models, it is implicitly assumed that there is a directional dependent net transport of ETF preferentially toward PIC. Unlike the consumption of adenosine triphosphates (ATPs) during the directed motion [12] of myosin on actin filaments, there is no free energy involved in the one-dimensional diffusion of ETF along DNA toward PIC. If one argues that the dynamics of ETF preferentially toward PIC is similar to that of a Brownian ratchet with net transport, then it has been shown earlier that [13] there is a necessity for the presence of a Langevin force directed toward PIC with nonzero correlation time and broken symmetry. These conditions further ensure that the origin of such directional preferences is in accordance with the second law of thermodynamics. So far no such time-correlated Langevin forces have been observed along DNA sequence.

Under cellular conditions, the dynamical trajectory of the protein molecule of interest will be often hindered by the presence of other/similar type of protein molecules on the same DNA chain. Recently effects of the presence of similar classes of protein molecules on the same DNA on the mean first passage time (MFPT) associated with the finding of specific site by the protein molecule of interest have been studied using continuum approximation methods [14]. However enhancing or retarding effect of the other classes of protein molecules present on the same DNA has not been addressed in detail so far.

Motivation for this paper mainly comes from the following biological facts. For an efficient binding of DBPs at their specific sites on DNA, apparently free-energy correlation or time-correlated Langevin forces along the DNA chain is required and the presence of other protein molecules on the same DNA chain also must be considered. Experimentally no such free-energy correlations or time-correlated Langevin forces with broken symmetry have been observed along DNA so far. This fact is reasonable since generation of such free-energy correlations or time-correlated Langevin forces along DNA must involve an external free-energy input and DBPs are not generally known for hydrolyzing ATPs to actively search for their specific sites on DNA. Only available sources of free energy are the specific/nonspecific binding affinities of other classes of protein molecules toward their binding sites on DNA. It is an intriguing question as to whether binding affinities of other classes of protein molecules toward their respective binding sites on DNA can be effectively used to trigger directional dependent net transport of protein molecule of interest along DNA. In this paper we answer this question in detail using a combination of theoretical and simulation tools. Other aim of this paper is to understand how the protein molecule of interest overcomes the retarding effects of other classes of protein molecules present on the same DNA chain to reach its target site in physiologically reasonable time scales. These studies are very important to further understand the *in vivo* mechanisms of many subcellular processes such as binding of transcription factors to their cis-acting elements.

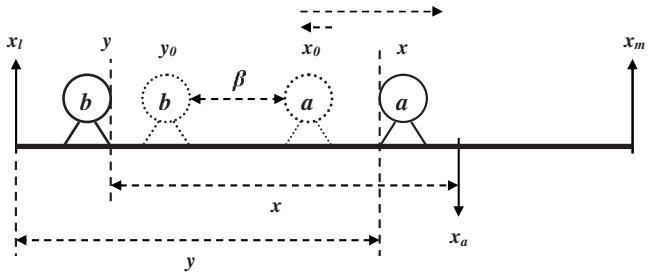

FIG. 1. Various boundary conditions described in the text. In the absence of b molecules, helical ends of DNA $\{x_l,x_m\}$ act as reflecting boundaries for a molecule whose initial position was at x_0 and current position is at x and its specific site is located at x_a. If there is another passive b molecule that does not have binding sites inside (x_l,x_m) and whose initial position y_0 was actively set such that $x_l < y_0 < x_0 < x_m$ and its current position is at y, then a molecule will have new reflecting boundaries at $\{y,x_m\}$ and for b molecule the new reflecting boundaries are at $\{x_l,x\}$ since both molecules a and b cannot occupy the same DNA site due to excluding effect. Confinement of a molecule inside the interval $\{y,x_m\}$ and dynamic reflections at the boundary y reduce the mean first passage time associated with a molecule to escape through the specific site x_a. As a result a molecule exhibits a net transport preferentially toward x_a on the cost of free energy spent on actively setting the initial position of b molecule.

II. THEORY

Consider a linear DNA whose helical ends are $\{x_l,x_m\}$. A three-dimensionally diffusing protein a molecule was nonspecifically bound at the position x_0 at time $t=0$ and is performing a one-dimensional random search along the DNA for its specific site which is located at x_a with unit base-pair step size (tracking or sliding) [4–6]. Current position of a molecule on DNA is x. Initial position of protein molecule on DNA was such that $x_l < x_0 < x_a < x_m$ (Fig. 1). We assume that the helical ends of DNA are the reflecting boundaries. This means that whenever protein a molecule hits these sites, it will be reflected back into the interval (x_l,x_m) and x_a is the only absorbing boundary. Dynamics of such protein molecule can be well described by Langevin equation as $d_t x = \sqrt{x_d}\xi_{x,t}$, where $\xi_{x,t}$ is Gaussian white-noise process with mean as $\langle \xi_{x,t} \rangle = 0$ and variance as $\langle \xi_{x,t}\xi_{x,t'} \rangle = \delta(t-t')$. Corresponding forward-type Fokker-Planck equation (FPE) can be given [15] as $\partial_t P(x,t|x_0,0) = 2^{-1}x_d \partial_x^2 P(x,t|x_0,0)$, where $P(x,t|x_0,0)$ is the probability of observing protein a molecule at position x at time t starting from x_0 at $t=0$ and x_d is the one-dimensional phenomenological diffusion coefficient associated with dynamics of protein a molecule along DNA. Initial condition for this Fokker-Planck-type equation is $P(x,0|x_0,0) = \delta(x-x_0)$ and the boundary conditions are $[\partial_x P]_{x=x_l} = [P]_{x=x_a} = 0$. Using the backward-type Fokker-

Planck equation $2^{-1}x_d d_x^2 T_a(x)=-1$ with boundary conditions $[d_x T_a]_{x=x_l}=[T_a]_{x=x_a}=0$, the MFPT associated with a molecule to locate its specific site starting from position x_0 on DNA can be given [15,16] as $T_a(x_0)=x_d^{-1}[(x_a^2-x_0^2)+2x_l(x_0-x_a)]$. One can also derive the expression for MFPT when the initial position x_0 is such that $x_l<x_a<x_0<x_m$. Here the boundary conditions are $[d_x T_a]_{x=x_m}=[T_a]_{x=x_a}=0$ and the MFPT is $T_a(x_0)=x_d^{-1}[(x_a^2-x_0^2)+2x_m(x_0-x_a)]$.

Let us assume that another protein b molecule was present on the same DNA whose initial position y_0 at time $t=0$ was actively set such that $x_l<y_0<x_0<x_m$. Its current position is at y. External free-energy sources are required for actively setting y_0 such that $y_0<x_0$ where the specific/nonspecific binding affinity of b molecule toward DNA is one of such sources. New reflecting boundaries for a molecule will be at $\{y,x_m\}$ and for the b molecule the new reflecting boundaries are at $\{x_l,x\}$ since both the molecules a and b cannot occupy the same DNA site due to excluding effect (Fig. 1). We further assume that this b molecule has no binding sites inside the interval (x_l,x_m). One-dimensional Brownian dynamics of both these molecules on the same DNA chain can be well described by a set of coupled Langevin-type equations [16] as $\{d_t x=\sqrt{x_d}\xi_{x,t}, d_t y=\sqrt{y_d}\xi_{y,t}\}$, where $\{\xi_{x,t},\xi_{y,t}\}$ are delta-correlated Gaussian white-noise processes with means as $\{\langle\xi_{x,t}\rangle,\langle\xi_{y,t}\rangle\}=\{0,0\}$ and variances as $\{\langle\xi_{x,t}\xi_{x,t}\rangle,\langle\xi_{y,t}\xi_{y,t}\rangle,\langle\xi_{x,t}\xi_{y,t}\rangle\}=\{1,1,0\}$. The corresponding forward-type Fokker-Planck equation associated with temporal evolution of the probability $P(x,y,t|x_0,y_0,0)$ of locating a molecule and b molecule at DNA-binding positions $\{x,y\}$ at time t, starting from $\{x_0,y_0\}$ at time $t=0$, can be written [15,16] as follows:

$$\partial_t P(x,y,t|x_0,y_0,0) = 2^{-1}[x_d\partial_x^2 P(x,y,t|x_0,y_0,0) + y_d\partial_y^2 P(x,y,t|x_0,y_0,0)]. \tag{1}$$

Here y_d is the one-dimensional phenomenological diffusion coefficient associated with the dynamics of b molecule on DNA chain. The initial and boundary conditions for Eq. (1) are as follows:

$$P(x,y,0|x_0,y_0,0) = \delta(x-x_0)\delta(y-y_0),$$

$$[\partial_x P]_{x=x_l} = [\partial_x P]_{x=y} = [P]_{x=x_a,y>x_a} = [\partial_y P]_{y=x} = [\partial_y P]_{y=x_m} = 0. \tag{2}$$

The MFPT $T_a(x_0,y_0)$ associated with the a molecule starting from the position x_0 to reach x_a in the presence of one b molecule inside the interval (x_l,x_0) obeys the following backward-type Fokker-Planck equation with similar boundary conditions [15] as given in Eq. (2):

$$2^{-1}[x_d\partial_x^2 T_a(x,y) + y_d\partial_y^2 T_a(x,y)] = -1,$$

$$[\partial_x T_a]_{x=x_l} = [\partial_x T_a]_{x=y} = [T_a]_{x=x_a,y>x_a} = [\partial_y T_a]_{y=x} = [\partial_y T_a]_{y=x_m} = 0. \tag{3}$$

Solution to Eq. (3) can be written in the form as $T_a(x,y)=\mathbf{H}_a(x,y)-x_d^{-1}(x^2+2C_1x+2C_2)$, where $\mathbf{H}_a(x,y)$ is solution of $x_d\partial_x^2\mathbf{H}_a(x,y)+y_d\partial_y^2\mathbf{H}_a(x,y)=0$ and C_1 and C_2 are constants. Particular solution to Eq. (3) that satisfies all the boundary conditions given by Eq. (2) is not known. However it is straightforward to derive that (a) when $x_l<y_0<x_0<x_m$, then the search space associated with this a molecule will be confined inside the interval (y,x_a) and the dynamics of the a molecule toward x_l will be retarded by the dynamic reflections of the b molecule that is present at y. As a consequence, the a molecule will exhibit a directional dependent net transport along the DNA chain preferentially toward its specific binding site x_a at the cost of the free energy spent on actively setting the initial position y_0 of b molecule on DNA such that $y_0<x_0$. This in turn decreases the resultant MFPT $[T_a(x_0,y_0)<T_a(x_0)]$. If $x_l<x_0<y_0<x_m$, then (b) there will be a retardation or shielding effect on the dynamics of the a molecule by the b molecule that results in a directional dependent net transport of the a molecule along DNA preferentially away from its specific site x_a or toward x_l at the cost of the free energy spent on actively setting the initial position y_0 of b molecule on DNA such that $x_0<y_0$. This in turn increases the resultant MFPT $[T_a(x_0,y_0)>T_a(x_0)]$. One should note that (c) the extent of this increase or retardation effects on the overall MFPT $T_a(x_0,y_0)$ should be a function of the distance $\beta=|x_0-y_0|$ of initial position of b molecule from the initial position of the a molecule since to make any impact on the dynamics of the a molecule, b molecule has to travel a distance of β. That is to say, one can write a functional relationship as follows:

$$|T_a(x_0,y_0) - T_a(x_0)| \propto \beta^{-1}. \tag{4}$$

The (d) limits $\lim_{\beta\to\infty} T_a(x_0,x_0-\beta)=\varepsilon_-$ and $\lim_{\beta\to\infty} T_a(x_0,x_0+\beta)=\varepsilon_+$ also should exist where $\varepsilon_{\mp}\neq 0$ and $\varepsilon_+>\varepsilon_-$. Using numerical simulations, we will show that the inequality $\varepsilon_+>\varepsilon_-$ partly arises as a consequence of our proposition (c). On any moment the passive b molecule does not transfer its specific/nonspecific binding free energy to the a molecule. Enhancing or shielding effects of b molecule on the dynamics of a molecule primarily originate as a consequence of transferring the momentum of the b molecules during the dynamic reflections at the boundary y. Momentum of the nonspecifically bound b molecules is mainly gained from the heat bath. Origin of the directional preference in the dynamics of the a molecule toward or away from its target site is clearly in accordance with the second law of thermodynamics since there are free-energy inputs involved in actively setting the initial position of the b molecule at y_0 in a purely enhancing/retarding mode with respect to the initial position x_0 of the a molecule which is also present on the same DNA chain.

III. SIMULATION RESULTS

To validate our propositions (a)–(d) we did random-walk simulations of the system of Eqs. (1)–(3) and computed the overall MFPT for a sample DNA chain. Results are shown in Fig. 2. For computation, we defined the helical ends of the DNA chain at $\{x_l,x_m\}\to\{-100,800\}$, the specific site of a molecule at $x_a=100$, and the initial position of a molecule at $x_0=90$ and we iterated the initial position of b molecule, y_0, inside interval $(-90,800)$ where $y_0\neq x_0$ at any moment due

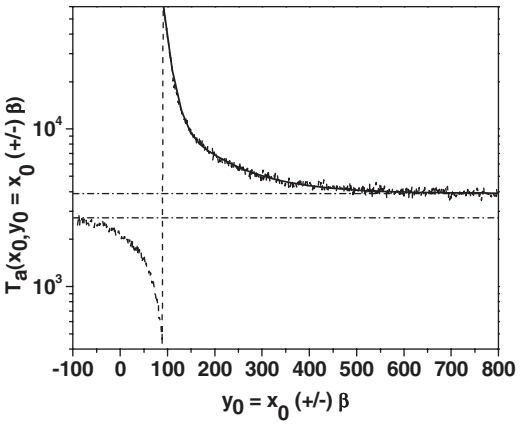

FIG. 2. Effect of the presence of other protein molecules on the MFPT (measured in the dimensionless number of steps) associated with the protein molecule of interest to locate its target site which is present on the same DNA. Here $\{x_l,x_m\}\rightarrow\{-100,800\}$ are the helical ends. The specific site is located at $x_a=100$ and the one-dimensional diffusion coefficient in the dimensionless form is assumed to be $x_d=1$. Further $x_0=90$ and y_0 (measured in the dimensionless position on the DNA lattice) is iterated in between the interval $[-90,800]$. In the absence of the b molecule, the MFPT becomes $T_a(x_0)=3900$ (dashed line). The MFPT was computed over 10^5 trajectories. Fitting (black solid line) of the retarding-effect data in the interval [90,800] to a function of the type $T_a(x_0,\beta) \sim m+ne^{-p_a\beta}+qe^{-p_b\beta}$ yielded the parametric estimates $m \sim (3.9\pm0.007)\times10^3$, $n\sim(1.84\pm0.5)\times10^4$, $p_a\sim0.009\pm0.0001$, $p_b\sim0.7\pm0.005$, and $q\sim(2.7\pm0.1)\times10^7$ with $R^2\sim0.996$.

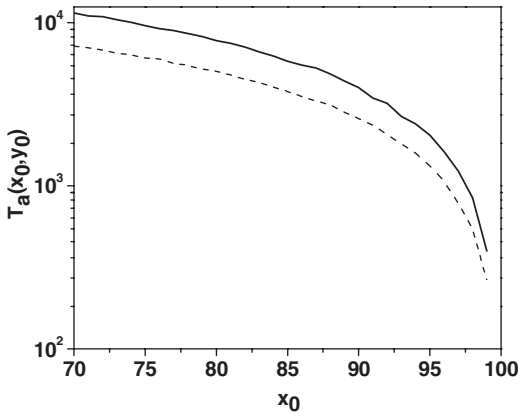

FIG. 3. Asymptotic behavior of $T_a(x_0,y_0)$ (measured in the dimensionless number of steps) at various initial positions x_0 (measured in the dimensionless position on the DNA lattice) with a fixed y_0. Here $\{x_l,x_m\}\rightarrow\{-100,800\}$ are helical ends. The specific site is located at $x_a=100$ and the one-dimensional diffusion coefficient in a dimensionless form is assumed to be $x_d=1$. Further $y_0=\{-90,800\}$ and x_0 is iterated inside the interval [70,99]. The MFPT was computed over 10^5 trajectories of a molecule. For a wide range of x_0, we observe the asymptotic values to be $\sim T_a(x_0)$ for $y_0=800$ (solid black line) and $\sim0.7T_a(x_0)$ for $y_0=-90$ (dashed black line).

to the excluding effect. When $x_l<y_0<x_0<x_m$ the reflecting boundaries for the a molecule are at $\{y,x_m\}$ and for the b molecule the reflecting boundaries are at $\{x_l,x\}$. When $x_l<x_0<y_0<x_m$, the reflecting boundaries for a molecule are at $\{x_l,y\}$ and for b molecule the reflecting boundaries are at $\{x,x_m\}$. Whenever a molecule finds its specific site x_a on DNA, it is assumed to bind there permanently. In other words x_a is an absorbing boundary for the dynamics of the a molecule. There are no absorbing boundaries or specific binding sites for the b molecule inside the interval (x_l,x_m). The MFPT was computed over 10^5 trajectories of the a molecule. Results indicate (Fig. 2) that our propositions (a)–(d) are indeed correct. The maximum enhancement or retardation of the MFPT occurs at $y_0=x_0\mp1$ for the cases $x_l<y_0<x_0<x_m$ and $x_l<x_0<y_0<x_m$, respectively. For all $\beta>1$ where $\beta=|x_0-y_0|$, we observed the following inequality:

$$T_a(x_0,x_0-\beta)<T_a(x_0,x_0+\beta). \tag{5}$$

The retarding or shielding effect of the b molecule when it is present at $y_0=x_0+\beta$ is much higher than that of the enhancing effect when it is present at $y_0=x_0-\beta$. From our proposition (d) it follows that $\lim_{\beta\rightarrow\infty}|T_a(x_0,x_0-\beta)-T_a(x_0,x_0+\beta)| =\Delta\varepsilon$, where $\Delta\varepsilon=|\varepsilon_--\varepsilon_+|>0$. When b molecule is present inside the interval (x_l,x_0), then the a molecule exhibits directional dependent net dynamics toward its specific site. This is mainly resulting from the confinement of search space of the a molecule inside the interval (y,x_m) and the dynamic reflections at boundary y. These enhancement or retardation effects approximately follow a biexponential-type trend with initial distance β of the b molecule from a molecule. Nonlinear least-squares fitting of the retardation effect data to a biexponential function of the form (Fig. 2) $T_a(x_0,x_0+\beta) \sim m+ne^{-p_a\beta}+qe^{-p_b\beta}$ yielded the parametric estimates as $p_a \sim0.009\pm0.0001$ and $p_b\sim0.7\pm0.005$. It is apparent to note that in the limit as $\beta\rightarrow\infty$ we recover $\lim_{\beta\rightarrow\infty}T_a(x_0,x_0+\beta) =m\sim\varepsilon_+$.

One can intuitively understand the origin of the inequality given by Eq. (5) as follows. Assume that there is no b molecule present inside the interval $\{x_l,x_m\}$. If there is a linear-type potential present along the DNA chain that favors the dynamics of this a molecule toward its specific site, then it can be shown that the MFPT $\vec{T}_a(x_0)$ associated with the escape of the a molecule through its specific site that is located at x_a starting from the position x_0 on DNA scales [9] with the distance $\alpha=|x_a-x_0|$ of its initial position x_0 from its specific site as $\vec{T}_a(x_0)\propto\alpha$. If the same linear-type potential drives our a molecule preferentially away from its specific site, then the MFPT $\overleftarrow{T}_a(x_0)$ scales with the distance of its initial position from the specific site as $\overleftarrow{T}_a(x_0)\propto e^{\alpha}$. As a result, we always have the inequality relationship $\vec{T}_a(x_0)<\overleftarrow{T}_a(x_0)$. There is a clear difference between the existence of linear-type potential favoring the dynamics of a molecule toward or away from its target site and enhancement or retardation effects observed in our model. In the former case, dynamics of a molecule is an asymmetric effect at each step, whereas in the latter case it is an overall or net effect. As per our proposition (d), we also observed the asymptotic limits $\varepsilon_-\sim0.7T_a(x_0)$ and $\varepsilon_+\sim T_a(x_0)$, and subsequently $\Delta\varepsilon=|\varepsilon_--\varepsilon_+|\sim0.3T_a(x_0)$. These asymptotic limits seem to be valid for all x_0 which is demonstrated in Fig. 3. So far we assumed that there is only one b molecule present in the system. If we increase the number of b molecules, then the enhancing or retarding effects increase proportionately. We computed the MFPT by

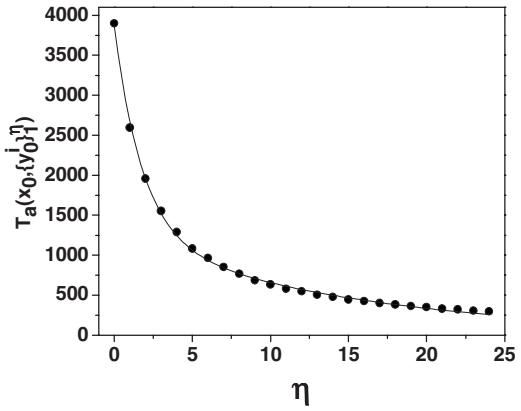

FIG. 4. Dependency of $T_a(x_0,y_0)$ (measured in the dimensionless number of steps) on the number of b molecules, η (a dimensionless quantity). Solid line is the nonlinear least-squares fit to the MFPT data (black circular dots). Here $\{x_l,x_m\}\rightarrow\{-100,800\}$ are helical ends. The specific site is located at $x_a=100$ and the one-dimensional diffusion coefficient in dimensionless form is assumed to be $x_d=1$. All the η numbers of b molecules are chosen from $(-99,-50)$, which results in net decrease in the resultant MFPT in a biexponential manner. Here β_i is chosen such that it approximately satisfies the asymptotic condition $\beta_i\rightarrow\infty$. MFPT was computed over 10^5 trajectories of a molecule. Fitting (solid black line) resulted in the biexponential factors of -0.6 ± 0.02 and -0.07 ± 0.003 with $R^2\sim0.998$.

keeping a constant x_0 and iterating the number of b molecules, η, in the asymptotic limits of $\mp\beta$. Numerical nonlinear least-squares fitting of these MFPT data with number of b molecules, η, is shown in Fig. 4. This result suggests that increasing or decreasing trend in MFPT with respect to η is approximately a biexponential type as follows:

$$T_a(x_0,\{y_0\mp\beta_i\}_{i=1}^{\eta})\sim T_a(x_0)\Omega_{\mp}(\eta)$$
$$=T_a(x_0)(C_3e^{\mp q_a\eta}+C_4e^{\mp q_b\eta}). \quad (6)$$

In the absence of b molecules ($\eta=0$) we have $\Omega_{\mp}(0)=1$ and right-hand side of Eq. (6) reduces to $T_a(x_0)$. As a result, we have the normalization condition $C_3+C_4=1$. Our further simulation results show that if there are η number of b molecules such that $x_l<\{x_0-\beta_i\}_{i=1}^{\eta}<x_0<x_m$ and σ number of b molecules such that $x_l<x_0<\{x_0+\beta_i\}_{i=1}^{\sigma}<x_m$, then their resultant or net effect on the overall MFPT can be given as follows:

$$T_a(x_0,(x_0>\{y_0-\beta_i\}_{i=1}^{\eta},x_0<\{y_0+\beta_j\}_{j=1}^{\sigma}))$$
$$\sim T_a(x_0)\Omega_-(\eta)\Omega_+(\sigma). \quad (7)$$

What will happen to the overall MFPT in Eq. (7) if we drive all these η and σ numbers of b molecules toward the limit $\beta\rightarrow\mp\infty$ with respect to initial position x_0 of a molecule? When all these retarding σ numbers of b molecules are sent toward the limit $\beta\rightarrow\infty$, from our proposition (d) and subsequent simulation results we find $\Omega_+(\sigma)\rightarrow1$ and right-hand side of Eq. (7) reduces to $T_a(x_0)\Omega_-(\eta)$ where we always have inequality relationship $\Omega_-(\eta)\leq1$. From our simulation results for single-enhancing-b-molecule case ($\eta=1$), we find the limit $\lim_{\beta\rightarrow\infty}\Omega_-(1)=\varepsilon_-\sim0.7$ (Fig. 2) and for $\eta=0$ we have $\Omega_-(0)=1$. Subsequently from Fig. 4 we also find that for all $\eta>1$ we have inequality $\Omega_-(1)>\Omega_-(\eta)$. These imply that when $\eta=\sigma$ and in the limit toward $\beta\rightarrow\infty$ for all these η and σ numbers of b molecules, the a molecule always exhibits directional dependent net transport preferentially toward its specific site on DNA. When all these other protein molecules are randomly distributed along DNA with equal stationary-state probabilities of finding them at a given DNA-binding site, then from inequality (5) we find that the a molecule always exhibits a directional dependent net transport preferentially away from its target site.

IV. DISCUSSION

Structural studies on *lac* repressor have revealed the presence of significant amount of thermally driven conformational fluctuations in its DNA-binding domains (DBDs) especially when this protein is nonspecifically bound with DNA. Upon finding the *lac*-operator sequence, DBD of *lac* repressor changes its conformation to more stable one with minimal degree of fluctuations [17]. These observations open up an argument or idea that these sequence dependent thermal fluctuations in DBDs might be effectively utilized by DBPs for actively searching for their specific sites on DNA via rapidly flipping between nonspecific binding sites and then switching to more stable conformation upon finding their specific sites. This means that DBPs might be acting as Maxwell's demons by extracting free energy from the heat bath to actively differentiate nonspecific binding sites from their specific sites on DNA. These arguments clearly violate second law of thermodynamics (18) since DBPs are not generally known for hydrolyzing ATPs to actively search for their specific sites on DNA. In other words, the second law must enforce a strict limit on the effects of such conformational fluctuations in DBDs on the efficiency of searching for their specific sites on DNA by DBPs. This also means that the shortest possible search time associated with site-specific DNA-protein interactions must be limited by the second law of thermodynamics (thermodynamic limit). Recently such a limit has been calculated in detail [18] and it seems that search time taken by DBPs to locate their specific sites on DNA could be closer to this thermodynamic limit only when the energy spectrum of such conformational fluctuations in DBDs of DBPs is tuned by the selective pressure in such a way that the overall site-specific binding time is minimum [18].

Under *in vivo* conditions, there will be different classes of DNA-interacting proteins (b molecules) undergoing one-dimensional diffusion along DNA simultaneously. From our theory, we learn that protein molecule of interest (a molecule) can be triggered to exhibit a directional dependent net transport along DNA by adjusting the distribution of the initial positions of b molecules on DNA with respect to the initial position of a molecule. In accordance with the second law, such redistribution of initial positions should involve free-energy inputs and no "demon" exists who can actively redistribute or recruit all the b molecules in a purely enhancing/retardation mode. Specific/nonspecific binding affinities of b molecules toward their binding sites on DNA can

be the sources of free-energy inputs for generating such initial distributions on the same DNA. Apparently this is the idea that is utilized in the enhancing effects of the CRMs on the RNAPII-promoter (PIC) interactions. In this case, the CRMs or enhancers are acting as demons by recruiting more TFs at their sites in enhancing mode at the cost of the binding affinities of TFs toward their cis-acting elements in accordance with the second law of thermodynamics. This is similar to actively recruiting more b molecules in the enhancing mode of our model where a molecule is the already formed ETF complex which is now tracking or looping toward PIC to initiate the transcription. In other words, influx of TFs toward ETF at the cost of their site-specific binding affinities results in directional dependent tracking or looping of ETF preferentially toward PIC. If there is no such free-energy input, then all these nonspecifically bound b molecules will be randomly distributed along DNA such that the stationary-state probabilities of finding them at any given DNA-binding site are equal. This means that the numbers of b molecules present inside a given segment of DNA will be directly proportional to the length of this DNA segment. In this condition, the overall retardation effect of the b molecules on the target-finding rate of the a molecule is always higher than the overall enhancing effect as given by inequality (5). As a consequence, the protein molecule of interest exhibits directional dependent net transport preferentially away from its target site. Extent of such directional preference strongly depends on the initial distance $\alpha=|x_0-x_a|$ of a molecule from its target site.

Two essential forces [16] are acting on the DNA-protein complex, viz., effective electrostatic interaction between the negatively charged backbone of the DNA chain and the positively charged amino acids and sequence dependent hydrogen-bonding interactions between the bases of DNA and amino acid side chains of protein molecule. Protein molecule can scan DNA for an average length of L bases for an average τ_L time, which is strongly influenced by these forces, and then dissociates to reassociate at different/same location. One should note that the one-dimensional mean-free-path length L is also strongly influenced by the presence of b molecules and also by the aforementioned forces. In the absence of other classes of protein molecules, the overall target-finding time can be given [6,7,16] by the following expression:

$$\tau_s(L) = (x_m - x_l)L^{-1}(\tau_L + \tau_{\rm ns}). \tag{8}$$

Here $\tau_L = L^2(6x_d)^{-1}$ is the time that is required to scan L bps of DNA and $\tau_{\rm ns}$ is the time required by our protein molecule to make nonspecific contact with DNA via three-dimensional routes. The term $(x_m-x_l)L^{-1}$ is the minimum number of association-dissociation events required by protein molecule of interest to scan the entire DNA. From the solution of $d_L\tau_s(L)|_{L=L_{\rm opt}}=0$ in Eq. (8), one can derive that [7] for an optimum target-finding rate, the one-dimensional scanning length L should be closer to $L_{\rm opt}=\sqrt{6x_d\tau_{\rm ns}}$. This also implies that for achieving a maximum target-finding rate, the protein molecule of interest should spend [7] equal amount of times in 1D and 3D routes ($\tau_{L_{\rm opt}}=\tau_{\rm ns}$). Our theory suggests that in the presence of b molecules, this rule will be strongly influenced by distance $\alpha=|x_0-x_a|$ of initial position of protein molecule of interest from its target site. If this distance α is higher, then retardation effects of other classes of proteins will result in directional dependent net transport of protein molecule of interest preferentially away from its target site. This also shifts the distribution $p(L)$ of one-dimensional scanning length L to lower side. Protein molecule of interest overcomes this retarding effect by spending more time on three-dimensional routes ($\tau_{L_{\rm opt}}<\tau_{\rm ns}$) or by increasing the number of dissociation-association events since the term $(x_m-x_l)L^{-1}$ in Eq. (8) increases as L decreases. This is mainly due to decrease in the mean-free-path length (L) of a molecule as $L\rightarrow L/(\sigma+\eta)$ upon increasing number of b molecules in the system. From this we obtain [14] the scaling law $\tau_L \propto L^2(\sigma+\eta)^{-2}$.

Contributions of 3D diffusion dominate when the distance of the protein molecule of interest from its target site is higher. If the protein molecule of interest moves closer to its target site, then the enhancing effects of the other classes of molecules slowly rise. As a result, there will be a net transport of the protein molecule of interest toward further closer to its target site and contribution of 1D diffusion increases. These results suggest that the distribution of the one-dimensional mean-free-path length L should be a dynamic quantity and also dependent on $\alpha=|x_0-x_a|$ as $p(L)\rightarrow p(L,\alpha)$. This implies that the optimum combination of 1D and 3D search times to achieve a minimum overall target-finding time ($\tau_{L_{\rm opt}}=\tau_{\rm ns}$) should be a dynamical rule in the presence of other classes of proteins which also depends on distance of protein molecule of interest from its target site on DNA. If this distance is higher, then the number of b molecules present in this distance will be proportionately higher. From our proposition (b) and subsequent simulation results as given in Eq. (6), we find that the MFPT associated with the a molecule to find its target site on DNA increases exponentially as a function of number of such b molecules. For avoiding such exponentially growing MFPTs, this a molecule will undergo more dissociation-association events which results in the inequality relationship $\tau_{L_{\rm opt}}<\tau_{\rm ns}$. This also means that $\tau_{L_{\rm opt}}\propto\alpha^{-1}$ since $\tau_{\rm ns}$ is a fixed quantity for a given DNA chain length. It is also apparent to note that the optimum 1D scanning length $L_{\rm opt}$ is directly connected with the recently calculated thermodynamic limit of the overall search time [18] since the extent of the thermally driven conformational fluctuations in DBDs of nonspecifically bound DBPs on DNA is negatively correlated with the one-dimensional scanning length L and also both these quantities minimize the overall search time. If the extent of such conformational fluctuations in DBDs is higher, then DBPs perform more dissociation-association events by rapidly flipping between nonspecific binding sites of the same DNA chain that in turn shifts the distribution of one-dimensional mean-free-path length L of DBPs on DNA to lower side. Presence of other molecules on the same DNA chain further intensifies this effect in a distance (α) dependent manner.

V. CONCLUSIONS

We have shown that depending on the distribution of the initial positions of other classes of protein molecules which

are present on the same DNA, the protein molecule of interest could exhibit a directional dependent net transport preferentially either toward or away from its target site. Using a combination of theoretical and simulation tools, we have demonstrated that the directional preference of the protein molecule of interest mainly arose as a consequence of the confinement of its search space and dynamic reflections by other protein molecules. In accordance with the second law of thermodynamics, generation of such initial distributions of other protein molecules purely in an enhancing or a retardation mode on DNA must involve an external free-energy input. In the mechanism of action of the cis-acting elements on the initiation of transcription of the corresponding genes, such free-energy inputs are mainly derived from the site-specific binding affinities of the inflowing transcriptional factors toward their respective cis-acting elements. In the absence of such site-specific binding affinities, initial distribution of the other protein molecules on DNA would be a random one with equal probabilities of finding them at a given DNA-binding site. Under such conditions, the protein molecule of interest exhibited a net transport preferentially away from its target site. This directionality mainly originated from unequal natures of the enhancing and retarding effects of randomly distributed other classes of protein molecules and the extent of such directional preference was positively correlated with the initial distance of the protein molecule of interest from its target site. Apparently the protein molecule of interest overcomes the retarding effects of other classes of protein molecules in a dynamical manner by increasing the number of dissociation-association events when it is far away from its target site and then by switching back to the sliding dynamics due to increase in enhancing effects as it gets closer to its target site.

ACKNOWLEDGMENTS

This work was supported by Indian Institute of Technology Madras (IITM). The author is a recipient of IC & SR new faculty start-up grant (Grant No. BIO/08-09/241/NFSC/RMUR) from IITM.

[1] B. Levin, *Genes VIII* (Prentice-Hall, Englewoods Cliff, NJ, 2003); B. Alberts, D. Bray, J. Lewis, K. Roberts, and J. D. Watson, *Molecular Biology of the Cell* (Garland, New York, 1994); M. Ptashne, *A Genetic Switch* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2004).

[2] A. D. Riggs, S. Bourgeois, and M. Cohn, J. Mol. Biol. **53**, 401 (1970).

[3] G. Adam and M. Delbruck, *Structural Chemistry in Molecular Biology* (Freeman, San Francisco, 1968); R. F. Bruinsma, Physica A **313**, 211 (2002).

[4] O. G. Berg, R. B. Winter, and P. H. von Hippel, Biochemistry **20**, 6929 (1981).

[5] O. G. Berg, R. B. Winter, and P. H. von Hippel, Biochemistry **20**, 6961 (1981).

[6] S. E. Halford and M. D. Szczelkun, Eur. Biophys. J. **31**, 257 (2002); S. E. Halford and J. F. Marko, Nucleic Acids Res. **32**, 3040 (2004).

[7] M. Slutsky and L. A. Mirny, Biophys. J. **87**, 4021 (2004); M. Slutsky, M. Kardar, and L. A. Mirny, Phys. Rev. E **69**, 061903 (2004).

[8] G. Kolesov, Z. Wunderlich, O. N. Laikova, M. S. Gelfand, and L. A. Mirny, Proc. Natl. Acad. Sci. U.S.A. **104**, 13948 (2007).

[9] R. Murugan, Biophys. Chem. **120**, 143 (2006).

[10] M. Ptashne and A. Gann, *Genes and Signals* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2001).

[11] R. Murugan, J. Theor. Biol. **248**, 696 (2007); E. M. Blackwood and J. T. Kadonaga, Science **281**, 60 (1998).

[12] G. Lan and S. X. Sun, Biophys. J. **88**, 999 (2005).

[13] M. O. Magnasco, Phys. Rev. Lett. **71**, 1477 (1993); R. D. Astumian, Science **276**, 917 (1997); P. Reimann, Phys. Rep. **361**, 57 (2002).

[14] I. M. Sokolov, R. Metzler, K. Pant, and M. C. Williams, Biophys. J. **89**, 895 (2005).

[15] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 2004); H. Risken, *Fokker-Planck Equations* (Springer, Berlin, 1996); N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 2004); S. Redner, *A Guide to First Passage Processes* (Cambridge University Press, Cambridge, England, 2001).

[16] R. Murugan, Phys. Rev. E **76**, 011901 (2007); R. Murugan, *ibid.* **69**, 011911 (2004).

[17] C. G. Kalodimos, N. Biris, A. M. J. J. Bonvin, M. M. Levandoski, M. Guennuegues, R. Boelens, and R. Kaptein, Science **305**, 386 (2004).

[18] L. Hu, A. Y. Grosberg, and R. Bruinsma, Biophys. J. **95**, 1151 (2008).