Generalized theory of site-specific DNA-protein interactions

R. Murugan*

Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai, 400005, India and Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, Maryland 20742, USA

(Received 2 January 2007; revised manuscript received 9 April 2007; published 2 July 2007)

We develop a generalized theory of the site-specific DNA-protein interactions, which includes both the static as well as the dynamical factors influencing the one-dimensional diffusion of the nonspecifically bound protein molecule which is in the process of searching for the specific site on the DNA lattice. We argue that the chemically driven condensation of the DNA molecule introduces a static distribution in the one-dimensional phenomenological diffusion coefficient associated with the protein molecule and the conformational dynamics of the DNA introduces temporal fluctuations in the one-dimensional diffusion coefficient over the static distribution. We further derive the generalized inequality conditions and the scaling laws which are required to enhance the three-dimensional diffusion controlled site-specific association rate to an arbitrary order. Our model predicts that when the degree of condensation of the DNA molecule under consideration is very high, then the probability distribution associated with the stationary state one-dimensional diffusion coefficient variable as well as the stationary state one-dimensional diffusion length variable will be a flat one. Further analysis reveals that to achieve a site-specific association rate which is higher than that of the three-dimensional diffusion controlled rate, the one-dimensional diffusion length associated with the dynamics of the nonspecifically bound protein molecule on the DNA lattice should fall in certain critical ranges. Comparison of our theoretical results with the recent experimental observations reveals that when the DNA molecule is under a stretched condition, then the static distribution of the one-dimensional diffusion coefficient associated with the dynamics of the protein molecule on the DNA lattice is a Gaussian and therefore the fluctuations in the one-dimensional diffusion coefficient generated by the dynamical factors are confined in a harmonic type potential.

DOI: 10.1103/PhysRevE.76.011901 PACS number(s): 87.15.Kg

I. INTRODUCTION

Interaction of a protein molecule with a specific site on the DNA lattice in the presence of an enormous amount of nonspecific sites is a fundamental phenomenon in molecular biology and biological physics (Refs. [1,2]) which plays important roles in the recognition of the promoter sequences by the RNA polymerase (RNAP) enzyme in case of transcription, recognition of the origin of replication by a DNA polymerase enzyme in case of replication and, recognition of the ribosome binding site (RBS) on the mRNA lattice by the ribosome assembly in case of translation of the genomic DNA. Apart from these, studying the site-specific DNA-protein interactions is very important and useful not only in the context of designing DNA-targeting molecules such as anticancer drugs, and probes in case of DNA microarrays, but also important in the bioinformatics field especially in devising the strategies or algorithms to identify the specific protein interacting *cis*-regulatory sequences on the DNA lattice such as transcription initiation sites and enhancer motifs in the eukaryotic genomes.

In the living cells, the DNA interacting proteins are synthesized in the cytoplasm from their corresponding mRNA templates by the ribosome machineries. Then these synthesized protein molecules might diffuse from the place of their origin through the cytoplasm and locate their specific sites on the DNA lattice via pure three-dimensional routes. Therefore earlier models described the site-specific interaction of the protein molecule with the DNA lattice as a three-dimensional diffusion controlled process. However, detailed experimental studies on the site specific binding of the Lac repressor with its corresponding operator sequence on the template DNA lattice (Refs. [2–5]) showed an association rate of $\sim 10^{10}$ mol^{-1} s^{-1} which is much higher than that of the three-dimensional diffusion controlled bimolecular collision rate limit of $\sim 10^{8}$ mol^{-1} s^{-1}. The possible explanation for this observed superdiffusive site-specific association rate could be the existence of an efficient mechanism (Ref. [2]) of searching for the specific site probably in the reduced dimensional space (Fig. 1). For example, consider a DNA lattice of N base pairs in length which contains a specific-site binding stretch of m base pairs in length on it. Here the specific-site binding stretch is simply the length of the DNA lattice spanned by the specifically or nonspecifically bound protein molecule (e.g., see Fig. 2). Now, if the protein molecule tries to find this specific site via pure three-dimensional routes, then the maximum achievable site-specific bimolecular association rate will be in the order of $\sim N^{-1} \times 10^{8}$ mol^{-1} s^{-1}. This is because the probability of occurrence of a site-specific contact between the protein molecule and the DNA lattice at each collision event is $N^{-1}$and the diffusion controlled collision rate limit in the aqueous medium is $\sim 10^{8}$ mol^{-1} s^{-1}. On the other hand, the nonspecific bimolecular association rate will be in the order of

*Present address: Department of Neurology, Children's Hospital Boston, Harvard Medical School, Boston, MA 02115. Electronic address: rmurugan@gmail.com

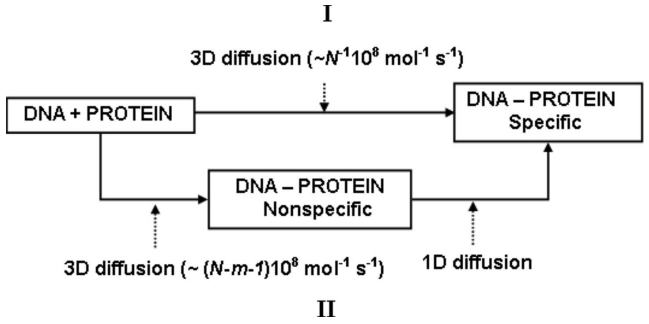

FIG. 1. Possible pathways of site-specific DNA-protein interactions. Pathway I: The protein molecule can find its specific site on the DNA lattice via pure three-dimensional (3D) routes. Here the maximum achievable site-specific association rate is $\sim N^{-1} \times 10^8$ mol^{-1} s^{-1} since the probability of finding the specific site at each three-dimensional collision between the protein molecule and the DNA lattice is N^{-1} where N is the size of the DNA under consideration in base pairs and the three-dimensional diffusion controlled bimolecular collision rate limit in the aqueous medium is $\sim 10^8$ mol^{-1} s^{-1}. Pathway II: The protein molecule nonspecifically binds with the DNA lattice via three-dimensional routes and then searches for the specific site via one-dimensional (1D) diffusion along the DNA lattice. When the specific-site binding stretch, which is simply the length of the DNA lattice spanned by the specifically or nonspecifically bound protein molecule, is m base pairs in length (see Fig. 2), then there are at least $N-m-1$ numbers of nonspecific binding sites on the DNA lattice and therefore the maximum achievable diffusion controlled nonspecific bimolecular association rate will be $\sim(N-m-1)\times 10^8$ mol^{-1} s^{-1}.

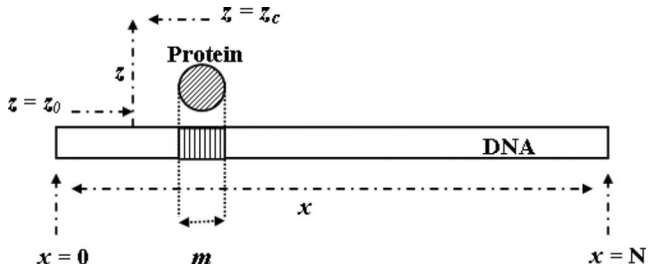

FIG. 2. Schematic diagram describing various boundaries associated with the dynamics of the nonspecifically bound protein molecule in the process of searching for the specific site on the DNA lattice. Here m is the length of the specific-site binding stretch which is simply the number of base pairs spanned by the bound protein molecule on the DNA lattice. The variable z denotes the contact distance between the charged groups of the protein molecule and the DNA backbone, and the variable x denotes the position of the protein molecule on the DNA lattice. Since the protein molecule and the DNA lattice cannot overlap each other, we set z_0 acts as a reflecting boundary and when the contact distance z between the protein molecule and the DNA lattice is such that $z \leq z_c$, we say that the protein molecule is nonspecifically associated with the DNA and, when $z \geq z_c$ we say that the protein molecule is physically dissociated from the DNA lattice and therefore the dynamical variable z describes the association-dissociation events in the DNA-protein interactions. Similarly the protein molecule may dissociate from the helical ends of the DNA lattice $x=\{0,N\}$ since the symmetry associated with the binding of the protein molecule with the DNA will be broken at these lattice positions.

$\sim(N-m-1)\times 10^8$ mol^{-1} s^{-1} (Fig. 1) since there will be at least $(N-m-1)$ numbers of nonspecific binding sites on the DNA lattice.

Therefore, the observed superdiffusive site-specific association rate may be explained by assuming a nonspecific binding of the protein molecule with the DNA lattice via three-dimensional routes which is then followed by a one-dimensional searching for the specific site. However, this one-dimensional searching mechanism should be efficient in such a way that the overall association rate is in the order of $\sim 10^{10}$ mol^{-1} s^{-1}. Using these ideas, the superdiffusion paradox was resolved by assuming a two-step (Refs. [3–5]) target finding mechanism where the protein molecule nonspecifically binds to the DNA via three-dimensional routes with a rate of $\sim(N-m-1)\times 10^8$ mol^{-1} s^{-1} in the first step where m is the length of the specific-site binding stretch and then searches for the specific site via one-dimensional facilitating diffusion processes such as sliding, hopping, and intersegmental transfers (Refs. [3–5]). Here one should note that the DNA interacting proteins are generally rich in positively charged amino acids whereas the phosphate groups on the DNA backbone are negatively charged. Due to this fact one can argue that the nonspecific interactions between the protein molecule and the DNA lattice might be mainly driven by the electrostatic attractive force which is almost homogeneous allover the entire DNA lattice. However, these electrostatic attractive forces are significantly influenced by the surrounding water molecules and the presence of other counterions. Particularly, when the ionic strength of the aqueous medium is higher, then the negative charges in the phosphate groups of the DNA lattice will be surrounded by a cloud of positively charged ions of the solution, and similarly the positive charges in the amino acid side chains present on the surface of the protein molecule will be surrounded by a cloud of negatively charged ions of the solution, which in turn leads to the formation of the loosely packed electrical double layer around the charged groups of the protein molecule as well as the DNA lattice. Therefore, in the presence of higher ionic strengths, the electrostatic attractive forces present in-between the protein molecule and the DNA lattice will be weakened in such a way that the resultant force will be a time dependent oscillating quantity with a mean value of zero. Here one should note that these time dependent oscillations or fluctuations in the resultant electrostatic potential originate mainly from the collisions of the protein molecule with the DNA lattice which is induced by the Brownian dynamics. For example, detailed experimental studies showed that (Refs. [4,5]) the increase in the ionic strength of the aqueous medium would eventually increase the rate of site-specific DNA-protein interactions and decrease the specificity of such site-specific interactions. Here one should note that the increase in the site-specific DNA-protein association rate is inversely proportional to the strength of the nonspecific electrostatic attractive forces present in-between the oppositely charged protein molecule and the DNA lattice since these attractive forces tend to retard the dynamics of the nonspecifically bound protein molecule which is in the process of searching for the specific site on the DNA lattice.

On the other hand, when the protein molecule finds its specific site on the DNA lattice via the one-dimensional routes, it is generally believed that the site-specific binding of the nonspecifically bound protein molecule with the DNA lattice is mainly driven by the hydrogen bonding interactions between the protein molecule and the specific site of the DNA lattice. However, when the protein molecule finds its specific site on the DNA lattice directly via the three-dimensional routes despite the one-dimensional scanning, then these site-specific hydrogen bonding interactions will accompany a serious dehydration penalty (e.g., see Ref. [6]) of the break of hydrogen bonds with water molecules. Therefore one can conclude that the hydrogen bonding interactions cannot be a major driving force for the site-specific DNA-protein interactions via the three-dimensional routes. Nevertheless, under these conditions, the site-specific interaction of the protein molecule with the DNA lattice may be driven by the gain of the configurational entropy of water as in the cases of protein folding (e.g., see Refs. [7]), protein aggregation, and receptor-ligand binding. Here we should note that when the site-specific DNA-protein interactions are driven by hydrogen bonding via one-dimensional routes, then the affinity and the specificity associated with the DNA-protein interactions will correlate negatively with each other. Contrasting from this, when the site-specific interactions between the protein molecule and the DNA lattice is driven by the gain of configurational entropy of water, then as in the case of receptor-ligand binding, the affinity and the specificity associated with the DNA-protein interaction will correlate positively with each other (Ref. [8]). Here one should note that these hydrogen bonding interactions may also be mediated by the water molecules which are present in the DNA-protein interface (Ref. [9]). However, these water mediated hydrogen bonding interactions are generally nonspecific in nature and they are mainly used as space fillers in the DNA-protein interface for stabilizing the nonspecifically bound DNA-protein complex (Ref. [9]).

II. FACTORS INFLUENCING THE PROTEIN DYNAMICS

Considering all the aforementioned interactions, we can conclude that the dynamics of the nonspecifically bound protein molecule which is in the process of searching for the specific site on the DNA lattice via the one-dimensional facilitating mechanisms will be influenced by the following four basic factors.

(1) Effective nonspecific attractive potential (Refs. [1,10–16]) present in-between the positively charged amino acids side chains of the protein molecule and the negatively charged phosphate groups of the DNA backbone. As we have already discussed in the previous section, the magnitude of these electrostatic attractive forces will be significantly influenced by the surrounding water molecules and the other counterions and, under extreme conditions such as in the presence of very high ionic strengths, the effective resultant potential may be a time dependent oscillating quantity around the mean value of zero and it also includes all the other probable interactions such as the entropically induced ones. However, here we assume that the ionic strength of the aqueous medium under consideration where the site-specific DNA-protein interaction is taking place is lower enough in such a way that the resultant potential is always attractive and almost time independent. We denote this effective attractive potential as $F_v(z)$ where z is the contact distance between the DNA helix and the protein molecule (see Fig. 2 for details), subscript v stands for the vertical direction, and the force generated by this potential is simply $-d_z F_v(z)$. When the contact distance between the protein molecule and the DNA lattice is $z \leq z_c$, we say that the protein molecule is nonspecifically associated with the DNA and, when $z \geq z_c$ we say that the protein molecule is physically dissociated from the DNA lattice. Here we should note that the dynamical variable z describes the association-dissociation events in the DNA-protein interactions.

(2) Sequence dependent interactions between the DNA-interacting amino acid side chains of the protein molecule (Refs. [17,18]) and the specific-site binding stretch sequence of the DNA lattice. Here one should note that the specific-site binding stretch is simply the number of base pairs spanned by a specifically or nonspecifically bound protein molecule (see Fig. 2) on the DNA lattice. For example, in case of prokaryotic RNA polymerase (RNAP), it is 50–60 base pairs (e.g., see Ref. [1]). This sequence dependent interaction includes hydrogen bonding, van der Waals, and water mediated bonding. As we have seen in the Introduction section, when the protein molecule finds its specific site on the DNA lattice via three-dimensional routes, then this sequence dependent potential will also include the water induced entropic effects. We put all these types of sequence dependent interactions together and denote the resultant potential of mean force as $F_h(x)$, where x is the base-pair position of the nonspecifically bound protein molecule along the DNA lattice, subscript h stands for the horizontal direction, and the force generated by this potential is given as $-d_x F_h(x)$. Here the variable x describes the dynamics associated with the protein molecule in the process of one-dimensional searching for the specific site on the DNA lattice.

If we denote the position of each base pair as x_i, where $i=1,2,3,\ldots,N$ and N is the total size of the DNA molecule in base pairs, then $F_h(x_i)$ is a local minimum in the sequence dependent potential $F_h(x)$ at the ith base-pair position. Here one should note that the potential at the ith base pair position $F_h(x_i)$ is the sum of the interactions between the protein molecule and all the base pairs in the specific-site binding stretch. For example, if we denote the interaction potential between the individual jth base pair and the bound protein molecule as $f_h(x_j)$ then we can write the sequence dependent potential $F_h(x_i)$ as the sum $F_h(x_i)=\Sigma_{j=i}^{i+m} f_h(x_j)$, where m is the length of the specific-site binding stretch (see Fig. 2) which is simply the number of base pairs spanned by the bound protein molecule on the DNA lattice. Here the movement of the protein molecule from the potential minimum $F_h(x_i)$ to $F_h(x_{i+1})$ along the DNA lattice will be impeded by a barrier $B_{i,i+1}$ and generally $B_{i,i+1} \neq B_{i+1,i}$. When there is a specific site at $x=x_a$, then one should note that $F_h(x_a)$ is the global minimum in the potential $F_h(x)$ and generally the inequalities $B_{a-1,a} \ll B_{a,a-1}$ and $B_{a+1,a} \ll B_{a,a+1}$ holds true which in turn

ensure the tight binding of the protein molecule at the specific site $x=x_a$ on the DNA lattice. It is obvious to note that under physiological conditions the nonspecifically bound protein molecule on the DNA lattice can perform an efficient one-dimensional random search for the specific site only when the inequality $B_{i,i+1} \leq k_B T$ holds true for any set of lattice positions $x_i, x_{i\pm1}$, where $i \neq a$, k_B is the Boltzmann constant, and T is the absolute temperature (e.g., see Refs. [17,18]). On the other hand, the binding specificity B_s of the protein molecule towards its target site which is situated at the position $x=x_a$ on the DNA lattice is directly proportional to the barrier heights $B_{a\pm1,a}$ as $B_s \propto B_{a\pm1,a}$ and the free energy difference $\delta F_{h,a,a\pm1}$ as $B_s \propto \delta F_{h,a,a\pm1} = |F_h(x_a) - F_h(x_{a\pm1})|$. It is obvious to note that the barrier heights $B_{a\pm1,a}$ as well as the free energy difference $\delta F_{h,a,a\pm1}$ should be in such a way that $B_{a\pm1,a} > k_B T$ and $|F_h(x_a) - F_h(x_{a\pm1})| = \delta F_{h,a,a\pm1} > k_B T$ for a tight binding of the protein molecule at the specific site on the DNA lattice to occur.

(3) Chemically driven condensation of the DNA molecule, e.g., condensation driven by the divalent cations such as Mg^{2+} under physiological conditions. Here one should note that the degree of condensation of the template DNA molecule is inversely proportional to the probability associated with the loss of the dissociated nonspecifically bound protein molecule from the template DNA into the bulk solution. In other words the time taken by the dissociated nonspecifically bound protein molecule to reassociate back with the DNA lattice will be reduced by the condensation of the DNA lattice. This in turn enhances the site-specific association rate. In addition to this, under condensed state, since two distal regions of the same DNA molecule can be brought closer together via the randomly occurring ring closure events, the phenomenological one-dimensional diffusion coefficient associated with the dynamics of the protein molecule will vary along the DNA lattice in a random manner. This is because the one-dimensional diffusion coefficient is directly proportional to the hopping distance of the protein molecule along the DNA lattice without an intervening dissociation event. Here one should note that the disorder introduced by the condensation of the DNA in the one-dimensional diffusion coefficient is a static one. In this context, it is interesting to note that the DNA condensing agent Mg^{2+} is an important component in almost all the reactions requiring the site-specific DNA-protein interactions, e.g., ligation, polymerase chain reaction, and restriction digestion.

(4) Thermally driven dynamical fluctuations in the DNA conformation and the presence of other or the same DNA interacting molecules or local entanglement of the DNA molecule in the trajectory of the protein molecule of our interest. These fluctuations apparently introduce a time dependent variation in the local conformational state of the DNA lattice which in turn introduces a dynamical disorder in the phenomenological one-dimensional diffusion coefficient which means that the one-dimensional diffusion coefficient will fluctuate in a time dependent manner over the static distribution generated by the factor 3.

III. MODELS ON DNA-PROTEIN INTERACTIONS

As we have seen in the previous section, here the factors 1–3 are static in nature and the factor 4 is a dynamical one. Earlier models (Refs. [10–16]) considered mainly the factor 1 while describing the site-specific DNA protein interactions and were therefore successful in explaining various experimental observations pertained to the effect of ionic strength on the site-specific association rate. Later, factor 2 was incorporated in some models (Refs. [17,18]). Recently a random jump model (Ref. [19]) has been developed which incorporated the factor 3. Therefore almost all the earlier models considered only the static factors 1–3 individually or in a combination, and none of the models considered all the factors 1–4 while describing the dynamics of the protein molecule on the DNA lattice. In this paper we develop such a generalized phenomenological model on site-specific DNA-protein interaction that includes all the aforementioned factors 1–4. The results obtained from our model are important not only to understand the physics of the DNA-protein interactions under *in vivo* conditions, but also have great applicability in the designing of DNA targeting molecules, e.g., anticancer drugs and probes. The organization of this paper is as follows. First we derive the set of Langevin type stochastic differential equations which includes all the factors 1–4 to describe the dynamics of the nonspecifically bound protein molecule which is in the process of searching for the specific site on the DNA lattice. Then we will derive the corresponding Fokker-Planck equation which describes the evolution of the probability distribution function associated with the dynamics of the nonspecifically bound protein molecule on the DNA lattice and we also solve the mean first passage time (MFPT) problem for such dynamics. Using these results we will derive the generalized inequality conditions and the scaling laws, which are required to enhance the three-dimensional diffusion controlled site-specific association rate to an arbitrary order. Finally we will discuss the consequences of our results in connection with the recent experimental observations.

IV. GENERALIZED THEORY ON DNA-PROTEIN INTERACTIONS

Let us start to develop our model by noting the fact that the factor 3 introduces a static distribution in the one-dimensional diffusion coefficient. We denote the phenomenological one-dimensional diffusion coefficient by the variable y. Here the factor 3 introduces a static distribution in the variable y within the interval $y_{\min} \leq y \leq y_{\max}$. When the degree of condensation associated with the DNA molecule under consideration is very high then one expects a flat distribution in the variable y, i.e., the stationary state probability $p_s(y)$ of observing the diffusion coefficient y is simply given as $p_s(y) = [y_{\max} - y_{\min}]^{-1}$. This is because the protein molecule can hop anywhere in the DNA molecule with equal probabilities when the degree of condensation of the DNA molecule is very high. However, for an arbitrary degree of condensation of the template DNA one can write the stationary state probability distribution associated with the diffusion coefficient variable y as $p_s(y) \propto e^{-(k_B T)^{-1} F_d(y)}$, where $F_d(y)$ is the potential of mean force acting on the variable y and, the subscript d stands for the diffusion process. The force generated by this potential is $-d_y F_d(y)$. On the other hand the

factor 4 introduces dynamical fluctuations in the one-dimensional diffusion coefficient over the static distribution generated by the factor 3. As a result, the one-dimensional diffusion coefficient will be a time dependent random quantity with definite boundaries, i.e., the variable y fluctuates within the interval $[y_{\min}, y_{\max}]$, whose mean value can be anywhere inside the interval $[y_{\min}, 2^{-1}(y_{\max}+y_{\min})]$, which in turn will be decided by the degree of condensation of the DNA molecule (factor 3). For example, when the probability distribution associated with the variable y is given as $p_s(y)=[y_{\max}-y_{\min}]^{-1}$ then it is easy to verify that the mean value is $\int_{y_{\min}}^{y_{\max}} y p_s(y) dy = 2^{-1}(y_{\max}+y_{\min})$. Similarly the variance associated with the time dependent fluctuations in the one-dimensional diffusion coefficient will be decided by the amplitude of the dynamical fluctuations in the DNA conformation (factor 4). When these static and the dynamical factors 1–4 are taken into account, the dynamics of a protein molecule on the DNA lattice under nonspecifically bound conditions can be well described by the following set of Langevin equations:

$$d_t x = - d_x F_h(x) + \sqrt{y}\,\eta(t),$$

$$d_t y = - d_y F_d(y) + \sqrt{y_m}\,\lambda(t), \tag{1}$$

$$d_t z = - d_z F_v(z) + \sqrt{z_m}\,\omega(t).$$

Here we should recall the fact that the variable x denotes the position of the protein molecule on the DNA lattice which is confined in the interval $0 \le x \le N$ where N is the size of the DNA lattice under consideration and $x=\{0,N\}$ are the helical ends of the DNA lattice, z denotes the distance between the charged groups on the protein molecule and the phosphate groups of the DNA backbone which is confined in the interval $z_0 \le z \le z_c$ (see Fig. 2), and y denotes the time dependent diffusion coefficient associated with the dynamics of the protein molecule on the DNA lattice which fluctuates in the interval $y_{\min} \le y \le y_{\max}$, where $y=\{y_{\min}, y_{\max}\}$ are acting as the reflecting boundaries. $F_d(y)$ is the y dependent potential for the dynamics of the protein molecule over the y variable, which originates from the conformational inhomogeneity of the DNA molecule (factor 3), and controls the shape of the distribution of y, $\omega(t)$ and $\eta(t)$and $\lambda(t)$ are the delta correlated Gaussian noises with means $\langle \omega(t) \rangle = 0$, $\langle \eta(t) \rangle = 0$, and $\langle \lambda(t) \rangle = 0$, and the variances satisfying the fluctuation-dissipation theorems as follows:

$$\langle \eta(t)\omega(t) \rangle = \langle \eta(t)\lambda(t) \rangle = \langle \omega(t)\lambda(t) \rangle = 0,$$
$$\langle \eta(t)\eta(t') \rangle = \langle \lambda(t)\lambda(t') \rangle = \langle \omega(t)\omega(t') \rangle = \delta(t-t'). \tag{2}$$

The constants z_m and y_m are the stationary state variances associated with the fluctuations in the corresponding variables. The Fokker-Planck equation describing the time evolution of the probability distribution function $(P_{x,y,z,t})$ associated with the protein molecule to be found at the point (x,y,z) at time t can be written as follows (Refs. [20–23]):

$$\begin{aligned} \partial_t P_{x,y,z,t} &= \partial_x [d_x F_h(x) P_{x,y,z,t}] \\ &+ \partial_y [d_y F_d(y) P_{x,y,z,t}] + \partial_z [d_z F_v(z) P_{x,y,z,t}] \\ &+ 2^{-1}(y \partial_x^2 P_{x,y,z,t} + y_m \partial_y^2 P_{x,y,z,t} + z_m \partial_y^2 P_{x,y,z,t}). \end{aligned} \tag{3}$$

Here the initial condition is $P_{x,y,z,0|x_0,y_0,z_0,0} = \delta(x-x_0)\delta(y-y_0)\delta(z-z_0)$. The generalized nonlinear partial differential equation (3) describes the evolution of the probability distribution function associated with the dynamics of the nonspecifically bound protein molecule in the process of searching for the specific site on the DNA lattice. The mean first passage time ($T_{x,y,z}$, denoted as MFPT) associated with the escape of the protein molecule from the DNA lattice, after making a nonspecific contact at the initial position (x,y,z) at time $t=0$ satisfies the following backward Fokker-Planck equation:

$$\left.\begin{aligned} &-[d_x F_h(x)]\partial_x T_{x,y,z} - [d_y F_d(y)]\partial_y T_{x,y,z} - [d_z F_v(z)]\partial_z T_{x,y,z} \\ &+ 2^{-1}(y \partial_x^2 T_{x,y,z} + y_m \partial_y^2 T_{x,y,z} + z_m \partial_z^2 T_{x,y,z}) \end{aligned}\right\} = -1. \tag{4}$$

Now, one should note that the total residence time $T_{x,y,z}$ associated with the protein molecule to stay on the DNA lattice consists of two independent components, viz., one-dimensional searching time $T_{x,y}$ along the DNA lattice and the dissociation time T_z. This is because at any moment the protein molecule can undergo only either a dissociation-association dynamics or a one-dimensional diffusion dynamics along the DNA lattice. Therefore, using the separation ansatz $T_{x,y,z}=T_{x,y}+T_z$ where $T_{x,y,z}$ is the mean time for which the protein molecule stays on the DNA lattice after making a nonspecific contact via three-dimensional routes, we can separate Eq. (4) into two auxiliary equations as follows:

$$\begin{aligned} &-[d_x F_h(x)]\partial_x T_{x,y} - [d_y F_d(y)]\partial_y T_{x,y} \\ &\quad + 2^{-1}(y \partial_x^2 T_{x,y} + y_m \partial_y^2 T_{x,y}) = -\alpha, \\ &-[d_z F_v(z)] d_z T_z + 2^{-1} z_m d_z^2 T_z = -\beta, \end{aligned} \tag{5}$$

where $\alpha+\beta=1$, α is the fraction of trajectories following $T_{x,y}$, and β is the fraction of trajectories following T_z. Now let us consider the dynamics of the protein molecule only over the z variable. Since the protein molecule and the DNA lattice cannot overlap each other, we set the effective potential $F_v(z)$ such that when $z \le z_0$, z_0 acts as reflecting boundary (Fig. 2), and when $z \ge z_c$ the effective attractive potential $F_v(z)$ will be so weak in such a way that the protein molecule can physically dissociate from the DNA lattice, i.e., z_c acts as absorbing boundary. Now noting the fact that $\alpha+\beta=1$ in Eq. (5), and using the absorbing and the reflecting boundary conditions for the corresponding MFPT components as $T_z|_{z=z_0} = d_z T_z t|_{z=z_c} = 0$ for T_z, we find that (Refs. [20,21])

$$T_z = \beta \frac{2}{z_m} \int_z^{z_c} e^{2/z_m \int_{z_0}^{y} d_w F_h(w) dw} dy \int_{z_0}^{y} e^{-2/z_m \int_{z_0}^{w'} d_w F_h(w) dw} dw'. \tag{6}$$

We have already seen in the Introduction section that the electrostatic attractive forces between the positively charged amino acid side chains of protein molecule and the negatively charged phosphate groups on DNA backbone will be weakened in the presence of the water molecules and higher strength of the other counterions in such a way that the effective force may be a time dependent fluctuating quantity around the mean value of zero. However, here the fluctuations in the effective attractive potential are thermally driven, which means that the free energy barrier associated with such fluctuations will be in the order of k_BT, where T is the absolute temperature. In this context, as we have already discussed in the previous sections, for an efficient one-dimensional search for the specific site on the DNA lattice by a protein molecule, the magnitude of the sequence dependent potential $F_h(x)$ as well as the effective z dependent interaction $F_v(z)$ should be comparable (Refs. [17,18]) with the thermal energy (k_BT) of the system and therefore one can assume $F_h(x)[k_BT]^{-1} \leq 1$ or $F_h(x) \approx 0$ under physiological conditions. Similarly when the degree of condensation of the DNA molecule under consideration is very high, then the probability distribution $p_s(y)$ associated with the one-dimensional diffusion coefficient variable y will be flat as $p_s(y) = [y_{\max} - y_{\min}]^{-1}$ and therefore $F_d(y) \approx 0$, i.e., the dynamics of the protein molecule over the y variable will be equally probable in the interval $y_{\min} \leq y \leq y_{\max}$. However, one should note that this assumption is not valid when the DNA is manually stretched or in a linear conformation. Now using these values in Eq. (5) and using the separation ansatz $T_{x,y} = T_x + T_y$ the general solution for $T_{x,y}$ can be given as follows:

$$T_{x,y} = \alpha[2^{-1}mx^2 + C_1x + C_2 - y_m^{-1}(6^{-1}my^3 - y^2) + C_3y + C_4]. \tag{7}$$

Here C_i, where $i = 1,2,3$ and m are constants to be determined from the boundary conditions. Now we can consider three different cases depending on the magnitude of the effective attractive potential $F_v(z)$ and the relative magnitudes of T_z and $T_{x,y}$ as follows.

V. PARTICULAR SOLUTIONS TO THE MPFT PROBLEM

Case I. The effective attractive potential $F_v(z)$ is strong enough to keep the protein molecule under the nonspecifically bound condition, i.e., $F_v(z)$ is such that $T_z > T_{x,y}$. However the protein molecule can dissociate from the DNA upon encountering the helical ends $x = \{0, N\}$ of the DNA lattice since the symmetry associated with the dynamical variable x will be broken at the helical ends, i.e., the helical ends $x = \{0, N\}$ of the DNA lattice act as absorbing boundaries for the dynamics of the protein molecule over the x variable. Then the corresponding boundary conditions for Eq. (7) become as follows:

$$T_{x,y}|_{x=0} = T_{x,y}|_{x_0=N} = \partial_y T_{x,y}|_{y=y_{\min}} = \partial_y T_{x,y}|_{y=y_{\max}} = 0. \tag{8}$$

Now defining the mean value of the time dependent phenomenological diffusion coefficient y as $\bar{y} = 2^{-1}(y_{\min} + y_{\max})$, and in the limiting condition $\lim_{y \to \bar{y}} T_{x,y}$, one can obtain an approximate solution to $T_{x,y}$ for the boundary conditions given by Eq. (8) as follows:

$$\lim_{y \to \bar{y}} T_{x,y} = T_{x,\bar{y}} \approx \bar{y}^{-1} x(N - x). \tag{9}$$

Here one should note that $T_{x,y} > T_{x,\bar{y}}$ when $y < \bar{y}$ and $T_{x,y} = T_{x,\bar{y}}$ when $y \geq \bar{y}$ and the MFPT ($T_{\bar{x},\bar{y}}$) averaged over the variable x can be given as follows:

$$T_{\bar{x},\bar{y}} \approx N^{-1} \int_0^N T_{x,\bar{y}} dx = 6^{-1} \bar{y}^{-1} N^2. \tag{10}$$

Equation (10) clearly states that in the limiting condition $\lim_{y \to \bar{y}} T_{x,y}$, the x averaged MFPT associated with the escape of the protein molecule through one of the helical ends of the DNA lattice is only dependent on the average value (i.e., $\bar{y}$) of the time dependent diffusion coefficient and it is independent of the initial value of the diffusion coefficient y as well as the variance y_m associated with the diffusion coefficient variable y. Here one should note that the variable y in Eq. (7) is simply the diffusion coefficient associated with the dynamics of the protein molecule in the three-dimensional space since the protein molecule was undergoing a three-dimensional diffusion process at time $t = 0$.

Case II. The effective attractive potential $F_v(z)$ is moderately weak, so that the protein molecule under nonspecifically bound conditions will dissociate from the DNA lattice after scanning an average of L number of base pairs, i.e., $F_v(z)$ is such that $T_z < T_{x,y}$. When $T_z < T_{x,y}$, and if we define the ratio $T_z/T_{x,y}$ as $\theta = T_z/T_{x,y}$, then one can easily show that the mean value of diffusion length L will be such that $L = N\theta = NT_z/T_{x,y}$. Suppose if the protein molecule nonspecifically binds at the position l_0 at time $t = 0$ where l_0 is such that the inequality conditions $0 < l_{\min} < l_0 < l_{\max} < N$ and $|l_{\min} - l_{\max}| = L < N$ holds true, then the mean time for which the nonspecifically bound protein molecule stays on the DNA lattice after each nonspecific contact can be given under the limiting condition $\lim_{y \to \bar{y}} T_{l_0,y}$ as follows:

$$\lim_{y \to \bar{y}} T_{l_0,y} = T_{l_0,\bar{y}} \approx \bar{y}^{-1} l_0(L - l_0). \tag{11}$$

Similar to Eq. (10), l_0 averaged $T_{l_0,\bar{y}}$ takes the following form:

$$T_{\bar{l}_0,\bar{y}} = L^{-1} \int_0^L T_{l_0,\bar{y}} dl_0 = 6^{-1} \bar{y}^{-1} L^2. \tag{12}$$

Here one should note that within the time $T_{\bar{l}_0,\bar{y}}$, when the inequality $T_z < T_{x,y}$ holds, the fraction of the DNA lattice scanned by the protein molecule is only $LN^{-1} = T_z/T_{x,y}$.

Case III. The attractive potential $F_v(z)$ is very strong so that the protein molecule under nonspecifically bound conditions will stay on the DNA lattice forever, i.e., $F_v(z)$ is such that $T_z \gg T_{x,y}$. Under this condition, one can calculate the

approximate mean time associated with the protein molecule to completely scan the DNA lattice after the occurrence of nonspecific contact as follows. From Eqs. (9) and (11) we can conclude that in the limiting condition $y \to \bar{y}$, where $y < \bar{y}$, Eq. (2) can be approximated as $\partial_t P_{x,t} = 2^{-1}\bar{y}\partial_x^2 P_{x,t}$. Since the protein molecule stays on the DNA lattice forever (i.e., practically for a long time and $T_z \gg T_{x,y}$), the helical ends of the DNA lattice i.e., $x=\{0,N\}$ should act as reflecting boundaries for the dynamics of the protein molecule over the x variable and therefore the probability distribution function $P_{x,t}$ associated with the dynamics of such a protein molecule with the initial condition $P_{x,t|x_0,0}=\delta(x-x_0)$ can be given as follows:

$$P_{x,t|x_0,0} = N^{-1} + 2N^{-1}\sum_{n=1}^{\infty} \cos(n\pi N^{-1}x_0)\cos(n\pi N^{-1}x) \times e^{-n^2(2N^2)^{-1}\pi^2\bar{y}t}. \quad (13)$$

Equation (13) can be obtained by solving the Fokker-Planck equation $\partial_t P_{x,t} = 2^{-1}\bar{y}\partial_x^2 P_{x,t}$ by the method of biorthogonal eigenfunction expansion with initial condition $P_{x,0|x_0,0}=\delta(x-x_0)$ and the reflecting boundary conditions $\partial_x P_{x,t}|_{x=0}=0$ and $\partial_x P_{x,t}|_{x=N}=0$. Equation (13) clearly states that under the limiting condition $\lim_{t>2N^2\bar{y}^{-1}\pi^{-2}} P_{x,t|x_0,0} \approx N^{-1}$, i.e., the time t_{III} taken by the protein molecule to scan the entire DNA lattice is simply given as $t_{III} \approx 2N^2\bar{y}^{-1}\pi^{-2} \approx T_{\bar{x},\bar{y}}$. Similarly when the inequality conditions $T_z > T_{x,y}$ and $T_z < T_{x,y}$ hold true, then the approximate probability distribution function $P_{x,t|x_0,0}$ associated with case I and II type dynamics of the protein molecule on the DNA lattice can be given as follows:

case I:

$$P_{x,t|x_0,0} = 2N^{-1}\sum_{n=1}^{\infty} \sin(n\pi N^{-1}x_0)\sin(n\pi N^{-1}x)e^{-n^2(2N^2)^{-1}\pi^2\bar{y}t}. \quad (14)$$

Equation (14) clearly states that the protein molecule stays on the DNA for a time period of $t_I \approx 2\bar{y}^{-1}\pi^{-2}N^2 \approx T_{\bar{x},\bar{y}}$ before it dissociates from the DNA lattice through one of the helical ends and since $t_I = t_{III}$ we can conclude that within the time t_I, the protein molecule can scan the entire DNA lattice.

case II:

$$P_{l,t|l_0,0} = 2L^{-1}\sum_{n=1}^{\infty} \sin(n\pi L^{-1}l_0)\sin(n\pi L^{-1}l)e^{-n^2(2L^2)^{-1}\pi^2\bar{y}t}. \quad (15)$$

From Eq. (15) we can conclude that the protein molecule stays on the DNA lattice for a time period of $t_{II} \approx 2L^2\bar{y}^{-1}\pi^{-2} \approx T_{l_0,\bar{y}}$, where $t_I \geq t_{II}$ and the fraction of the DNA lattice explored within the time t_{II} is only $LN^{-1} = T_z/T_{x,y}$.

Now let us assume that the specific site for the protein molecule is situated at the lattice position x_a such that $0 \leq x_a \leq N$. The mean time t_m required by the protein molecule to locate its specific site x_a after a nonspecific contact is simply given as $t_m \approx t_I = t_{III} \approx T_{\bar{x},\bar{y}}$ for case I and case III, and $t_m \approx NL^{-1}[t_{II} + (k_{3d}N)^{-1}]$ for case II type dynamics of the protein molecule on the DNA lattice. The target finding time t_m for case II can be derived as follows. Since the DNA of N base pairs in length can have a maximum NL^{-1} number of nonoverlapping segments of length L, the probability associated with the protein molecule to find its target site after each nonspecific contact with the DNA lattice is LN^{-1}. Therefore, the protein molecule would require at least $NL^{-1}t_{II}$ amount of time to scan the entire DNA lattice which means that a minimum of NL^{-1}association-dissociation events are necessary. However, each dissociation-association event is separated by a time lag of $(k_{3d}N)^{-1}$, i.e., the time required by the protein molecule to make a nonspecific contact with the DNA lattice via a three-dimensional route where k_{3d} is the three-dimensional diffusion-controlled collision rate. Upon including this time lag between two successive dissociation-association events we finally obtain the expression for the time that is taken by the protein molecule to scan the entire DNA lattice after making the first nonspecific contact via case II type dynamics as follows:

$$t_m \approx NL^{-1}[t_{II} + (k_{3d}N)^{-1}]. \quad (16)$$

VI. RESULTS AND DISCUSSION

Now we derive the scaling laws and inequality conditions which are required to enhance the three-dimensional association rate to an arbitrary order as follows. When the protein molecule targets its specific site simultaneously via both by three-dimensional as well as one-dimensional routes, the overall site specific association rate k_{ov} can be given as follows (Ref. [16]):

$$k_{ov} = k_{3d}[1 + N(1 + k_{1d}^{-1}k_{3d}N)^{-1}]. \quad (17)$$

Here k_{3d} is the three-dimensional specific-site finding rate and k_{1d} is the one-dimensional specific-site finding rate which is $k_{1d} \approx t_m^{-1}$ in the present context, where t_m is the mean time required by the protein molecule to scan the entire DNA lattice after making the first nonspecific contact with the DNA lattice. Suppose if we denote the number of times the overall site-specific association rate k_{ov} needs to be enhanced over the three-dimensional diffusion controlled rate k_{3d} as $\varepsilon = k_{ov}k_{3d}^{-1}$, then we obtain the inequality condition $\varepsilon \leq [1 + N(1 + k_{1d}^{-1}k_{3d}N)^{-1}]$. When the protein molecule follows case I or case III type dynamics on the DNA lattice, then as we have already derived in the previous section, we can write $k_{1d}^{-1} = 2\pi^{-2}\bar{y}^{-1}N^2$. Since the mean value of the phenomenological one-dimensional diffusion coefficient $\bar{y}$ can also be expressed as $\bar{y} = k_{3d}3^{-1}k(k+1)(2k+1) \approx 2k_{3d}3^{-1}k^3$ (e.g., see Ref. [19]), where k is the average jump size associated with the dynamics of the nonspecifically bound protein molecule (Refs. [19,24–27]) on the DNA lattice, the inequality becomes

$$k_{1d} \geq \{Nk_{3d}[N - (\varepsilon - 1)]^{-1}(\varepsilon - 1)\}. \quad (18)$$

When $N \gg \varepsilon$, one can conclude that to facilitate the three-dimensional diffusion controlled site-specific association rate

k_{3d} over ε times (i.e., $k_{3d} \mapsto \varepsilon k_{3d}$ where $\varepsilon > 1$), the one-dimensional target finding rate should be such a way that $k_{1d} \geq k_{ov}\varepsilon$. Denoting the average jump size k required to attain ε times of rate enhancement over the three-dimensional diffusion controlled rate as k_ε, one can show that the following inequality $k_\varepsilon \geq \{3^{1/3}\pi^{2/3}N(\varepsilon-1)^{1/3}[N-(\varepsilon-1)]^{-1/3}\}$ holds true. When $\varepsilon \gg 1$ and $N \gg \varepsilon$, this inequality simplifies to $k_\varepsilon \geq (3^{1/3}\pi^{2/3}\varepsilon^{1/3}N^{2/3})$, i.e., the average jump size required to enhance the diffusion controlled site-specific association rate of the protein molecule with the DNA molecule scales with the size of the DNA as $k_\varepsilon \propto N^{2/3}$. Since k_ε should be such a way that $k_\varepsilon \leq N$, the inequality $k_\varepsilon \geq (3^{1/3}\pi^{2/3}\varepsilon^{1/3}N^{2/3})$ implies that $1 \leq \varepsilon \leq 3^{-1}\pi^2 N$, i.e., the overall target finding rate k_{ov} is such that $k_{3d} \leq k_{ov} \leq 3^{-1}\pi^2 N k_{3d}$. However, when the protein molecule follows a case III type dynamics, then it has been shown earlier (Refs. [24–26]) that there exists a critical jump size $k_c \approx 2N^{2/3}$ beyond which the site-specific association rate cannot be enhanced, implying that $\varepsilon \leq 3^{-1}2^3\pi^2$ or $\varepsilon \leq 30$ as predicted by the earlier studies (Ref. [19]). When the protein molecule follows case II type dynamics, then one can derive the inequality condition $k_{3d} \leq k_{ov} \leq 3^{-1}\pi^2 L k_{3d}$, where $0 < L \leq N$. Now noting the definition $\varepsilon = k_{ov}k_{3d}^{-1}$ and substituting the equalities $k_{1d}^{-1} \approx NL^{-1}[t_{II} + (k_{3d}N)^{-1}]$, $N-(\varepsilon-1) \approx N-\varepsilon$, and $(\varepsilon-1) \approx \varepsilon$ for the case II type dynamics into the required inequality condition given by Eq. (18), we finally obtain the inequality condition $[2\varepsilon k_{3d}N(\pi^2\bar{y})^{-1}L^2 - L + \varepsilon] \leq 0$, which indicates that to achieve ε times enhancement of the site-specific association rate over the three-dimensional diffusion controlled rate, the average diffusion length L associated with the dynamics of the nonspecifically bound protein molecule on the DNA lattice should be such a way that $L_{\min} < L < L_{\max}$ where $L_{\min}$ and $L_{\max}$ are the two roots of the quadratic equation $2\varepsilon k_{3d}N(\pi^2\bar{y})^{-1}L^2 - L + \varepsilon = 0$. One also should note that the average diffusion length L_{ave} which is required to enhance the three-dimensional site-specific association rate to an arbitrary order of ε times, can be given as $L_{ave} = 2^{-1}[L_{\min} + L_{\max}] \approx (4\varepsilon k_{3d}N)^{-1}\pi^2\bar{y}$, which means that the diffusion length L is also a time dependent random quantity and it is also directly proportional to the diffusion coefficient y as $L_{ave} \propto \bar{y}$ or simply $L \propto y$.

Similar to the stationary state distribution function $p_s(y)$ for the y variable, the stationary state probability distribution function $p_s(L)$ associated with the one-dimensional diffusion length L is also flat and therefore one can conclude that $p_s(L) = [L_{\max} - L_{\min}]^{-1}$. Nevertheless, this is not valid when the DNA molecule under consideration is stretched or in a linear conformation. Here one should note that the recent experimental observations reveal (Ref. [24]) not only the existence of such $L_{\min}$ and $L_{\max}$ but also confirms the existence of $y_{\min}$ and $y_{\max}$ in the one-dimensional diffusion coefficient y as predicted by our generalized model. However, one cannot compare the predicted $L_{\min}$ and $L_{\max}$ with the observed values from the single molecule studies since all the experimental observations were carried out with a manually stretched DNA molecule (Ref. [27]). When the DNA molecule under consideration is in the stretched conformation, then the probability distribution $p_s(y)$ associated with the diffusion coefficient variable y will be dependent on y, i.e., $p_s(y)$ will not be a flat distribution and $p_s(y) \neq [y_{\max} - y_{\min}]^{-1}$, and therefore the probability distribution $p_s(L)$ associated with the diffusion length variable L will also be dependent on L, i.e., $p_s(L) \neq [L_{\max} - L_{\min}]^{-1}$. In this context, recent single molecule studies on the site-specific interaction of the protein molecule with a stretched DNA lattice showed (Ref. [27]) a Gaussian type distribution for the one-dimensional phenomenological diffusion coefficient y as well as the diffusion length L, which in turn suggests that the associated potential $F_d(y)$ is such that $F_d(y) \propto y^2$, i.e., a harmonic type potential of mean force confines the dynamics of the y variable, and therefore $p_s(y) \propto e^{-y^2}$ contrasting from our assumption that $F_d(y) \approx 0$ for a DNA lattice with higher degree of condensation. Since the one-dimensional diffusion length L is directly proportional to the one-dimensional diffusion coefficient y as $L \propto y$, one can conclude that $p_s(L) \propto e^{-L^2}$ since $p_s(y) \propto e^{-y^2}$ is similar to the observations in the recent single molecule studies (Ref. [27]) on the one-dimensional diffusion of the protein molecule in the process of searching for the specific-site on the stretched DNA lattice.

VII. CONCLUSIONS

According to the current view on site-specific DNA-protein interactions, the protein molecule first nonspecifically binds with the DNA lattice, and then searches for the specific site via various facilitating one-dimensional diffusion mechanisms such as sliding, hopping, and intersegmental transfers. Under such nonspecifically bound conditions, while the protein molecule is in the process of searching for the specific site, various static factors such as electrostatic attractive potential present between the positively charged amino acids and the negatively charged phosphate groups on the DNA backbone, water induced entropic effects, sequence dependent potential, and the condensation of the DNA molecule under consideration, and the dynamical factors such as local conformational fluctuations in the DNA molecule, the presence of other or the same types of DNA interacting molecules on the DNA lattice are acting on the protein molecule. Here we have developed a generalized theory which includes both the aforementioned static as well as dynamical factors influencing the one-dimensional diffusion of the nonspecifically bound protein molecule which is in the process of searching for the specific site on the DNA lattice. We have argued that the chemically driven condensation of the DNA molecule leads to the static distribution in the one-dimensional diffusion coefficient and the local conformational dynamics of the DNA molecule leads to the dynamical fluctuations in the one-dimensional diffusion coefficient over the static distribution. We further derived the generalized inequality conditions and the scaling laws which are required to enhance the three-dimensional diffusion controlled site-specific association rate to an arbitrary order.

Our model predicted that when the degree of condensation of the DNA molecule under consideration is very high, then the probability distribution associated with the stationary state one-dimensional diffusion coefficient variable as

well as the stationary state one-dimensional diffusion length variable will be a flat one. Further analysis revealed that to achieve a site-specific association rate which is higher than that of the three-dimensional diffusion controlled rate the one-dimensional diffusion length associated with the dynamics of the nonspecifically bound protein molecule on the DNA lattice should fall in certain critical ranges. Comparison of our theoretical results with the recent experimental observations has revealed that when the DNA molecule is under a stretched condition, then the static distribution of the one-dimensional diffusion coefficient associated with the dynamics of the protein molecule on the DNA lattice is a Gaussian and therefore the fluctuations in the one-dimensional diffusion coefficient generated by the dynamical factors are confined in a harmonic type potential.

ACKNOWLEDGMENTS

The author carried out a part of this work in the Muñoz group at the University of Maryland, College Park, and sincerely thanks Professor Victor Muñoz for offering a research position and the financial support.

[1] B. Levin, *Genes VIII* (Prentice-Hall, Englewood Cliffs, NJ, 2003); M. Ptashne and A. Gann, *Genes and Signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
[2] G. Adam and M. Delbruck, *Structural Chemistry in Molecular Biology* (Freeman, San Francisco, 1968).
[3] A. D. Riggs, S. Bourgeois, and M. Cohn, J. Mol. Biol. **53**, 401 (1970).
[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. F. Sneddon, D. J. Tobias, and C. L. Brooks III, J. Mol. Biol. **209**, 817 (1989).
[7] H. Ohtaka, A. Schon, and E. Freire, Biochemistry **42**, 13659 (2003); Y. Harano and M. Kinoshita, Biophys. J. **89**, 2701 (2005).
[8] R. Murugan, Biophys. Chem. **116**, 105 (2005).
[9] N. M. Luscombe, R. A. Laskowski, and J. M. Thornton, Nucleic Acids Res. **29**, 2860 (2001).
[10] O. G. Berg and C. Blomberg, Biophys. Chem. **4**, 367 (1976).
[11] O. G. Berg, Chem. Phys. **31**, 47 (1978).
[12] S. E. Halford and M. D. Szczelkun, Eur. Biophys. J. **31**, 257 (2002).
[13] R. F. Bruinsma, Physica A **313**, 211 (2002).
[14] I. M. Sokolov, R. Metzler, K. Pant, and M. C. Williams, Biophys. J. **89**, 895 (2005).
[15] H. X. Zhou, Biophys. J. **88**, 1608 (2005).
[16] S. E. Halford and J. F. Marko, Nucleic Acids Res. **32**, 3040 (2004).
[17] M. Slutsky and L. A. Mirny, Biophys. J. **87**, 4021 (2004).
[18] M. Slutsky, M. Kardar, and L. A. Mirny, Phys. Rev. E **69**, 061903 (2004).
[19] R. Murugan, Phys. Rev. E **69**, 011911 (2004).
[20] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 2002).
[21] H. Risken, *Fokker-Planck Equations* (Springer, Berlin, 1996).
[22] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 2004).
[23] S. Redner, *A Guide to First Passage Processes* (Cambridge University Press, Cambridge, England, 2001).
[24] R. Murugan, Phys. Rev. E **73**, 051915 (2006).
[25] R. Murugan, Biophys. Chem. **120**, 143 (2006).
[26] R. Murugan, J. Phys. A **39**, 1575 (2006); **39**, L199 (2006).
[27] Y. M. Wang, R. H. Austin, and E. C. Cox, Phys. Rev. Lett. **97**, 048302 (2006).