Packaging effects on site-specific DNA-protein interactions

Rajamanickam Murugan*
Department of Biotechnology, Indian Institute of Technology Madras (IITM), Chennai, Tamil Nadu 600036, India
(Received 29 January 2009; revised manuscript received 26 April 2009; published 19 June 2009)

We show that the rate of site-specific association of a protein molecule of interest with the DNA chain can be $\sim 10^2$ times higher than that of the three-dimensional diffusion-controlled collision rate limit $\sim 10^8$ $\mathrm{mol}^{-1}\,\mathrm{s}^{-1}$ only when the protein molecule of interest searches for its specific site on the DNA chain in a reduced dimensional space with a dimensionality d_r of $d_r < 1$. Upon considering the concurrent dynamics of the linear DNA chain that is embedded in a d-dimensional space along with the one-dimensional diffusion dynamics of the nonspecifically bound protein molecule on the DNA chain, we derive the generalized scaling law $\varepsilon \sim 2^{3(2-d)+3}$, where ε is the number of times by which the rate of site-specific association of the protein molecule with the DNA chain can be enhanced over the three-dimensional diffusion-controlled collision rate limit and d is the dimensionality of the reduced search space. Using the analogy between the self-intersection loop length in the theory of random walks and the ring-closure events in the theory of site specific interactions of a protein molecule with the DNA chain, we further show that the extent of packaging and volume compression of the genomic DNA inside the living cell is designed in such a way that the efficiency of the protein molecule in the process of searching for its specific site on the genomic DNA is a maximum. Our simulation results suggest that the volume compression factor θ which is the ratio between the total volume of the living cell and the volume occupied by the DNA chain along with all the other bound protein molecules should be such that $\theta \geq 100$ for an efficient site specific interaction of a protein molecule of interest with the linear DNA chain that is embedded in a three-dimensional space. Our theoretical and simulation results agree well with the *E. coli* cellular system.

DOI: 10.1103/PhysRevE.79.061920 PACS number(s): 87.10.−e, 87.15.kj

I. INTRODUCTION

Site-specific interaction of a protein molecule with the DNA lattice is a fundamental process in molecular biology and biological physics [1]. Transcription of the genomic DNA starts with the recognition of the promoter sequences by the RNA polymerase (RNAP) enzyme complex (RNAP in prokaryotes and RNAPII in eukaryotes), and replication of the genomic DNA starts with recognition of the origin of replication by the DNA polymerase enzyme complex. It was believed earlier that the site-specific interaction of a protein molecule with the DNA chain is a single-step three-dimensional diffusion-controlled bimolecular rate process. However detailed experimental studies [2–4] on the interaction of the *lac*-repressor protein with its specific operator sequence on the DNA chain showed a bimolecular association rate of $\sim 10^{10}$ $\mathrm{mol}^{-1}\,\mathrm{s}^{-1}$ which is $\sim 10^2$ times higher than that of the three-dimensional diffusion-controlled collision rate limit of $\sim 10^8$ $\mathrm{mol}^{-1}\,\mathrm{s}^{-1}$. Clearly this observation ruled out the possibility of a single-step recognition mechanism which is exclusively based on the three-dimensional diffusion process. It seems that such higher site-specific association rates could originate when the protein molecules are searching for their specific sites on DNA in a reduced one- or two-dimensional space [2]. Based on this idea, a two-step model on the site-specific interaction of the protein molecule of interest with the DNA lattice was proposed. According to this model, the protein molecule of interest nonspecifically binds with DNA via three-dimensional diffusion in the first step, and then it searches for its specific site via one-dimensional diffusion along the template DNA under nonspecifically bound conditions [3,4] in the second step.

The bimolecular rate associated with the nonspecific interaction step which is mediated by the three-dimensional diffusion is directly proportional to the total length of the DNA under consideration. When the length of the DNA molecule of interest is N base pairs (bps), then the maximum achievable bimolecular rate for the nonspecific interaction step can be of the order of $\sim N \times 10^8$ $\mathrm{mol}^{-1}\,\mathrm{s}^{-1}$. This follows from the fact that the DNA chain of N bps in length has at least $(N-m-1)$ numbers of nonspecific binding sites for the protein molecule of interest whose recognition stretch on the DNA chain under consideration is only m bps. In real situations generally $N \gg m$ since [1] the genome size will be in the order of $N \sim 10^6$ bps for bacteria such as *E. coli*, and the recognition stretch of the protein molecule will be in the order of $m \sim 10^2$ bps. Particularly for the RNAP enzyme complex the length of the recognition [1] stretch m on the DNA chain is $m \sim 60$ bps. One should note that the protein molecule of interest can also find its specific site on the DNA chain even in the first nonspecific interaction step with a maximum achievable rate of $\sim N \times 10^8$ $\mathrm{mol}^{-1}\,\mathrm{s}^{-1}$. However the probability or steric factor of such a site specific bimolecular collision event is very low as $\sim(1/N)$. The one-dimensional searching step is the rate limiting one which decides how many times by which the overall site-specific association rate can be higher than that of the three-dimensional diffusion-controlled collision rate. Assume that the protein molecule has already made a nonspecific contact with the DNA chain and it is currently scanning the genomic DNA for its specific site via one-dimensional diffusion-mediated random search with unit bps step size. Under this

*rmurugan@gmail.com

condition, using the results from the theory of random walks and mean first passage time (MFPT) calculations one can show [3,4] that the average search time τ_s that is required by such a protein molecule to locate its specific site on the DNA chain via one-dimensional diffusion with unit base-pair step size scales with the size of DNA such as a stiff rod as $\tau_s \propto N^2$, and this type of searching is not an efficient one. Earlier studies suggested that the efficiency of the one-dimensional diffusion-mediated searching process could be enhanced [3,4] by various facilitating processes such as correlated sliding, hopping, and intersegmental transfers via ring-closure events. These facilitating processes eventually increase the one-dimensional diffusion coefficient that in turn decreases the time that is required by the protein molecule to locate its specific site on the genomic DNA. In the presence of either these facilitating processes or a linear-type free-energy potential (a correlated walk) which induces a drift motion that favors the dynamics of the protein molecule towards its specific binding site, the overall search time τ_s that is required by the nonspecifically bound protein molecule to locate its specific site on DNA via one-dimensional diffusion-mediated search scales with the total length of the DNA chain as $\tau_s \propto N$.

The molecular structure, spatial arrangement, and the extent of packaging or compression of the genomic DNA inside the living cell seem to be well designed in such a way that the overall efficiency of the protein molecule in locating its specific site on the genomic DNA is a maximum. The genomic structure in prokaryotes has been shown to be designed such that the genes corresponding to the DNA binding proteins (DBPs) and their respective specific sites on the genomic DNA are generally colocalized [5]. This ensures faster searching of DBPs for their specific sites on the genomic DNA within physiologically reasonable time scales. Eukaryotic genomes are more complex than that of the prokaryotes and also larger in size. Additional driving factors such as *cis*-acting regulatory elements (enhancers) are required for efficient RNAPII-promoter interactions and subsequently for the initiation of transcription of the gene of interest in eukaryotes. Apparently these *cis*-acting elements can increase the probability of transcription of the associated genes only when the jump size associated with the dynamics of the corresponding transcription factors (TFs) on the genomic DNA is equal to or higher than a critical value k_c [6] that scales with the length of the genomic DNA N as $k_c \sim 2N^{2/3}$. Recently influence of the spatial organization of the DNA chain on the rate of site-specific interaction of the protein molecule with DNA has been studied [7] in detail.

In this paper we address an important question how the extent of compression or packaging of the genomic DNA inside the living cell influences the efficiency and rate of the site-specific interaction of a protein molecule of interest with DNA. The outline of this paper is as follows. We will derive a functional relationship between the degree of packaging or volume compression of the genomic DNA inside the living cell and the jump size associated with the dynamics of the nonspecifically bound protein molecules which are in the process of searching for their specific sites on DNA via ring-closure events. For this purpose we use the analogy between the self-intersection loop length in the theory of random walks and ring-closure events which are observed in the site-specific interactions of the protein molecules of interest with the DNA chain. We further derive a generalized functional relationship between the reduced dimensionality d and the number of times ε by which the rate of site-specific association of the protein molecule with DNA can be enhanced over the three-dimensional diffusion-controlled collision rate limit. We finally apply these results to *E. coli* bacterial system and demonstrate that the extent of packaging or compression of the genomic DNA of this bacterium is in such a way that the efficiency associated with the searching of the nonspecifically bound protein molecules for their specific sites which are present on the genomic DNA is a maximum.

II. THEORY

Consider a linear DNA polymer of N bps length that contains a specific binding site for a DBP. Initially this DBP was in solution. Upon colliding with the DNA chain which is mediated by a three-dimensional diffusion, the DBP of our interest nonspecifically binds with the DNA chain at a bimolecular collision rate of $\sim N \times 10^8$ $\mathrm{mol}^{-1}\ \mathrm{s}^{-1}$. Here the nonspecific interactions of the protein molecule with the DNA chain mainly originate from the electrostatic forces present in between the negatively charged phosphate backbone of the DNA chain and the positively charged side chains of the aminoacids which are present in the DNA-binding domains (DBDs) of DBPs. We assume that this electrostatic attractive potential is approximately an invariant quantity along the entire DNA chain. As a result, the nonspecifically bound protein molecule of our interest will stay in this electrostatic attraction domain that is closer to the DNA chain for a longer time. Under such nonspecifically bound conditions, the thermally driven diffusion dynamics of the protein molecule on the DNA chain will be mostly confined to the two-dimensional cylindrical surface of the DNA molecule. Since most of the DBPs such as RNAPII complex enfold the DNA chain upon their nonspecific binding, one can approximately assume the dynamics of the protein molecule of interest on the two-dimensional cylindrical surface of the DNA molecule as a one-dimensional diffusion dynamics along the DNA chain. Here one should note that the one-dimensional diffusion-mediated searching of the nonspecifically bound protein molecules for their specific sites on the genomic DNA is often interrupted by many dissociation and association events since the electrostatic forces are weakened by the solvent water molecules which are present at the interface of DNA-protein complex and also by the ions which are present in the bulk medium [8].

Under such conditions, the total free energy that is associated with all the bonding- and nonbonding-type nonspecific interactions which are present in between the protein molecule and the DNA chain can be assumed to be comparable with that of the thermal free energy ($RT = k_B N_A T \sim 0.591$ kcal mol^{-1} at $T = 298$K, where $R = k_B N_A \sim 1.986$ cal $\mathrm{mol}^{-1}\ \mathrm{K}^{-1}$ is the universal gas constant, k_B is the Boltzmann constant, N_A is the Avogadro number, and T is the absolute temperature in K), and one can ignore the finer details about the interactions at the interface of the nonspecifi-

cally bound protein molecule and the DNA chain. On the other hand, it seems that for an efficient site-specific association of a protein molecule with the DNA chain the free energy associated with all the bonding- and nonbonding-type nonspecific interactions should be comparable with that of the thermal free energy [10]. Upon making a nonspecific contact, the protein molecule of interest randomly scans the DNA chain for an average length of L bps for a time of τ_L [9–11], and then it dissociates from the DNA chain to reassociate back at a different or the same location. The above said condition can also be created artificially by manipulating the ionic strength of the medium in which the site specific association of the protein molecule of interest with the DNA chain is taking place. Here we use the fact that an increase in the ionic strength would eventually weaken the electrostatic forces which are present in between the protein molecule and the DNA chain [12]. That is to say, the dynamics of the protein molecule of interest on the DNA chain can be assumed as a one-dimensional diffusion of a random walker along a linear lattice that is embedded in a three-dimensional space or box under constant electrostatic potential along the DNA chain. Throughout this paper, we measure the dimensional quantities in terms of bps using the transformation rule 1 bps $\approx 3.4\times 10^{-10}$ m.

The total search time that is required by the protein molecule to locate its specific site on the DNA chain can be derived as follows. Let us denote the time that is required by the protein molecule of interest to make a nonspecific contact with the DNA chain via three-dimensional diffusion-mediated search as τ_{ns}. From the theory of random walks [11,12] and MFPT calculations one can derive the time that is required by the nonspecifically bound protein molecule to scan an average length of L bps of DNA as $\tau_L=L^2(6x_d)^{-1}$. If the protein molecule scans L bps upon each of its nonspecific contact with the template DNA, then the minimum amount of search time τ_s that is required by the protein molecule to locate its specific site on the DNA chain of N bps length can be given as follows:

$$\tau_s = NL^{-1}(\tau_{L-1} + \tau_{ns}) = \bar{\tau}_{1\mathrm{D}} + \bar{\tau}_{3\mathrm{D}}. \quad (1)$$

Here $\tau_L=L^2(6x_d)^{-1}$ s is the time taken [11,12] by the protein molecule to randomly scan an average length of L (bps), x_d(bps^2 s^{-1}) is the one-dimensional phenomenological diffusion coefficient associated with the dynamics of the protein molecule of interest on the DNA chain and the time that is required for the nonspecific binding is defined as $\tau_{ns}=(k_tN)^{-1}=\tau_tN^{-1}$ s, where $k_t=\tau_t^{-1}$ is the three-dimensional diffusion-controlled bimolecular collision rate ($\sim 10^8$ bps^{-1} s^{-1}). One should note that the prefactor NL^{-1} in Eq. (1) is the minimum number of dissociation and association events followed by a nonoverlapping mean-free-path length of L bps that is required by the protein molecule of interest to randomly scan the entire DNA chain. We have used τ_{L-1} in Eq. (1) mainly to account for the fact that the site of nonspecific contact on DNA will be automatically checked by the protein molecule of interest for its specific site and the protein molecule would need to search only L-1 bps among L bps. Many interesting results can be derived from Eq. (1) as follows:

(a) Overall the protein molecule spends $\bar{\tau}_{1\mathrm{D}}=NL^{-1}\tau_L$ amount of time in the one-dimensional diffusion-mediated random search along the DNA chain and $\bar{\tau}_{3\mathrm{D}}=NL^{-1}\tau_{ns}$ amount of time in the three-dimensional diffusion-mediated random search before locating its specific site on the DNA chain of interest.

(b) When there is no electrostatic attraction present between the protein molecule and the DNA chain, then we find that $L\rightarrow 1$ and Eq. (1) reduces to $\tau_s=(N\tau_{ns})=\tau_t$ which is the maximum average time that is required by the protein molecule to find its specific site on DNA via the three-dimensional diffusion-mediated search. Under this condition, the minimum achievable site specific association time is $\tau_s \sim \tau_{ns}$ since the protein molecule of interest can also find its specific site on the DNA chain upon making the first contact with a probability of $\sim(1/N)$. As a result one can conclude that the overall search time τ_s will be in such way that $\tau_t \geq \tau_s \geq \tau_{ns}$ which also means that $1\geq(\tau_s/\tau_t)\geq(1/N)$ after some algebraic manipulation. If we define the number of times by which the rate of site specific association of the protein molecule of interest with the DNA chain can be higher than that of the three-dimensional diffusion-controlled rate k_t as $\varepsilon=\tau_t/\tau_s$ (bps), then by inverting the inequality $1 \geq(\tau_s/\tau_t)\geq(1/N)$ we find that $1\leq\varepsilon\leq N$.

(c) When $L\gg 1$ Eq. (1) can be approximated as $\tau_s \sim NL^{-1}(\tau_L+\tau_{ns})$. Upon solving $\partial_L\tau_s=0$ for the optimum value of L, it can be shown [9] that the condition that is required to achieve an overall minimum search time is given as $\tau_{L_{opt}}=\tau_{ns}$ where $L_{opt}=\sqrt{6N^{-1}x_d\tau_t}$ and subsequently one finds that $\tau_{L_{opt}}=\tau_t/N$. From these results one can derive the minimum achievable search time for the protein molecule of interest to find its target site on DNA as $\tau_{s,\min}=2\tau_tL_{opt}^{-1}$ and the maximum achievable enhancement factor over the three-dimensional diffusion-controlled collision rate as $\varepsilon_{\max}=\tau_t\tau_{s,\min}^{-1}$. This also means that under such optimum conditions, the one-dimensional diffusion coefficient x_d can be expressed as $x_{d,o}=\varepsilon_{\max}^2 2N(3\tau_t)^{-1}$. From the theory of random walks [12,13] one can also find an expression for the one-dimensional diffusion coefficient as $x_{d,r}\approx\Sigma_{i=-k}^{+k}(i^2w_i)$, where k is the unbiased jump size associated with the dynamics of the protein molecule along the DNA chain. Here the jump size k that is directly proportional to the degree of condensation of the DNA chain means that the protein molecule of interest that is nonspecifically bound with the DNA chain can jump from its current position x to anywhere in the range of positions $x\pm k$ with equal probabilities which are given as $\phi_i=1/(2k)$ for $i=\pm 1,\pm 2,\ldots,\pm k$. Here the jumps associated with the protein molecule on the DNA chain are mainly driven by the ring-closure events in the three-dimensional space which are consequences of the condensation of the DNA polymer. In the context of site-specific interaction of a protein molecule with the DNA chain, the transition rates w_i associated with the jumping of the protein molecule of interest from one position of the DNA chain to another position of the same DNA chain are the three-dimensional diffusion-controlled collision rates as $w_i=(\phi_ie_i)=k_t$. Here $\phi_i=1/(2k)$ are the probabilities associated with the protein molecule to jump in an unbiased manner to the lattice positions $x\pm i$ starting from x and $e_i=2kk_t$ is the rate associated with such a

jump with jump size of k bps since it is similar to that of the nonspecific interactions between $2k$ numbers of nonspecific binding sites and a protein molecule of interest. Upon substituting these expressions in $x_{d,r}$, we find that $x_{d,r} \approx 2k^3(k_t/3)$ where $k \gg 1$ is the average unbiased jump size (bps) associated with the dynamics of the nonspecifically bound protein molecules which are in the process of searching for their specific sites on the DNA chain. Upon equating the expressions for the one-dimensional diffusion coefficients as $x_{d,o}=x_{d,r}$ and subsequently solving the equation for the jump size variable k, we find the optimum jump size $k_{\max}$ that is required to attain a maximum of the enhancement factor ε as $k_{\max}=\varepsilon_{\max}^{2/3}N^{1/3}$. Here one should note that $k_{\max}$ should be in such a way that $k_{\max} \leq N$ which means that $\{\varepsilon_{\max}^{2/3}N^{1/3}\} \leq N$. This follows from the fact that $k_{\max}$ cannot be higher than that of the length N of the DNA chain under consideration. Upon solving the inequality $\{\varepsilon_{\max}^{2/3}N^{1/3}\} \leq N$ for $\varepsilon_{\max}$ we recover our earlier result (b) as $\varepsilon_{\max} \leq N$. These results should also be true even when the linear DNA chain under consideration is embedded in an n-dimensional space. When the nonspecifically bound protein molecule randomly searches for its specific site along the linear DNA chain that is embedded in an n-dimensional space, then we have $x_{d,r,n} \approx 2^{n-1}\Sigma_{i=-k}^{+k} i^2 e_i \phi_i$ where $e_i=2kk_t$ is the overall rate associated with the transition and the probability associated with the protein molecule to jump to the site $x \pm i$ starting from the position x is $\phi_i=1/(2^n k)$ since there are $2k$ numbers of possible jumps where $i=1,2 \ldots k$ and each jump is associated with 2^{n-1} numbers of additional degrees of freedoms. As a result we will again have $x_{d,r,n}=x_{d,r}$. In deriving the expression for the one-dimensional diffusion coefficient $x_{d,r}$ we have assumed that the DNA chain is a static one which may not be correct. When the concurrent dynamics of the DNA chain is also considered, then the diffusion coefficient $x_{d,r}$ will be enhanced to a factor of at least 2 [14] as $x_{d,r} \to 2x_{d,r}$. Under this condition we have $k_{\max}=\varepsilon_{\max}^{2/3}(N/2)^{1/3}$ and subsequently upon insisting the condition that $k_{\max} \leq N$ and solving this inequality for $\varepsilon_{\max}$ we find that $\varepsilon_{\max} \leq \sqrt{2}N$ which explicitly follows from $\varepsilon_{\max}^{2/3}(N/2)^{1/3} \leq N$.

(d) When the nonspecific interaction that is present in between the protein molecule and the DNA chain is very strong, then $L \to N$ and Eq. (1) reduces to $\tau_s \sim (\tau_N + \tau_{ns})$. Upon substituting $x_d=x_{d,r}$ in the expression of τ_N in τ_s we get $\tau_s=\tau_t[(N^3+4k^3)/(4k^3N)]$. Under this condition, when $N \gg k$ the maximum achievable enhancement factor ε_a will be such that $\varepsilon_a \sim N^2/(4k^3)$. When we insist the condition that this maximum possible enhancement factor ε_a should be such that $\varepsilon_a \leq \varepsilon$, where ε is the required value of the enhancement factor for which the system needs to be tuned up, then upon solving the inequality $\{N^2/(4k^3)\} \leq \varepsilon$ for the jump size k we arrive at the following scaling relationship between the critical jump size $k_c(A)$ (bps) that is required to achieve a given preset site-specific association rate that is ε (bps) times higher than that of the three-dimensional diffusion-controlled collision rate and the size of the DNA chain N under consideration as follows [12]:

$$k_c(A) \geq \{(\varepsilon/4)^{1/3}N^{2/3}\}. \tag{2}$$

Equation (2) states that any required value of ε can be achieved by tuning the jump size $k_c(A)$ which is not correct in all the situations since it is meaningless to have a jump size such that $k_c(A) > N$ for a given DNA chain of N bps in length. In the following sections we will show that it is meaningless even when the jump size $k_c(A)$ is increased such that $k_c(A) > \{2N^{2/3}\}$. Since from the result (c) we have $x_{d,r,n}=x_{d,r}$, Eq. (2) is valid even when the DNA chain under consideration is embedded in an arbitrary d-dimensional lattice box and the random walk or jump of the protein molecule along such a DNA chain is considered instead of the random walks or jumps in a one-dimensional space. When the concurrent dynamics of the DNA chain is also considered, then we have $x_d \to 2x_{d,r}$ and Eq. (2) becomes as $k_c(A) \geq \{(\varepsilon/8)^{1/3}N^{2/3}\}$.

One can also derive the scaling law similar to the one given by Eq. (2) from a purely random-walk perspective as follows. Consider a linear lattice of N units in size where the lattice position $x=0$ is a reflecting boundary and the lattice position $x=N$ is the "only" absorbing boundary for the random walker which is present inside this lattice interval $(0,N)$. This means that the random walker can escape only through the lattice point $x=N$ and it will be reflected back into the interval $(0,N)$ whenever $x>N$ and $x<0$. We assume that the random walker was present initially at the lattice position $x=x_0$ at time $t=0$, where $0<x_0<N$, and it is currently at the lattice position x at time t. The dynamics of such a random walker can be well described by the Langevin-type equation as $d_t x=\sqrt{D}\xi_{x,t}$, where $\xi_{x,t}$ is the Gaussian white noise with the mean as $\langle \xi_{x,t} \rangle=0$ and the variance as $\langle \xi_{x,t}\xi_{x,t'} \rangle=\delta(t-t')$ and D is the one-dimensional phenomenological diffusion coefficient. The Fokker-Planck equation corresponding to the temporal evolution of the probability of finding the random walker at the lattice point x at time t which started from the lattice point $x=x_0$ at time $t=0$ can be written as $\partial_t P(x,t|x_0,0)=(D/2)\partial_x^2 P(x,t|x_0,0)$. Here the initial condition is given as $P(x,0|x_0,0)=\delta(x-x_0)$ and the boundary conditions are given as $[\partial_x P]_{x=0}=[\partial_x P]_{x>N}=[P]_{x=N}=0$. The MFPT associated with the escape of such a random walker through the lattice position N can be derived from the corresponding backward-type Fokker-Planck equation $x_d d_x^2 T(x)=-2$ with the similar boundary conditions $[d_x T]_{x=0}=[d_x T]_{x>N}=[T]_{x=N}=0$ as follows [12,13]:

$$T(x_0) = D^{-1}(N^2 - x_0^2). \tag{3}$$

Here we have $D=\Sigma_{i=-1}^{+1}(i^2 w_i)=1$ by definition, where $w_{\pm 1}=e_{\pm 1}\phi_{\pm 1}=(1/2)$. Since the three-dimensional diffusion-mediated transitions are not allowed in the present case, one should note that the rates of transitions associated with the random walker under consideration in a dimensionless form are defined as $e_{\pm 1}=1$ and the probabilities associated with these unbiased transitions are defined as $\phi_{\pm 1}=(1/2)$. When the three-dimensional diffusion-mediated transitions are also considered, then we have $e_{\pm 1}=2r_t$ where we have defined the dimensionless rate associated with the three-dimensional diffusion-mediated transition as r_t. We measure the MFPT $T(x_0)$ in terms of the dimensionless number of steps which is required by the random walker to find the lattice point N. In spite of an unit-step size assumption, if the random walker jumps on the linear lattice under consideration in an unbiased manner with an average step size of k, then the MFPT that is

required by such a random walker to escape only through the lattice point N starting from the lattice point x_0 can be derived as follows [14]. As we have already defined in the previous sections, here the jump size k means that the random walker can jump from its current position x to anywhere in the range of positions $x \pm k$ with equal probabilities which are given as $\phi_i = 1/(2k)$ for $i = \pm 1, \pm 2, \ldots, \pm k$. Assume that the current position of the random walker is $x=(N-1)$. In the next step the random walker can jump to any one of the possible positions $(N-1) \pm i$ with equal probabilities where $i=1,2,\ldots,k$. Among such $2k$ numbers of possibilities only one will be productive in the sense that the random walker can escape only when it hits the absorbing point N. Since all those trajectories hitting the lattice positions $(N-1)+i$, where $i=2,3,\ldots,k$ will be reflected back into to the same lattice interval $(0,N)$, the MFPT that is required by the random walker to hit these positions will be added up to the resultant MFPT with an appropriate weighting factors [14] as $\mu_i = i/k$. Upon summing over all the possible MFPTs with appropriate weighting, we finally arrive at the following expression for the overall MFPT $T_R(x_0)$ where $0 \leq x_0 < N$:

$$T_R(x_0) \approx D_k^{-1} \sum_{i=1}^{k} \mu_i [(N+i)^2 - x_0^2]$$

$$= D_k^{-1} T(x_0) + 2N + (3/2)k(k+1)(2k+1)^{-1}. \quad (4)$$

Here $D_k = \Sigma_{i=-k}^{+k} (i^2 w_i)$ where we have defined $w_i = (e_i \phi_i) = 1/(2k)$. Since the three-dimensional diffusion-mediated transitions are not allowed in the present case, the unbiased transition rates associated with the random walker of interest are defined as $e_i = 1$ and the unbiased transition probabilities associated with the random walker are defined as $\phi_i = 1/(2k)$. When $x_0 = 0$ and the jump size k is also sufficiently large, then we can obtain the following approximation for the overall MFPT as $T_R(0) \approx 3N^2 k^{-2} + 2N + (3/4)k$. Subsequently upon solving $\partial_k T_R(0) = 0$ for the jump size k, we find the critical or optimum value of the jump size $k = k_c$ that is required to attain an overall minimum value of MFPT as follows [we denote this as $k_c(B)$]:

$$k_c(B) \sim 2N^{2/3}. \quad (5)$$

Here the one-dimensional diffusion coefficient which is a function of the jump size k is defined explicitly as $D_k = 6^{-1}(k+1)(2k+1)$. From Eq. (5) we learn that by manipulating the jump size variable k, one cannot decrease the MFPT to zero. In other words, increasing the jump size variable k beyond certain level will not be an effective measure of increasing the efficiency or rate of the one-dimensional diffusion-mediated search process. This also means that the inequality condition given by Eq. (2) will be meaningful only when $k_c(A) \leq k_c(B)$ since in case of Eq. (2) we actually tune the one-dimensional jump size k to achieve the required value of the enhancement factor ε. One can derive the followings from Eqs. (2)–(5).

(a) At sufficiently larger jump sizes, the overall MFPT $T_R(x_0)$ associated with the escape of the random walker only through the lattice point N will be almost independent of the initial position x_0 of random walker on the linear lattice under consideration.

(b) There is a lower limit in the MFPT that is required by the random walker to escape only through the lattice point N as $\lim_{k \to k_c} T_R(x_0) \sim 2N$ beyond which it cannot be decreased by increasing or manipulating the jump size variable k.

(c) Comparison of Eqs. (2) and (5) suggests that when the protein molecule searches for its specific site on DNA in a reduced one-dimensional space, then the maximum achievable enhancement (ε) of the site-specific association rate over the three-dimensional diffusion-controlled collision rate limit is such that $\varepsilon \leq 32$. This inequality can be obtained by solving the inequality $\{(\varepsilon/4)^{1/3} N^{2/3}\} \leq \{2N^{2/3}\}$ for the enhancement factor ε that in turn follows from the inequality condition $k_c(A) \leq k_c(B)$ which is insisted on the one-dimensional diffusion-mediated search process so that it is meaningful. Here one should note that when the concurrent dynamics of the DNA chain is also considered along with the one-dimensional diffusion dynamics of the protein molecule of interest, then we will obtain the inequality $\{(\varepsilon/8)^{1/3} N^{2/3}\} \leq \{2N^{2/3}\}$. Subsequently upon solving this inequality for the enhancement factor ε we find that the maximum achievable enhancement factor ε will be such that $\varepsilon \leq 64$.

What are all the consequences of the observations (a)–(c) in the context of site-specific interaction of a protein molecule with the DNA chain? One can consider the DNA molecule as a linear lattice on which the protein molecule of interest searches for its specific site via one-dimensional diffusion. We have shown in the earlier sections that the overall site-specific association rate can be higher than that of the three-dimensional diffusion-controlled rate only when the rate of this one-dimensional search is faster than that of the three-dimensional diffusion-controlled collision rate. From the theory of random walks we learn that the one-dimensional search rate can be increased by increasing the jump size associated with the dynamics of the protein molecule on the DNA chain however only up to certain limit that is given by the critical jump size limiting condition in Eq. (5). Within this jump size limit, we learn from (c) that the overall site-specific association rate can be increased only up to ~ 32 times (~ 64 times upon including the concurrent dynamics of the DNA chain) higher than that of the three-dimensional diffusion-controlled rate. However experiments on the site-specific DNA-protein interactions showed an association rate that is $\sim 10^2$ times faster than that of the three-dimensional diffusion-controlled rate limit. This implies that the protein molecule of interest might search for its specific site on DNA possibly in a reduced d-dimensional space with a dimensionality of $d < 1$. One also should note that the jump size k is directly proportional to the degree of supercoiling or condensation of the DNA chain. This means that the rate of site-specific association of the protein molecule with the DNA chain can be enhanced by manipulating the spatial organization of the DNA chain only up to certain limit.

III. RESULTS AND DISCUSSION

One can also derive the critical jump size limit from the point of view of polymer confinement. Consider a linear

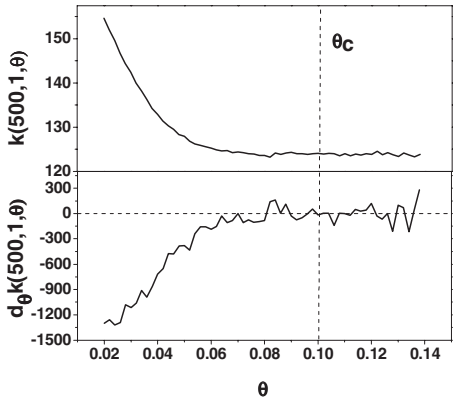

FIG. 1. Dependency of the jump size k on the volume compression parameter θ is shown in the upper panel. Here a polymer chain of 500 units in length is embedded in a one-dimensional lattice box with side length ranging from 10 to 70. We generalize the jump size as sojourn distance of random walker before visiting same site again via formation of self-intersection loops or ring-closure events. We sampled 10^5 such polymeric trajectories which are all starting from the left end of the lattice box and computed the average of these self-intersection loop lengths. As we increase the compression parameter θ, the average of the self-intersection loop length reaches the limiting value as $k(500,1,\theta>0.01)\rightarrow k_c(500,1)\sim 2\times 500^{2/3}$ when $\theta>0.01$. Lower panel is the numerical derivative of the average of the self-intersection loop lengths. This simulation result is inline with the critical jump size limit that is predicted by Eq. (5) for a one-dimensional case when the compression parameter is such that $\theta>\theta_c\sim 0.01$. Here both the variables k and θ are in dimensionless form.

polymer chain of N units in length which is "embedded" in a hypothetical one-dimensional lattice box of M units in length. Assume that this hypothetical polymer chain can self-intersect without excluded volume effect and also $M<N$. We insist this inequality constraint mainly to be consistent with the fact that the length of the genomic DNA is higher than the cellular dimensions. Now one can ask a question: what is the average of the loop length associated with the self-intersections of this embedded polymeric trajectory or trace in the one-dimensional lattice box under consideration particularly in the limit towards $(M/N)\rightarrow\infty$? Here the self-intersection loop length is the sojourn distance of the random walker (polymer) before revisiting a site again, and $\theta=(M/N)$ is the ratio of volume compression of the embedded polymer chain. One also should note that the ends of this one-dimensional lattice box are acting as reflecting boundaries for the embedded polymer. This problem can be easily simulated. Figure 1 summarizes the results from such simulation. The simulation settings are $N=500$, and M is iterated over a range of values. Here the average of the self-intersection loop length is taken over 10^5 such polymeric trajectories. All the trajectories start from the left end of the lattice box. Our simulation results demonstrate that when $(M/N)\geq\theta_c$ where $\theta_c\sim 0.01$ is some critical value of this volume compression ratio, the average of the loop length associated with the self-intersections of the embedded polymer is equal to the critical jump size limit given by Eq. (5) for the case of one-dimensional random walk. When the polymeric trajectory is embedded in or projected on a one-dimensional lattice box, then one can consider the jumps with jump size k as the ring-closure events since a self-intersection loop will be folded on itself upon projection. Here the jump size k can be thought as the maximum sojourn distance of the random walker along the one-dimensional lattice box before revisiting a site again.

To check the validity of Eq. (5) in higher dimensions we did random-walk simulations in various dimensions. We again considered a polymer of $N=500$ units in length. We embedded this polymer chain in cubical boxes of various dimensions and side lengths in such a way that each monomer or unit is confined in one volume unit of the d-dimensional lattice box under consideration. For example, in a three-dimensional lattice box, each step of the polymeric trajectory is fixed along the highest diagonal of the unit cube where the length of the diagonal is $\lambda_3=\sqrt{3}$. In general for a d-dimensional cube, each step of the polymeric trajectory is fixed along the highest diagonal length $\lambda_d=\sqrt{d}$ so that each polymeric unit occupies one volume unit in the d-dimensional cubical lattice box with side M. Here we defined the compression parameter as $\theta=V_B/V_N$, where $V_N=N=500$ volume units and $V_B=M^d$ is the total volume of the lattice box. For example, if we have a three-dimensional cubical box with sides equal to 10 units in length in which a polymer chain of 500 units in length is embedded, then we have $V_B=10^3$ and $V_N=500$. As a result we have $\theta=10^3/500=2$ ($\theta=1/5$ for the two-dimensional and $\theta=1/50$ for the one-dimensional cases) since each monomer of the embedded polymer chain occupies one volume unit (it is a single point in the one-dimensional, unit area in the two-dimensional and unit volume in the three-dimensional cases and so on). From the theory of DNA-protein interactions, we learn that the protein molecule can jump to a distal site mainly via ring-closure events [3]. In the random-walk terminology we can consider the self-intersections of the polymeric trajectory of DNA as ring-closure events and the mean value of the loop lengths of these ring-closures or self-intersections can be thought as the average jump size k. As a consequence, the jump size k should be a function of the dimensionality d, length N, and compression parameter θ as $k\rightarrow k(N,d,\theta)$. We are particularly interested in the limiting jump size as $k(N,d,\theta_c)\rightarrow k_c(N,d)$ at large values of θ as $\theta\geq\theta_c$. To check the existence of such a limit θ_c in various dimensions, we sampled the self-intersection loop lengths over 10^5 numbers of independent trajectories which were all starting from the origin (all with 500 steps in length), computed the average of the self-intersection loop length and plotted it as a function of the compression ratio θ. Figure 2 shows such a plot for various dimensions. We can summarize the results of this simulation study as follows:

(a) When $\theta>\theta_c$, then we observed the limit $k(N,d,\theta\geq\theta_c)\rightarrow k_c(N,d)$, where θ_c is a function of the dimensionality d. This critical value θ_c seems (Fig. 3) to be dependent on the dimensionality d in an exponential manner as $\theta_c\propto e^d$. For $d=3$, we observed a critical compression ratio of $\theta_c\sim 100$ irrespective of N (Fig. 4). This means that if the genomic DNA is packaged inside a three-dimensional living cell with a volume compression ratio of $\theta\geq\theta_c\sim 100$, then the protein molecules can locate their specific sites on DNA with a maxi-

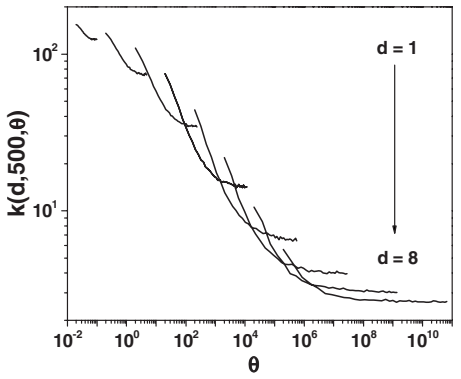

FIG. 2. Dependency of the average self-intersection loop lengths or jump sizes via ring-closure events on the compression parameter θ in various dimensions. Simulation results suggested that beyond certain critical values of compression parameter $\theta \to \theta_c$ associated with the polymer chain, the average of the self-intersection loop length tends to a limit in all the dimensions where the value of this limit is a function of the dimensionality d as $k(N,d,\theta > \theta_c) \to k_c(N,d)$. We find this limit explicitly as $k_c(N,d) \sim 2^{2-d}N^{2/3}$. In this simulation, we have embedded a polymer chain of $N=500$ units in length inside a d-dimensional lattice box with side length ranging from 10 to 70 and computed $k(N,d,\theta)$ in various dimensions. Here both the variables k and θ are in dimensionless form.

mum efficiency irrespective of the total size of the genomic DNA.

(b) Direct observation (Fig. 5) and nonlinear least-squares fitting results suggested a functional form $k_c(N,d) \sim 2^{1-d}k_c(N,1)$, where $k_c(N,1) \sim 2N^{2/3}$ as in Eq. (5). One can interpret this result as follows. From Eq. (5) we learn that when a one-dimensional chain is embedded in a one-dimensional lattice box, then the equilibrium value of the MFPT is attained at the critical jump size $k_c(B) \sim 2N^{2/3}$. When the same chain is embedded in a d-dimensional lattice box, then the degrees of freedom associated with each of the one-dimensional jump will be increased to $\sim 2^{d-1}$ times than that of the one-dimensional case. Particularly when a polymer of $N=500$ units is embedded in a one-dimensional lattice box, then we observed the critical jump size value of $k_c(500,1) \sim 126$. When the same chain is embedded in a d-dimensional lattice box, then the increase in the degrees of freedom of jumps in turn reduces the required critical jump size as $\sim 2^{-d+1} \times 126$. Since $k_c(N,1) = k_c(B)$ is a critical jump size for the case of one-dimensional random walk as given in Eq. (5), $k_c(N,3)$ can be thought as the observed or required critical jump size when the same linear polymer chain in embedded in a three-dimensional lattice box.

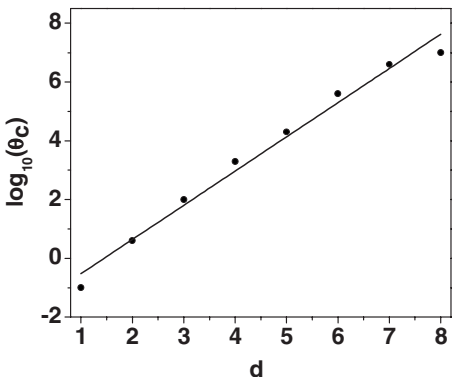

FIG. 3. Dependency of the critical compression ratio θ_c on the dimensionality d. Here a polymer chain of 500 units in length is embedded in a d-dimensional cubical lattice box. An iteration over θ is carried out at various dimensions. The value at which the derivative of the average self-intersection loop lengths with respect to θ vanishes is plotted as a function of the dimensionality d. Results suggested an approximate functional relationship as $\theta_c \propto e^d$. Solid line is a linear least-squares fitting with equation of type $\log_{10}(\theta_c) = -(1.7 \pm 0.3) + (1.2 \pm 0.03)d$ with $R^2 \approx 0.98$. Here both the variables d and θ_c are in dimensionless form.

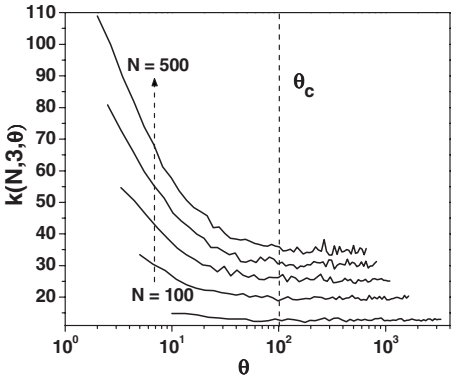

FIG. 4. This figure demonstrates that irrespective of the length of the polymer chain, the critical compression parameter is invariant in a given dimensional (d) space. Here the settings are dimensionality $d=3$ and polymer length N is iterated from 100 to 500. These polymer chains were embedded in a cubic lattice box with sides ranging from 10 to 70 and the average self-intersection loop lengths were computed over 10^5 such polymeric trajectories which were all starting from the origin. Here both the variables k and θ are in dimensionless form.

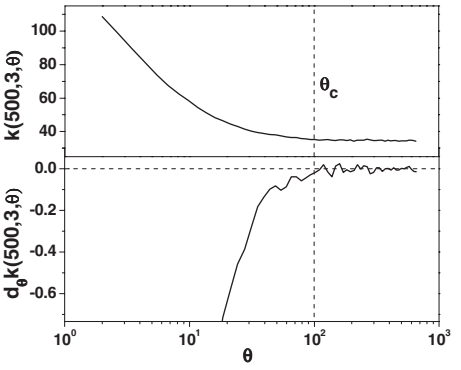

FIG. 5. Dependency of the jump sizes on the compression parameter θ in a three-dimensional lattice box. Here a polymer chain of 500 units in length is embedded in a three-dimensional lattice box with various side lengths ranging from 10 to 70. We sampled 10^5 such polymeric trajectories which were all starting from the origin of lattice box and computed the average of these self-intersection loop lengths. Lower panel is the numerical derivative of the average self-intersection loop lengths which suggested that $d_\theta k(500,3,\theta > \theta_c) \to 0$. This simulation result suggested a critical compression parameter of $\theta_c \sim 100$ for the three-dimensional case. Here both the variables k and θ are in dimensionless form.

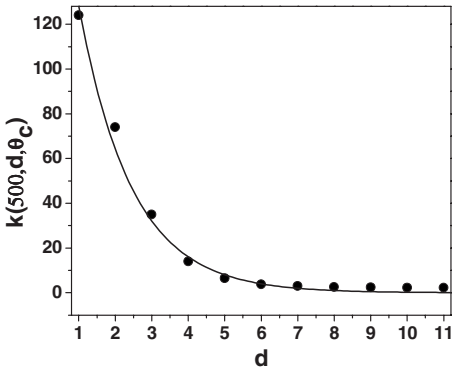

FIG. 6. Dependency of the critical jump size for a fixed polymeric length on the dimensionality d. Here a polymer chain of $N=500$ units in length is embedded in various d-dimensional cubic lattices with sides ranging from 10 to 70. The critical jump sizes were computed by averaging the self-intersection loop lengths over 10^5 polymeric trajectories which were all starting from the origin. Results suggested a functional form as $k_c(N,d) \sim 2^{1-d} k_c(N,1)$, where $k_c(N,1) \sim 2N^{2/3}$, and subsequently we have the limiting condition as $k(N,d,\theta_c) \rightarrow k_c(N,d) \sim 2^{2-d} N^{2/3}$. Here both the variables k and d are in dimensionless form.

(c) Here one should note that when the protein molecule searches for its specific site on the DNA chain via a combination of one and three-dimensional-mediated diffusion process, then the required one-dimensional jump size to achieve an enhancement of ε times over the three-dimensional diffusion-controlled collision rate is given by Eq. (2) as $k_c(A)$. Upon comparing the critical jump size $k_c(A)$ with the critical jump size $k_c(B)$ which is given by Eq. (5), we can conclude that the inequality $k_c(A) \leq k_c(B)$ should be true for a meaningful and tunable range of jump sizes. We find from the result (b) that this $k_c(B)$ transforms as $k_c(B) \rightarrow 2^{1-d} k_c(B)$ upon embedding the same polymer chain in a d-dimensional lattice box. As a consequence of our earlier result (c) from the theory section as $x_{d,r,n}=x_{d,r}$, the inequality $k_c(A) \leq k_c(N,d)$ should be still true for an arbitrary d that is given in an explicit form as $\{(\varepsilon/4)^{1/3} N^{2/3}\} \leq \{2^{2-d} N^{2/3}\}$. From this we get $(\varepsilon/4)^{1/3} \leq 2^{2-d}$. When there is a change in the dimensionality d at the right-hand side of this inequality, then there will be a concurrent change in the enhancement variable ε at the left side of the inequality so that the entire inequality is valid for all values of d and ε. Upon solving the inequality $(\varepsilon/4)^{1/3} \leq 2^{2-d}$ by taking logarithms at both the sides and then inverting it for the dimensional variable d, we arrive at the following generalized inequality condition (Fig. 6):

$$\{2 - \ln|(\varepsilon/4)^{(\ln 2)^{-1}/3}|\} \geq d. \tag{6}$$

(d) We find from Eq. (6) that the reduced dimensionality d of the search space that is required to achieve the enhancement of $\varepsilon=90$ over the three-dimensional diffusion-controlled collision rate as observed in many experimental studies [15] is $d_r \sim (1/2)$. For $d=1$ we have the inequality relationship as $\varepsilon \leq 32$. Here one should note that when the concurrent dynamics of the DNA chain is also considered, then Eq. (6) will become as $\{2-\ln|(\varepsilon/8)^{(\ln 2)^{-1}/3}|\} \geq d$. As a result, the reduced dimensional space (Fig. 6) that is required to achieve the observed enhancement factor of $\varepsilon=90$ in the presence of concurrent dynamics of the DNA chain is $d_r \sim 0.83$ and for $d=1$ we have the inequality relationship as $\varepsilon \leq 64$. Here one should note that the fractal dimension d_r that is required for achieving the enhancement factor $\varepsilon=90$ could originate from the sequence-dependent topology of the interaction energy landscape of the DNA chain and the protein molecule of interest. Here the interaction energy is the sum of energies associated with all the bonding and nonbonding interactions present in between the DBD of the protein molecule and its recognition stretch sequence on the DNA chain. Since the free energy associated with the interactions between the protein molecule and the DNA chain depends on the sequence, we can conclude that the free-energy landscape or profile associated with these nonspecific interactions might possess some sort of self-similar and fractal properties along the DNA sequence. We hypothesize that the protein molecule that is diffusing along the DNA chain might be capable of dynamically detecting or reading the information that is hidden inside this sequence-dependent fractal topology of the interaction energy landscape. The following observations seem to support our arguments indirectly. Firstly the structural studies on *lac*-repressor have revealed the presence of a significant amount of thermally driven conformational fluctuations in its DNA binding domains (DBDs) especially when this protein is nonspecifically bound with the DNA chain. Upon finding the *lac*-operator sequence on the template DNA which is its specific site, the DBDs of *lac*-repressor changes its conformation to more stable one with minimal degree of fluctuations [16]. This observation indirectly suggests that the protein molecule of interest might dynamically resonate with the sequence-dependent interaction energy landscape of the DNA chain via thermally driven conformational fluctuations. The efficiency of these thermally driven conformational fluctuations in the DBDs of the protein molecule to resonate with the sequence-dependent interaction energy landscape of the DNA chain seems to be strictly restricted by the second law of thermodynamics [17]. Apart from this observation, the sequence of the DNA chain is also known to be following a self-similar fractal-type pattern [18]. Further the self-similarity in the sequence pattern of the coding exons seems [18] to be different from that of the noncoding introns. Based on these observations, we further speculate that this self similar fractal pattern of the sequence of the DNA chain might be strongly connected with the thermally driven dynamic interactions of the protein molecule with the DNA chain. That is to say, the information content that is hidden inside the self-similar fractal pattern of the sequence-dependent interaction energy landscape of the DNA chain might be playing critical roles in accelerating the site-specific interaction of the protein molecule of interest with the DNA chain beyond the three-dimensional diffusion-controlled collision rate limit.

(e) On solving Eq. (6) for the enhancing factor ε we find the generalized inequality condition or scaling law as $\varepsilon \leq \{2^{3(2-d)+2}\}$. This means that by reducing or manipulating the dimensionality of the search space, the rate associated with the site-specific interaction of a protein molecule with the DNA chain can be enhanced to a maximum of $\sim 2^8$ times over the three-dimensional diffusion-controlled collision

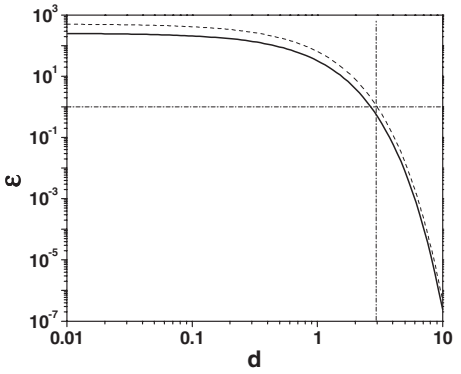

FIG. 7. Dependency of the number of times by which the rate of site-specific association of the protein molecule with the DNA chain can be enhanced over the three-dimensional diffusion-controlled collision rate on a reduced dimensional space with dimensionality of d in which the protein molecule of interest searches for its target site on DNA. The theoretically derived functional form is $\varepsilon \sim 2^{3(2-d)+2}$ (solid line). When the concurrent dynamics of the DNA chain is also considered, then our theory showed a functional form as $\varepsilon \sim 2^{3(2-d)+3}$ (dashed line). This result suggests that the maximum achievable enhancement of the site-specific association rate over the three-dimensional diffusion-controlled rate by reduction in the dimensionality of search space is $\varepsilon_{\max} \sim 2^8$. When the concurrent dynamics of the DNA chain is also considered, then this maximum limit will become as $\varepsilon_{\max} \sim 2^9$. Here both d and ε are in dimensionless form.

rate. This follows from the inequality relationship $\varepsilon \leq 2^8$ for the hypothetical limit $d \to 0$. Here one should note that the maximum achievable enhancement factor for the hypothetical limit as $d \to 0$ in the presence of concurrent dynamics of the DNA chain under consideration is $\varepsilon \leq 2^9$ which follows from the fact that (Fig. 7) $\varepsilon \leq \{2^{3(2-d)+3}\}$ under such conditions.

The genomic DNA of *E. coli* can be thought as an embedded linear polymer chain inside a closed three-dimensional cellular lattice box (Fig. 8). From our theory we can conclude that the critical compression ratio θ_c that is required to achieve the critical jump size condition in the three-dimensional space should be such that $\theta_c \geq 100$. We should note that when the volume compression ratio is such that $\theta_c \geq 100$, then the rate associated with the site-specific interaction of a protein molecule with the DNA chain will be a maximum. Using the *E. coli* cellular model system for which a detailed cellular statistics is available, we show that the nature has designed the volume compression ratio of the living cells inline with our theoretical predictions as follows (Fig. 8). From literature we find that [19] the average total volume of an *E. coli* cell is $V_C \sim 1.0 \times 10^{-18}$ m^3. The average volume occupied by the genomic DNA and all the other specifically and nonspecifically bound DBPs is $V_G \sim 1.6 \times 10^{-19}$ m^3, whereas the total size of the *E. coli* genome is $N_G \sim 4.6 \times 10^6$ bps [20]. Here the genomic DNA of length $L_G \sim 1.6 \times 10^{-3}$ m is confined inside a cylindrical shaped cellular box with a length of $l_c \sim 2 \times 10^{-6}$ m and a width of $w_c \sim 0.8 \times 10^{-6}$ m. For convenience we can measure these dimensional quantities in terms of bps using the transformation rule 1 bps $\approx 3.4 \times 10^{-10}$ m. According to this

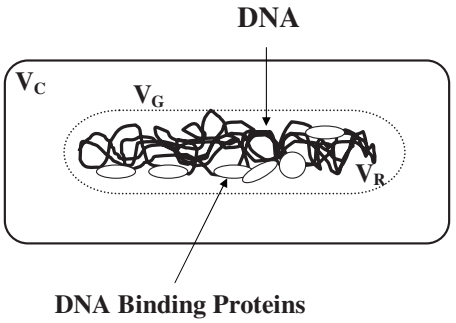

FIG. 8. Total volume of *E. coli* bacterial cell is $V_C \sim 1.0 \times 10^{-18}$ m^3. Volume occupied by the genome and bound DBPs is $V_G \sim 1.6 \times 10^{-19}$ m^3 and size of the genome is $N_G \sim 4.6 \times 10^6$ bps. Here genomic length of $L_G \sim 1.6 \times 10^{-3}$ m is confined inside a cylindrical-shaped cellular volume with length of $l_c \sim 2 \times 10^{-6}$ m and width of $w_c \sim 0.8 \times 10^{-6}$ m. Dotted line is the approximate boundary of the total volume occupied by the genomic DNA and all the specifically and nonspecifically bound DNA-binding proteins under normal cellular conditions. Using these values, we found the volume compression ratios as $\theta = V_G/V_N \sim 180$ and $\sigma = V_G/V_R \sim 1124$. These results demonstrate that the volume compression ratio for the genomic DNA of *E. coli* is closer to or beyond the critical value that is required to maximize the rate site specific interaction of a protein molecule with the genomic DNA for the three-dimensional case as $\theta = V_G/V_N \sim 180 > \theta_c(3)$.

rule we have the transformed values as $V_C \sim 2.54 \times 10^{10}$ bps^3 and $V_G \sim 4.07 \times 10^9$ bps^3. If we consider the genomic DNA as a cylindrical shaped object with diameter of $D_d \sim 5$ bps and length of $N_G \sim 4.6 \times 10^6$ bps, then the total volume occupied by only the genomic DNA is $V_R = \pi(D_d/2)^2 N_G \sim 22.6 \times 10^6$ bps^3. Using these values, we can find the compression ratio θ that is the ratio between the total volume occupied by both the genomic DNA and DNA-interacting proteins and the volume of only the genomic DNA as $\theta = V_G/V_N \sim 180$. Similarly we can also find that the ratio between the total cell volume and the volume occupied by the genomic DNA as $\sigma = V_G/V_R \sim 1124$. These results demonstrate that the compression ratio for the genomic DNA of *E. coli* is closer to or beyond the required critical value that is required for the three-dimensional case as $\theta = V_G/V_N \sim 180 > \theta_c$. In other words, the extent of packaging of the genomic DNA inside this bacterial cell satisfies the criteria $\theta \to \theta_c$ to achieve the critical jump size limit that in turn is required to achieve a maximum site-specific association rate. This result is in line with our theoretical predictions. This also means that the extent of packaging of the genomic DNA inside the *E. coli* bacterial cell under physiological conditions is such that the efficiency associated with the searching of the protein molecule for its target site on the genomic DNA is a maximum.

IV. CONCLUSIONS

In this paper, we have shown that the rate of site-specific association of a protein molecule of interest with the DNA chain could be $\sim 10^2$ times higher than that of the three-dimensional diffusion-controlled collision rate limit

$\sim 10^8$ mol^{-1} s^{-1} only when the protein molecule of interest searches for its specific site on the DNA chain in a reduced dimensional space with a dimensionality of $d_r < 1$. Upon taking the concurrent dynamics of the linear DNA chain that is embedded in a d-dimensional space into consideration along with the one-dimensional diffusion dynamics of the protein molecule on the DNA chain, we derived the generalized scaling law $\varepsilon \sim 2^{3(2-d)+3}$, where ε is the number of times by which the rate of site-specific association of the protein molecule with the DNA chain can be enhanced over the three-dimensional diffusion-controlled collision rate limit and d is the dimensionality of the search space. Using the analogy between the self intersection loop length in the theory of random walks and the ring-closure events in the theory of site specific interactions of a protein molecule with the DNA chain, we further showed that the extent of packaging or volume compression of the genomic DNA inside the living cell is in such a way that the efficiency of the protein molecule in the process of searching for its specific site on the genomic DNA is a maximum. Our simulation results suggested that the volume compression factor θ which is the ratio between the total volume of the living cell and the volume occupied only by the DNA chain along with other bound protein molecules should be such that $\theta \geq 100$ for an efficient site specific interaction of a protein molecule with the linear DNA chain that is embedded in a three-dimensional space. Our theoretical and simulation results agreed well with the *E. coli* cellular system.

ACKNOWLEDGMENTS

This work was supported by Indian Institute of Technology Madras (IITM). The author thanks IITM for Grant No. BIO/08-09/241/NFSC/RMUR.

[1] B. Levin, *Genes VIII* (Prentice Hall, New York, 2003); M. Ptashne and A. Gann, *Genes and Signals* (Cold Spring Harbor, New York, 2001).

[2] G. Adam and M. Delbruck, *Structural Chemistry in Molecular Biology* (Freeman, San Francisco, 1968).

[3] O. G. Berg, R. B. Winter, and P. H. von Hippel, Biochemistry **20**, 6929 (1981); R. B. Winter, O. G. Berg, and P. H. von Hippel, *ibid.* **20**, 6961 (1981); A. D. Riggs, S. Bourgeois, and M. Cohn, J. Mol. Biol. **53**, 401 (1970).

[4] R. F. Bruinsma, Physica A **313**, 211 (2002); 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); T. Hu and B. I. Shklovskii, Phys. Rev. E **74**, 021903 (2006).

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

[6] R. Murugan, J. Theor. Biol. **248**, 696 (2007).

[7] Z. Wunderlich and L. A. Mirny, Nucleic Acids Res. **36**, 3570 (2008).

[8] S. F. Sneddon, D. J. Tobias, and C. L. Brooks III, J. Mol. Biol. **209**, 817 (1989); H. Ohtaka, A. Schon, and E. Freire, Biochemistry **42**, 13659 (2003); Y. Harano and M. Kinoshita, Biophys. J. **89**, 2701 (2005).

[9] M. Slutsky and L. A. Mirny, Biophys. J. **87**, 4021 (2004).

[10] H. X. Zhou, Biophys. J. **88**, 1608 (2005); M. Slutsky, M. Kardar, and L. A. Mirny, Phys. Rev. E **69**, 061903 (2004).

[11] R. Murugan, Phys. Rev. E **69**, 011911 (2004); I. M. Sokolov, R. Metzler, K. Pant, and M. C. Williams, Biophys. J. **89**, 895 (2005).

[12] R. Murugan, Phys. Rev. E **76**, 011901 (2007).

[13] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 2002); H. Risken, *Fokker Plank 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, London, 2001).

[14] R. Murugan, Biophys. Chem. **120**, 143 (2006); J. Phys. A **39**, 1575 (2006); **39**, L199 (2006).

[15] Y. M. Wang, R. H. Austin, and E. C. Cox, Phys. Rev. Lett. **97**, 048302 (2006); N. Shimamoto, J. Biol. Chem. **274**, 15293 (1999).

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

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

[18] Y. Xiao, R. Chen, R. Shen, J. Sun, and J. Xu, J. Theor. Biol. **175**, 23 (1995); C. K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Halvin, F. Sciortino, M. Simons, and H. E. Stanley, Nature (London) **356**, 168 (1992); H. E. Stanley and P. Meakin, *ibid.* **335**, 405 (1988).

[19] D. S. Goodsell, Trends Biochem. Sci. **16**, 203 (1991); K. R. Albe, M. H. Butler, and B. E. Wright, J. Theor. Biol. **143**, 163 (1990); M. B. Elowitz M. G. Surette, P. E. Wolf, J. B. Stock, and S. Liebler, J. Bacteriol. **181**, 197 (1999); http://redpoll.pharmacy.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi

[20] J. G. Sutcliffe, Cold Spring Harb Symp. Quant Biol. **43**, 77 (1979).