Theory of Site-Specific DNA-Protein Interactions in the Presence of Conformational Fluctuations of DNA Binding Domains

R. Murugan*
Department of Biotechnology, Indian Institute of Technology Madras, Chennai, India

ABSTRACT We develop a theory that explains how the thermally driven conformational fluctuations in the DNA binding domains (DBDs) of the DNA binding proteins (DBPs) are effectively coupled to the one-dimensional searching dynamics of DBPs for their cognate sites on DNA. We show that the rate γ_{opt}, associated with the flipping of conformational states of DBDs beyond which the maximum search efficiency of DBPs is achieved, varies with the one-dimensional sliding length L as $\gamma_{opt} \propto L^{-2}$ and with the number of roadblock protein molecules present on the same DNA m as $\gamma_{opt} \propto m^2$. The required free energy barrier E_{RTO} associated with this flipping transition seems to be varying with L as $E_{RTO} \propto \ln L^2$. When the barrier height associated with the conformational flipping of DBDs is comparable with that of the thermal free energy, then the possible value of L under in vivo conditions seems to be $L \leq 70$ bps.

INTRODUCTION

Site-specific interaction of a protein molecule with the DNA chain in the presence of an enormous amount of nonspecific binding sites is a fundamental process in molecular biology and biological physics (1). This is evident from the fact that the basic processes in molecular biology such as the initiation of replication and transcription of the genomic DNA are based on the site-specific interactions of the DNA polymerase enzyme with the respective origin of replication and RNA polymerase enzyme with the respective promoter sequences of genes that are all located on the genomic DNA (1–4). It was thought earlier that the site-specific interactions of a protein molecule with the DNA chain are mediated via three-dimensional diffusion-controlled collision routes (2,3). Later experimental studies (2,3) on site-specific binding of the Lac repressor protein with its corresponding Operator sequence, which is located on a DNA chain, showed a bimolecular site-specific collision rate of $\sim 10^{10}$ mol^{-1} s^{-1}, which is $\sim 10^2$ times faster than that of the three-dimensional diffusion controlled rate in aqueous conditions $\sim 10^8$ mol^{-1} s^{-1}. Searching for the specific sites that are located on the DNA chain, by the corresponding protein molecules, via a combination of three-dimensional and (reduced) one-dimensional routes, could explain (2,3) these observed higher bimolecular collision rates.

Winter et al. (3) suggested that various facilitating processes such as sliding, hopping, and intersegmental transfers can enhance the rate of site-specific interactions of the protein molecule with the DNA chain over the three-dimensional diffusion-controlled rate limit. The protein molecule that is diffusing along the DNA polymer can randomly switch between different modes of these facilitating dynamics depending on the prevailing local environment. Here the sliding mode of dynamics indicates the diffusion of the protein molecule along the DNA chain with unit basepair (bps) step size whereas the protein molecule can leap over few bps at a time in the hopping-mode. These sliding and hopping modes dominate whenever the DNA molecule is somewhat stretched and loosely packed. On a highly condensed or super-coiled DNA chain, the diffusing protein molecule can undergo intersegmental transfers via ring closure events that can occur whenever two distal segments of the same DNA chain come closer upon condensation. The protein molecule can leap over a few hundreds to thousands of bps during these intersegmental transfer events. All these facilitating modes reduce the overall search-time that is taken by the protein molecule to locate its specific site on DNA mainly by fine-tuning the ratio of the search-times spent on one-dimensional and three-dimensional routes. Slutsky and Mirny (5) and Murugan (4) have shown that the minimum of this overall search-time can be achieved when the protein molecule spends equal amount of time both in the one-dimensional and three-dimensional routes. Detailed theoretical studies of Coppey et al. (6) and Lomholt et al. (7) as well as the single molecule experimental studies of van den Broek et al. (8), Sokolov et al. (9), Bonnet et al. (10), and Wang et al. (11) substantiated the ideas of Winter et al. (3) and further suggested that the spatial organization and packaging (4) of the DNA molecule can significantly enhance the rate of site-specific interactions of the protein molecule with DNA.

Recent experimental observations by Kalodimos et al. (12) and related theoretical studies of Hu et al. (13) revealed the presence of thermally driven conformational fluctuations in the DNA binding domains (DBDs) of the nonspecifically bound DNA binding proteins (DBPs). Upon finding the specific sites, these conformational fluctuations in the DBDs of DBPs are damped-out, which results in the formation of a tight site-specific DNA-protein complex. These

Submitted February 12, 2010, and accepted for publication April 13, 2010.

*Correspondence: rmurugan@gmail.com

Editor: David P. Millar.

0006-3495/10/07/0353/7 $2.00

doi: 10.1016/j.bpj.2010.04.026

results suggested (12,13) the presence of at least two different conformations of DBDs—namely the plus or fast-diffusing state, and the minus or slow-diffusing state. In the fast-diffusing state, the protein molecule is somewhat loosely packed, or less ordered in structure, so that it can freely slide along the DNA. This means that the fast-diffusing state is less sensitive to the DNA sequence on which it slides and the DBPs cannot distinguish their specific sites from the nonspecific sites whenever their DBDs are in the fast-diffusing state. This is in contrast to the slow-diffusing state, in which the protein molecule is more ordered in structure and closely associated with the DNA sequence. When DBDs are in the slow-diffusing state, the DBPs slowly diffuse along the DNA chain and tightly bind with DNA upon locating their specific sites. The DBDs of DBPs undergo thermally driven conformational fluctuations between these plus (+) and minus (–) states.

Apparently, in the presence of such thermally driven conformational fluctuations in the DBDs, the DBPs can efficiently locate their specific binding sites by freely flipping between one-dimensional (Fig. 1) and three-dimensional modes. Because the fluctuation-induced flipping between plus- and minus-states is thermally driven, the maximum enhancement of the efficiency of searching for the specific sites which are located on the DNA chain by the respective protein molecules seems to be strictly restricted by the second law of thermodynamics (thermodynamic limit) (12). Recently, such a limit has been calculated in detail (13), and it seems that the search-time taken by DBPs to locate their specific sites on DNA could be closer to this thermodynamic limit only when the energy spectrum of such conformational fluctuations in the DBDs of DBPs is tuned by the selective pressure in such a way that the overall site-specific binding time is minimum.

It is still not clear how these thermally driven conformational fluctuations in the DBDs of DBPs are efficiently coupled to the searching dynamics of DBPs as they move toward their specific sites on the DNA chain. In this context, it is also not clear how the sliding length associated with the dynamics of the nonspecifically bound DBPs is influenced by the rate that is associated with the thermally driven flipping between the conformational states of DBDs of DBPs.

It is, additionally, of great importance to reveal the optimum-flipping rate required to achieve the minimum search time associated with the protein molecule of interest, in locating its specific binding site for a given sliding length, in the presence of other roadblock protein molecules also present on the same DNA. In this article, we address these issues in detail.

THEORY

Consider a protein molecule that is searching for its specific binding site on DNA via a combination of one-dimensional and three-dimensional routes. Assume that the total length of the DNA chain is N bps. In line with two recent studies (12,13), we assume that the DBD of the protein molecule can exist in two possible states—namely, the fast-diffusing plus-state which is less sensitive to the DNA sequence, and the slow-diffusing minus-state, which is more sensitive to DNA sequence and binds tightly upon detecting its target site on DNA. Flipping between these two different states is a thermally driven process. We assume that these different states are characterized by distinct one-dimensional diffusion coefficients $D_{\pm}$ associated with them and $D_+ > D_-$. Upon making a nonspecific contact, the protein molecule scans the DNA chain for an average sliding length of L bps and then dissociates from DNA to reassociate back at the same or different location of the same DNA chain after a brief three-dimensional excursion. Such events are possible only within the capturing domain, which is characterized by the electrostatic attractive force field that is present in between the negatively charged phosphate backbone of the DNA chain and the positively charged side chains of the amino acids present at the DBDs of DBPs (Fig. 1).

The conformational transition (+) ⇄ (−) associated with the DBDs of DBPs is a stochastic process that can occur

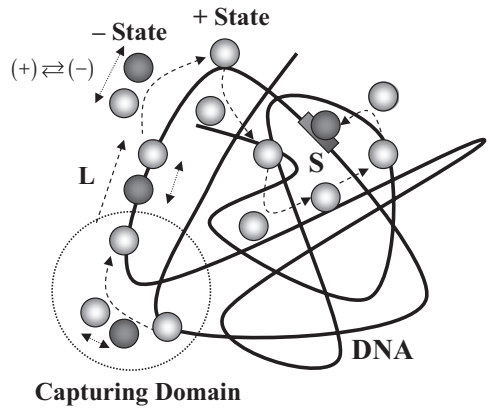

FIGURE 1 The DNA binding domain (DBD) of the DNA binding protein (DBP) molecule can exist in two possible states—namely, the plus or fast-diffusing state (*light shading*), and the minus or slow-diffusing state (*dark shading*). Flipping between these two different states with rate γ (s^{-1}) is a thermally driven process. These different states are mainly characterized by distinct one-dimensional diffusion coefficients $D_{\pm}$ associated with them and $D_+ > D_-$ for nonspecific DNA sequences. Upon making a nonspecific contact, the protein molecule of interest scans the DNA chain for an average sliding length of L bps and then dissociates from DNA to reassociate back at the same or different location of the same DNA chain after a brief three-dimensional excursion. Such events are possible only within the capturing domain, which is characterized by the electrostatic attractive force field present in-between the phosphate backbone of the DNA polymer and the negative side chains of the amino acids present at the DBDs of DBPs. When the DBDs are in a fast-diffusing state, DBPs are less sensitive to the DNA sequence and freely diffuse along DNA. When the DBDs are in the slow-diffusing state, they are more sensitive to the DNA sequence and slowly diffuse along the DNA. Upon detecting the specific site (S), the protein molecule flips to the slow-diffusing state and forms a tight complex. The transition (+) ⇄ (−) associated with the DBDs of DBPs is a stochastic event that can occur irrespective of whether the DBPs are sliding along the DNA polymer or are in the three-dimensional excursion after a recent dissociation event.

irrespective of whether the DBPs are sliding along the DNA polymer or they are in the three-dimensional excursion after a recently occurred dissociation event. Clearly, (N/L) numbers of such dissociation-association events, which are followed by three-dimensional diffusion-mediated nonspecific binding events, are required by the protein molecule of interest to locate its specific binding site.

The overall search time τ_s associated with the site-specific binding of the protein molecule of interest with the DNA chain can be given as $\tau_s = (N/L)(\tau_L + \tau_{ns})$, where$\tau_L$ is the time that is required by the protein molecule to scan a sliding length of L bps and τ_{ns} is the time that is required by the protein molecule to make a nonspecific contact with the DNA chain via three-dimensional diffusion. In the absence of the thermally driven conformational fluctuations in the DBDs of DBPs, the scan timeτ_L can be given (14–17) as $\tau_{L\pm} \sim L^2(6D_\pm)^{-1}$. In the presence of thermally driven flipping between plus-state and minus-state, the dynamics of the nonspecifically bound protein molecule of interest on the DNA chain can be described by the following coupled differential Chapman-Kolmogorov equation (14–17):

$$\partial_t \begin{bmatrix} p_+(x,t|x_0,0) \\ p_-(x,t|x_0,0) \end{bmatrix} = \begin{bmatrix} -\gamma + (D_+/2)\partial_x^2 & \gamma \\ \gamma & -\gamma + (D_-/2)\partial_x^2 \end{bmatrix} \times \begin{bmatrix} p_+(x,t|x_0,0) \\ p_-(x,t|x_0,0) \end{bmatrix}. \tag{1}$$

Here, $p_\pm(x,t|x_0,0)$ is the probability of finding the protein molecule at the DNA position x at time t starting from the DNA position x_0 at time $t = 0$, and γ is the transition rate associated with the thermally driven flipping between plus- and minus-states of the DBDs of DBPs under consideration. The initial condition is

$$p_+(x,0|x_0,0) = p_-(x,0|x_0,0) = \delta(x - x_0)/2,$$

and boundary conditions can be given as

$$[p_+]_{x=0} = [p_-]_{x=0} = [p_+]_{x=L} = [p_-]_{x=L} = 0. \tag{2}$$

The overall mean first passage time, $\overline{T}(x)$, which is required by the nonspecifically bound protein molecule to scan of L bps of the DNA chain in the presence of flipping dynamics between two different conformational states of DBDs, can be derived from the following backward-type differential Chapman-Kolmogorov equation (14–21):

$$\begin{bmatrix} -\gamma + (D_+/2)d_x^2 & \gamma \\ \gamma & -\gamma + (D_-/2)d_x^2 \end{bmatrix} \begin{bmatrix} T_+(x) \\ T_-(x) \end{bmatrix} = -\begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}. \tag{3}$$

Here, boundary conditions for Eq. 3 directly follow from Eq. 2:

$$[T_+]_{x=0} = [T_-]_{x=0} = [T_+]_{x=L} = [T_-]_{x=L} = 0. \tag{4}$$

We should note that the overall mean first passage time that is required by the protein molecule to escape from the interval [0,L], starting from the position x that is anywhere inside [0,L], can be given as

$$\overline{T}(x) = T_+(x) + T_-(x).$$

Using this, one can derive the solution of Eq. 3 corresponding to the boundary conditions, which are given by Eq. 4 as

$$\overline{T}(x) = \frac{x(L-x)}{\overline{D}_A} + \frac{1}{8\gamma}\left(\frac{D_\delta}{\overline{D}_A}\right)^2 \times \left(1 - \frac{\sinh\left(2\sqrt{\gamma/\overline{D}_G}x\right) + \sinh\left(2\sqrt{\gamma/\overline{D}_G}(L-x)\right)}{\sinh\left(2\sqrt{\gamma/\overline{D}_G}L\right)}\right). \tag{5}$$

Here, we have defined

$$D_\delta = (D_+ - D_-),$$

$$\overline{D}_A = (D_+ + D_-)/2,$$

and

$$\overline{D}_G = 2D_+D_-/(D_+ + D_-).$$

The three-dimensional plot of $\overline{T}(x)$ as a function of both the variables x and γ is shown in Fig. 2. One can derive many interesting results from Eq. 5 as follows. The initial position averaged mean exit time, $\widehat{T}(L,\gamma)$, which is required by the protein molecule of interest to scan an average sliding length of L bps before dissociating from the DNA chain in the presence of thermally driven conformational fluctuations of DBDs of DBPs, can be given as

$$\widehat{T}(L,\gamma) = L^{-1}\int_0^L \overline{T}(x)dx = \frac{L^2}{6\overline{D}_A} + \frac{1}{8\gamma}\left(\frac{D_\delta}{\overline{D}_A}\right)^2 \times \left(1 + \frac{1 - \cosh\left(2\sqrt{\gamma/\overline{D}_G}L\right)}{L\sqrt{\gamma/\overline{D}_G}\sinh\left(2\sqrt{\gamma/\overline{D}_G}L\right)}\right). \tag{6}$$

Noting the limits as

$$\lim_{\gamma\to 0}\widehat{T}(L,\gamma) = \widehat{T}(L,0) = L^2/(6\overline{D}_G)$$

and

$$\lim_{\gamma\to\infty}\widehat{T}(L,\gamma) = \widehat{T}(L,\infty) = L^2/(6\overline{D}_A),$$

one can conclude that upon increasing the flipping rate γ as $\gamma \to \infty$, the overall effective diffusion coefficient transform as $\overline{D}_G \to \overline{D}_A$, where $\overline{D}_A \geq \overline{D}_G$ and we have $\widehat{T}(L,0) \geq \widehat{T}(L,\infty)$.

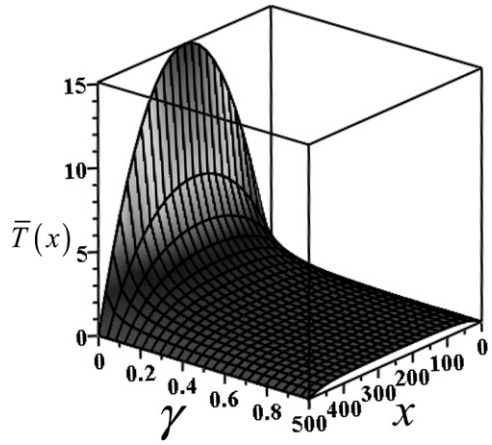

FIGURE 2 Mean first passage time $\overline{T}(x)$ (s) required by the protein molecule to escape from the interval [0,L] of the DNA chain as given by Eq. 5. Here we used the experimental values (11) of the lower and upper limits of the diffusion coefficients $D_+ \sim 11.2 \times 10^5$ bps^2 s^{-1} and$D_- \sim 2.07 \times 10^3$ bps^2 s^{-1}. We set the one-dimensional sliding length as $L = 500$ bps. The variable x (bps) is the landing or initial position of the protein molecule inside the interval $[0, L]$ and γ(s^{-1}) is the rate of flipping between the plus (+) and minus (−) states of the DNA binding domain of the protein molecule.

From these limiting conditions, the maximum achievable reduction η of the search time upon coupling the thermally driven conformational fluctuations in the DBDs of DBPs with the search dynamics of protein molecules on the DNA chain can be given as

$$\eta = \widehat{T}(x_a, \infty)/\widehat{T}(x_a, 0) = \overline{D}_A/\overline{D}_G = (D_+ + D_-)^2/(4D_+D_-). \quad (7)$$

The function $\widehat{T}(L, \gamma)$ is a monotonically decreasing one with γ, as $\partial_\gamma \widehat{T}(L, \gamma)$ vanishes only at the flipping rate $\gamma = 0$, which is a point of inflection. This follows from the fact that the solution to $\partial_\gamma \widehat{T}(L, \gamma) = 0$ can be given as

$$\gamma_s = 9\overline{D}_G\left(e^{2w_r} - 1\right)^2 \Big/ \left(4L^2\left(e^{2w_r} + 4e^{w_r} + 1\right)^2\right). \quad (8)$$

Here, w_r is the real root of

$$we^{2w} + 4we^{w} + w - 3e^{2w} + 3 = 0,$$

where $w_r = 0$ and subsequently one finds that $\gamma_s = 0$ and the second derivative will be

$$\partial_\gamma^2 \widehat{T}(L, \gamma) < 0$$

for all $0 < \gamma < \infty$. This means that the mean first passage time $\widehat{T}(L, \gamma)$ attains the minimum only in the limit $\gamma \to \infty$. It is evident from Eq. 6 that

$$\left|\widehat{T}(L, \gamma) - \widehat{T}(L, \infty)\right| \to 0$$

only when $\gamma \sim (\overline{D}_G/L^2)$, which follows from the inequality condition that

$$\left\{L\sqrt{\gamma/\overline{D}_G}\sinh\left(2\sqrt{\gamma/\overline{D}_G}L\right)\right\} \geq \left\{-1 + \cosh\left(2\sqrt{\gamma/\overline{D}_G}L\right)\right\},$$

where we have $L > 0$. This also means that the inequality

$$\left\{L\sqrt{\gamma/\overline{D}_G}e^{2\sqrt{\gamma/\overline{D}_G}L}\right\} \geq \left\{-1 + e^{2\sqrt{\gamma/\overline{D}_G}L}\right\}$$

holds true in the limit $L \to \infty$, and we find that

$$\widehat{T}(L, \gamma) \to \widehat{T}(L, \infty)$$

is faster when

$$\gamma \sim (\overline{D}_G/L^2)$$

and therefore,

$$\gamma_{\text{opt}} \sim (\overline{D}_G/L^2).$$

This is reasonable, because to make any significant effect on the overall scanning time, the timescale associated with the flipping dynamics of the DBDs must be much less than that of the timescale associated with the scanning dynamics of the DBPs along the DNA chain in the absence of conformational flipping. This means that the inequality

$$\gamma_{\text{opt}} \sim \left[\widehat{T}(L, 0)\right]^{-1}$$

should be true to attain the overall minimum scanning time $\widehat{T}(L, \infty)$. Here one should note that $\widehat{T}(L, \gamma)$ is a monotonically increasing function of L and $\partial_L \widehat{T}(L, \gamma) = 0$ only at $L = 0$, which follows from its solution that is given as

$$L_s = w_l\sqrt{\overline{D}_G(4\gamma)^{-1}},$$

where w_l is the real root of the equation

$$3e^{2w}\overline{D}_\delta^2 + e^{2w}w^3\overline{D}_G\overline{D}_A - 6e^{w}w\overline{D}_\delta^2 + 2e^{w}w^3\overline{D}_G\overline{D}_A + w^3\overline{D}_G\overline{D}_A - 3\overline{D}_\delta^2 = 0. \quad (9)$$

Noting that $w_l = 0$ for a sufficiently large sliding length L and when $\gamma \sim (\overline{D}_G/L^2)$, Eq. 6 can be approximated as follows, which in fact is clearly demonstrated in Fig. 3:

$$\widehat{T}(L, \gamma) \approx L^2/(6\overline{D}_A) + (D_\delta/\overline{D}_A)^2(8\gamma)^{-1}\left(1 - \left(L\sqrt{\gamma/\overline{D}_G}\right)^{-1}\right). \quad (10)$$

One should note that the approximation given by Eq. 10 is not valid when $\gamma < (\overline{D}_G/L^2)$, which is apparent from Fig. 3. Upon substituting the expression for the one-dimensional scanning time, $\widehat{T}(L, \gamma)$, which is required by DBPs to scan L bps of the DNA chain given by Eq. 10 in the expression for the overall search time associated with the

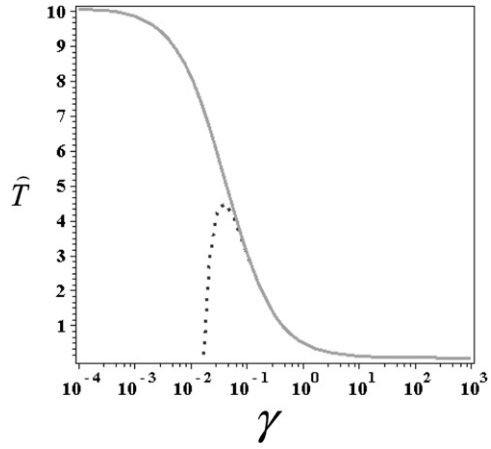

FIGURE 3 Semilog plot of initial position averaged mean first -passage time $\widehat{T} \rightarrow \widehat{T}\ (L,\gamma)$ (s) required to scan L bps of the DNA chain as a function of the flipping rate $\gamma(s^{-1})$ associated with transition between the plus (+) and minus (−) states of the DNA binding domain of the protein molecule as given in Eq. 6 (*solid line*) and the approximation that is given by Eq. 10 (*dotted line*). Here we used the experimental values (11) of the lower and upper limits of the diffusion coefficients $D_+ \sim 11.2 \times 10^5$ bps^2 s^{-1} and $D_- \sim 2.07 \times 10^3$ bps^2 s^{-1} and we set the one-dimensional sliding length as $L = 500$ bps. With this setting, we find that $[\widehat{T}\ (500,0)]^{-1} \sim 0.1\ s^{-1}$, and clearly the optimum flipping rate γ_{opt} should be such that $\gamma_{opt} \geq 10^2[\widehat{T}\ (500,0)]^{-1} \sim 10\ s^{-1}$ to attain the minimum possible scanning time of $\widehat{T}\ (500, \infty) \sim 0.074$ s.

protein molecule of interest to locate its specific binding site on the DNA chain of length N bps, we arrive at

$$\tau_s(L,\gamma) = (N/L)\left(\widehat{T}\ (L,\gamma) + \tau_{ns}\right) \sim N\left(L/(6\overline{D}_A) + \sigma/L - \theta/L^2\right). \quad (11)$$

From Eq. 11, one can conclude that the overall minimum search time is $\tau_s\ (L,\infty)$. Here we have defined

$$\sigma = \left((D_\delta/\overline{D}_A)^2/(8\gamma) + \tau_{ns}\right)$$

and

$$\theta = (D_\delta/\overline{D}_A)^2/\left(8\gamma\sqrt{\gamma/\overline{D}_G}\right).$$

When $\gamma \sim (\overline{D}_G/L^2)$, the approximate optimum one-dimensional sliding length L_{opt} that is required to achieve the overall minimum search time in the presence of thermally driven flipping of conformational states of DBDs of DBPs can be derived by solving $\partial_L \tau_s\ (L,\gamma) = 0$ for L as

$$L_{opt} = \lambda^{1/3} + 2\sigma\overline{D}_A\lambda^{-1/3}.$$

Here we have defined

$$\lambda = 2\sqrt{9\theta^2\overline{D}_A^2 - 2\sigma^3\overline{D}_A^3} - 6\theta\overline{D}_A.$$

From the limits of $\widehat{T}\ (L,\gamma)$ that is given in Eq. 6 as $\gamma \rightarrow 0$ and $\gamma \rightarrow \infty$, one finds that

$$\sqrt{6\overline{D}_A\tau_{ns}} \leq L_{opt} \leq \sqrt{6\overline{D}_G\tau_{ns}},$$

which depends on the value of the flipping rate. It is apparent from our earlier arguments that the overall possible minimum search time $\tau_s(L,\infty)$ that is associated with the site-specific binding of DBPs with the DNA chain can be achieved only when the flipping rate is such that

$$\gamma_{opt} \sim \left[\widehat{T}\ (L,0)\right]^{-1}.$$

Fig. 3 also suggests that

$$\gamma_{opt} \geq \left\{10^2\left[\widehat{T}\ (L,0)\right]^{-1}\right\}.$$

In the presence of m numbers of roadblock protein molecules (9) on the same DNA chain, the sliding length L associated with the searching dynamics of the protein molecule of interest for its specific site that is located on the same DNA chain under consideration varies as $L < Lm^{-1}$. As a result, the optimum-flipping rate that is required to achieve the overall minimum search time varies with m as $\gamma_{opt} \propto m^2$.

RESULTS AND DISCUSSION

Single-molecule experiments (11) on the diffusion of *Lac*I repressor protein molecule on the stretched nonspecific DNA sequence revealed the values of the one-dimensional diffusion coefficient in a wide range from $D_- \sim 2.3 \times 10^{-12}$ cm^2 s^{-1} to $D_+ \sim 1.3 \times 10^{-9}$ cm^2 s^{-1}. By using the transformation 1 bps ~ 3.4×10^{-8} cm, we find that $D_+ \sim 11.2 \times 10^5$ bps^2 s^{-1} and $D_- \sim 2.07 \times 10^3$ bps^2 s^{-1}. This observation suggests the possible existence of at least two different conformational states of DBDs with such different diffusion coefficients. However, such studies showed (11) a unimodal type distribution of diffusion coefficients rather than a bimodal type distribution corresponding to these two different states. Analogous to the observations on the downhill folding proteins (22), the unimodal type distribution of the diffusion coefficients could be possible when the free energy barrier that separates these two different states of DBDs of DBPs is comparable with that of the thermal free energy. From single-molecule experiments, we also find (11) that the free sliding length L that is associated with the one-dimensional diffusion dynamics of the *Lac*I protein molecule on a stretched nonspecific DNA chain ranges from L_{min} ~ 120 nm to L_{max} ~ 2920 nm. By using the transformation rule 1 bps ~ 0.34 nm, we find that these lengths correspond to $L_{min} = 353$ bps and $L_{max} \sim 8588$ bps, respectively. Upon substituting these values in the expression for the optimum flipping rate

$$\gamma_{opt} \geq \left\{10^2\left[\widehat{T}\ (L,0)\right]^{-1}\right\}$$

required to attain the minimum possible overall search time, we find that γ_{opt} ranges from ~0.03 s^{-1} to ~20 s^{-1} for a stretched nonspecific DNA chain. One should note that the in vivo experiments (23) at single-cell and single-molecule levels showed an effective diffusion coefficient

that is associated with the one-dimensional diffusion dynamics of the DBP of interest on the DNA chain as $\overline{D}_G \sim 0.046\ \mu m^2\ s^{-1} \sim 4 \times 10^5\ bps^2\ s^{-1}$ and the corresponding optimum flipping rate γ_{opt} ranges from ~3 s^{-1} to ~1926 s^{-1}. In this context, we should note that the *Escherichia coli* bacterial cell contains the genomic DNA chain (24–27) of length $N \sim 4.6 \times 10^6$ bps that is loaded with $m \sim 3 \times 10^4$ numbers of roadblock protein molecules (25,26) in its logarithmic growth phase. This mean that the sliding length associated with the one-dimensional diffusion dynamics of the protein molecule of our interest on the same DNA chain will be $L \sim 10^2$ bps. When the effective diffusion coefficient associated with the one-dimensional dynamics of the protein molecule of interest on the genomic DNA chain inside the *E. coli* cell is in the order of $\overline{D}_G \sim 0.046\ \mu m^2\ s^{-1}$, then the optimum rate that is associated with the conformational flipping of DBDs of the nonspecifically bound DBPs to achieve the overall minimum search time becomes ~2.4×10^4 s^{-1}. This is clearly within the physiologically relevant timescales.

From the theory of reaction rates, we find the general expression for the flipping rate to be

$$\gamma = \gamma_0 e^{-E_{RT}}\left(s^{-1}\right),$$

where E_{RT} is the free energy barrier associated with the flipping transition that is measured in terms of RTs (1 *RT* ~ 0.591 kcal/mol at 298 K) at a given temperature and γ_0 is the flipping rate when $E_{RT} \to 0$. Analogous to the downhill folding rate limit, one can conclude that the rate that is associated with the thermally driven flipping between the plus- and minus-states of DBDs of DBPs will be ~10^6 s^{-1} when these states are separated by a free energy barrier E_{RT}, which is comparable with that of the thermal free energy (27,28). Under such conditions the flipping rate will be closer to

$$\gamma = \gamma_0 \sim 10^6\ s^{-1},$$

which is much higher than that of the required optimum-flipping rate $\gamma_{opt} \sim 2.4 \times 10^4\ s^{-1}$ for the genomic DNA of the bacterium *E. coli*. One also should note that this value of the flipping corresponds to a barrier height of $E_{RT} \sim 3.7\ RT$, which is ~2.2 kcal/mol at $T = 298$ K. In general, we have

$$E_{RTO} \leq \ln\left(10^{-2} L^2 \gamma_0 / (6\overline{D}_G)\right),$$

where E_{RTO} is the optimum barrier height (kcal/mol) that separates the plus- and minus-states of DBDs of DBPs to achieve a one-dimensional sliding length of L bps along the DNA chain (Fig. 4). Because the barrier height of a downhill folding protein will be $\leq 3\ RT$, upon solving

$$3 \leq \ln\left(10^{-2} L^2 \gamma_0 / (6\overline{D}_G)\right)$$

for L as well as from Fig. 4 we find that the corresponding free sliding length L should be such that $L \leq 70$ bps under in vivo conditions. These results further suggest that the in vivo conditions of *E. coli* bacterial cell are optimized by the evolution to attain the maximum efficiency of searching for the specific sites on DNA, by effectively coupling the thermally driven conformational fluctuations in the DBDs of DBPs with the one-dimensional diffusion dynamics of the protein molecules along the DNA chain.

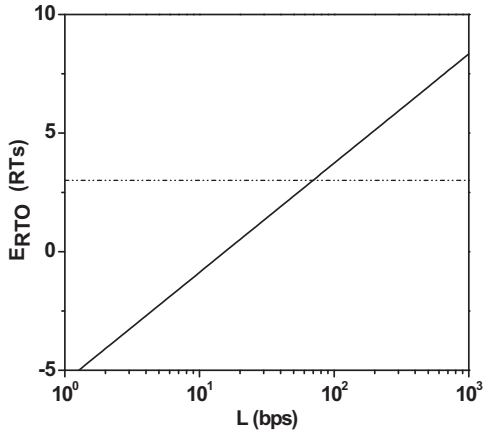

FIGURE 4 Optimum required free energy barrier E_{RTO} (RTs where 1 *RT* ~ 0.591 kcal/mol at 298 K) that separates the plus (+) and minus (−) states of the DNA binding domains (DBDs) for a given sliding length L (bps). Here we used the in vivo effective diffusion coefficient $\overline{D}_G \sim 0.046\ \mu m^2\ s^{-1}$ (18) and the folding rate limit for a downhill folding (21) protein as $\gamma_0 \sim 10^6\ s^{-1}$ and the barrier height is given by the expression $E_{RTO} \leq \ln(10^{-2} L^2 \gamma_0 / (6\overline{D}_G))$ (*solid line*). Because the barrier height of a downhill folding protein will be $\leq 3\ RT$, we find that the corresponding optimized sliding length L of proteins on DNA should be such that $L \leq$ 70 bps under in vivo conditions.

CONCLUSIONS

In summary, in this article we have developed a theory that explains how the thermally driven conformational fluctuations in the DBDs of DBPs are effectively coupled to the one-dimensional diffusion-mediated search dynamics of DBPs for their cognate sites on the DNA chain. Our theory suggested that the optimum rate associated with the flipping of conformational states of DBDs beyond which the maximum search efficiency of DBPs is achieved varies with the one-dimensional sliding length L as $\gamma_{opt} \propto L^{-2}$ and with the number of roadblock protein molecules present on the same DNA m as $\gamma_{opt} \propto m^2$. The required free energy barrier that is associated with this flipping transition seems to be varying with L as $E_{RTO} \propto \ln L^2$. When the barrier height is comparable with that of the thermal free energy as in case of downhill folding proteins, then our theory predicts the possible value of L under in vivo conditions as $L \leq 70$ bps.

REFERENCES

1. Levin, B. 2003. Genes VIII. Prentice Hall, New York.
2. Ptashne, M., and A. Gann. 2001. Genes and Signals. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

3. Winter, R. B., O. G. Berg, and P. H. von Hippel. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor-operator interaction: kinetic measurements and conclusions. *Biochemistry*. 20:6961–6977.

4. Murugan, R. 2009. Packaging effects on site-specific DNA-protein interactions. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys*. 79:061920–061928.

5. Slutsky, M., and L. A. Mirny. 2004. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J*. 87:4021–4035.

6. Coppey, M., O. Bénichou, …, M. Moreau. 2004. Kinetics of target site localization of a protein on DNA: a stochastic approach. *Biophys. J*. 87:1640–1649.

7. Lomholt, M. A., B. van den Broek, …, R. Metzler. 2009. Facilitated diffusion with DNA coiling. *Proc. Natl. Acad. Sci. USA*. 106: 8204–8208.

8. van den Broek, B., M. A. Lomholt, …, G. J. Wuite. 2008. How DNA coiling enhances target localization by proteins. *Proc. Natl. Acad. Sci. USA*. 105:15738–15742.

9. Sokolov, I. M., R. Metzler, …, M. C. Williams. 2005. Target search of *N* sliding proteins on a DNA. *Biophys. J*. 89:895–902.

10. Bonnet, I., A. Biebricher, …, P. Desbiolles. 2008. Sliding and jumping of single *Eco*RV restriction enzymes on non-cognate DNA. *Nucleic Acids Res*. 36:4118–4127.

11. Wang, Y. M., R. H. Austin, and E. C. Cox. 2006. Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys. Rev. Lett*. 97:048302–048305.

12. Kalodimos, C. G., N. Biris, …, R. Kaptein. 2004. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science*. 305:386–389.

13. Hu, L., A. Y. Grosberg, and R. Bruinsma. 2008. Are DNA transcription factor proteins Maxwellian demons? *Biophys. J*. 95:1151–1156.

14. Murugan, R. 2004. DNA-protein interactions under random jump conditions. *Phys. Rev. E*. 69:011911–011917.

15. Murugan, R. 2006. Critical jump sizes in DNA-protein interactions. *Biophys. Chem*. 120:143–147.

16. Murugan, R. 2006. Mean first passage time calculation for a random walk with random step size. *J. Phys. A*. 39:1575–1582.

17. Murugan, R. 2007. Generalized theory of site-specific DNA-protein interactions. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys*. 76:011901–011909.

18. Gardiner, C. W. 2002. Handbook of Stochastic Methods. Springer, Berlin.

19. Risken, H. 1996. Fokker Plank Equations. Springer, Berlin.

20. Van Kampen, N. G. 2004. Stochastic Processes in Physics and Chemistry. North-Holland, Amsterdam.

21. Redner, S. 2001. A Guide to First Passage Processes. Cambridge University Press, London.

22. Garcia-Mira, M. M., M. Sadqi, …, V. Muñoz. 2002. Experimental identification of downhill protein folding. *Science*. 298:2191–2195.

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

24. Goodsell, D. S. 1991. Inside a living cell. *Trends Biochem. Sci*. 16: 203–206.

25. Azam, T. A., A. Iwata, …, A. Ishihama. 1999. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol*. 181:6361–6370.

26. Johnson, R. C., and M. F. Bruist. 2005. *In* The Bacterial Chromosome. N. P. Higgins, editor. ASM Press, Washington DC.

27. Muñoz, V. 2007. Conformational dynamics and ensembles in protein folding. *Ann. Rev. Biophys. Biomol. Struct*. 36:395–412.

28. Yang, W. Y., and M. Gruebele. 2003. Folding at the speed limit. *Nature*. 423:193–197.