THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

DNA Looping and DNA Conformational Fluctuations Can Accelerate Protein Target Search

Published as part of The Journal of Physical Chemistry virtual special issue "Dave Thirumalai Festschrift".

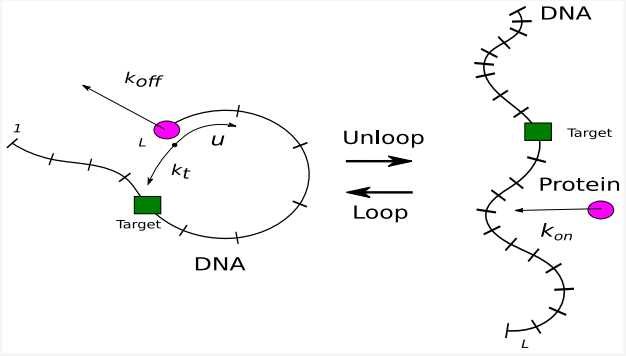
Cayke Felipe, Jaeoh Shin, and Anatoly B. Kolomeisky*

Cite This: *J. Phys. Chem. B* 2021, 125, 1727–1734

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Protein searching and binding to specific sites on DNA is a fundamentally important process that marks the beginning of all major cellular transformations. While the dynamics of protein–DNA interactions in *in vitro* settings is well investigated, the situation is much more complex for *in vivo* conditions because the DNA molecules in live cells are packed into chromosomal structures where they are undergoing strong dynamic and conformational fluctuations. In this work, we present a theoretical investigation on the role of DNA looping and DNA conformational fluctuations in the protein target search. It is based on a discrete-state stochastic analysis that allows for explicit calculations of dynamic properties, which is also supplemented by Monte Carlo computer simulations. It is found that for stronger nonspecific interactions between DNA and proteins the search occurs faster on the DNA looped conformation in comparison with the unlooped conformation, and the fastest search is observed when the loop is formed near the target site. It is also shown that DNA fluctuations between the looped and unlooped conformations influence the search dynamics, and this depends on the magnitude of conformational transition rates and on which conformation is more energetically stable. Physical–chemical arguments explaining these observations are presented. Our theoretical study suggests that the geometry and conformational changes in DNA are additional factors that might efficiently control the gene regulation processes.

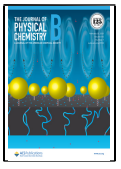


INTRODUCTION

DNA is the molecule that contains all the genetic information to support life. This information is "read" by protein transcription factors that bind to specific sites on DNA, initiating cascades of biochemical and biophysical processes. All these events together form a complex gene regulation system that is specific for each living organism.[1] In recent years, significant progress has been achieved in uncovering dynamic properties of subcellular structures that are relevant for the flow of genetic information.[2−4] These experimental advances support the idea that not only the DNA sequence but also its organization inside the living cells might play an important role in regulation of genetic processes. It was found, for example, that the DNA spatial configurations influence the amplitudes of gene expression for both activator and repressor systems.[5−7] These observations suggest that it is important to understand at the microscopic level how the protein search dynamics is affected by the DNA spatial organization and conformational dynamics.

Protein search for specific sites on DNA has been extensively studied using various experimental and theoretical approaches.[8−16] Many aspects of protein–DNA interactions are now much better understood. It has been argued, for example, that due to nonspecific interactions the protein molecules can be engaged in different searching modes.[15] For strong interactions, the search is mostly one-dimensional (1D) since once the protein binds to DNA it diffuses along the chain until the target is found. For weak interactions, protein can only occasionally bind to DNA without sliding, and here the search can be viewed as a three-dimensional (3D) process. The most interesting behavior is observed for intermediate strengths of nonspecific interactions when the protein is engaged in a combination of 1D+3D search process. In this regime, the protein molecule binds to DNA, slides some distance on the DNA chain, dissociates back into the solution, and then diffuses in the bulk until the search cycle is repeated after rebinding to DNA. Multiple search cycles might be observed before the target is located, and it was found that in this regime

Received: October 23, 2020
Revised: February 1, 2021
Published: February 11, 2021



https://dx.doi.org/10.1021/acs.jpcb.0c09599
J. Phys. Chem. B 2021, 125, 1727–1734

the search could be the fastest, corresponding to a so-called facilitated diffusion.[10,11,15]

While the protein search dynamics has been considered before in simplified systems, which are relevant to *in vitro* situations, the problem has been extended in multiple ways to understand the dynamics in more realistic cellular environment. Several theoretical studies on the role of protein conformational dynamics, the effect of molecular crowding on DNA, and the role of the protein intermolecular interactions have been presented.[17−21] Moreover, the role of sequence specificity of the protein−DNA interactions[22,23] has been investigated. Additionally, the role of DNA conformation on the search dynamics has been studied both experimentally and theoretically,[14,24−32] which is particularly important as multiple studies have shown that the genomic organization might strongly influence the protein search in live cells.[2,26,28,33−35] DNA loops are frequently found in both prokaryotic and eukaryotic cells, being formed actively by different proteins such as CTCF or condensin.[36] While these DNA loops might work as repressors by sterically blocking the protein transcription factors from finding to their target sites[7] or just allowing another repressor molecule to bind easier,[37] most of them are associated with enhancer−promoter activity.[6] Additionally, DNA loops may create either a set of static (fixed) loops or dynamic looping interactions that affect all processes in the system.[38,39] It was found that the geometry of DNA chains in the chromatin configurations affects the dynamics of interactions with protein transcription factors.[30−32] However, the microscopic details of how the topological state of DNA and its conformational dynamics are influencing the protein target search remain not well understood.

In this work, we present a theoretical study on the role of the two most important features of DNA in live cells, formation of loops and conformational fluctuations, in the protein finding specific target sites on DNA. Using a discrete-state stochastic model and Monte Carlo computer simulations, we investigate how the search process can be facilitated in the looped configuration in comparison with linear configuration without loops. Our goal is to consider a relatively simple model of protein target search in order to clarify the molecular picture of underlying processes. We found that search dynamics depends on the strength of nonspecific protein−DNA interactions and on the distance between the target and the loop intersection positions. In addition, the effect of conformational transitions between DNA looped and unlooped states has been investigated, and we suggest that the most efficient search is observed for fast DNA looping/unlooping transitions. The importance of our theoretical findings for understanding the mechanisms of genetic regulation is also discussed.

THEORETICAL MODEL

Search on DNA Looped Conformations. Let us start by investigating a model of the protein search for the target on the looped DNA chain as illustrated in Figure 1a. To simplify theoretical calculations, we consider a single DNA and a single protein molecule that is searching for the target located inside the DNA chain. But the real concentrations of protein and DNA molecules can be easily recovered from our single-molecule approach. The DNA molecule is viewed as a lattice segment consisting of L sites. In our theoretical model, one lattice site corresponds to approximately 10 base pairs (bp), which is a typical size of the DNA segment covered by the bound protein molecule, and it is also the average size of the

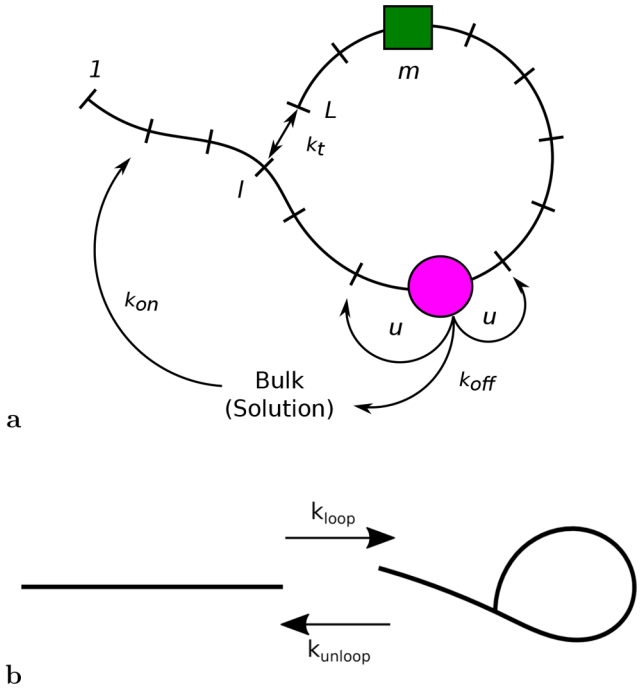


Figure 1. (a) Schematic view of the model where the DNA loop is formed by coupling at the sites l and L. The green square located at the site m is the target site. The magenta circle is the searching protein. Black arrows indicate allowed transitions to different states. (b) Simplified view of the conformational transitions in the DNA molecule. The free DNA (left) can form a loop with rate k_{loop}, while the looped DNA (right) can go back to its linear configuration with a rate k_{unloop}.

target sequences on DNA.[40] The end of the DNA molecule at the site L is coupled to the site l, which yields a DNA loop of size $p = L - l + 1$ (Figure 1a). For $l = 1$, the DNA molecule becomes circular. The target is located on the site m ($1 \leq m \leq L$).

The protein molecule starts in the bulk solution where it diffuses until it binds to DNA. It is assumed that the probability of binding to any site on DNA is the same, and the association rate per each DNA site is given by k_{on}: see Figure 1a. While on DNA, the protein can slide along the contour with a rate u in any direction, or it can dissociate back into the solution with a dissociation rate k_{off}. Once the protein molecule is dissociated, it is assumed that it diffuses fast enough so that the location of the consecutive binding site is uncorrelated with the location of the dissociation site. This means that our approach neglects the spatial correlation effects that lead to the protein molecule recapturing the DNA chain and that might be important for real *in vivo* cellular processes.[14,41] Additionally, near the loop intersection with the main DNA chain, the sites l and L are spatially close, and the protein can hop reversibly between these sites with a transfer rate k_t: see Figure 1a. When the transfer rate is zero ($k_t = 0$), the DNA is found in the free linear conformation. For now, we assume that the DNA loop is very stable and DNA remains in the looped conformation until the target is found.

To investigate the protein search dynamics for the model presented in Figure 1a, we use a method of first-passage probabilities that has been successfully applied for analyzing various protein search phenomena.[15,19,20,42,43] It is assumed that the protein can be found in one of $L + 1$ states that we

https://dx.doi.org/10.1021/acs.jpcb.0c09599
J. Phys. Chem. B 2021, 125, 1727−1734

label as n ($0 \le n \le L$). The state $n = 0$ corresponds to the protein being in the bulk solution, while the states $1 \le n \le L$ describe the protein on the site n of the DNA chain (Figure 1a). One can define then a probability density function, $F_n(t)$, of reaching the target site m at time t for the first time given that, initially, it was an the site n. The temporal evolution of these first-passage probability functions is controlled by the following backward master equations,

$$\frac{\partial F_n(t)}{\partial t} = k_{off}F_0(t) + u[F_{n-1}(t) + F_{n+1}(t)] - (k_{off} + 2u)F_n(t) \quad (1)$$

for $n = 2, 3, 4, ..., L - 1$ and $n \neq l$. For the boundary and special states ($n = 1, l, L$), we have

$$\frac{\partial F_1(t)}{\partial t} = k_{off}F_0(t) + uF_2(t) - (k_{off} + u)F_1(t) \quad (2)$$

$$\frac{\partial F_l(t)}{\partial t} = k_{off}F_0 + u[F_{l-1}(t) + F_{l+1}(t)] + k_tF_L(t) - (k_{off} + 2u + k_t)F_l(t) \quad (3)$$

$$\frac{\partial F_L(t)}{\partial t} = k_{off}F_0(t) + uF_{L-1}(t) + k_tF_l(t) - (k_{off} + u + k_t)F_L(t) \quad (4)$$

The backward master equation that describes the temporal evolution of the protein in the bulk ($n = 0$) is given by

$$\frac{\partial F_0(t)}{\partial t} = k_{on}\sum_{n=1}^{L} F_n(t) - Lk_{on}F_0(t) \quad (5)$$

Initially, the protein is found in the solution ($n = 0$ state). Finally, $F_m(t) = \delta(t)$, which means that if the initial position of the searching protein is already at the target site at $t = 0$, the process ends immediately.

It is convenient to analyze this problem using Laplace transformations of the first-passage probability functions, $\tilde{F}_n(s) \equiv \int_0^\infty \exp(-st)F_n(t)\,dt$. Then the backward master equations will be modified as follows,

$$(s + 2u + k_{off})\tilde{F}_n(s) = k_{off}\tilde{F}_0(s) + u[\tilde{F}_{n-1}(s) + \tilde{F}_{n+1}(s)] \quad (6)$$

$$(s + u + k_{off})\tilde{F}_1(s) = k_{off}\tilde{F}_0(s) + u\tilde{F}_2(s) \quad (7)$$

$$(s + 2u + k_t + k_{off})\tilde{F}_l = k_{off}\tilde{F}_0 + u[\tilde{F}_{l-1}(s) + \tilde{F}_{l+1}(s)] + k_t\tilde{F}_L(s) \quad (8)$$

$$(s + u + k_t + k_{off})\tilde{F}_L(s) = k_{off}\tilde{F}_0(s) + u\tilde{F}_{L-1}(s) + k_t\tilde{F}_l(s) \quad (9)$$

$$(s + Lk_{on})\tilde{F}_0(s) = k_{on}\sum_{n=1}^{L}\tilde{F}_n(s) \quad (10)$$

In the Supporting Information, we show how to solve exactly these equations to obtain analytical expressions for the Laplace transforms of the first-passage probability functions. This allows us to explicitly describe the system's search dynamics, including all the moments of the search time. Here we will focus on the mean search time since it is experimentally the most relevant quantity of the process. The mean search time to find the target when the initial state is $n = 0$ (starting from the bulk solution) can be simply obtained as

$$T_{loop} = T_0 = \int_0^\infty tF_0(t)\,dt = -\left.\frac{\partial\tilde{F}_0(s)}{\partial s}\right|_{s=0} = \frac{1}{k_{on}}\frac{L}{S(0)} + \frac{1}{k_{off}}\left(\frac{L}{S(0)} - 1\right) \quad (11)$$

where $S(s)$ is an auxiliary function (given in the Supporting Information). It can be argued that $S(0)$ corresponds to the average number of different sites explored by the protein in each search cycle. This also allows us to understand better the meaning of eq 11. The first term corresponds to the total time that the protein is found in the solution during the overall search process. This is because $1/k_{on}$ is the average time to be in the state not bound to DNA per one search cycle and $L/S(0)$ is the number of search cycles, that is, the number of times that the protein comes to DNA. The second term gives the time for the protein to be found on DNA since $1/k_{off}$ is the residence on DNA per cycle, and $(L/S(0) - 1)$ is the number of dissociations from DNA during the overall search process. It is less by one than the number of search cycles because the last association to DNA should be successful (the target will be found) and there will be no dissociations from DNA.

Search on Conformationally Fluctuating DNA. Now let us consider a more realistic situation with the DNA molecule transitioning between the looped and unlooped conformations as shown in Figure 1b. The linear DNA chain can transit into the looped configuration with a rate k_{loop}, while the opposite transition is taking place with a rate k_{unloop}. We assume that the identical loop is always formed at the same location. If the looping/unlooping transitions are very slow, the protein search should be effectively described by the search on the DNA configuration where the system started. This is because the target will be found before the conformational transition can happen. The situation is different when the conformational looping/unlooping dynamics is very fast. In this case, the system can be viewed as the effective looped DNA conformation with the rescaled transfer rate, $k_{t,eff}$. Then the protein search will be the same as in the always looped conformation. The search dynamics for intermediate rates k_{loop} and k_{unloop} is expected to be more complex.

The transition rates k_{loop} and k_{unloop} are related by the free energy changes in the system due to looping/unlooping, and they specify the equilibrium constant $K_{conform}$ for the conformational fluctuations,

$$K_{conform} \equiv \frac{k_{loop}}{k_{unloop}} = \exp[-G(p)] \quad (12)$$

where $G(p)$ is the free energy difference (in the unit of thermal energy, k_BT) between the looped conformation with the loop size $p = L - l + 1$ and the free DNA chain. It can be approximated using simple polymer physics arguments, assuming a circular shape for the DNA loop,[40]

$$G(p) = \frac{A}{p} + \alpha \ln p + \varepsilon \quad (13)$$

The first term describes the contribution to the free energy change due to the bending energy with the parameter A being proportional to the bending stiffness of DNA. It can be written as $A = 2\pi^2 l_p$, where l_p is the bending persistence length of the

https://dx.doi.org/10.1021/acs.jpcb.0c09599
J. Phys. Chem. B 2021, 125, 1727−1734

DNA. In our explicit calculations, we will utilize $A = 300$ in dimensionless units, which corresponds to $l_p \simeq 150$ bp in real units.[40] The second term corresponds to the entropic contribution to the free energy change. The parameter α is the scaling factor, which for ideal polymer chains is equal to 3/2. The last term describes the enthalpy of the loop formation,[40] and for explicit calculations, we take $\varepsilon = -5k_BT$, unless otherwise specified. In addition, we assume that the sliding rate u, association rate k_{on}, and dissociation rate k_{off} are the same in both DNA conformations.

In the next section, we present specific calculations of protein target search dynamics on looped DNA and on the DNA with looped conformational transitions using realistic parameters of the system.

RESULTS AND DISCUSSIONS

Because our theoretical approach could provide analytical results for dynamic properties of the system, it allows us to obtain a comprehensive description of the protein target search processes. We first analyze the search on the always looped DNA molecule. Figure 2 presents the mean search times as a

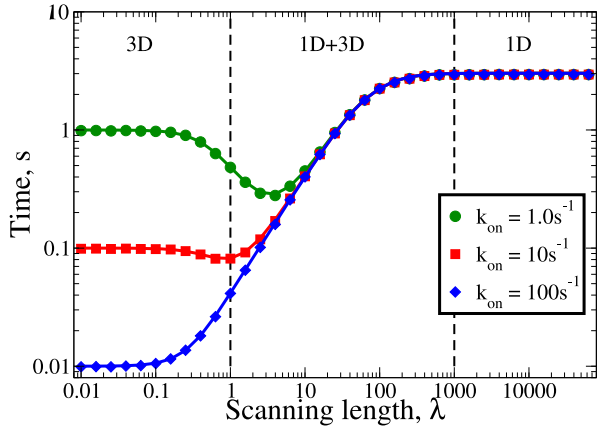


Figure 2. Mean search times as a function of the sliding length λ for the always looped DNA molecules. Analytical results are obtained using eq 11, and they are shown by solid lines. Monte Carlo simulations data are shown by symbols for three different values of the association rate k_{on}. The following parameters were used for calculations: $L = 500$, $m = 250$, $l = 200$, $u = 6.0 \times 10^3$ s^{-1}, and $k_t = 2.0 \times 10^3$ s^{-1}.

function of the scanning length $\lambda \equiv \sqrt{u/k_{off}}$, which is defined as the average distance on DNA that the protein scans every search cycle. The diffusion rate for the protein to slide along the DNA chain is known to be 10^5–10^6 bp^2 s^{-1},[10] and for specific calculations, we take this parameter to be $u = 6.0 \times 10^3$ s^{-1} because in our model each lattice site corresponds to 10 bp. The dissociation is controlled by the protein–DNA interactions: for stronger (weaker) protein–DNA interactions, smaller (larger) k_{off} is expected. In our study, the dissociation rate k_{off} is varied leading to changing λ. Note that the scanning length λ is closely related to the parameter $S(0)$ that describes the average number of visited sites per cycle, but it is not the same. It can be shown that $1 \leq S(0) \leq L$, while the scanning length can be smaller than one or larger than the length of the DNA segment L. It is convenient to utilize the scanning length λ because it correlates better with the strength of nonspecific interactions between the protein and DNA molecules. The stronger the interaction, the larger the scanning length.

Figure 2 shows that, similarly to other protein target search systems,[15,17−20,40,43] there are three dynamic regimes depending on the strength of nonspecific interactions. When the scanning length λ is larger than the total DNA length L (strong interactions), the protein is mostly involved in 1D search. This is because after binding to DNA from the solution, the protein will only diffuse along the DNA chain until the target is found. This search regime is slow since the protein molecule redundantly visits the same sites multiple times. In this case, the overall search times are determined by the residence times on DNA, and they are essentially independent of the association rates k_{on}. More efficient search dynamics is observed for very small scanning lengths $\lambda < 1$ (weak interactions). In this regime, the protein is able occasionally to bind to DNA but it cannot slide and after some time it dissociates back into the solution. The search times depend strongly on the association rates because $1/k_{on}$ gives the average time for the protein to be diffusing in the bulk solution for each search cycle. The faster the protein diffuses in 3D, the faster it will explore the whole DNA via binding/unbinding events and locate the target. For the intermediate scanning lengths, $1 < \lambda < L$, which also correspond to the intermediate strength of interactions, the protein is performing a combined 1D+3D search, and for some range of parameters, this is the most optimal search regime. In this case, the protein is able to scan some length on DNA every search cycle without being trapped on DNA for long periods of time, and the fast diffusion in the bulk allows it to return quickly to DNA to continue the scanning. This explains the nonmonotonic behavior of the mean search times as a function of the scanning length λ for some range of parameters.

We can focus now on the effect of the looped conformational state on the search time by varying the transfer rate k_t that allows the protein to jump reversibly between the sites l and L (see Figure 1a). Figure 3 shows the search time for

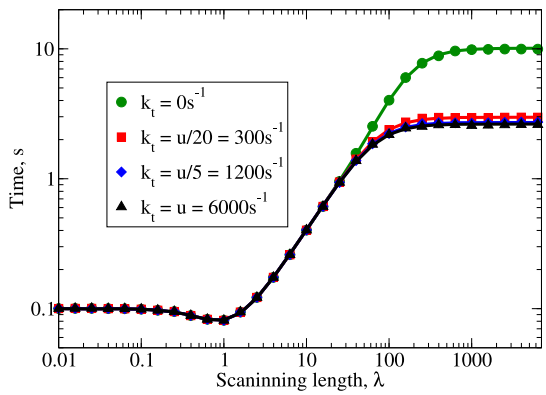


Figure 3. Mean search times as a function of the scanning length λ for different transfer rates, k_t. Analytical results are obtained from eq 11, and they are shown by solid lines. Monte Carlo simulations data are shown by symbols. The following parameters were used for calculations: $u = 6.0 \times 10^3$ s^{-1}, $k_{on} = 10$ s^{-1}, $L = 500$, $l = 100$, and $m = 450$.

several values of k_t. It is found that even small values of the transfer rate ($k_t \ll u$) can significantly accelerate the search dynamics in the regimes with intermediate and strong nonspecific interactions. An order of magnitude acceleration in the search might be observed (Figure 3). The reason is that the looped conformation provides an additional channel to

https://dx.doi.org/10.1021/acs.jpcb.0c09599
J. Phys. Chem. B 2021, 125, 1727−1734

transfer between the l and L sites near the loop intersection (see Figure 1a), allowing the protein to scan the DNA molecule more efficiently and to move faster between distant parts of the DNA chain. Thus, in these regimes, the search dynamics is accelerated for *any* value of the transfer rate k_u, and the search times are effectively independent of k_u, as long as $k_t \neq 0$.

We note that in ref 25 it has been found that the association rate of protein to a target site on the circular DNA is lower, albeit modestly, than that for the linear counterpart. In that case, the authors consider only a special case where the target site is located at the center of the DNA chain. The difference in the association rates originates from the excluded interaction between the search protein and the ends of the linear DNA, which will be negligible for a long DNA. For small values of the protein−DNA interactions, the variation in the transfer rate does not affect the search dynamics as much. The reason is that in these situations the protein molecule moves mostly in the bulk and the additional paths to explore DNA are less helpful to the search since the protein is not sliding along the DNA chain. However, we remark here that there could be another mechanism of enhancing the search efficiency, namely, "hopping" between spatially close segments of the DNA even without direct physical connectivity.[14,24,41,44] Therefore, the results of our model would describe the lower bound of the search efficiency.

Our theoretical analysis also suggests that the search dynamics depends on the relative positions of the target m and the loop intersection with DNA l, from where the transitions to the site L might happen (see Figure 1a). The corresponding results are presented in Figures 4 and 5. From

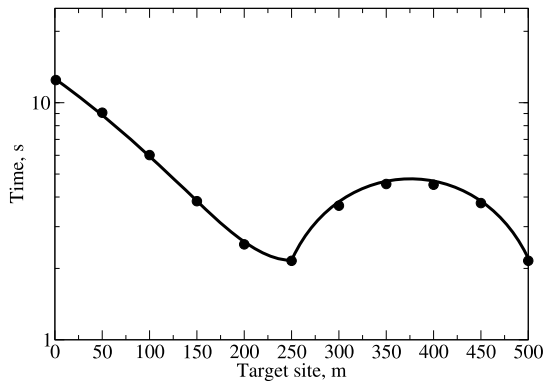


Figure 4. Mean search times as a function of the target position for strong nonspecific protein−DNA interactions (large scanning lengths, $\lambda \gg L$). Analytical result [from eq 11] is shown by solid lines, and the Monte Carlo simulations data are shown by symbols. The following parameters were used for calculations: $u = 6 \times 10^3$ s^{-1}, $k_t = u$, $k_{on} = 10$ s^{-1}, $L = 500$, $l = 250$, and $\lambda = 10^4$.

Figure 4, one can see that the largest search times are observed for $m = 1$ when the target is at the DNA free end. Increasing m accelerates the search, and the fastest dynamics is observed for $m = l$. Further increase in m first slows down the search, but then the search starts to become faster again because the target is approaching the loop intersection ($m \approx L$). This observation can be explained by noting that due to the reversible transfer at the loop intersection the DNA region around this location is explored more frequently (see Figure 1a). This means that when the target is not far away from the loop intersection the search will be completed faster.

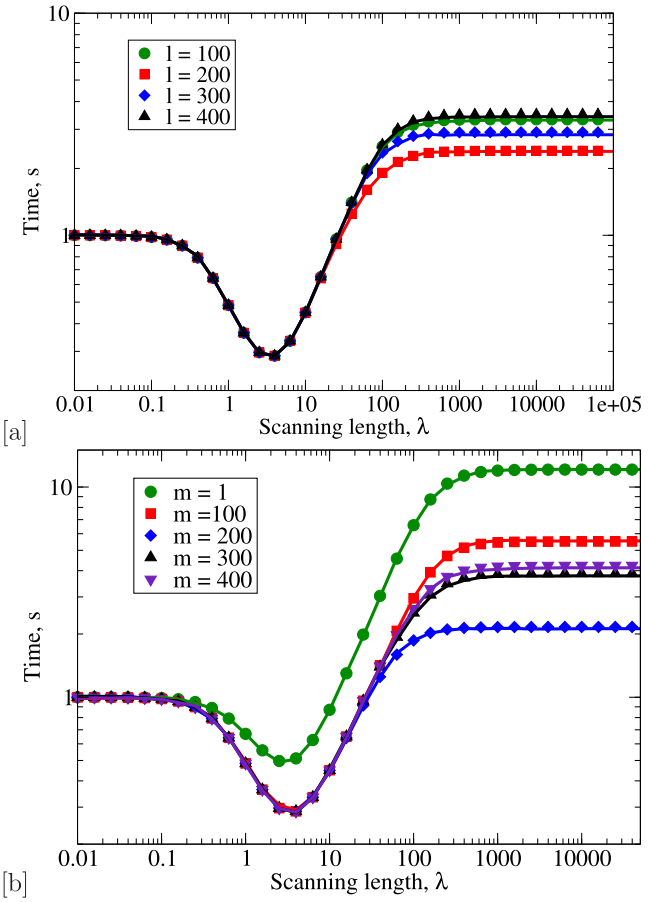


Figure 5. Mean search times as a function of the scanning length λ for different locations of (a) the DNA loop intersection and (b) the target. Analytical results are shown in solid lines, and Monte Carlo simulations data are shown in symbols. The following parameters were used for calculations: (a) $u = 6 \times 10^3$ s^{-1}, $k_t = 2 \times 10^3$ s^{-1}, $k_{on} = 1$ s^{-1}, $L = 500$, and $m = 225$; (b) $u = 6 \times 10^3$ s^{-1}, $k_t = 2 \times 10^3$ s^{-1}, $k_{on} = 1$ s^{-1}, $L = 500$, and $l = 225$.

The effect of the distance between the target and the loop intersection on the search dynamics is also illustrated in a slightly different way in Figure 5. For a fixed location of the target, one can vary the position of the loop intersection l, which also corresponds to changing the size of the DNA loop. These calculations are shown in Figure 5a. One can see again that the changes in the search times are observed for strong nonspecific interactions (large scanning lengths), and the fastest dynamics is found when the loop intersection is close to the target ($m \approx l$). Similar results are found if one fixes the size of the loop and varies the position of the target m as shown in Figure 5b. The search is the most efficient when the target is close to the loop intersection.

Our calculations show that, when the target is not far away from the loop intersection, it might lead to a significant acceleration in the search dynamics, modifying the search times by up to 1 order of magnitude. As shown in the Figure 6 for the heat map for the acceleration, the fastest search is clearly observed when $m \approx l$ and $m \approx L$. We can conclude then that the DNA looped configuration is generally better for the search dynamics, and the most optimal conditions are found when the loop intersection is close to the target sequence. This effect is observed for intermediate and strong nonspecific protein−DNA interactions.

https://dx.doi.org/10.1021/acs.jpcb.0c09599
J. Phys. Chem. B 2021, 125, 1727−1734

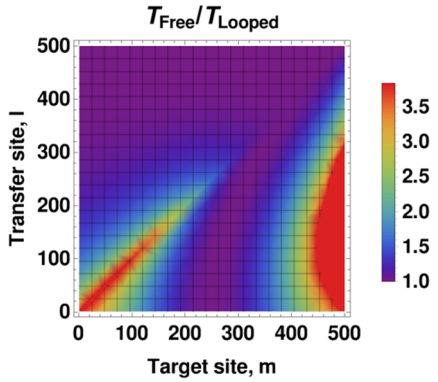


Figure 6. Acceleration of the mean search times on the DNA chain with the loop as compared to the free DNA, T_{free}/T_{loop}. The following parameters were used for calculations: $u = 6 \times 10^3$ s^{-1}, $k_{on} = 10$ s^{-1}, $L = 500$, $\lambda = 10^4$, and $k_t = u$.

A more realistic description of the protein search on DNA should include the conformational transitions between the DNA looped and free states, as schematically presented in Figure 1b. It is clear that search dynamics will be affected by looping/unlooping transitions for intermediate and strong protein−DNA interactions when the protein molecule can slide long distances on DNA. The results of computer simulations for this situation are presented in Figure 7. For

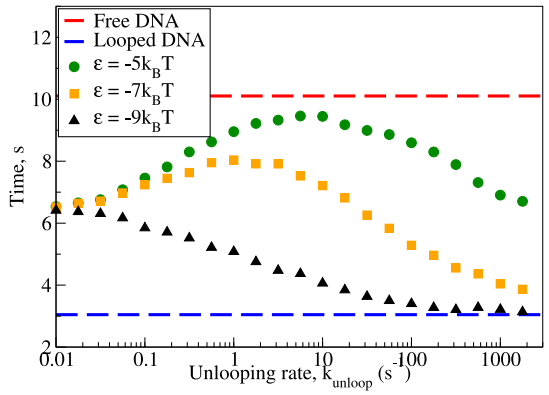


Figure 7. Mean search times as a function of the unlooping rates for the system when DNA fluctuates between the free and the loop conformations. Initially, the DNA molecule starts with equal probability in any of these conformations. The dashed lines correspond to analytic limiting results for the free and for the looped conformations, respectively, while the symbols correspond to computer simulations for the systems with different enthalpic contributions due to creating the DNA loop ε. The following parameters were used for calculations: $u = 6 \times 10^3$ s^{-1}, $k_t = 6 \times 10^3$ s^{-1}, $k_{on} = 10$ s^{-1}, $L = 500$, $m = 450$, $l = 200$, and $\lambda = 10^4$.

all situations, the search times are bounded between the results for the free and those for the looped DNA conformations. This means that the protein search is *always* faster on the conformationally fluctuating DNA in comparison with the search on the free DNA chain.

When the looping/unlooping transition rates are small, the search is mostly determined by the initial DNA conformation. In Figure 7, the system started with equal probability in one of two conformations, and the search times are essentially given by $T \simeq (T_{unlooped} + T_{looped})/2$. In this case, the target is found before the conformational equilibrium is established. Increasing the unlooping rate (and proportionally the looping rate via eq 12) has a more complex effect on the search times. For very high unlooping/looping rates, the conformational equilibrium is quickly reached before the target is located, and the system behaves like a single looped state with a new effective transfer rate, $k_{t,eff}$. In this situation, the search dynamics is also fast and it is close to the case of the protein search on the always looped DNA as discussed above. This is because, as we already argued, the magnitude of the transfer rate is essentially not affecting the search, but the topological change of having the loop versus not having the loop is important.

The most interesting behavior is observed for intermediate k_{unloop}, as shown in Figure 7. When the free conformation state is energetically more preferred ($\varepsilon = -5k_BT$, $K_{conform} \simeq 0.01$) the protein can transfer from the looped into the free state and it would stay there for significant periods of time. The fraction of time when the DNA molecule is in the free (unlooped) state, f_{free}, can be estimated from the equilibrium constant, $f_{free} = 1/(K_{conform} + 1)$. In this case, the system is found in the free state 99% of the time, and for this reason, the search times are start to increase with k_{unloop} because $k_{unloop} \gg k_{loop}$ and the probability to shift the system into the free state is very large. Then the search times are approaching closer to T_{free}. The target is typically found before the conformational transition can happen. In other words, for intermediate k_{unloop}, the search times are faster than the times to establish the conformational equilibrium. However, when the free state becomes energetically less favorable ($\varepsilon = -9k_BT$, $K_{conform} \simeq 0.58$), the fraction of time the DNA chain is free decreases to 63%. In this case, the search times do not increase with k_{unloop} (because $k_{unloop} \approx k_{loop}$), and now they are getting closer to T_{looped}. In all situations, however, for very fast looping/unlooping transitions, the equilibrium is reached faster than the typical search times, and the system can be viewed as the single looped conformation. As the result, a nonmonotonic dependence of the search times as a function of the unlooping transition rates is observed for the cases when the free conformational state is energetically more stable, but not for the cases when the looped conformations starting to become more relevant. These observations show that the DNA conformational dynamics results in a very nontrivial effect on the protein target search. But in all cases, importantly, it is always faster to search on the conformationally fluctuating DNA than on the free DNA.

The presented theoretical approach allows us to explicitly evaluate the role of DNA geometry and conformational dynamics on the protein search for specific sequences on DNA. Our results suggest that tuning the DNA geometry and the degree of conformational fluctuations might be an additional efficient method of how the nature might regulate genetic processes. One could imagine that changing the size of DNA loops and the location of the loop intersections might activate or repress the specific genes. Modifying the conformational dynamics might also lead to nonmonotonic changes in the search times, allowing for fine-tuning the gene activation processes. It will be interesting to explore these ideas in more direct experimental studies.

SUMMARY AND CONCLUSIONS

In this work, we developed a theoretical approach to evaluate the role of DNA geometry and conformational dynamics on protein target search phenomena. By constructing a discrete-state stochastic model for the underlying processes, the dynamics of the system is explicitly computed using a method

https://dx.doi.org/10.1021/acs.jpcb.0c09599
J. Phys. Chem. B 2021, 125, 1727−1734

of first-passage probabilities. Our theoretical calculations are also supported by Monte Carlo computer simulations. It is found that the formation of the DNA looped configuration might significantly affect the search dynamics, and this depends on the strength of nonspecific protein−DNA interactions. For weak interactions, the DNA geometry does not much influence the protein search, but for stronger interactions it becomes much more important because in this case the protein molecule spends most of the time on DNA. In these dynamic regimes, the effect is stronger depending on how close the target is to the loop intersection. Taking into account the effect of conformational transitions between the free and the DNA looped conformations is also investigated. Here we found that the conformational fluctuations also modify the search, and the effect is more complex. When the conformational transitions are fast, the system behaves like the always looped DNA state and the search is fast. For intermediate values of the looping/unlooping rates, the search dynamics to a large degree depends on what conformational state is energetically more favorable and how fast the conformational equilibrium is reached. In addition, it is argued that varying the geometry and the degree of conformational fluctuations of the DNA molecules might serve as a very efficient and sensitive tool in genetic regulations and manipulations. Our theoretical calculations strongly suggest that topological features of DNA molecules and dynamic transformations between different conformational states are critically important for understanding the molecular details of protein−DNA interactions. Our main idea is that the protein target search is always faster on the looped DNA molecule or on the conformationally fluctuating DNA because of the appearance of additional channels for finding the target due to topological changes in the system.

While the presented theoretical approach provides a reasonable minimal description of DNA geometry and conformational dynamics, it is important to emphasize its limitations. Several important features of DNA molecules in chromatin structures such as DNA supercoiling,[45] sequence heterogeneity,[43] and the possibility of multiple DNA loops are not considered at all. In addition, the description of the free energy associated with the formation of DNA loops is rather oversimplified. Furthermore, the protein diffusion rates on the free and on the looped DNA chains are probably not the same as was assumed in this work. Spatial correlation effects and crowding are also not taken into account in our approach. However, despite these shortcomings our theoretical model provides some clear physical insights on the complex mechanisms of protein target search in live cells. Most importantly, it gives quantitative predictions that can be probed in experiments. It will be important to extend our theoretical approach by taking more realistic features of protein−DNA dynamics in living cells and to compare with experimental observations.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c09599.

Details of analytical calculations for mean search times (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Anatoly B. Kolomeisky − *Department of Physics and Astronomy, Rice University, Houston, Texas 77005, United States; Center for Theoretical Biological Physics, Department of Chemistry, and Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0001-5677-6690; Email: tolya@rice.edu

Authors

Cayke Felipe − *Department of Physics and Astronomy, Rice University, Houston, Texas 77005, United States; Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States*

Jaeoh Shin − *Center for Theoretical Biological Physics and Department of Chemistry, Rice University, Houston, Texas 77005, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.0c09599

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work was supported by the Welch Foundation (C-1559), by the NSF (CHE-1953453 and MCB-1941106), and by the Center for Theoretical Biological Physics sponsored by the NSF (PHY-2019745).

■ REFERENCES

(1) Lodish, H.; Berk, A.; Kaiser, C. A.; Krieger, M.; Scott, M. P.; Bretscher, A.; Ploegh, H.; Matsudaira, P., et al. *Molecular cell biology*; Macmillan, 2008.

(2) Stråringlfors, A.; Ekwall, K. *Reviews in Cell Biology and Molecular Medicine*; American Cancer Society, 2011.

(3) Goldberg, A. D.; Allis, C. D.; Bernstein, E. Epigenetics: A Landscape Takes Shape. *Cell* **2007**, *128*, 635−638.

(4) Rothbart, S. B.; Strahl, B. D. Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2014**, *1839*, 627−643.

(5) Adhya, S. Multipartite genetic control elements: communication by DNA loop. *Annu. Rev. Genet.* **1989**, *23*, 227−250.

(6) Kustu, S.; North, A. K.; Weiss, D. S. Prokaryotic transcriptional enhancers and enhancer-binding proteins. *Trends Biochem. Sci.* **1991**, *16*, 397−402.

(7) Cournac, A.; Plumbridge, J. DNA Looping in Prokaryotes: Experimental and Theoretical Approaches. *J. Bacteriol.* **2013**, *195*, 1109−1119.

(8) Berg, O. G.; Winter, R. B.; Von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **1981**, *20*, 6929−6948.

(9) Halford, S. E.; Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic acids research* **2004**, *32*, 3040−3052.

(10) Mirny, L.; Slutsky, M.; Wunderlich, Z.; Tafvizi, A.; Leith, J.; Kosmrlj, A. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A: Math. Theor.* **2009**, *42*, 434013.

(11) Kolomeisky, A. B. Physics of protein−DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2088−2095.

(12) Sheinman, M.; Bénichou, O.; Kafri, Y.; Voituriez, R. Classes of fast and specific search mechanisms for proteins on DNA. *Rep. Prog. Phys.* **2012**, *75*, 026601.

(13) Coppey, M.; Bénichou, O.; Voituriez, R.; Moreau, M. Kinetics of target site localization of a protein on DNA: a stochastic approach. *Biophys. J.* **2004**, *87*, 1640−1649.
(14) Lomholt, M. A.; van den Broek, B.; Kalisch, S.-M. J.; Wuite, G. J.; Metzler, R. Facilitated diffusion with DNA coiling. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 8204−8208.
(15) Veksler, A.; Kolomeisky, A. B. Speed-Selectivity Paradox in the Protein Search for Targets on DNA: Is It Real or Not? *J. Phys. Chem. B* **2013**, *117*, 12695−12701.
(16) Godec, A.; Metzler, R. Universal proximity effect in target search kinetics in the few-encounter limit. *Phys. Rev. X* **2016**, *6*, 041037.
(17) Kochugaeva, M. P.; Shvets, A. A.; Kolomeisky, A. B. How conformational dynamics influences the protein search for targets on DNA. *J. Phys. A: Math. Theor.* **2016**, *49*, 444004.
(18) Shin, J.; Kolomeisky, A. B. Molecular search with conformational change: One-dimensional discrete-state stochastic model. *J. Chem. Phys.* **2018**, *149*, 174104.
(19) Shin, J.; Kolomeisky, A. B. Target search on DNA by interacting molecules: First-passage approach. *J. Chem. Phys.* **2019**, *151*, 125101.
(20) Felipe, C.; Shin, J.; Loginova, Y.; Kolomeisky, A. B. The effect of obstacles in multi-site protein target search with DNA looping. *J. Chem. Phys.* **2020**, *152*, 025101.
(21) Gomez, D.; Klumpp, S. Facilitated diffusion in the presence of obstacles on the DNA. *Phys. Chem. Chem. Phys.* **2016**, *18*, 11184−11192.
(22) Cencini, M.; Pigolotti, S. Energetic funnel facilitates facilitated diffusion. *Nucleic Acids Res.* **2018**, *46*, 558−567.
(23) Barbi, M.; Place, C.; Popkov, V.; Salerno, M. Base-sequence-dependent sliding of proteins on DNA. *Phys. Rev. E* **2004**, *70*, 041901.
(24) Wunderlich, Z.; Mirny, L. A. Spatial effects on the speed and reliability of protein−DNA search. *Nucleic Acids Res.* **2008**, *36*, 3570−3578.
(25) Alsallaq, R.; Zhou, H.-X. Protein association with circular DNA: rate enhancement by nonspecific binding. *J. Chem. Phys.* **2008**, *128*, 115108.
(26) Koslover, E. F.; de la Rosa, M. A. D.; Spakowitz, A. J. Theoretical and computational modeling of target-site search kinetics in vitro and in vivo. *Biophys. J.* **2011**, *101*, 856−865.
(27) Foffano, G.; Marenduzzo, D.; Orlandini, E. Facilitated diffusion on confined DNA. *Phys. Rev. E* **2012**, *85*, 021919.
(28) Bauer, M.; Metzler, R. In vivo facilitated diffusion model. *PLoS One* **2013**, *8*, No. e53956.
(29) Amitai, A. Chromatin configuration affects the dynamics and distribution of a transiently interacting protein. *Biophys. J.* **2018**, *114*, 766−771.
(30) Cortini, R.; Filion, G. J. Theoretical principles of transcription factor traffic on folded chromatin. *Nat. Commun.* **2018**, *9*, 1740.
(31) Mondal, A.; Bhattacherjee, A. Understanding the role of DNA topology in target search dynamics of proteins. *J. Phys. Chem. B* **2017**, *121*, 9372−9381.
(32) Mondal, K.; Chaudhury, S. Effect of DNA Conformation on the Protein Search for Targets on DNA: A Theoretical Perspective. *J. Phys. Chem. B* **2020**, *124*, 3518−3526.
(33) van Holde, K.; Zlatanova, J. Unusual DNA structures, chromatin and transcription. *BioEssays* **1994**, *16*, 59−68.
(34) Choi, J.; Majima, T. Conformational changes of non-B DNA. *Chem. Soc. Rev.* **2011**, *40*, 5893−5909.
(35) Nyberg, M.; Ambjörnsson, T.; Stenberg, P.; Lizana, L. Modelling Protein Target-Search in Human Chromosomes. *Phys.Rev. Research* **2021**, *3*, 013055.
(36) Hofmann, A.; Heermann, D. W. The role of loops on the order of eukaryotes and prokaryotes. *FEBS Lett.* **2015**, *589*, 2958−2965.
(37) Vilar, J. M.; Leibler, S. DNA Looping and Physical Constraints on Transcription Regulation. *J. Mol. Biol.* **2003**, *331*, 981−989.
(38) Ito, Y.; Nativio, R.; Murrell, A. Induced DNA demethylation can reshape chromatin topology at the IGF2-H19 locus. *Nucleic Acids Res.* **2013**, *41*, 5290−5302.
(39) Rao, S. S.; Huntley, M. H.; Durand, N. C.; Stamenova, E. K.; Bochkov, I. D.; Robinson, J. T.; Sanborn, A. L.; Machol, I.; Omer, A. D.; Lander, E. S.; Aiden, E. L. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **2014**, *159*, 1665−1680.
(40) Shin, J.; Kolomeisky, A. B. Facilitation of DNA loop formation by protein−DNA non-specific interactions. *Soft Matter* **2019**, *15*, 5255−5263.
(41) Hu, T.; Grosberg, A. Y.; Shklovskii, B. How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophys. J.* **2006**, *90*, 2731−2744.
(42) Shvets, A. A.; Kolomeisky, A. B. The Role of DNA Looping in the Search for Specific Targets on DNA by Multisite Proteins. *J. Phys. Chem. Lett.* **2016**, *7*, 5022−5027.
(43) Shvets, A. A.; Kolomeisky, A. B. Sequence heterogeneity accelerates protein search for targets on DNA. *J. Chem. Phys.* **2015**, *143*, 245101.
(44) Loverdo, C.; Benichou, O.; Voituriez, R.; Biebricher, A.; Bonnet, I.; Desbiolles, P. Quantifying hopping and jumping in facilitated diffusion of DNA-binding proteins. *Phys. Rev. Lett.* **2009**, *102*, 188101.
(45) Baranello, L.; Levens, D.; Gupta, A.; Kouzine, F. The importance of being supercoiled: how DNA mechanics regulate dynamic processes. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2012**, *1819*, 632−638.

https://dx.doi.org/10.1021/acs.jpcb.0c09599
J. Phys. Chem. B 2021, 125, 1727−1734