

The effect of obstacles in multi-site protein target search with DNA looping

Cite as: J. Chem. Phys. **152**, 025101 (2020); <https://doi.org/10.1063/1.5135917>

Submitted: 08 November 2019 . Accepted: 13 December 2019 . Published Online: 08 January 2020

Cayke Felipe , Jaeoh Shin , Yulia Loginova, and Anatoly B. Kolomeisky 

COLLECTIONS

 This paper was selected as an Editor's Pick



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[On the transferability of ion parameters to the TIP4P/2005 water model using molecular dynamics simulations](#)

The Journal of Chemical Physics **152**, 024501 (2020); <https://doi.org/10.1063/1.5124448>

[Polymer physics across scales: Modeling the multiscale behavior of functional soft materials and biological systems](#)

The Journal of Chemical Physics **151**, 230902 (2019); <https://doi.org/10.1063/1.5126852>

[Dissipation and dynamics in ultrafast intersystem crossings](#)

The Journal of Chemical Physics **152**, 024104 (2020); <https://doi.org/10.1063/1.5125005>



Lock-in Amplifiers

Zurich Instruments

Watch the Video 

The effect of obstacles in multi-site protein target search with DNA looping

Cite as: J. Chem. Phys. 152, 025101 (2020); doi: 10.1063/1.5135917

Submitted: 8 November 2019 • Accepted: 13 December 2019 •

Published Online: 8 January 2020



View Online



Export Citation



CrossMark

Cayke Felipe,¹  Jaeoh Shin,^{2,3}  Yulia Loginova,⁴ and Anatoly B. Kolomeisky^{1,2,3,5,a)} 

AFFILIATIONS

¹Department of Physics, Rice University, Houston, Texas 77005, USA

²Department of Chemistry, Rice University, Houston, Texas 77005, USA

³Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, USA

⁴Department of Chemistry, Moscow State University, Moscow, Russia

⁵Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77005, USA

^{a)}tolya@rice.edu

ABSTRACT

Many fundamental biological processes are regulated by protein-DNA complexes called *synaptosomes*, which possess multiple interaction sites. Despite the critical importance of synaptosomes, the mechanisms of their formation are not well understood. Because of the multisite nature of participating proteins, it is widely believed that their search for specific sites on DNA involves the formation and breaking of DNA loops and sliding in the looped configurations. In reality, DNA in live cells is densely covered by other biological molecules that might interfere with the formation of synaptosomes. In this work, we developed a theoretical approach to evaluate the role of obstacles in the target search of multisite proteins when the formation of DNA loops and the sliding in looped configurations are possible. Our theoretical method is based on analysis of a discrete-state stochastic model that uses a master equations approach and extensive computer simulations. It is found that the obstacle slows down the search dynamics in the system when DNA loops are long-lived, but the effect is minimal for short-lived DNA loops. In addition, the relative positions of the target and the obstacle strongly influence the target search kinetics. Furthermore, the presence of the obstacle might increase the noise in the system. These observations are discussed using physical-chemical arguments. Our theoretical approach clarifies the molecular mechanisms of formation of protein-DNA complexes with multiple interactions sites.

Published under license by AIP Publishing. <https://doi.org/10.1063/1.5135917>

I. INTRODUCTION

Various genetic modifications are critically important for genome maintenance of living systems and for the ability of cellular organisms to adapt during the course of evolution.^{1,2} The examples of such processes include site-specific recombinations, gene regulation, genome rearrangements, and genome integration.^{4,3-6} Large protein-DNA molecular assemblies known as *synaptic complexes* or *synaptosomes* typically control them.^{1-4,7} Although these complexes play important biological roles, it is still unclear how they form so quickly and efficiently given the very complex nature of the cellular environment.³

The proteins in the synaptosomes can associate to DNA molecules at multiple sites. It is assumed that the overall process of formation of synaptosomes consists of several steps of

sequential protein binding to the corresponding sites on DNA.^{1,2,4,7} This suggests that the final steps of the synaptosome assembly take place when the protein is already bound to DNA at some locations, as shown in Fig. 1. Consequently, the search for the presently unoccupied specific sites should involve the formation of protein-facilitated DNA loops via nonspecific protein-DNA interactions and protein sliding in the looped configurations. The appearance of such complex topological structures significantly complicates the understanding of the molecular mechanisms of the formation of synaptic complexes.

It is known that the single-site proteins find their specific sites on DNA via a combination of three-dimensional (3D) bulk diffusion and one-dimensional (1D) sliding of nonspecifically bound proteins along the DNA chain.^{4,8-19} This alternating mechanism can dramatically accelerate the search process, and it was

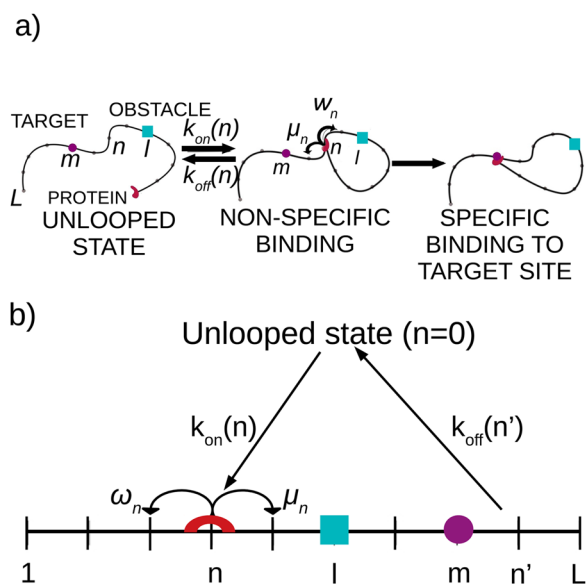


FIG. 1. (a) Schematic view of two-site protein search for the specific site on DNA when one site is already bound to the DNA end. Red arc describes the protein's molecule. Purple circle describes the target site. Cyan square describes the obstacle particle. Black arrows indicate allowed transitions. (b) A corresponding discrete-state stochastic model of the search process.

intensively investigated in recent years.^{16,18,19} However, the appearance of DNA loops for multisite proteins makes the search process significantly more complex. The theoretical methods used for investigating the target search of single-site proteins cannot be used for these systems. We recently introduced a theoretical framework for taking into account the appearance of DNA loops in the multisite protein target search.^{20,21} This theoretical approach uses a discrete-state stochastic description, and it combines analytical first-passage probabilities calculations with extensive Monte Carlo computer simulations. After identifying several dynamic search regimes, it was shown that at some conditions, the loop formation can lead to the accelerated target search.^{20,21}

A large number of various macromolecules occupy the cytoplasm and cover DNA in live cells.^{1,2,22–24} This should influence the target search dynamics in various ways, including the anomalous diffusion²⁵ in bulk and the protein sliding motion along the DNA chain. The role of these biological obstacles on DNA has been investigated theoretically for the single-site proteins.^{24,26–28} These studies identified the mobility of obstacles and the strength of nonspecific protein-DNA interactions as key features influencing the target search dynamics. However, the role of obstacles for multisite protein search has never been investigated.

In this paper, we present a theoretical study on the role of a static obstacle in the target search of multisite proteins for specific sites on DNA, considering specifically the formation of DNA loops and sliding in the looped configurations. Using a combination of analytical and computer simulations, the multisite proteins search dynamics is analyzed for general sets of conditions. Our theoretical calculations suggest that the lifetimes of DNA-looped

configurations and relative positions of the target and the obstacle are the most important factors determining the search dynamics of multisite proteins. In addition, adding the obstacle might increase the overall level of the noise in the system. Physical-chemical arguments to explain these observations are also presented.

II. THEORETICAL MODEL

A. Description

A model for the multisite protein target search with the obstacle is presented in Fig. 1. We will concentrate on the last stages of the formation of the protein-DNA complex, which involve DNA loops and sliding in the looped configurations. Let us consider a single protein molecule with 2 binding sites, and one of them is already associated tightly to the end of the DNA chain with L binding sites [see Fig. 1(a)]. We set the size of each binding site to be 10 base pairs (bp) to reflect the typical size of the region of protein-DNA interactions. One of these sites at the location m ($1 \leq m \leq L$) is the target sequence for the second site on the protein molecule. In addition, there is an obstacle particle at the site $l \neq m$ that stays forever at the same location. In this study, we consider the static obstacles because, as was shown recently, they influence the search dynamics more than the mobile obstacles.²⁶ Due to nonspecific interactions, the protein molecule can nonspecifically associate to DNA at any site, and this leads to the formation of DNA loops [Fig. 1(a)]. In the looped conformation, the protein might slide along the DNA chain, but it cannot pass the obstacle.

It is important to note that since we are trying to understand the complex mechanisms of multisite protein search using a minimal theoretical model, there are several approximations in our approach.²¹ They include the assumption that the chain segments in the DNA loop can quickly relax to equilibrium, which is reasonable for not very long DNA chains. In addition, we assume that the location of nonspecific bindings is not correlated. Furthermore, the possible appearance of DNA supercoiling and twists during the motion in looped configurations and DNA sequence effects are also neglected. A more detailed discussion on these issues and on the validity of these approximations can be found in Ref. 21.

Now, we can construct a discrete-state description of the model, as presented in Fig. 1(b). The overall system consisting of the single two-site protein bound to the end of DNA and the DNA chain can be viewed as $L + 1$ discrete chemical states. The state n ($1 \leq n \leq L$) describes the DNA looped state when the loop of size n is formed, and there are $L - n$ DNA sites that are not in the loop. For $n = 0$, we have a state without the loop, i.e., the protein is not associated with DNA via nonspecific interactions (but recall that one site is always bound to the DNA end, see Fig. 1). From this unlooped conformation, the protein molecule can associate to DNA at the site n with a rate $k_{on}(n)$. The corresponding dissociation rate from the state n is $k_{off}(n)$. The protein binding to DNA is associated with a binding energy ϵ (enthalpic contribution), where $\epsilon < 0$ describe the attractions and $\epsilon > 0$ correspond to repulsions. In the looped conformations, the protein can slide along the DNA with position-dependent rates: The transition $n \rightarrow n - 1$ (reducing the loop size) takes place with a rate w_n , while the transition $n \rightarrow n + 1$ (increasing the loop size) takes place with a rate μ_n (Fig. 1).

The formation of DNA loops and sliding in the loop configurations modify the free energy of the system, and this is the reason for position-dependent association/dissociation and sliding rates. If all these transitions are slower than the chain relaxation, the free energy of the system can be well approximated as²¹

$$G(n) = G_0(n) + \epsilon = \frac{A}{n} + \alpha \log[n] + \epsilon, \quad (1)$$

where $G_0(n)$ is the free energy of the formation of the loop of size n without enthalpic contribution. The first term in this expression corresponds to the bending energy of DNA, the second term describes the entropy cost of creating the loop (see also Ref. 29), and the last term is the enthalpic contribution due to the formation of the non-specific protein-DNA bond. The coefficient A is the bending stiffness of the DNA chain, and for a circular loop, it is given by $A = 2\pi^2 l_p$, where l_p is the persistence length. Below, for specific calculations, we set the chain length as $L = 300$ and the parameter $A = 300$, which in real units corresponds to $L = 3000$ bps and $l_p = 150$ bps. Note that in the presence of confinement, the bending energy term can be strongly modified, which will affect the loop formation dynamics.³⁰ The coefficient α is related to a scaling exponent for the radius of gyration, and for an ideal Gaussian chain, it is equal to $\alpha = 3/2$. The free energy profile for these parameters is presented in Fig. 2 (with $\epsilon = -2 k_B T$). We note that this free energy expression is one of the simplest approximations which neglects many details, including the self-avoiding effects²⁹ (see Ref. 21 for more discussion on this issue).

Knowing the free-energy profile of the system allow us to write explicit expressions for the transition rates,

$$k_{\text{on}}(n) = k_{\text{on}}^{(0)} \exp[-\theta G_0(n)], \quad (2)$$

$$k_{\text{off}}(n) = k_{\text{off}}^{(0)} \exp[(1 - \theta)G_0(n)], \quad (3)$$

$$\mu_n = \mu_0 \exp[-\theta_t \Delta G(n + 1)], \quad (4)$$

$$w_n = \mu_0 \exp[(1 - \theta_t) \Delta G(n)]. \quad (5)$$

Here, $k_{\text{on}}^{(0)}$, $k_{\text{off}}^{(0)}$, and μ_0 are the binding, unbinding, and hopping rates, respectively, in the absence of the loop formation [$G(n) = 0$], and $\Delta G(n) \equiv G(n) - G(n - 1)$. Additionally, parameters θ and θ_t specify how the free energy change is distributed between the

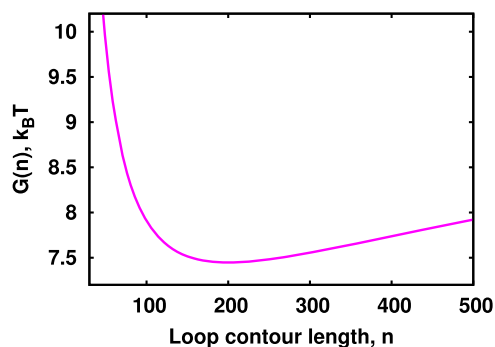


FIG. 2. Free-energy profile of the two-site protein target search as a function of the loop contour length n . The following parameters were used: $A = 300$, $\alpha = 3/2$, and $\epsilon = -2 k_B T$.

binding and unbinding transitions and for the sliding rates in different directions, respectively. For convenience, we set both of them to be $\theta = \theta_t = 0.5$. It has been argued before that changing the values for these parameters does not change the qualitative behavior of the system.²¹ In the following calculations, we also take $\mu_0 \simeq 60 \text{ s}^{-1}$ from the experimental measurements of the diffusion constant of p53 protein,³¹ and both transition rates $k_{\text{on}}^{(0)}$ and $k_{\text{off}}^{(0)}$ are varied for a wide range of values.

B. Analytical calculations in limiting cases

It is not possible to explicitly analyze the discrete-state stochastic model for general sets of parameters. However, there are two limiting situations that allow for full analytical solutions, providing a valuable physical insights into the molecular mechanisms of the target search.

Let us start with the case of short-lived DNA loops. Then, the protein molecule after binding to DNA does not have much time to slide and it quickly dissociates. Clearly, the obstacle does not affect much the dynamics of the target search. As the protein barely slides in this regime, it effectively does not encounter the obstacle at all times. This is a so-called “no-sliding limit” that was analytically fully investigated in Refs. 20 and 21. For the mean search time T , it was shown that

$$T = \frac{1 + \sum_{i \neq \{m, l\}} \frac{k_{\text{on}}(i)}{k_{\text{off}}(i)}}{k_{\text{on}}(m)}. \quad (6)$$

This expression has a simple physical meaning. Because there are no correlations in binding/unbinding events, on average, the protein has to visit every site on DNA (except the one occupied by the obstacle) before the target can be found.

Another limiting situation that can be analytically solved is the case with long-lived DNA loops. In this regime, once the protein binds, it will stay on DNA for a long time. Then, there are two possible situations. If the protein lands on the DNA segment where the obstacle does not prevent it from sliding to the target, it will eventually find it without dissociating. However, if the protein binds to the DNA segment where the obstacle is the barrier for reaching the target that cannot be passed, the protein will spend a lot of time on this segment before dissociating. We can divide the DNA chain into two segments separated by the obstacle, as shown in Fig. 3. The region that does not include the target is labeled S_1 , and the other region is

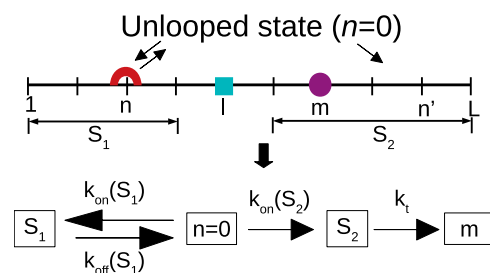


FIG. 3. A discrete-state stochastic model for the multisite protein target search in the limit of long-lived DNA loops. The system is described as three distinct states, labeled S_1 , $n = 0$, and S_2 , respectively.

labeled S_2 (see Fig. 3). We consider stationary state conditions which means that as soon as the protein reaches the target site, the system starts immediately in the state $n = 0$.³² From the steady-state flux J to the target, the mean search time can be estimated as $T = 1/J$. Let us define $P_0(t)$, $P_1(t)$, and $P_2(t)$ as the probabilities to be in the state $n = 0$, in the state S_1 , or in the state S_2 , respectively, at time t . Then, the dynamics in the system at all times is described by the following (forward) master equations:

$$\frac{\partial P_0(t)}{\partial t} = -[k_{\text{on}}(S_1) + k_{\text{on}}(S_2)]P_0(t) + k_{\text{off}}(S_1)P_1(t) + k_t P_2(t), \quad (7)$$

$$\frac{\partial P_1(t)}{\partial t} = -k_{\text{off}}(S_1)P_1(t) + k_{\text{on}}(S_1)P_0(t), \quad (8)$$

$$\frac{\partial P_2(t)}{\partial t} = -k_t P_2(t) + k_{\text{on}}(S_2)P_0(t). \quad (9)$$

In these expressions, $k_{\text{on}}(S_1)$, $k_{\text{on}}(S_2)$, and $k_{\text{off}}(S_1)$ are transition rates to associate and dissociate from each region, which can be determined from the following expressions:

$$k_{\text{on}}(S_1) = \sum_{n \in S_1} k_{\text{on}}(n), \quad (10)$$

$$k_{\text{on}}(S_2) = \sum_{n \in S_2} k_{\text{on}}(n), \quad (11)$$

$$k_{\text{off}}(S_1) = \frac{\sum_{n \in S_1} k_{\text{off}}(n) \exp[-G(n)]}{\sum_{n \in S_1} \exp[-G(n)]}. \quad (12)$$

The physical meaning of these expressions is the following: The total rate to reach the segment S_1 or S_2 is the sum of all corresponding association rates to each segment. The total dissociation rate out of the segment S_1 is the sum over all dissociation rates from all sites in the segment weighted by the probability to be found at each site. The parameter k_t is the average rate to slide to the target at the site m after reaching the region S_2 . The explicit expression for this parameter can be obtained using the analysis developed in Ref. 33,

$$k_t^{-1} = \frac{1}{k_{\text{on}}(S_2)} \sum_{n \in S_2} k_{\text{on}}(n) T(n), \quad (13)$$

where $k_{\text{on}}(n)/k_{\text{on}}(S_2)$ is the probability to bind to DNA segment S_2 at the site n and $T(n)$ is the mean first-passage time to reach the target site m starting at the initial position n , which is given by³³

$$T(n) = \begin{cases} \sum_{K=n}^{m-1} \left(\sum_{i=1}^K \frac{1}{\mu_i} \prod_{j=i+1}^K \frac{w_j}{\mu_j} \right), & \text{if } n < m \\ \sum_{K=L+1-n}^{L-m} \left(\sum_{i=1}^K \frac{1}{\omega_{L+1-i}} \prod_{j=i+1}^K \frac{\mu_{L+1-j}}{w_{L+1-j}} \right), & \text{if } n > m \end{cases}. \quad (14)$$

Under stationary conditions, we have $\frac{\partial P_0(t)}{\partial t} = \frac{\partial P_1(t)}{\partial t} = \frac{\partial P_2(t)}{\partial t} = 0$, and Eqs. (7)–(9) can be easily solved, producing stationary probabilities P_0 , P_1 , and P_2 . This allows us to calculate the flux to the target as $J = k_t P_2 + k_{\text{on}}(m) P_0$. Finally, the mean search time is given by

$$T = J^{-1} = \left[\frac{1}{k_t} + \frac{k_{\text{on}}(S_1) + k_{\text{off}}(S_1)}{k_{\text{off}}(S_1) k_{\text{on}}(S_2)} \right] \frac{k_{\text{on}}(S_2)}{k_{\text{on}}(S_2) + k_{\text{on}}(m)}. \quad (15)$$

III. RESULTS

To test analytical predictions in the limiting cases and to probe the dynamics in the system for the general range of parameters, extensive Monte Carlo computer simulations that utilized a Gillespie algorithm were performed.³⁴ We introduce a scanning length, $\lambda_0 \equiv \sqrt{\frac{\mu_0}{k_{\text{off}}^{(0)}}}$, as an important parameter that quantifies the search process. It is defined as the average length the protein slides along the DNA chain before dissociating under no potential [$G_0(n) = 0$]. The actual sliding length will be modified by the effect of the real free-energy profile $G(n)$. The scanning length also correlates with the lifetime of DNA looped configurations: the longer λ_0 , the more the systems spends in the looped states.

The results of computer simulations and analytical predictions are presented in Fig. 4 where the mean search times as a function of the scanning length λ_0 are plotted for different association rates. Three search dynamic regimes can be identified. For short scanning lengths ($\lambda_0 \leq 1$), which corresponds to short-lived DNA looped conformations, the search is independent of the obstacle and it also does not depend on the scanning length. This is a “no-sliding” regime where the protein does not slide much on DNA. In this dynamic regime, our analytical predictions perfectly agree with the results of computer simulations. Increasing the scanning length ($1 \leq \lambda_0 \leq L$, intermediate lifetimes for DNA-looped states) accelerates the search dynamics. This is because in addition to coming directly from the solution, the protein can also reach the target via sliding along the DNA chain. However, the obstacle again does not affect the search dynamics—this can be seen by comparing the results for our model with the obstacle with the model without obstacles that was investigated before (dashed curves in Fig. 4).²¹ The situation dramatically changes in the regime of long scanning lengths ($\lambda_0 > L$), which corresponds to the long-lived DNA-looped conformations. Here, the obstacle significantly slows down the search dynamics. Since in this regime, the protein mostly reaches the target via sliding, the protein occasionally maybe trapped in configurations where the obstacle prevents reaching the target (like the region S_1 in Fig. 3). One could

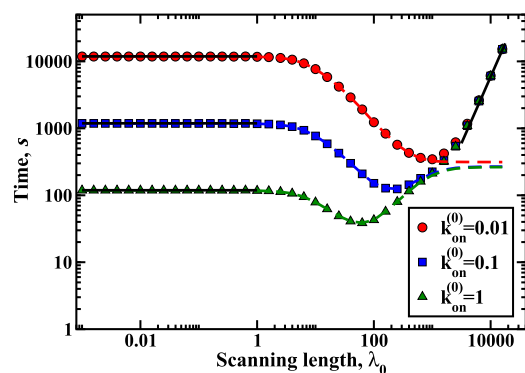


FIG. 4. Mean search times as a function of the sliding length λ_0 in the absence (dashed lines) and in the presence of obstacle (symbols and solid curves) for three different values of the protein association rates. The black curves are the analytic results for 2 limiting regimes: for no sliding regime ($\lambda_0 < 1$) and for sliding dominated regime ($\lambda_0 > L$). The following parameters were used in calculations: $\mu_0 = 60 \text{ s}^{-1}$, $L = m = 300$, and $l = 150$.

also see that in this dynamic regime, our analytical calculations fully agree with computer simulations.

The next question we investigated is the role of relative positions of the target and the obstacle in the dynamic regimes with dominating long-lived DNA loops. The results are presented in Fig. 5 where the target position is varied, while the obstacle position is fixed in the middle of the DNA stand. The search time is significantly larger if the target is located for $1 \leq m < l$ in comparison with positioning the target to $l < m \leq L$: In Fig. 4, we employed $l = 150$ and $L = 300$. To understand this result, let us recall that according to Fig. 2, it costs much more energy to make short loops ($1 \leq n \leq 150$), and this means that the protein has a lower probability to reach this part of DNA. The protein can also dissociate more easily from this region. So putting the target in this segment will slow down the search due to slow association rates and multiple visits. However, the segment $150 \leq n \leq 300$ is energetically much more favorable, and putting the target there will lead to faster search due to faster association rates and not so many repeat visits. Our calculations also show that in all situations for large scanning lengths (long-lived DNA loops), the search time in the presence of the obstacle is always slower than in the case without the obstacle, as illustrated in Fig. 5.

In the related study, we fixed the position of the target m and varied the position of the obstacle l , and the results are illustrated in Fig. 6, where the target is at the DNA end, $m = L$. When the obstacle is found far away from the target at the other end of DNA, the search is not affected by the presence of the obstacle. This is because the formation of short loops is energetically unfavorable so that the protein does not go frequently to this region. In addition, the obstacle is too far away from the target to affect the sliding in the direction of the target. However, moving the obstacle closer to the target increases the mean search time since the obstacle now works as the barrier that cannot be passed over. The protein must dissociate more frequently in order not to be trapped in the region $n < l$.

Until now, we focused on the mean times as a measure of the searching dynamics of multisite proteins, but the distribution of search times should provide a more comprehensive description

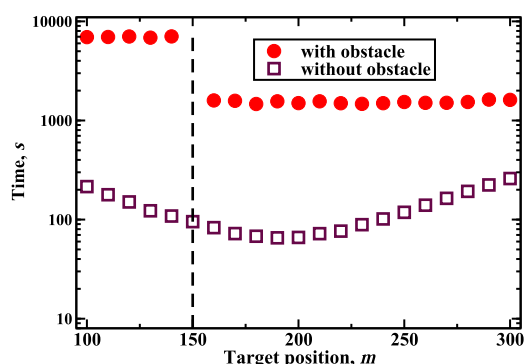


FIG. 5. Mean search time as a function of the target position m for large scanning lengths with an obstacle (filled circles) and without an obstacle (empty squares). The following parameters were used in calculations: $\lambda_0 = 5000$, $l = 150$, $k_{\text{on}}^{(0)} = 1 \text{ s}^{-1}$, $L = 300$, and $\mu_0 = 60 \text{ s}^{-1}$. The vertical dashed line indicates the obstacle position.

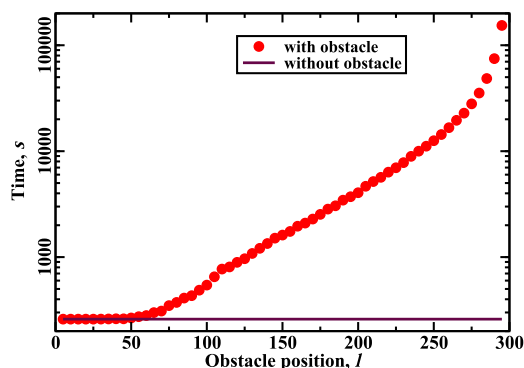


FIG. 6. Mean search time as a function of the obstacle position l for the fixed position of the target. Circles are for the model with the obstacle, while the solid line is for the model without the obstacle for the same conditions. The following parameters were used in calculations: $\lambda_0 = 5000$, $L = 300$, $m = 300$, $k_{\text{on}}^{(0)} = 1 \text{ s}^{-1}$, and $\mu_0 = 60 \text{ s}^{-1}$.

of the process.^{35–37} In order to understand how the presence of the obstacle influences the distribution of search times, we analyzed the relative standard deviation (RSD), which is the ratio of the standard deviation of search times to the mean search time, and the results are presented in Fig. 7. This quantity also measures the degree of stochastic noise in the system. One can see that for short and intermediate scanning lengths, which correspond to the case when the lifetimes of the DNA-looped states are not large, the RSD is close to 1 and it does not depend on the presence of the obstacle. We also find that the distribution of search times is close to the exponential with a relatively small degree of noise in the system. However, for the regime when the long-lived DNA loops dominate (large scanning lengths), the presence of the obstacle affects the distribution of search times. The distribution becomes wider (see Fig. 8), which also corresponds to the increase in the stochastic noise in

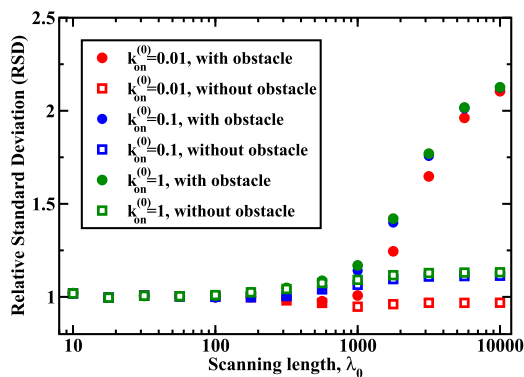


FIG. 7. Relative standard deviation of the search times as a function of the scanning length λ_0 for different values of association rates. Filled circles correspond to the systems with the obstacle, and empty squares describe the systems without obstacle. The following parameters were used in calculations: $L = 300$, $m = 300$, $l = 150$, and $\mu_0 = 60 \text{ s}^{-1}$.

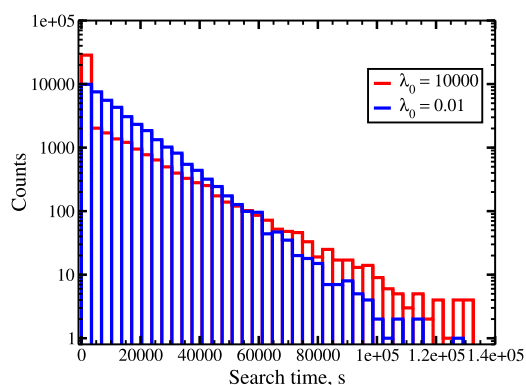


FIG. 8. The distribution of the search times for two values of λ_0 . The following parameters were used in calculations: $L = 300$, $m = 300$, $l = 150$, and $\mu_0 = 60 \text{ s}^{-1}$.

the system. In this regime, when the protein binds to the DNA segment from which it cannot slide to the target, it has to dissociate to bulk for the eventual association to the correct sequence on the DNA. These multiple association and dissociation events result in the distribution of the search time deviating from the exponential distribution. In this case, the mean value of the search time is no longer a good measure of the kinetics, but the full distribution should be considered (see Fig. 8),^{35–37} which is a future direction of our study.

We can evaluate the distribution of the search times using Monte Carlo simulations. The results are presented in Fig. 8. One can see that increasing the scanning length makes the distribution of search times wider, which corresponds to increasing the noise in the system.

Our theoretical calculations provide important molecular insights into the role of obstacles in the multisite protein target search. They suggest that from the biological point of view, the presence of obstacle does not always lead to significant delays in the formation of protein-DNA complexes, in contrast to naive expectations. If the DNA loops lifetimes are not too short and too long, the system can provide a fast target search and it can also avoid the negative effect of the obstacle. This range of parameters also does not increase the noise in the system, which is important for robustness of biological systems. In addition, the negative effects of the obstacles can be moderated by proper arrangements of specific target sites with respect to them. It will be interesting to check if real biological systems satisfy these conditions.

IV. SUMMARY AND CONCLUSIONS

In this work, a theoretical analysis on the role of static obstacles in the formation of protein-DNA complexes with multiple interaction sites is presented. We concentrated on the final stages of the multisite protein target search that involves the formation of DNA loops and sliding in the looped configurations. A discrete-state stochastic model for this process is developed, and it is investigated using exact analytical calculations and Monte Carlo computer simulations. It is found that the static obstacle does not interfere with the target search dynamics if the lifetimes of the DNA-looped

configurations are not too long. However, in the situations when the DNA loops are long lived, the search dynamics is significantly slowed down, and the stochastic noise in the system also increases. In addition, the search dynamics might be affected by the relative location of the target and the obstacle. It is argued that the biological system might prefer conditions with intermediate lifetimes of DNA-looped states since it leads to the fastest search dynamics without much effect from the obstacles.

Although our theoretical approach identifies the most relevant features of the multisite protein search dynamics in the presence of obstacles, we note that the presented theoretical picture is rather limited with many approximations. Our method does not take into account the sequence heterogeneity of DNA molecules, although this might be an important effect.³⁸ In addition, more advanced descriptions of the free-energy profile for polymers with loops are available. Our theoretical analysis will not work for very long DNA chains because the DNA chain will not be able to relax fast enough. Another complication is the assumption of no correlation in the protein consecutive bindings. Furthermore, our theoretical approach neglects the possibility of DNA supercoiling and twists, which should complicate the overall search dynamics. Despite these considerations, the presented theoretical model clarifies some important molecular details of very complex biological processes of formation of protein-DNA complexes. It will be crucial to test our theoretical conclusions in experiments as well as in more advanced theoretical studies.

ACKNOWLEDGMENTS

The work was supported by the Welch Foundation (Grant No. C-1559), by the NSF (Grant Nos. CHE-1664218 and MCB-1941106), and by the Center for Theoretical Biological Physics sponsored by the NSF (Grant No. PHY-1427654).

REFERENCES

- 1 B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 6th ed. (WW Norton & Company, NY, 2014).
- 2 H. Lodish, A. Berk, C. A. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh, P. Matsudaira *et al.*, *Molecular Cell Biology* (Macmillan, 2008).
- 3 N. D. Grindley, K. L. Whiteson, and P. A. Rice, *Annu. Rev. Biochem.* **75**, 567 (2006).
- 4 S. E. Halford and J. F. Marko, *Nucleic Acids Res.* **32**, 3040 (2004).
- 5 R.-S. Mani and A. M. Chinnaiyan, *Nat. Rev. Genet.* **11**, 819 (2010).
- 6 F. Bushman, M. Lewinski, A. Ciuffi, S. Barr, J. Leipzig, S. Hannenhalli, and C. Hoffmann, *Nat. Rev. Microbiol.* **3**, 848 (2005).
- 7 A. Cournac and J. Plumbridge, *J. Bacteriol.* **195**, 1109 (2013).
- 8 A. D. Riggs, S. Bourgeois, and M. Cohn, *J. Mol. Biol.* **53**, 401 (1970).
- 9 O. G. Berg and C. Blomberg, *Biophys. Chem.* **4**, 367 (1976).
- 10 R. B. Winter and P. H. Von Hippel, *Biochemistry* **20**, 6948 (1981).
- 11 P. H. von Hippel and O. G. Berg, *J. Biol. Chem.* **264**, 675 (1989).
- 12 M. Coppey, O. Bénichou, R. Voituriez, and M. Moreau, *Biophys. J.* **87**, 1640 (2004).
- 13 M. A. Lomholt, B. van den Broek, S.-M. J. Kalisch, G. J. Wuite, and R. Metzler, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8204 (2009).
- 14 A. Vekslar and A. B. Kolomeisky, *J. Phys. Chem. B* **117**, 12695 (2013).
- 15 A. Esadze, C. A. Kemme, A. B. Kolomeisky, and J. Iwahara, *Nucleic Acids Res.* **42**, 7039 (2014).
- 16 L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj, *J. Phys. A: Math. Theor.* **42**, 434013 (2009).

- ¹⁷O. Bénichou, C. Loverdo, M. Moreau, and R. Voituriez, *Rev. Mod. Phys.* **83**, 81 (2011).
- ¹⁸A. B. Kolomeisky, *Phys. Chem. Chem. Phys.* **13**, 2088 (2011).
- ¹⁹M. Sheinman, O. Bénichou, Y. Kafri, and R. Voituriez, *Rep. Prog. Phys.* **75**, 026601 (2012).
- ²⁰A. A. Shvets and A. B. Kolomeisky, *J. Phys. Chem. Lett.* **7**, 5022 (2016).
- ²¹J. Shin and A. Kolomeisky, *Soft Matter* **15**, 5255 (2019).
- ²²E. G. Marklund, A. Mahmutovic, O. G. Berg, P. Hammar, D. van der Spoel, D. Fange, and J. Elf, *Proc. Natl. Acad. Sci. U. S. A.* **110**, 19796 (2013).
- ²³C. Brackley, M. Cates, and D. Marenduzzo, *Phys. Rev. Lett.* **111**, 108101 (2013).
- ²⁴M. Bauer, E. S. Rasmussen, M. A. Lomholt, and R. Metzler, *Sci. Rep.* **5**, 10072 (2015).
- ²⁵L. Liu, A. G. Cherstvy, and R. Metzler, *J. Phys. Chem. B* **121**, 1284 (2017).
- ²⁶A. Shvets, M. Kochugaeva, and A. B. Kolomeisky, *J. Phys. Chem. B* **120**, 5802 (2015).
- ²⁷A. A. Shvets and A. B. Kolomeisky, *J. Phys. Chem. Lett.* **7**, 2502 (2016).
- ²⁸D. Gomez and S. Klumpp, *Phys. Chem. Chem. Phys.* **18**, 11184 (2016).
- ²⁹A. Hanke and R. Metzler, *Biophys. J.* **85**, 167 (2003).
- ³⁰J. Shin, A. G. Cherstvy, and R. Metzler, *ACS Macro Lett.* **4**, 202 (2015).
- ³¹A. Tafvizi, F. Huang, J. S. Leith, A. R. Fersht, L. A. Mirny, and A. M. Van Oijen, *Biophys. J.* **95**, L01 (2008).
- ³²J. Shin and A. B. Kolomeisky, *J. Phys. Chem. B* **122**, 2243 (2018).
- ³³X. Li and A. B. Kolomeisky, *J. Chem. Phys.* **139**, 144106 (2013).
- ³⁴D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
- ³⁵O. Pulkkinen and R. Metzler, *Phys. Rev. Lett.* **110**, 198101 (2013).
- ³⁶A. Godec and R. Metzler, *Phys. Rev. X* **6**, 041037 (2016).
- ³⁷D. S. Grebenkov, R. Metzler, and G. Oshanin, *Commun. Chem.* **1**, 96 (2018).
- ³⁸A. A. Shvets and A. B. Kolomeisky, *J. Chem. Phys.* **143**, 245101 (2015).