THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Dynamics of the Protein Search for Targets on DNA in the Presence of Traps

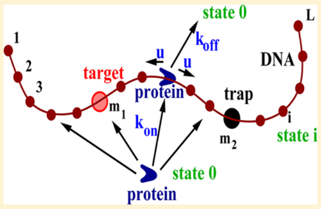
Martin Lange,[†,§] Maria Kochugaeva,[†,‡] and Anatoly B. Kolomeisky*,[†,‡]

[†]Department of Chemistry, Rice University, Houston, Texas 77005, United States

[‡]Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States

[§]Johannes Gutenberg University, Mainz 55122, Germany

ABSTRACT: Protein search for specific binding sites on DNA is a fundamental biological phenomenon associated with the beginning of most major biological processes. It is frequently found that proteins find and recognize their specific targets quickly and efficiently despite the complex nature of protein−DNA interactions in living cells. Although significant experimental and theoretical efforts were made in recent years, the mechanisms of these processes remain not well-clarified. We present a theoretical study of the protein target search dynamics in the presence of semispecific binding sites which are viewed as traps. Our theoretical approach employs a discrete-state stochastic method that accounts for the most important physical and chemical processes in the system. It also leads to a full analytical description for all dynamic properties of the protein search. It is found that the presence of traps can significantly modify the protein search dynamics. This effect depends on the spatial positions of the targets and traps, on distances between them, on the average sliding length of the protein along the DNA, and on the total length of DNA. Theoretical predictions are discussed using simple physical−chemical arguments, and they are also validated with extensive Monte Carlo computer simulations.



INTRODUCTION

All major biological processes are governed by protein−DNA interactions.[1] Many of them begin after a protein molecule first searches and then binds to a short segment of DNA with a specific sequence, which is known as a specific binding site. This allows proteins to effectively transfer the genetic information contained in DNA by initiating a cascade of biochemical processes relevant for the successful functioning of biological cells. Clearly, the protein search for targets on DNA is one of the most important phenomena in nature.[1] This subject was investigated for many years,[2−5] but significant progress was achieved in recent years with a development of advanced experimental and theoretical methods.[6−26] However, many details of the mechanisms of the protein search for targets on DNA still remain not well-understood.[27−31]

One of the most fascinating observations in this field is that many proteins can find and recognize their specific binding sites much faster than expected if the search would take place only via 3D bulk diffusion.[3,5,6,27,28] This surprising result is called a *facilitated diffusion*, and it is frequently argued that this happens due to the protein search being a combination of 3D and 1D modes.[3,5,6,27,28] More specifically, the proposed picture assumes that the protein molecule associates nonspecifically to DNA, scans some length of DNA by sliding, dissociates from DNA, and repeats these actions several times until the target is located. Several experiments support this point of view.[7,10,11,18] These observations also suggest that the specific nature of the protein−DNA interactions should have a stronger effect on the protein search dynamics. Indeed, in real systems there are many sequences that have structures and chemical compositions similar to the specific sites.[20,33] The protein molecule can be trapped in these semispecific sites for large periods of time, and it is not clear then how a fast search can be accomplished. However, the majority of theoretical models for the facilitated diffusion ignore this effect, assuming that the nonspecifically bound proteins slide along the homogeneous DNA chain with the same diffusion constant.[6,28] There are several theoretical studies that take into account the possibility of trapping.[12,32] They argue that the bound proteins can fluctuate between several conformations while still being associated with the DNA chain, and this leads to the avoidance of these semispecific binding sites. However, the molecular mechanisms of this avoidance are not clear, and there is no experimental proof for this.

In this paper, we present a theoretical investigation of the protein search for targets on DNA with semispecific binding sites that are viewed as traps. Our analysis uses a discrete-state stochastic approach[15,16] that explicitly takes into consideration major physical and chemical processes in the system. It allows us to obtain a full analytical description for all dynamic properties in the protein search by utilizing a method of first-passage processes. The application of the discrete-state stochastic method to the protein search without traps uncovered three dynamic regimes, depending on the relative values of the important length scales in the system.[16] For the protein sliding length λ larger than the length of the DNA chain L, the protein is involved in a 1D search process with a random-walk dynamics. When the sliding length is larger than the target

Received: July 28, 2015
Revised: August 27, 2015
Published: September 2, 2015

ACS Publications

DOI: 10.1021/acs.jpcb.5b07303

size but smaller than the length of DNA, the search mechanism combines 1D and 3D motions. For even smaller sliding lengths the diffusion along the DNA chain is not possible, and the protein searches for the target only via 3D motions. By generalizing and extending this method to the system with semispecific sites, we show that the traps have a strong effect on the search dynamics. Surprisingly, there are many counter-intuitive observations when the presence of the traps might accelerate the search. We also test our theoretical predictions with Monte Carlo computer simulations.

■ THEORETICAL METHODS

The discrete-state stochastic approach[15,16] can be generalized for analyzing the protein search with an arbitrary number of targets and traps on DNA. To capture the main features of the process, we consider a simpler model where a single protein molecule searches for one specific binding site on a single DNA chain, which also has one semispecific binding site, as presented in Figure 1. To simplify things even further, we assume that traps are irreversible; i.e., if the protein molecule binds to the semispecific site it will never dissociate. This is a strong assumption, but because the search times in many systems are quite short and experiments can be done only for finite periods of time, this should approximate the protein search dynamics reasonably well.

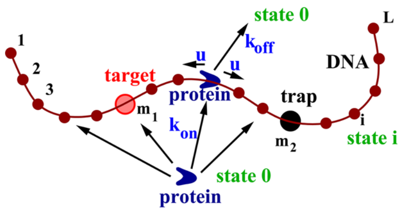


Figure 1. General scheme for the protein target search on DNA with a trap. There are $L - 2$ nonspecific, 1 specific, and 1 trap binding site on the DNA chain. The target is at site m_1, and the trap is at site m_2. A protein molecule can slide along the DNA chain with rate u, or might dissociate into the solution with the rate k_{off}. The bulk solution is labeled as a state 0. From the solution the protein can associate to any site on DNA with the total rate k_{on}.

We consider a single DNA molecule with $L - 2$ nonspecific binding sites and two special sites: one of them placed at m_1 is a specific target for the protein, while another one (at m_2) is a semispecific trap: see Figure 1. For convenience, we always assume that $m_1 < m_2$, i.e., the trap is always to the right of the target (Figure 1). But the specific order of the target and trap, obviously, does not affect the physics of this phenomenon. In *in vitro* experiments the protein molecule moves much faster in the bulk solution than on DNA, and we assume that it can reach any site on DNA with the same probability. A total protein association rate to DNA is equal to k_{on}, while the bound protein can dissociate into the solution with a rate k_{off}, as shown in Figure 1. In addition, the DNA-bound protein can slide along the chain with rate u in both directions (Figure 1). This rate can be viewed as a 1D diffusion constant for moving on DNA. The protein search always starts from the solution that we label as a state 0. There are two possible final outcomes of the search process. The protein molecule can find the target, and this is a successful event. Or, the protein might fall into the trap and never leave it: this is not a successful event. Thus, the probability to reach the target in this model is always less than one.

The protein search for the target can be associated with a first-passage process of reaching the specific binding site, and this provides a direct way of evaluating the dynamics of the system.[16] We introduce a function $F_n(t)$ which is defined as a probability to reach the target at time t for the first time, while not being trapped to the semispecific site, if initially (at $t = 0$) the protein molecule starts at the state n ($n = 0, 1, ..., L$). It is important to note that this is a *conditional* probability for the protein molecules that are not captured by the trap. The temporal evolution of these probabilities can be described via a set of backward master equations[15,16]

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

for $2 \leq n \leq L - 1$ and $n \neq m_1$ or m_2. At DNA boundaries the dynamics is slightly different

$$\frac{dF_1(t)}{dt} = uF_2(t) + k_{off}F_0(t) - (u + k_{off})F_1(t) \tag{2}$$

and

$$\frac{dF_L(t)}{dt} = uF_{L-1}(t) + k_{off}F_0(t) - (u + k_{off})F_L(t) \tag{3}$$

In addition, for the bulk solution ($n = 0$) we have

$$\frac{dF_0(t)}{dt} = \frac{k_{on}}{L}\sum_{n=1}^{L} F_n(t) - k_{on}F_0(t) \tag{4}$$

Also, there are additional constraints in the system, which can be written as

$$F_{m_1}(t = 0) = \delta(t) \qquad F_{m_2}(t) = 0 \tag{5}$$

The physical meaning of these expressions is the following. If the protein molecule starts at $t = 0$ at the target site m_1, the search process is successfully finished immediately. But if the protein at any time binds to the trap site (m_2), it will never find the target.

To solve eqs 1−5, we reformulate the problem in the language of Laplace transformations, i.e., with $\widetilde{F}_n(s) = \int_0^\infty e^{-st}F_n(t)\, dt$.[16] Then the set of backward master equations can be transformed into a set of simpler algebraic expressions

$$(s + 2u + k_{off})\widetilde{F}_n(s) = u[\widetilde{F_{n+1}}(s) + \widetilde{F_{n-1}}(s)] + k_{off}\widetilde{F}_0(s) \tag{6}$$

for $2 \leq n \leq L - 1$ and $n \neq m_1$ or m_2, and

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

$$(s + u + k_{off})\widetilde{F}_L(s) = u\widetilde{F_{L-1}}(s) + k_{off}\widetilde{F}_0(s) \tag{8}$$

$$(s + k_{on})\widetilde{F}_0(s) = \frac{k_{on}}{L}\sum_{n=1}^{L}\widetilde{F}_n(s) \tag{9}$$

$$\widetilde{F}_{m_1}(s) = 1 \qquad \widetilde{F}_{m_2}(s) = 0 \tag{10}$$

The solutions of these equations can be found by assuming a general form of the solution as $\widetilde{F}_n(s) = Ay^n + B$.[16] This yields

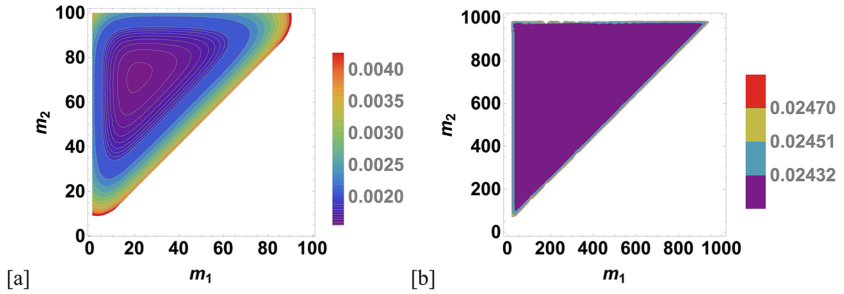


Figure 2. Contour maps for the mean search times to reach the specific binding site as a function of the positions of the target m_1 and trap m_2. Parameters used for calculations are $k_{on} = u = 10^5$ s^{-1} and $k_{off} = 10^3$ s^{-1}. The length of the DNA chain is (a) $L = 100$ and (b) $L = 1000$.

$$\widetilde{F}_0(s) = \frac{k_{on}(k_{off} + s)S_1(s)}{Ls(s + k_{on} + k_{off}) + k_{on}k_{off}S_2(s)} \tag{11}$$

with

$$S_1(s) = \frac{(1 + y)(1 - y^{m_1+m_2-1})}{(1 - y)(1 + y^{2m_1-1})(1 + y^{m_2-m_1})} \tag{12}$$

$$S_2(s) = \frac{(1 + y)[2(1 - y^{2L+m_1-m_2}) + (1 - y^{m_2-m_1})(y^{2m_1-1} + y^{1+2(L-m_2)})]}{(1 - y)(1 + y^{2m_1-1})(1 + y^{1+2(L-m_2)})(1 + y^{m_2-m_1})} \tag{13}$$

and

$$y(s) = \frac{s + 2u + k_{off} - \sqrt{(s + 2u + k_{off})^2 - 4u^2}}{2u} \tag{14}$$

Explicit analytical expressions for the first-passage probability functions in the Laplace form provide us with a complete description for all dynamic properties in the protein search. More specifically, a function Π [with $y(s = 0) \equiv y$]

$$\Pi = \widetilde{F}_0(s = 0) = \frac{S_1(0)}{S_2(0)} = \frac{(1 - y^{m_1+m_2-1})[1 + y^{1+2(L-m_2)}]}{2(1 - y^{2L+m_1-m_2}) + (1 - y^{m_2-m_1})[y^{2m_1-1} + y^{1+2(L-m_2)}]} \tag{15}$$

is the overall probability (at all times) for the protein molecule to reach the target starting from the bulk solution. It is generally less than 1 because of the possible falling into the trap. For a symmetric distribution of the target and the trap with respect to the middle of the DNA chain, when $m_2 = L - m_1+1$, it follows from eq 15 that the probability to reach the specific site is always $\Pi = 1/2$.

A mean first-passage time to reach the target, T_0, which we also identify as the average search time, is given by

$$T_0 = -\frac{\frac{\partial \widetilde{F}_0(s)}{\partial s}|_{s=0}}{\Pi} \tag{16}$$

It is important to note that this is a *conditional* mean first-passage time, which means that the average is taken *only* over the successful trajectories that lead to the target. The search trajectories that end up in the trap are ignored for the calculation of the mean search times. The explicit expression for T_0 can be written as

$$T_0 = \frac{Lk_{off} + k_{on}(L - S_2(0))}{k_{on}k_{off}S_2(0)} + \Pi\frac{\mathrm{d}}{\mathrm{d}s}\left[\frac{S_2(s)}{S_1(s)}\right]|_{s=0} \tag{17}$$

It can be shown that the first term on the right side of eq 17 corresponds to the search time for the system with two targets (at the positions m_1 and m_2),[34] while the second term corrects this result by accounting for the fact that the site at m_2 is the irreversible trap. From this point of view, the search time can be presented as

$$T_0(\text{target/trap}) = T_0(2\text{ targets}) + \Pi\frac{\mathrm{d}}{\mathrm{d}s}\left[\frac{S_2(s)}{S_1(s)}\right]|_{s=0} \tag{18}$$

The reason that our system with the target and the trap is related to the search on the DNA chain with 2 targets is due to the fact that targets and traps are special positions on DNA that guide the dynamics. In the system with two targets, there are two probability fluxes for going from the solution to these special sites, and the smallest time (or the largest flux) determines the overall search time. In our model with the target and trap we also have two probability fluxes to the special sites, but only one of them, to the target, defines the search time.

■ RESULTS AND DISCUSSION

Spatial Distribution of Targets and Traps. The first question we would like to address is the effect of the spatial distributions of targets and traps in the protein search dynamics. The average times to reach the target as a function of the positions m_1 and m_2 for different DNA lengths L are presented in Figure 2. One can see that there are optimal positions for the target and for the trap (Figure 2a), namely, $m_1 = L/4$ and $m_2 = 3L/4$, for which the mean first-passage times are minimal. These are exactly the same optimal positions for the system with two targets.[34] The following arguments can be used to explain this effect. Because at these conditions the search is taking place mostly through 1D diffusion, if the starting position of the protein on DNA is the site $n > (m_1 + m_2)/2$, on average, it will not reach the specific binding site. This means that the problem with the target at the site m_1 and the trap at the site m_2 is identical to the search on DNA of the length $(m_1 + m_2)/2$ with only 1 target at m_1. In this case, the most optimal position for the target is in the middle of the DNA segment,[16] i.e., $m_1 = (m_1 + m_2)/4$. This leads to the relation $m_2 = 3m_1$. Now, these optimal positions must also be symmetric with respect to the middle of the chain because the exchange of the locations of the target and the trap should not affect the outcome. This yields $m_1 + m_2 = L$. It can be easily shown then that putting the target and the trap to sites $m_1 = L/4$ and $m_2 = 3L/4$ satisfies these requirements.

However, the most optimal distribution is not observed for all conditions. Increasing the length of DNA (see Figure 2b) completely changes the picture. Now, any position of the target and the trap along the DNA chain, as long as they are not at the boundaries ($m_1 \neq 1$ and $m_2 \neq L$), leads to the same search times. This is taking place because for $1 < \lambda < L$ (where the protein sliding length is given by $\lambda = \sqrt{u/k_{\text{off}}}$) the search follows the sliding regime: the protein molecule scans the length λ on the DNA before dissociating, and it repeats this searching cycle many times (L/λ on average) before reaching the target. After each dissociation, the protein does not have any memory of what part of DNA it just scanned. With equal probability it can bind to any site on DNA. As a result, the absolute positions of the target and traps are not important anymore for the search optimization.

Dynamic Phase Diagram. In the next step, we investigate how the presence of the semispecific sites influences the mechanisms of the search in different regimes. The results are presented in Figure 3. First of all, the general features of the dynamic phase diagram do not change with the addition of the trap sites. There are still 3 search regimes depending on the relative values of the scanning length $\lambda = \sqrt{u/k_{\text{off}}}$, the DNA length L, and the target size, which is taken to be equal to 1.[16] When $\lambda > L$ we have the phase in which the protein molecule binds to DNA and moves along the chain until it encounters the specific binding site. This is the 1D random-walk search regime with the expected quadratic scaling on the search times as a function of the DNA length L.[16] For $1 < \lambda < L$ the system is in the sliding regime, where the protein molecule binds nonspecifically to DNA, scans a distance $\sim \lambda$, and dissociates, and this search cycle, on average, is repeated several times until the target is found. This search mechanism can be viewed as a combination of 3D and 1D motions. In this phase, the scaling of the search times is linear with L because the number of search cycles is proportional to the DNA length, i.e., $T_0 \sim L/\lambda$.[16] For even smaller scanning lengths, $\lambda < 1$, the search becomes purely 3D because the protein bound to DNA cannot slide along the chain. Again, the linear scaling of the search times with L is observed because, on average, the protein has to visit $L - 1$ sites before it associates to the target.[16]

One can also see from Figure 3 that adding the trap decreases the search times for the system that originally had only a single target. However, it comes with a price of lowering the probability to reach the target: see Figure 4. For the most optimal positions of the targets and traps this probability is always equal to 1/2 (Figures 3a and 4). This can be easily explained if we notice that the most optimal distribution is symmetric ($m_1 = L/4$ and $m_2 = 3L/4$), which means that exactly half of the trajectories are successful and another half end up in the trap. For other spatial distributions the probability to reach the target depends on the search regime. In the random-walk dynamic phase ($\lambda > L$), this can be explained by purely geometric factors because of the one-dimensional nature of the search process. As we already discussed above, if the protein starts the search at the position $n > (m_2 + m_1)/2$ (see Figure 1), then on average it will be trapped since the distance to the semispecific site is shorter. So the probability to reach the target in this regimes can be estimated as

$$\Pi = \frac{m_1 + m_2}{2L} \tag{19}$$

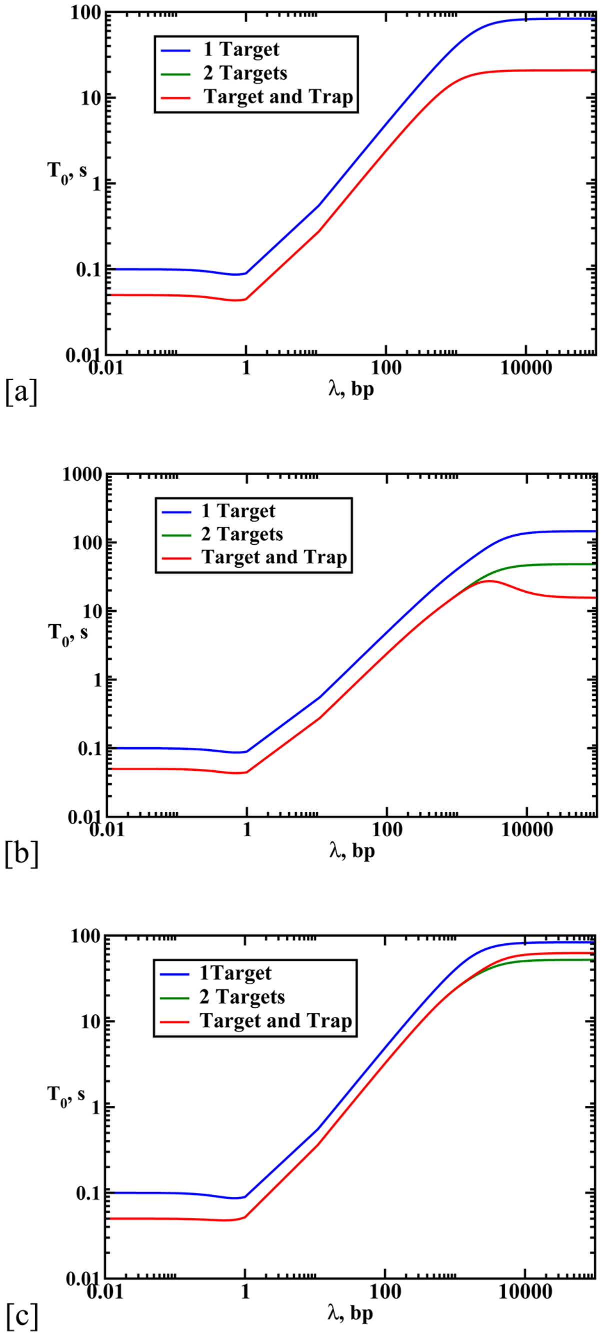


Figure 3. Dynamic phase diagrams for the protein search on DNA with one target at position m, with two targets at positions m_1 and m_2 and with the target and the trap at positions m_1 and m_2. Parameters used for calculations are $k_{\text{on}} = u = 10^5\ \text{s}^{-1}$ and $L = 10\,000$: (a) $m = L/2$, $m_1 = L/4$, and $m_2 = 3L/4$; (b) $m = L/4$, $m_1 = L/4$, and $m_2 = L/2$; and (c) $m = L/2$, $m_1 = L/2$, and $m_2 = L$.

For the most optimal positions $m_1 = L/4$ and $m_2 = 3L/4$ this yields $\Pi = 1/2$, while for $m_1 = L/4$ and $m_2 = L/2$ this gives $\Pi = 3/8$, and for $m_1 = L/2$ and $m_2 = L$ we obtain $\Pi = 3/4$. These calculations fully agree with the results presented in Figure 4 in the limit of very large scanning lengths λ. For very small $\lambda < 1$, when the search follows the 3D mechanism, the spatial positions of the target and trap are not important. In this case, exactly half of the trajectories will be successful, yielding $\Pi = 1/2$. For the intermediate sliding regime ($1 < \lambda < L$), the probability to reach the target obviously has the lower and

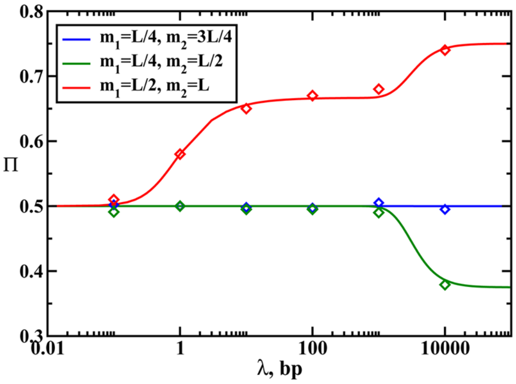


Figure 4. Probability to reach the target as a function of the scanning length for different distributions of the target and trap sites. Parameters used for calculations are $k_{on} = u = 10^5$ s^{-1}, $L = 10\,000$, and k_{off} is changing. Symbols are from Monte Carlo computer simulations.

upper bounds between 1/2 and $\frac{m_1 + m_2}{2L}$ (which could be larger or smaller than 1/2), and the explicit values of Π depend on the relative contribution of 1D and 3D fluxes into the target site.

It is also important to compare the protein search dynamics on DNA with two targets with the search in the system that has only one target and one trap at the same positions, as shown in Figure 3. In the jumping and sliding regimes ($\lambda < L$), there is no difference in the search times between two systems because the location of the special sites does not influence the search mechanisms. The protein mostly reaches the specific binding site from the bulk solution. For the system with two targets, one-half of search trajectories will go to one of the targets, and the second half will finish at the second specific site. The symmetry requires that both of these sets of trajectories have the same mean times because the targets are indistinguishable. For the protein search on DNA with the target and the trap the dynamics is similar: half of all trajectories will end up at the trap, and they will not be counted. But another half of trajectories that reach the specific binding site have the same mean search time as in the case of two targets on DNA.

The situation is different in the random-walk phase ($\lambda > L$). Here, the search times for the target and trap system could be the same as for the two targets system (Figure 3a). This is the case for the symmetric locations of the special sites. The presence of the trap could also slow down the search (Figure 3c), or surprisingly, it could even accelerate the search (Figure 3b). It is interesting to note that, in the target and trap system where the search is faster, the probability to find the target is lower (see Figure 4). Thus, this effect can be also explained by the geometric arguments, as discussed above. The traps effectively remove trajectories with longer search times.

How Traps Accelerate the Search. We have already shown that the addition of the semispecific site strongly affects the protein search dynamics for specific binding sites. To quantify this effect we introduce two auxiliary functions, r_1 and r_2, which are defined as the ratio of the search times on DNA with one or two targets and for the system with the target and trap for the fixed values of the positions m_1 and m_2, respectively

$$r_1 = \frac{T_0(1\text{ target})}{T_0(\text{target/trap})} \qquad r_2 = \frac{T_0(2\text{ targets})}{T_0(\text{target/trap})} \tag{20}$$

These acceleration functions for various sets of parameters are presented in Figures 5 and 6.

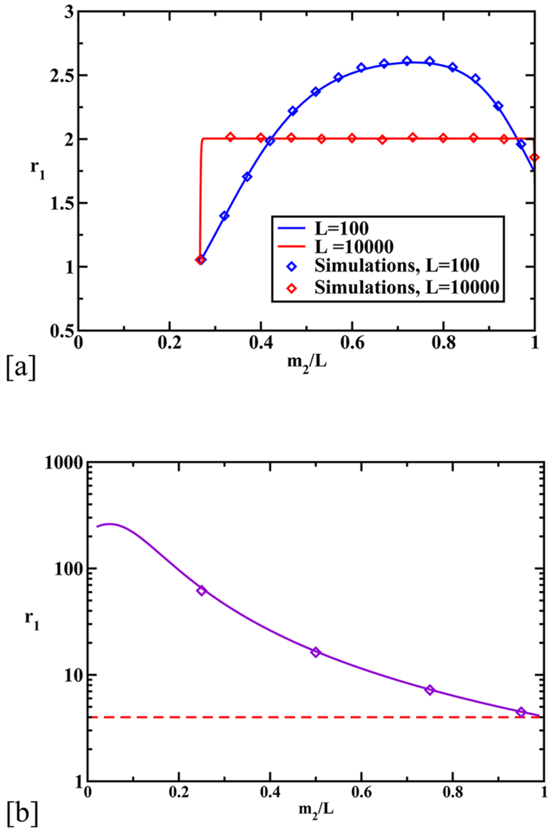


Figure 5. Acceleration parameter r_1 as a function of the trap location for the fixed position of the target. Parameters used for calculations are $k_{on} = u = 10^5$ s^{-1}. (a) The target is at $m_1 = L/4$ and $k_{off} = 10^3$ s^{-1}. The target is at $m = L/2$ for the reference single target system. (b) The target is at $m_1 = 1$, $L = 100$, and $k_{off} = 0.1$ s^{-1}. The target is at $m = 1$ for the reference single target system. Dashed line corresponds to $r_1 = 4$. Symbols are from Monte Carlo computer simulations.

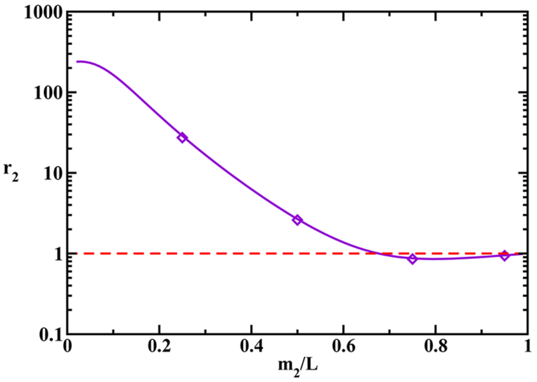


Figure 6. Acceleration function r_2 as a function of the position of the trap for the fixed position of the target. Parameters used for calculations are $L = 100$, $m_1 = 1$, $k_{on} = u = 10^5$ s^{-1}, and $k_{off} = 0.1$ s^{-1}. Dashed line corresponds to $r_2 = 1$. Symbols are from Monte Carlo computer simulations.

First, we analyze how the traps influence the search on DNA with only one target as presented in Figure 5. One can see that it is usually faster to find the target if there is a semispecific site in the system. For not very long DNA, there is an optimal position of the trap that provides the shortest search times (Figure 5a,b). The acceleration can even reach very high values if the target is far away from the middle of the chain and the trap is close to the target: see Figure 5b. These observations can be explained using the following arguments. For this set of parameters, the protein reaches the specific binding site mostly via 1D sliding along the DNA chain ($L \sim \lambda$). Introducing the trap site into the system has two opposite effects: it removes many long-time trajectories from the search, lowering the mean search time. However, it also decreases the flux into the target site from one side of DNA which makes the search longer. Balancing these two effects leads to the optimal position of the trap. These arguments also suggest that the maximal acceleration can be achieved if the target sits at the boundary, i.e., $m_1 = 1$ (see Figure 5b). In the limiting case of $m_1 = 1$ and $m_2 = L$ the acceleration is equal to 4 (Figure 5b). This is because this system of the target and trap on DNA with the length L can be mapped into the search on DNA of the length $L/2$ with the target and without the trap. Since the 1D search in the random-walk regime has a quadratic scaling of the search times ($T_0 \sim L^2$), the search time acceleration becomes $r_1 \simeq \frac{L^2}{(L/2)^2} = 4$.

Increasing the length of DNA L shifts the system to the sliding search regime, and the maximal acceleration in the search is equal to 2 (Figure 5a). The spatial positions of the target and traps are not important because of 3D + 1D search mechanisms, as explained above. In this regime, the trap effectively removes half of all search trajectories, which is equivalent to lowering the number of search cycles also in 2 times. Because of the linear scaling for the search times in this dynamic phase, this yields $r_1 = 2$.

The comparison of the protein search dynamics on DNA that has one target and one trap with the system with two targets is shown in Figure 6. Here the effect of the trap sites is more complex. For relatively short distances between the target and the trap, the search is much faster than that for the case of two targets. This result is unexpected, but it can be explained using the geometric arguments. The trap effectively removes a significant part of DNA from the search, and because for this set of parameters (Figure 6) the search is mostly one-dimensional, this leads to large accelerations. One could also think about two closely located targets as one "effective" new target, and all our arguments why it is faster to search by adding the trap to the system with one target can be applied now. However, moving the trap site away from the target lowers this effect and starting from some distance the search in the system with two targets is faster because the number of specific sites is larger. One can see in Figure 6 that for $m_2/L > 0.67$ the acceleration parameters goes below the unity. Thus, our calculations clearly show that the spatial distribution of the targets and traps controls the search dynamics.

SUMMARY AND CONCLUSIONS

We presented a theoretical investigation on the role of semispecific binding sites in the protein search for targets on DNA. Our approach is based on the discrete-state stochastic method that connects the search process with the first-passage events. The advantage of this approach is that it provides a full analytical description for all dynamic properties in the system. We determined that the protein search dynamics is governed by several important length scales such as the DNA length, the average sliding length of the protein along the DNA chain, the distance between the targets and traps, and the distance to the DNA ends from the specific and semispecific sites. It was found that there is the optimal spatial distribution of the target and traps that for short DNA leads to the smallest search time, while for long DNA the search is not affected by exact positions of specific and semispecific binding sites. This was explained by exploring the dynamic phase diagram which shows three different regimes for the protein search depending on the relative values of the relevant lengths scales in the system. We also analyzed the probability of reaching the target, and it was found that it varies for different dynamic search regimes. Furthermore, we investigated the acceleration in the search due to the presence of the trap sites. Adding the semispecific site in most cases decreases the search time for the system with only one target. For the system with two targets on DNA the substitution of one them with the trap leads to more complex behavior. For a short distance between the special sites the search is accelerated, while for large distances the search becomes slower. These phenomena are explained by noting that the significant fraction of the search trajectories is removed from the search due to falling into the trap.

Our theoretical approach provides a simple and clear picture of the complex biological processes during the protein search for the specific binding sites on DNA. At the same time, it should be noted that the presented theoretical method is not exact, and it involves several approximations. The conformational freedom of DNA chains and the intersegment transfer processes are neglected. We also assume that the protein moves faster in the bulk solution than on DNA. In the real biological cells all these assumptions probably are not valid, but it is not clear how this would affect the overall search dynamics. But the most serious issue in our work is the assumption of the trap irreversibility. In reality, the protein molecules cannot be absorbed by these traps for an infinite amount time, and they will be eventually released. In addition, some of these traps are not very strong. It will be critically important to address these issues in more advanced theoretical and experimental studies.

AUTHOR INFORMATION

Corresponding Author

*E-mail: tolya@rice.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was supported by the Welch Foundation (Grant C-1559), by the NSF (Grant CHE-1360979), and by the Center for Theoretical Biological Physics sponsored by the NSF (Grant PHY-1427654). We also would like to thank Prof. J. Iwahara for stimulating discussions and for introducing us to this problem.

REFERENCES

(1) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*, 5th ed.; Garland Science: New York, 2007.

(2) Riggs, A. D.; Bourgeois, S.; Cohn, M. The lac Repressor-Operator Interaction: III. Kinetic Studies. *J. Mol. Biol.* **1970**, *53*, 401−417.

(3) 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.

(4) Berg, O. G.; von Hippel, P. H. Diffusion-Controlled Macromolecular Interactions. *Annu. Rev. Biophys. Biophys. Chem.* **1985**, *14*, 131−160.

(5) Winter, R. B.; Berg, O. G.; von Hippel, P. H. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 3. The

Escherichia Coli *lac* Repressor-Operator Interaction: Kinetic Measurements and Conclusions. *Biochemistry* **1981**, *20*, 6961−6977.

(6) Halford, S. E.; Marko, J. F. How do Site-Specific DNA Binding Proteins Find Their Targets? *Nucl. Acid Res.* **2004**, *32*, 3040−3052.

(7) Gowers, D. M.; Wilson, G. G.; Halford, S. E. Measurement of the Contributions of 1D and 3D Pathways to the Translocation of a Protein along DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (88), 15883−158.

(8) Kolesov, G.; Wunderlich, Z.; Laikova, O. N.; Gelfand, M. S.; Mirny, L. A. How Gene Order Is Influenced by the Biophysics of Transcript Ion Regulation. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 13948−13953.

(9) Wang, Y. M.; Austin, R. H.; Cox, E. C. Single-Molecule Measurements of Repressor Protein 1D Diffusion on DNA. *Phys. Rev. Lett.* **2006**, *97*, 048302.

(10) Elf, J.; Li, G.-W.; Xie, X. S. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science* **2007**, *316*, 1191−1194.

(11) Tafvizi, A.; Huang, F.; Leith, J. S.; Fersht, A. R.; Mirny, L. A.; van Oijen, A. M. Tumor Suppressor p53 Slides on DNA with Low Friction and High Stability. *Biophys. J.* **2008**, *95*, L01−L03.

(12) Benichou, O.; Kafri, Y.; Sheinman, M.; Voituriez, R. Searching Fast for a Target on a DNA without Falling to Traps. *Phys. Rev. Lett.* **2009**, *103*, 138102.

(13) Hu, T.; Grosberg, A. Y.; Shklovskii, B. I. How Proteins Search for Their Specific Sites on DNA: the Role of DNA Conformation. *Biophys. J.* **2006**, *90*, 2731−2744.

(14) Givaty, O.; Levy, Y. Protein Sliding along DNA: Dynamics and Structural Characterization. *J. Mol. Biol.* **2009**, *385*, 1087−1097.

(15) Kolomeisky, A. B.; Veksler, A. How to Accelerate Protein Search on DNA: Location and Dissociation. *J. Chem. Phys.* **2012**, *136*, 125101.

(16) 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.

(17) Mechetin, G. V.; Zharkov, D. O. Mechanisms of Diffusional Search for Specific Targets by DNA-Dependent Proteins. *Biochemistry (Moscow)* **2014**, *79*, 496−505.

(18) Hilario, J.; Kowalczykowski, S. C. Visualizing Protein-DNA Interactions at the Single-Molecule Level. *Curr. Opin. Chem. Biol.* **2010**, *14*, 15−22.

(19) Afek, A.; Schipper, J. L.; Horton, J.; Gordan, R.; Lukatsky, D. V. Protein-DNA Binding in the Absence of Specific Base-Pair Recognition. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 17140−17145.

(20) Hammar, P.; Leroy, P.; Mahmutovic, A.; Marklund, E. G.; Berg, O. G.; Elf, J. The lac Repressor Displays Facilitated Diffusion in Living Cells. *Science* **2012**, *336*, 1595−1598.

(21) Koslover, E. F.; Diaz de la Rosa, M. A.; Spakowitz, A. J. Theoretical and Computational Modeling of Target-Site Search Kinetics in Vitro and In Vivo. *Biophys. J.* **2011**, *101*, 856−865.

(22) Marklund, E. G.; Mahmutovic, A.; Berg, O. G.; Hammar, P.; van der Spoel, D.; Fange, D.; Elf, J. Transcription-Factor Binding and Sliding on DNA Studied Using Micro- and Macroscopic Models. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 19796−19801.

(23) Brackley, C. A.; Cates, M. E.; Marenduzzo, D. Facilitated Diffusion on Mobile DNA: Configurational Traps and Sequence Heterogeneity. *Phys. Rev. Lett.* **2012**, *109*, 168103.

(24) Sokolov, I. M.; Metzler, R.; Pant, K.; Williams, M. C. Target Search of N Sliding proteins on a DNA. *Biophys. J.* **2005**, *89*, 895−902.

(25) Shimamoto, N. One-dimensional Diffusion of Proteins along DNA. *J. Biol. Chem.* **1999**, *274*, 15293−15296.

(26) Gorman, J.; Greene, E. C. Visualization One Dimensional Diffusion of Proteins along DNA. *Nat. Struct. Mol. Biol.* **2008**, *15*, 768−774.

(27) Mirny, L. A.; Slutsky, M.; Wunderlich, Z.; Tafvizi, A.; Leith, J. S.; Kosmrlj, A. How a Protein Searches for its Site on DNA: The Mechanism of Facilitated Diffusion. *J. Phys. A: Math. Theor.* **2009**, *42*, 434013.

(28) Kolomeisky, A. B. Physics of Protein-DNA Interactions: Mechanisms of Facilitated Target Search. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2088−2095.

(29) Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. The Role of DNA Shape in Protein-DNA Recognition. *Nature* **2009**, *461*, 1248−1254.

(30) Marcovitz, A.; Levy, Y. Frustration in Protein-DNA Binding Influences Conformational Switching and Target Search Kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17957−17962.

(31) Dahirel, V.; Paillusson, F.; Jardat, M.; Barbi, M.; Victor, J.-M. Nonspecific DNA-Protein Interaction: Why Proteins Can Diffuse Along DNA. *Phys. Rev. Lett.* **2009**, *102*, 228101.

(32) Bauer, M.; Rasmussen, E. S.; Lomholt, M. A.; Metzler, R. Real Sequence Effects on the Search Dynamics of Transcription Factors on DNA. *Sci. Rep.* **2015**, *5*, 10072.

(33) Townson, S. A.; Samuelson, J. C.; Bao, Y.; Xu, S.-Y.; Aggarwal, A. K. BstYI Bound to Noncognate DNA Reveals a "Hemispecific" Complex: Implications for DNA Scanning. *Structure* **2007**, *15*, 449−459.

(34) Lange, M.; Kochugaeva, M.; Kolomeisky, A. B. Protein Search for Multiple Targets on DNA. *J. Chem. Phys.* **2015**, *143*, 105102.

(35) Esadze, A.; Kemme, C. A.; Kolomeisky, A. B.; Iwahara, J. Positive and Negative Impacts of Nonspecific Sites During Target Location by a Sequence-Specific DNA-Binding Protein: Original of the Optimal Search at Physiological Ionic Strength. *Nucleic Acids Res.* **2014**, *42*, 7039−7046.